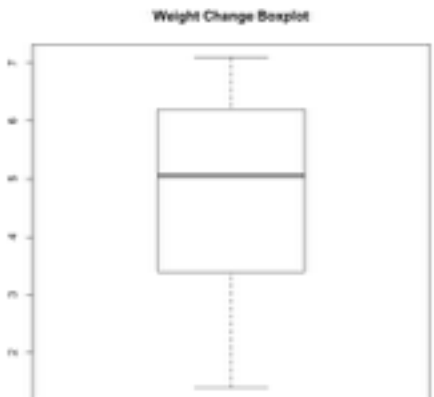
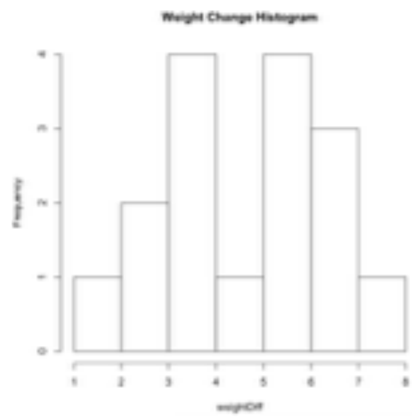
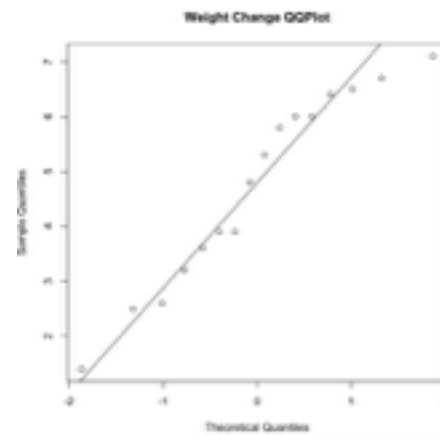
Harris Christiansen
christih@purdue.edu
STAT350 - Lab 6
2015-04-01

A. Food Intake And Weight Gain (all code can be found in appendix)
   1. This experiment should use a matched pairs t procedure because you are comparing each subjects weight after the 8 weeks to their weight before, while the weight person to person might vary greatly.

   2. Because we are looking at data for non-obese adults in a certain age range, all experiencing the same dietary change, I would expect the data to be normally distributed.







      The data appears reasonably normal and should be testable using a matched pairs t procedure.

   3. Paired t-test

      data: weightData$wta and weightData$wtb
      t = 10.8406, df = 15, p-value = 1.71e-08
      alternative hypothesis: true difference in means is not equal to 0
      95 percent confidence interval:
      3.801008 5.661492 (8.362 to 12.455 lbs)
      sample estimates:
      mean of the differences 4.73125 (10.408 lbs)

      The Paired t-test tells us that 95% of subjects statistically gained between 3.801 and 5.661 kgs.

   4. u is the population mean difference between weight before and weight after for the participants.

      H_0: u = 16
      H_a: u ≠ 16

$t_t = 10.8406$
DF = 15

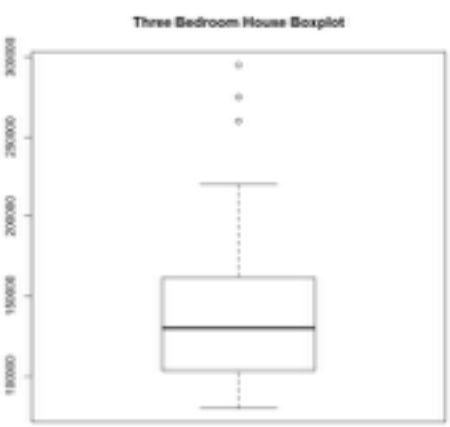
P-value = 0.0000000171

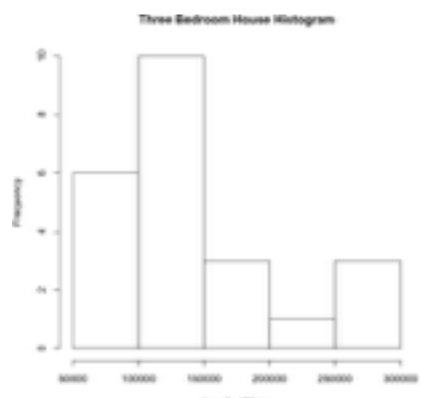
a = 0.05
Conclusion: Since the P-value < 0.05, we should reject H_0. The data provides strong evidence that the weight gained by eating an additional 56,000 calories does not raise your weight by 16 lbs, for the specific target group (age and weight)

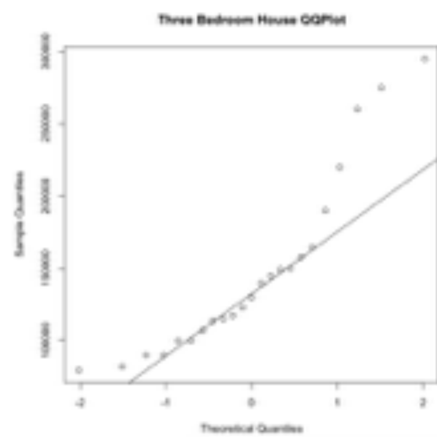
5. Yes, both tests are saying the same thing, that it is **not** likely that the average weight gained by eating an extra 56,000 calories is 16 lbs. 16lbs does not fall into the 95% confidence interval, and is rejected by the null hypothesis test.

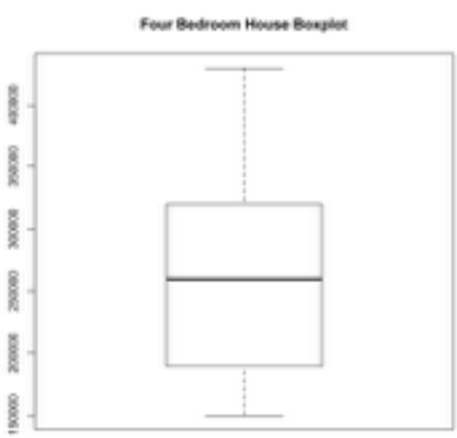
B. House Prices (all code can be found in appendix)
1. This experiment should use a 2-sample t procedure, because you are comparing two separate samples (four bedroom homes and three bedroom homes).

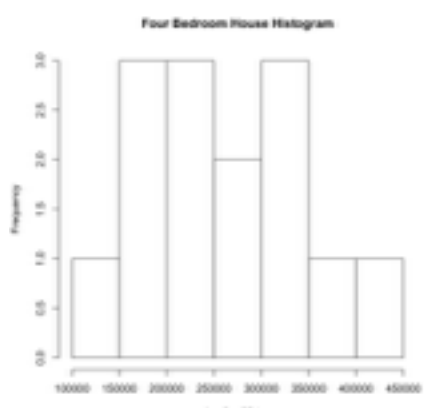
2. From the necessary figures for determining normality (see below), I can conclude that both three-bedroom and four-bedroom house data should be normal enough to preform a t-test on, however three-bedroom house data is slightly long/right skewed, while four bedroom house data is more normal.

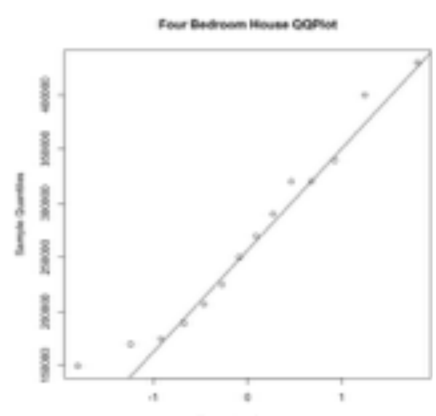












3. We are justified in using a two sample t test even though the data is not an SRS because we just want to compare 3 bedroom house prices with that of 4 bedroom house prices. The house prices are collected from the **same geographical area** and allow us to make our conclusion based on data from houses currently on the market.

4. Yes, a one-sided alternative for this analysis would allow you to conclude if one data sample has a smaller mean than the other, particularly meaningful to us for concluding of on average 3 bedroom houses are cheaper than 4 bedroom.

5. Welch Two Sample t-test

   data: threeBed$Price and fourBed$Price
   t = -4.4753, df = 20.976, p-value = 0.0001045
   alternative hypothesis: true difference in means is less than 0
   99 percent confidence interval:
   -Inf -52150.14
   sample estimates:
   mean of x mean of y
   147560.8 266792.9

   u_3 - u_4 is the population mean difference between Prices of 3 bedroom houses and 4 bedroom houses.

   H_0: $u_3 - u_4 = 0$
   H_a: $u_3 - u_4 < 0$

   $t_t = -4.4753$
   DF = 20.976

   P-Value = 0.0001045

   a = 0.01
   Conclusion: Since the P-Value < 0.01, we should fail to reject H_0 and therefor conclude that 3 bedroom houses are on average cheaper than 4 bedroom houses.

6. 99% Confidence Interval: -Inf -52150.14

7. Interpretation: We are 99% certain that 3 bedroom houses are $52150.14 cheaper than 4 bedroom houses.

8. Yes, both the 99% confidence interval and the null hypothesis test determine that 3 bedroom houses are substantially less than 4 bedroom houses.

Appendix:

```
##### Problem A #####

##### Part 2 #####
weightData <- read.table(file="wtgain.txt",header=T)
```

```r
weightDiff = weightData$wta - weightData$wtb
boxplot(weightDiff,main="Weight Change Boxplot") #Boxplot
hist(weightDiff,main="Weight Change Histogram") # Histogram
# QQPlot
qqnorm(weightDiff,main="Weight Change QQPlot")
qqline(weightDiff)

##### Part 3 #####
t.test(weightData$wta, weightData$wtb, conf.level=0.95, paired=TRUE,
 alternative="two.sided")

##### Problem B 4 #####

##### Part 2 #####
homeData <- read.table(file="houseprice.txt",header=T)
attach(homeData)
threeBed <- subset(homeData, Bedroom=="3")
fourBed <- subset(homeData, Bedroom=="4")
## 3 Bedroom
boxplot(threeBed$Price,main="Three Bedroom House Boxplot") #Boxplot
hist(threeBed$Price,main="Three Bedroom House Histogram") # Histogram
# QQPlot
qqnorm(threeBed$Price,main="Three Bedroom House QQPlot")
qqline(threeBed$Price)
## 4 Bedroom
boxplot(fourBed$Price,main="Four Bedroom House Boxplot") #Boxplot
hist(fourBed$Price,main="Four Bedroom House Histogram") # Histogram
# QQPlot
qqnorm(fourBed$Price,main="Four Bedroom House QQPlot")
qqline(fourBed$Price)

##### Part 5 #####
t.test(threeBed$Price, fourBed$Price, conf.level=0.99, paired=F,
 alternative = "less", var.equal=F)
```