Cumulative AP Practice Test 4 Solutions

Page 799

AP4.1 e.
AP4.2 c.
AP4.3 e.
AP4.4 a.
AP4.5 b.
AP4.6 e.
AP4.7 d.
AP4.8 a.
AP4.9 e.
AP4.10 a.
AP4.11 b.
AP4.12 d.
AP4.13 e.
AP4.14 d.
AP4.15 b.
AP4.16 a.
AP4.17 d.
AP4.18 c.
AP4.19 a.
AP4.20 b.
AP4.21 e.
AP4.22 e.
AP4.23 b.
AP4.24 c.
AP4.25 d.
AP4.26 e.
AP4.27 c.
AP4.28 e.
AP4.29 c.
AP4.30 b.
AP4.31 d.
AP4.32 a.
AP4.33 e.
AP4.34 b.
AP4.35 a.
AP4.36 b.
AP4.37 d.
AP4.38 a.
AP4.39 b.
AP4.40 d.

AP4.41 **State:** We want to perform a test of

$$H_o: \mu_1 - \mu_2 = 0$$
$$H_a: \mu_1 - \mu_2 \neq 0$$

where

μ_1 = mean difference in electrical potential for diabetic mice (right hip – front feet)

μ_2 = mean difference in electrical potential for normal mice (right hip – front feet)

at the 5% significance level. **Plan:** If conditions are met, we should carry out a two-sample *t* test.

- **Random:** The researchers randomly selected diabetic mice and normal mice for the study.
- **Normal:** We are told that graphs of the data show no outliers or strong skewness, so we should be safe using *t* procedures even if the two population distributions aren't perfectly Normal.
- **Independent:** The data came from independent samples of diabetic and normal mice. Knowing one mouse's difference in electric potential should give no additional information about another mouse's difference in electric potential, so the individual measurements should be independent. There are more than 10(24) = 240 diabetic mice and 10(18) = 180 normal mice.

Do: We will conduct a two-sample *t*-test for $\mu_1 - \mu_2$ on our calculator.

- **Test statistic:** $t = 2.549$
- ***P*-value:** Using df = 38.46, the *P*-value is 0.0149.

```
2-SampTTest
  μ1≠μ2
  t=2.549830747
  p=.0148807394
  df=38.45983024
  x̄1=13.09
↓ x̄2=10.022
```

Conclude: Since the *P*-value of 0.015 < 0.05, we reject the null hypothesis and conclude that diabetic mice and normal mice differ in the mean electrical potential between their right hip and front feet.

AP4.42 (a) We want to test

$$H_o: p_1 - p_2 = 0$$
$$H_a: p_1 - p_2 < 0$$

where

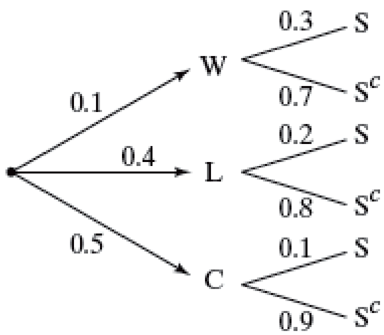
p_1 = proportion of women who were physically active as teens that would suffer a cognitive decline

p_2 = proportion of women who were sedentary as teens that would suffer a cognitive decline

(b) You would perform a two-sample *z* test for $p_1 - p_2$ (a two-proportion *z* test). (c) No. Because the participants were mostly white women, the findings may not be generalizable to women in other racial and ethnic groups. (d) Two variables are confounded when their effects on the response variable (measure of cognitive decline) cannot be distinguished from one another. For example, women who were physically active as teens might have also done other things like engaging in mentally stimulating activities that those who were sedentary during their teen years did not do. We would be unable to determine if it was their physically active youth or the other activities that slowed their level of cognitive decline.

AP4.43 (a) Since the first question called it a "fat tax", people may have reacted negatively because they believe this is a tax on those who are overweight. The second question provides extra information that gets people thinking about the obesity problem in the U.S. and the increased health care that could be provided as a benefit with the tax money, which might make them respond more positively to the proposed tax. The question should be worded in a more straightforward manner. For example, "Would you support or oppose a tax on non-diet sugared soda?" (b) This method samples only people at fast-food restaurants. They may go because they like the sugary drinks and thus don't want to pay a tax on their favorite beverages. It is likely that the proportion of those who would oppose such a tax will be overestimated with this method. A random sample of all New York State residents should be taken to provide a better estimate of the level of support for such a tax. (c) Use a stratified random sampling method in which each state is a stratum. That is, take a random sample of residents within each state.

AP 4.44 Let's start by making a tree diagram that shows the various probabilities. Let W = Mr. Worcester arrives first, L = Mr. Legacy arrives first, C = Dr. Currier arrives first, and S = the coffee is strong.



(a) $P(S) = (0.1)(0.3) + (0.4)(0.2) + (0.5)(0.1) = 0.16$

(b) $P(C \mid S) = \dfrac{P(C \cap S)}{P(S)} = \dfrac{(0.5)(0.1)}{(0.1)(0.3) + (0.4)(0.2) + (0.5)(0.1)} = \dfrac{0.05}{0.16} = 0.3125$

AP4.45 (a) A linear model is not appropriate. The scatterplot exhibits a strong curved pattern.
(b) Model B would be better to use if you wanted to predict seed weight from seed count. The scatterplot of ln(weight) versus ln(count) shows a much more linear pattern and its residual plot shows a random scatter about the residual = 0 line.

(c) $$\ln(\text{weight}) = 15.491 - 1.5222\ln(3700) \Rightarrow \ln(\text{weight}) = 2.984$$
$$\text{weight} = e^{2.984} \approx 19.77\text{mg}$$

(d) About 86.3% of the variability in ln(seed weight) is accounted for by the least-squares regression line using ln(seed count) as the explanatory variable.

AP4.46 (a) Shape: Normal; Center: $\mu_{\bar{x}}$ = 4 inches; Spread**:** $\sigma_{\bar{x}} = \dfrac{0.02}{\sqrt{25}}$ = 0.004 inches.
(b) $P(\bar{x} < 3.99) + P(\bar{x} > 4.01)$ = 1 – normalcdf(3.99,4.01,4.00,.004) = 0.0124.
(c) $P(4.00 < \bar{x} < 4.01)$ = normalcdf(4, 4.1, 4, .004) = 0.4938.
(d) Let X = the number of samples (out of 5) in which the mean is between 4.00 and 4.01. The random variable X is binomially distributed with n = 5 and p = 0.4938.
$P(X \geq 4) = 1 - P(X < 4) = 1 - P(X \leq 3) = \text{binomcdf}(5, .4938, 3) = 0.1798.$

(e) Since $P(\overline{x} < 3.99) + P(\overline{x} > 4.01) = 0.0124$ from part (b), whereas $P(X \geq 4) = 0.1798$ from part (d), Method 1 would give more convincing evidence that the machine needs to be shut down, since it is much less likely to occur by chance if the machine is working correctly. (f) Answers will vary. For example, stop the production process if there are 5 consecutive sample means between 4.00 and 4.01. Using this rule: Probability = $(0.4938)^5 = 0.029$. An analogous rule would be to consider 5 consecutive sample means between 3.99 and 4.00. Another possibility is to consider getting two consecutive samples means between 4.005 and 4.01. Note that $P(4.005 < \overline{x} < 4.01) = 0.0994$. If two consecutive samples are taken, then $P(\text{both between 4.005 and 4.01}) = (0.0994)^2 = 0.0099 \approx 0.01$