There are 7 questions, worth a total of 70 points.

You may use any calculator, R and any internet resources for this exam, although you are not allowed to ask someone else for help.

(10) 1. Consider the confidence interval for the mean of a variable. For each part, say if the confidence interval gets wider, gets narrower, or stays the same:
   (a) The sample size \( n \) increases.
   (b) The sample standard deviation \( s \) increases.
   (c) The sample mean \( \bar{x} \) increases.
   (d) The level of confidence \( C \) increases.

   **Solution:** (a) narrower (b) wider (c) same (d) wider

(10) 2. In 1798, Henry Cavendish measured the density of the Earth using a torsion balance. His measurements are the first column (*density*) in the dataset *Cavendish* in the HistData library. Each measurement is the density of the earth as a multiple of the density of water.

   (a) State and carry out a hypothesis test that his results are significantly different from the modern accepted density of 5.517 g/cm\(^3\).
   (b) The third value, 4.88, was later discovered to be incorrect. Cavendish actually measured 5.88 and copied it wrong. Change the value to 5.88, and run your test again. How does this affect the results?

   **Solution:** (a) \( t = -1.6834 \) with 28DF. His results are not significantly different from 5.517 (P = 0.1034). (b) Now P=0.3753 and his results differ even less significantly from the modern value.

(10) 3. The file *normtemp.csv* on our web page contains measurements of the heart rate and body temperature of 130 men and women.

   (a) Is there a significant difference in heart rate between men and women?
   Include the P-value in your answer.

   (b) Is there a significant difference in body temperature between men and women?
   Include the P-value in your answer.
There is a significant difference in body temperature ($P = .02$) but not in heart rate ($P = .53$).

4. In 1876, Darwin studied a variety of plants to compare cross-fertilized plants with self-fertilized plants. In one experiment, he planted thirty pairs of “candytuft” flowers, putting one self- and one cross-fertilized plant together in a pot. When the plants reached maturity, he measured their heights in inches. This data is on our web page as candytuft.csv.

(a) Which type of plant is taller, on average?

(b) State a hypothesis test that the mean heights of self- and cross-fertilized plants are different.

(c) Perform the test and report the results.

Using a matched pairs t-test gives $t = 3.0227$ with 29DF and $P = 0.005196$. Cross-fertilized plants are significantly taller.

(Independent samples $P = 0.001163$ is incorrect.)

5. (a) Look at the candytuft data from question 4 with appropriate graphical or summary methods. How comfortable are you using a t-test on this data? Why or why not?

(b) Test for a difference in plant heights using the Wilcoxon test and report the P-value. Does this change your conclusions from question 4?

There are two outliers in the self-fertilized plant measurements. One is a bit high, probably ok. The other is a value of 0 (in fact, the plant died), and is a major outlier. Using a t-test with this outlier is questionable. It is less of a problem with Wilcoxon, which reports $P = 0.0004877$. The conclusions don’t change.

(Unpaired test gives $P = 0.0001002$, which is the wrong thing to do.)

The next problem uses the dataset oly12 from the VGAMdata package that you will probably need to install. It has individual competitor information from the Summer 2012 London Olympic Games.

6. According to this data, which country won the most medals? How many did that country win?

(You need to sum Gold, Silver, and Bronze)
### Solution:

```r
oly12 %>% group_by(Country) %>% summarize(medals = sum(Gold + Silver + Bronze)) %>% arrange(desc(medals))
```

The US won the most medals, with 79.

---

(10) 7. Which country is the heaviest? Compute the mean weight of male athletes for all countries with at least 10 competitors, and report the top three.

### Solution:

```r
oly12 %>% group_by(Country) %>%
  filter(!is.na(Weight) & Sex == "M") %>%
  summarize(nc = n(), mw = mean(Weight)) %>%
  filter(nc > 5) %>%
  arrange(desc(mw))
```

# A tibble: 97 x 3
## Country  nc    mw
## <fctr> <int> <dbl>
1 Estonia  21 96.76190
2 Montenegro 16 95.50000
3 Iceland  15 94.93333

---

---