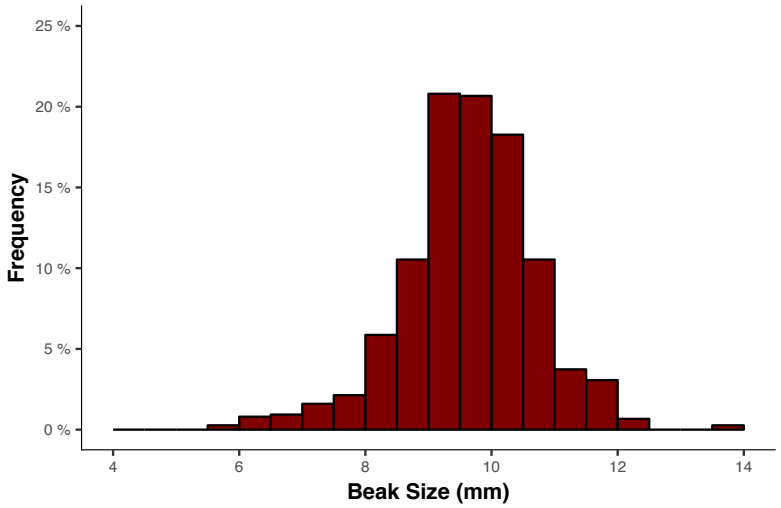
Data Analysis in Evol/Evol - HW Week1

Jing-Yi Lu

1. **Use R to read in the Darwin's finch data, uploaded here: dfdata.xls**
   Preview the document and reproduce figure 1.4-1. Compute the following summary statistics: mean, variance, coefficient of variation, and standard error.



Mean: 9.650.
Variance: 1.124.
Coefficient of variation: 11%
Standard error: 0.039.

* R Codes see appendix

2. **87, q.8**
   a. Increased by 10 times. ($\bar{X}' = 10\bar{X}$)
   b. Increased by 10 times. ($S_{X'} = |10|S_X$)
   c. Increased by 10 times.
   d. Increased by 10 times.
   e. Remained. ($CV_{X'} = \frac{S_{X'}}{\bar{X}'} = \frac{10S_X}{10\bar{X}} = CV_X$)
   f. Increased by 100 times. (${S_{X'}}^2 = 100{S_X}^2$)

3. **88, q.10: are more countries showing a decline or an increase in growth rate? Explain how you worked that out.**
   More countries are showing an increase in growth rate.
   I map the growth rate = 0 point onto the cumulative relative frequency curve for the corresponding frequency. The frequency is ~ 0.1 and far smaller than 0.5, indicating the countries with negative growth rate changes are less than a half in this survey.

4. **91, q.19**
   a. Females (1.7 > 1.5)

b. Females might live longer and/or be able to reproduce at older age compared to males. On the other hand, males might have a higher mortality rate at the younger age before bearing an offspring.
c. Females (4.3 > 3.5)

q.20
0

5. **110, q.15**
a. Standard error of the mean in men: 0.10
Standard error of the mean in women: 0.06
b. Standard deviation. The standard deviation is the spread of sample (the number of sexual partners.)
c. Standard error. The standard error is the spread of sampling estimates, that is, the precision of the estimates. Therefore, the standard error represents the uncertainty of the sample mean.
d. The discrepancy might be the results of confounding factor. As the survey samples people of a range of age, the results might be biased by the different age distribution of the two gender groups. That is, if the number of sexual partners is age-dependent generally, and the two groups have distinct age sampling, the estimates of two groups would also reflect the effects of age.

Appendix. R codes for beak size data

```
# Packages
library(tidyverse)
library(ggplot2)

# Retrieve data
Daphne_Island <- read.csv("dfdata.csv", header = T)

# Calculate the summary statistics for beak size
Summary_beak_size <- Daphne_Island %>%
  select(Beak_Size) %>%
  summarise(Beak_mean = mean(Beak_Size), Beak_variance = var(Beak_Size), CV =
sd(Beak_Size)/mean(Beak_Size), Beak_se = sd(Beak_Size)/sqrt(length(Beak_Size)))

# Reproduce the frequency histogram
Beak_size_histogram <- Daphne_Island %>%
  ggplot(aes(x = Beak_Size)) +
  geom_histogram(breaks = seq(4,14,0.5), fill = "#800000", colour = "black") +
  theme(panel.background = element_blank(),
        axis.text=element_text(size=7), axis.title=element_text(size=10,face="bold"),
        axis.line.x.bottom = element_line(color="black", size = 0.25),
        axis.line.y.left = element_line(color="black", size = 0.25)) +
  scale_x_continuous(breaks = seq(4,14,2)) +
  scale_y_continuous(breaks = seq(0, 187.5, 37.5), labels = paste(seq(0, 25, 5), "%"), limits = c(0,187.5))
+ xlab("Beak Size (mm)") + ylab("Frequency")
ggsave("Beak Size Frequency.pdf", dpi = 300, height = 8.4, width =12.7, units = "cm")
```