RLP-STATSA · RLP Statistics A — Exploring Data, Study Design and Probability

Unit 3: Designing Studies: Samples and Experiments

Practice Workbook. The same problems, in the same order, as the Practice sections of the online lessons.

Name _____ Period _____

Lesson 3.1 — Populations, Samples and Bias

FLUENCY

1. (Hypothetical data.) A survey of 400 randomly chosen households in a Missouri city finds that 23% have solar panels. Identify the population, the parameter, the sample and the statistic.

2. Parameter or statistic? The mean daily high of the 30 randomly chosen days in a sample.

3. Parameter or statistic? The proportion of all 9,132 days from 2000 to 2024 on which St. Louis Lambert recorded rain.

4. Name the sampling method: a school lists all 1,200 students, picks a random start between 1 and 20, and surveys every 20th student.

5. Name the sampling method: a district randomly selects 6 of its 40 homerooms and surveys every student in them.

6. Name the sampling method: a county takes a separate simple random sample of voters in each of its five townships.

7. Name the sampling method: a reporter interviews shoppers at one mall on a Saturday.

8. Does increasing the sample size reduce bias?

APPLICATION

9. (Hypothetical.) A radio station asks listeners to text whether the city should build a new stadium; 78% of the 2,300 texts say yes. Identify the type of sample and the likely direction of bias.

10. A phone survey of a county's residents uses a list of landline numbers only. What is the bias, and who is likely missing?

11. A survey question reads: 'Given the dangerous condition of our roads, do you support a small tax to fix them?' Identify the problem and rewrite the question.

12. The American Community Survey requires households to respond by law and follows up by mail, internet, phone and in-person visits. Which kind of bias is this design mainly fighting?

CHALLENGE

13. A district of 3,000 high-school students (four grades of roughly equal size) wants to estimate average nightly sleep, and suspects seniors sleep less. Design a stratified sample of 200 and explain why it beats a simple random sample of 200.

REVIEW

14. Use technology to find the least-squares line and the correlation coefficient for (1, 8), (2, 3), (4, 6), (11, 12), (12, 16), (13, 13).

15. (Unit 2) A regression of y on x has r = −0.6. What fraction of the variation in y is explained by the linear model?

16. (Unit 1) Which summary is resistant: the SD or the IQR?

Lesson 3.2 — Simulating Sampling Variability and the Margin of Error

FLUENCY

1. Simulated samples of 100 days give $\hat{p}$'s with SD 0.046. What SD would you expect for samples of 400? Of 1,600?

2. Samples of 25 give $\hat{p}$'s with SD about 0.092. Estimate the margin of error (about 2 SDs).

3. For samples of 100 days, the middle 95% of simulated $\hat{p}$'s ran from 0.22 to 0.40, around a true p of 0.3117. Is random sampling biased here?

4. To halve a margin of error, by what factor must the sample size grow?

5. True or false: a margin of error of ±3 points means the survey cannot be off by more than 3 points.

6. A simulated sampling distribution of $\hat{p}$ for n = 25 is lumpy and slightly skewed right, while n = 400 gives a smooth, symmetric one. Why?

APPLICATION

7. A friend claims it rains on one day in five (20%) of St. Louis days. A simulation of 1,000 random samples of 100 days from a population with $p = 0.20$ produced $\hat{p} \geq 0.30$ in 15 samples (seed 2026). A real random sample of 100 St. Louis days gave $\hat{p} = 0.30$. What do you conclude?

8. Using the scaling rule, about how many randomly sampled days would give a margin of error of 0.02 for the proportion of rainy days, if $n = 100$ gives about 0.092?

9. Describe how you would use simulation to estimate a margin of error for the mean daily high temperature at St. Louis from a random sample of 30 days, using the 2000-2024 record as the population.

CHALLENGE

10. A newspaper runs an online poll with 50,000 responses and reports a margin of error of ±0.4 points. Explain why this margin of error is misleading.

REVIEW

11. (Unit 2) Missouri's cancer death-rate line (CDC, age-adjusted rate per 100,000) is predicted rate = 210.66 − 2.353x (x = years since 1999). Predict 2015.

12. (Unit 1) For 3, 6, 6, 7, 13 find the median and IQR.

Lesson 3.3 — Experiments, Observational Studies and Confounding

FLUENCY

1. Survey, observational study or experiment? Researchers compare asthma rates of children living near busy highways with those living far from them.

2. Survey, observational study or experiment? 60 volunteers are randomly assigned to type an essay with or without background music, and errors are counted.

3. Survey, observational study or experiment? A random sample of 1,000 adult residents of a state is asked whether they support a ballot measure.

4. What does random assignment accomplish in an experiment?

5. What is a placebo, and why use one?

6. In a double-blind experiment, who does not know which treatment each subject received?

APPLICATION

7. Students who eat school breakfast have higher average test scores than those who don't. Name two confounding variables and say whether this shows breakfast improves scores.

8. Design a randomized experiment to test whether a 10-minute walk before a quiz raises scores, using 40 volunteer students.

9. In the 1954 polio vaccine field trial, some areas used 'observed controls': second graders whose parents consented got the vaccine, and all first and third graders were watched as controls (17.1 cases per 100,000 among the vaccinated, 45.5 among controls). In the randomized, placebo-controlled areas the rates were 16.4 and 57.1. Why is the observed-control comparison weaker?

CHALLENGE

10. A coach wants to compare two training methods, but her 40 athletes come from two practice groups, and group 1 is much stronger at the start. Describe a blocked design and explain what blocking adds.

REVIEW

11. For the data 12, 18, 18, 20, 22, 26, 27, 28, 34: find the mean, median, range and interquartile range. (Find the quartiles as the medians of the lower and upper halves, leaving out the overall median when the count is odd.)

12. (Unit 2) Classify: cities with more police officers have more crimes, because larger cities have more of both.

Lesson 3.4 — Comparing Treatments with a Randomization Test

FLUENCY

1. Group A: 12, 15, 14, 17. Group B: 10, 11, 13, 10. Find the difference in means (A – B).

2. In a randomization test, what does each shuffle of the pooled responses represent?

3. In 1,000 shuffles, 37 gave a difference at least as large as the observed one. What is the approximate one-sided P-value?

4. A randomization test gives P ≈ 0.42. What should you conclude?

5. With the same group sizes and the same spread of responses, how does a larger observed difference between treatment means affect the randomization P-value?

6. Write a correct sentence interpreting a one-sided P-value of 0.013 from the music experiment.

APPLICATION

7. (Illustrative data) Sixteen paper helicopters were randomly assigned to long or short blades, and flight times (seconds) recorded. Long: 2.41, 2.58, 2.36, 2.62, 2.49, 2.55, 2.44, 2.60. Short: 2.30, 2.47, 2.28, 2.39, 2.52, 2.33, 2.41, 2.35. Find the difference in means and interpret the exact one-sided randomization P-value: 127 of the 12,870 possible splits give a difference at least that large.

8. State the scope of the helicopter conclusion.

9. A student compares 15 volunteers who chose to study with music with 15 who chose silence, runs a randomization test and gets P = 0.02. What is wrong with concluding that music caused the difference?

CHALLENGE

10. Treatment: 5, 8, 9. Control: 2, 4, 6. List all C(6, 3) = 20 ways to choose 3 values for 'treatment', compute the difference in means for each, and find the exact one-sided P-value.

REVIEW

11. (Unit 2) Why is a randomized experiment the only way to show causation directly, when a correlation alone cannot?

12. (Unit 1 skill) A group of scores has mean 13.2 and SD 1.93 words. Find the z-score of a score of 16.

Unit 3 checkpoint

1. A statistic is…
   A. A number describing a population
   B. A number computed from a sample
   C. Always equal to the parameter
   D. A kind of bias
2. A school randomly picks 5 classrooms and surveys every student in them. This is a…
   A. Stratified sample
   B. Cluster sample
   C. Systematic sample
   D. Convenience sample
3. True or false: A voluntary-response sample of 10,000 is more trustworthy than a random sample of 500.
4. If samples of 100 give a margin of error of 0.092, samples of 400 give about _____.
5. True or false: A margin of error includes errors from nonresponse and question wording.
6. Which study can show cause and effect?
   A. A survey
   B. An observational study
   C. A randomized experiment
   D. A voluntary poll
7. Random selection allows you to…
   A. Conclude causation
   B. Generalize to the population
   C. Avoid all bias
   D. Skip the analysis
8. An inactive treatment given to a control group is a _____.
9. A randomization test gives P = 0.013. Which is correct?
   A. There is a 1.3% chance the null is true
   B. If there were no effect, a difference this large would occur about 1.3% of the time
   C. The effect is 1.3%
   D. The null is proven false
10. True or false: A P-value of 0.40 proves the treatment has no effect.
11. Students choose whether to join a program, and joiners score higher. The main problem is…
    A. Too few students
    B. Self-selection (confounding)
    C. The P-value
    D. Rounding
12. To halve a margin of error, multiply the sample size by _____.