Week 4 Quiz

Back to Week 4

11/12 points earned
(91%)

Quiz passed!

1 / 1
points

1.
Suppose you want to construct a confidence interval for a population proportion. Which of the following, if it were true, would **prevent** you from being able to assume that the distribution of the sample **proportion** is nearly normal?

- *n* = 104. Out of these 104 there are an equal number of successes and failures (52 each).
- *n* = 104. Out of these 104 there are only a few successes (15), but relatively many failures (89).
- *n* = 104. These observations are a simple random sample and make up less than 10% of the population.
- None of these options.

Correct Response
Recognize that the Central Limit Theorem (CLT) is about the distribution of point estimates, and that given certain conditions, this distribution will be nearly normal.

- In the case of the proportion the CLT tells us that if

(1) the observations in the sample are independent,

(2) the sample size is sufficiently large (checked using the success/failure condition: $np \geq 10$ and $n(1-p) \geq 10$),

then the distribution of the sample proportion will be nearly normal, centered at the true population proportion and with a standard error of $\sqrt{\frac{p(1-p)}{n}}$.

$$\hat{p} \sim N\left(mean = p, SE = \sqrt{\frac{p(1-p)}{n}}\right)$$

Review the associated learning objective.

---

1 / 1 points

2.

When checking conditions for calculating a confidence interval for a proportion, you should use which number of successes and failures?

- Observed

Correct Response

For confidence intervals use $\hat{p}$ (observed sample proportion) when calculating the standard error and checking the success/failure condition. For hypothesis tests use p_0 (null value) when calculating the standard error and checking the success/failure condition.

Use the observed number of successes and failures when calculating a confidence interval for a proportion, but not when doing a hypothesis test. In a hypothesis test for a proportion, you should use *np0* and $n(1 - p_0)$ successes and failures; that is, the expected number based on the **null proportion**.

- Not applicable. The number of successes and failures (observed or otherwise) is not part of the conditions required for calculating a confidence interval for a proportion.
- Depends on the context
- Expected (based on the null value)

---

0 / 1 points

3.

In May 2011, Gallup asked 1,721 students in grades five through twelve if their school teaches them about money and banking. Researchers are interested in finding out if a majority of students receive such education. Which of the following is the correct set of hypotheses?

- $H_0 : p < 0.5$; $H_A : p > 0.5$

Incorrect Response

This question revisits the setup of hypothesis testing within the categorical data / proportions.

Null hypothesis should set the parameter equal to a null value.

- $H_0 : p = 0.5$; $H_A : p > 0.5$
- $H_0 : \hat{p} = 0.5$; $H_A : \hat{p} \neq 0.5$
- $H_0 : \mu = 0.5$; $H_A : \mu > 0.5$

---

1 / 1 points

4.

The campaign manager for a congressional candidate claims that the candidate has **more than** 50% support from the district's electorate. A newspaper collects a simple random sample of 500 likely voters in this district and estimates the support for this candidate to be 52%. The p-value for the hypothesis test evaluating the campaign manager's claim is 0.19. Which of the below is **correct**?

- 95% of random samples of size 500 will estimate the support for this candidate to be 52%.
- If in fact 50% of likely voters support this candidate, the probability of obtaining a random sample of 500 likely voters where 52% or more support the candidate is 0.19.

Correct Response

p-value = P(observed or more extreme test statistic | H_0 true)

- The success-failure condition is not met, so this p-value is not reliable.
- The data provide convincing evidence for the campaign manager's claim.

---

1 / 1 points

5.

Voters in the U.S. state of New Jersey voted on a measure that would legalize sports betting. Before the vote, SurveyUSA asked a random sample of registered voters the following question: "'If betting on sports were to become legal in New Jersey, would you yourself place bets?" The distribution of responses by age group is shown in the table below. Consider a hypothesis test evaluating whether there is a difference between proportions of 18 to 34 year olds and 35+ year olds who would never place bets. Using the information from the table, calculate the standard error for this hypothesis test. Choose the closest answer.

		age group		
		18 to 34	35+	total
response	regularly	3	30	33
	occasionally	69	128	197
	never	71	224	295
	not sure	1	7	8
	total	144	389	533

- 0.5535
- 0.0024
- 0.0485

Correct Response

First calculate

$$\hat{p}_{pool}$$

$$= \frac{71 + 224}{144+389} \approx 0.55$$

Then $SE =$

$$\sqrt{\frac{0.55 \times (1-0.55)}{144} + \frac{0.55 \times (1-0.55)}{389}}$$

$$= 0.0421$$

- 0.5758
- 0.4931

---

1 / 1 points

6.

To evaluate the following hypotheses

$H_0 : p = 0.3$

$H_A : p \neq 0.3$

we use a random sample of 50 observations where $\hat{p} = 0.36$. Which of the following is the correct standard error? Choose the closest answer.

- 0.0679
- 0.0648

Correct Response

The question refers to the following learning objective(s):

Note that the reason for the difference in calculations of standard error is the same as in the case of the single proportion: when the null hypothesis claims that the two population proportions are equal, we need to take that into consideration when calculating the standard error for the hypothesis test, and use a common proportion for both samples.

For a hypothesis test,

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.3 \times 0.7}{50}} \approx 0.0648$$

- 0.0092
- 0.0096
- 0.0297
- 0.0042

---

1 / 1 points

7.

At the beginning of a semester an anonymous survey was conducted on students in a statistics class. Two of the questions on the survey were about gender and whether or not students have equal, more, or less energy in the afternoon compared to the morning. Below are the results.

	Equal	Less	More
Female	18	37	24
Male	9	15	24

What test should we perform to see if gender and energy level are associated?

- hypothesis test for a single mean
- ANOVA
- Chi-square test of independence

Correct Response

The question refers to the following learning objective(s):

- Use a chi-square test of goodness of fit to evaluate if the distribution of levels of a single categorical variable follows a hypothesized distribution.
- When evaluating the independence of two categorical variables where at least one has more than two levels, use a chi-square test of

independence.

- Z test
- Comparing two means
- Chi-square test of goodness of fit
- F test
- Comparing two proportions

---

1 / 1 points

8.

A variety of studies suggest that 10% of the world population is left-handed. It is also claimed that artists are more likely to be left-handed. In order to test this claim we take a random sample of 40 art students at a college and find that 6 of them (15%) are left handed. Which of the following is the correct set-up for calculating the p-value for this test?

- In a bag place 40 chips, 6 red and 34 blue. Randomly sample 40 chips, with replacement, and record the proportion of red chips in the sample. Repeat this many times, and calculate the proportion of samples where at least 10% of the chips are red.
- Roll a 10-sided die 40 times and record the proportion of times you get a 1. Repeat this many times, and calculate the proportion of simulations where the sample proportion is 15% or more.

Correct Response

The question refers to the following learning objective(s):

In hypothesis testing for one categorical variable, generate simulated samples based on the null hypothesis, and then calculate the number of samples that are at least as extreme as the observed data.

In this problem, generating simulated samples based on the null hypothesis corresponds to rolling the 10-sided die 40 times. Note that we're treating a rolled 1 as "left-handed" where a rolled 2 through rolled 9 are considered "right-handed".

- Roll a 10-sided die 40 times and record the proportion of times you get a 1. Repeat this many times, and calculate the proportion of simulations where the sample proportion is 10% or more.

- Randomly sample 40 non-art students, and record the number of left-handed students in the sample. Repeat this many times and calculate the proportion of samples where at least 15% of the students are left-handed.

✔ 1 / 1 points

9.

True or false: The χ^2 statistic is always non-negative.

- False
- True

Correct Response

✔ 1 / 1 points

10.

80% of Americans start the day with a cereal breakfast. Based on this information, determine if the following statement is true or false.

"The sampling distribution of the proportions of Americans who start the day with a cereal breakfast in random samples of size 40 is right skewed."

- True
- False

Correct Response

The question refers to the following learning objective(s):

Note that if the CLT doesn't apply and the sample proportion is low (close to 0) the sampling distribution will likely be right skewed, if the sample proportion is high (close to 1) the sampling distribution will likely be left skewed.

S-F condition not met, and the true population is closer to 1 than 0, so the sampling distribution will be left skewed.

---

1 / 1 points

11.

At a stop sign, some drivers come to a full stop, some come to a 'rolling stop' (not a full stop, but slow down), and some do not stop at all. We would like to test if there is an association between gender and type of stop (full, rolling, or no stop). We collect data by standing a few feet from a stop sign and taking note of type of stop and the gender of the driver. Below is a contingency table summarizing the data we collected. If gender is not associated with type of stop, how many males would we expect to not stop at all? Choose the closest answer.

		gender	
		female	male
stop	full stop	6	6
	rolling stop	16	15
	no stop	4	3

- 3.64
- 6.24
- 5.76
- 3.36

Correct Response

Calculate expected counts in two-way tables as

$$E = \frac{rowtotal \times columntotal}{grandtotal}$$

---

1 / 1
points

12.

We would like to test the following hypotheses

H_0 :p = 0.05

H_A :p < 0.05

The sample size is 150 and the sample proportion is 8%, i.e. $\hat{p}$ = 0.08. Which of the below is the correct test for this situation?

- chi-square test of independence
- randomization test for a proportion

Correct Response

Use simulation methods when sample size conditions aren't met for inference for categorical variables.

- Note that the t-distribution is only appropriate to use for means. When sample size isn't sufficiently large, and the parameter of interest is a proportion or a difference between two proportions, we need to use simulation.

In hypothesis testing

- for one categorical variable, generate simulated samples based on the null hypothesis, and then calculate the number of samples that are at least as extreme as the observed data.

- for two categorical variables, use a randomization test.

- z-test for a proportion
- t-test for a mean
- z-test for comparing two proportions