Answer Key

Section I

1. A	9. E	17. A	25. B	33. B
2. E	10. A	18. A	26. D	34. A
3. C	11. E	19. A	27. E	35. E
4. B	12. C	20. D	28. D	36. B
5. A	13. B	21. B	29. A	37. E
6. A	14. C	22. E	30. D	38. E
7. C	15. E	23. C	31. B	39. E
8. E	16. E	24. B	32. A	40. E

Answers Explained

Section I

1. **(A)** The slope, 6.2, gives the predicted increase in the y-variable for each unit increase in the x-variable.

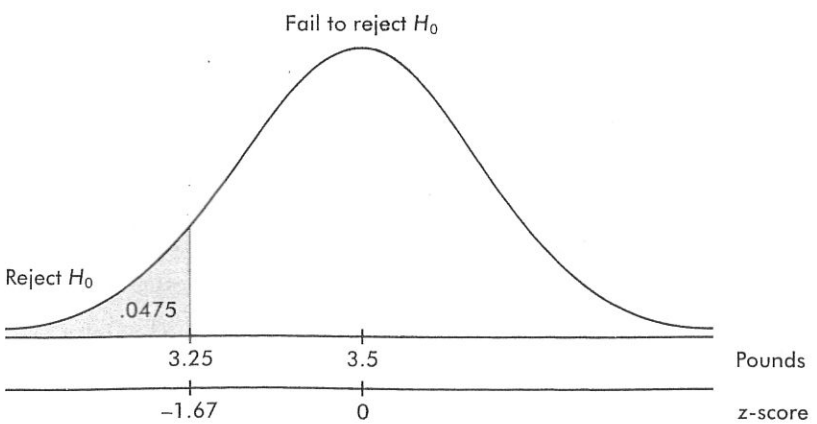
2. **(E)** For a simple random sample, every possible group of the given size has to be equally likely to be selected, and this is not true here. For example, with this procedure it will be impossible for all the early arrivals to be together in the final sample. This procedure is an example of systematic sampling, but systematic sampling does not result in simple random samples.

3. **(C)** Power $= 1 - \beta$, and β is smallest when α is more and n is more.

4. **(B)** The critical z-scores for 60% to the right and 70% to the left are -0.253 and 0.524, respectively. Then $\{\mu - 0.253\sigma = 3, \mu + 0.524\sigma = 6\}$ gives $\mu = 3.977$ and $\sigma = 3.861$.

5. **(A)** The cumulative proportions of 0.25 and 0.75 correspond to $Q_1 = 57$ and $Q_3 = 75$, respectively, and so the interquartile range is $75 - 57 = 18$.

6. **(A)** We have H_0: $\mu = 3.5$ and H_a: $\mu < 3.5$. Then $\sigma_{\bar{x}} = \frac{0.9}{\sqrt{36}} = 0.15$, the z-score of 3.25 is $\frac{3.25-3.5}{0.15} = -1.67$, and Table A gives .0475. (A t-test on the TI-83 gives .0522.)



7. **(C)** A placebo is a control treatment in which members of the control group do not realize whether or not they are receiving the experimental treatment.

8. **(E)** In experiments on people, subjects can be used as their own controls, with responses noted before and after the treatment. However, with such designs there is always the danger of a placebo effect. In this case, subjects might well have slower reaction times after drinking the alcohol because they think they should. Thus the design of choice would involve a separate control group to use for comparison. Blocking is not necessary for a well-designed experiment, and there is no indication that it would be useful here.

9. **(E)** Since $(-2, 4)$ is on the line $y = 7x + b$, we have $4 = -14 + b$ and $b = 18$. Thus the regression line is $y = 7x + 18$. The point $(\bar{x}, \bar{y})$ is always on the regression line, and so we have $\bar{y} = 7\bar{x} + 18$.

10. **(A)** A simple random sample can be any size.

11. **(E)** With such small populations, censuses instead of samples are used, and there is no resulting probability statement about the difference.

12. **(C)** Half the area is on either side of 23, so 23 is the median. The distribution is skewed to the right, and so the mean is greater than the median. With half the area to each side of 23, half the applicants' ages are to each side of 23. Histograms such as this show relative frequencies, not actual frequencies.

13. **(B)** While the procedure does use some element of chance, all possible groups of size 75 do not have the same chance of being picked, so the result is not a simple random sample. There is a real chance of selection bias. For example, a number of relatives with the same name and all using the same long-distance carrier might be selected.

14. **(C)** While t-distributions do have mean 0, their standard deviations are greater than 1.

15. **(E)**
$$P(\text{smoker}|\text{cancer}) = \frac{P(\text{smoker} \cap \text{cancer})}{P(\text{cancer})} = \frac{.014}{.035} = .4$$

16. **(E)** The critical z-scores are $\frac{80{,}000 - 75{,}000}{12{,}000} = 0.42$ and $\frac{100{,}000 - 75{,}000}{12{,}000} = 2.08$, with corresponding right tail probabilities of .3372 and .0188. The probability of being less than 100,000 given that the mileage is over 80,000 is $\frac{.3372 - .0188}{.3372} = .94$.

17. **(A)** $(1.645)\frac{\sqrt{(0.8)^2 + (0.5)^2}}{\sqrt{n}} \leq 0.25$ gives $\sqrt{n} \geq 6.208$ and $n \geq 38.5$.

18. **(A)** The correlation coefficient is not changed by adding the same number to every value of one of the variables, by multiplying every value of one of the variables by the same positive number, or by interchanging the x- and y-variables.

19. **(A)** The binomial distribution with $n = 2$ and $p = .8$ is $P(0) = (.2)^2 = .04$, $P(1) = 2(.2)(.8) = .32$, and $P(2) = (.8)^2 = .64$, resulting in expected numbers of $.04(200) = 8$, $.32(200) = 64$, and $.64(200) = 128$. Thus,

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp} = \frac{(10-8)^2}{8} + \frac{(80-64)^2}{64} + \frac{(110-128)^2}{128}$$

20. **(D)** Option II gives the highest expected return: $(50{,}000)(.5) + (10{,}000)(.5) = 30{,}000$, which is greater than 25,000 and is also greater than $(100{,}000)(.05) = 5000$. Option I guarantees that the \$20,000 loan will be paid off. Option III provides the only chance of paying off the \$80,000 loan. The moral is that the highest expected value is not automatically the "best" answer.

21. **(B)** $P(X \cap Y) = P(X|Y)\,P(Y) = (.28)(.40) = .112$. Then $P(Y|X) = \frac{P(X \cap Y)}{P(X)} = \frac{.112}{.35} = .32$.

22. **(E)** This study is an experiment because a treatment (extensive exercise) is imposed. There is no blinding because subjects clearly know whether or not they are exercising. There is no blocking because subjects are not divided into blocks before random assignment to treatments. For example, blocking would have been used if subjects had been separated by gender or age before random assignment to exercise or not.

23. **(C)** From the shape of the normal curve, the answer is in the middle. The middle two-thirds is between z-scores of ± 0.97, and $35 \pm 0.97(10)$ gives (25.3, 44.7).

24. **(B)** The interquartile range is the length of the box, so they are not all equal. More than 25% of the patients in the A group had over 210 minutes of pain relief, which is not the case for the other two groups. There is no way to positively conclude a normal distribution from a boxplot.

25. **(B)** A scatterplot would be horizontal; the correlation is zero.

26. **(D)** $\frac{.54}{.54+.19+.07} = .675$

27. **(E)** The midpoint of the confidence interval is .08.

28. **(D)** While the sample proportion is between 64% and 70% (more specifically, it is 67%), this is not the meaning of ±3%. While the percentage of the entire population is likely to be between 64% and 70%, this is not known for certain.

29. **(A)** With $df = 43.43$ and $t = -3.94$, the P-value is $.000146 < .01$. [On the TI-84, use `tcdf`.]

30. **(D)** Increasing the sample size by a multiple of d^2 divides the standard deviation of the set of sample means by d.

31. **(B)** In this binomial situation, the probability that a car does not receive a ticket is $1 - 3 = .7$, the probability that none of the five cars receives a ticket is $(.7)^5$, and thus the probability that at least one receives a ticket is $1 - (.7)^5$.

32. **(A)** $b_1 = r\frac{s_y}{s_x} = .42\frac{1.7}{11{,}500} = 0.000015$

33. **(B)** If the standard deviation of a set is zero, all the values in the set are equal. The mean and median would both equal this common value and so would equal each other. If all the values are equal, there are no outliers. Just because the sample happens to have one common value, there is no reason for this to be true for the whole population. Statistics from one sample can be different from statistics from any other sample.

34. **(A)** The first study is an experiment with two treatment groups and no control group. The second study is observational; the researcher did not randomly divide the subjects into groups and have each group watch a designated number of hours of television per night.

35. **(E)** This refers only to very particular random variables, for example, random variables whose values are the numbers of successes in a binomial probability distribution with large n.

36. **(B)** A Type I error means that the null hypothesis is correct (the weather will remain dry), but you reject it (thus you needlessly carry around an umbrella). A Type II error means that the null hypothesis is wrong (it will rain), but you fail to reject it (thus you get drenched).

37. **(E)** If the sample statistic is far enough away from the claimed population parameter, we say that there is sufficient evidence to reject the null hypothesis. In this case the null hypothesis is that $\mu = 9500$. The P-value is the probability of obtaining a sample statistic as extreme as the one obtained if the null hypothesis is assumed to be true. The smaller the P-value, the more significant the difference between the null hypothesis and the sample results. With $P = .0069$, there is strong evidence to reject H_0.

38. **(E)** There are 10 + 12 = 22 students in the combined group. In ascending order, where are the two middle scores? At least 5 third graders and 6 fourth graders have heights less than or equal to 49 inches, so at most 11 students have heights greater than or equal to 49 and thus the median is less than or equal to 49. At least 5 third graders and 6 fourth graders have heights greater than or equal to 47 inches, so at most 11 students have heights less than or equal to 47 and thus the median is greater than or equal to 47. All that can be said about the median of the combined group is that it is between 47 and 49 inches.

39. **(E)** Good experimental design aims to give each group the same experiences except for the treatment under consideration, Thus, all three SRSs should be picked from the same grade level.

40. **(E)** The stemplot does not indicate what happened for any individual school.