Exam 2: 105 points possible

YOUR NAME HERE

2025-10-12

This is a take home exam for BIOL 305. It will be graded out of 100 points, such that the maximum possible grade is 105%. Please read each question thoroughly and be sure to address all parts of the question posed. A slight departure from the homeworks, but each question *must* be answered with a complete sentence communicating the answer, even if this sentence is brief.

I want you to do these exams on your own. This is an honor system for the exams, but please work on your own. This is an assessment of *your* understanding of the material. The exam is open resource, just please tell me what resources you used if you rely on one heavily.

General point penalties: You will lose 1 pt for each:

- incorrectly rounded value
- incorrectly rounded test statistic (i.e., Z)
- incorrectly reported p-value
- incomplete sentences / answers that are only a code output (even if that output is correct - you *must* contextualize and communicate your answer)
- **Please use spellcheck**; I will not take off for minor mistakes, but if there are lot of mistakes I will take off points. *RStudio has a built in spellcheck just like Word!*

Communicating results:

For questions that want you to communicate all the results, please ensure that you are reporting your *test statistics* and your *p-values* as well. For example:

| Kearney minnows are significantly smaller than minnows in other populations ($Z = 6.37$, $p < 0.0001$).

Libraries go here

```r
# tidyverse always required
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
# need anything else...?
```

1 [16 pts] Comparing yourself to others

As a kid, your parents often tell you not to compare yourself to others. However, you now have the statistical tools to compare yourself to others scientifically, and you decide to compare your exam scores to your classmates. You are convinced that you are doing exceptionally well in the class, as you got an 80.0% on the last exam and the average was a 70.0%. You find out from your professor that the standard deviation for the class was 8.0%.

A: [4 pts] Choose an α value for evaluating your test, and choose whether you want to use a one-tailed or a two-tailed test to evaluate your results. Justify your choice.

B: [4 pts] Calculate a Z-score for your exam score. What does a Z-score represent?

C: [2 pts] Calculate a p value for your Z-score.

D: [4 pts] Write out a conclusion about your test. It should contain (1) a full sentence stating your conclusion and (2) your Z-score and your p-value in parentheses. See instructions for an example.

E: [2] What is the grade that corresponds to your α value? I.e., translate your α value back into an exam score.

2 [56 pts] Fish watchin'

You go to the Bonneville Dam and discover they have windows for watching the fish swim through the fish ladder. Much to the chagrin of your family, you decide to do 15 minute counts on the number of salmon passing through the fish ladder for 3 hours, obtaining the following data:

```r
fish_counts <- c(16,18,25,24,
                 18,11,6,17,
                 21,16,32,26)
times <- c("9:00", "9:15", "9:30", "9:45",
           "10:00", "10:15", "10:30", "10:45",
           "11:00", "11:15", "11:30", "11:45")

fish_data <- cbind(times, fish_counts) |>
  as.data.frame() |>
  mutate(fish_counts = as.numeric(fish_counts))

fish_data
```

```
##    times fish_counts
## 1   9:00          16
## 2   9:15          18
## 3   9:30          25
## 4   9:45          24
## 5  10:00          18
## 6  10:15          11
## 7  10:30           6
## 8  10:45          17
## 9  11:00          21
## 10 11:15          16
## 11 11:30          32
## 12 11:45          26
```

A: [2 pts] What did the command "mutate" do in the above code block?

B: [5 pts] What is the explanatory variable in the above dataset? [2 pts] What type of data are these? [3 pts] (Nominal, ordinal, integer, or ratio? Also, categorical or continuous?)

C: [5 pts] What is the response variable in the above dataset? [2 pts] What type of data are these? [3 pts]

D: [5 pts] Plot a histogram of the response data. [3 pts] Use the arguments `main =` and `xlab =` to make the histogram prettier and to have it make more sense; you will have to designate names in "quotes". [2 pts]

E: [5 pts] Plot a cumulative frequency plot of these data. [3 pts] Use the arguments `main =` and `xlab =` to make the graph prettier and to have it make more sense; you will have to designate names in "quotes". [2 pt]

F: [2 pts] What is the kurtosis of the response variable?

G: [2 pts] What is the skew of the response variable?

H: [4 pts] Are there any outliers in these data? Create the appropriate graph [3 pts] and perform the appropriate command to find out [1 pt].

I: [2 pts] Now perform and interpret a Shapiro-Wilkes test on the response variable.

J: [4 pts] Are the data normally distributed? Explain, referencing the *three* pieces of evidence you've obtained above.

K: [6 pts] Your annoying friend who is always trying to one-up you says they saw 35 fish in 15 minutes. You think they are lying, but want to show scientifically why you doubt their claim. What test would you use to evaluate this claim? Describe the kind of test [2 pts], the α value you would use and why [2 pts], and whether this would be a one-tailed or a two-tailed test [2 pts]. (Note: the salmon are one-tailed, but this does not affect the statistical test).

L: [4 pts] What are the *null* and *alternative* hypotheses for part K?

M: [5 pts] Perform the test you outlined in parts K & L; report the test statistic (i.e., Z) and the P value.

N: [5 pts] What is the statistical conclusion of this test? State both a statistical conclusion, with a Z value and a P value, as well as a statement relating back to your original hypothesis in part K.

3 [11 pts] A statistical disagreement

You and your friends are discussing the number of cookies you eat on average every day. Between the eight of you, you find that you eat 3 cookies on average, with a standard deviation of 2 cookies.

You then find out that Dr. Cooper eats 6 cookies on average every day (if his wife is out of town). You decide to see how unusual this is by using a one-tailed test to see if the number of cookies Dr. Cooper consumes is higher than your group's average.

A: [2 pts] What are the null and alternative hypotheses in this situation?

B: [9 pts] You run the test with $\alpha = 0.05$. One of your friends says no, Dr. Cooper's behavior is not unusual with $p = 0.07$, whereas another friend says Dr. Cooper eats way too many cookies with $p < 0.0001$. Perform the test yourself, and determine why your friends got different answers. (Full points for performing the test [3], correctly interpreting the test [2], and correctly identifying why the two answers disagree [4]).

4 [22 pts] Belltower traffic

You are counting the number of students passing the belltower every passing period, obtaining the following dataset:

```
people_passing <- c(2,1,6,13,30,22,15,10,5,11,1,14,
                    23,6,25,134,4,8,1,13,8,5,6,12,6)

belltower_data <- people_passing |>
  as.data.frame() |>
  mutate(time = 1:25) |>
  # this command re-orders the columns
  select(time, people_passing)

belltower_data
```

```
##    time people_passing
## 1     1              2
## 2     2              1
## 3     3              6
## 4     4             13
## 5     5             30
## 6     6             22
## 7     7             15
## 8     8             10
## 9     9              5
## 10   10             11
## 11   11              1
## 12   12             14
## 13   13             23
## 14   14              6
## 15   15             25
## 16   16            134
## 17   17              4
## 18   18              8
## 19   19              1
## 20   20             13
## 21   21              8
## 22   22              5
## 23   23              6
## 24   24             12
## 25   25              6
```

A: [2 pts] Which variable is the response variable? What type of data are these?

B: [7 pts] Is the response variable normally distributed? [2 pts] Support your answer with a histogram [1 pt], a cumulative frequency plot [1 pt], assessments of skewness and kurtosis [2 pts], an indication of if outliers exist [1 pt]. You will lose 1 pt overall if you do *not* change the histogram and plot titles and axes labels as per the previous problem(s).

C: [7 pts] If a transformation is needed, what transformation is required? Note that as soon as you find a transformation that works, you do not need to do any others. Confirm your transformation works by using at least *three* methods to check the distribution (i.e., histogram, qqplot, cumulative frequency plot, Shapiro-Wilkes test, etc.)

NOTE: Test only the following transformations:

- $\sqrt{x}$
- x^2
- x^3
- $log(x+1)$ (can be done with function `log1p` or manually as `log(x+1)`)

D: [6 pts] What is the 95% confidence interval for the mean of these data?

HINT: You may have to *back-transform* your answers using one of the following commands:

- $\sqrt{x}$: x^2
- x^2: $\sqrt{x}$
- x^3: $x^{\frac{1}{3}}$
- $log(x+1)$: $exp(1)^x - 1$ (can be done with function `expm1(x)`)