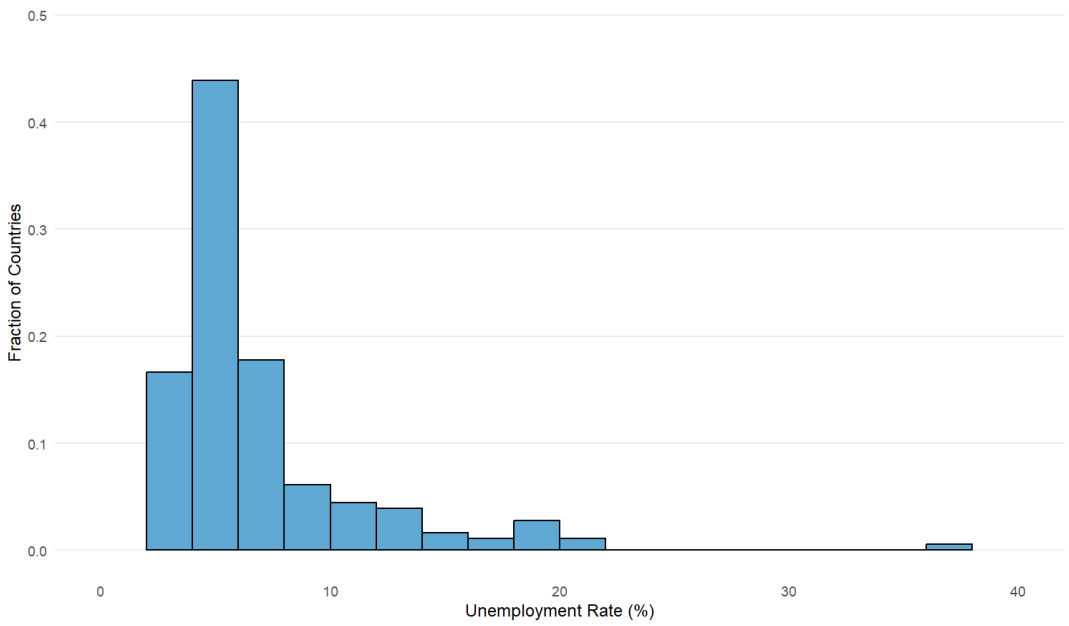
Statistics with Recitation — Quiz 1

September 23, 2025

Answer Key

1. (**6 points**) The unemployment rate is defined as the percentage of the labor force that is unemployed and actively seeking work. This rate is often used as an indicator of economic conditions in a country. The relative frequency histogram below shows the distribution of unemployment rates across 180 countries in 2022.



   (a) (**3 points**) Estimate Q1, the median, and Q3 from the histogram.

   (b) (**3 points**) Would you expect the mean of this data set to be smaller or larger than the median? Explain your reasoning.

Suggested Answers for Problem 1:

(a) Q1 $\approx 5(\%)$, Q2 $\approx 5(\%)$, Q3 $\approx 7(\%)$.

Grading Criterion

- (3 pts): Correct estimates of Q1, median, and Q3
- (1–2 pts): Partial correct answer.
- (0 pts): Incorrect or missing

(b) The distribution is strongly right-skewed. In such cases, the mean is pulled toward the tail, so the mean is expected to be larger than the median.

Grading Criterion

- (3 pts): Correct direction (mean > median) with reasoning
- (1–2 pts): Correct but incomplete reasoning
- (0 pts): Incorrect

2. (**10 points**) A metropolitan transit agency wants to survey riders in a large city to understand satisfaction and priorities. The bus and rail network covers 24 districts with different demographics and route densities. For each part below, identify the sampling method described, and briefly discuss one statistical pro and one con in this city's context.

   (a) (**2 points**) Sample 400 riders selected at random from the list of all monthly pass holders.

   (b) (**2 points**) Split the city into the 24 districts and randomly sample 25 riders *from each district*.

   (c) (**2 points**) Randomly select 4 districts and survey *every rider* boarding at stops in those districts for one week.

   (d) (**2 points**) Randomly select 6 districts, then within each selected district randomly sample 8 routes, then on each selected route randomly select 10 riders to survey.

   (e) (**2 points**) Station surveyors at the 10 busiest downtown stops during rush hour and survey the first 40 riders who agree.

Suggested Answers for Problem 2:

(a) **Method**: Random sampling
Pro: unbiased if the pass holders cover (or are random in) the target population.
Con: under-covers occasional/cash riders; responses may not represent all riders.

Grading Criterion

- (1 pts): Correct sampling method.
- (1 pts): Reasonable pro and con.

(b) **Method**: Stratified random sampling
Pro: ensures representation across all districts; increase precision if districts differ.
Con: equal allocation may be inefficient if district rider counts vary greatly.

Grading Criterion

- (1 pts): Correct sampling method.
- (1 pts): Reasonable pro and con.

(c) **Method**: Cluster sampling
Pro: logistically efficient, since survey teams focus on a few areas.
Con: higher variance if districts vary greatly; estimates depend heavily on which districts are picked.

Grading Criterion

- (1 pts): Correct sampling method.
- (1 pts): Reasonable pro and con.

(d) **Method**: Multistage sampling
Pro: practical and cost-effective while spreading across multiple districts/routes.
Con: needs correct weights for unbiased estimates.

Grading Criterion

- (1 pts): Correct sampling method.
- (1 pts): Reasonable pro and con.

(e) **Method**: Convenience sampling
Pro: very easy and cheap to implement.
Con: severely biased toward downtown, rush-hour, and willing respondents.

Grading Criterion

- (1 pts): Correct sampling method.
- (1 pts): Reasonable pro and con.

3. (**14 points**) In an educational psychology study on the effects of study environment on test performance, high school students were randomly assigned to three groups: (1) study in a quiet library, (2) study in a classroom with background music, or (3) study at home with no restrictions. After one week of preparation under these conditions, all students completed the same standardized math test. The average scores of the groups were compared, and the library group scored the highest.

   (a) (**2 points**) What type of study is this?

   (b) (**3 points**) Identify the explanatory variable(s). What is the type of these variables?

   (c) (**3 points**) Identify the response variable.

   (d) (**3 points**) Comment on whether the results of this study can be generalized to all high school students, and why.

   (e) (**3 points**) Comment on whether the results of the study can be used to establish causal relationships, and why.

Suggested Answers for Problem 3:

(a) This is a randomized controlled experiment. Students were randomly assigned to study environments before the outcome was measured.

Grading Criterion

- (3 pts): Correct
- (0 pts): Incorrect answer

(b) The explanatory variable is study environment with three levels (library, music classroom, home). Type: categorical (nominal).

Grading Criterion

- (4 pts): Correctly identifies both the variable and its type
- (2 pts): Correctly identifies either the variable or its type
- (0 pts): Incorrect answer

(c) The response variable is the math test score.

Grading Criterion

- (3 pts): Correct
- (0 pts): Incorrect answer

(d) No. Generalization requires a random sample from the population. The context does not state that students were randomly sampled (they may be a convenience group), so the results may not represent all high school students.

Grading Criterion

- (4 pts): Correct answer with correct reasoning
- (2 pts): Correct answer, but incomplete/incorrect reasoning
- (0 pts): Incorrect answer

(e) Yes. Because the study used random assignment to treatment groups, differences in scores can be attributed to the study environment-supporting causation.

Grading Criterion

- (4 pts): Correct answer with correct reasoning
- (2 pts): Correct answer, but incomplete/incorrect reasoning
- (0 pts): Incorrect answer