Answer Key

1. **C**
2. **E**
3. **B**
4. **C**
5. **D**
6. **D**
7. **E**
8. **D**
9. **E**
10. **E**
11. **E**
12. **E**

Answers Explained

Multiple-Choice

1. **(C)** There are two observations of each mouse, a *before* time and an *after* time. These two times are dependent so a paired t-test is appropriate, *not* a two-sample test.

2. **(E)** The expected counts if dog bites occur equally during all moon phases are each $\frac{1}{4}(32 + 27 + 47 + 38) = 36$. A chi-square goodness-of-fit test gives

$$\chi^2 = \sum \frac{(obs - \exp)^2}{\exp} = \frac{(32-36)^2}{36} + \frac{(27-36)^2}{36} + \frac{(47-36)^2}{36} + \frac{(38-36)^2}{36}$$

$= 6.167$, and with $df = 4 - 1 = 3$, $P(\chi^2 > 6.167) = .1038$. With this large a P-value ($.1038 > .10$), there is not sufficient evidence to conclude that dog bites are related to moon phases.

3. **(B)** $df = (\text{rows} - 1)(\text{columns} - 1) = (2 - 1)(5 - 1) = 4$

4. **(C)** Picking separate samples from each of 16 populations and classifying according to one variable (perception of quality education) is a survey design which is most appropriately analyzed using a chi-square test of homogeneity of proportions.

5. **(D)** With $77 + 85 + 23 + 15 = 200$ samples, the expected counts if the blood type distribution on the island is the same as that of the general population are 46% of 200 = 92, 40% of 200 = 80, 10% of 200 = 20, and 4% of 200 = 8. A chi-square goodness-of-fit test gives

$$\chi^2 = \sum \frac{(obs - \exp)^2}{\exp} = \frac{(77-92)^2}{92} + \frac{(85-80)^2}{80} + \frac{(23-20)^2}{20} + \frac{(15-8)^2}{8}$$

$= 9.333$, and with $df = 4 - 1 = 3$, $P(\chi^2 > 9.333) = .0252$. With a P-value this small ($.0252 < .05$), there is sufficient evidence at the 5% significance level that blood type distribution on the island is different from that of the general population.

6. **(D)** With $df = (3 - 1)(5 - 1) = 8$, $P(\chi^2 > 13.95) = .083$. Since $.05 < .083 < .10$, there is evidence at the 10% significance level, but not at the 5% significance level, of a relationship between education level and sports interest.

7. **(E)** With a P-value this small (less than .05), there is evidence in support of the alternative hypothesis H_a: the distributions of music preferences are different, that is, they differ for at least one of the proportions.

8. **(D)** With $1 + 3 + 3 + 9 = 16$, according to the geneticist the expected number of fruit flies of each species is $\frac{1}{16}(2000) = 125$, $\frac{3}{16}(2000) = 375$, $\frac{3}{16}(2000) = 375$, $\frac{9}{16}(2000) = 1125$. A chi-square goodness-of-fit test gives $\chi^2 =$

$$\sum\frac{(obs-\exp)^2}{\exp} = \frac{(110-125)^2}{125} + \frac{(345-375)^2}{375} + \frac{(360-375)^2}{375} + \frac{(1185-1125)^2}{1125}$$

$= 8$, and with $df = 4 - 1 = 3$, $P(\chi^2 > 8) = .0460$. With a P-value this small $(.0460 < .05)$, there is sufficient evidence at the 5% significance level to reject the geneticist's claim.

9. **(E)** A chi-square test of independence gives $\chi^2 = 2.852$, and with $df = 2$, we find $P = .2403$, and since $.2403 > .10$, there is not evidence at the 10% significance level of a relationship between taste preference and the presence of the marker.

10. **(E)** A chi-square test of homogeneity gives $\chi^2 = 5.998$, and with $df = 3$, the P-value is $.1117$. With a P-value this large $(.1117 > .10)$ there is not evidence of a difference in cafeteria food satisfaction among the class levels.

11. **(E)** The relevant P-value is 0.065 which is less than 0.10 but greater than 0.05.

12. **(E)** In computer printouts of regression analysis, "S" typically gives the standard deviation of the residuals.

Free-Response

1. First, state the hypotheses: H_0: The colors of the sugar shells are distributed according to 35% cherry red, 10% vibrant orange, 10% daffodil yellow, 25% emerald green, and 20% royal purple, and H_a: The colors of the sugar shells are not distributed as claimed by the manufacturer. Or [H_0: $P_{CR} = .35$, $P_{VO} = .10$, $P_{DY} = .10$, $P_{EG} = .25$, $P_{RP} = .20$, and H_a: at least one proportion is different from this distribution.]

   Second, identify the test and check the assumptions: χ^2 goodness-of-fit test. We are given a random sample, and calculate that all expected cells are at least 5: 35% of 300 = 105, 10% of 300 = 30, 25% of 300 = 75, and 20% of 300 = 60.

   Third, calculate the test statistic χ^2 and the P-value: A calculator gives $\chi^2 = \sum\frac{(obs-\exp)^2}{\exp} = 6.689$, and with $df = 5 - 1 = 4$, $P = .153$.

   Fourth, linking to the P-value, give a conclusion in context: With a P-value this large $(.153 > .10)$, there is not evidence that the distribution is different from what is claimed by the manufacturer.

2. (a) Design I, with a single sample from one population classified on two variables (smoking and fitness), will result in a test of independence. Design II, with independent samples from two populations each with the single variable (fitness), will result in a test of homogeneity.

   (b) Design II, with its test of homogeneity, and using an equal sample size from each of the two populations (smokers and non-smokers), is best for comparing proportions of smokers who have different fitness levels with proportions of non-smokers who have different fitness levels.

   (c) Design I, which classifies one population on the two variables, smoking and fitness, is the only one of these two designs which will give data on the conditional distribution of people with given fitness levels who are smokers or are not smokers.

3. (a) First, state the hypotheses: H_0: Happiness level is independent of busy/idle choice for high school students and H_a: Happiness level is not independent of busy/idle choice for high school students.

   Second, identify the procedure and check the conditions: This is a chi-square test of independence. It is given that there is a random sample, the data are measured as "counts," and the expected counts are all at least 5 (put the observed counts in a matrix; then `χ²-Test` on the TI-84 gives expected counts of

14.6	19.6	21.8	20.2	21.8
11.4	15.4	17.2	15.8	17.2

   Third, calculate the test statistic and the P-value: Calculator software (`χ²-Test` on the TI-84) gives $\chi^2 = 14.54$ with $P = .0058$ and $df = 4$.

   Fourth, give a conclusion in context with linkage to the P-value: With a P-value this small ($.0058 < .01$), there is strong evidence of a relationship between happiness level and busy/idle choice for high school students.

   (b) No, it is not reasonable to conclude that encouraging high school students to keep more busy will lead to higher happiness levels. This was not an experiment with students randomly chosen to sit or walk. The students themselves chose whether or not to sit or walk so no cause-and-effect conclusion is possible. For example, it could well be that the happier students choose to walk, whereas the less happy students choose to sit.

4. (a) First, state the hypotheses: H_0: Eating breakfast and morning energy level are independent and H_a: Eating breakfast and morning energy level are not independent.

   Second, identify the test and check the assumptions: This is a χ^2 test of independence on

111	120	120
60	50	40

where we are given a random sample, and a calculator gives that all expected cells are at least 5:

119	119	112
51	51	48

Third, calculate the test statistic χ^2 and the P-value: Calculator software (such as `χ2-Test` on the TI-84) gives $\chi^2 = 4.202$ and $P = .1224$.

Fourth, linking to the P-value, give a conclusion in context: With this large a P-value ($.1224 > .10$), there is not evidence of a relationship between eating breakfast and morning energy level.

(b) Yes, the conclusion changes. With $n = 1000$, the observed numbers are:

220	240	240
120	100	80

with $\chi^2 = 8.403$ and $P = .0150$. With a P-value this small ($.0150 < .05$), now there is evidence of a relationship between eating breakfast and morning energy level.

5. (a) First, state the hypotheses: H_0: The different treatments lead to the same satisfaction levels and H_a: The different treatments lead to different satisfaction levels.

Second, identify the test and check the assumptions: χ^2 test of homogeneity. We are given a random sample, and a calculator gives that all expected cells are at least 5:

55.6	55.6	27.8
23.6	23.6	11.8
20.8	20.8	10.4

Third, calculate the test statistic χ^2 and the P-value: A calculator gives $\chi^2 = 10.9521$, and with $df = 4$, $P = .0271$.

Fourth, linking to the P-value, give a conclusion in context: With this small a P-value ($.0271 < .05$), there is evidence (at the 5% significance level) that the different treatments do lead to different satisfaction levels.

(b) An example of a possible confounding variable is severity of the acne outbreak. It could be that those with more severe cases have less satisfaction no matter what the treatment and are also the ones who are encouraged to use oral medications or laser therapy. So it would be wrong to conclude that oral medications or laser therapy are the causes of less satisfaction.

6. (a) Each additional year in age of teenagers is associated with an average of 0.4577 more texts per waking hour.

   (b) The scatterplot of texts per hour versus age should be roughly linear, there should be no apparent pattern in the residual plot, and a histogram of the residuals should be approximately normal.

   (c) First, state the hypotheses: H_0: $\beta = 0$, H_a: $\beta \neq 0$, where β is the slope of the regression line that relates average texts per hour to age.

   Second, identify the test and check the assumptions: This is a test of significance for the slope of the regression line, and we are given that all conditions for inference are met.

   Third, calculate the test statistic t and the P-value: The computer printout gives that $t = 2.45$ and $P = .016$.

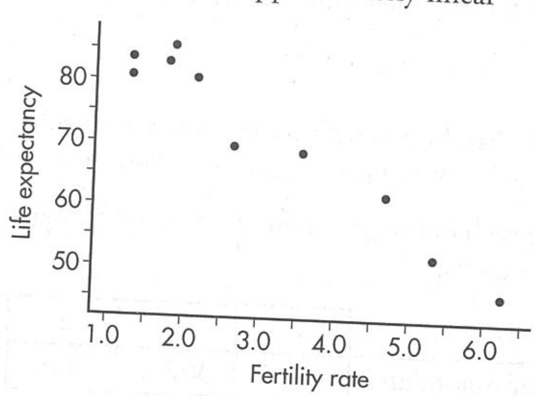
   Fourth, linking to the P-value, give a conclusion in context: With this small a P-value ($.016 < .05$), there is evidence of a linear relationship between average texts per hour and age for teenagers ages 13–17.

   (d) R-Sq = 5.8%, so even though there is evidence of a linear relationship between average texts per hour and age for teenagers ages 13–17, only 5.8% of the variability in average texts per hour is explained by this regression model (or "is accounted for by the variation in age.").

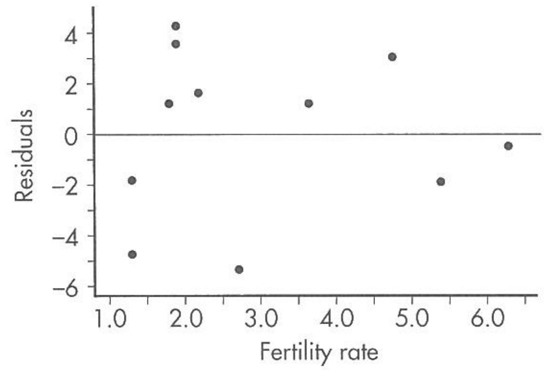
7. First, state the hypotheses: H_0: $\beta = 0$, H_a: $\beta \neq 0$. (If asked to state the parameter, then state "where β is the slope of the regression line that relates average fertility rate to women's life expectancy.")

   Second, identify the test and check the assumptions: This is a test of significance for the slope of the regression line. We are given that the data come from a random sample of countries.

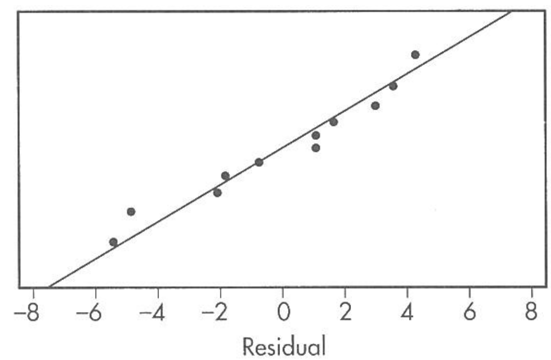
   Scatterplot is approximately linear



No apparent pattern in residual plot



Distribution of residuals is approximately normal
(Checked by normal probability plot)



Third, calculate the test statistic t and the P-value: Using a calculator (for example, `LinRegTTest` on the TI-84) gives that $t = -12.53$ and $P = .0000$.

Fourth, linking to the P-value, give a conclusion in context: With a P-value this small ($.0000 < .05$), there is evidence of a linear relationship between average fertility rate (children/woman) and life expectancy (women).

Investigative Task

1. (a) The binomial distribution with $p = 0.2$ and $n = 3$ results in: $P(0) = (.8)^3 = .512$, $P(1) = 3(.2)(.8)^2 = .384$, $P(2) = 3(.2)^2(.8) = .096$, $P(3) = (.8)^3 = .008$.

   (b) Multiplying each of the probabilities in (a) by 800 gives the expected number of occurrences:

	0	1	2	3
Expected (if binomial)	409.6	307.2	76.8	6.4

   (c) First, state the hypotheses: H_0: The number of researchers able to determine the composition of a placebo follows a binomial with $p = 0.2$ and H_a: The number of researchers able to determine the composition of a placebo does not follow a binomial with $p = 0.2$.