

Exam 2 Practice Key (code)

BIOL 305

2025-10-12

```
## -- 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
```

Please complete the following problems as practice for Exam 2. *Note* that you may need to load specific packages from the previous exercises to get your code to work.

All questions must be answered in full sentences.

Example answers: For your conclusion, answers should look like this:

The mean height of the class is significantly higher than the height of the general population ($Z = 2.50$, $p = 0.006$).

The mean length of the leg of the high-elevation Woodhouse's Toads (*Anaxyrus woodhousii*) is indistinguishable from those of the general population ($Z = 0.12$, $p = 0.45$).

Minnows

Consider this scenario: You have discovered a never-before-documented population of minnow in the Kearney Canal near campus. During your first sampling trip, you notice that the total length (i.e., measured from the tip of the snout to the very tip of the tail) of the fish you measure appear to be smaller than the average total length of the species as recorded among all known individuals across their range. The mean total length noted in one publication is 85.00 mm with a standard deviation of 4.50. Below are your data from 20 minnows that you captured during your first sampling trip to the Kearney Canal:

Table 1: Fish data for this problem.

Fish ID	Length (mm)
1	89.58
2	75.44
3	86.86
4	74.71
5	69.70
6	100.34
7	73.70
8	69.56
9	96.24

Fish ID	Length (mm)
10	79.35
11	61.37
12	62.82
13	95.45
14	98.71
15	100.34
16	57.57
17	70.54
18	78.65
19	65.39
20	65.57

NOTE: you will have to create the numeric object for this problem. use c to link things together; e.g.:

```
## [1] 1 5 8 9
```

1. State the null and alternative hypotheses for this study.

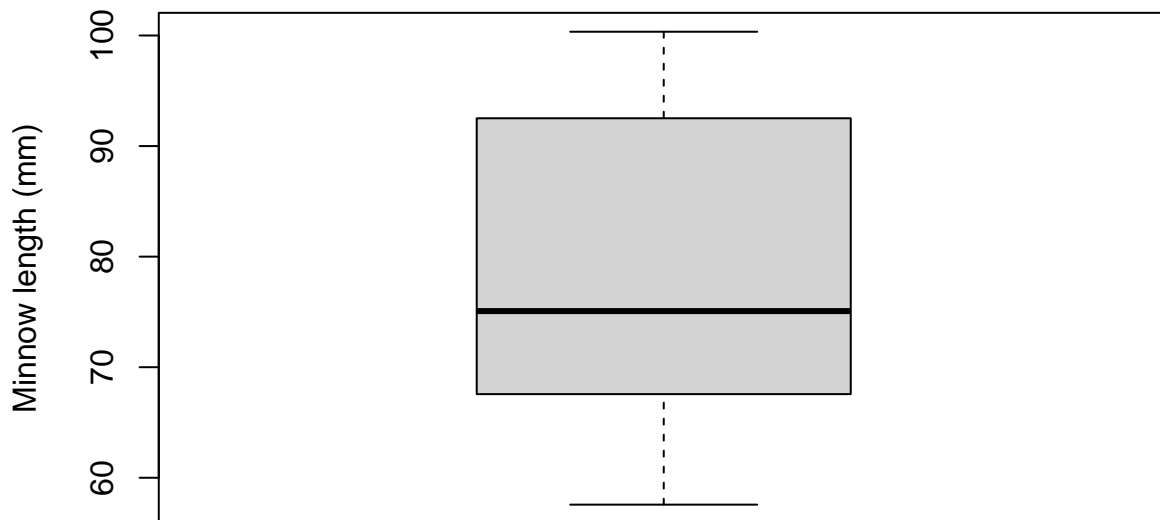
H_0 : The mean length of minnows in the Kearney Canal is not smaller than the mean length of other minnow populations.

H_A : The mean length of minnows in the Kearney Canal is less than the length of other minnow populations.

2. Calculate the first quartile, median, third quartile, and interquartile range of your response variable. Create a boxplot. Based on your results, are there any outliers in your data? Explain.

```
##      0%      25%      50%      75%     100%
##  57.57  66.57  75.07  93.98 100.34
```

```
## [1] 27.42
```



The median length of minnows is 75.07 mm, with a lower quartile of 66.57 mm and an upper quartile of 93.98 mm. The interquartile range is 27.42 mm. There are no outliers.

3. Calculate the z -score for this scenario.

```
## [1] -6.37
```

The Z score is -6.37.

4. What is the probability that, by random chance alone, you would find your observed mean or something more extreme?

```
## [1] 9.711376e-11
```

The p value is < 0.0001 .

5. Assume that you set your α for your study prior to your data collection to 0.05. Based on this information and the p value you obtained, is your null hypothesis supported or rejected? Report your conclusion in a full sentence, including the Z score and the p value in your answer.

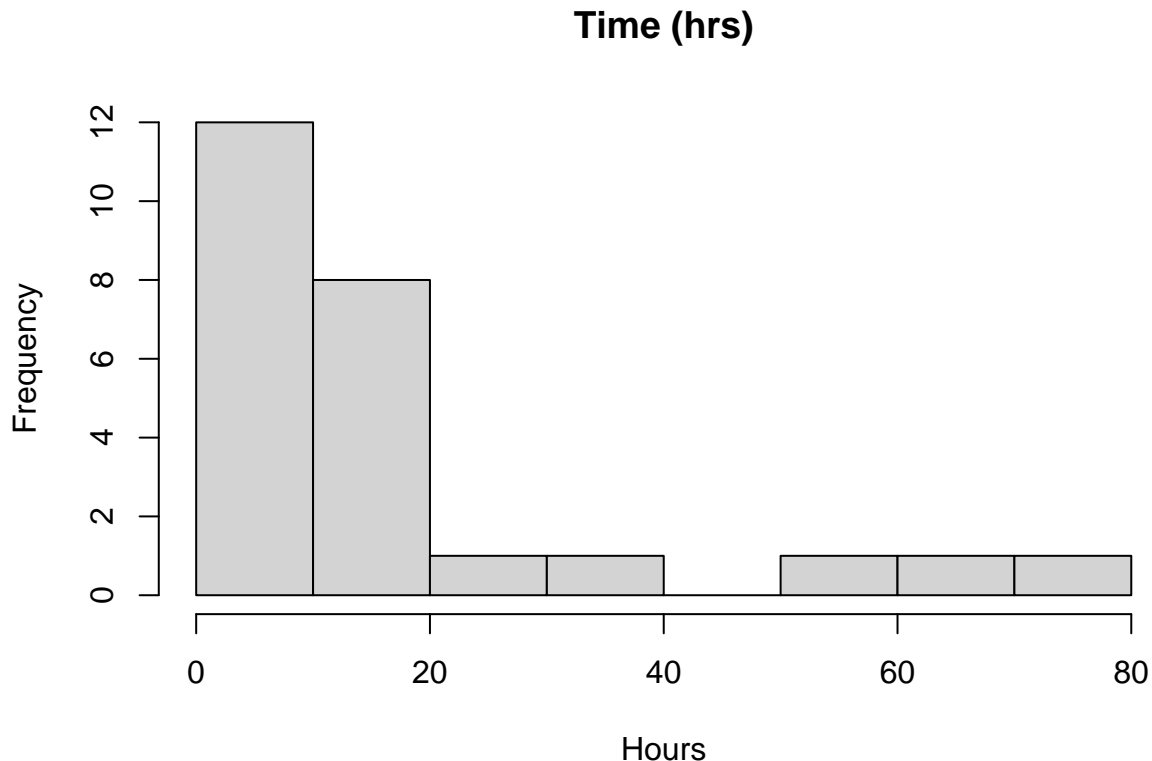
The mean length of minnows in the Kearney Canal is less than the mean length of published minnow populations ($Z = 6.37$, $p < 0.0001$).

Time spent on canvas

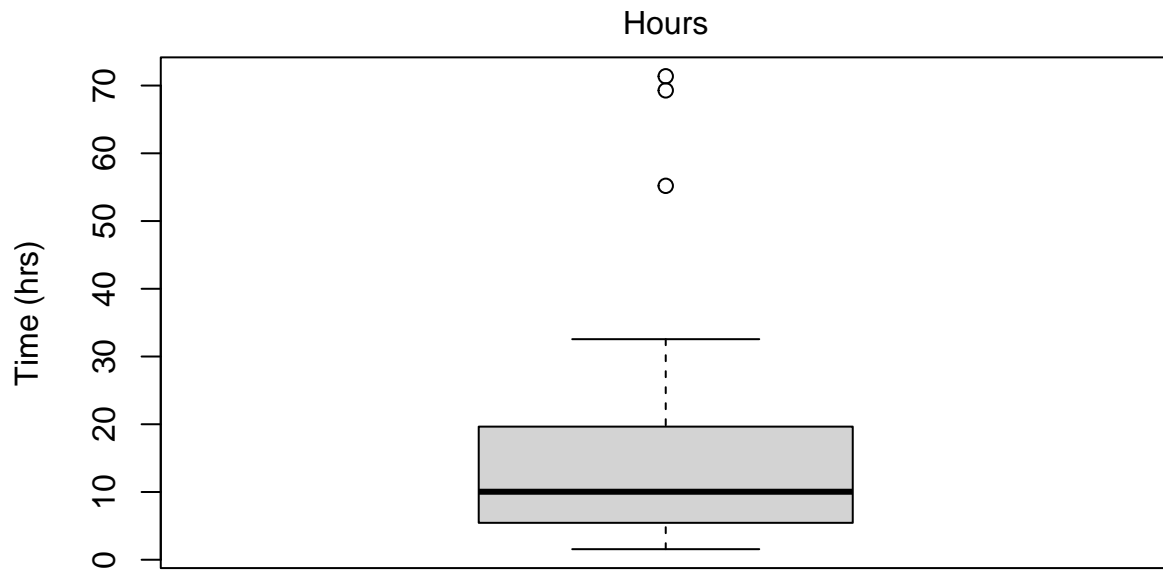
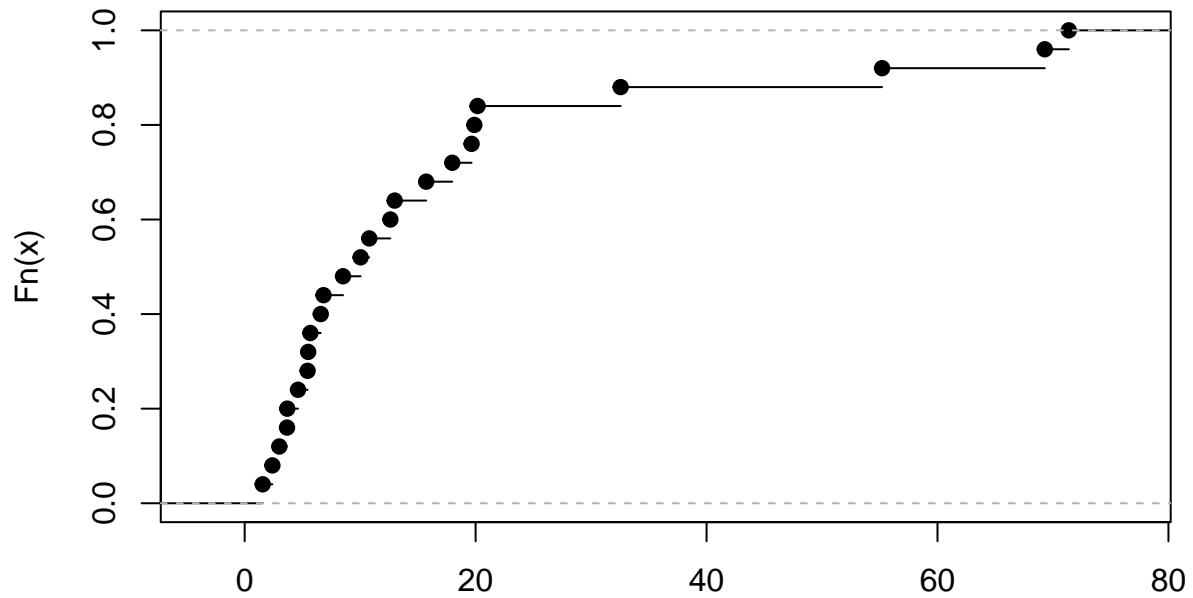
You are curious about how much time your classmates spend on Canvas for BIOL305. Let's say you get the following data for the amount of time folks spend on Canvas:

```
## student time.hrs
## 1      1      19.65
## 2      2      69.29
## 3      3       6.83
## 4      4       5.50
## 5      5      17.98
## 6      6      19.89
```

1. Which variable is the response variable? What type of data are these?
2. Is the response variable normally distributed? Support your answer with a histogram, a cumulative frequency plot, assessments of skewness and kurtosis, and an indication of if outliers exist. Change the title and axes labels of the plots to make them intelligible.



Cumulative summary (time in hours)



Shapiro-Wilkes test:

```
##
##  Shapiro-Wilk normality test
##
## data:  time.hrs
## W = 0.70202, p-value = 7.759e-06
```

Kurtosis:

```
## [1] 2.31
```

Skewness:

```
## [1] 1.86
```

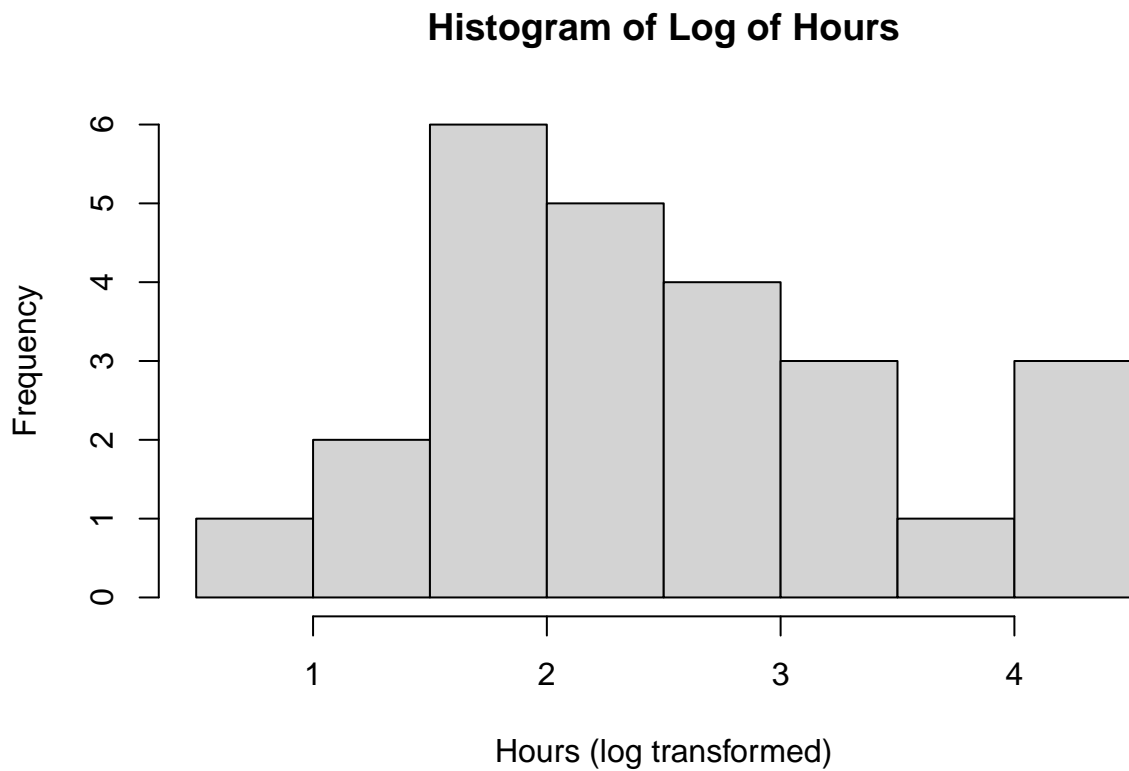
Outliers:

```
## [1] 69.29 71.37 55.20
```

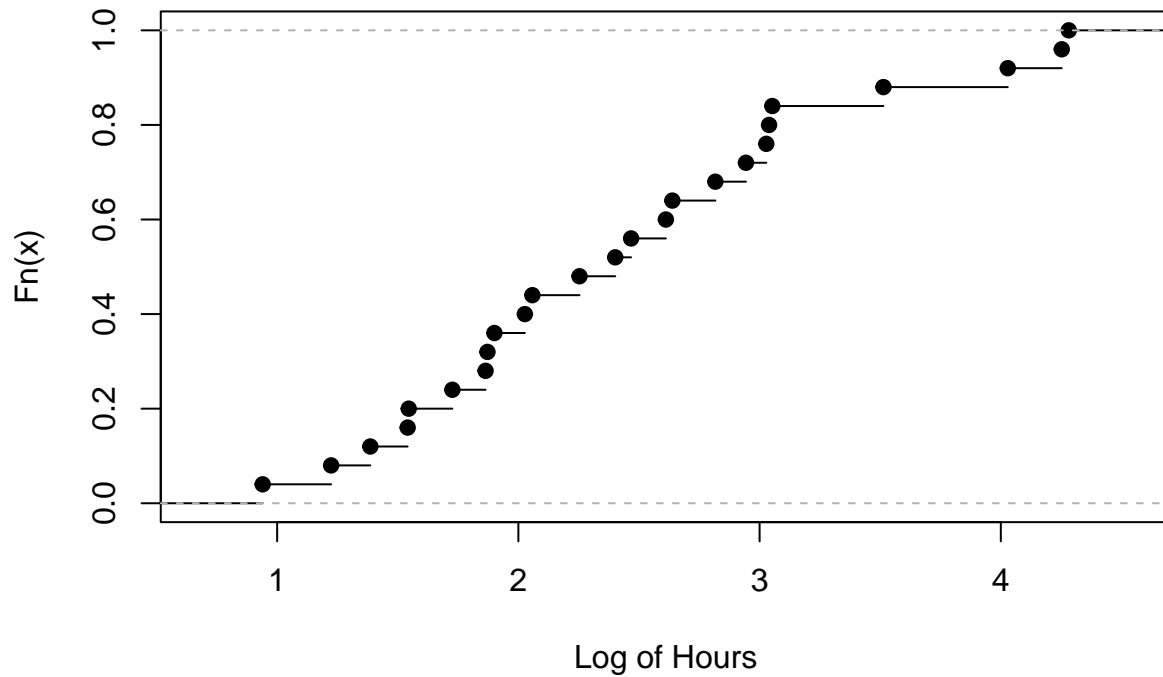
The data are not normal. (NOTE - you will have to cite THREE pieces of evidence above in *your* answer).

1. If a transformation is needed, which transformation is required? *Note* - you only have to do a histogram and a `shapiro.test` to assess a transformation. As soon as you find a transformation that works, you do not have to do another.

Transformation:



Cumulative Frequency of Time in Hours



```
##  
## Shapiro-Wilk normality test  
##  
## data: log.hrs  
## W = 0.9581, p-value = 0.3779
```

A log transformation successfully makes the data normal.

1. What is the 95% confidence interval for the mean of these data? Report as the mean \pm the confidence interval, and answer in a full sentence. Don't forget to mention the units of your answer. HINT: you may have to back-transform your data.

The average is 10.66 hours, with a confidence interval of 7.14 to 15.7 hours.