Homework 7 Soln: Hypothesis Tests for 2 Independent and Dependent Populations due Fri 11/16 at 10 am

*Submit your homework to your TA's mailbox (and Canvas if you wish) before the due date/time. The mailboxes are to the left as you enter the Medical Science Center (1300 University Ave.) from the main University Ave. entrance.

*No late homework will be accepted for credit.

*If a problem asks you to use R, include a copy of the code and output. Please edit your code and output to be only the relevant portions.

*If a problem does not specify how to compute the answer, you may use any appropriate method but show the work that is necessary for the TA to follow and evluate your reasoning. I may ask you to use R or manual calculations on exam, so practice accordingly.

*Staple multiple papers together; present solutions to all prompts clearly; and put the discussion you attend on top to get full "Completion Points".

Total Points: 45 Completion Points: 10; Accuracy Points: 35

1. A researcher is interested in comparing the weight gain of young rabbits fed two different diets, diet I and diet II. The researcher selects 7 pairs of male litter mates (that is, 7 pairs of brothers) and randomly selects one rabbit from each pair for diet I and the other for diet II. The weight gains are recorded for each rabbit over an 8 week period. The weight gains for all rabbits (in grams) are given below:

Litter #	1	2	3	4	5	6	7
Diet I	368	293	401	314	384	404	267
Diet II	422	298	423	346	375	431	290

(a) It is claimed that the mean weight gain from diet II is 5 grams greater than the mean weight gain from diet I. Evidence against the claim will be provided when the weight gain from II exceeds the weight gain from I by more than 5 grams. Carry out a test for the claim at a 10% level by hand and then check your work in R. Interpret the results.

8 points; 2 points Hyp; 2 points t; 2 points pvalue and conclusion; 2 points check in R *ANSWER: This is a paired-sample problem as two brothers are dependent on each other.*
$X_1 = \{368, 293, 401, 314, 384, 404, 267\}$. *Let* μ_1 *is the population mean for the diet I.*
$X_2 = \{422, 298, 423, 346, 375, 431, 290\}$. *Let* μ_2 *is the population mean for the diet II.*
The difference values are $\{-54, -5, -22, -32, 9, -27, -23\}$. $H_0 : \mu_D = \mu_1 - \mu_2 = -5$ *versus* $H_1 : \mu_D < -5$.
$T = \frac{\bar{D}-(-5)}{S_D/\sqrt{n}} \sim T_{n-1}$ *under* H_0. $\bar{d} = -22$, $s_d = 20$, $T_{obs} = \frac{-22-(-5)}{20/\sqrt{7}} = -2.249$, *the p-value is* $P(t_6 \leq -2.249) = P(t_6 \geq 2.249) = pvalue$ *i.e.*$0.025 <$ *p-value* < 0.05. *We have sufficient evidence to reject the null at the 10% level. Evidence suggests the mean weight gain from diet II exceeds the mean weight gain from diet I by more than 5 grams.*

(b) State the assumptions necessary for performing the test in (a).

4 points; 2 points normality of differences; 2 points indep of differences *ANSWER: The assumption needed for the paired-sample T-test is the differences* $\{D_1, \ldots, D_n\}$ *are independently identically distributed as* $N(\mu_D, \sigma_D^2)$. *We are also assuming we do now know the population standard deviation* σ.

(c) Construct a 90% confidence interval for the difference in mean weight gains for the two diets. Compare this confidence interval to the conclusions from the test you performed above.

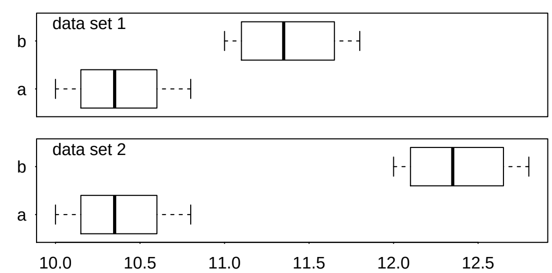
The $100(1-\alpha)\%$ *CI for the difference is* $\bar{d} \pm T_{(n-1,\alpha/2)}\frac{s_d}{\sqrt{n}}$ *where* s_d *is the sample sd of the differences and* $t_{(6,.05)} = 1.943$. *Hence the 90% CI of the difference between mean weight gains is* $(-36.69, -7.31)$ *grams. We notice that both interval values are below -5 for both; so we would also reject the null proposed in part (b) using the confidence interval.*

2. Data set 2 is collected to compare two treatments "a" and "b". The observations are collected independently for each treatment. Also the samples corresponding to treatments "a" and "b" are independent. Comparison of the treatments were previously done based on data set 1 under same independence assumptions. With data set 1 (top panel) the Wilcoxon Rank Sum/Mann-Whitney test test gives a p-value $p = 0.028$ and the t-test gives $p = 0.006$.

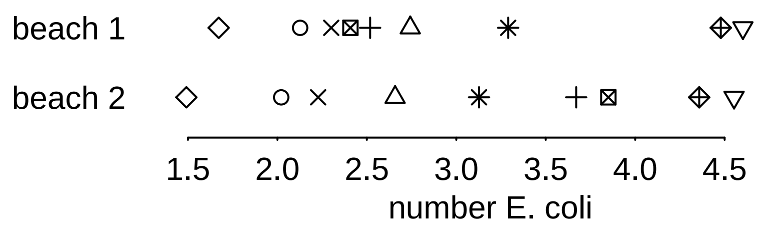
With data set 2 (bottom panel) which has the same number of values in each of treatments a and b (ie, $n_{a1} = n_{a2}$ and $n_{b1} = n_{b2}$)
the Wilcoxon Rank Sum/Mann-Whitney test test p-value is
☐ smaller than 0.028
☐ 0.028
☐ larger than 0.028

and the t-test p-value is
☐ smaller than 0.006
☐ 0.006
☐ larger than 0.006

Justify your choices.
4 points; 2 points each *ANSWER: For the Wilcoxon Rank Sum/Mann-Whitney test test, the p-value will be* 0.028 *again, because the Wilcoxon Rank Sum/Mann-Whitney test test is a rank based test and the ranks will not change. For the t-test, the p-value will be smaller than* 0.006 *because the distance between the means has increased and the variances are roughly same separately for both the samples.*



3. In a study of water quality in the Chicago area, concentration of *E. Coli* bacteria was measured in two specific locations along the lake ("Beach 1" and "Beach 2"). These measurements were taken after rainfalls, when a higher volume of water flows from the city into the lake. Rainfall dates were chosen to be at least a week apart, so that measurements taken from different dates can be considered as independent. The concentrations were measured in number of *E. Coli* bacteria per ml of water. They are shown below. Data taken on the same rainfall date are shown with the same symbol (with 9 rainfall dates total).



The investigator wants to know if the concentration of *E. Coli* in the water is the same at these two beach locations.

(a) Determine which test might be appropriate to answer the investigators questions: check all that apply in the list below. Justify your choices.

☐ Two independent-sample t-test ☐ Welch's t-test ☐ Wilcoxon Rank Sum/Mann-Whitney test

☐ Paired-sample t-test ☐ Sign test ☐ Wilcoxon Signed Rank test

☐ Bootstrap test for one sample ☐ Bootstrap test for two samples

4 points; 1 for each *The data are paired by sampling date. There are 4 appropriate choices: Paired-sample t-test, Sign test and Wilcoxon Signed Rank test, bootstrap test for one sample.*

4. In an effort to link cold environment to an increase in mean blood pressure, two random samples of 5 rats each were exposed to different environments. One sample of rats was held in a normal environment of 26°C and the other was held aat 15°C. Blood pressures were measured for rats of both groups after 1 day and are given below:

26°C BP:	214	194	221	198	212
15°C BP:	225	215	253	272	254

(a) If the scientists want to assume that the necessary populations are normal, what test[s] would be reasonable to run? Explain why. Identify the hypotheses of interest and then run this/these tests in R and report the test statistic, degrees of freedom, and resulting p value (it may also be useful to compute these by hand for practice, but we will not be grading you on it).

ANSWER: If we assume normality of each population and independence of the two populations (and independence within), then need to decide whether to do an equal variance or Welch's two sample t test. Because the sd of the rats held at 15 is more than double the 26 degree rats, I would choose the Welch's T. Note, it would also be ok to do a bootstrap, permutation, or wilcoxon rank sum- they are just typically a bit less poserful.

8 points; 2 for hyp; 2 for t, 2 for t; 2 for pvalue (equal variance assumption would be wrong -4.) *ANSWER: Hypothesis:* $H_o : \mu_{15} - \mu_{26} = 0$ *vs* $H_o : \mu_{15} - \mu_{26} > 0$ *where* μ_{15} *and* μ_{26} *stand for the true average blood pressure of rats in 15 and 26 degree climates, respectively.* $t_{obs} = 3.1078$, *df* $= 5.8054$, *and pvalue*=0.01092

(b) If the researcher does not want to assume that the relevant populations are normally distributed, what 3 tests could they perform?

3 points; 1 for each *We have two independent samples, so they could perform a permutation, bootstrap, or wilcoxon-rank sum test.*

(c) Perform each of the three tests listed above (using set.seed(1) and B=10000 for any that require computer simulation) and (i) identify the assumptions (ii) report the observed test statistic and (iii) the resulting p value.

ANSWER: Bootstrap: independence between and within; $t_{obs} = 3.10784$, *p value=193/10000=0.0193; Permutation:independence between and within; p value=74/10000=0.0074; Wilcoxon: we're also* *assuming the distributions have same shape and spread and symmetric if we want to talk about the means; W=24, p value=0.00794*

(d) Based on your findings from the tests in parts a and b, what conclusion would you draw and what recommendations would you give to the scientist?

We have strong enough evidence at the 5% level regardless of the test that we run. Depending on the assumptions they want to make, it wouldn't be unreasonable to run any of the forementioned tests.

5. A reporter for Time Magazine was interested in residents' levels of worry about being the victim of crime in their neighborhood. They performed a telephone poll of 500 adult Americans, 140 from urban areas, 160 from suburban, and 200 from rural areas. The number of adults who reported worring about being a victim of crime was urban:83, suburban: 92, and rural: 86.

(a) Perform a hypothesis test at the 5% level of significance to determine if there is evidence of a difference in the proportion of urban and suburban residents who worry about being the vicim of crime? (Be sure to state your hypotheses, assumptions, and show your computations.)

ANSWER: $H_o: \pi_U - \pi_S = 0$ *vs* $H_o: \pi_U - \pi_S \neq 0$. *We check our assumptions that number of successes and number of failures is larger than 5 for each sample: Urban:* $n_u\hat{p_U} = 83$, $n_U(1-\hat{p_U}) = (140-83) = 57$, *and Suburban:*$n_S\hat{p_S} = 92$, $n_S(1-\hat{p_S}) = (160-92) = 68$. *Yes, so the CLT can reasonably assumed and we'll use a normal approximation. Under the null of equal proportion, we find a pooled estimate of success:* $\pi_V = \frac{83+92}{140+160} = \frac{175}{300} = 0.5833$. *We then can find an estimate for the* $SE_{diff} = \sqrt{0.5833 * (1-0.5833)(1/140+1/160)} = 0.05705443$. *Our observed difference:* $83/140 - 92/160 = 0.01785$ *So* $Z = \frac{0.01785-0}{0.05705443} = 0.313$ *so our p value:* $2 * P(Z \geq 0.313) = 2 * 0.377 = 0.754$

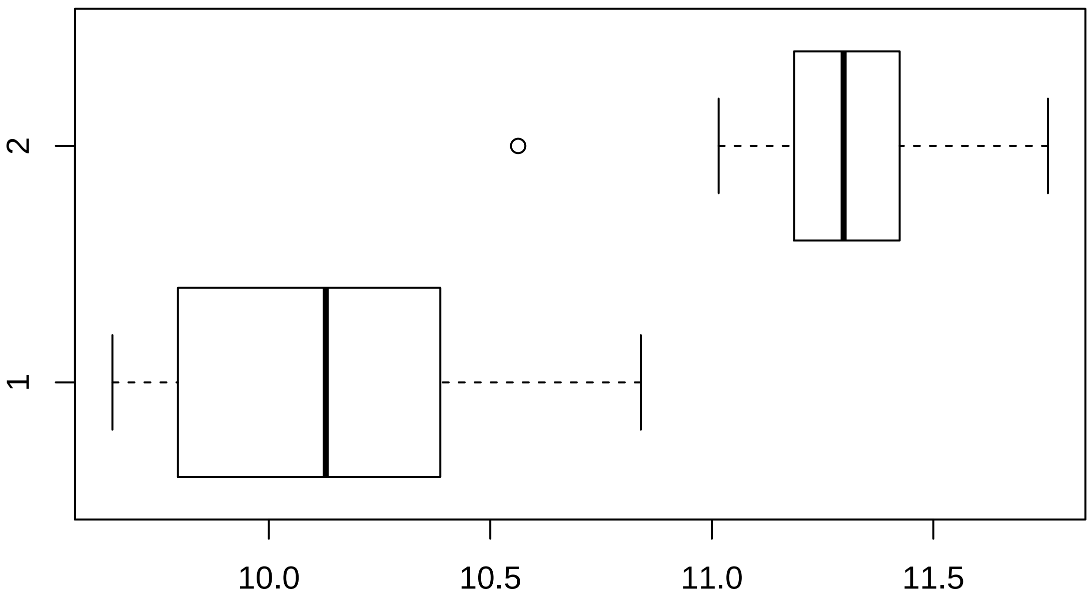
(b) Create a 95% confidence interval for the difference in proportion of rural and urban residents who worry about being a victim of crime.(Be sure to state your assumptions and show your computations.)

4 points; 2 for assuptions; 2 for computation *We need to assume that the 2 populations are independent and observations within samples are independent. We also check our assumptions that number of successes and number of failures is larger than 5 for each sample: Urban:* $n_u\hat{p_U} = 83$, $n_U(1-\hat{p_U}) = (140-83) = 57$, *and Rural:*$n_R\hat{p_R} = 86$, $n_R(1-\hat{p_R}) = (200-86) = 114$. *Yes, so the CLT can reasonably assumed and we'll use a normal approximation. Our observed difference:* $86/200 - 83/140 = -0.1628571$ *and the* $SE_{diff} = \sqrt{\frac{(86/200)*(114/200)}{200} + \frac{(83/140)*(57/140)}{140}} = 0.05431$ *and* $z_{.025} = 1.96$ *So a 95% CI is* $-0.162857 \pm 1.96 * 0.05431 = (-0.269, -0.0564)$

```
#Question 1
B1<-c(368, 293, 401, 314, 384, 404, 267)
B2<-c(422, 298, 423, 346, 375, 431, 290)
(Diff<-B1-B2)

## [1] -54  -5 -22 -32   9 -27 -23

mean(Diff); sd(Diff)

## [1] -22

## [1] 20

t.test(B1, B2, paired=TRUE, conf.level=.90, mu=-5, alternative="less")

##
##  Paired t-test
```

```
##
## data:  B1 and B2
## t = -2.2489, df = 6, p-value = 0.03277
## alternative hypothesis: true difference in means is less than -5
## 90 percent confidence interval:
##       -Inf -11.11647
## sample estimates:
## mean of the differences
##                     -22
```

```r
#Question 2
dataA<-rnorm(10,10.2, .3); data1b<-rnorm(10,11.25, .3)
boxplot(dataA, data1b, horizontal = TRUE)
```



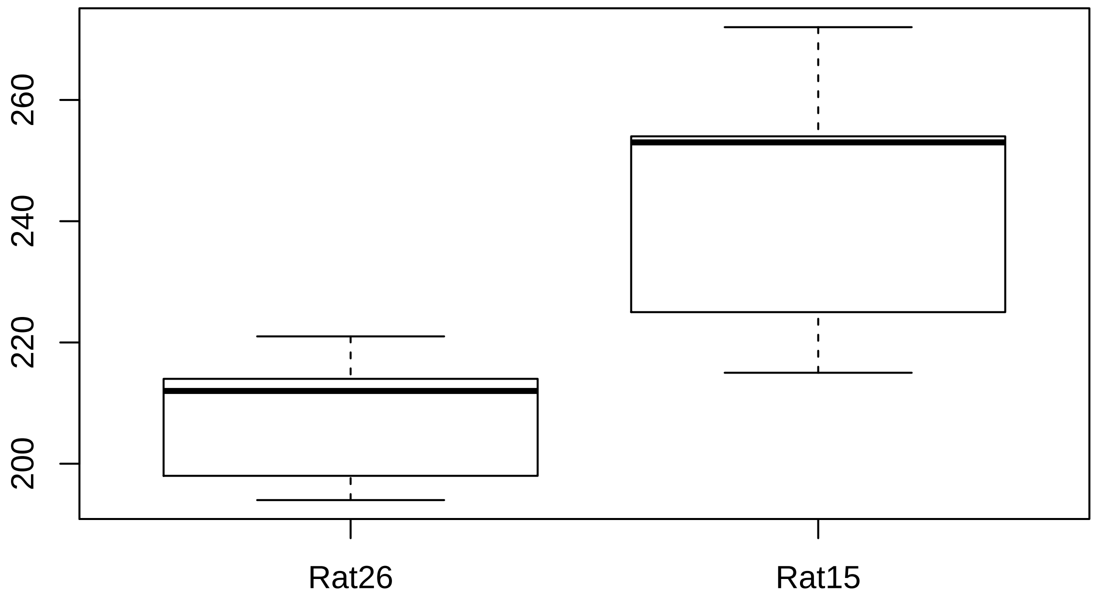

```r
wilcox.test(dataA, data1b)
```

```
##
##  Wilcoxon rank sum test
##
## data:  dataA and data1b
## W = 1, p-value = 2.165e-05
## alternative hypothesis: true location shift is not equal to 0
```

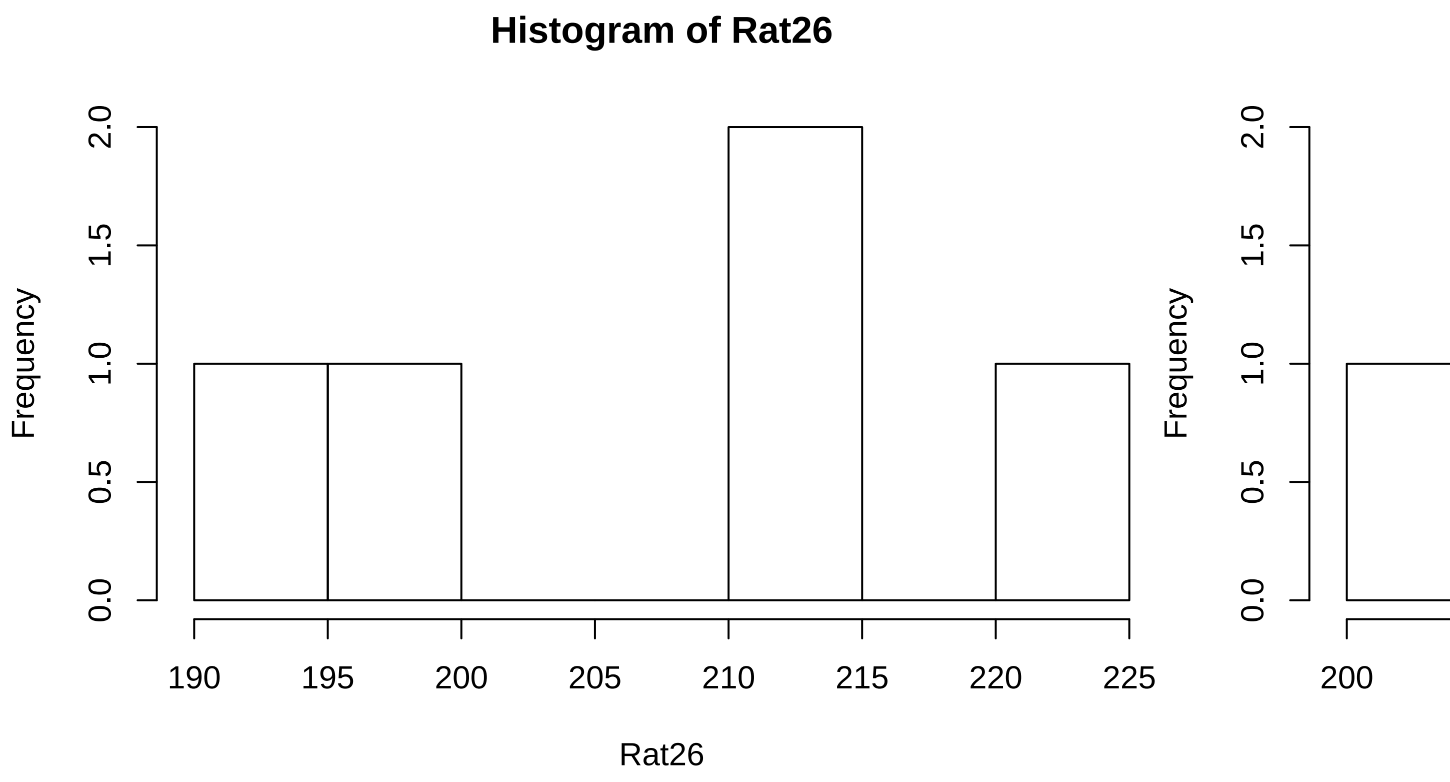
```r
wilcox.test(dataA, data1b+5)
```

```
##
##  Wilcoxon rank sum test
##
## data:  dataA and data1b + 5
## W = 0, p-value = 1.083e-05
## alternative hypothesis: true location shift is not equal to 0
```

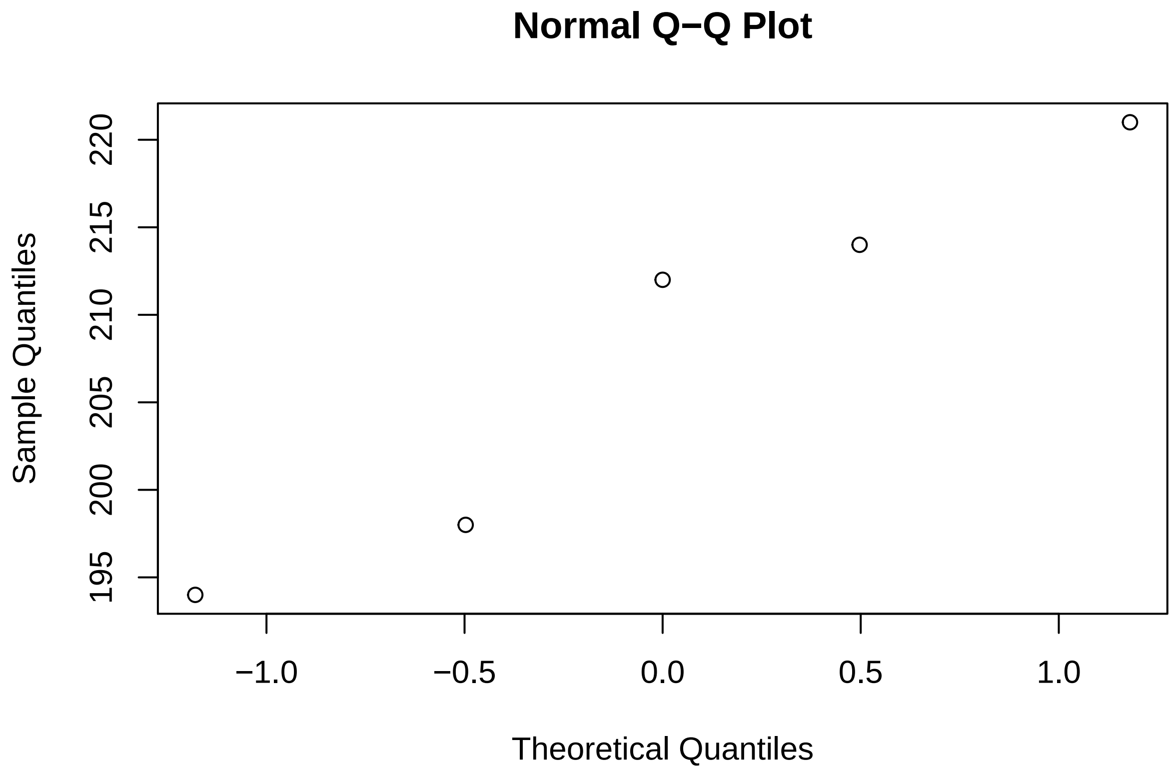
```r
#Question 4
Rat26<-c(214, 194, 221, 198, 212)
Rat15<-c(225, 215, 253, 272, 254)
boxplot(Rat26, Rat15, names=c("Rat26", "Rat15"))
```

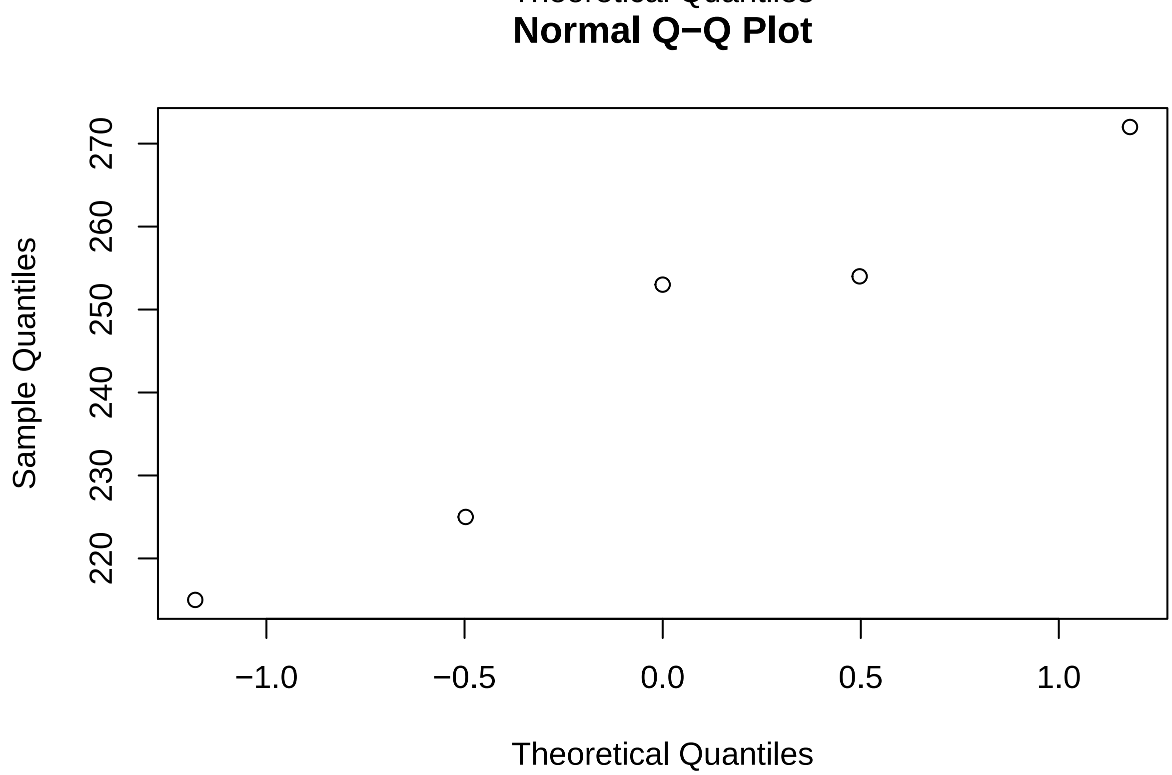



```
hist(Rat26); hist(Rat15)
```




```
qqnorm(Rat26); qqnorm(Rat15)
```






```
sd(Rat26); sd(Rat15)
```

```
## [1] 11.36662
```

```
## [1] 23.27445
```

```r
t.test(Rat26, Rat15, var.equal=FALSE, alternative = "less")
```

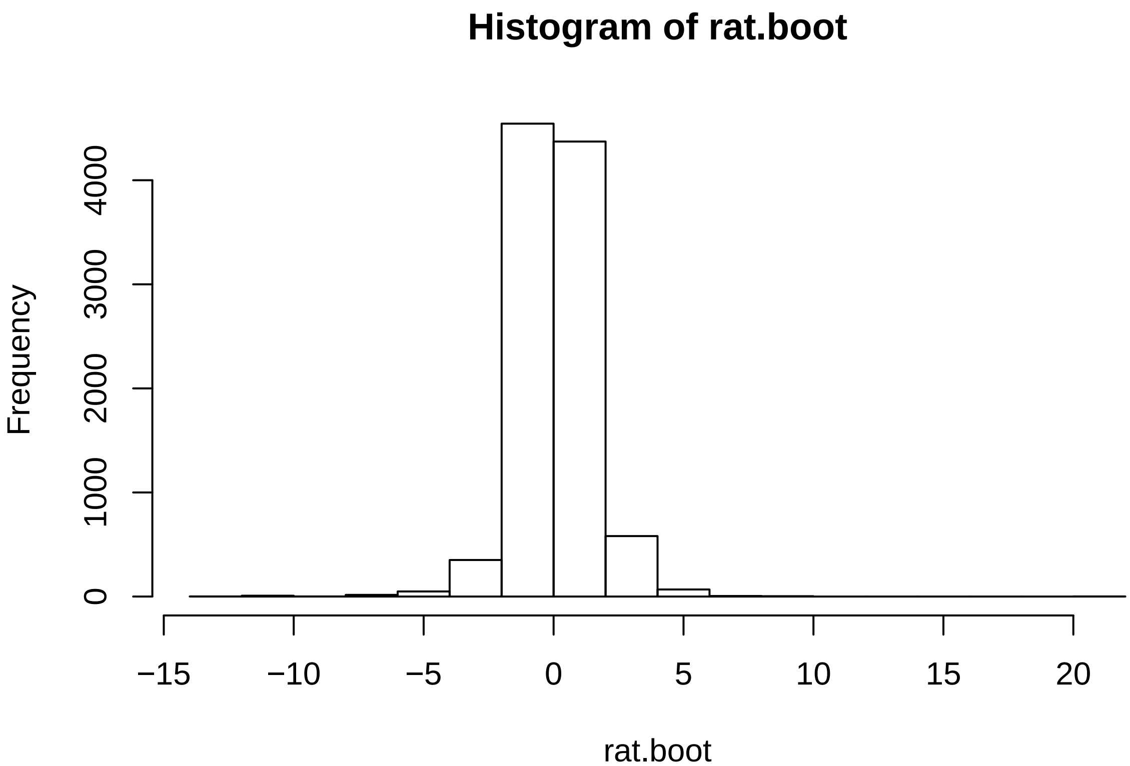
```
##
##  Welch Two Sample t-test
##
## data:  Rat26 and Rat15
## t = -3.1078, df = 5.8054, p-value = 0.01092
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -13.35509
## sample estimates:
## mean of x mean of y
##     207.8     243.8
```

```r
t.test(Rat15, Rat26, var.equal=FALSE, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  Rat15 and Rat26
## t = 3.1078, df = 5.8054, p-value = 0.01092
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  13.35509      Inf
## sample estimates:
## mean of x mean of y
##     243.8     207.8
```

```r
#function to perform permutation test#
#dat1 is data from first group, dat2 is from second#
#nperm is the number of times to do the random splitting#
permtwo <- function(dat1, dat2, nperm) {
    permstat <- NULL
    for(i in 1:nperm) {
        n1 <- length(dat1)  #find length of first sample
        n2 <- length(dat2)  #find length of second sample
        alldat <- c(dat1, dat2)  #combine into 1 large sample
        samp <- sample(alldat, replace=FALSE)  #shuffle the combined vector of values
        pmean1 <- mean(samp[1:n1])   #find mean of first n1 observations as x1*-bar
        pmean2 <- mean(samp[(n1+1):(n1+n2)]) #find mean of last n2 observations as x2*-bar
        permstat[i] <- pmean1 - pmean2
}
    return(permstat)
}

# Here's one way to do the bootstrap for a difference of two means in R:
# dat1 and dat2 are data from the two groups. nboot is the number of resamples.
#Notice obsdiff is computed as mean(dat1)-mean(dat2) - order matters!
boottwo = function(dat1, dat2, nboot) {
  bootstat = numeric(nboot)
  obsdiff = mean(dat1) - mean(dat2)
  n1 = length(dat1)
  n2 = length(dat2)
  for(i in 1:nboot) {
    samp1 = sample(dat1, size = n1, replace = T)
```

```r
    samp2 = sample(dat2, size = n2, replace = T)
    bootmean1 = mean(samp1)
    bootmean2 = mean(samp2)
    bootvar1 = var(samp1)
    bootvar2 = var(samp2)
    bootstat[i] = ((bootmean1 - bootmean2) - obsdiff)/sqrt((bootvar1/n1) + (bootvar2/n2))
}
  return(bootstat)
}

B=10000
set.seed(1)
rat.boot<-boottwo(Rat15, Rat26,B)
hist(rat.boot)
```



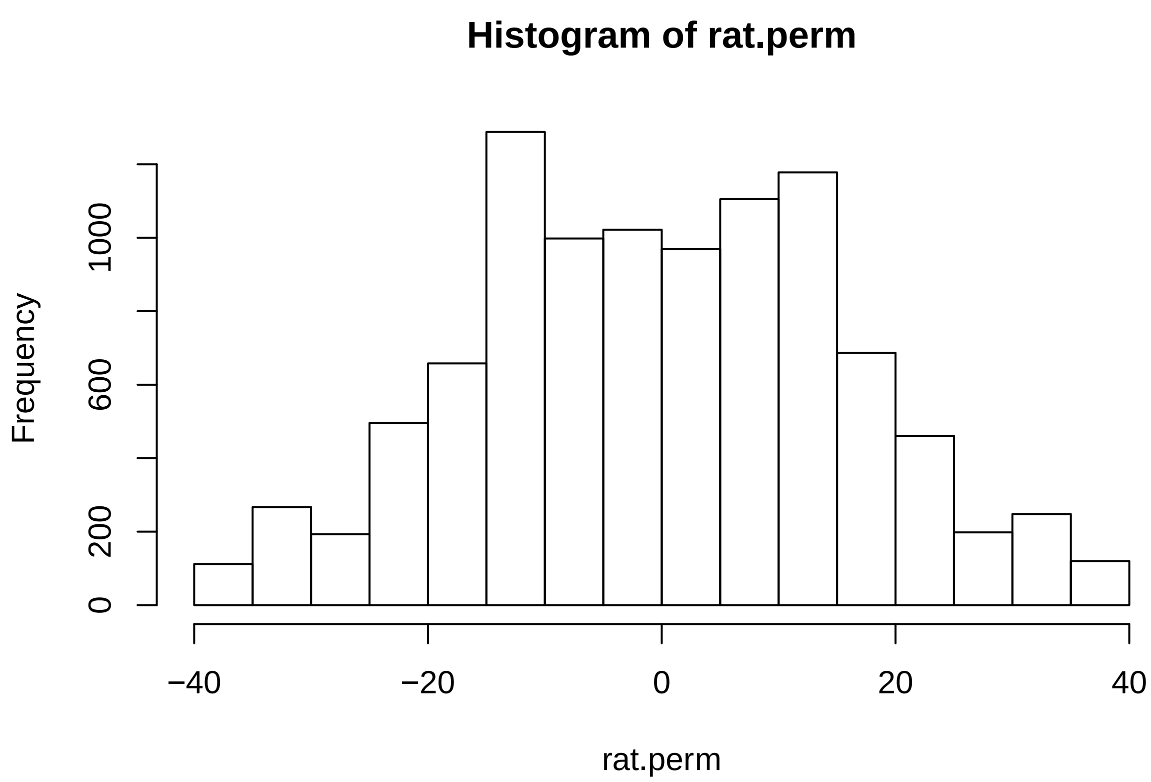

```r
(rat.t.obs=(mean(Rat15)-mean(Rat26))/sqrt(var(Rat15)/length(Rat15)+var(Rat26)/length(Rat26)))   #3.10784
```

```
## [1] 3.10784
```

```r
boot.mup<-sum(rat.boot >= rat.t.obs) #193
boot.low<-sum(rat.boot <=rat.t.obs) #9807
(boot.pval=boot.mup/B)   #0.0193
```

```
## [1] 0.0193
```

```r
rat.perm<-permtwo(Rat15, Rat26, B)
hist(rat.perm)
```




```
(rat.diff.obs=mean(Rat15)-mean(Rat26))   #diff=36
```

```
## [1] 36
```

```
(perm.mup<-sum(rat.perm >=rat.diff.obs))   #74
```

```
## [1] 74
```

```
(perm.pval=perm.mup/B)   #0.0074
```

```
## [1] 0.0074
```

```
wilcox.test(Rat15, Rat26, alternative="greater")   #W = 24, p-value = 0.007937
```

```
##
##  Wilcoxon rank sum test
##
## data:  Rat15 and Rat26
## W = 24, p-value = 0.007937
## alternative hypothesis: true location shift is greater than 0
```

```
#Question  5
prop.test(x=c(83,92), n=c(140,160), correct=FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(83, 92) out of c(140, 160)
## X-squared = 0.097959, df = 1, p-value = 0.7543
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.09390325  0.12961753
```

```
## sample estimates:
##    prop 1    prop 2
## 0.5928571 0.5750000
```

```r
prop.test(x=c(86, 83), n=c(200,140), correct=FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(86, 83) out of c(200, 140)
## X-squared = 8.7371, df = 1, p-value = 0.003118
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.26930368 -0.05641061
## sample estimates:
##    prop 1    prop 2
## 0.4300000 0.5928571
```