AP Stats Diagnostic Test #1 KEY (TPS4e)

Answer Key for Practice Exam 1
Multiple-Choice

1. Since the process produces weights that are approximately normally distributed. $P\left(z < \frac{7.5-8}{0.27}\right) = .0320$. Since the chocolate bars are independent:

P(both are less than 7.5) = $(0.032)^2 = 0.001$

ANS: A

2. For $n = 10$, there are $n-2$ degrees of freedom in regression on slope. Therefore, df = 8 and $t^* = 2.449$. Weeks of experience is the predictor, so the slope coefficient is 0.1637

ANS: C

3. If a sampling process is biased, taking a larger sample will not correct the problem.

ANS: E

4. Use the formula $\text{Bound} = z^* \frac{s}{\sqrt{n}}$. Thus, $500 = 2.326\left(\frac{2650}{\sqrt{n}}\right)$ and $n > (12.33)^2 = 152.03$.

ANS: D

5. E(red) = $n_1 p_1 + n_2 p_2 = (50)(.5) + (50)(.3) = 40$

$Var(\text{red}) = Var(\text{holiday 1}) + Var(\text{holiday 2}) = n_1 p_1 (1-p_1) + n_2 p_2 (1-p_2)$

SD(red) = $\sqrt{50(0.5)(0.5)+50(0.3)(0.7)} = 4.796$

ANS: B

6. The height of the tree depends on its distance from the shoreline, therefore distance is the explanatory variable and height is the response variable.

ANS: C

7. There are four varieties of tomato plant, two types of soil, and two fertilizers: $4 \times 2 \times 2 = 16$ treatments.

ANS: A

8. To increase power you can increase n and increase α.

ANS: E

9. (A) wrong minimum

(B) correct median, wrong Q_1

(C) correct median, wrong Q_1

(E) wrong maximum

ANS: D

10. The monthly median draft number is trending lower as the year progresses. People born the near end of the year tended to have lower lottery numbers than those born near the first of the year. No regression can be computed using only the medians.

ANS: C

11. This is a binomial distribution with $n = 20$ and $p = 0.15$.

$$P(x \leq 1) = P(x = 0) + P(x = 1) = {}_{20}C_0(0.85)^{20} + {}_{20}C_1(0.15)^1(0.85)^{19}$$

ANS: C

12. The *P*-value is the measure of the probability of getting a sample result at least as extreme as the observed result you have, given that the population mean is really what is hypothesized.

ANS: A

13. Since you are selecting a random sample of marinas and then asking everyone within each marina, this is an example of cluster sampling (taking all from some).

ANS: D

14. The null hypothesis always includes an equal sign. The one-sided alternate hypothesis in this case is less than 48 ounces.

ANS: A

15. If the week plays no role in the number of tickets issued, then each week would be expected to yield 141 tickets. The question is how well do the data "fit" this expected distribution. You are also dealing with counts, i.e., the number of tickets issued. This is chi-square goodness-of-fit .

ANS: B

16. (A) False. The confidence either contains 0 or it does not.
(B) False. The data are clearly paired.
(C) False. Even though the mean difference is positive, it is not positive enough to be significant.
(D) False. A sample of size 10 can be sufficient to make a conclusion.
(E) True. Since the confidence interval contains zero, there is no statistically significant difference between the two means. We cannot conclude that the pill is effective in improving gas mileage. The *P*-value is 0.083, which does not show strong evidence against the null hypothesis.

ANS: E

17. By looking at outliers, it is clear that the two lowest values can be used to eliminate choices (E) and (B). The lowest value is less than 15 and the next value is around 20. The median for choice (D) is too low for the histogram and the median for choice (A) is too high.

ANS: C

18. (A) The range of retail stores is larger (40 > 24).
(B) It is skewed toward the left (the lower numbers)
(C) Median retail = 73 and median manufacturer-direct = 84
(D) *IQR* retail = 78 – 62.5 = 15.5 and *IQR* manufacturer-direct = 87 – 81 = 6
(E) True. Both the median and the mean ratings are lower for retail than for manufacturer-direct.

ANS: E

19. Confidence intervals are used to estimate the population mean when only a sample of the data is known. The administrator has information on the entire population (a census).

ANS: A

20. Since the female age distribution is strongly skewed to the left, samples of size 20 will still leave the sampling distribution somewhat skewed to the left, based on the Central Limit Theorem, and the mean is equal to the mean of the original population.

ANS: C

21. Since $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, multiplying the sample size by 4 would divide the standard deviation of the sampling distribution for the mean by $\sqrt{4n} = 2\sqrt{n}$, i.e., in half.

ANS: B

22. This is a matched-pairs situation since you have before and after scores on 15 subjects. Thus, a one-sample *t* test would be used (a *t* test on paired differences).

ANS: E

23. A statistic is <u>resistant</u> if changing one of the extreme values in the distribution does not change the value of the statistic. Only the median and *IQR* exhibit this particular property.

ANS: D

24. From the *z*-table or by using a calculator, InvNormal(0.20) = –0.84 and InvNormal (0.90) = 1.28

$$z = \frac{x-\mu}{\sigma} \Rightarrow -0.84 = \frac{8.61-\mu}{\sigma} \text{ and } 1.28 = \frac{10.48-\mu}{\sigma}$$

Solving the system: $\begin{cases} -0.84\sigma = 8.61 - \mu \\ 1.28\sigma = 10.48 - \mu \end{cases}$, $\mu = \$9.35$ and $\sigma = \$0.88$

ANS: A

25. This is a matched pairs design testing the difference between the old and new methods on the 9 subjects. Using a right-tail one-sample t-test with degrees of freedom = 9 – 1 = 8, the t_{Critical} value is 1.860.

ANS: D

26. For Course A: $z = \dfrac{80-76}{8} = 0.5$. Therefore, Course B needs to be played with a z-score of the same value, 0.50.

Therefore, $0.5 = \dfrac{x-80}{6}$. Solve for x.

ANS: B

27. This is a binomial distribution with $n = 200$ and $p = 0.78$.

For the sampling distribution of proportions: $\mu_p = p$ and $\sigma_p = \sqrt{\dfrac{p(1-p)}{n}}$

$\mu_p = 0.78$ and $\sigma_p = \sqrt{\dfrac{0.78(0.22)}{200}}$

ANS: D

28. A Type II error occurs when a false null hypothesis is not rejected. In this case the null hypothesis is that the oysters have at least 5 ppb of PCB and are therefore unsafe. The error is not finding evidence that the oysters are safe and thus declaring them unsafe (fail to reject the null) when, in fact, they are fine to eat. Perfectly good oysters are thrown out and the fishermen don't get paid.

ANS: E

29. The distribution of hours spent surfing the Web exhibits more variation (has a larger *IQR*) than does the distribution of hours spent watching TV.

ANS: D

30. The mean or the expectation on each roll is $E(\text{winnings}) = (\$5)\left(\frac{1}{3}\right) + (-\$2)\left(\frac{2}{3}\right) = \$\frac{1}{3}$

with a standard deviation of $\sigma(\text{winnings}) = \sqrt{\left(5-\frac{1}{3}\right)^2\left(\frac{1}{3}\right) + \left(-2-\frac{1}{3}\right)^2\left(\frac{2}{3}\right)} = \$3.30$

For 30 turns, the expected winnings are $E(total) = 30\left(\frac{\$1}{3}\right) = \$10$

The standard deviation of the 30 turns is $\sigma(\text{total}) = \sqrt{30(3.30)^2} = \$18.07$

$$P(\text{total} > \$15) = P\left(z > \frac{\$15 - \$10}{\$18.07}\right) = P(z > 0.277) \approx 0.3910$$

ANS: <u>E</u>

31. P(airport and not SUV) = (0.70)(0.80) = 0.56
P(downtown and not SUV) = (0.30)(0.60) = 0.18
P(not SUV) = 0.56 + 0.18 = 0.74

ANS: <u>A</u>

32. A confidence interval is given by statistic ± margin of error. The more variation that exists in the parent population, the higher will be the margin of error.

ANS: <u>E</u>

33. $$P(\text{private} \mid \text{left}) = \frac{P(\text{private} \cap \text{left the company})}{P(\text{left the company})} = \frac{(0.3)(0.35)}{(0.3)(0.35) + (0.7)(0.2)} = 0.4286$$

ANS: <u>B</u>

34. x = number of customers and y = daily sales. Start with $\hat{y} = a + bx$.

$b = r\left(\frac{s_y}{s_x}\right) = 0.862\left(\frac{7.442}{182}\right) = 0.0352$ and $a = \bar{y} - b\bar{x} = 38.778 - (0.0352)(721.47) = 13.382$

Therefore the equation is $\hat{y} = 13.382 + 0.0352x$. If x = 800, then $\hat{y}$ =41.546.

ANS: <u>E</u>

35. Transforming the data from hours to minutes will change all of the numerical values except the z-score for each data point since no data point would change its position relative to the mean. The z-score is given in terms of standard deviations from the mean, which have no units such as hours or minutes.

ANS: E

36. Since the four fields are different and we want to account for these differences we will block by field and plant all six varieties of seed in each field. The varieties of seed corn are the treatments that are being applied to the fields.

ANS: B

37. Based on the boxplots of yield versus field, the median yields do not appear to be much different for the four fields. There is also a lot of overlap in the four boxplots. Looking at the boxplots of yield versus variety of seed corn, we see substantial differences in the yields. Variety 6 has the lowest median yield followed by variety 4. These are significantly lower that the remaining four varieties. Varieties 1, 2, 3, and 5 exhibit smaller differences between them, but without some summary statistics we are not able to perform a test of significance to see if there really is a statistically significant difference between these latter four.

ANS: B

38. This question is asking for an interpretation of the confidence level not the interval. The 98% level indicates how often, on average, the procedure will produce an interval that will capture the true population parameter of interest.

ANS: C

39. If a linear model were a good fit, the scatterplot would exhibit no pattern in the residuals. The curved nature of the residual plot indicates that a nonlinear model is more appropriate.

ANS: E

40. We would reject the null hypothesis if the P-value < the α level of significance. With a P-value of 0.0760, we would fail to reject H_0 at the 5% level, eliminating choices A, B, and C. The P-value would be sufficient to reject at the 10% level. Choice (D) indicates that they are independent (i.e., no change), which is equivalent to failing to reject H_0. Choice (E) rejects H_0 and concludes that a difference exists.

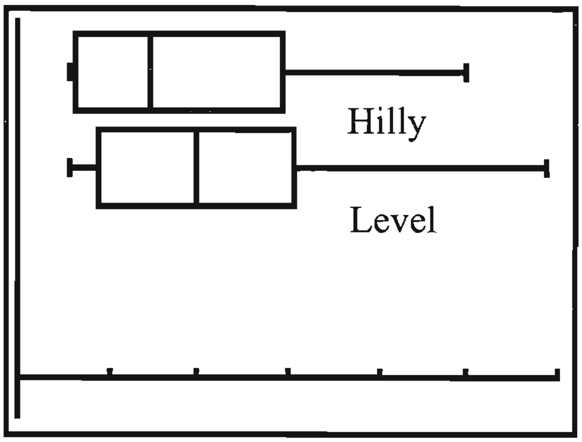
ANS: E

Textbook Correlation Multiple-Choice Practice Exam 1

Question	Correct Answer	Textbook Section	Question	Correct Answer	Textbook Section
1	A	2.2 & 5.3	21	B	7.3
2	C	12.1	22	E	9.3
3	E	4.1	23	D	1.3
4	D	8.3	24	A	2.2
5	B	6.3	25	D	9.1 & 9.3
6	C	3.1 & 4.2	26	B	2.2
7	A	4.2	27	D	7.2
8	E	9.1	28	E	9.1 & 9.3
9	D	1.2 & 1.3	29	D	1.3
10	C	3.1	30	E	6.2
11	C	6.3	31	A	5.2
12	A	9.1	32	E	8.1
13	D	4.1	33	B	5.3
14	A	9.1 & 9.3	34	E	3.2
15	B	11.1	35	E	1.3 & 2.1
16	E	10.2	36	B	4.2
17	C	1.2 & 1.3	37	B	1.3 & 10.2
18	E	1.2 & 1.3	38	C	8.1
19	A	8.1	39	E	3.2 & 12.2
20	C	7.1 & 7.3	40	E	11.2

Answer Key for Practice Exam 1
Free Response

(a)



(b) In viewing graphical displays of data we want to look for measures of center, shape, and spread. The median for the corn yields from hilly ground is about 10 (hundred) bushels smaller than the median for the corn yields from level ground. Both distributions are skewed to the right. The interquartile range is approximately equal for both corn yield distributions. Neither distribution has an outlier. Though it appears that level ground produces slightly higher yields, the data are not conclusive since there is so much overlap between the two boxplots.

(c) You should perform a two-sample t test on the difference in means.

(d) The conditions for conducting the test appear to be met. The plots of land were randomly selected, and the boxplots indicate that no apparent outliers exist. We can also consider the plots on the hillsides to be independent of the plots on level ground.

2.

(a) This is an experiment because treatments (herbal compound and a placebo compound) were imposed on two randomly assigned groups.

(b) Since subjects may respond favorably to any treatment, even a placebo, researchers will need to measure for this effect. In addition, medical conditions often improve without treatment. Without this group for comparison, the researchers will not be able to assess the effectiveness of the herbal compound in preventing the common cold, i.e., will not be able to determine if any differences exist between the two groups.

(c) Let p_1 = true proportion (rate) of people who would drink the herbal compound that would contract a common cold and p_2 = true proportion (rate) of people who would drink the placebo compound that would contract a common cold.

$H_o : p_1 = p_2 \text{ versus } H_a : p_1 < p_2$

(d) The company should not advertise that the herbal product reduces the chance of contracting a cold for everyone. Only college-age students were used in the study and the results for this group may not be applicable to groups of other ages. College students may be more (or less) susceptible to the common cold than other groups, given their close living proximity in college housing.

3.
(a)

Let p = true proportion of adults who would agree with the statement, "I should exercise more than I do."

A one-sample confidence interval for proportions will be constructed

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Random: We are told that a random sample of adults was taken.
- Normal: $n \cdot \hat{p} = 1500(0.68) = 1020$ and $n(1-\hat{p}) = 1500(0.32) = 480$
  Since both are greater than 10, the sample size is large enough to use z procedures.
- Independent: There are more than $10(1500) = 15,000$ adults in the United States.

$$0.68 \pm 2.054\sqrt{\frac{(0.68)(0.32)}{1500}} \Rightarrow (0.655, 0.705)$$

We are 96% confident that the true proportion of adults who would agree with the statement, "I should exercise more than I do," lies between 0.655 and 0.705.

(b) There is a good possibility of response bias. Given all of the media attention concerning the issue of overweight Americans, some of the respondents in the survey may feel compelled to agree with the statement in order to make themselves appear more health-conscious than they really are. This would lead to an overestimation of the proportion of people who would agree with the above statement.

4.

(a)

We will perform a chi-square test for independence.

H_o: Location of drink and type of drink purchased are independent (no relationship exists).
H_a: Location of drink and type of drink purchased are not independent (a relationship exists).

- Random: We are told that a random sample of students at this school was taken.
- Large Sample Size: All expected cell counts are at least 5.

$$\text{Expected counts} \begin{bmatrix} 69.8 & 17.1 & 55.1 \\ 34.4 & 8.4 & 27.2 \\ 75.7 & 18.5 & 59.7 \end{bmatrix}$$

- Independent: We must assume that there are more than $10(366) = 3660$ students in the school.

All the conditions are met so we can use the χ^2 test.

By calculator, the value of $\chi^2 = 38.89$. The P-value is less than 0.0000000734.
Degrees of freedom = (row − 1)(column − 1) = $(3-1)(3-1) = 4$.

If there was no relationship between location and type of drink selected, one could expect to find a distribution of counts at least this extreme in about 7 cases in 100,000,000. This is so rare as to not be plausible at any reasonable alpha value. Therefore we reject H_o. There is sufficient evidence to conclude that a student's beverage choice depends on location.

(b) Since we took a random sample of all students in the school and if we believe that the one-week period is representative of all weeks in the school, then $\frac{142}{366}$ is a reasonable estimate of the proportion of all students who buy their beverages in the lobby of the gym.

6.

(a) Let x = the number of games in which the team scores the first overtime goal.
P(the team scores first) = 0.72

Let the digits 00 through 71 represent scoring the first goal in overtime. Let 72 through 99 represent not scoring the first goal in overtime. Two digits will represent one game and it will take 12 digits to represent the six games played in overtime. Starting at the left, taking two digits at a time, count the number of games in which the team scores first until you have counted six games. For each trial, record the number of games in which the team scored the first goal. Mark the end of each trial with a vertical line. Above each trial write the number of games in which the team scored the first overtime goal.

(b)

5 5 6 5
8 4 1 7 7 0 6 7 5 7 1 7 |6 1 3 1 5 5 8 2 5 1 5 0| 6 8 1 4 3 5 4 1 0 5 0 9| 2 0 3 1 0 6 4 4 9 0 5 0| 5 9

4 5 2 5
5 9 8 8 4 3 1 1 8 0| 5 3 1 1 5 8 4 4 6 9 9 4| 8 6 8 5 7 9 6 7 0 5 8 1| 1 8 4 5 1 4 7 5 0 1 1 1| 3 0 0 6

6 4 4 5
6 3 3 9 5 5 5 0| 4 1 1 5 8 6 6 0 6 5 8 9| 1 3 1 1 9 7 1 0 2 0 8 5| 9 4 0 9 1 9 3 2 0 9 4 8| 8 7 4 9 8 7

# Games	Frequency
0	0
1	0
2	1
3	0
4	3
5	6
6	2

Based on the above simulation,

$P(\text{win at least 5 games}) = P(x \geq 5) = \frac{8}{12}$

(c) If the teams were fairly evenly matched, then each team would be expected to win approximately 3 out of 6. This situation would best be represented by the histogram in Distribution B. If one team was clearly better than the other, it would be expected to win more games than its opponent and this is represented by Distribution A.