

Lab Assignment 6

Summer Heschong

2025-02-19

```
#Setup
```

```
#load packages  
library(here)
```

```
## here() starts at /Users/summerheschong/stats_spring25
```

```
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    3.5.1      v tibble    3.2.1  
## v lubridate  1.9.4      v tidyr     1.3.1  
## v purrr      1.0.2
```

```
## -- 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
```

```
library(moments)
```

```
#load data  
vertebrate_data <- read.csv(here('Data/Raw/and_vertrebrates.csv'))
```

(1) Coastal Giant Salamander

```
##a. Are salamander data normally distributed?
```

```
#filter dataset  
salamander_data <- vertebrate_data %>%  
  filter(species == 'Coastal giant salamander' &  
         (unittype == 'C' | unittype == 'P' | unittype == 'SC'))  
  
#evaluate normality with figure  
fig1 <- ggplot(salamander_data, aes(x = length_2_mm, fill = unittype)) +  
  geom_histogram(bins = 45, binwidth = 6, color = 'black') +
```

```

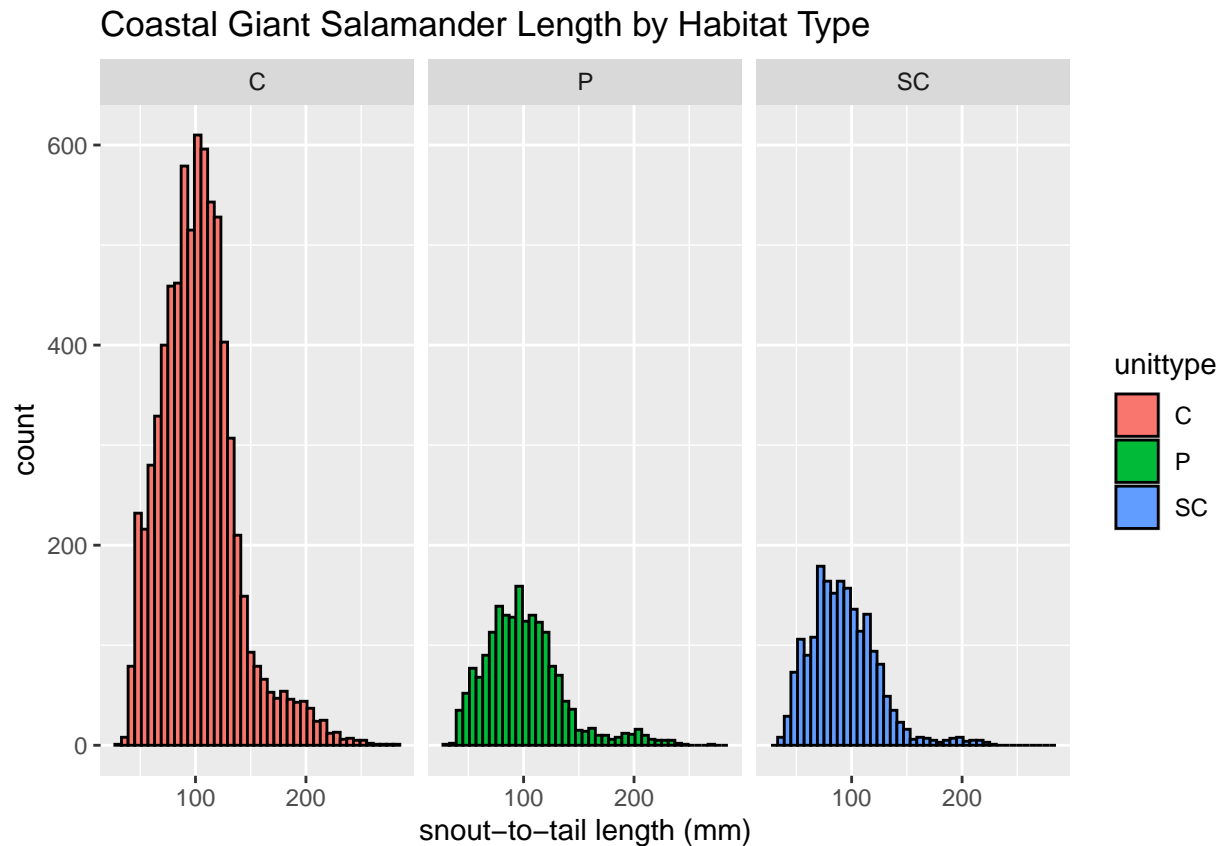
facet_wrap(~unittype) +
  labs(title = 'Coastal Giant Salamander Length by Habitat Type',
        x = 'snout-to-tail length (mm)')
fig1

```

```

## Warning: Removed 220 rows containing non-finite outside the scale range
## ('stat_bin()').

```



```

#calculate skewness
skewness(salamander_data$length_2_mm, na.rm = TRUE)

```

```
## [1] 0.9913912
```

```

#calculate kurtosis
kurtosis(salamander_data$length_2_mm, na.rm = TRUE)

```

```
## [1] 4.793014
```

```

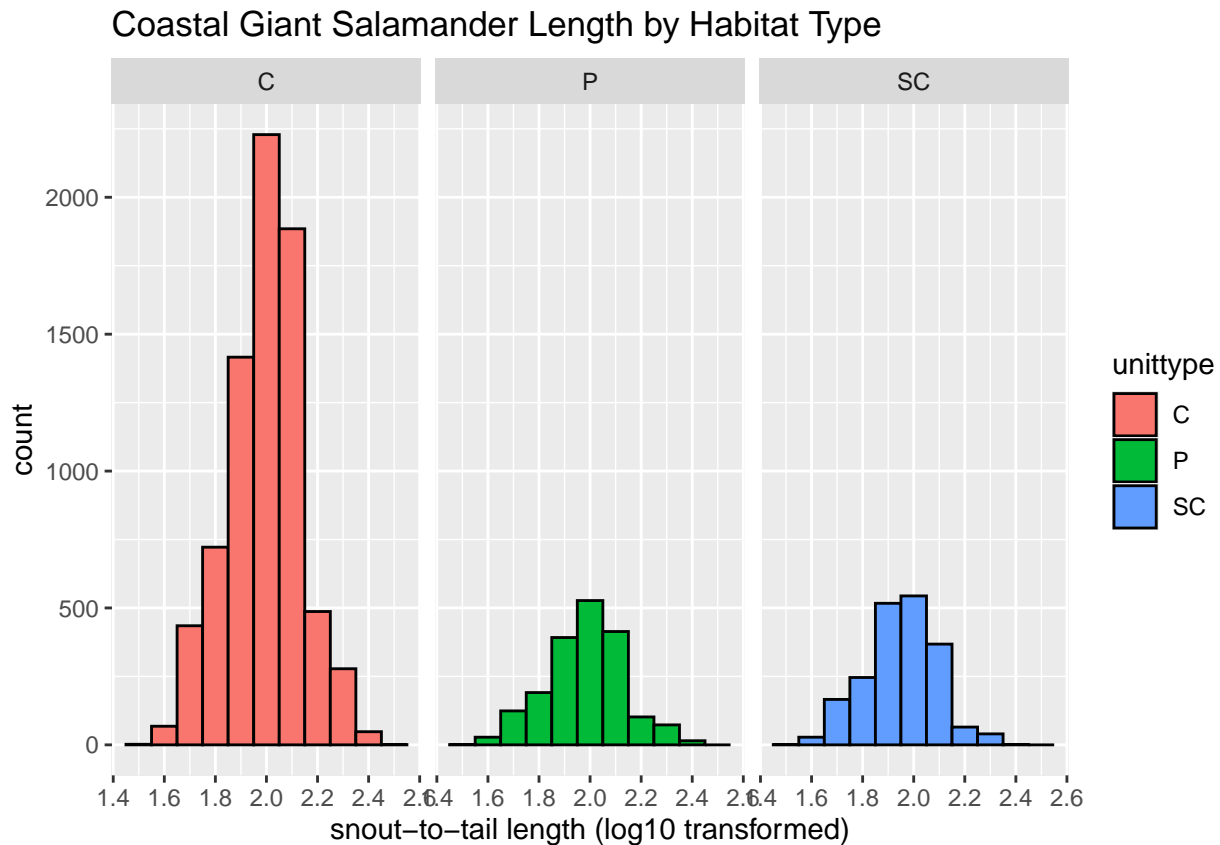
#not normal -> apply a log-transformation
salamander_data$length_2_mm <- log10(salamander_data$length_2_mm)

#re-evaluate normality

```

```
fig2 <- ggplot(salamander_data, aes(x = length_2_mm, fill = unittype)) +
  geom_histogram(bins = 45, binwidth = 0.1, color = 'black') +
  facet_wrap(~unittype) +
  labs(title = 'Coastal Giant Salamander Length by Habitat Type',
       x = 'snout-to-tail length (log10 transformed)')
fig2
```

```
## Warning: Removed 220 rows containing non-finite outside the scale range
## ('stat_bin()').
```



```
skewness(salamander_data$length_2_mm, na.rm = TRUE)
```

```
## [1] -0.125949
```

```
kurtosis(salamander_data$length_2_mm, na.rm = TRUE)
```

```
## [1] 3.05272
```

Answer: Based on fig1 the data looks like it is not normally distributed with a longer right tail. The skewness and kurtosis values confirm this. I applied a log-transformation and re-evaluated the data. In fig2 the data looks pretty normally distributed. The new kurtosis value (3) confirms this, and though the skewness value is not equal to 0 it is close.

##b. Conduct a Bartlett's test for equal variance

```

#perform Bartlett test
salamander_var <- bartlett.test(salamander_data$length_2_mm,
                                salamander_data$unitttype)
salamander_var

##
## Bartlett test of homogeneity of variances
##
## data: salamander_data$length_2_mm and salamander_data$unitttype
## Bartlett's K-squared = 11.213, df = 2, p-value = 0.003674

```

Answer: These data do not have approximately equal variances because the p-value for the Bartlett test was less than 0.05.

#(2) Cutthroat Trout ##a. Are trout data normally distributed?

```

#filter dataset
trout_data <- vertebrate_data %>%
  filter(species == 'Cutthroat trout')

#evaluate normality with figure
fig3 <- ggplot(trout_data, aes(x = length_1_mm, fill = reach)) +
  geom_histogram(bins = 55, binwidth = 4, color = 'black') +
  facet_wrap(~reach) +
  labs(title = 'Cutthroat Trout Length by Reach',
        x = 'snout-to-fork length (mm)')
fig3

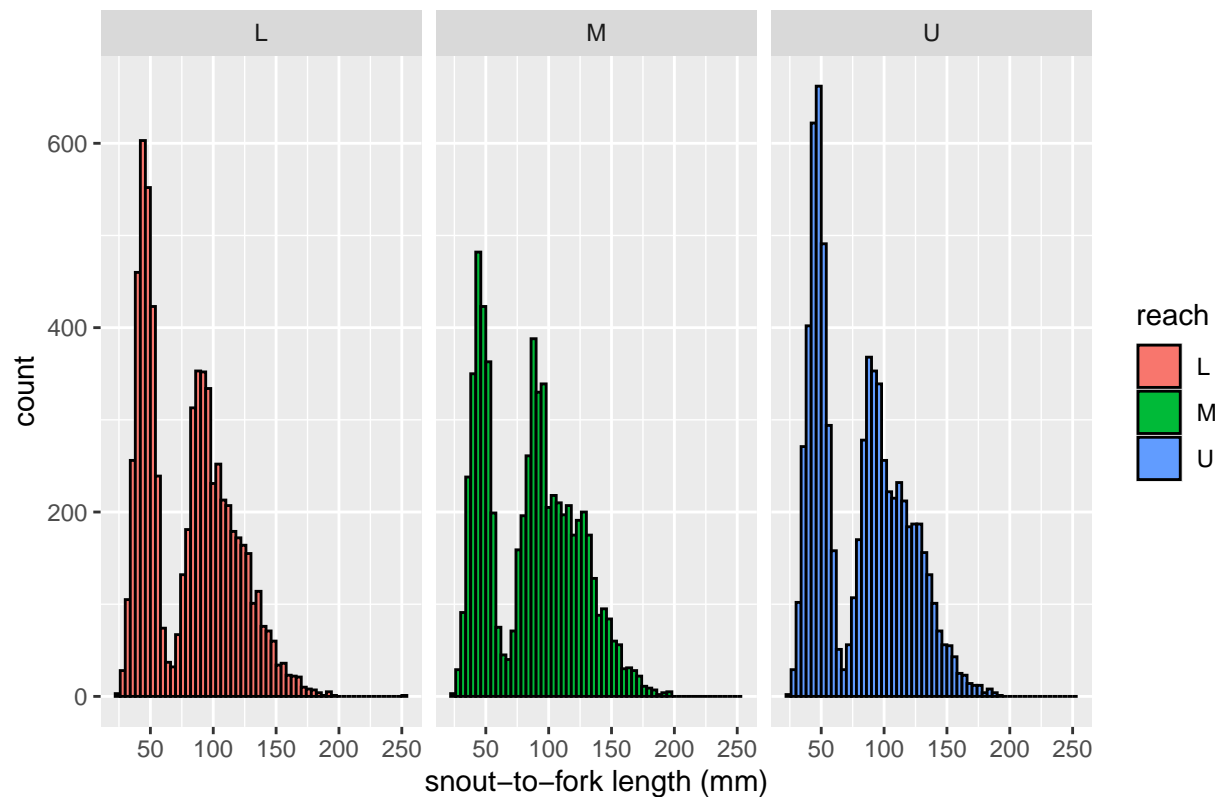
```

```

## Warning: Removed 5 rows containing non-finite outside the scale range
## ('stat_bin()').

```

Cutthroat Trout Length by Reach



```
#calculate skewness and kurtosis
skewness(trout_data$length_1_mm, na.rm = TRUE)
```

```
## [1] 0.3531134
```

```
kurtosis(trout_data$length_1_mm, na.rm = TRUE)
```

```
## [1] 2.177088
```

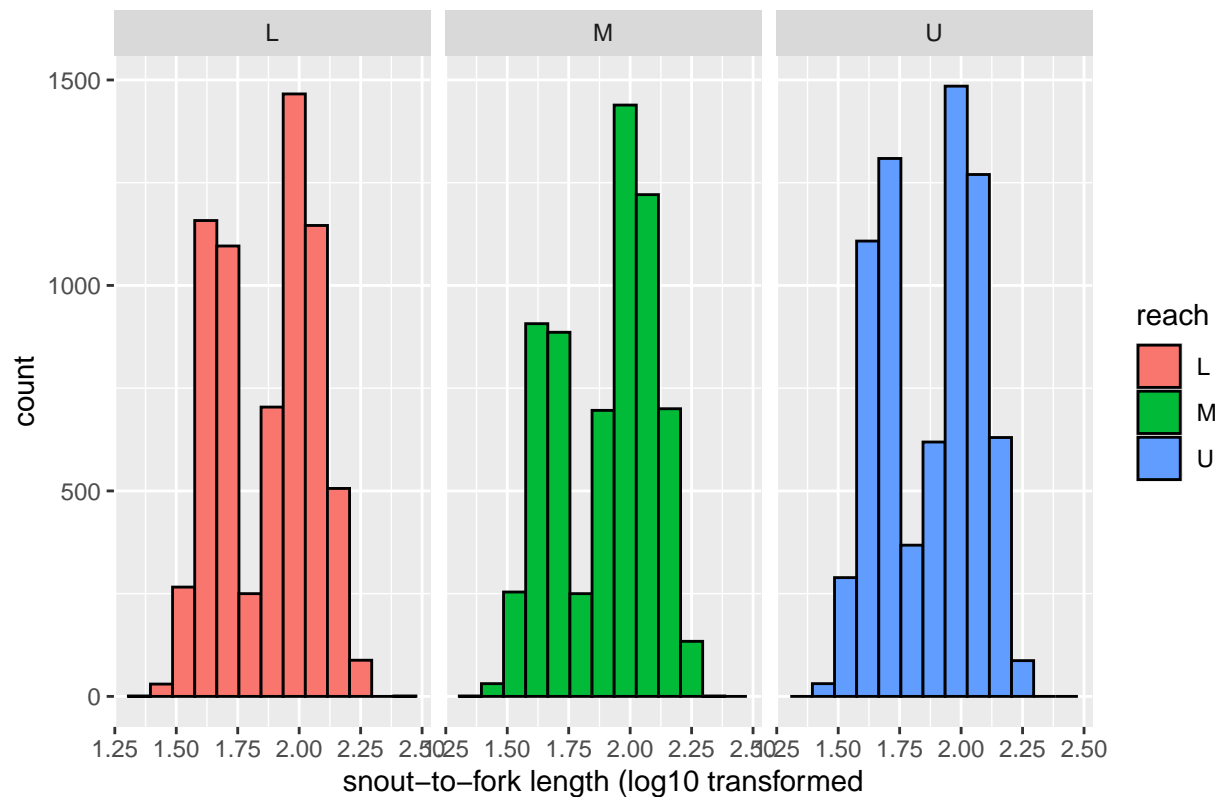
```
#not normal -> apply log-transformation
trout_data$length_1_mm <- log10(trout_data$length_1_mm)
```

```
#re-evaluate normality
```

```
fig4 <- ggplot(trout_data, aes(x = length_1_mm, fill = reach)) +
  geom_histogram(bins = 55, binwidth = 0.09, color = 'black') +
  facet_wrap(~reach) +
  labs(title = 'Cutthroat Trout Length by Reach',
       x = 'snout-to-fork length (log10 transformed)')
fig4
```

```
## Warning: Removed 5 rows containing non-finite outside the scale range
## ('stat_bin()').
```

Cutthroat Trout Length by Reach



```
skewness(trout_data$length_1_mm, na.rm = TRUE)
```

```
## [1] -0.1864256
```

```
kurtosis(trout_data$length_1_mm, na.rm = TRUE)
```

```
## [1] 1.756056
```

Answer: These data are not normally distributed as you can see in fig3. This is backed up by the kurtosis and skewness values. I applied a log-transformation to the data and re-evaluated. The data still do not appear to be normally distributed after the log-transformation. The skewness is less than 0 and the kurtosis is less than 3.