Solutions For Homework 2

1. Load the NHIS sample from the course web page, drop observations where weight is missing, restrict the sample to men and draw a sample of size 25. (Please show your sample and your code)

```r
> NHIS <- read.table("C:\Teaching\Econ 113 Spring 2012\NHIS 2007 data.csv", header=TRUE,sep="",")
> NHIS_m <- subset(NHIS,SEX==1 & weight < 500)
> men_samp <- sample(NHIS_m$weight,25)
> men_samp
 137 200 220 240 190 163 220 165 245 220 295 275 176 155 180 295 165 166 130 180 210 190 150 190 232
```

2. Compute the mean and the standard error of the mean for this sample. (Please show your results and your code).

```r
> mean(men_samp)
[1] 199.56
> var(men_samp)^.5/25^.5
[1] 9.098
```

3. Do you believe your estimate from question 2 is an unbiased estimate of the average weight of an American male? Why or why not?

Yes because I have drawn a random sample from the population and I believe that people are honestly reporting their weights so the two assumptions under which the sample mean is an unbiased estimate of the population mean are met.
4. Use your sample from part 1 to conduct a significance test at the 5 percent level to test the null hypothesis that the average weight in the population is 185 pounds.

- State your null and alternative hypothesis,

\[ H_0 : \mu_{\text{height}} = 185. \]
\[ H_1 : \mu_{\text{height}} \neq 185. \]

- Estimate your t-statistic

\[ t_{\bar{x}} = \frac{\bar{x} - \mu_0}{se(\bar{x})} = \frac{199.56 - 185}{9.098} = 1.60 \]

- Conduct your significance test (under the assumption that weight is distributed normally) - Which distribution did you check your t-statistic against and why?

In this case the degrees of freedom is 24 (n-1). When we look up the cut off for a two sided 5 percent significance test on the table for the Student’s t-distribution we get a cut off of 2.064. I checked it against the Student’s t distribution because when the variable we are estimating the mean for is normally distributed the t-statistic follows the Students t-distribution.

- What do you conclude about the null hypothesis?

In this case I conclude that there is no compelling evidence to reject the null hypothesis that the mean weight in the population is 185 pounds.

5. Draw a sample of size 200 and following the steps from the prior question test the null hypothesis that the average weight of men is 185 pounds. Did you find evidence against the null. What distribution did you get your critical values from and why?

- State your null and alternative hypothesis,

\[ H_0 : \mu_{\text{height}} = 185. \]
$H_1 : \mu_{\text{height}} \neq 185.$

- Estimate your t-statistic

$$t_{\hat{\mu}} = \frac{\hat{\mu}_x - \mu_0}{SE(\hat{\mu}_x)} = \frac{193.7 - 185}{2.63} = 3.30$$

- Conduct your significance test (under the assumption that weight is distributed normally) - Which distribution did you check your t-statistic against and why?

In this case the degrees of freedom is 199 (n-1). When we look up the cut off for a two sided 5 percent significance test on the table for the normal distribution we get a cut off of 1.96. So I reject the null. I checked the t-statistic against the normal distribution because with a sample of size 200 the Lindberg-Levy Central Limit Theorem tells us that the t-statistic is distributed normal with mean of 0 and standard deviation of 1.

- What do you conclude about the null hypothesis?

In this case I conclude that there is compelling evidence that the null is false. I conclude that the average male probably doesn’t weight 185 pounds.

6. If people are not sure exactly how much they weigh but they are not systematically misrepresenting their weight, then the random variable reported weight . Lets assume the reporting error $u$ follows the normal distribution and has mean 0 and is independent of X and the sample drawn from X is independent and identically distributed. Show what impact will this have on how precise our estimate of the average weight is. Work this out explicitly in terms of the $Var(\hat{\mu}_x)$

$$Var(\hat{\mu}_x) = Var\left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var(x_i)$$
\[
\begin{align*}
&= \frac{1}{n^2} \sum_{i=1}^{n} Var(x_{i}^* + u_i) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} (Var(x_{i}^*) + Var(u_i)) \\
&= \frac{1}{n^2} \left( n \sigma_x^2 + n \sigma_u^2 \right) \\
&= \frac{\sigma_x^2 + \sigma_u^2}{n} \\
\end{align*}
\]

7. Take 1000 random samples of size 200 from the sample of males weights. Compute the sample mean from each sample and plot the histogram of these estimates of the mean. Does the distribution of the means look normally distributed? Do you think it should? (Please show your code and the histogram)

```r
means <- 1:1000

z <- 0
while(z<1000) {
    z <- z+1
    means[z] <- mean(sample(NHIS_m$weight,200))
}
hist(means,50)

h<-hist(means, breaks=40, col="red", xlab="Weight", main="Sample Estimates of Weight of Men (n=200) with Normal Distribution Over it",
xfit<-seq(min(means),max(means),length=40)
yfit<-dnorm(xfit,mean=mean(means),sd=sd(means))
yfit <- yfit*diff(h$mids[1:2])*length(means)
lines(xfit, yfit, col="blue", lwd=2)
```
The Estimates of Weight of Men (n=200) with Normal Distribution