Homework 9

Alizeh Azhar

10/27/2021

Problem 1: Always, Sometimes, or Never (8 points)

The following statements are always, sometimes, or never true. Mark each statement with Always, Sometimes, or Never, depending on the truth of each statement

1. The Bernoulli distribution is based on the geometric distribution
   Answer
   Always
2. For any two events A and B, $P(A \text{ or } B) = P(A) + P(B)$.
   Answer
   Sometimes
3. X and Y are two variables. Small values of X result in large values of Y , and large values of X result in small values of Y . X and Y are negatively related.
   Answer
   Always
4. The standard normal distribution is a discrete distribution.
   Answer
   Never
5. Box plots are good for investigating the distribution of a categorical variable.
   Answer
   Sometimes
6. If X is right-skewed, the median is less than the mean.
   Answer
   Always
7. A sample is a subset of a population.
   Answer
   Always
8. In a valid probability distribution, the sum of the probabilities of the disjoint outcomes must sum to one.
   Answer
   Always

Problem 2: OIS Exercise 5.19 page 202 (4 points)

Teens were surveyed about cyberbullying, and 54% to 64% reported experiencing cyberbullying (95% confidence interval). Answer the following questions based on this interval.

(a) A newspaper claims that a majority of teens have experienced cyberbullying. Is this claim supported by the confidence interval? Explain your reasoning.

Answer
This claim is reasonable, since the entire in- terval lies above 50%.

(b) A researcher conjectured that 70% of teens have experienced cyberbullying. Is this claim supported by the confidence interval? Explain your reasoning.
Answer
The value of 70% lies out- side of the interval, so we have convincing evidence that the researcher's conjecture is wrong.

(c) Without actually calculating the interval, determine if the claim of the researcher from part (b) would be supported based on a 90% interval. Explain your reasoning.
Answer
A 90% confidence interval will be narrower than a 95% confidence interval. Even without calculating the interval, we can tell that 70% would not fall in the interval, and we would reject the researcher's conjecture based on a 90% confidence level as well.

(d) Would a hypothesis test of $H_0 : p = 0.70$ be rejected or fail to rejected, based on the confidence interval? Explain your reasoning.
Answer
The hypothesis would be rejected. This is because 0.70 is not contained in the 95% confidence interval.

Problem 3: OIS Exercise 6.20, page 226 (4 points)

The United States federal government shutdown of 2018 – 2019 occurred from December 22, 2018 to January 25, 2019, a span of 35 days. A Survey USA poll of 614 randomly sampled Americans during this time period reported that 48% of those who make less than $\$40,000$ per year and 55% of those who make $\$40,000$ or more per year said that the government shutdown has not at all affected them personally. A 95% confidence interval for $(p < 40k - p \geq 40k)$, where p is the proportion of those who said the government shutdown has not at all affected them personally, is $(-0.16, 0.02)$. Based on this information, determine if the following statements are true or false, and explain your reasoning if you identify the statement as false.

(a) At the 5% significance level, the data provide convincing evidence of a real difference in the proportion who are not affected personally between Americans who make less than $\$40,000$ annually and Americans who make $\$40,000$ annually (1 pt).
Answer
False. The interval contains point 0. This indicates that there is no significant difference between this proportion.

(b) We are 95% confident that 16% more to 2% fewer Americans who make less than $\$40,000$ per year are not at all personally affected by the government shutdown compared to those who make $\$40,000$ or more per year (1 pt).
Answer
False. We are 95% confident that 2% more to 16% fewer Americans who make less than $\$40,000$ per year are not at all personally affected by the government shutdown compared to those who make $\$40,000$ or more per year

(c) A 90% confidence interval for $(p < 40k - p \geq 40k)$ would be wider than the $(-0.16, 0.02)$ interval (1 pt).
Answer
False. The 90% confidence interval is always narrower than 95% confidence as the critical value is small for 90% confidence interval.

(d) A 95% confidence interval for $(p \geq 40k - p < 40k)$ is $(-0.02, 0.16)$ (1 pt).
Answer
True.

Problem 4: OIS Exercise 5.32, page 205 (4 pts)

It is believed that nearsightedness affects about 8% of all children. In a random sample of 194 children, 21 are nearsighted. Conduct a hypothesis test for the following question: do these data provide evidence that the 8% value is accurate?

Answer

The Hypothesis test for the question is:

$$H_0 : p = 0.08$$

vs

$$H_A : p \neq 0.08$$

```
#proportion
p_hat <- 21/194
p_hat
```

```
## [1] 0.1082474
```

```
#standard error
s_e <- sqrt((0.08*(1-0.08))/194)
s_e
```

```
## [1] 0.01947772
```

```
#test statistic
z_stat <- (p_hat-0.08)/s_e
z_stat
```

```
## [1] 1.450243
```

```
#p- value
p_value <-  pnorm(-z_stat, 0, 1)+ pnorm(z_stat,0,1, lower.tail = F);p_value
```

```
## [1] 0.1469908
```

The p- value is 0.147 and α is 0.05. 0.147>0.05. We fail to reject the null hypothesis. There is not enough evidence to suggest that the true proportion of children affected by nearsightedness is statistically different from 0.08