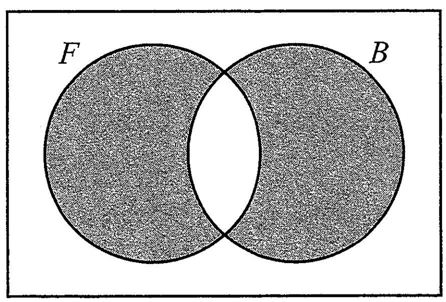
Sample Examination One – Answers

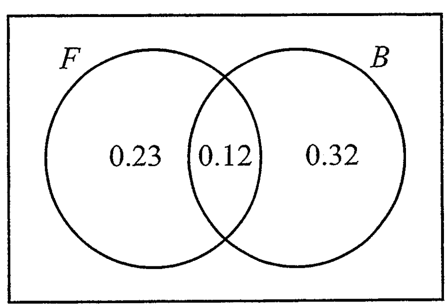
1. **(C)** For the sedans, the distance from the median to the maximum is greater than the distance from the minimum to the median. So the distribution for the sedans is positively skewed. For the SUVs the distance from the minimum to the median is greater than the distance from the median to the maximum. So the distribution for the SUVs is negatively skewed. The range for the sedans is approximately $275 - 135 = 140$, and the range for the SUVs is approximately $380 - 160 = 220$. Thus the range is less for the sedans than it is for the SUVs.

2. **(E)** Since the treatments (organic or non-organic feed) are imposed on randomly assigned groups of cows, rather than merely observing the health characteristics of cows who are already being given one of the two types of feed, this is an experiment. The experiment is designed to study the effect of the type of feed on the health of the cows, and so the type of feed is the explanatory variable, and the level of health is the response variable.

3. **(B)** For a randomly chosen male student, let F be the event that the student plays football and B be the event that the student plays basketball. We are given that $P(F) = 0.35$, $P(B) = 0.44$, and $P(F \cap B) = 0.12$. We require the probability associated with the shaded region in the Venn diagram below.



The probability that a randomly chosen male plays football but not basketball is $0.35 - 0.12 = 0.23$. The probability that a randomly chosen male plays basketball but not football is $0.44 - 0.12 = 0.32$. These probabilities are shown in the diagram below.



Therefore, the required probability is $0.23 + 0.32 = 0.55$.

4. **(C)** The confidence interval for a population proportion is given by $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. We can assume that the values of $\hat{p}$ for the two samples will be roughly the same. The z^* value for a 90% confidence interval is 1.645 and the z^* value for a 95% confidence interval is 1.960. (The z^* value for a 90% confidence interval is given by invNorm(0.95) or invNorm(0.95,0,1) on the TI-83 and TI-84 calculators, and invNorm(0.95,0,1) on the TI-Nspire. For the 95% confidence interval, replace 0.95 with 0.975. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) Thus, since the z^* value is smaller for the 90% confidence intervals, the 90% confidence intervals will be narrower than the 95% intervals. Looking again at the formula for the confidence interval we note that, since the n is in the denominator of the fraction, the larger the value of n, the narrower the confidence interval will tend to be. Thus, the narrowest of the four confidence intervals will be the 90% interval for the larger sample.

   Alternatively, for a given sample, the 95% confidence interval for the population proportion will be wider than the 90% confidence interval in order to increase the confidence that the true population proportion is captured. Also, for a given confidence level, larger samples, by giving us more information, allow us to be more specific about the likely location of the true population proportion. And so, here, the narrowest interval is the one for the smaller confidence level and the larger sample.

5. **(B)** Using $b = r\frac{s_y}{s_x}$, $b = (-0.675)\left(\frac{0.469}{0.383}\right) = -0.827$.

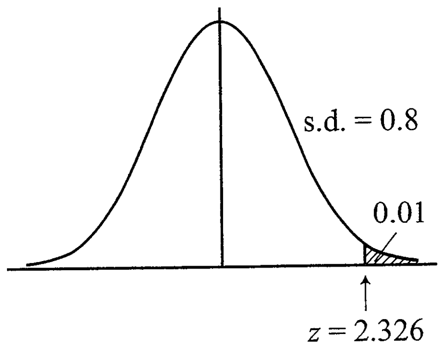
6. **(E)** Since the counts for multiple categories (canola oil, olive oil, sunflower oil, and some other sort of oil) are to be compared over three samples, the chi-square test for homogeneity is a suitable test to use. Note that the chi-square test for goodness of fit is used when the counts from a *single* sample are to be compared to a hypothesized set of proportions, and therefore is not suitable here.

7. **(C)** The results given here are an example of *categorical* data, and so a pie chart could be used. Stemplots, histograms, boxplots, and scatterplots are used for displaying *quantitative* data (with scatterplots being used for bivariate data sets).

8. **(E)** The expected number of males in the sample is $800(0.68) = 544$ and the expected number of females is $800(0.32) = 256$. Amongst the 544 males the expected number of employed people is $544(0.9) = 489.6$. (Note that this does not have to be a whole number, since it is an expected value.) Amongst the 256 females the expected number of employed people is $256(0.76) = 194.56$. Thus, the expected number of employed people in the sample is $489.6 + 194.56 = 684.16$.

9. **(A)** The means of two independent random samples are being compared, and so a two-sample t-test is used. Furthermore, the researcher wishes to establish whether there is a *difference* between the mean times for men and women, and so this is a two-tailed test. Thus, the hypotheses given in (A) are correct. (Note that the hypotheses given in (C) and (D) are for a paired t-test, and so are incorrect here.)

10. **(B)** The z-value for the length of the smallest eligible pod is 2.326. (This value is given by invNorm(0.99) or invNorm(0.99,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.99,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd, DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) The relevant normal distribution is shown in the graph below.



This tells us that the required length is 2.326 standard deviations above the mean. Therefore, since the standard deviation is 0.8 inches, the required length is $2.326(0.8) = 1.86$ inches above the mean.

11. **(E)** Larry's grade in English is at the first (lower) quartile. His grade in History is between the first quartile and the median. His grade in Math is at the median. His grade in Chemistry is between the median and the third (upper) quartile. Therefore, the required list of subjects is Chemistry, Math, History, English.

12. **(B)** If, assuming that the null hypothesis is true, a sample result such as the one obtained is very unlikely, then we have convincing evidence that the null hypothesis is not true. In other words, we have convincing evidence that the alternative hypothesis is true. Option (A) is incorrect because we can never have convincing evidence that the parameter takes *exactly* the value given in the null hypothesis. We could become increasingly convinced that the value of the parameter is *close* to the value hypothesized, but not that the value of the parameter is *equal* to the value hypothesized. Option (C) is incorrect because the results cannot *prove* that something is the case; they can only *provide evidence* of something. (Also, providing evidence that the alternative hypothesis is false would be providing evidence that the null hypothesis is true, which, as explained above, is not possible.) Option (D) is incorrect because the null hypothesis could, by chance, be rejected when it is true. (This occurrence is known as a Type I error.) Option (E) is incorrect for two reasons: first, because we can fail to reject the null hypothesis even when it is false, this being called a Type II error; and, second, because it is not possible even to *find evidence* that the null hypothesis is true (as explained above).

13. **(D)** The distribution of the number of left-handed people in the sample is binomial with $n = 10$ and $p = 0.13$. Thus, the probability that exactly two people in the sample are left-handed is $\binom{10}{2}(0.13)^2(0.87)^8$.

14. **(E)** The method given in option (E) divides the population into strata according to the type of goods primarily ordered, and then randomly selects 1000 customers from each stratum. This is stratified random sampling. Option (A) describes a convenience sample, option (B) describes a simple random sample, option (C) describes a systematic sample, and option (D) describes a cluster sample.

15. **(A)** The classroom activity estimates the standard deviation of the sample proportion, $\hat{p}$, when $n = 10$ and $p = 300/500 = 0.6$. This standard deviation is given by

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.6)(0.4)}{10}}.$$

16. **(E)** Since the population standard deviation is unknown, the sample standard deviation, s, must be used when calculating the confidence interval. When this is done, the t-distribution must be used in place of the standard normal (z) distribution. Thus the required confidence interval is given by $\overline{x} \pm t^* \cdot \frac{s}{\sqrt{8}}$. Also, in order to use the t-distribution, since the sample is small, we must assume that the population is normally distributed.

17. **(A)** The idea behind an experiment such as this is that the groups (here the treatment group and the placebo group) should be as alike as possible in all respects except that the people in the first group receive the drug and the people in the second group do not. That way, any significant difference between the responses of the two groups can be attributed to the drug, since this is (as much as possible) the only difference between the two groups. Thus, here, the second group receives the "placebo" treatment so that the people in *both* groups experience the psychological effect of taking tablets, and so, therefore, that the two groups are comparable in all respects apart from the consumption or non-consumption of the drug.

18. **(B)** The number of runners whose times were less than or equal to 20 minutes was $80(0.2625) = 21$ and the number of runners whose times were less than or equal to 22 minutes was $80(0.5125) = 41$. Therefore, the number of runners whose times were more than 20 minutes and at most 22 minutes was $41 - 21 = 20$.

19. **(B)** The first (lower) quartile is the time below which 25% of the runners achieved. Thus the first quartile is between 18 and 20 minutes. The third (upper) quartile is the time below which 75% of the runners achieved. Thus the third quartile is between 24 and 26 minutes. According to this information, the smallest possible value of the interquartile range is $24 - 20 = 4$ minutes, and the largest possible value of the interquartile range is $26 - 18 = 8$ minutes. The only one of the five options that is between 4 and 8 minutes is option (B).

20. **(E)** The statement in (A) is correct, since the p-value for the test is given to be 0.127, which is greater than 0.05. The statement in (B) is correct, since a two-sided t-test with 90 degrees of freedom is being performed; when the value of the t-statistic is positive, failure to reject H_0 is equivalent to this value being less than the positive critical value for a single-tail probability of 0.025. The statement in (C) is correct, since zero being in the confidence interval is equivalent to non-rejection of H_0 at the 0.05 level. The statement in (D) is a correct interpretation of the p-value: the probability of getting a test statistic at least as extreme as the one obtained, given that H_0 is true. The attempted reversal in (E) of this statement, however, is incorrect: no probability can be attached to the population means.

21. **(B)** Statistic 1 and Statistic 3 have similar amounts of variability. However, the center of the distribution of Statistic 1 seems to be at, or close to, 5, the true value of the parameter, and this cannot be said for Statistic 3. So Statistic 1 is preferable to Statistic 3. The center of the distribution of Statistic 2 also seems to be at, or close to, 5, but the variability of Statistic 2 seems to be less than that of Statistic 1. Therefore, Statistic 2 is preferable to Statistic 1, and the required list is 2, 1, 3.

22. **(D)** The power of a hypothesis test is the probability of correctly rejecting H_0 given that H_0 is false, for a specific alternative value of the parameter. The relevant parameter in this question is the population mean, μ, and we are told to assume that the true value of μ is greater than the hypothesized value, 8. Thus, for this value of μ, the power of the test is the probability that H_0 will be (correctly) rejected.

The test statistic here will be either z or t. Changing the significance level to 0.1 will increase the size of the rejection region at the right end of the distribution of the test statistic, thus increasing the probability that H_0 will be rejected; so the power of the test is increased. (Another way to think of this is that changing the significance level has opposite effects on the probabilities of Type I and Type II errors, and the probability of a Type I error is exactly the significance level of the test. Thus, changing the significance level from 0.05 to 0.1 increases the probability of a Type I error and decreases the probability of a Type II error. Decreasing the probability of a Type II error is increasing the power of the test.) This change has resulted in a test that has greater power.

If the test is changed to the two-tailed version, the 5% critical region will be apportioned equally between the left and right extremes of the distribution of the test statistic. Thus, the size of the critical region on the right will be reduced, making it less likely that H_0 will be rejected. Therefore, this change has resulted in a test that has smaller power.

Increasing the sample size from 50 to 100 means that more information is being provided, and therefore it is more likely that the hypothesis test will reach the correct conclusion (that μ is greater than 8). Thus the power of the test has increased.

(A more precise way of explaining why an increase in the sample size will increase the power of the test is as follows. We'll assume that a t-test is being performed. (A very similar argument will apply in the case of a z-test.) The test statistic is given by

$$t = \frac{\overline{X} - 8}{s/\sqrt{n}}.$$

For the sake of argument, let's assume that the true population mean is 8.5. For either sample size the value of $\overline{X}$ is then likely to be around 8.5. However, the value of $s/\sqrt{n}$ is almost certain to be smaller when $n = 100$ than when $n = 50$. Consequently, the value of t is almost certain to be larger for the larger sample size. Therefore, for the larger sample size there is a greater probability of H_0 being rejected, and thus a greater power.)

23. **(D)** In systematic sampling, the members of the population are numbered, and then, for example, every 100th item in the population will be included in the sample. This method can often provide a sample that represents the population well. Meanwhile, virtually any sampling procedure that involves randomness, including simple random sampling, has some chance of producing a sample that does *not* represent the population well. Therefore it is not true to say that a simple random sample will always represent the population better than a systematic sample.

24. **(E)** This is a chi-square test for goodness of fit. The observed counts are 192, 133, 118, and 57. The corresponding expected counts are $500(0.4) = 200$, $500(0.3) = 150$, $500(0.18) = 90$, and $500(0.12) = 60$. So the value of the test statistic is

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$
$$= \frac{(192-200)^2}{200} + \frac{(133-150)^2}{150} + \frac{(118-90)^2}{90} + \frac{(57-60)^2}{60} = 11.108.$$

25. **(C)** This process consists of repeated trials, where the outcome of each trial is either "success" (commercial vehicle) or "failure" (non-commercial vehicle). The probability of success on each trial is the same $(1/5)$, and the outcomes of the trials are independent of each other. The trials are repeated until a success is obtained, and the random variable we're interested in is defined to be the number trials up to and including the first success. Thus, the random variable has a geometric distribution.

26. **(D)** This is a two-proportion z-test, and so the test statistic is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

where $\hat{p}_c$ represents the proportion of saucers with faults for the combined samples. The values of $\hat{p}_1$, $\hat{p}_2$, and $\hat{p}_c$ are given by

$$\hat{p}_1 = \frac{17}{80} = 0.2125,$$

$$\hat{p}_2 = \frac{28}{100} = 0.28,$$

$$\text{and } \hat{p}_c = \frac{17+28}{80+100} = \frac{45}{180} = 0.25.$$

The aim is to determine whether the population proportion for the new machine (p_1) is *less* than that for the current machine (p_2), and so this is a one-tailed test, and we are testing the lower tail. Therefore, the p-value for the test is given by

$$P\left(z < \frac{0.2125 - 0.28}{\sqrt{(0.25)(0.75)\left(\frac{1}{80}+\frac{1}{100}\right)}}\right) = P(z < -1.039) = 0.149.$$

(This p-value is obtained on the TI-83 and TI-84 calculators by entering normalcdf(-999,-1.039) or normalcdf(-999,-1.039,0,1) and on the TI-Nspire by entering normalCdf(-999,-1.039,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

Since the p-value is greater than 0.05, we do not have convincing evidence at the 0.05 significance level that the proportion of faulty saucers is less for the new machine than for the current machine, and the new machine will not be incorporated.

27. **(A)** For the point given, $x = 22836$ and $y = 57693$. The residual is given by $y - \hat{y}$ where $\hat{y}$ is the value predicted by the least squares regression line for the number of used cars sold when $x = 22836$. The equation of the least squares regression line is given by the computer output to be $\hat{y} = 32971 + 0.8566x$. Therefore, when $x = 22836$, $\hat{y} = 32971 + 0.8566(22836) = 52532$ and the residual is $y - \hat{y} = 57693 - 52532 = 5161$.

28. **(C)** This is the t-test for the slope of the population regression line, and the p-value for the two-tailed version of the test is given in the column labeled "P" in the computer output. The test we're doing is for the *slope* of the regression line as opposed to the y-intercept, and so we require the p-value associated with the slope, which is the coefficient of "new_cars." Thus the required p-value is 0.092, and there is evidence of a non-zero slope of the population regression line at the 0.1 level, but not at the 0.05 level.

29. **(C)** One of the conditions for an experiment to be double-blind is that the experimental subjects do not know which treatment they are receiving (the other being that the people who measure the response variable don't know which experimental subjects received which treatment). In an experiment to investigate whether taking vitamin C speeds recovery from a cold it is possible for the experimental subjects not to know what treatments they are receiving, since those who are not to receive the vitamin C can be given a placebo. However, in all four of the other options the experimental subjects must know what treatments they are receiving: you know whether or not you're listening to music as you type, you know whether or not you're exercising, you know what quantities of water you're drinking, and you know whether or not you're using keyboard shortcuts.

30. **(C)** This is a chi-square test for independence, and the passengers' responses could be recorded in a table such as the one shown below.

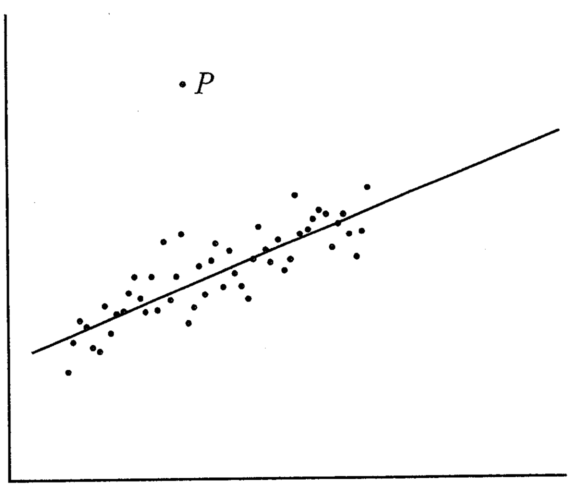
		How Made Payment		
		Ticket Office	On Platform	Online
How Happy	Very Happy			
	Happy			
	Neutral			
	Unhappy			

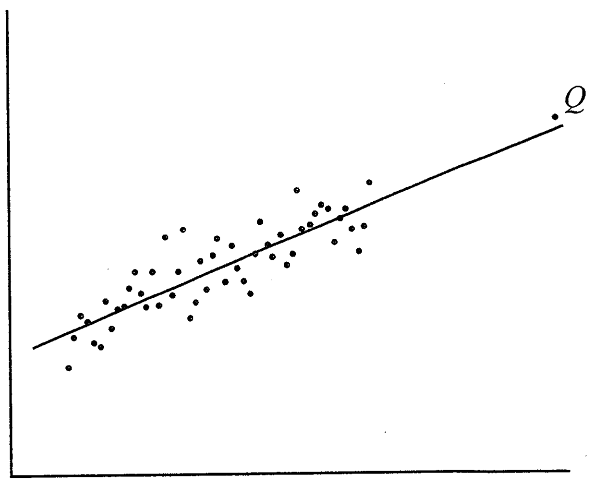
The table has four rows and three columns, and it is known that, for an m by n table, the number of degrees of freedom of the chi-square distribution used in the test is $(m-1)(n-1)$. Thus, here, the number of degrees of freedom is $3 \cdot 2 = 6$.

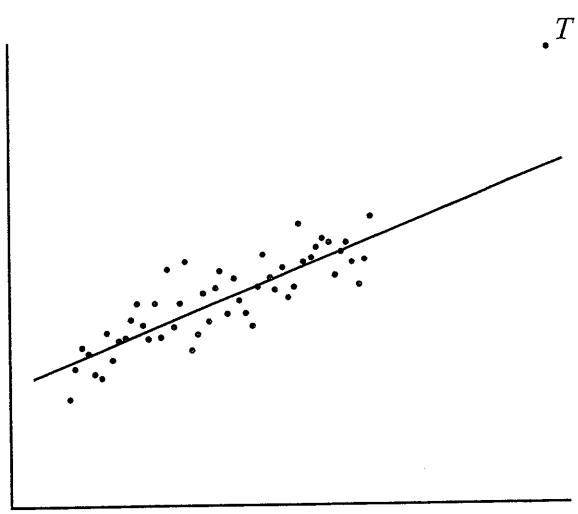
31. **(C)** Increasing a quantity by 5% is the same as multiplying the quantity by 1.05. So if we denote a salary by x, the new salary is given by $x_{\text{new}} = 1.05x$. We know that multiplying all the values in a data set by a constant has the effect that both the mean and the standard deviation of the data set are multiplied by that constant. Thus, here, both the mean and the standard deviation are multiplied by 1.05; that is, they are both increased by 5%. However, since the standard deviation is multiplied 1.05, the variance, being the square of the standard deviation, is multiplied by $(1.05)^2$, and so the variance is not increased by 5%. Turning our attention to the median and the interquartile range, the person whose salary was the median salary before the increase has the median salary after the increase. Since this person's salary is increased by 5%, the median is increased by 5%. Likewise, the people whose salaries were at the upper and lower quartiles of the distribution still have the upper and lower quartile salaries after the increase. Thus the upper and lower quartiles are increased by 5%, and their difference, the interquartile range, is increased by 5%, also.

32. **(C)** Blocking ensures that the sets of experimental units assigned to the various treatments are similar with respect to the blocking variable, which here is age. Let us illustrate this by means of the scenario given in the question. The blocking described could be achieved as follows: Sort the experimental subjects according to their ages. The two oldest people form Block 1, the next two Block 2, and so on. Then, from each block, the two people are randomly assigned to the treatments: one of the people receives the new drug and the other receives the current drug. Thus the set of people receiving the new drug consists of one person from the two oldest, one person from the next two, and so on; and the same applies to the set of people receiving the current drug. Thus, in terms of age, the people who receive the new drug are similar to the people who receive the current drug.

33. **(B)** Although the question is worded in terms of the effect of *removing* points, it is slightly easier to think in terms of *adding* points. In the graphs below, the least squares regression line shown is the correct one for the data set *without* the labeled points, *P*, *Q*, and *T*.







Consider, first, the effect of adding the point P, as shown in the first graph. Remember that the least squares regression line minimizes the sum of the squares of the vertical distances of the points from the line. When P is added, the regression line will move upward in order to reduce the vertical distance of P from the line. However, the line will move only a small amount, for otherwise the combined effect of the large number of increased vertical distances from the other points in the data set would outweigh the effect of the reduced distance from P. Thus, the least-squares regression line will change only a small amount, even though P has a large residual. So statement I is false.

In the second graph, the point Q is close to the least squares regression line, and so the addition of Q to the data set will have an extremely small effect on the line. Thus Q is an example of a point that is an outlier in the x-direction but is not influential, and statement II is false.

In the third graph the point T is influential. (This is the case because T is an outlier in the x-direction and is far from the regression line that applies to the other points in the data set. When T is added, the slope of the least squares regression line increases in order to decrease the vertical distance of T from the line whilst not greatly increasing the sum of the squares of the vertical distances of the other points from the line.) It is clear from the diagram that when T is added, the resulting set of points shows a linear relationship that is less strong than when T is omitted. Thus the removal of T produces a stronger (positive) linear relationship, and so the absolute value of the correlation coefficient has increased. So statement III is true.

34. **(C)** The confidence interval for the population mean when the population standard deviation is known is given by

$$\overline{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}.$$

Using $\overline{x}$ as an estimate of μ, we need the margin of error, with 95% confidence, to be less than 30. Thus, we need

$$z^* \cdot \frac{\sigma}{\sqrt{n}} < 30.$$

Replacing this inequality with an equation, we have

$$z^* \cdot \frac{\sigma}{\sqrt{n}} = 30,$$

and so

$$n = \left(\frac{z^* \sigma}{30}\right)^2.$$

The value of z^* is 1.96. (This value is given by invNorm(0.975) or invNorm(0.975,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.975,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd, DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

39. **(C)** Since the sample size is known to be between 5 and 30, inclusive, the number of degrees of freedom for the test must be between 4 and 29, inclusive. Looking at the table of t distribution critical values, the row for df = 4 shows a tail probability of 0.1 at $t = 1.533$ and a tail probability of 0.05 at $t = 2.132$. Thus, since this is a one-tailed test, when using 4 degrees of freedom the t-value of 1.58 would give us a p-value between 0.05 and 0.1. If we now look at the row for df = 29, we see a tail probability of 0.1 at $t = 1.311$ and a tail probability of 0.05 at $t = 1.699$. Therefore, when df = 29, we come to the same conclusion, that the p-value is between 0.05 and 0.1. Furthermore, a glance at the table for all the other numbers of degrees of freedom between 4 and 29 shows us that the same conclusion will be reached for all values in that range. We can therefore conclude that the p-value is between 0.05 and 0.1.

40. **(B)** The probability that the two randomly chosen people are of the same ethnicity is given by

$$\begin{aligned}&P(\text{both Caucasion}) + P(\text{both Hispanic}) + P(\text{both African American}) + P(\text{both Asian})\\&= (0.38)(0.38) + (0.34)(0.34) + (0.16)(0.16) + (0.12)(0.12)\\&= 0.3.\end{aligned}$$

Section II—Part A

Question One

(a) In the undergraduate program, just less than half of the students are female. In the master's program a smaller proportion are female (just over 40%), and in the doctoral program an even smaller proportion (around 35%) are female.

(b) No. It's possible that the number of students in the master's program is greater than in the doctoral program, resulting in a greater number of males in the master's program than in the doctoral program, even though the *proportion* of males is greater in the doctoral program.

(c) No. Since it is possible for a student to be enrolled in the doctoral program and to be female, the two events are not mutually exclusive.

(d) No. If a student is selected at random from the doctoral program, it is less likely that the student will be female than if the student were selected at random from the university as a whole. The two given events are therefore not independent.

Question Two

(a) It is quite possible, for example, that the students who choose to attend extra study sessions tend to be more hardworking than those who do not, and that it's the stronger work ethic of these students that is causing the high AP grades, not the extra study sessions.

(b) Write the names of the 200 students on identical slips of paper, and place the slips of paper in a hat. Randomly pick 100 of the slips from the hat. The students whose names appear on these slips will form Group A, and the remaining 100 students will form Group B.

(c) Since the students were randomly assigned to the two groups, the only differences between the two sets of people are

- chance differences resulting from the randomization process, and
- the fact that those in Group A attend the extra study sessions and those in Group B do not.

Since we're told that Group A's results were *significantly* higher on average than Group B's, we know that chance differences between the treatment groups is very unlikely to be the sole cause of the difference in results. Thus we have evidence that attending extra study sessions causes an increase in AP scores.

Question Three

(a) No. Since the mean is 160 and the standard deviation is 142, the z-value for zero is $(0-160)/142=-1.13$. Therefore, if the distribution of household water use were approximately normal, then the proportion of households using less than zero gallons of water during the year would be approximately $P(Z<-1.13)=0.13$. (This value is given by normalcdf(-999,-1.13) or normalcdf(-999,-1.13,0,1) on the TI-83 or TI-84 calculators, and by normalCdf(-999,-1.13,0,1) on the TI-Nspire. The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) Since it is not possible for 13% of households to use less than zero gallons of water we conclude that the distribution of household water use cannot be approximately normal.

(b) Yes. The Central Limit Theorem tells us that, since the sample size is 50 which is large (greater than 30), the sampling distribution of the sample mean is approximately normal.

(c) Since we are told that the distribution of indoor water use is approximately normal, we know that the distribution of the sample mean, $\overline{X}$ is approximately normal, also. The mean of the distribution of $\overline{X}$ is $\mu=57$ and its standard deviation is $\sigma/\sqrt{n}=12/\sqrt{50}=1.697$. So

$$P(\overline{X}>59)=P\left(Z>\frac{59-57}{1.697}\right)=P(Z>1.179)=\mathbf{0.119}.$$

(This probability is given by normalcdf(1.179, 999) or normalcdf(1.179, 999,0,1) on the TI-83 or TI-84 calculators, and by normalCdf(1.179, 999,0,1) on the TI-Nspire. The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

Question Four

(a) Check of conditions and naming of type of interval:

This is a two-proportion z-interval.

We are told that the samples were randomly selected.

Since $n_1\hat{p}_2 = 48(37/48) = 37 > 10$, $n_1(1-\hat{p}_1) = 48(11/48) = 11 > 10$, $n_2\hat{p}_2 = 45(20/45) = 20 > 10$ and $n_2(1-\hat{p}_2) = 45(25/45) = 25 > 10$ the samples are large enough.

Mechanics:

The 95% confidence interval for $p_1 - p_2$ is

$$\begin{aligned}&(\hat{p}_1-\hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\\ &=\left(\frac{37}{48}-\frac{20}{45}\right) \pm 1.96\sqrt{\frac{(37/48)(11/48)}{48}+\frac{(20/45)(25/45)}{45}}\\ &=(\mathbf{0.139,0.514}).\end{aligned}$$

(The value of z^*, 1.96, is obtained using invNorm(0.975) or invNorm(0.975,0,1) on the TI-83 and TI-84 calculators and using invNorm(0.975,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

Interpretation of interval:

We are 95% confident that the proportion of all girls in the school district who do not consume the recommended amount of vitamin A minus the proportion of all boys in the school district who do not consume the recommended amount of vitamin A is between 0.139 and 0.514.

(b) Yes. Since zero is not included in the interval, the sample results provide convincing evidence (at the 0.05 significance level) of a difference between the proportions of boys and girls in the school district who do not consume the recommended amount of vitamin A.

Question Five

(a) H_0: $\mu = 0.1$,
H_a: $\mu < 0.1$,
where μ is the mean level of asbestos at the site (in f/cc).

(b) A Type I error is obtaining convincing evidence that the mean asbestos level is less than 0.1 f/cc when in fact it is not.

A Type II error is failing to obtain convincing evidence that the mean asbestos level is less than 0.1 f/cc when in fact it is.

The consequence of a Type I error is that work at the site will continue when in fact the asbestos level is not below the permissible exposure limit. The workers will be exposed to harmful levels of asbestos.

The consequence of a Type II error is that work at the site will stop even though the level of asbestos in the air is below the permissible exposure limit. The workers will not be allowed to work at the site even though the level of asbestos is actually within the legal limit.

Section II—Part B

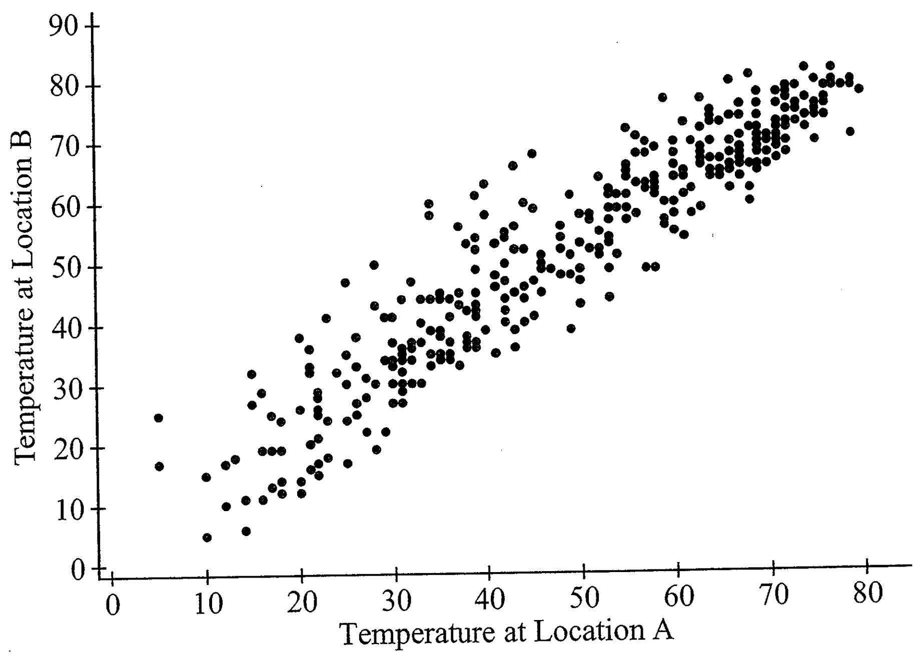
Question Six

(a) Any two of the following would be sufficient.

- The distribution of temperatures at Location A is negatively skewed.
- The distribution of temperatures at Location A is bimodal.
- The center of the distribution of temperatures at Location A is around 50 degrees.

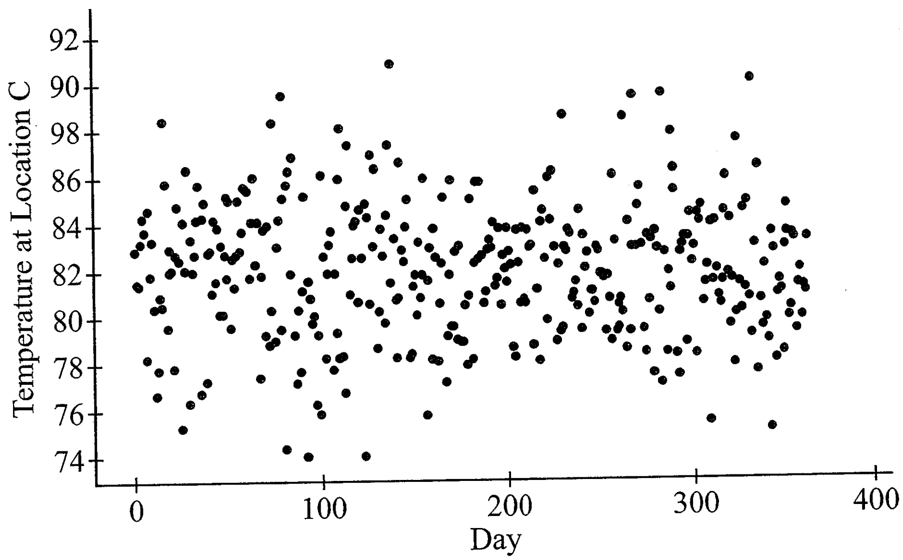
(b) Temperatures at Location A drop to their minimum around day 50 and rise to their maximum around day 200.

(c) Since, roughly speaking, temperatures rise at Location B when they rise at Location A, and fall at Location B when they fall at Location A, the scatterplot will show a positive correlation between the two variables. Values on each axis will range from around 0 to around 80. The scatterplot is shown below.



(d) A paired t-test would be appropriate.

(e) Since the given scatterplot shows a correlation of approximately zero between the temperatures at Location A and Location C, it is possible that the temperatures at Location C remain roughly the same whatever the temperature at Location A. This would be a consequence of Location C having temperatures that are roughly the same throughout the year. Temperatures at Location C range from about 74 to 91. Thus the time plot for Location C could be roughly as shown below.



Sample Examination Two – Answers

1. **(E)** The total number of goals scored is $3 \cdot 0 + 10 \cdot 1 + 6 \cdot 2 + 4 \cdot 3 + 2 \cdot 4$ and the total number of games played is $3 + 10 + 6 + 4 + 2$. So the mean number of goals scored is

$$\frac{3 \cdot 0 + 10 \cdot 1 + 6 \cdot 2 + 4 \cdot 3 + 2 \cdot 4}{3 + 10 + 6 + 4 + 2}.$$

2. **(E)** The treatments are shown in the table below.

		Distance	
		2	4
Weight	0	T1	T2
	2	T3	T4
	4	T5	T6
	6	T7	T8

Since there are 8 cells in the table, there are 8 treatments.

3. **(B)** We use *t* distributions in tests for means when the population standard deviation (or standard deviations) is (or are) unknown. Note that (C) is incorrect as it is a test for a proportion, and that (D) and (E) are chi-square tests.

4. **(D)** The brothers' performances are compared using *z*-scores.

For Robert, $z = (84 - 77.1)/7.3 = 0.945$.

For Justin, $z = (93 - 82.0)/7.8 = 1.410$.

For Bryan, $z = (81 - 73.2)/6.9 = 1.130$.

So the required list is Justin, Bryan, Robert.

5. **(B)** As stated on the formula sheet provided with the exam, for a discrete random variable, $\text{Var}(X) = \sigma_x^2 = \sum (x_i - \mu_x)^2 p_i$. So here, $\sigma_x = \sqrt{\sum (x_i - \mu_x)^2 p_i} = \sqrt{(0 - 1.62)^2 (0.08) + (1 - 1.62)^2 (0.22) + (2 - 1.62)^2 (0.70)}$.

6. **(C)** The hypothesis test will need to determine whether the results of the study provide convincing evidence that the mean number of eggs laid in the summer is greater than the mean number of eggs laid in the winter. Since the study is conducted using a single sample of ducks, and the numbers of eggs laid in the two seasons are recorded for each duck, paired data will be produced. Thus a paired *t*-test would be appropriate for analyzing the results.

7. **(C)** The Central Limit Theorem states that, if the sample size is large, the sampling distribution of the sample mean is approximately normal, whatever the shape of the population distribution. Note that for a random sample from a large population, the sampling distribution of the sample mean *always* has mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ (whether or not the sampling distribution of the sample mean is approximately normal). Thus the statements in (A) and (B), if read in isolation, are true; however, they are not consequences of the Central Limit Theorem.

8. **(B)** The test statistic is given by

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{33/98 - 0.25}{\sqrt{\frac{(0.25)(0.75)}{98}}} = 1.983.$$

The critical value of z for a one-tailed test at the 0.05 significance level is 1.645.

(This value is given by invNorm(0.95) or invNorm(0.95,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.95,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

Since the value of the test statistic (1.983) is greater than the critical value (1.645), the null hypothesis is rejected, and there is sufficient evidence to conclude that more than 25 percent of the adults in the town consume vegetables three or more times per day.

9. **(C)** We use here the conditional probability formula,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

We're told that $P(A \mid B) = P(B \mid A)$, and so

$$\frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A)}.$$

Since we're told that $P(A \cap B) \neq 0$, we can divide both sides of this equation by $P(A \cap B)$, giving

$$\frac{1}{P(B)} = \frac{1}{P(A)}.$$

From this it follows that $P(A) = P(B)$.

10. **(A)** The scatterplot shows a strong negative linear relationship between the variables, so the only value of the correlation coefficient from the ones available that could be correct is -0.91.

11. **(B)** The decision to make sale of the fish legal, or not, is based on whether the results of the study provide convincing evidence that the mean naphthalene level is less than 3.3. So the hypotheses used for the test will be $H_0: \mu = 3.3$ and $H_a: \mu < 3.3$. A Type I error is rejecting H_0 when H_0 is true, which here is obtaining convincing evidence that fish from the region are safe to eat, when in fact they are not. Having obtained this evidence, the authorities will then legalize sale of the fish.

12. **(B)** In the cumulative relative frequency curve given in the question, the increase in the cumulative relative frequency for each income bracket is given by the relative frequency for that bracket. (For example, the increase in cumulative relative frequency for the 60–80 thousand dollar income bracket is the relative frequency for that bracket, which is the number of families in that bracket divided by the total number of families in the country.) Thus, the income bracket that contains the greatest number of families is the one with the greatest increase in cumulative relative frequency, which the graph shows to be the $20,001 to $40,000 bracket.

13. **(E)** The chi-square test statistic is given by

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}.$$

14. **(A)** The null hypothesis for this chi-square test for homogeneity is that the true proportions falling into the three response categories are the same for low video quality as for high video quality. Stated another way, this null hypothesis is that video quality has no effect on the rating of content quality. Now the *p*-value for any hypothesis test is the probability of getting sample results at least as extreme as those obtained, given that the null hypothesis is true. So here the *p*-value can be said to be the probability of getting observed counts that are at least as far from the expected counts as was the case in this study given that video quality has no effect on the rating of content quality.

15. **(E)** It is quite possible, for example, that Tammi's sample will contain zero, one, or two ninth graders. None of these outcomes would provide adequate representation of that grade. Note that the statement in (C) is the definition of a simple random sample of size 20 in this context. The statement in (D), though not the definition of a simple random sample of size 20, is a consequence of that definition.

16. **(D)** Note, first, that the normal distribution is continuous; a discrete random variable cannot be approximately normally distributed unless it has a large number of possible outcomes. The distribution of the tail lengths of fully grown males of the common raccoon species is continuous, and it is entirely feasible that this distribution is bell-shaped. So (D) should certainly be considered possible as the answer to the question. We will now eliminate the other options. In (A), the distribution of the scores on a very easy 5-question mental math quiz is discrete (with only six possible outcomes) and likely to be negatively skewed (since the quiz was easy). In (B), the distribution of the scores on a large number of rolls of a cube whose faces are numbered 1 through 6 is again discrete with only six possible outcomes, and will be

close to being a discrete uniform distribution. In (C), the distribution of the number of attempts it took to pass the drivers' test for drivers in New York State is again a discrete distribution, and likely to be positively skewed, since the great majority of drivers in the state pass the test on their first attempt. Finally, in (E), for the survey where the possible responses were "yes," "no," and "maybe," the variable is categorical, not quantitative, and so the normal distribution cannot apply. Thus, with all other options eliminated, the answer to the question is (D).

17. **(E)** Since $r = 0.65$, $r^2 = 0.4225$. Thus, 42.25% (that is, less than half) of the variation in first grade math score can be explained by the regression line of first grade math score on time spent playing with brick-based construction toys.

    Note that (A) is incorrect since correlation does not imply causation. In (B), the values of one of the variables (time spent playing with brick-based construction toys) are being divided by 60: this has no effect on the value of the correlation coefficient. Turning our attention to (C), the average increase in first grade math score associated with a 1-unit increase in time spent playing with brick-based construction toys is the slope of the least squares regression line of math score on time spent playing with brick-based toys, not the value of the correlation coefficient. Finally, considering (D), the positive correlation indicates that, for children in the study, those with greater time spent playing with brick-based construction toys *tended* to have higher first grade math scores; but it cannot be said that for *any pair* of children included in the study, the one with the greater time spent playing with brick-based construction toys will have the higher first grade math score.

18. **(B)** Any well-designed experiment will include some random assignment to treatments, since the mathematical methods used to analyze the results of experiments depend on this. However, many experiments do not include a control group; an experiment that compares the effectiveness of two different teaching methods would be an example of this. Placebos are only applicable where the experimental units (that is, "subjects," in the case of people) might respond to the mere knowledge of the fact that they are receiving the treatment (in some way that affects the response variable). No experiment, for example, where the experimental units are inanimate objects will include a group that receives a placebo. Many experiments do not involve blocking; the completely randomized design is a common approach. Finally, there are many experiments that do not involve blinding, since, very often, it is not possible for the subjects not to know what treatment they are receiving. Take, for example, an experiment that studies the effect of exercise; the subjects obviously must know whether or not they are exercising, and so a double blind study is not possible in this context.

19. **(C)** In a t-test for a single population mean, the number of degrees of freedom is $n - 1$, where n is the sample size. Thus, increasing the sample size changes the number of degrees of freedom used in the test. Similarly, in a t-test for the slope of the population regression line, the number of degrees of freedom is $n - 2$, where n is the number of observations. Thus, in this case also, increasing the number of observations changes the number of degrees of freedom used in the test. However, in a chi-square test for independence the number of degrees of freedom is given by $(m-1)(n-1)$, where m is the number of rows and n is the number of columns in the table. Increasing the sample size will increase the counts within the cells of

the table, but will not change the number of rows or columns, and so the number of degrees of freedom used in the test will not change. Thus, statements I and III are true and statement II is false.

20. **(D)** First, since the outlier is an extreme value at the *low* end of the distribution, removing the outlier will increase the mean. Second, removal of the outlier will decrease the spread of the distribution, and therefore will decrease the standard deviation.

21. **(C)** In the design in (A), the people included in the survey are restricted to those who use the Whole Nutrition food market. It is very likely that people who use that store will have attitudes that differ from those of the population of the town as a whole with regard to recycling. Thus the design in (A) is unlikely to produce an accurate result.

    In the design in (B), attempts are made to diversify the set of people included. However, the question they are asked, by including the phrase "improving the environment by recycling greater proportions of household waste," is encouraging the respondents to say that they *have* changed their approach to recycling. Thus, the design in (B) is also likely to give an inaccurate result.

    In (D), even though every household in the town is *potentially* included, the people are asked to respond online. It is very possible that those who choose to respond will differ in their attitudes to recycling from those who do not. Therefore this design, too, is unlikely to give an accurate result.

    The design in (E) is clearly too imprecise to meet the administrators' requirements.

    In (C), however, a *random* set of households is selected, strong attempts are made to get a response from every household included in the sample, and the less biased version of the question is used. Therefore, of the five designs given, this is the one that is likely to produce the most accurate answer to the administrators' question.

    It is worth noting here that in this context a sample survey is possibly not a good approach to getting the information required. It could be argued that people will tend to say that they recycle more of their waste than they actually do, even when the question is worded in as unbiased a way as is possible. Some kind of study design that actually looks at how much material is being recycled might give a more accurate answer to the administrators' question.

22. **(A)** As stated on the formula sheet provided with the exam, the standard deviation of the sampling distribution of the sample proportion is given by

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

23. **(B)** The definition of 95% confidence is that in repeated sampling, 95% of the confidence intervals found will capture the true value of the parameter being estimated. Jimmy could be

just one of the 5% of the students taking part in this exercise (and working correctly) whose confidence intervals will not capture the true population mean.

24. **(C)** Let X be the number of pages in a randomly selected romance novel and Y be the number of pages in a randomly selected detective novel. We are given that $\sigma_X = 47$ and $\sigma_{Y=173}$. Since we are told that the novels are chosen independently and at random, X and Y are independent, and so $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$. Thus,

$$\sigma_{X+Y} = \sqrt{47^2 + 173^2} = 179.271.$$

25. **(A)** The confidence interval for the slope of the population regression line, β, is given by

$$b \pm t^* s_b,$$

where b is the estimated slope, s_b is its standard error, and t^* is the critical value of t for the appropriate number of degrees of freedom and the confidence level required. Here we are given that $b = 0.0623$ and $s_b = 0.0224$, and the number of degrees of freedom is $n - 2 = 28 - 2 = 26$. The critical value of t for a 95% confidence level and 26 degrees of freedom, found in the table of t critical values provided with the exam, is 2.056.

(Alternatively, this critical value can be found on some operating systems for the TI-84 calculator, and on the TI-Nspire, by entering invT(0.975,26). On those TI-84's on which it is installed, the invT function can be located using 2nd,DISTR. On the TI-Nspire (in the *Scratchpad* or *Calculator* application), the invt function is located by selecting menu,Statistics,Distributions.)

Thus, the required confidence interval is given by $0.0623 \pm (2.056)(0.0224)$.

26. **(D)** The proportion of "couples only" units for houses is $14/47 = 0.298$. The proportion of "couples only" units for apartments is $28/73 = 0.384$. Since $0.384 > 0.298$, the proportion of "couples only" units is higher for apartments than for houses.

27. **(E)** The standard deviation of the sampling distribution of the sample mean is given by $\sigma/\sqrt{n}$. We are given that $\sigma = 15$, and so when $n = 9$ the standard deviation of the sampling distribution of the sample mean is $15/\sqrt{9} = 5$, and when $n = 16$ the standard deviation of the sampling distribution of the sample mean is $15/\sqrt{16} = 3.75$.

Now, whatever the sample size, the sample mean is clearly more likely to be between 90 and 110 than between 95 and 105 or between 98 and 102, since both of the latter two intervals are subsets of the larger interval.

So we look now at the interval 90 to 110. When $n = 9$, the z values corresponding to the endpoints are $\pm 10/5 = \pm 2$, and when $n = 16$, the relevant z values are $\pm 10/3.75 = \pm 2.667$. Therefore, since the z values are greater in absolute value for $n = 16$ than for $n = 9$, the greater probability is for $n = 16$.

28. **(D)** If we were to list the files in order of size, the median file size would be very close to the size of the $464/2 = 232$nd file, the first quartile would be very close to the size of the $464/4 = 116$th file, and the third quartile would be very close to the size of the $464(3/4) = 348$th file. Looking at the frequencies given in the histogram, since $112 < 116$, the first quartile is greater than or equal to 15. Also, since $112 + 217 + 51 = 380$, which is greater than 348, the third quartile is less than 45. Therefore the IQR, which is $Q_3 - Q_1$, is less than $45 - 15 = 30$.

Note that (A) is not correct: $112 + 217 = 329$, which is greater than 232, and so the median is *at most* 30. Looking at (B), the maximum file size is less than or equal to 315 and the minimum is greater than zero, and so the range is less than 315 (and therefore not greater than 320). Option (C) is incorrect since, as explained above, the first quartile is greater than 15. Finally, it is clear from the histogram that the distribution is positively skewed, and so (E) is not correct.

29. **(C)** Since the null hypothesis for this two-sample t test is $H_0: \mu_1 = \mu_2$, the test statistic is

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

Thus,

$$t = \frac{2.886 - 3.114}{\sqrt{\frac{(0.525)^2}{10} + \frac{(0.644)^2}{15}}}.$$

30. **(A)** Think of the students being chosen one-by-one. The probability that the first student chosen is a girl is $8/12$. Given that this happens, there are then 11 students remaining, of whom 7 are girls; so the probability that the second student is also a girl is $7/11$. Therefore, the probability that the first two students chosen are both girls is $(8/12)(7/11)$. Continuing this process, we see that the probability that all of the students chosen are girls is

$$\left(\frac{8}{12}\right)\left(\frac{7}{11}\right)\left(\frac{6}{10}\right)\left(\frac{5}{9}\right) = 0.141.$$

Alternatively, this question can be answered using combinations. (Please note that an understanding of this approach is not required for AP Statistics. However, some students might be interested in solving the problem this way.) The total number of ways of choosing a team of 4 students is ${}_{12}C_4 = 495$. The number of ways of choosing four girls is ${}_{8}C_4 = 70$. So the probability that the team consists entirely of girls is $70/495 = 0.141$.

(Using the TI-83 and TI-84 calculators, the value of ${}_{12}C_4$ is found by typing 12 nCr 4. The nCr function is accessed by MATH,PRB,3. On the TI-Nspire, this value is found by typing nCr(12,4), and the nCr function is accessed by selecting menu,Probability,Combinations.)

31. **(B)** Since the p-value is greater than α, we do not reject H_0, and we do not have convincing evidence that H_a is true. Note that we never accept H_0, and so option (A) is incorrect. We'll look now at option (C). We never have evidence that H_0 is true. Saying that "we have convincing evidence that H_a is false," is equivalent to saying that we have convincing evidence that H_0 is true. Thus option (C) is incorrect.

32. **(E)** When $x = 66{,}000{,}000$,

$$\begin{aligned}\text{predicted value of } \log(y) &= -0.285 + 0.2255 \log(66{,}000{,}000) \\ &= 1.47831\end{aligned}$$

Therefore, the value of y (the number of lightning flashes per minute) predicted by the model is $10^{1.47831} = 30.1$.

33. **(A)** A random pattern in the residual plot tells us that a linear model is appropriate.

34. **(C)** The possible outcomes, and the total scores for each, are shown in the table below.

		Score on First Cube					
		1	2	3	4	5	6
Score on Second Cube	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Since there are 36 possible outcomes, and 2 of those outcomes result in a total score of 11, the probability that the total score is 11 is $2/36 = 1/18$.

Alternatively, we could let X be the score on the first cube and Y be the score on the second cube. Then,

$$\begin{aligned}P(\text{total score} = 11) &= P(X = 6 \text{ and } Y = 5) + P(X = 5 \text{ and } Y = 6) \\ &= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{2}{36} = \frac{1}{18}.\end{aligned}$$

35. **(B)** Since a confidence interval for the difference of two population means is to be calculated, and the population standard deviations are unknown, a two-sample t-interval will be constructed (and therefore the statement in (C) is true). For construction of a two-sample t-interval, it is necessary to assume that the samples were randomly and independently selected (and so the statement in (A) is true). However, because the sample sizes are 97 and 94, which are large, the Central Limit Theorem allows us to construct this interval even if the population

distributions are not approximately normal. Thus, the statement in (B) is false, making (B) the correct response to the question.

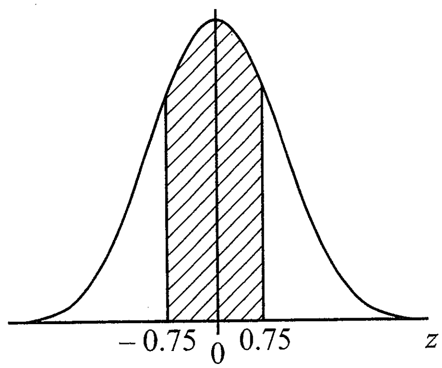
To see that that statements in (D) and (E) are true we recall the formula for a two-sample t-interval:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

The formula tells us that the quantity $\bar{x}_1 - \bar{x}_2$ is at the center of the interval (so the statement in (D) is true), and that if n_1 and n_2 are increased, the fractions within the square root are likely to decrease, thereby producing a narrower confidence interval (and so the statement in (E) is true).

36. **(D)** The student is thinking of foot size as the explanatory variable and speed as the response variable, and is claiming that, all else being equal, an increase in foot size will be associated with an increase in speed. The teacher, however, points out the association between *age* and foot size, and suggests that the greater age of the large-footed students could be causing the increase in speed, and so that speed could have nothing to do with foot size. Age is being named as a confounding variable.

37. **(E)** A mass that is 54 milligrams away from the mean mass has a z value of $54/72 = 0.75$. The required probability is therefore given by the area of the shaded region in the diagram below.



This area can be obtained by subtracting the area to the left of -0.75 from the area to the left of 0.75. Thus the required probability is $P(z < 0.75) - P(z < -0.75)$.

38. **(A)** A 95% confidence interval is equivalent to a two-tailed hypothesis test at the 0.05 significance level. We'll illustrate this in the context of this question. We have the confidence interval for μ, (58.770, 61.428). Therefore, taking any number that lies within the confidence interval, such as 59, for example, we would say that it is feasible that $\mu = 59$. More precisely, the null hypothesis $H_0: \mu = 59$ would not be rejected at the 0.05 level when tested against $H_a: \mu \neq 59$. On the other hand, taking a value that is not in the interval, such as 57, the null hypothesis $H_0: \mu = 57$ *would* be rejected at the 0.05 level when tested against $H_a: \mu \neq 57$. So option (A) is correct.

We'll now look at the other options. Looking at (B), it is clear from the confidence interval that we have evidence that μ is greater than 57, but not that μ is *less* than 57. Thus option (B) is incorrect. Turning our attention to (C), as explained above, the confidence interval tells us that the null hypothesis $H_0: \mu = 59$ would *not* be rejected at the 0.05 level when tested against $H_a: \mu \neq 59$. So option (C) is incorrect. Looking at (D), the confidence interval tells us that $H_0: \mu = 60$ would not be rejected when tested against $H_a: \mu \neq 60$. It does not necessarily follow from this that $H_0: \mu = 60$ would be rejected when tested against $H_a: \mu > 60$. However, since the center of the interval, which is the sample mean that was obtained in this study, is close to 60, it is clear that we do not have convincing evidence that $\mu > 60$. So H_0 would *not* be rejected, and option (D) is incorrect. Finally, for (E), the entire confidence interval lies below 62, so we clearly have strong evidence that $\mu < 62$. In other words, the null hypothesis $H_0: \mu = 62$ would be rejected when tested against $H_a: \mu < 62$, and option (E) is incorrect.

39. **(B)** Random assignment ensures that any differences between the treatment groups occur completely by chance. The mathematics that analyzes the results of the experiment then labels any difference between the results for the two groups as "significant" if such a difference is very unlikely to have appeared solely as a result of chance differences between the groups. Note that option (A) is incorrect since, while we hope that the treatment groups will be similar, we cannot expect them to be *exactly* the same with respect to any variable.

40. **(D)** The hypothesis test in I is comparing the sample proportions from two random samples selected independently from two populations. Thus it is appropriate to use a two-proportion *z*-test.

    In III, an experiment is performed in which the customers are randomly assigned to two groups, and the proportions choosing to use cash are compared for the two groups. A two-proportion z-test is also appropriate in this experimental context.

    However, in II, the responses are in three categories, and the proportions of adults falling into the *three* categories are compared for the two samples. A two-proportion *z*-test could not be used to make this comparison. (A chi-square test for homogeneity could be used for a comparison such as this.)

Section II—Part A

Question One

(a) The median for the same-sex adult group (around 24.5) is greater than the median for the different-sex adult group (around 21.5). The median for the different-sex adult group (around 21.5) is greater than the median for the control group (around 10).

The interquartile range for the same-sex adult group (around $31 - 14 = 17$) is slightly greater than the interquartile range for the different-sex adult group (around $28 - 12 = 16$). The interquartile range for the different-sex adult group (around 16) is greater than the interquartile range for the control group (around $14 - 6 = 8$).

The distributions for the same-sex adult and different-sex adult groups are negatively skewed. The distribution for the control group is positively skewed (or roughly symmetrical, with two high outliers).

(b) For the same-sex adult group, since the distribution is negatively skewed, the mean is likely to be less than the median.

(c) If there had been no control group, the experiment would only have enabled the researchers to compare between children exposed to same-sex and different-sex models showing nurturing behaviors. By including the control group the researchers are able to compare the same-sex and different-sex groups to children who did not see any nurturing behavior shown by an adult to the furry toys.

Question Two

(a) The equation of the least squares regression line is

$$\hat{y} = -9.47 + 32.461x,$$

where x = screen size, and y = price.

(b) We know that the slope of the least squares regression line is 32.461. Therefore, when x increases by 4, $\hat{y}$ increases by $4(32.461) = 129.844$. The GPS unit with the larger screen is expected to cost **$129.84** more than the unit with the smaller screen.

(c) When $x = 6.5$, $\hat{y} = -9.47 + 32.461(6.5) = 201.5265$. The unit with screen size 6.5 is predicted to cost **$201.53**.

(d) No. When the unit with screen size 7 inches is removed, it is clear from the graph that the correlation for the remaining points would be close to zero. Therefore, the slope of the least squares regression line for the remaining points would also be close to zero, which is very different from the clear positive slope of the line shown.

Question Three

(a) Let X be the number of the 8 patients who stay in the ICU for less than 24 hours. Then the distribution of X is binomial, with $n = 8$ and $p = 0.278$. So

$$\begin{aligned} P(X \leq 2) &= P(X=0)+P(X=1)+P(X=2) \\ &= (0.722)^8 + \binom{8}{1}(0.278)(0.722)^7 + \binom{8}{2}(0.278)^2(0.722)^6 \\ &= \mathbf{0.608} \end{aligned}$$

(In the calculation above, the expression

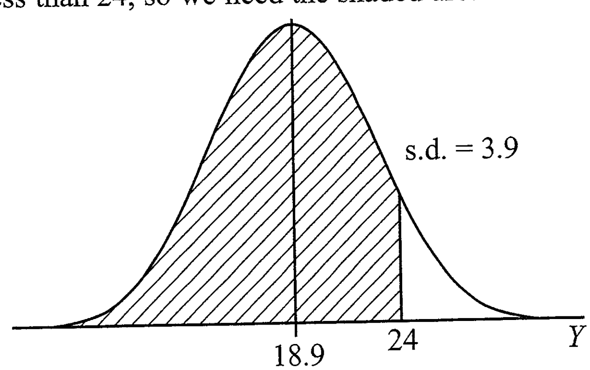
$$\binom{8}{2}$$

is another way of writing $_8C_2$. Using the TI-83 and TI-84 calculators, the value of $_8C_2$ is found by typing 8 nCr 2. The nCr function is found by selecting MATH,PRB,3. On the TI-Nspire (in the *Scratchpad* or *Calculator* application), this value is found by typing nCr(8,2), and the nCr function is found by selecting menu,Probability,Combinations.

Alternatively, the answer to the question can be found directly, using a calculator. On the TI-83 or TI-84 calculators, use binomcdf(8,.278,2). The binomcdf function can be accessed by 2nd,DISTR. On the TI-Nspire (in the *Scratchpad* or *Calculator* application), use binomCdf(8,.278,0,2). The binomCdf function can be found by selecting menu,Statistics,Distributions.)

(b) If a random variable X has a binomial distribution with n trials and a probability of p of success on each trial, then $\mu_X = np$ and $\sigma_X = \sqrt{np(1-p)}$. Here, $n = 20$ and $p = 0.278$, so $\mu_X = 20(0.278) = \mathbf{5.56}$ and $\sigma_X = \sqrt{20(0.278)(0.722)} = \mathbf{2.004}$.

(c) Let the length of stay (in hours) for a randomly chosen elective patient be Y. The distribution of Y is approximately normal with mean 18.9 and standard deviation 3.9. We require the probability that Y is less than 24, so we need the shaded area shown in the diagram below.



At $Y = 24$, $Z = (24 - 18.9)/3.9 = 1.308$. So we require $P(Z < 1.308) = \mathbf{0.905}$.

(This value is obtained on the TI-83 and TI-84 calculators by entering normalcdf(-999,1.308) or normalcdf(-999,1.308,0,1), and on the TI-Nspire by entering normalCdf(-999,1.308,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

Question Four

(a) The proportions of the entire inventory falling into the four categories are given by $124327/198233 = 0.627$, and so on. The required proportions are given in the table below.

Category	Journals	Books	DVDs	Other
Proportion of Inventory	0.627	0.284	0.053	0.036

(b) Hypotheses:

Let p_1, p_2, p_3, and p_4 be the proportions of all uses falling into the given categories.

H_0: $p_1 = 0.627$, $p_2 = 0.284$, $p_3 = 0.053$, $p_4 = 0.036$

H_a: H_0 is false.

Check of conditions and naming of type of test:

This is a chi-square test for goodness of fit.

We are told that the librarians are willing to treat the uses of the library during that week as a random sample from the set of all uses of the library.

The observed and expected counts are shown in the table below.

Category	Journals	Books	DVDs	Other
Observed	600	353	58	30
Expected	652.71	295.64	55.173	37.476

(The expected counts are calculated as follows. The total number of uses during the week was $600 + 353 + 58 + 30 = 1041$. The first expected count is therefore $1041(0.627) = 652.71$. The other expected counts are calculated in the corresponding way.)

Since all the expected counts are greater than 5, the sample is large enough for the chi-square test to be appropriate.

Mechanics:

$$\chi^2 = \sum \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$$
$$= \frac{(600 - 652.71)^2}{652.71} + \cdots + \frac{(30 - 37.476)^2}{37.476}$$
$$= 4.256 + 11.127 + 0.145 + 1.491$$
$$= 17.020.$$

We use the chi-square distribution with $4 - 1 = 3$ degrees of freedom.

So the p-value is $P(\chi_3^2 > 17.020) = 0.001$.

(On the TI-83, TI-84, and TI-Nspire calculators, this result is obtained using χ^2cdf(17.020,999,3). On the TI-83/84, the χ^2cdf function is accessed by 2nd,DISTR, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) the χ^2Cdf function is accessed by selecting menu,Statistics,Distributions.)

Conclusion:

Since the *p*-value is 0.001, which is less than 0.05, we have convincing evidence that the proportions of all uses that fall into the four categories are different from the proportions of items in those categories.

Question Five

(a) A t distribution should be used, since the standard deviation of the lengths of all the songs on the site is unknown.

(b) In 95% of random samples of 15 songs from the site, the confidence intervals found will capture the true mean length of all the songs on the site.

(c) The formula for this 95% confidence interval is

$$\overline{x} \pm t^* \cdot \frac{s}{\sqrt{n}}.$$

The number of degrees of freedom used is $n - 1 = 15 - 1 = 14$. So the t^* quantity in the formula is the critical value of t, with 14 degrees of freedom, for a 95% confidence interval. The table provided with the exam gives this value to be 2.145.

(Alternatively, this critical value can be found on some operating systems for the TI-84 calculator, and on the TI-Nspire, by entering invT(0.975,14). On those TI-84's on which it is installed, the invT function is accessed using 2nd,DISTR. On the TI-Nspire (in the *Scratchpad* or *Calculator* application), the invt function is accessed by selecting menu,Statistics,Distributions.)

We are told that Chin-Sun's confidence interval was 242.733 ± 19.209. So, referring to the formula above, we see that $\overline{x} = 242.733$ and

$$t^* \cdot \frac{s}{\sqrt{n}} = 19.209.$$

So, since $n = 15$ and $t^* = 2.145$,

$$2.145 \cdot \frac{s}{\sqrt{15}} = 19.209.$$

Solving for s,

$$s = \frac{19.209\sqrt{15}}{2.145} = 34.684.$$

The standard deviation of song lengths in Chin-Sun's sample was **34.684** seconds.

Section II—Part B

Question Six

(a) Write the names of the 32 employees on identical slips of paper. Place the slips of paper in a hat. Randomly choose 8 slips from the hat, without replacement. The employees whose names are shown on the slips selected should be included in the sample.

(b) Employees paid at the different grade levels earn differing amounts, and therefore are likely to differ in their spending patterns. Stratifying by grade level ensures an adequate representation of each grade level in the sample, and therefore ensures a good representation of the differing spending patterns.

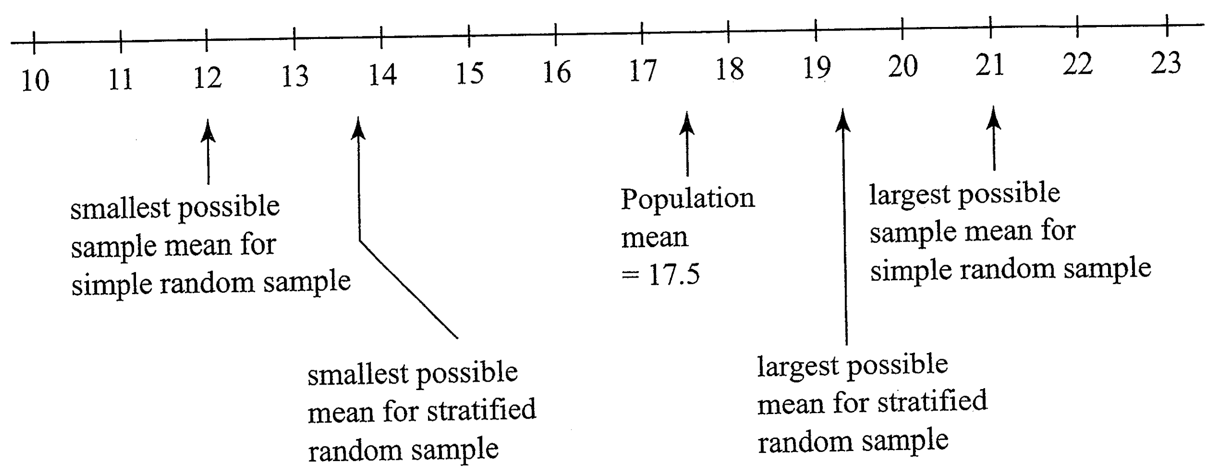
(c) The lowest possible sample mean for the stratified sampling method given is found by including in the sample the 4 "Individual" level employees with the lowest spending values, the 3 "Professional" level employees with the lowest spending values, and the "Managerial" level employee with the lowest spending value. This gives

$$\bar{x}_{\min} = \frac{9 + 10 + 11 + 13 + 15 + 15 + 17 + 20}{8} = \mathbf{13.75}.$$

The highest possible sample mean for the stratified sampling method given is found by including in the sample the 4 "Individual" level employees with the highest spending values, the 3 "Professional" level employees with the highest spending values, and the "Managerial" level employee with the highest spending value. This gives

$$\bar{x}_{\max} = \frac{17 + 17 + 17 + 18 + 19 + 20 + 20 + 26}{8} = \mathbf{19.25}.$$

The number line, with these values added, is shown below.



(d) The answer to part (c) suggests that it is the stratified sampling method that produces the sample mean with the smaller variability, since this sample mean has a range of possible values $(19.25 - 13.75 = 5.5)$ which is smaller than the range of possible values for the sample mean for the simple random sampling method $(21 - 12 = 9)$.

(e) The aim of whichever sampling method is chosen is to produce a sample mean that is close to the true population mean spending value. The stratified random sampling method produces a sample mean that has a smaller variability than the sample mean produced by the simple random sampling method. This tells us that, when taking one stratified random sample according to the method given in the question, the sample mean produced is likely to be closer to the true population mean than when one simple random sample is taken. Therefore, for estimating the population mean spending value, the stratified random sampling method described in the question would be preferable to simple random sampling.

Sample Examination Three – Answers

1. **(D)** The upper half of the boxplot is much longer than the lower half, and so the distribution is skewed to the right. Referring to the boxplot again, the IQR is $Q_3 - Q_1 \approx 2.7 - 0.7 = 2$. So the IQR of the pay raises is approximately \$2000.

2. **(A)** This is cluster sampling, with each volume forming a cluster. (To avoid confusion, note that the word "volume" here refers to one of the 30 books that make up the encyclopedia.) Do not get cluster sampling confused with stratified random sampling. If the researcher believed that the various volumes were *different* in terms of the proportion of diagrams and pictures, then she might take a random sample of pages from *every volume*. That would be stratified random sampling.

3. **(A)** The null hypothesis must reflect the company's claim that the mean breaking strength is, in fact, 80 pounds. Since the group wants to find out whether there is evidence that the cloth's breaking strength is *less* than 80 pounds, the alternative hypothesis is as given in (A).

4. **(D)** Suppose that an employee is chosen at random. Let M be the event that the employee opts for medical insurance and let L be the event that the employee opts for life insurance.

   We are given that $P(M) = 0.78$, $P(L) = 0.42$, and $P(M \cup L) = 0.82$. The "addition rule" states that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, or, solving for $P(A \cap B)$, that $P(A \cap B) = P(A) + P(B) - P(A \cup B)$. So, applying that rule, $P(M \cap N) = P(M) + P(N) - P(M \cup N) = 0.78 + 0.42 - 0.82 = 0.38$.

5. **(E)** By drawing a vertical line up from 43 on the horizontal axis to the curve and then drawing a horizontal line from this point of intersection to the vertical axis, we see that the percentage of members *less than* or equal to 43 years old is approximately 25%, that is, approximately 40 members. So the number of members *more than* 43 years old is approximately 120.

6. **(B)** Two independent random samples are being taken and their means are being compared. So the two-sample t-test for means would be used.

   Let us now look at why the other options are incorrect. Considering (D) and (E), note that the coach wishes to compare the thrower's average (mean) length of throw in the morning with his/her average (mean) length of throw in the afternoon. The tests in (D) and (E) are for proportions, and are therefore not applicable to this situation. The test in (A) is a *one*-sample test which therefore cannot be used to make this comparison. Lastly, looking at (C), the paired t-test is used when you have a single sample with two data values for each element in the sample or when you have an experiment that uses matched pairs. Neither of these scenarios is present in this context, and so option (C) is incorrect.

7. **(B)** If, for example, she sets 6 places, then the probability that all the children can be seated is the probability that 6 or fewer children show up. This is the sum of the probabilities up to and

including 6, which is 0.78, and this is not quite enough. Diana wants to be 80% sure, so she will have to set 7 places.

8. **(A)** This is a one proportion z-test. The null hypothesis is $H_0: p = 0.1$, and the z-statistic is calculated assuming H_0 to be true. So the z-statistic is $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} = \dfrac{0.08 - 0.1}{\sqrt{\dfrac{(0.1)(0.9)}{200}}}$.

We're doing a two-tailed test since we're asking whether the proportion of defective springs has *changed*, and the p-value is twice the area in the *left* tail of the normal distribution, since the sample proportion is *less* than the hypothesized population proportion. Hence, the p-value is $2 \cdot P\left(z < \dfrac{0.08 - 0.1}{\sqrt{\dfrac{(0.1)(0.9)}{200}}}\right)$. (The conditions for use of the one-proportion z-test are satisfied since we are given that the sample is random, and $np = 200(0.1) = 20 \geq 10$, $n(1-p) = 200(0.9) = 180 \geq 10$.

Note: If we were asked for a *confidence interval* (with, let's say, a 95% confidence level), then the formula would be $0.08 \pm 1.96\sqrt{\dfrac{(0.08)(0.92)}{200}}$. The square root term used in the hypothesis test is different from the one in the confidence interval.

9. **(A)** P has an extremely large x-coordinate relative to those of the other data points. Removal of a point with an extreme x-coordinate will have a large effect on the regression line if the point is not close to the regression line that would pertain to the remaining points. This is clearly the case here. What is more, it is obvious that without P the regression line would have a greater slope than the one shown in the diagram. However, Q has an x-coordinate that is near the mean x-coordinate for all the data points and has an extreme y-coordinate. So the effect of the removal of Q would be to lower the regression line, but only very slightly as there are so many other points.

10. **(C)** This is a chi-square test for goodness of fit.

	Blue	Brown	Green
Expected Count	16	14	10
Observed Count	10	19	11

So we have $\chi^2 = \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}} = \dfrac{(10-16)^2}{16} + \dfrac{(19-14)^2}{14} + \dfrac{(11-10)^2}{10} = 4.136.$

The number of degrees of freedom is $3 - 1 = 2$. So p-value $= P(\chi_2^2 \geq 4.136) = 0.126$.

(On the TI-83, TI-84, and TI-Nspire calculators, this result is obtained using χ^2cdf(4.136,999,2). On the TI-83/84, the χ^2cdf function is accessed by 2nd,DISTR, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) the χ^2Cdf function is accessed by selecting menu,Statistics,Distributions.)

11. **(E)** The number of trials, X, up to and including the first success has a geometric distribution. $P(X \geq 3) = 1 - P(X = 1) - P(X = 2) = 1 - (0.2) - (0.8)(0.2) = 0.64$.

 Alternatively, taking three or more trials to get the first success is equivalent to failing on the first two trials. The probability that this happens is $(0.8)(0.8) = 0.64$.

12. **(B)** A small p-value indicates strong evidence against H_0 and in favor of H_a. We have no evidence at all to support the alternative hypotheses in (A) and (E). With a sample mean of 10.8 we clearly have stronger evidence for $\mu > 10$ than for $\mu < 11$; so the p-value for (B) is smaller than the p-value for (C). The p-value for (D) is twice the p-value for (C). So the p-value for (B) is smallest.

13. **(D)** The mean, the standard deviation, and the range will all be affected by the outlier. Thus, options (A), (B), (C), and (E) are all incorrect, and so the correct response is (D).

14. **(C)** The number of underclassmen who take the bus to school is $638 \cdot (0.831) = 530$ and the number of upperclassmen who take the bus to school is $523 \cdot (0.709) = 371$. So the required probability is $530/(530 + 371) = 0.588$.

15. **(C)** In any hypothesis test, the p-value is the probability of getting a value of the test statistic at least as extreme (in favor of H_a) as the one obtained, given that H_0 is true. Here H_0 states that $p_M = p_W$, and so the correct response is (C).

16. **(B)** We are told that the Board wishes to determine the opinions of the parents of students in grades K-12 on the issue. Therefore, sending the message to parents of students who are not in the 12th grade is not a source of bias, and so (B) is the correct response.

 The statement in (A) is correct (and therefore (A) is an incorrect response) since the email message includes the sentence, "The reduction in parking would allow for a substantial expansion in student activities." This comment is likely to influence the reader in favor of not allowing 12th graders to drive to school. Likewise, the issues sited in (C), (D), and (E) are clearly all potential sources of bias, and so these options are incorrect.

17. **(C)** The correlation is unaffected by adding a constant to all the x-values (or y-values) or by multiplying all the x-values (or y-values) by a constant. For any given student, let x be the height and y be the weight. When the heights are converted to centimeters and the weights are converted to kilograms, the x-values are multiplied by 2.54 and the y-values are multiplied by 0.454. So the value of the correlation is unchanged.

18. **(E)** Saying that the distribution of A's throws is positively skewed does not tell us anything about how far A is throwing. It merely tells us that the spread of the throws above the median is greater than the spread of the throws below the median. Thus, the correct response is (E).

 Let us look now at why the statements in the other options are correct (making these incorrect responses). Clearly the scenarios in (A) and (B) would be valid reasons for choosing thrower A over thrower B. Considering (D), since a javelin competition is judged according to a thrower's maximum throw, the maximum for thrower A being greater than the maximum for thrower B is a good reason for choosing thrower A over thrower B. Looking at (C), suppose, for example, that the upper quartiles for throwers A and B were 45 meters and 40 meters respectively. This would mean that 25% of the time thrower A is throwing the javelin at least 45 meters, while 25% of the time thrower B is throwing at least 40 meters, which is clearly less good than A's achievement.

19. **(B)** The two explanatory variables are the drug administered and the number of doses.

20. **(C)** Note that the confidence interval tells us something about the population *mean*.
The statement in (C), however, tries to say something about the actual flight times, without reference to the population mean. This is therefore a false statement, and thus (C) is the correct response.

    Options (A), (B), (D), and (E) are all correct statements regarding the population mean (making these responses incorrect).

21. **(B)** Both statistics (the sample mean and the sample median) are unbiased since their expected values (means) are both 3, which is the value of the parameter that is being estimated. We prefer the statistic that has the smaller standard deviation, since this tells us that this statistic is more likely than the other statistic to be *close* to μ.

22. **(B)** To say that two events are mutually exclusive means that they can't both happen. In this case, A and B *can* both happen (the card can be the green 12) so A and B are not mutually exclusive. To say that two events A and B are independent is to say that knowing that one of them is happening does not change the probability that the other is happening; in other words that $P(A \mid B) = P(A)$, or, equivalently, $P(B \mid A) = P(B)$. Here, the probability that the card is a 12 is 4/48 = 1/12, and the probability that the card is a 12 given that it is green is also 1/12. So the events are independent.

23. **(D)** We know that the study is conducted in at least a *single* blind manner as we are told that the people who measure the muscle mass do not know which subjects are in which group. However, it can't be *double* blind as the subjects clearly know whether or not they are doing the warm-up exercises.

    The statement in (A) is correct (making (A) an incorrect response), since the treatments (doing or not doing the warm-up exercises) are imposed on the volunteers. The statement in (B) is clearly correct (making (B) an incorrect response) since the volunteers are randomly assigned to the groups. The statement in (C) is correct (making (C) an incorrect response) since one of the groups does no warm-up exercises, thus making this group a control group. The statement in (E) is correct (making (E) an incorrect response) since there is indeed no blocking in the description of the experiment.

24. **(B)** If $y = ax + b$, where a and b are constants and $a > 0$, then $s_y = a \cdot s_x$. So here $s_y = 0.7s_x = 0.7 \cdot 8.8 = 6.16$.

Alternatively, multiplying the x-values by 0.7 has the effect of multiplying the standard deviation by 0.7, and adding 30 has no effect on the standard deviation. So the new standard deviation is $0.7 \cdot 8.8 = 6.16$.

25. **(C)** r^2 is the proportion of the y-variability that can be explained by the regression line. (Here's an explanation of this in the context of this question: If the math scores and the physics scores are linearly related, then as we think of the math scores varying we expect the physics scores to vary accordingly. However, the actual physics scores do not vary exactly according to the approximate linear relationship: some are above the line, some below. So some of the variation in physics scores can be explained by their approximate linear relationship to the varying math scores, some cannot. The proportion of the variation of the physics scores that *can* be explained by the variation of the math scores is r^2.)

Let's look now at the other options. Option (A) is incorrect as a high value of r^2 could indicate a large *negative* correlation. Option (B) is incorrect because a high value of r^2 (or, equivalently, a value of r that is close to 1 or −1) does not mean that a straight line is the *best* model – you might still find a curve that fits the data better. The value of r^2 can give no information about the presence (or not) of outliers, and so option (D) is incorrect. Finally, considering option (E), a comparison between the physics scores and the math scores would be achieved by looking at the pattern of points in the scatterplot relative to the line $y = x$. Assuming that the math scores are denoted by x and the physics scores by y, if the points are generally above the line $y = x$ then, generally speaking, the physics scores are higher than the math scores. If the points are generally below the line $y = x$, then, generally speaking, the physics scores are lower than the math scores. However, the value of r^2 gives no information about a comparison between the scores in the two subjects since, for example, a value of r^2 close to 1 could indicate a strong linear relationship between the two variables with the points generally above the line $y = x$ or with the points generally below the line $y = x$.

26. **(D)** The column total for "Democrat" is 182, the row total for "Female" is 210, and the grand total is 400. So the expected number of female Democrats is $182 \cdot 210/400 = 95.55$.

27. **(D)** In order to show why the result does not imply that in order to live longer one should start to regularly consume substantial amounts of olive oil, we have to explain how olive oil consumption can be *associated* with long life without olive oil consumption actually *causing* long life. The argument in (D) is the only one that does this. Note that option (C) is not adequate, because it contains no suggestion that regular exercise (or any other life prolonging habit) is *connected with olive oil consumption*, and therefore does not explain how olive oil consumption can be associated with long life without causing it.

28. **(E)** A scatterplot deals with the situation where you have one set of objects (or people) and *two* measurements for each object. For example, a scatterplot could illustrate, for one set of children, the connection between their on-task times and their spelling scores. However, here we have two sets of children being compared, and so a scatterplot would not be suitable; thus, (E) is the correct response. (All of the graphs given in the other four options would be suitable for making this comparison.)

29. **(D)** Let the amount of flour per bag have mean μ. We are told that the distribution has standard deviation $\sigma = 0.4$. We estimate μ using the sample mean, $\bar{x}$. If the sample size is large (we hope to get an answer bigger than 30), then the distribution of $\bar{x}$ is approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$. For any normally distributed random variable, with 95% probability the random variable will be within 1.96 standard deviations of the mean. (This value, 1.96, is given by invNorm(0.975) or invNorm(0.975,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.975,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) So, with 95% probability, $\bar{x}$ will be within $1.96 \cdot \sigma/\sqrt{n}$ of μ. So, referring to the question, we need

$$1.96 \cdot \frac{\sigma}{\sqrt{n}} = 0.1,$$

that is,

$$1.96 \cdot \frac{0.4}{\sqrt{n}} = 0.1.$$

So

$$\sqrt{n} = \frac{1.96 \cdot 0.4}{0.1},$$

giving

$$n = \left(\frac{1.96 \cdot 0.4}{0.1}\right)^2 = 61.47.$$

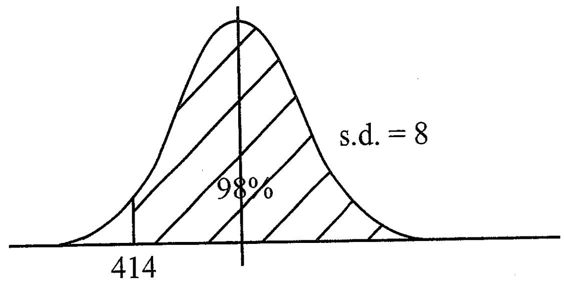
So we need $n = 62$.

30. **(E)** All three statements are true. Note that the question states that we are talking about small samples – if the sample were large we would not need to make the assumptions in I and II. With regard to II, it's important to remember that the *t*-test does indeed require the assumption that the population distribution is normal.

31. **(D)** The number of seeds that germinate is binomially distributed with $n = 20$ and $p = 0.7$. The mean of this binomial distribution is given by $np = 20 \cdot 0.7 = 14$ and the standard deviation is given by $\sqrt{np(1-p)} = \sqrt{20 \cdot 0.7 \cdot 0.3} = 2.049$.

Be careful not to confuse the number of seeds that germinate with the *proportion* of seeds that germinate. The values given in (A) are the correct mean and standard deviation for the *proportion* of seeds that germinate.

32. **(D)**



We are given that 98% of the cans contain more than 414 ml, which means that 2% contain less than 414 ml. The z-value for 414 is −2.054. (This value is given by invNorm(0.02) or invNorm(0.02,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.02,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) Thus, using $z = (x - \mu)/\sigma$, we see that

$$-2.054 = \frac{414 - \mu}{8}.$$

Solving for μ, we get $\mu = 414 + 2.054 \cdot 8 = 430.4$.

33. **(D)** The confidence interval is given by $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, and the margin of error is

$z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. So in order to make the margin of error small, we need to minimize z^* and maximize n. The larger value of n available to us is 500, and z^* will be minimized by taking the smallest available confidence level – so we choose 95%.

34. **(C)** This question is about blocking. The belief is that the younger a person is, the more likely he is to respond to a drug of this sort – so it makes sense to block by age. This is done by forming a block consisting of the three young men, a block consisting of the three middle-aged men, and a block consisting of the three elderly men. The three treatments are then randomly assigned within each block. The reason this is helpful is that we are now giving drug A to one young man, one middle aged man, and one elderly man; and the same for each of the other two drugs. So, in terms of age, the treatment groups are very similar. Therefore, if, for example, the men treated with drug A do better than the others, we can more readily attribute this to the greater effectiveness of drug A, and not to differences in age.

Another way of looking at this, is that in each of the procedures described in (A), (B), (D), and (E), there is some possibility that the treatment groups will differ from each other in terms of age. Then, if, for example, one treatment group responds significantly better, on average, than another treatment group, it would be impossible to tell whether the difference comes about as a result of the greater effectiveness of one of the drugs, or of the age differences between the groups.

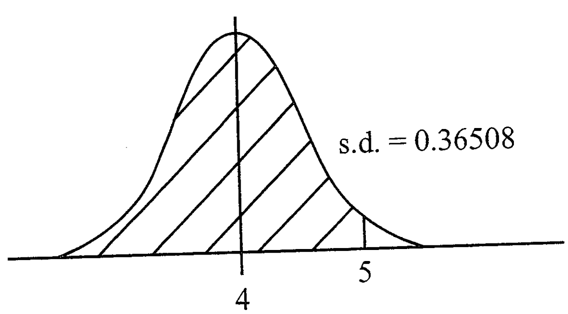
35. **(A)** We are given that the p-value is 0.063, which is not particularly small (this being judged by the fact that 0.063 is greater than the significance level, 0.05). This means that if the population mean were 40,000, then it would not be particularly unlikely that the sample mean would be as far away from 40,000 as the one obtained in the study. Thus, we do not have sufficient evidence to conclude that the mean life of the tires is not equal to 40,000, and option (A) is correct. Note that we never "accept H_0," so option (C) is incorrect. (In this context, accepting H_0 would mean that we have sufficient evidence to conclude that the mean life of the tires was actually equal to 40,000 miles *exactly*. This is very, very unlikely to be the case.) Turning our attention to (B), the alternative hypothesis in the question tells us that a two-tailed test is being used. The conclusion in (B) is the conclusion for the equivalent *one*-tailed test, and is therefore incorrect.

36. **(B)** The quantity $\sigma/\sqrt{n}$ is the standard deviation of the sampling distribution of the sample mean.

37. **(B)** Firstly, increasing the significance level, α, increases the probability of a Type I error and decreases the probability of a Type II error. So we choose $\alpha = 0.05$. Secondly, the probability of a Type II error is reduced by increasing the sample size, so we choose $n = 40$. Thirdly, looking at the true value of μ, we want to reduce the probability of a Type II error, so we want to increase the probability of correctly rejecting H_0 in favor of H_a, which says $\mu < 50$. The further the true value of μ is to the left of 50, the more likely we are to reject H_0. So we select $\mu = 46$.

38. **(E)** The confidence interval for the slope is $b \pm t^* \cdot s_b$. First, b is the slope of the regression line, and its value is given by the table to be 0.264. Second, t^* is the critical value of t for a 99% confidence level with $n - 2 = 22$ degrees of freedom, which the t-table provided with the exam gives to be 2.819. Third, s_b is the standard error of the slope of the regression line, whose value is given in the table as 0.109.

39. **(E)** Since the sample size is large ($n = 40$), the Central Limit Theorem tells us that the distribution of the sample mean, $\overline{x}$, is approximately normal. The mean of the sampling distribution of $\overline{x}$ is $\mu = 4$ and the standard deviation is $\sigma/\sqrt{n} = 2.309/\sqrt{40} = 0.36508$. We need $P(\overline{x} < 5)$.



For $\overline{x} = 5$, $z = (5 - 4)/0.36508 = 2.739$. So $P(\overline{x} < 5) = P(Z < 2.739) = 0.997$. (This answer is obtained on the TI-83 and TI-84 calculators by entering normalcdf(-999,2.739) or normalcdf(-999,- 2.739,0,1) and on the TI-Nspire by entering normalCdf(-999, 2.739,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

40. **(A)** The random pattern in the residual plot obtained by plotting $\log y$ against $\log x$ indicates a linear relationship between $\log y$ and $\log x$. At this point, we can either use memory of the fact that this linear relationship indicates a power relationship between y and x, that is $y = ax^p$, or we can use the following mathematical argument:

The linear relationship between $\log y$ and $\log x$ can be written in the form

$$\log y = A + B \log x.$$

Using the definition of logs: $y = 10^{A+B\log x}$

Using $x^{a+b} = x^a \cdot x^b$: $y = 10^A \cdot 10^{B\log x}$

Using $B \log x = \log x^B$: $y = 10^A \cdot 10^{\log x^B}$

Using $10^{\log x^B} = x^B$: $y = 10^A \cdot x^B$

Writing 10^A as a and B as p, we have the relationship given in (A).

Section II—Part A

Question One

(a) Outliers are values that are more than $1.5 \cdot (\text{IQR})$ above Q_3 or more than $1.5 \cdot (\text{IQR})$ below Q_1. Here, $\text{IQR} = Q_3 - Q_1 = 667.5 - 463 = 204.5$. So $Q_3 + 1.5 \cdot (\text{IQR}) = 667.5 + 1.5 \cdot (204.5) = 974.25$. The maximum file size is given to be 774, and 774 is less than 974.25, therefore there is no outlier at the upper end of the distribution. However, $Q_1 - 1.5 \cdot (\text{IQR}) = 463 - 1.5 \cdot (204.5) = 156.25$. The minimum file size is given to be 99 MB, which is smaller than 156.25, so there is at least one outlier at the lower end of the distribution.

[*There are other methods of identifying outliers that are also acceptable.*]

(b) Since maximum − median = 774 − 566 = 281 and median − minimum = 566 − 99 = 467, and 467 is greater than 281, the left tail is longer than the right tail and the distribution is skewed to the left.

(c) The table tells us that 463 MB is the first (lower) quartile file size and that 667.5 is the third (upper) quartile file size. Therefore, approximately 25% of the files have sizes less than 463 MB and approximately 75% of the files have sizes less than 667.5 MB. From this we can deduce that approximately 50% of the files have sizes between 463 MB and 667.5 MB.

(d) The standard deviation, 167.26 MB, is a typical deviation of the file sizes in this data set from the mean file size.

[*Did you remember to put your interpretation of the standard deviation in context? An interpretation without the words "file size" would not get full credit!*]

Question Two

(a) $r = \pm\sqrt{0.82} = \pm 0.906$. Since the prices are declining as the age increases (or, alternatively, since the slope of the regression line is negative), the correlation is negative, and so $r = \mathbf{-0.906}$. There is a strong negative linear relationship between the age and price of this type of car.

[*Did you include strength, direction, linearity, and context?*]

(b) The equation of the regression line is $\hat{y} = 25844.789 - 4764.155x$, where x is the age of the car and $\hat{y}$ is the predicted price.

[*Definition of x and $\hat{y}$ is required.*]

The slope is -4764.155, which tells us that when the age of the car increases by one year, the predicted price decreases by $4764.155.

[*Note the word "predicted." Words like "average," "expected," or "estimated" would also be acceptable.*]

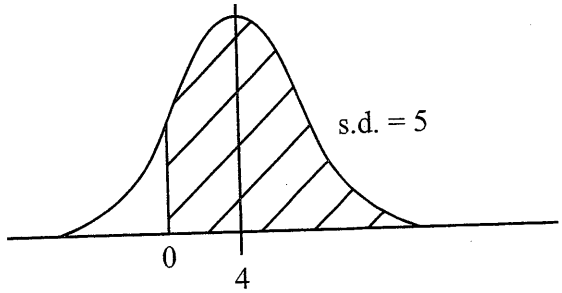
(c) The intercept is 25844.789. The predicted price when the age is zero is $25844.789. Zero is outside the range of x-values on which the regression line is based, so there is no reason to think that the regression line should apply when the age is zero. So this value ($25844.789) cannot be usefully applied to the prices of this type of car.

(d) The residual plot shows a random scatter, therefore a straight line is appropriate for modeling the prices of cars of this type that are at least one year and at most three years old.

Question Three

(a) The two funds have the same expected value of the monthly gain, but Fund A has the lower standard deviation, and so is likely to be less erratic – a more reliable form of income than Fund B.
[*Alternatively you could say that since Fund A has the smaller standard deviation, this fund is less likely to make a loss. Whatever your approach, you* do *need to mention that in terms of their means, the two funds are equal.*]

(b) We need the gain for Fund A to be greater than zero.



For a zero gain, the z-value is given by $z = (0-4)/5 = -0.8$. So the probability that Fund A gains money is $P(Z > -0.8) = 0.788$. (This answer is obtained on the TI-83 and TI-84 calculators by entering normalcdf(-999,-.8) or normalcdf(-999, -.8,0,1), and on the TI-Nspire by entering normalCdf(-999, -.8,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

(c) The number of months for which Fund A gains money during the next four months is binomially distributed with $n = 4$ and $p = 0.78814$. So the required probability is

$$\binom{4}{2}(0.78814)^2(0.21186)^2 = \mathbf{0.167},$$

(In the calculation above, the expression $\binom{4}{2}$ is another way of writing ${}_4C_2$. Using the TI-83 and TI-84 calculators, the value of ${}_4C_2$ is found by typing 4 nCr 2. The nCr function is found by selecting MATH,PRB,3. On the TI-Nspire (in the *Scratchpad* or *Calculator* application), this value is found by typing nCr(4,2), and the nCr function is found by selecting menu,Probability,Combinations.
Alternatively, the answer to the question can be found directly, using a calculator. On the TI-83 or TI-84 calculators, use binompdf(4,.78814,2). The binompdf function can be accessed by 2nd,DISTR. On the TI-Nspire (in the *Scratchpad* or *Calculator* application), use binomPdf(4,0.78814,2). The binomPdf function can be found by selecting menu,Statistics,Distributions.)

(d) Let the gain in the first month on a $1000 investment in Fund A be X, and let the equivalent gain for Fund B be Y. Then the gain for this investor is $Z = 8X + 3Y$.

So $\mu_z = 8\mu_X + 3\mu_Y = 8 \cdot 4 + 3 \cdot 4 =$ **$44**

and $\sigma_Z = \sqrt{8^2\sigma_X^2 + 3^2\sigma_Y^2} = \sqrt{8^2 \cdot 5^2 + 3^2 \cdot 18^2} =$ **$67.20**.

Question Four

(a) Print the names of all 50 gymnasts on pieces of paper and place them in a hat. Then pick at random 25 names from the hat. These 25 gymnasts will form the group that receives the new treatment, and the remainder will be in the group that receives the current treatment.

[*Be sure to specify which group the first 25 names will go into. Very often students will say they go into "one group" without saying which group this is. Also, note that the method where you flip a coin for each gymnast and stop when you have 25 in one of the groups is not completely random assignment. For further details, please see*

"Know how to achieve random assignment" in Part II, Section 3 of "Top Tips for AP Statistics" in the Question Book.]

(b) If some of the gymnasts were allowed to choose which group they would go into then it is quite possible, for example, that the more enthusiastic gymnasts who practice more would be keen to try out the new treatment. So then any greater toughness in the new treatment might be masked by the fact that those gloves were used more heavily. Random assignment produces two groups of subjects that are expected to be similar in all respects.

[*Be careful not to suggest that random assignment might eliminate differences between the groups.*]

(c) Despite the random assignment, in the first design there can be large differences between the individuals in the two groups and therefore the heaviness of use could differ greatly between the two treatments. The effect on the results of any difference in the toughness of the two treatments cannot be distinguished from the effect of the differences in the heaviness of use, and therefore any difference in toughness is harder to detect. However, under the second design, the two treatments will be subjected to almost identical use, and so any differences in the wear on the two types of glove can be attributed almost exclusively to the differences in the effectiveness of the two treatments. (The random assignment of the gloves to the hands means that any factors resulting from use by dominant or non-dominant hands are expected to be equally distributed between the two treatments.)

[*Remember – when comparing two things you must consider both in what you write.*]

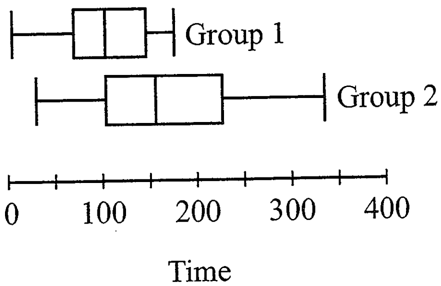
Question Five

Hypotheses:
Let μ_1 and μ_2 be the population mean times when other participants are not present and present, respectively.
$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 < \mu_2$

Check of conditions and naming of type of test:
This is a two-sample t-test.

The conditions for this two-sample t-test are:
1. Random assignment of subjects to treatments
2. Normally distributed populations or large treatment groups

We are told that the groups were assigned randomly. We do not have large treatment groups and so we will consider boxplots of the values for the groups.



Since the distributions of the group times are roughly symmetrical and there are no outliers in either distribution, it appears reasonable to assume that the populations are normally distributed.

Mechanics:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{145.5 - 220.063}{\sqrt{\frac{51.419^2}{16} + \frac{89.925^2}{16}}} = -2.879.$$

We use the t distribution with 23.861 degrees of freedom (given by the calculator). So the p-value is $P(t_{23.861} < -2.879) = 0.004$. (This p-value can be obtained on the TI-83, TI-84, and TI-Nspire calculators by entering tcdf(-999,-2.879,23.861). The tcdf function is accessed on the TI-83/84 by 2nd,DISTR, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) tCdf is accessed by selecting Menu,Statistics,Distributions. In order to find the number of degrees of freedom given above, you can use the statistical tests facility of a calculator. On the TI-83 and TI-84 calculators, first select STAT,Edit and enter the results for the two groups into L1 and L2. Then select STAT,TESTS,2-SampTTest. For Inpt select Data, for List1 select L1, for List2 select L2, for Freq1 select 1, and for Freq2 select 1. Select <μ2 for the alternative hypothesis, select No for Pooled, and then select Calculate and press Enter. All the statistics relevant to the test are now displayed. On the TI-Nspire, from the home screen tab to the Add Lists & Spreadsheet icon. Press enter, and enter the two data sets into lists A and B. Name the lists List_1 and List_2. Then select menu, Statistics, Stat Tests, 2-Sample t Test. Select Data for the data input method and tab to OK. For List 1 click the down arrow and select List_1. For List 2 click the down arrow and select

List_2. For Frequency 1 enter 1 and for Frequency 2 enter 1. For the alternative hypothesis enter Ha: $\mu1 < \mu2$, for Pooled select No. Then tab to OK, press enter, and the results of the hypothesis test are displayed in columns C and D of the table.)

Conclusion:
Since the *p*-value is 0.004, which is less than 0.05, we reject H_0 We have statistically significant evidence that the presence of other participants brings about a higher mean time for this task.

Section II—Part B

Question Six

(a) Hypotheses:
Let p be the proportion of students at Central HS who are in favor of the idea.
$H_0: p = 0.5$
$H_a: p > 0.5$

Check of conditions and naming of test:
This is a one-sample z-test for a proportion.

The conditions for this test are:
1. Simple random sample
2. np_0 and $n(1 - p_0)$ both greater than 10.

We are told that the sample of students is random. Moreover, $np_0 = 40 \cdot 0.5 = 20$ and $n(1 - p_0) = 40 \cdot 0.5 = 20$, which are both greater than 10.

Mechanics:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{26/40 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{40}}} = 1.897.$$

So the p-value is $P(Z > 1.897) = 0.029$. (This p-value is obtained on the TI-83 and TI-84 calculators by entering normalcdf(1.897,999) or normalcdf(1.897,999,0,1) and on the TI-Nspire by entering normalCdf(1.897,999,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

Conclusion:
Since the p-value is 0.029 which is less than 0.05, we reject H_0. We have statistically significant evidence that a majority of the students at Central HS are in favor of the idea.

(b) 1-3 will denote "D" (0 point), 4-6 will denote "N" (1 point), and 7-9 will denote "A" (2 points); ignore 0's.
[*Any assignment of digits that gives the correct probabilities is acceptable.*]

(c) By my assignment, and starting in the top left corner of the table and moving to the right, the responses are:
N (1) N (1) A (2) D (0) A (2) A (2) N (1) D (0) N (1) D (0) A (2) A (2) N (1) A (2) A (2) N (1) D (0) A (2) N (1) N (1),
giving a total score of 24.

(d) The 200 runs of the simulation give us an approximate probability for each of the total scores, given that the three responses are favored by equal proportions of the school. So, for example, $P(\text{total score} = 30) \approx 1/200 = 0.005$. Therefore, considering the results of the 200 runs of the simulation, if the three responses are favored by equal proportions of the school, the probability of obtaining a total score of at least 24 in a sample of 20 students is approximately 30/200 = 0.15, which is not particularly small. It is therefore plausible, given that the three responses are favored by equal proportions of the school, to get a sample of 20 students with a total score of 24. Thus we do not have convincing evidence that the student body as a whole is in favor of the idea.

Sample Examination Four – Answers

1. **(B)** $(20/100)\cdot 24 = 4.8$. So the 20th percentile travel time is approximately the 5th time, counting from the smallest to the largest. So the 20th percentile time is approximately 40 minutes.

2. **(B)** We need the probability of success to be 1/3. In the strategy in (B), three digits represent success and six digits represent failure, and so this gives a probability of success that is equal to 1/3, as is required; thus, (B) is the correct response.

   In the strategy in (A), three digits represent success and seven digits represent failure. This gives a probability of success that is equal to 3/10, which is incorrect. In the strategy in (C), 33 digit-pairs represent success and 67 digit-pairs represent failure. This gives a probability of success that is equal to $33/100 = 0.33$, which is incorrect. In the strategy in (D), 30 digit-pairs represent success and 70 digit-pairs represent failure. This gives a probability of success that is equal to $30/100 = 0.3$, which is incorrect. In the strategy in (E), 334 three-digit numbers represent success and 666 three-digit numbers represent failure. This gives a probability of success that is equal to $334/1000 = 0.334$, which is incorrect.

3. **(E)** Whenever possible, it is preferable to perform a census (that is, gather information about the whole population) rather than take a sample. The strategy in (E) achieves this.

4. **(E)** The only way to establish causation is by performing an experiment. Randomly assigning some teenagers to play video games and some not to play video games constitutes an experiment; all the other options are modifications of the observational study described in the question.

5. **(E)** The quantity $36/50 = 0.72$ is the *sample* proportion, and so should not be used in the hypotheses. The alternative hypothesis should be $p > 0.65$, since the proportion of employees who eat the meals is suspected to have *increased* beyond the proportion of 0.65 that has been assumed.

6. **(A)** r^2 is the proportion of the variation in y that can be explained by the least squares regression line of y on x.

7. **(D)** The data given are *paired*, and so of the three options given, only the paired t-test and the t-test for the slope of the regression line can be applied. The two-sample t-test is used when you have two independent random samples, which is not the case here.

8. **(D)** The number of sixes rolled has a binomial distribution with $n = 5$ and $p = 1/6$. So, denoting the number of sixes rolled by X, $P(X = 1) = {}_5C_1(1/6)(5/6)^4 = 0.4019$, and $P(X = 2) = {}_5C_2(1/6)^2(5/6)^3 = 0.1608$. Therefore, $P(X = 1 \text{ or } X = 2) = 0.4019 + 0.1608 = 0.5627$.

9. **(E)** In order to use the t distribution we have to know either that the population is normally distributed or that the sample is large (usually requiring $n > 30$). Here $n = 15$, and so we examine the distribution of the sample values in order to check that it is feasible that the population is normally distributed.

10. **(C)** It is clear from the boxplot that the distribution is positively skewed. Hence the mean is greater than the median. It can probably be seen intuitively that the mean cannot be as large as 22, but to prove this, consider the fact that the five number summary (read from the boxplot) is 0, 5, 10, 20, 30. Since there are 23 measurements, this tells us that the 1st, 6th, 12th, 18th, and 23rd values are 0, 5, 10, 20, and 30, respectively. In other words, the dataset, in order from the smallest to the largest, is:

$$\underline{0} * * * * \underline{5} * * * * * \underline{10} * * * * * \underline{20} * * * * \underline{30},$$

where the stars represent any numbers, so long as the list remains in numerical order. Therefore the largest possible value of the mean is achieved by the following dataset:

$$\underline{0}\ 5\ 5\ 5\ 5\ \underline{5}\ 10\ 10\ 10\ 10\ 10\ \underline{10}\ 20\ 20\ 20\ 20\ 20\ \underline{20}\ 30\ 30\ 30\ 30\ \underline{30}.$$

The mean of this set of values is 15.43, and so this is the largest possible mean for a dataset of size 23 having the boxplot given in the question.

11. **(D)** The confidence interval for the difference of two proportions is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

The value of z^* is 2.576. (This value is given by invNorm(0.995) or invNorm(0.995,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.995,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) Therefore, since $\hat{p}_1 = 83/120$, $\hat{p}_2 = 78/140$, $n_1 = 120$ and $n_2 = 140$ the interval given in (D) is correct.

Note that (C) and (E) make use of the proportion of germinating seeds for the *combined* samples. This is not valid in the context of confidence intervals.

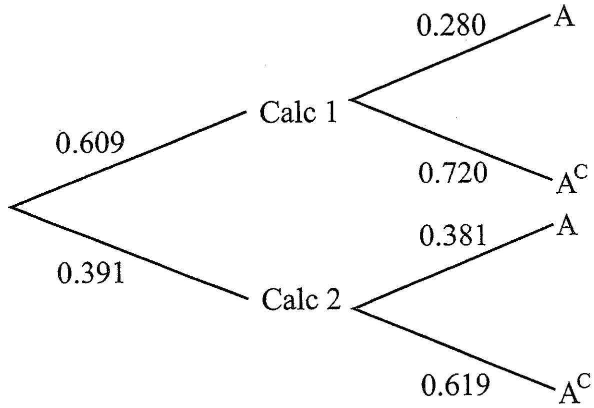
12. **(E)** We are testing whether the three category proportions are the same for the two populations. This is a chi-square test for homogeneity.

Note that the tests given in (A) and (D) are applicable to studies involving a *single* sample. This study involves two samples, and therefore these responses are incorrect. Options (B) and (C) are tests for means. This study concerns proportions, and so these responses are incorrect.

13. **(D)** Replication is exactly as described in option (D): using enough subjects so that chance variations between the groups become negligible.

14. **(E)** The distribution with the smallest standard deviation is the one with the greatest concentration close to its mean. This criterion is satisfied by the distribution represented by the histogram in (E), since the great majority of the data values are at most one unit from the mean. All of the other options show deviations from the mean that are, on average, greater than those in (E).

15. **(C)**



$$P(\text{Calc 1} \mid \text{A}) = \frac{P(\text{Calc 1} \cap \text{A})}{P(\text{A})} = \frac{0.609 \cdot 0.28}{0.609 \cdot 0.28 + 0.391 \cdot 0.381} = 0.534.$$

16. **(D)** The computer output gives "Prob < t" as 0.0297: this is the p-value for the one-tailed test specified in the question. It is also clear from the output that the students' scores have decreased on average, since the PostTest mean is less than the PreTest mean. So we have evidence at the 5% level, but not the 1% level, that the students' mental arithmetic abilities have decreased. Thus, option (D) is correct (and option (E) is incorrect).

    Options (A) and (B) are incorrect since the PostTest mean is less than the PreTest mean, and so we cannot possibly have evidence that on average the students' mental arithmetic scores have increased. Turning our attention to (C), we note that this option refers to the equivalent *two-tailed* test. The p-value for the one-tailed test is 0.0297, and so the p-value for the two-tailed test would be 2(0.0297) = 0.0594. Hence we would not have evidence at the 5% level of a change in the students' average score; thus, option (C) is incorrect.

17. **(B)** $\frac{\sigma}{\sqrt{n}}$ is the standard deviation of the distribution of the sample *mean*, not the standard deviation of the sample.

    Let us now look at why the statements given in the other options are correct (making these incorrect responses). The distribution of $\overline{x}$, the sample mean, has mean μ and standard deviation $\sigma/\sqrt{n}$; thus, the statements in (A), (D), and (E) are correct. Moreover, the Central Limit Theorem tells us that for large values of n, the distribution of $\overline{x}$ is approximately normal. Hence the statement in (C) is correct.

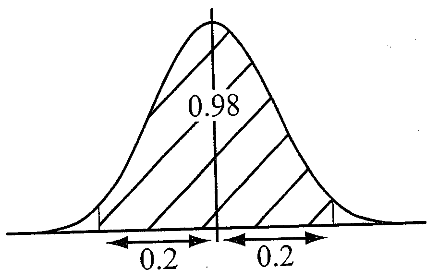
18. **(A)** To say that the distribution is negatively skewed means that it has a longer tail to the left than to the right. Option (A) is the only response that satisfies this. Note that (D) and (E) are incorrect as they tell us nothing about the *spread* of relatively high scores, as compared to the *spread* of relatively low scores. Note, also, that (C) is incorrect, since the histogram of a negatively skewed distribution will have its highest point towards the right, not the left.

19. **(A)** There is no reason to conclude that the study was conducted poorly. Type I and Type II errors occur by *chance*. Typically, when a Type I or II error occurs, the survey or experiment is well designed and conducted, and the results are correctly analyzed, however it happens by chance that the results lead to the correct conclusion not being reached by the significance test.

Let us now look at why the statements in (B)-(E) are correct (making these responses incorrect). A Type II error is failing to reject H_0 when H_0 is false, and this is failure to reach a correct conclusion; thus, the statements in (B), (C) and (D) are correct. Turning our attention to (E), we note that a Type I error is rejecting H_0 when H_0 is *true*. We are told here that a Type II error has occurred, and therefore we know that H_0 is *false*. Therefore, a Type I error cannot have occurred, and the statement in (E) is true.

20. **(B)** The difference between experiments and observational studies is that in experiments the conditions are *imposed* on the participants, whereas in observational studies the participants have made their own choices as to what they do. Thus, (B) is the correct response.

21. **(A)**



The areas of the left and right tails in the diagram are 0.01. From this we find that the z-value for the upper bound of the shaded region in the diagram is 2.326. (This value is given by invNorm(0.99) or invNorm(0.99,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.99,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) The quantity z is the number of standard deviations from the mean. Hence $0.2 = 2.326\sigma$, which gives $\sigma = 0.086$.

22. **(C)** $t = \dfrac{b-\beta}{s_b} = \dfrac{4.169-0}{2.142} = 1.946.$

23. **(B)** Bias lies in the sampling *method*. Bias is present when something in the sampling *design* makes it likely that the sample will over- or under-emphasize some relevant characteristic of the population. Thus, (B) is the correct response. The phenomenon in (A) – getting a sample that over-emphasizes some characteristic of the population – can occur purely as a result of sampling variability, and is therefore not indicative of bias.

24. **(B)** The z-score for a particular value tells us how many standard deviations that value is above (or below) the mean, with negative z-scores signifying values below the mean. Therefore, with a z-score of -0.290, this item is 0.290 standard deviations below the mean. Hence the price of the item is $3.4573 - 0.290 \cdot 1.613 = 2.99$.

25. **(C)** The statement in (C) is a correct interpretation of the 95% confidence level. Looking at the other options, remember that any mention of the particular confidence interval that has been calculated – as in (A) and (B) – makes an attempted interpretation of the confidence level incorrect. Considering (D), note that the confidence interval that has been constructed gives us some information about the mean of the *population.* Any attempt to apply the confidence interval to a *sample* mean, as in (D), is incorrect. Finally, considering (E), note, again, that the confidence interval tells us something about the population *mean.* Any attempt to relate the interval to the actual population measurements without reference to the population mean, as in (E), is incorrect.

26. **(C)** Recall that when X and Y are independent random variables and a and b are constants, $\sigma_{aX+bY} = \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}$. Here, if X is the amount (in grams) of chips and Y is the amount (in grams) of salsa, then the number of calories is $5X + 0.5Y$. Therefore the standard deviation of the number of calories is $\sqrt{5^2(3.7)^2 + 0.5^2(2.8)^2} = 18.55$.

27. **(A)** The total score for the girls is $12 \cdot 8.25 = 99$. The total score for the boys is $20 \cdot 7.3 = 146$. There are $12 + 20 = 32$ students altogether. Therefore, the mean score for all of the students is

$$\frac{99 + 146}{32} = 7.66.$$

28. **(C)** The probability that the largest score is a six is the probability that at least one six is rolled, and this probability is $1 - P(\text{no sixes}) = 1 - \left(\frac{5}{6}\right)^4 = 0.518$.

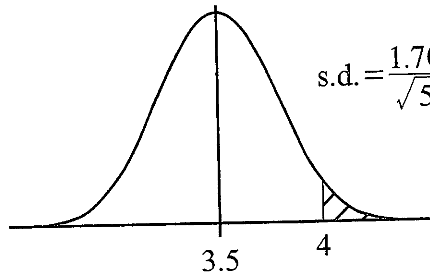
29. **(B)** Since 18 out of the 30 items of clothing in the sample have been worn, the sample proportion is $\hat{p} = 18/30 = 0.6$. Denoting the proportion of all returned items of clothing that have been worn by p, we need to test the null hypothesis $H_0: p = 0.5$ against the alternative hypothesis $H_a: p > 0.5$. We do this using a one-proportion z-test, for which the test statistic is

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.6 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{30}}} = 1.095.$$

Therefore the p-value for the test is given by $P(Z > 1.095) = 0.137$. (This value is obtained on the TI-83 and TI-84 calculators by entering normalcdf(1.095,999) or normalcdf(1.095,999,0,1) and on the TI-Nspire by entering normalCdf(1.095,999,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

30. **(B)** $s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{3.24 + 3.24 + 17.64 + 0.04 + 0.64}{4}} = 2.49.$

31. **(B)** For 50 rolls of the cube, the probability that the total score is greater than or equal to 200 is the probability that $\bar{X} \geq 4$, where $\bar{X}$ is the mean score for the 50 rolls. Since $n = 50 > 30$, the distribution of $\bar{X}$ is approximately $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(3.5, \frac{1.708}{\sqrt{50}}\right)$. So $P(\bar{X} \geq 4)$ is approximately the shaded area in the diagram below.



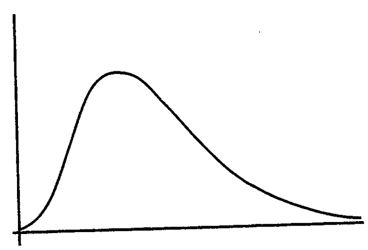
The z-value for the lower bound of the shaded region is $z = \frac{4-3.5}{\left(1.708/\sqrt{50}\right)} = 2.06998$. So approximately $P(\overline{X} \geq 40) = P(Z \geq 2.06998) = 0.019$ (This value is obtained on the TI-83 and TI-84 calculators by entering normalcdf(2.06998,999) or normalcdf(2.06998,999,0,1) and on the TI-Nspire by entering normalCdf(2.06998,999,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

32. **(C)** The 95% confidence interval is given by $\overline{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$. Therefore, when n is multiplied by 4, the width of the confidence interval is multiplied by 1/2.

33. **(A)** We will denote the y-intercept and the slope of the regression line by a and b, respectively. We are given $\overline{x} = 70.2$, $\overline{y} = 160.3$, $b = 2.75$. The formula sheet gives $b_0 = \overline{y} - b_1\overline{x}$, and so in our notation, $a = \overline{y} - b\overline{x} = 160.3 - 2.75 \cdot 70.2 = -32.75$. Hence the equation of the regression line is $\hat{y} = -32.75 + 2.75x$. Substituting $x = 68$, we get $\hat{y} = -32.75 + 2.75 \cdot 68 = 154.25$.

34. **(B)** As indicated in the hypotheses, the hypothesis test for the slope of the regression line is also a test for correlation. The computer output gives the p-value for this test as 0.094, which is the p-value for the two-tailed test. (The p-value given by a computer can be assumed to be the one for a two-tailed test unless it is specified as being for the one-tailed version.) Therefore, since $0.094 > 0.05$, we do not have sufficient evidence at the 5% level to conclude that there is a nonzero correlation between the results of the two methods. Note that (C) is "accepting H_0," which is never correct.

35. **(A)** The probability density function of the chi-square distribution with 8 degrees of freedom is shown in the diagram below.



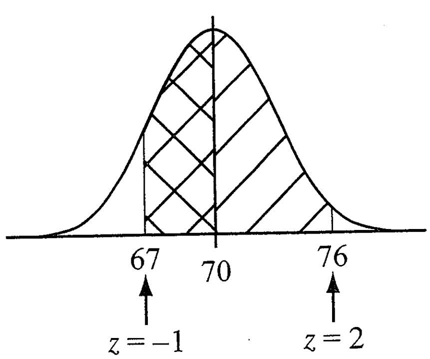
This distribution is not symmetrical. The distributions given in options (B)–(E) are all symmetrical.

36. **(C)** This question is about the placebo effect. The point of any experiment is to keep all variables equal (or as close as possible) between the two groups except for the one being tested, which here is the chemical in the drug. But in the design given, there is an additional difference between the two groups in that the members of Group A are taking a tablet whereas those in Group B are not – and we know that the mere taking of a tablet can have a positive effect. Hence (C) provides a strong argument against the claim. Option (A) is incorrect since the mathematics involved in significance testing always takes into account the sizes of the groups/samples. Looking at (B), we have no reason to think that the people in Group A might be eating less than the people in Group B since the groups are chosen at random, unless the taking of the tablets is causing a reduction in eating, which would in fact be a weight-reducing effect of the tablets. The warmer weather cited in (D) would affect the experimental subjects in both groups, and so cannot account for a greater weight loss on average in Group A. Finally, considering (E), if some of the people in Group A had not taken the drug (which in fact constitutes a breakdown in the experimental procedure), then the observed weight loss for Group A would most likely have been caused by a very significant weight loss in those in Group A who did take the drug. This would reinforce the evidence for the effectiveness of the tablets.

37. **(B)** First recall, for example, that zero being contained in the 95% confidence interval is equivalent to the non-rejection of H_0 at the 5% level for the two-tailed test. Here the p-value for the *one*-tailed test is given as 0.043, and so the p-value for the two-tailed test is 0.086. Therefore H_0 is rejected at the 10% level, but is not rejected at the 8% level. Equivalently, zero is not contained in the 90% confidence interval but *is* contained in the 92% confidence interval. Thus, (B) is the correct response.

38. **(D)** The computer output gives the equation of the regression line to be $\hat{y} = 602.509 + 24.927x$, where x is the size and $\hat{y}$ is the predicted rent. So for $x = 32$ we get $\hat{y} = 602.509 + 24.927 \cdot 32 = 1400.173$. Therefore, for the point (32,1450) the residual is $y - \hat{y} = 1450 - 1400.173 = 49.827$. Option (D) gives the closest value to this result.

39. **(B)**



We require $P(X < 70 \mid 67 < X < 76) = \dfrac{P(X < 70 \text{ and } 67 < X < 76)}{P(67 < X < 76)}$.

Now, saying that $X < 70$ <u>and</u> $67 < X < 76$ is equivalent to saying that $67 < X < 70$. So we require

$$\frac{P(67 < X < 70)}{P(67 < X < 76)} = \frac{P(-1 < z < 0)}{P(-1 < z < 76)} = 0.417,$$

(This value is obtained on the TI-83 and TI-84 calculators by entering normalcdf(-1,0)/ normalcdf(-1,2) or normalcdf(-1,0,0,1)/ normalcdf(-1,2,0,1) and on the TI-Nspire by entering normalCdf(-1,0,0,1)/normalCdf(-1,2,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

40. **(A)** Consider the chi-square goodness of fit test of H_0: the four outcomes on the spinner are equally likely *versus* H_a: the four outcomes on the spinner are not equally likely. The quantities f_1, f_2, f_3, and f_4 are the observed frequencies of the four outcomes and, assuming H_0 to be true, the expected frequencies are all 30. Therefore, the quantity

$$Q = \frac{(f_1 - 30)^2}{30} + \frac{(f_2 - 30)^2}{30} + \frac{(f_3 - 30)^2}{30} + \frac{(f_4 - 30)^2}{30}$$

is the chi-square statistic $\sum \frac{(O - E)^2}{E}$, where O represents the observed frequency and E represents the expected frequency.

We know H_0 is true, since the question states that the four outcomes are equally likely. Also, the conditions for the chi-square goodness of fit test are satisfied since the expected frequencies are all greater than 5. So the distribution of Q is chi-square with 3 degrees of freedom (since there are 4 categories in this goodness of fit test).

So the required probability is $P(\chi_3^2 > 6.25) = 0.1001$. (On the TI-83, TI-84, and TI-Nspire calculators, this result is obtained using χ^2cdf(6.25,999,3). On the TI-83/84, the χ^2cdf function is accessed by 2nd,DISTR, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) the χ^2Cdf function is accessed by selecting menu,Statistics,Distributions.) Thus, the correct response is (A).

Section II—Part A

Question One

(a) Approximately 91% of the country's residents are less than or equal to 70 years old.

(b) Looking at the graph, the cumulative relative frequency for Age = 40 is approximately 0.58. This tells us that the proportion of residents over 40 years old is approximately $1 - 0.58 = 0.42$. Therefore the number of residents over 40 years old is approximately $32465233 \cdot (0.42) \approx$ **13.6 million**.

[Note that you do not have to get this answer, exactly. So long as you have shown your work on the graph and your calculations are consistent with the work shown you will receive full credit. This applies to parts (a)-(c) of this question.]

(c) Using the graph, the age for cumulative relative frequency = 0.75 is approximately 52, and the age for cumulative relative frequency = 0.25 is approximately 17.5. So the interquartile range is approximately $52 - 17.5 =$ **34.5 years.**

(d) The fact that the graph is approximately a straight line for 0-50 on the horizontal axis tells us that the numbers of residents in the age groups 0-5, 5-10, 10-15, and so on up to 45-50, are all approximately equal.

[To understand this, note first that the vertical increment between, for example, Age = 10 and Age = 15 on the cumulative relative frequency graph is the proportion of residents who are between 10 and 15 years old. The fact that the graph is approximately a straight line tells us that these vertical increments are approximately equal, and therefore that the proportions of residents in the age groups 0-5, 5-10, and so on up to 45-50, are equal. It follows that the numbers of residents in these age groups are equal.]

Question Two

(a) The random pattern in the residual plot tells us that the use of a straight line to model the relationship between log(number of subscribers) and Years since 1920 is appropriate.

[*Note: Do not attempt to use the value of r^2 to answer this question. The value of r^2 (or of r) can tell you how close the points are to being in a straight line, but it cannot tell you whether a straight line is an appropriate model.*]

(b) The equation of the regression line is:
Predicted $\log(\text{number of subscribers}) = 3.364928 + 0.0315783(\text{Yrs since } 1920)$.

[*Note that if you use symbols such as x and y in your equation, then you must say what these symbols represent.*]

(c) Substituting Yrs since $1920 = 110$, we get
$\log(\text{number of subscribers}) = 3.364928 + 0.0315783(110) = 6.838541$.

Therefore, number of subscribers $= 10^{6.838541} =$ **6,895,107**.

[*Note that different answers will be obtained according to the degree of rounding in the work. If your work is correct, and your answer agrees with your work, then you will receive full credit.*]

(d) The year 2030 lies outside the range of years in the original data set, and we have no reason to believe that the linear relationship between log(number of subscribers) and Yrs since 1920 will apply outside that range. Therefore, using this regression model to predict the number of subscribers in 2030, known as extrapolation, is not reasonable.

Question Three

(a) It could be argued that people who feel strongly against the increased police presence are more likely to respond than those who are neutral or in favor of the policy, in which case the survey would be likely to overestimate the level of opposition to the policy.

(b) The group could send the survey to a large random sample of residents. After a period of time, they could make follow-up phone calls to those residents who have not responded, in order to obtain responses from all of the residents in the sample.

(c) Let the proportion of residents in opposition to the policy be p. We estimate p using the sample proportion, $\hat{p}$. If the sample size, n, is large (we hope to get an answer that satisfies the condition that np and $n(1-p)$ are both greater than 10), then the distribution of $\hat{p}$ is approximately normal with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. For any normally distributed random variable, with 95% probability the random variable will be within 1.96 standard deviations of the mean. (This value is given by invNorm(0.975) or invNorm(0.975,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.975,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

So, with 95% probability, $\hat{p}$ will be within $1.96\cdot\sqrt{\frac{p(1-p)}{n}}$ of p.
Therefore, referring to the question, we need

$$1.96\cdot\sqrt{\frac{p(1-p)}{n}}<0.02.$$

From this we get

$$\sqrt{n}>\frac{1.96}{0.02}\sqrt{p(1-p)},$$

and so, squaring both sides, we see that we require

$$n>\left(\frac{1.96}{0.02}\right)^2 p(1-p).$$

We need a value of n that will satisfy this inequality for all values of p in the range $0.32\le p\le 0.45$. Experimenting with different values of p, we see that $p(1-p)$ is largest when p is as close as possible to 0.5, in other words when $p=0.45$.

So we need $n>\left(\frac{1.96}{0.02}\right)^2\cdot 0.45\cdot(1-0.45)=2376.99$

So $n=$ **2377** is the sample size required.

Question Four

(a)

Sample	Sample Mean	Sample Midrange
1, 2, 2*, 3	2	2
1, 2, 2*, 3*	2	2
1, 2, 2*, 4	2.25	2.5
1, 2, 3, 3*	2.25	2
1, 2, 3, 4	2.5	2.5
1, 2, 3*, 4	2.5	2.5
1, 2*, 3, 3*	2.25	2
1, 2*, 3, 4	2.5	2.5
1, 2*, 3*, 4	2.5	2.5
1, 3, 3*, 4	2.75	2.5
2, 2*, 3, 3*	2.5	2.5
2, 2*, 3, 4	2.75	3
2, 2*, 3*, 4	2.75	3
2, 3, 3*, 4	3	3
2*, 3, 3*, 4	3	3

(b)

Sample Midrange	2	2.5	3
Probability	4/15	7/15	4/15

(c) $E(\text{sample mean}) = 2\cdot\frac{2}{15} + 2.25\cdot\frac{1}{5} + 2.5\cdot\frac{1}{3} + 2.75\cdot\frac{1}{5} + 3\cdot\frac{2}{15} = \mathbf{2.5}.$

$E(\text{sample midrange}) = 2\cdot\frac{4}{15} + 2.5\cdot\frac{7}{15} + 3\cdot\frac{4}{15} = \mathbf{2.5}.$

(d) The fact that the expected value of the sample mean is equal to the population mean tells us that use of the sample mean to estimate the population mean is appropriate, since, on average, the sample mean will be equal to the population mean.

The fact that the expected value of the sample midrange is equal to the population mean tells us that use of the sample midrange to estimate the population mean is also appropriate, since, on average, the sample midrange will be equal to the population mean.

(e) The standard deviation of a statistic is a typical deviation of the values of the statistic from the expected value of the statistic. Since the expected values of both statistics are equal to the population mean, and since the standard deviation of the sample mean is smaller than that of the sample midrange, we can conclude that, on average, the deviation of the sample mean from the population mean will be less than the deviation of the sample midrange from the population mean. Therefore the sample mean is preferable to the sample midrange as an estimator of the population mean.

Question Five

Hypotheses:

Let p_M and p_F be the proportions of all male and female parakeets (respectively) that can mimic speech.

$H_0: p_M = p_F$

$H_a: p_M \neq p_F$

Check of conditions and naming of test:

This is a two-proportion z-test.

We are told that we can consider the sets of male and female parakeets as random samples from the populations of male and female parakeets.

$n_M \hat{p}_M = 33 \cdot \frac{14}{33} = 14 > 10.$ $n_M(1 - \hat{p}_M) = 33 \cdot \frac{19}{33} = 19 > 10.$

$n_F \hat{p}_F = 44 \cdot \frac{13}{44} = 13 > 10.$ $n_F(1 - \hat{p}_F) = 44 \cdot \frac{31}{44} = 31 > 10.$

[Alternatively, one could check that $n_M \hat{p}_c$, $n_M(1 - \hat{p}_c)$, $n_F \hat{p}_c$, and $n_F(1 - \hat{p}_c)$ are all greater than 10, where $\hat{p}_c$ is the proportion of successful parakeets for the combined samples. Here $\hat{p}_c = (14 + 13)/(33 + 44) = 27/77$, making $n_M \hat{p}_c = 33(27/77) = 11.57$, $n_M(1 - \hat{p}_c) = 33(50/77) = 21.43$, $n_F \hat{p}_c = 44(27/77) = 15.43$, and $n_F(1 - \hat{p}_c) = 44(50/77) = 28.57$, which are all greater than 10.]

Mechanics:

$z = \dfrac{(\hat{p}_M - \hat{p}_F) - (p_M - p_F)}{\sqrt{\hat{p}_c(1 - \hat{p}_c)\left(\frac{1}{n_M} + \frac{1}{n_F}\right)}}$, where $\hat{p}_c$ represents the proportion of successful parakeets for the combined samples.

As explained above, $\hat{p}_c = 27/77$. So $z = \dfrac{(14/33 - 13/44) - (0)}{\sqrt{\left(\frac{27}{77}\right)\left(\frac{50}{77}\right)\left(\frac{1}{33} + \frac{1}{44}\right)}} = 1.172.$

So the p-value is 0.241. (This value is obtained on the TI-83 and TI-84 calculators by entering 2*normalcdf(1.172,999) or 2*normalcdf(1.172,999,0,1) and on the TI-Nspire by entering 2*normalCdf(1.172,999,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

Conclusion:

Since the p-value is 0. 241, which is greater than 0.05, we do not reject H_0. We do not have sufficient evidence to conclude that there is a difference between male and female parakeets in their ability to mimic speech.

Section II—Part B

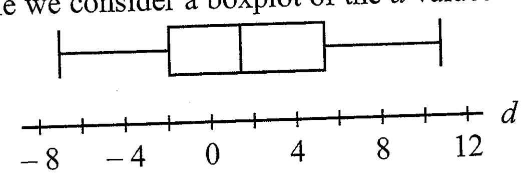
Question Six

(a)

Student	1	2	3	4	5	6	7	8	9	10	11	12
Teacher X	99	94	88	80	83	84	82	82	82	91	90	65
Teacher Y	97	87	87	76	88	87	80	83	71	84	97	65
$d = X - Y$	2	7	1	4	−5	−3	2	−1	11	7	−7	0

Hypotheses:
Let μ_d be the mean difference in grades given by the two teachers.
$H_0: \mu_d = 0$
$H_a: \mu_d > 0$
[*Note: If you defined d as (Score from Teacher Y – Score from Teacher X), then your alternative hypothesis should be* $H_a: \mu_d < 0$.]

Check of conditions and naming of test:
This is a paired t-test.
We need to assume that the students given are a random sample of all possible students taught this course by these two teachers.
Since this is a small sample we consider a boxplot of the d values:



Since the distribution of the differences is roughly symmetrical and has no outliers, it appears reasonable to assume that the population of d-values is normally distributed. Therefore we can proceed with the paired t-test.

Mechanics:

$$t = \frac{\bar{d} - \mu_d}{\left(\frac{s_d}{\sqrt{n}}\right)} = \frac{1.5 - 0}{\left(\frac{5.231}{\sqrt{12}}\right)} = 0.993.$$

There are $n - 1 = 11$ degrees of freedom. So the p-value is $P(t_{11} > 0.993) = 0.171$. (This p-value can be obtained on the TI-83, TI-84, and TI-Nspire calculators by entering tcdf(0.993,999,11). The tcdf function is accessed on the TI-83/84 by 2nd,DISTR, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) tCdf is accessed by selecting Menu,Statistics,Distributions.)

Conclusion:
Since the p-value is 0.171, which is greater than 0.05, we do not have sufficient evidence to conclude that the mean grade given by Teacher X is higher than the mean grade given by Teacher Y.

(b) Since the p-value is 0.015, which is less than 0.05, we have sufficient evidence to conclude that the mean grade given by Teacher A is higher than the mean grade given by Teacher B.

(c) No, the answer to part (b) does not imply that Teacher A is more generous than Teacher B. Since the students are assigned to the teachers according to what other courses they are taking, it is quite possible that the students taught by Teacher A are smarter on average than those taught by Teacher B, and so the fact that Teacher A is giving higher grades than Teacher B does not mean that Teacher A is more generous than Teacher B.

(d) The majority of the points in the scatterplot for Teacher A lie above the line $y = x$, which means that for most students, Teacher A is giving grades that are higher than those given by other teachers. In the scatterplot for Teacher B, the points are roughly evenly distributed above and below the line $y = x$, so generally speaking Teacher B is giving grades that are roughly the same as those given by other teachers. Therefore, in this sense, Teacher A appears to be more generous than Teacher B.

Sample Examination Five – Answers

1. **(B)** The 70th percentile weight is the weight below which 70% of all package weights lie. Now, 70% of 14,532 is $14532 \cdot (0.7) = 10172.4$. The number of packages with a weight less than 10 pounds is $3550 + 4215 = 7765$, and the number of packages with a weight less than 20 pounds is $3550 + 4215 + 2662 = 10427$. Therefore, the 70th percentile package weight lies between 10 and 20 pounds. (B) is the only response that satisfies this condition.

2. **(D)** The least squares regression line is defined to be the line that minimizes the sum of the squares of the vertical distances of the points from the line.

3. **(A)** For any number produced by the random number generator, the probability that the number is less than 3 is 2/5. Considering a number less than 3 as "success," we are interested in the number of successes in 10 trials, where the probability of success is the same for each trial and the outcomes of the trials are independent. Thus the number of successes is binomially distributed with $n = 10$ and $p = 2/5$, and so the required probability is the one given in (A).

4. **(C)** By emphasizing the cost of the renovations, the manager's question encourages the response "No." Thus, (C) is the correct response.

   Option (A) is incorrect since, although 12 members of the club failed to respond, these 12 members form only a small proportion of the club's membership, and therefore this is unlikely to be a source of bias. Option (B) is incorrect since the population of interest is the membership of the club, and all the members of the club are included in the survey. Options (D) and (E) are incorrect: the study does indeed make use of a census, but this is preferable to any form of sampling.

5. **(D)** Let X be the number of students who attend. We need
$$P(X \le 4 \mid X \ge 2) = \frac{P(X \le 4 \cap X \ge 2)}{P(X \ge 2)} = \frac{0.16 + 0.18 + 0.22}{0.16 + 0.18 + 0.22 + 0.16 + 0.08} = 0.7.$$

6. **(C)** By assigning each subject a random number between 0001 and 9999, ignoring repeats, and assigning the subjects with the lowest 50 numbers to Group 1, a completely randomized design is achieved. Note that in (B) the assignment is not completely random. Consider, for example, the last two subjects in the alphabetical list. These two subjects are very likely to be assigned to the same group. This cannot be the case in a completely randomized design. Options (D) and (E) are incorrect since the processes given don't involve any randomization. Option (A) is incorrect since the method described is unlikely to result in groups of equal size.

7. **(C)** Consider the third quartile, Q_3. Three quarters of the observations are smaller than Q_3 and one quarter are larger. Thus Q_3 is a measure of location and not a measure of spread. (The third quartile is not to be confused with the interquartile range (third quartile − first quartile), which *is* a measure of spread.)

8. **(B)** In the chi-square goodness of fit test the number of degrees of freedom is given by $k - 1$, where k is the number of outcome categories. So here the number of degrees of freedom is $3 - 1 = 2$. Thus, p-value $= P(\chi_2^2 \geq 4.19) = 0.123$. (On the TI-83, TI-84, and TI-Nspire calculators, this result is obtained using χ²cdf(4.19,999,2). On the TI-83/84, the χ²cdf function is accessed by 2nd,DISTR, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) the χ²Cdf function is accessed by selecting menu,Statistics,Distributions.) Since this value is greater than 0.05, we do not have sufficient evidence (at the 5% level) to conclude that the magazine's claim is incorrect. Note that (C) is incorrect, since we can never have convincing evidence that the null hypothesis is true. This would be "accepting H_0," which is never allowed.

9. **(B)** This is a two-proportion z-test. In significance tests for proportions, there can be no assumption of normality in a population, since our interest in the population is only with regard to a yes/no (success/failure) scenario. These tests use the normal distribution since, under the condition that the counts are large enough (as stated, for example, in response (A)), the distributions of the *sample proportions* are normal.

10. **(A)** The "R-sq" value given in the computer output is 22.1%. Thus, $r = \pm\sqrt{0.221} = \pm 0.470$. We know that the correlation between the total rainfall and the total sunshine is negative, since the output gives a negative slope (-1.670) for the regression line. Therefore, $r = -0.470$.

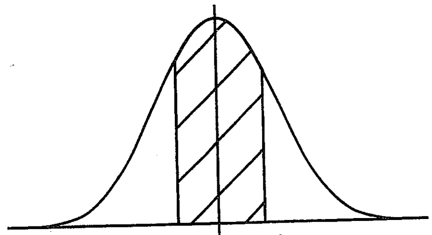
11. **(E)** The statistic referred to as "S" in the computer output is the standard error about the least-squares regression line, given by

$$s = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}}.$$

The quantity $\sum (y - \hat{y})^2$ is the sum of the squares of the residuals, which are the vertical displacements of the points from the regression line. Thus s is a measure of the variability of the y-values from those predicted by the regression line.

12. **(B)** This is the significance test for the slope of the regression line, and the slope of the regression line for this sample is the coefficient of Rainfall in the regression equation. The p-value for the test is given in the "P" column of the computer output, in the "Rainfall" row, and the value given in the computer output in the context of this significance test is always the one for the two-tailed version of the test. Thus the required p-value is 0.0178.

13. **(E)**



Since the mean of the given distribution is 120, we can see (as shown above) that 110 and 130 are equal distances to the left and the right, respectively, of the mean. The required probability is given by the area under the curve between 110 and 130, and it is easy to see that this is twice the area between 120 and 130.

Option (B) is incorrect, since the events "$X \geq 110$" and "$X \leq 130$" are not independent. (Knowing that $X \geq 110$ makes it less likely that $X \leq 130$.) Option (C) would be correct if the inequality in the subtracted probability were "*less* than or equal to," but this is not the case!

14. **(D)** When a population whose mean is μ and whose standard deviation is σ is normally distributed, it follows that the distribution of $(\overline{X}-\mu)/(\sigma/\sqrt{n})$ is normal (where $\overline{X}$ is the mean of a random sample and n is the sample size). However, when σ is unknown, the quantity $(\overline{X}-\mu)/(\sigma/\sqrt{n})$ cannot be used in the significance test, and so the quantity $(\overline{X}-\mu)/(s/\sqrt{n})$ is used in its place, and this quantity is t-distributed (with $n-1$ degrees of freedom) if we know that the population is normally distributed. Thus (D) is the correct response. Note that (E) is incorrect since, as we know that the population is normally distributed, the sampling distribution of the sample mean, $\overline{X}$, is normal.

15. **(D)** Let P(King) be p. Then P(Steward) $= p$, also. Moreover, P(Peasant) $= P$(Thief) $= 4p$. Since "King," "Steward," "Peasant," and "Thief" are the only four possibilities, $p + p + 4p + 4p = 1$. Therefore $10p = 1$, and so $p = 0.1$. Thus P(Peasant) $= 4(0.1) = 0.4$.

16. **(E)** This is a paired t confidence interval. We use a t-distribution with $n - 1 = 10 - 1 = 9$ degrees of freedom. Obtained from the table of t-distribution critical values, the critical value of this distribution for a 95% confidence interval is 2.262. So the confidence interval is given by $\overline{d} \pm t^* \cdot \frac{s_d}{\sqrt{n}} = 0.48 \pm 2.262 \cdot \frac{0.766}{\sqrt{10}}$.

17. **(C)** We use a z confidence interval for $\mu_2 - \mu_1$ since the population standard deviations are known. Thus the required confidence interval is given by $(\overline{x}_2 - \overline{x}_1) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$. The

value of z^* is 1.96. (This value is given by invNorm(0.975) or invNorm(0. 975,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0. 975,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) The values of $\sigma_1, \sigma_2, n_1, n_2$ are given in the question. So all that remains for us is to calculate the value of $\overline{x}_2 - \overline{x}_1$. Now $\overline{x}_1$ will be the central value in the confidence interval for μ_1, and $\overline{x}_2$ will be the central value in the confidence interval for μ_2. Therefore $\overline{x}_1 = (14.223 + 16.164)/2 = 15.1935$, and $\overline{x}_2 = (16.388 + 18.930)/2 = 17.659$. So the required confidence interval is $(17.659 - 15.1935) \pm 1.96\sqrt{\frac{2.1^2}{18} + \frac{2.9^2}{20}} = (0.867, 4.064)$.

18. **(E)** Correlation does not imply causation: the fact that TV-watching has been found to be negatively correlated with test scores does not imply that watching a large amount of TV has *caused* low test scores. Therefore (E) is the correct response.

    The statement in (A) is correct (and therefore option (A) is incorrect) because the correlation of -0.898 does indicate a strong linear relationship between the variables. The statement in (B) is correct (and therefore option (B) is incorrect) since r^2 here is $(-0.898)^2 = 0.806$; thus, 80.6% (that is, *most*) of the variability in test scores can be explained by the regression line. The statements in (C) and (D) are correct (and therefore these options are incorrect) since the negative correlation indicates that a large amount of TV-watching was associated with low test scores, and vice versa.

19. **(C)** If the members of the treatment group were to undergo, on average, a significantly smaller amount of burning than the members of the control group, this would provide convincing evidence that the combination of the chemical and the liquid base (along with any possible placebo effect) provides protection against the sun. We will now think about why (A) and (B) are incorrect. The essence of experimental design is to obtain two groups that differ *only* with respect to the factor that is to be tested. So, in the context of this experiment, if the control group were to receive the liquid base (and no chemical), and if we were to make sure that that the complete lotion and the liquid-base-only were identical in color, smell, and so on, then the only difference between the two groups would be the presence of the chemical in the lotion given to the treatment group. Thus, with that design, the effectiveness of the chemical itself could be tested. However, with the experiment as it is described in the question, the groups differ with respect to the liquid base, the chemical, *and* the placebo effect (if you believe the placebo effect could happen in this context), and so it is impossible to distinguish between the effects of these three factors.

20. **(D)** In a significance test, the *p*-value is the probability, given that H_0 is true, that the observed outcome would take a value at least as extreme as the one actually observed.

21. **(E)** If we were given that the distribution of the taxable values of the homes in the city were normally distributed, then we would be able to determine the percentage of homes that were within two standard deviations of the mean. However, we are not given here that the distribution is normal. In fact, since we are told that the distribution is positively skewed, we know that the distribution is *not* normal. Therefore the percentage cannot be estimated from the information given.

22. **(B)** The probability of a Type I error is always equal to the significance level of the test (which is here given to be 0.05). So an increase in the sample size will have no effect on the probability of a Type I error. Therefore (B) is the correct response.

An increase in the sample size *will*, however, decrease the probability of a Type II error; thus, (C) is an incorrect response. What is more, power is the probability that a Type II error is *not* made, and therefore an increase in the sample size will result in an increase in the power; thus, (D) is an incorrect response. Turning our attention to (E), the standard deviation of the sampling distribution of the sample proportion is $\sqrt{p(1-p)/n}$, where n is the sample size, and therefore this standard deviation is reduced by an increase in n; thus (E) is an incorrect response. Finally, considering (A), the confidence interval for the population proportion is $\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$. When the sample size, n, is increased, the sample proportion, $\hat{p}$, is unlikely to be changed greatly (what is more, any increase in the value of $\hat{p}$ will be compensated by a decrease in the value of $(1-\hat{p})$, and vice versa) and so the increase in the value of n in the denominator of the fraction is likely to result in a decrease in the width of the confidence interval; thus, response (A) is incorrect.

23. **(B)** Consider the sequence of t-distributions, as we increase the number of degrees of freedom. With low numbers of degrees of freedom, the t-distributions are noticeably different from the standard normal distribution and have greater variability than the standard normal distribution. As the number of degrees of freedom increases, the variability of the t-distributions decreases, and the t-distributions get closer to the standard normal distribution. Therefore, of the distributions given, the one with the smallest standard deviation is $N(0,1)$, followed by t_{10}, and then by t_5. (Please see "Understand where the t-distributions come from" in the Top Tips section of the Question Book for a graph that compares two of the t-distributions with the standard normal distribution.)

24. **(B)** Cluster sampling is carried out as follows. The population is divided into clusters, some of the clusters are selected at random, and then individuals from the selected clusters form the sample. Under the strategy given in (B), the resulting sample is likely to be representative of the population of third graders in terms of reading ability. In (C) and (D), certain parts of the population will necessarily be excluded from the sample, and so these options are not appropriate. To use the strategy given in (E) would defeat the object of cluster sampling, which is to cut down the work and expense involved: if there are many (small) clusters, then a large number of clusters would have to be selected in order to adequately represent the population, which would be time-consuming and expensive. Likewise, (A) is incorrect: the point of a cluster sample is to avoid the work and expense involved in sampling methods such as a simple random sampling.

25. **(D)** We can show that A and B are not independent a number of different ways: by showing that $P(A \cap B) \neq P(A) \cdot P(B)$, or that $P(A \mid B) \neq P(A)$, or that $P(B \mid A) \neq P(B)$, or by many other similar methods. We will use the first one.

The total number of cars sold is $977 + 421 + 244 + 361 + 610 + 326 = 2939$.

So $P(A) = (977 + 421)/2939 = 0.476$, $P(B) = (977 + 244 + 610)/2939 = 0.623$, and $P(A \cap B) = 977/2939 = 0.332$.

Therefore $P(A) \cdot P(B) = (0.476)(0.623) = 0.297$, and this is not equal to $P(A \cap B)$. So the events A and B are not independent.

To show that events A and B are not mutually exclusive, we have to show that (when one vehicle is picked at random) both events can happen, in other words that the vehicle can be both a sedan and bought from the downtown office. This clearly is possible, and so the events A and B are not mutually exclusive.

26. **(A)** The power of the test is the probability, when H_a is true, that H_0 is correctly rejected in favor of H_a.

27. **(C)** Let the proportion of all male students who approve be p_1 and the proportion of all female students who approve be p_2. When a random sample of 80 male students is taken, the proportion, $\hat{p}_1$, of the sample who approve has standard deviation $\sqrt{p_1(1-p_1)/80}$. Similarly, the proportion, $\hat{p}_2$, of the sample of 100 female students who approve has standard deviation $\sqrt{p_2(1-p_2)/100}$, and these two proportions are independent. Therefore, using the formula $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$, we see that the standard deviation of $\hat{p}_1 + \hat{p}_2$ is $\sqrt{p_1(1-p_1)/80 + p_2(1-p_2)/100}$. To obtain the *mean* of the two sample proportions we divide $(\hat{p}_1 + \hat{p}_2)$ by 2, and so the standard deviation is divided by 2, also. Hence the standard deviation of $(\hat{p}_1 + \hat{p}_2)/2$ is $(1/2)\sqrt{p_1(1-p_1)/80 + p_2(1-p_2)/100}$.

Now the question requires the standard *error* of the mean of the two sample proportions. The standard error of a statistic is obtained by replacing, in the formula for the standard deviation of the statistic, all population parameters by their estimates obtained from the sample(s). Therefore, here, p_1 is replaced by $23/80$ and p_2 is replaced by $38/100$. This gives the required standard error to be

$$\frac{1}{2}\sqrt{\frac{(23/80)(57/80)}{80} + \frac{(38/100)(62/100)}{100}}.$$

28. **(E)** Within each block, we need the plots to be similar in terms of the amounts of water and light received. Looking at (E), we see that this is the case with all six blocks listed: plots 1 and 4 receive similar amounts of light and similar amounts of water, plots 2 and 3 receive similar amounts of light and similar amounts of water, and so on. Thus option (E) gives a suitable blocking scheme. All the other options give schemes that are unsuitable. For example, the first block in (A) contains plots 1 and 9, which receive different amounts of light; the first block in (B) contains plots 1 and 2, which receive different amounts of water; likewise, every block given in (A)-(D) contains plots that receive differing amounts of light and/or water. Thus (E) is the only option that provides a suitable blocking scheme.

29. **(D)** A (two-sided) 95% confidence interval is equivalent to a two-tailed test with a 5% significance level. So here, since 158.4 is contained in the confidence interval, we do not have convincing evidence at the 5% level that $\mu \neq 158.4$.

Note that (A) is incorrect because we can never obtain convincing evidence that μ is *exactly* 158.4: this would be accepting H_0, which is never allowed. Options (B) and (C) are clearly incorrect since the confidence interval for μ contains 158.4.

For anyone who is concerned that the statement in (E) might be true, consider the following (somewhat tricky) argument. In terms of critical values and tail probabilities, a one-tailed test at the 5% significance level is equivalent to a two-tailed test at the 10% significance level. Furthermore, a two-tailed test at the 10% significance level is equivalent to a 90% (two-sided) confidence interval. So we will use the information given to calculate a 90% confidence interval for μ. The center of the confidence interval given is $(158.28 + 160.92)/2 = 159.60$, and the margin of error (half of the width of the confidence interval) is $160.92 - 159.60 = 1.32$. Now, this (given) confidence interval was calculated using the t-distribution with 19 degrees of freedom, of which the critical value for a 95% confidence level is 2.093. The critical value of that same distribution for a 90% confidence level is 1.729. So the margin of error of the 90% confidence interval will be $1.32(1.729/2.093) = 1.090$. Therefore, the 90% confidence interval for μ is $159.60 \pm 1.090 = (158.51, 160.69)$. 158.4 is not in this interval, and thus we have convincing evidence at the 10% level that $\mu \neq 158.4$. Equivalently, since every value in the 90% confidence interval is *greater* than 158.4, we have convincing evidence at the 5% level that $\mu > 158.4$. Hence the statement in (E) is false.

30. **(D)** Since the total number of umbrellas is 300, we can calculate the total number of black umbrellas by $300 - 35 - 79 = 186$. Using this result, we can calculate the number of black telescopic umbrellas by $186 - 59 = 127$. From this, we see that the total number of telescopic umbrellas is $127 + 47 + 13 = 187$. Hence the expected number of black telescopic umbrellas is $(186)(187)/300 = 115.94$.

31. **(A)** Recall the following three facts: First, for *any* two random variables X and Y, $E(X+Y) = E(X) + E(Y)$. Second, if X and Y are *independent*, then $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$. Third, that if X and Y are *independent* and normally distributed then the random variable $X+Y$ is normally distributed. Since here it's a married couple being chosen at random (as opposed to a man being chosen at random and a woman being chosen at random), the height of the man and the height of the woman are not independent. Thus, only statement I can be concluded from the information given.

32. **(C)** The 95% confidence interval for μ is given by $\overline{x} \pm t^*(s/\sqrt{n})$. So the center of the confidence interval is at the sample mean, $\overline{x}$, and the distance from the center of the interval to the upper (or lower) end of the interval is given by $t^*(s/\sqrt{n})$. Only (C) and (E) give intervals for which the center is at the sample mean, and so (A), (B) and (D) are eliminated. Now t^*, the critical value of the t-distribution with 14 degrees of freedom, is 2.145, and $n = 15$. Looking at the dotplot given, we estimate the standard deviation of the sample (a typical deviation from the sample mean) to be somewhere between 15 and 30. (The true value of the sample standard deviation is close to 25.) So the distance from the center of the interval to the upper end of the interval is between $2.145(15/\sqrt{15}) = 8.3$ and $2.145(30/\sqrt{15}) = 16.6$. The interval given in (B) satisfies this condition, and the one given in (E) does not.

33. **(E)** Recall that if a and b are constants, x and y are defined as they are in the question, and $y = ax + b$, then $\overline{y} = a\overline{x} + b$ and, if $a > 0$, $s_y = as_x$. In other words, multiplying by a multiplies the mean and the standard deviation by a, and adding b adds b to the mean but has

no effect on the standard deviation (and similarly subtracting b would subtract b from the mean and would have no effect on the standard deviation). Now consider (E). We start with a mean of 52.3 and a standard deviation of 10.8. Subtracting 52.3 will reduce the mean by 52.3 making the mean now 0 and leaving the standard deviation at 10.8. We then multiply by $7.9/10.8$, which will transform the standard deviation to 7.9 and will leave the mean at 0. Finally, adding 80.6 will increase the mean to 80.6 and will leave the standard deviation at 7.9. These are the required mean and standard deviation.

34. **(D)** Suppose, for example, the sample size used was 100. A chi-square test for goodness of fit can be used, setting up a table as shown below.

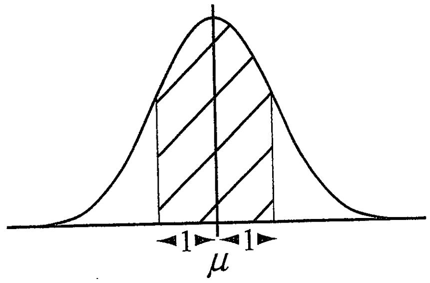
	Observed Count	Expected Count
Males	*	50
Females	*	50

The chi-square goodness of fit test can then be performed in the usual way. (For interest, it is worth noting that the p-value obtained for the chi-square goodness of fit test will be exactly the same as the p-value obtained for the one-proportion z-test.)

Note that (E) is incorrect since only one species is being considered. If two or more species were being compared in terms of the proportions of the species that are males, then the chi-square test for homogeneity could be used. Likewise, (C) is incorrect: there is only one sample in this study. Options (A) and (B) are incorrect since this study is concerned with the *proportion* of males in the species; thus, tests for means cannot be used to investigate the biologist's question.

35. **(D)** By definition, the sample of 81 employees will be a simple random sample if (and only if) every subset of 81 employees has an equal chance of being the sample. This is not the case here since, for example, a sample that includes 51 manual workers is not possible. Thus, not every subset of 81 employees has an equal chance of being the sample, and so the sample obtained will not be a simple random sample. (Considering option (A), we note that, since exactly 1/10 of each type of employee is included in the sample, every employee has an equal probability (1/10) of being included in the sample. However, this does *not* enable us to conclude that the sample obtained will be a simple random sample.)

36. **(D)** When a population with mean μ and standard deviation σ is normally distributed, and a random sample of size n is taken from the population, the distribution of the sample mean is $N(\mu, \sigma/\sqrt{n})$. So here the sample mean battery life has distribution $N(\mu, 4.3/\sqrt{10})$. The graph of this distribution is shown below.



The required probability is the area of the shaded region. Since the right boundary of this region is 1 month above the mean, the z-value there is given by

$$z = \frac{1}{4.3/\sqrt{10}} = 0.7354.$$

Similarly, the z-value at the left boundary is -0.7354. Therefore the required probability is $P(-0.7354 < Z < 0.7354) = 0.538$. (This value is obtained on the TI-83 and TI-84 calculators by entering normalcdf(-.7354,.7354) or normalcdf(-.7354,.7354,0,1) and on the TI-Nspire by entering normalCdf(-.7354,.7354,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

37. **(A)** First look at the scatterplot shown above the question. Remember that a point will have a positive residual if it lies above the regression line, and a negative residual if it lies below the regression line. So if we look across the scatterplot from left to right, we see negative residuals, then primarily positive residuals, then primarily negative residuals. Looking at the residual plot in (A) we have exactly that: negative residuals, then primarily positive residuals, then primarily negative residuals. This is not the case in any of the other options.

38. **(C)** A sample proportion of $\hat{p} = 107/200 = 0.535$ has been obtained. The formula for the 95% confidence limits for p is $\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$. We use $\hat{p} = 0.535$, $z^* = 1.96$ and $n = 200$. (The z^* value is given by invNorm(0.975) or invNorm(0.975,0,1) on the TI-83 and TI-84 calculators, and by invNorm(0.975,0,1) on the TI-Nspire. The invNorm function is accessed on the TI-83/84 by 2nd,DISTN, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.) This gives a confidence interval of (0.466, 0.604). In other words, we are 95% confident that p is between 0.466 and 0.604, as stated in (C).

Let's be clear that (D) is incorrect. The equivalent true statement would be that if the procedure of flipping the coin 200 times were to be repeated a large number of times, then 95% of the resulting confidence intervals would contain the true proportion of "heads" for this coin. Both this correct statement and the correct statement in (C) tell us something about the *population* proportion. The statement in (D) tries to relate the confidence interval to *sample* proportions. Likewise, the statement in (E) tries to relate the confidence interval to a sample proportion, and therefore is incorrect.

39. **(C)** In any cumulative relative frequency graph, such as the one given in (C) for example, the cumulative relative frequency plotted for a measurement of 6, say, is the proportion of all measurements that are less than or equal to 6. Likewise, the cumulative relative frequency plotted for a measurement of 7 is the proportion of all measurements that are less than or equal to 7. So for this data set, the difference between these two cumulative relative frequencies must be the proportion of all measurements that are equal to 7, which is governed by the height of the block in the histogram above 7 on the horizontal axis. Therefore the amounts by which the cumulative relative frequencies increase (the vertical increments in the cumulative relative frequency graph) are governed by the heights of the blocks in the histogram. The heights of the blocks in the histogram increase up to 8 on the horizontal axis, and therefore we are looking for a cumulative relative frequency graph for which the vertical increments are increasing up to 8 on the horizontal axis. This is the case only for (C).

40. **(E)** We will first calculate the probability of getting red, green, blue, in that order. There are 10 beads in total and five of them are red, and so the probability that the first bead *is* red is 5/10. If the first bead is red, then there are 9 beads remaining and 3 of them are green. So the probability that the next bead is green is 3/9. Once this has happened, there are 8 beads remaining and 2 of them are blue. So the probability that the last bead is blue is 2/8. Therefore the probability of getting red, green, blue, in that order, is given by

$$P(\text{RGB}) = \left(\frac{5}{10}\right)\left(\frac{3}{9}\right)\left(\frac{2}{8}\right).$$

Now we'll consider a different order of those three colors: RBG. By the same logic, the probability of this is given by

$$P(\text{RBG}) = \left(\frac{5}{10}\right)\left(\frac{2}{9}\right)\left(\frac{3}{8}\right),$$

which has the same value as that of $P(\text{RGB})$. It should be clear now that every different order of the three colors has the same probability.

So all that remains is to calculate the number of possible orders of the three colors. The orders can be listed as follows: RGB, RBG, GRB, GBR, BRG, BGR. So there are 6 possible orders. Therefore we need 6 times the probability calculated above, which is the quantity given in (E).

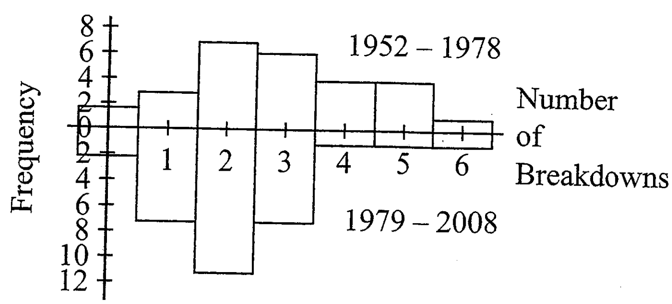
Section II—Part A

Question One

(a)

Number of Breakdowns	Frequency 1952-1978	Frequency 1979-2008
0	2	2
1	3	7
2	7	11
3	6	7
4	4	1
5	4	1
6	1	1

(b) A back-to-back histogram displaying these data is shown below.



[Several other types of display are possible here; for example, histograms drawn side-by-side and with the same scales, a bar graph with columns in pairs (shaded differently to indicate the two distributions), parallel boxplots with a single scale, or frequency polygons drawn on a single pair of axes.]

(c) Based on the display in part (b), the median of the distribution for 1979-2008 appears to be around 2 breakdowns per year, and this is less than the median of the distribution for 1952-1978, which appears to be around 3 breakdowns per year. Therefore the occurrence of breakdowns does seem to have decreased since the time of the refurbishment.

(d) The distribution of the annual numbers of breakdowns in the years 1952 to 1978 is roughly symmetrical. The distribution of the annual numbers of breakdowns in the years 1979 to 2008 is positively skewed (skewed to the right).

[The histogram for 1952-1978 could be said to give an impression of positive skewness, and therefore a statement that this distribution is positively skewed would be accepted. Inspection of the boxplot for this distribution shows that the distribution, in terms of the locations of its maximum, minimum, median, and quartiles, is very close to being symmetrical.]

Question Two

(a) Number the students in the school from 1 to N, where N is the total number of students. Use a random number generator to select an integer between 1 and N, inclusive. The student with this number will be the first student in the sample. Repeat this process, ignoring repeated numbers, until 40 students have been selected.

(b) Using a method similar to the one used in (a), randomly select 10 students from the 9th grade. In the same way, randomly select 10 students from the 10th grade, 10 students from the 11th grade, and 10 students from the 12th grade. These 40 students will form the sample.

(c) It is sensible to stratify by grade level if we believe that the four grades might differ from one another in terms of the amount of time spent studying and that within each grade the amount of time spent studying varies little between students. For example, the statistics students might well suspect, in advance of doing the survey, that older students tend to study longer than younger students, with some degree of consistency of study time within the grades.

(d) Assume that the criteria mentioned in part (c) are satisfied. Then, using a stratified sample in the way described in place of a simple random sample will reduce the variability of the estimate of the population mean, in the sense that if many of these stratified samples were taken, their means would vary less than the means of a large set of simple random samples. This would come about as a result of the fact that each stratified sample will have 10 students from each grade, whereas the simple random samples would have varying numbers of students from the four grades. Thus stratified sampling will produce a more reliable estimate of the population mean.

Question Three

(a) For Game A, the probability of a positive gain is $P(X > 0) = 0.20 + 0.13 + 0.07 = 0.4$.
For Game B, the probability of a positive gain is $P(Y > 0) = 0.16 + 0.10 + 0.05 = 0.31$.

(b) The expected value of the money gained in any play of Game B is
$\mu_Y = (-5)(0.37) + (-2)(0.32) + (1)(0.16) + (4)(0.10) + (25)(0.05) = -\$0.68$

The variance of the money gained in any play of Game B is

$$\sigma_Y^2 = (-5-(-0.68))^2(0.37) + (-2-(-0.68))^2(0.32) + (1-(-0.68))^2(0.16) + (4-(-0.68))^2(0.10) + (25-(-0.68))^2(0.05) = 43.0776$$

So the required standard deviation is $\sigma_Y = \sqrt{43.0776}$ = **$6.56**

(c) No, the friend's argument in not valid, since the *amount* of the gain or loss is relevant in the choice of game, rather than just whether the gain is positive or negative. If a large number of plays is made of Game A, we expect the mean gain per play to be around the expected value for Game A, which is $-\$0.77$. Likewise, if a large number of plays is made of Game B we expect the mean gain per play to be around $-\$0.68$. Therefore, Game B is the better choice.

[*It is worth noting here that due to the large standard deviations (compared to the sizes of the means) it is still quite possible, in a large number of plays, that Game A could turn out to be the more profitable (or less loss-making) choice. For example, suppose that 100 plays were to be made at Game A and 100 plays were to be made at Game B. Using the Central Limit Theorem we can show that the probability of the mean gain at Game A being greater (less negative) than the mean gain at Game B is 0.45. Nonetheless, this confirms that more often than not, when a large number of plays is made, Game B is the better choice.*]

Question Four

Hypotheses:
H_0: The distributions of car type are the same in the populations of reservations at all three locations.
H_a: The distributions of car type are not all the same.

Check of conditions and naming of type of test:
This is a chi-square test for homogeneity.

The conditions for the chi-square test for homogeneity are:
1. Random samples
2. Large sample sizes

We are told that we can consider the samples of reservations as random samples from the populations of reservations at the three locations. So condition 1 is satisfied.

The expected counts are shown in the table below.

	Type			
	Compact	Standard	Full Size	Other
Location 1	81.98	67.9	46.31	53.82
Location 2	90.83	75.23	51.31	59.63
Location 3	89.19	73.87	50.38	58.55

Since all the expected counts are greater than 5, the sample sizes are large enough for the chi-square test to be appropriate. Thus condition 2 is satisfied.

Mechanics:

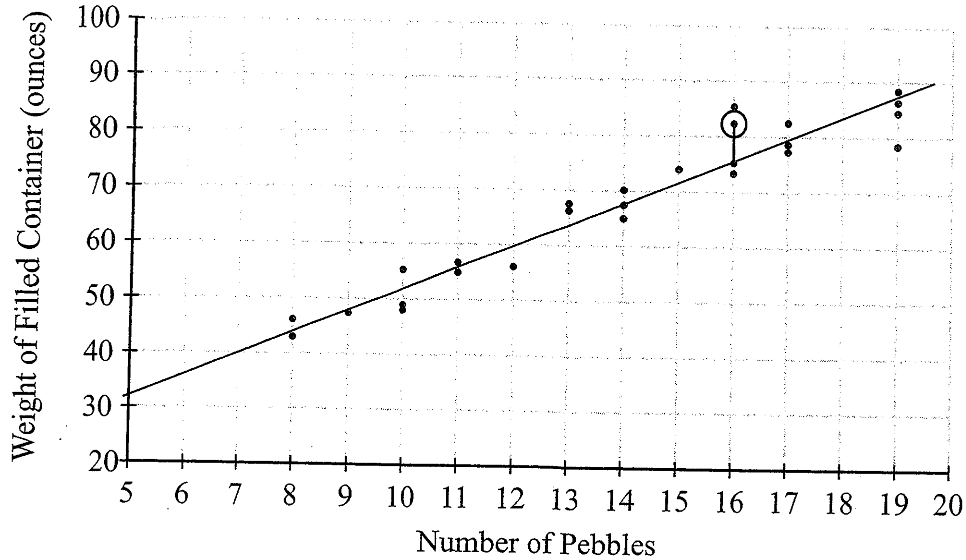
$$\chi^2 = \sum \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$$
$$= 0.436 + 0.743 + 0.859 + 0.499 + 0.008 + 0.102 + 0.362 + 0.094 + 0.520 + 1.319 + 2.237 + 0.974$$
$$= 8.154$$

We use the chi-square distribution with $(3-1)(4-1) = 6$ degrees of freedom.
So the *p*-value is $P(\chi^2_6 > 8.154) = 0.227$. (On the TI-83, TI-84, and TI-Nspire calculators, this result is obtained using χ^2cdf(8.154,999,3). On the TI-83/84, the χ^2cdf function is accessed by 2nd,DISTR, and on the TI-Nspire (in the *Scratchpad* or *Calculator* application) the χ^2Cdf function is accessed by menu,Statistics,Distributions.)

Conclusion:
Since the *p*-value is 0.227, which is greater than 0.05, we do not reject H_0. We do not have statistically significant evidence of a difference in the distributions of the car types between the populations of reservations at the three locations.

Question Five

(a) The scatterplot, with the least squares regression line added, is shown below.



[*The easiest way to draw the line is to plot the points for* $x = 5$ *and* $x = 20$ *first, and then to join them with a straight line. (Be sure to use a straight edge for drawing the line!) The y-coordinate of the point with x-coordinate* 5 *is found by substituting* $x = 5$ *into the equation of the regression line, giving* $\hat{y} = 13.419 + 3.865 \cdot (5) = 32.7$. *Similarly, the y-coordinate of the point with x-coordinate* 20 *is* $\hat{y} = 13.419 + 3.865 \cdot (20) = 90.7$.]

(b) The point (16, 82) has been circled and the line segment corresponding to the residual for this point has been drawn on the scatterplot given in the solution to part (a).

The predicted value of y for $x = 16$ is given by $\hat{y} = 13.419 + 3.865 \cdot (16) = 75.26$. So the residual for the point (16, 82) is $82 - 75.26 =$ **6.74 ounces.**

(c) The weight of the filled container for this student was 6.74 ounces more than the weight predicted by the regression line for a student who collected 16 pebbles.

(d) The values of the slope and the intercept of the least squares regression line both have meaningful interpretations in this context. The slope, 3.865, is the number of ounces by which the predicted weight of the filled container increases for each increase of one in the number of pebbles collected. The intercept, 13.419, is the value predicted by the regression line for the weight (in ounces) of the "filled container" when the number of pebbles collected is zero. This value, 13.419 ounces, is therefore an estimate of the weight of the empty container.

[*Note that since the slope is an estimate of the extra weight of the filled container when the number of pebbles collected is increased by one, it would make sense for this value,* 3.865 *ounces, to be the mean weight of all the pebbles collected by the children. If we assume that the weight of the empty container is* 13.419 *(as estimated by the regression line), then this is indeed the case. Therefore, it is acceptable to use this alternative interpretation of the slope in part (d).*]

Section II—Part B

Question Six

(a) Hypotheses:
Let p be the proportion of the members of this gym who have exercise equipment at home.

$H_0: p = 0.6$
$H_a: p < 0.6$

Check of conditions and naming of test:
This is a one-proportion z-test.

[*Note that since the type of test is named in the question it is unnecessary to name it in your solution. Nonetheless we include it here for the sake of completeness.*]

The conditions for this test are:
1. Simple random sample
2. np_0 and $n(1-p_0)$ both greater than 10

It is stated in the question that we can assume that the 83 people in Amy's sample form a random sample of the gym's members. Moreover, $np_0 = 83 \cdot 0.6 = 49.8$ and $n(1-p_0) = 83 \cdot 0.4 = 33.2$, which are both greater than 10.

Mechanics:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{42/83 - 0.6}{\sqrt{\frac{(0.6)(0.4)}{83}}} = -1.748.$$

Therefore the p-value is $P(Z < -1.748) = 0.040$. (This p-value is obtained on the TI-83 and TI-84 calculators by entering normalcdf(-999,-1.748) or normalcdf(-999,-1.748,0,1) and on the TI-Nspire by entering normalCdf(-999,-1.748,0,1). The normalcdf function is accessed on the TI-83/84 by 2nd,DISTR, and the normalCdf function is accessed on the TI-Nspire (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

Conclusion:
Since the p-value is 0.040, which is less than 0.05, we reject H_0. The result provides convincing evidence that fewer than 60% of the gym's members have exercise equipment at home.

(b) In the proposed one-proportion z-test, p_0 would be 0.6 and n would be 10, making $np_0 = 10 \cdot (0.6) = 6$ and $n(1-p_0) = 10 \cdot 0.4 = 4$, neither of which is greater than 10. Consequently, this one-sample z-test cannot be used.

(c) For each person in the sample, the probability of having exercise equipment at home is 0.6. Moreover, since we're considering a *random* sample of 10 of the gym's members, and since we're told that we can assume that the gym has a large number of members, the outcomes for the 10 people in the sample are independent. Therefore, denoting the number of people in the sample who have exercise equipment at home by X, the distribution of X is binomial, with $n = 10$ and $p = 0.6$.

Hence,

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$
$$= (0.4)^{10} + \binom{10}{1}(0.6)(0.4)^9 + \binom{10}{2}(0.6)^2(0.4)^8$$
$$= \mathbf{0.012}.$$

(In the calculation above, the expression $\binom{10}{2}$ is another way of writing ${}_{10}C_2$. Using the TI-83 and TI-84 calculators, the value of ${}_{10}C_2$ is found by typing 10 nCr 2. The nCr function is found by selecting MATH,PRB,3. On the TI-Nspire (in the *Scratchpad* or *Calculator* application), this value is found by typing nCr(10,2), and the nCr function is found by selecting menu,Probability,Combinations.

Alternatively, the answer to the question can be found directly, using a calculator. On the TI-83 or TI-84 calculators, use binomcdf(10,.6,2). The binomcdf function can be accessed by 2nd,DISTR. On the TI-Nspire, use binomCdf(10,.6,0,2), and the the binomCdf function can be found (in the *Scratchpad* or *Calculator* application) by selecting menu,Statistics,Distributions.)

(d) Yes. The answer to part (c) tells us that if exactly 60% of the gym's members had exercise equipment at home then the probability of getting 2 or fewer in a random sample of 10 of the gym's members having exercise equipment at home is 0.012, which is small. In other words, if exactly 60% of the gym's members had exercise equipment at home, then a sample result as low as 2, as obtained in Leonard's sample, would be very unlikely. Therefore we reject the hypothesis that 60% of the gym's member's have exercise equipment at home in favor of an alternative hypothesis that fewer than 60% of the gym's member's have exercise equipment at home.