Practice Exam #1 (Sta-209, F25)

Ryan Miller

Directions

- Answer each question using *no more than specified number of sentences* and not attempt to avoid these guidelines by using run-on sentences. Answers that are unnecessarily verbose may result in point loss.
- Do not include superfluous information in your answers, you may be penalized if you make an inaccurate statement even if you go on to provide a correct answer. Your answers should be clear, concise, and include only what is needed to answer the question that was asked.

Formula Sheet

Below are some formulas that have appeared in our lecture slides:

Mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Pearson's Correlation Coefficient:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$$

Simple linear regression (theoretical model):

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Simple linear regression (fitted model):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Coefficient of Determination (R^2):

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Standard Errors:

Statistic	Standard Error	Conditions
$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$	$np \geq 10$ and $n(1-p) \geq 10$
$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$	normal population or $n \geq 30$
$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$n_i p_i \geq 10$ and $n_i(1-p_i) \geq 10$ for $i \in \{1,2\}$
$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	normal populations or $n_1 \geq 30$ and $n_2 \geq 30$

Section 1 - True/False

Directions: Clearly indicate whether each of the following statements is true or false. You do not need to explain your reasoning and you will not receive a better score for doing so.

1. A *large p-value*, such as $p = 0.90$, should be interpreted as strong evidence that the null hypothesis is *most likely true*.

- FALSE - a large p-value only indicates that the sample data would not be unexpected if the null hypothesis were.

2. A *small p-value*, such as $p = 0.01$, should be interpreted as strong evidence that the null hypothesis is *most likely false*.

- TRUE - a small p-value indicates the sample data are unlikely to occur if the null hypothesis were true, which is taken as evidence against the null hypothesis.

3. In hypothesis testing, the null hypothesis is set up to reflect what the researchers suspect is true about the population they are studying.

- FALSE - the null hypothesis is a "stawman" that the researchers want to disprove.

4. The T-test is only appropriate when the sample size is relatively large.

- FALSE - the T-test was created to work on small, Normally distributed samples

5. Suppose we suspect a two-sided coin is biased to land on heads more often than it should. If we flip the coin 15 times and observe 12 heads, the one-sided p-value corresponding to this scenario can be expressed via $Pr(\hat{p} \geq 12/15 | p = 0.5)$.

- TRUE - the p-value is a conditional probability of an outcome at least as extreme as the one that was observed given the null hypothesis is true.

6. The Z-test is only appropriate for categorical data when we've observed a relatively large number of cases in each category involved in the test.

- TRUE - for categorical data the Z-test uses a Normal distribution to approximate a distribution of proportions (which are discrete), so it only works when there's enough data for this approximation to be reasonable.

7. The Z-test uses a Normal distribution as the probability model used to compute the p-value

- TRUE - this is how the Z-test is defined

8. A *very small p-value*, such as $p = 0.0001$ indicates a very important scientific discovery.

- FALSE - p-value measures evidence against the null hypothesis, it doesn't tell you anything about scientific importance

9. Consider $H_0 : p = 0.5$ and $H_a : p \neq 0.5$ and suppose we collect a sample of size $n = 50$ and *fail to reject the null hypothesis* after finding a sample proportion of $\hat{p} = 0.6$. If we collect a larger sample, such as $n = 100$, and observe the same sample proportion, $\hat{p} = 0.6$, it is possible that we have found enough evidence in the new sample to reject the null hypothesis.

- TRUE - in the Z-test the SE will decrease when the sample size is larger, thereby giving us a larger test statistic (Z-value) for the same sample proportion

10. Consider $H_0 : p = 0.5$ and $H_a : p \neq 0.5$ and suppose we collect a sample of size $n = 50$ and *fail to reject the null hypothesis* $H_0 : p = 0.5$ after finding a sample proportion of $\hat{p} = 0.6$. If we collect another sample of size $n = 50$ and observe $\hat{p} = 0.58$ it is possible that we have found enough evidence in the new sample to reject the null hypothesis.

- FALSE - if $\hat{p} = 0.6$ wasn't enough evidence to reject H0 neither will $\hat{p} = 0.58$ if the sample size (and therefore the SE) remains the same.

11. In hypothesis testing, the null hypothesis is a claim we make about the sample data.

- FALSE - H0 is a claim made about the population, not the sample.

Section 2 - Conceptual Questions

Directions: Answer each question using no more than 3-sentences. Do not include unnecessary details, as you will be penalized for any inaccurate statements, regardless of whether they are relevant or not. Aim to clearly answer the question that was posed, not to demonstrate your knowledge of related topics.

1. One of your friends overheard that you know statistics, and they approach you for advice on whether they should use *Pearson's correlation* or *Spearman's correlation* to analyze data they've collected for another class. You sit down with this friend, open up `R` Studio, and load their data. Briefly explain what you'd do next to determine which type of correlation coefficient they should use.

- They should create a scatter plot of the data to determine whether it exhibits a linear or non-linear trend. They might add a smoothing line (loess) and a linear regression line to the plot to help facilitate this comparison. If the data show a linear trend they should use Pearson's correlation, and if it shows a non-linear trend they should use Spearman's.

2. Briefly explain *why* seeing a p-value meeting an established threshold for statistical significance is reason to reject the null hypothesis.

- A small p-value indicates a small probability of seeing an outcome like the one that was observed in the sample if the null hypothesis were true. So, either the sample was a really rare occurrence, or the assumption of this probability calculation was wrong (ie: H_0 wasn't actually true). We set up a statistical significance threshold for how low the probability needs to be for us to believe the latter.

3. In your own words, explain the concept of a *null distribution* in hypothesis testing. That is, what is the null distribution and how does it relate to the p-value?

- The null distribution shows the distribution of outcomes that could have been observed in a study if the null hypothesis were true. That is, it shows us possible outcomes and how frequently they might occur. The null distribution gives us something to compare the actual observed outcome in the study with to calculate the p-value because the p-value is just the probability of a certain outcome assuming the null hypothesis is true.

4. For each of the following scenarios provide the name of an appropriate data visualization (graph). Be specific.

- A: One-sample categorical data
  - Bar chart
- B: Two-sample categorical data
  - Stacked, conditional or clustered bar chart
- C: One-sample quantitative data
  - Histogram or box plot
- D: Two-sample quantitative data
  - Side-by-side box plots or histograms
- E: Two quantitative variables
  - Scatter plot

Section 3 - Application

In a 1997 research paper, psychologists Ruback and Juieng performed a series of studies investigating how quickly a person leaves their parking spot.

In one of these studies, Ruback and Juieng observed 200 drivers departing from a public parking lot. For each departing driver they recorded the time (in seconds) from when the driver first entered their vehicle to when they had fully exited their parking space. The main explanatory variable in the study was whether or not another car was waiting for the driver's parking space while they were exiting.

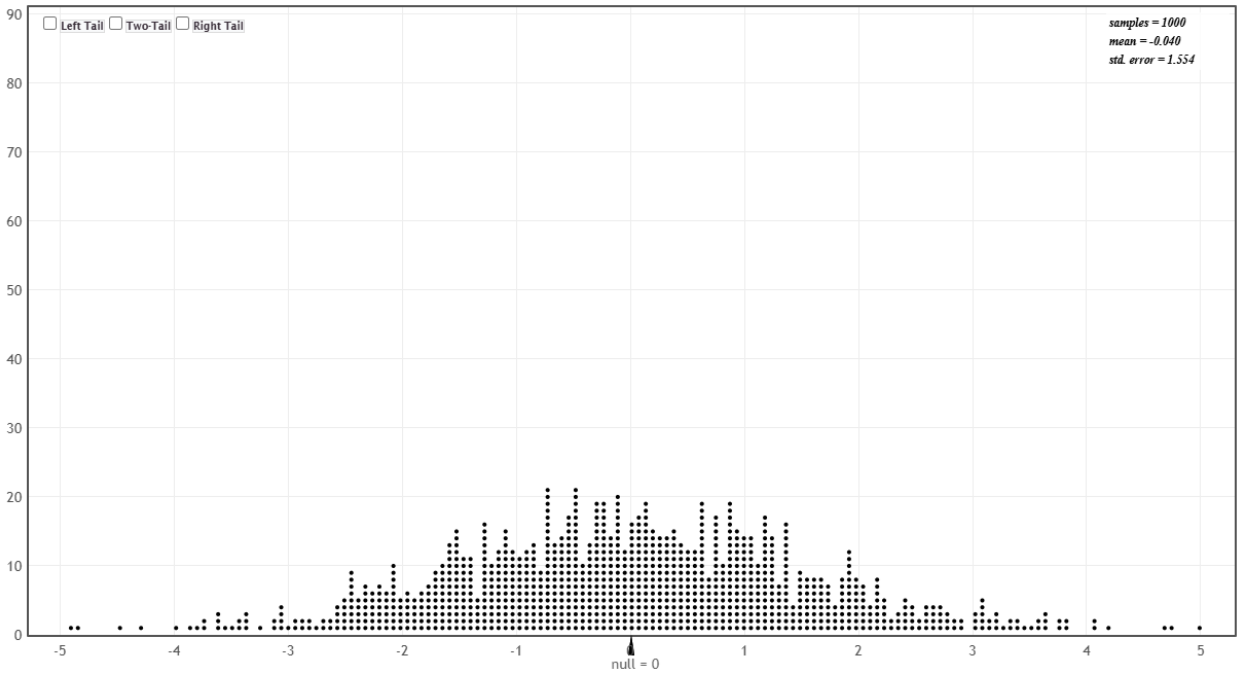
A summary of their data is given below:

Waiting	Mean Time	Median Time	Std Dev	IQR	N
no	45.55	46.13	10.53	13.93	132
yes	49.42	48.11	9.42	9.57	68
total	46.87	46.68	10.31	12.30	200

1. Suppose the researchers would like to evaluate whether the presence of another driver waiting influences the speed at which a person exists a public parking lot. State the null and alternative hypotheses these researchers should consider using the proper statistical notation.

- $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$ where μ_1 and μ_2 are the respective population means of drivers when someone is waiting and not waiting.

2. Shown below are 1000 outcomes that were simulated using the null hypothesis you proposed in Part 1. Use this distribution to estimate the two-sided p-value. You may choose to annotate the figure to help me understand how you are estimating the p-value.



- There are approximately 11 simulated outcomes (out of 1000 simulations) at least as extreme as the difference in means observed in the sample. So the p-value is estimated as 0.011

3. Express the p-value in this application as a formal probability statement, such as $Pr(\ldots)$.

- $Pr(\overline{x}_1 - \overline{x}_2 \geq 3.87 | \mu_1 = \mu_2)$ is acceptable (note that its a one-sided p-value). You could multiple by two to make it two-sided, or you could add an "or" statement in the event.

4. Provide a one-sentence conclusion based upon the p-value and hypotheses you've provided in earlier parts of this question. Your conclusion should include all of the components of a proper conclusion discussed in our labs and examples.

- The study provides strong evidence (p = 0.011) that drivers leave their parking space slower when another driver is waiting (mean = 49.42 sec) than when no one is waiting (mean = 45.55 sec).

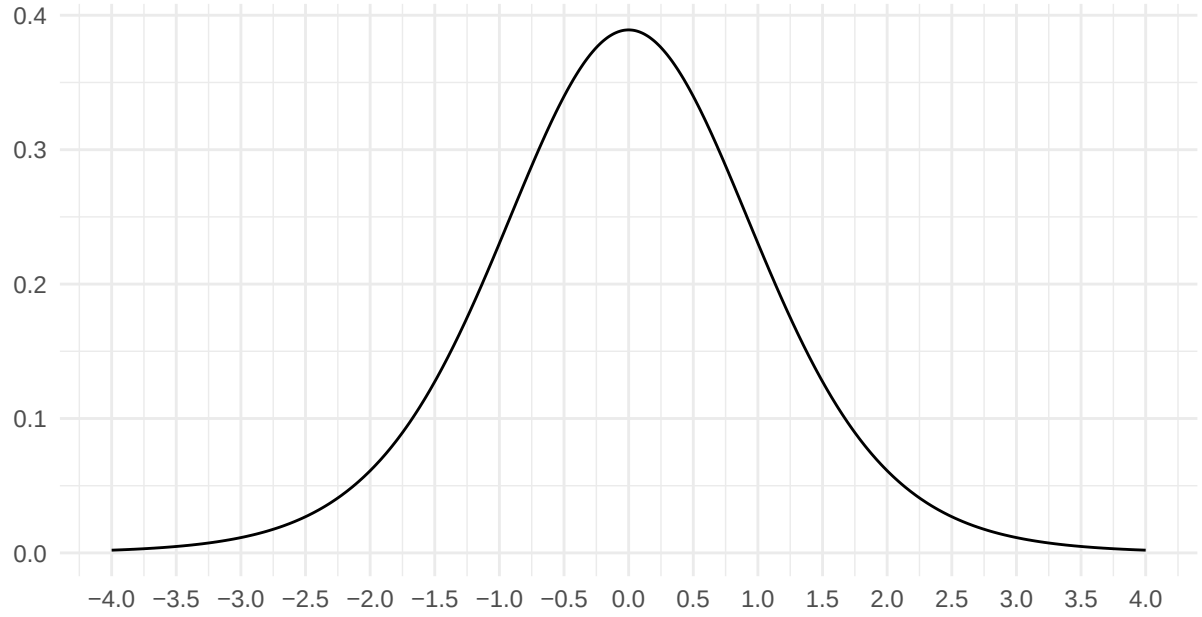
5. Suppose you'd like to use a statistical test based upon a probability model rather than simulation to evaluate the hypotheses you provided earlier. What is the name of the statistical test you should use? Provide the name of the test as well as a brief explanation of why it is an appropriate choice.

- T-test (two-sample T-test more specifically). The test is appropriate because we have a quantitative outcome compared across two samples that are both relatively large (n_1 and n_2 are both larger than 30).

6. The test statistic in this scenario uses the standard error formula $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. Calculate this test statistic. Show your work.

- $T = \frac{\bar{x}_1 - \bar{x}_2 - 0}{SE} = \frac{3.87}{\sqrt{10.53^2/132 + 9.42^2/68}} = 2.647$

7. Shown below is the probability model for the test statistic you calculated in the previous part. Shade the area of this distribution that represents the *two-sided p-value* in this application.



- The areas outside of $T = -2.647$ and $T = 2.647$ should be shaded.

8. The researchers in this study balanced their sample by following an equal number of male and female drivers into the parking lot, but they could not control whether or not a vehicle was waiting for each of these drivers. For the male drivers, they observed 38 instances of a vehicle waiting, and for the female drivers they observed 30 instances of a vehicle waiting. Consider the null hypothesis that male and female drivers were equally likely to have a vehicle waiting. Fill in the missing components (question marks) in the `R` code given below to test this hypothesis:

```
prop.test(x = ?, n= ?, alternative = "two.sided")
```

- `x = c(38, 30)`, `n = c(100, 100)`

9. The p-value produced by the `prop.test()` function from the previous part is $p = 0.296$. Provide an appropriate one-sentence conclusion based upon this p-value.

- The study provides insufficient evidence ($p = 0.296$) that male drivers encounter a vehicle waiting for their parking spot at a different rate than female drivers.