Week 4 Quiz

Quiz, 12 questions

10/12 points (83.33%)

Congratulations! You passed!

Next Item

1 / 1 point

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 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
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.

- *n* = 104. Out of these 104 there are an equal number of successes and failures (52 each).

Week 4 Quiz

Quiz, 12 questions

10/12 points (83.33%)

0 / 1 point

2.

When performing a hypothesis test on proportions (either where $H_0 : p = p_0$ or where $H_0 : p_1 = p_2$) you should use the **observed** number of successes and failures when checking conditions.

- True

This should not be selected

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 np_0 and $n(1 - p_0)$ successes and failures; that is, the expected number based on the **null proportion**.

- False
- Depends on the context
- Observed proportion for one sample, expected proportion for two samples.

1 / 1 point

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?

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

Correct

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

The wording of the question tells us we're interested in whether the true proportion of students receiving this education is greater than 50% (i.e. makes them "a majority").

Week 4 Quiz

Quiz, 12 questions

10/12 points (83.33%)

1 / 1 point

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**?

- 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

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.
- 95% of random samples of size 500 will estimate the support for this candidate to be 52%.
- The data provide convincing evidence for the campaign manager's claim.

1 / 1 point

5.

Week 4 Quiz

Quiz, 12 questions

10/12 points (83.33%)

Gallup conducts an annual poll of U.S. residents. Approximately 1,000 residents across all 50 states and Washington D.C. are asked "Do you believe the use of marijuana should be made legal?" The distribution of responses by date of survey is shown in the table below. Imagine a hypothesis test evaluating whether there is a difference from 2012 to 2013 between proportions of "yes" responses. Using the information in the table below, calculate the standard error for this hypothesis test. Choose the closest answer.

		time of survey	
		Nov 2012	October 2013
response	yes	493	596
	no	514	401
	undecided	30	31
	total	1037	1028

- 0.5274
- 0.022

Correct

First calculate

$$\hat{p}_{\text{pool}} = \frac{493 + 596}{1037+1028} \approx 0.53$$

Then *SE* =

$$\sqrt{\frac{0.53 \times (1-0.53)}{1037} + \frac{0.53 \times (1-0.53)}{1028}}$$

$$= 0.022$$

- 0.5798
- 0.4754
- 0.00048

0 / 1 point

6.

Week 4 Quiz
Quiz, 12 questions

10/12 points (83.33%)

"In statistical inference for proportions, standard error (SE) is calculated differently for hypothesis tests and confidence intervals." Which of the following is the **best** justification for this statement?

- Because in hypothesis testing, we assume the null hypothesis is true, hence we calculate SE using the null value of the parameter. In confidence intervals, there is no null value, hence we use the sample proportion(s).
- Because in hypothesis testing we're interested in the variability of the true population distribution, and in confidence intervals we're interested in the variability of the sampling distribution.

This should not be selected

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.

Review the associated learning objective.

- Because if we used the same method for hypothesis tests as we did for confidence intervals, the calculation would be impossible.
- Because statistics is full of arbitrary formulas.

✔ 1 / 1 point

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
- Chi-square test of independence

Correct

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.

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

---

1 / 1 point

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?

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

Week 4 Quiz

Quiz, 12 questions

10/12 points (83.33%)

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".

✔ 1 / 1 point

9.

One of the early studies linking smoking and lung cancer compared patients hospitalized with lung cancer to similar patients without lung cancer (hospitalized for other reasons), and recorded whether each patient smoked. For a hypothesis test testing whether the proportion of smokers is higher for the patients with lung cancer than for patients without lung cancer, the p-value is less than 0.000001. Does this provide significant evidence that smoking causes lung cancer?

		smoking status		
		smoker	non-smoker	total
cancer	lung cancer	647	2	649
	not sure	622	27	649
	total	1269	29	1298

Study reference: Doll, R. & Hill, A.B. (1950) "Smoking and carcinoma of the lung: preliminary report", *British Medical Journal.*

- No, with the given p-value we would fail to reject H0 in favor of HA.
- Whether or not we can conclude that smoking causes lung cancer depends on the statistical method the researchers used to obtain the p-value.
- Yes, with the given p-value we would reject H0 in favor of HA, and conclude that smoking causes lung cancer.
- ◉ Based on this study we cannot conclude that smoking causes lung cancer,

  regardless of the p-value.

Correct

These data was not (and could not have ethically been!) collected through a controlled experiment; therefore no causal statements can be made using the study. Recall that a controlled experiment (as opposed to an observational study) is needed in order to determine causality.

Week 4 Quiz

Quiz, 12 questions

1 / 1 point

10/12 points (83.33%)

10.

Suppose in a population 20% of people wear contact lenses. What is the expected shape of the sampling distribution of proportion of contact lens wearers in random samples of 30 people from this population?

- nearly normal
- left-skewed
- uniform
- right-skewed ✓

Correct

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 0 than 1, so the sampling distribution will be right skewed.

---

1 / 1 point

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. What are the hypotheses for testing for an association between gender and type of stop?

- H_0: Males and females are equally likely to come to a full stop.

  H_A: Males and females are not equally likely to come to a full stop.

- H_0: Males and females are equally likely to come to a rolling stop.

  H_A: Males are more likely than females to come to a rolling stop.

- H_0: Gender and type of stop are independent. ✓

  H_A: Gender and type of stop are associated.

Week 4 Quiz

Quiz, 12 questions

10 / 12 points (83.33%)

Correct

When evaluating the independence of two categorical variables where at least one has more than two levels, use a chi-square test of independence.

H_0 : The two variables are independent.

H_A : The two variables are dependent.

- H_0: Gender and type of stop are associated.

  H_A: Gender and type of stop are independent.

---

1 / 1 point

12.

True / False: A randomization test for $H_0 : p_1 = p_2$ uses a randomization distribution centered at 0.

- True

Correct

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.

- False

Week 4 Quiz

Quiz, 12 questions

10/12 points (83.33%)