Homework#2

1. Fish Oil and Blood Pressure (One-Sample Analysis)

```
group1 <- subset(ex0112, Diet=="FishOil")
```

a) Compute:
- Sample average, $\bar{Y}_1$:
  ```
  > sample.mean <- mean(group1$BP)
  > sample.mean
  [1] 6.571429
  ```
- Sample Standard Deviation, s_1:
  ```
  > sample.sd <- sd(group1$BP)
  > sample.sd
  [1] 5.8554
  ```
- Degrees of freedom:
  = Sample size - # mean parameters
  ```
  > n <- length(group1$BP) #find sample size
  > n
  [1] 7
  ```
  = 7 – 1
  = **6**

b) Compute standard error for the average from this group (Fish Oil).
- SE($\bar{Y}_1$)
  = $s_1/\sqrt{n}$
  Where, S is the sample std. deviation and n is the sample size
  = $5.8554/\sqrt{7}$
  = **2.213**

c) Construct a 95% CI :
- CI = point estimate of parameter $\pm$ $t_d(0.975)$. SE($\bar{Y}_1$)
  = $\bar{Y}_1 \pm t_d(0.975)$. SE($\bar{Y}_1$)
  ```
  > qt(0.975, 6)
  [1] 2.446912
  ```
  = 6.57 $\pm$ (2.45*2.213)
  = **1.15 and 11.99**

d) For the hypothesis that μ_1 is zero, construct the t-statistic $\bar{Y}_1/\mathrm{SE}(\bar{Y}_1)$
- t-statistic
  = $\bar{Y}_1/\mathrm{SE}(\bar{Y}_1)$
  = 6.57/ 2.213
  = **2.97**

- Two-sided p-value as the proportion of values from a t_d- distribution farther from 0 than this value.
  Now that we have calculated t-statistic, we can get the two-sided p-value by:

  ```
  >2*pt(2.97, 6, lower.tail=FALSE)
  [1] 0.02495752
  ```

  = **0.02495752**

Answer these questions:

- Is there any evidence that the mean reduction for this group is different from zero?
  - There is moderate evidence that the mean reduction for the fish oil group is different from zero (p-value = 0.02498).

- What is the typical reduction in blood pressure expected from this type of diet (for individuals like these men)?
  - Expected typical reduction in blood pressure for Fish Oil diet: **6.57 mm** of mercury.

- Provide a 95% confidence interval.
  - 95% CI: from **1.15** to **11.99**

Statistical Conclusion:

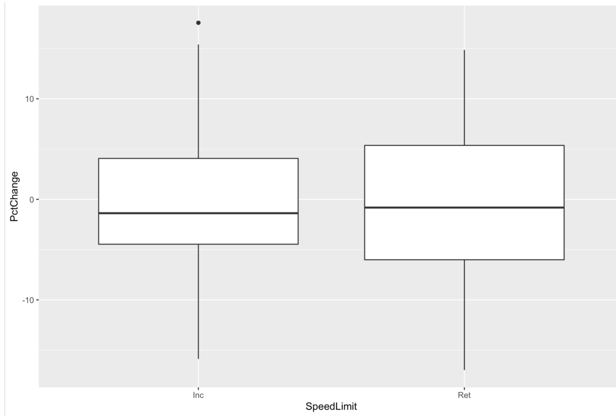
There is moderate evidence that the mean reduction for the fish oil group is different from zero (p-value = 0.02498). The mean reduction in blood pressure with Fish Oil diet was estimated as 6.57 mm of mercury (95% CI: from 1.15 to 11.99).

2. ***Speed Limits and Traffic Fatalities:***
   - *What evidence is there that the percentage change was greater in states that increased their speed limits?*
   - *How much of a difference is there?*
   - *Write a brief statistical report detailing the answers to these questions.*

Your statistical report need not be more than one page, including any figures you choose to include. The report should include statistical conclusions answering the research questions. Please follow the guidelines on the syllabus for writing the statistical conclusions. Your report should also include a scope of inference (see the case studies in the textbook for examples).

Let's have a look at the box plot to compare the percentage change in states that increased their speed limits with those that retained it.

```
> qplot(SpeedLimit, PctChange, data=ex0223, geom="boxplot")
```



By the boxplot, it seems like there wasn't much of a percentage change in states that increased their speed limits vs. those that retained their speed limits. We can do a hypothesis test to see if there's evidence that the pct. Change was greater in the states that increased their speed limits.

$$H_0: \mu_{Inc} - \mu_{Ret} = 0$$
$$H_A: \mu_{Inc} - \mu_{Ret} > 0$$

Doing a one-sided t-test:

```
>t.test(PctChange~SpeedLimit,data = ex0223,var.equal=TRUE, alternative="greater")
```

- This answers the first research question:
  Statistical conclusion:
  There is **no** evidence that the mean percent increase was higher in the states that increased their speed limits vs. those that retained (**one-sided p-value: 0.44**). The mean percent increase in fatalities in states that increase their speed limit was estimated to exceed by **0.39** pct. from those that retained the speed limits.

- To answer the second research question, we need to find the confidence interval, which will give the estimate of the difference. We want two-sided confidence intervals hence we'd do a two-sided t-test,
  ```
  >t.test(PctChange~SpeedLimit,data = ex0223, var.equal=TRUE)
  ```
  Statistical Conclusion:
  There is **no** evidence that the mean percent increase was higher in the states that increased their speed limits vs. those that retained (**two-sided p-value: 0.87**). The mean percent increase in fatalities in states that increase their speed limit was estimated to exceed by **0.39** pct. from those that retained the speed limits (**95% CI**: **-4.4 to 5.2** pct.).

Scope of Inference:

The samples of this study were not random; the inference to the population is speculative. The inference may only apply to the actual subjects & may not be representative of their respective groups. For eg., the states that retained speed limits might have faced more traffic fatalities due to other confounding variables like increase in road damages, drastic weather conditions, increase in construction zones. If so, statistical statements about the population could be biased. Hence, we do not have the statistical security to infer back to the population.

3.

(a) The sampling distribution of the sample determines the extent to which the sample statistic differs from one another and the population parameter.
The **sampling distribution of sample minimum** would be the distribution of the minimum sample statistic obtained by each sample randomly drawn from the same population.

(b) The **mean of sampling distribution of sample minimum:** From the sampling distribution of sample minimum that we obtain from repeated sampling of population, mean of sampling dist. of sample minimum would be the population mean μ.