Math 3070 § 1. **Plasma Acid Example: Data of Unequal** Name: _____ 𝔈𝔵𝔞𝔪𝔭𝔩𝔢 _____
Treibergs   **Length, Two Sample** t**-Test with** $\sigma_1 \neq \sigma_2$. June 1, 2011

Data File Used in this Analysis:

```
# M 3070 - 1     Plasma Ascorbic Acid Data        6-16-11
# Treibergs
#
# From Walpole, Myers, Myers, Ye, "Probability and Statistics for Engineers
# and Scientists, 7th ed.," Prentice Hall, 2002.
#
# data from Virginia Poly study of plasma ascorbic acid levels of pregnant
# women who are smokers  versus nonsmokers. 33 healthy women in last
# trimester of age 15 to 32 had plasma ascorbic acid (milligrams per 100
# milliliters) measured in blood samples. Is there sufficient evidence to
# conclude that there is a difference of ascorbic acid levels in smokers
# and nonsmokers? Assume data sets from normal pops with unequal variances.
NonSmoker Smoker
0.97    .48
0.72    .71
1.00    .98
0.81    .68
0.62   1.18
1.32   1.36
1.24    .78
0.99   1.64
0.90
0.74
0.88
0.94
1.16
0.86
0.85
0.58
0.57
0.64
0.98
1.09
0.92
0.78
1.24
1.18
```

R Session:

```
R version 2.11.1 (2010-05-31)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
```

```
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.34 (5589) i386-apple-darwin9.8.0]

[Workspace restored from /home/1004/ma/treibergs/.RData]

> ########################### READ DATA ###################################
> tt <- read.table("M3074PlasmaAcidEg.txt",header=TRUE)
Error in scan(file, what, nmax, sep, dec, quote, skip, nlines, na.strings,  :
  line 13 did not have 2 elements

> ######################## FIX DATA FILE. FIRST METHOD ####################
> # Two fixes in wordprocessor:
> # First Fix: make single vector and a second factor indicating smoking
```

Data File Used in this Analysis:

```
# M 3070 - 1    Plasma Ascorbic Acid Data 1     6-16-11
# Treibergs
#
# From Walpole, Myers, Myers, Ye, "Probability and Statitics for Engineers
# and Scientists, 7th ed.," Prentice Hall, 2002.
#
# data from Virginia Poly study of plasma ascorbic acid levels of pregnant
# women who are smokers  versus nonsmokers. 33 healthy women in last
# trimester of age 15 to 32 had plasma ascorbic acid (milligrams per 100
# milliliters) measured in blood samples. Is there sufficient evidence to
# conclude that there is a difference of ascorbic acid levels in smokers
# and nonsmokers? Assume data sets from normal pops with unequal variances.
# Variables
#    Plasma   plasma ascorbic acid (milligrams per 100 milliliters)
#    Smoker/Nonsmoker   factor: Y=yes, N=no
Plasma Smoker/Nonsmoker
.97 N
0.72 N
1.00 N
0.81 N
0.62 N
1.32 N
1.24 N
```

```
0.99 N
0.90 N
0.74 N
0.88 N
0.94 N
1.16 N
0.86 N
0.85 N
0.58 N
0.57 N
0.64 N
0.98 N
1.09 N
0.92 N
0.78 N
1.24 N
1.18 N
0.48 Y
0.71 Y
0.98 Y
0.68 Y
1.18 Y
1.36 Y
0.78 Y
1.64 Y
```

Resume R Session:

```
> tt <- read.table("M3074PlasmaAcidData1.txt",header=TRUE)
> tt
   Plasma Smoker.Nonsmoker
1    0.97                N
2    0.72                N
3    1.00                N
4    0.81                N
5    0.62                N
6    1.32                N
7    1.24                N
8    0.99                N
9    0.90                N
10   0.74                N
11   0.88                N
12   0.94                N
13   1.16                N
14   0.86                N
15   0.85                N
16   0.58                N
17   0.57                N
18   0.64                N
19   0.98                N
20   1.09                N
```

```
21   0.92              N
22   0.78              N
23   1.24              N
24   1.18              N
25   0.48              Y
26   0.71              Y
27   0.98              Y
28   0.68              Y
29   1.18              Y
30   1.36              Y
31   0.78              Y
32   1.64              Y

> attach(tt)
> names(tt)
[1] "Plasma"            "Smoker.Nonsmoker"
> S.N <- factor(Smoker.Nonsmoker)
>
> layout(matrix(1:4,ncol=2,byrow=TRUE))
> stripchart(Plasma~Smoker.Nonsmoker,method="stack",col=c(4,2),pch=19,ylim=c(0,3),
+ ylab="Smoke",main="Plasma Ascorbic Acid")
> plot(Plasma~Smoker.Nonsmoker, data = tt, col = c("lightblue","pink"),
+ main = "Plasma Ascorbic Acid", horizontal = TRUE)
> # M3074PlasmaAcid2.pdf
>
> # Means are clese but variances differ.
```


Plasma Ascorbic Acid
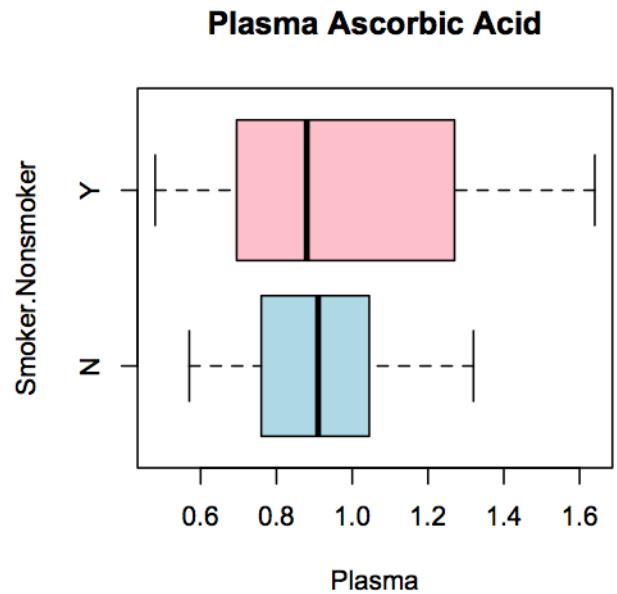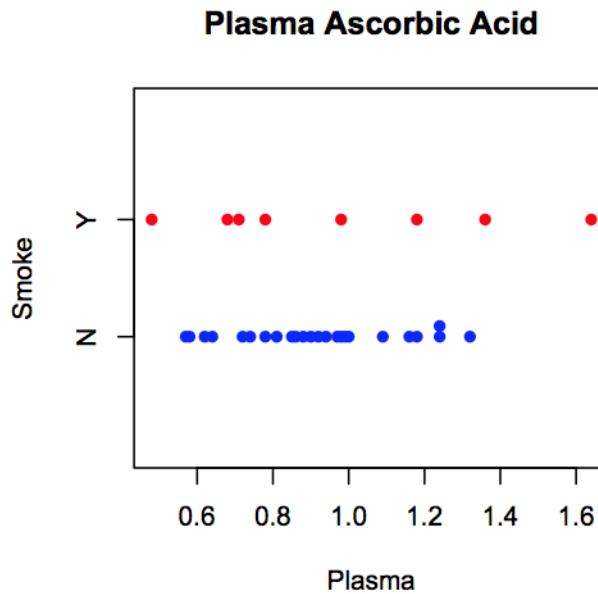
```
> ######################## PICK OFF EACH FACTOR ###############################
> NonSmoker <- Plasma[S.N=="N"]
> Smoker <- Plasma[S.N=="Y"]
> t.test(Smoker,NonSmoker)

        Welch Two Sample t-test

data:  Smoker and NonSmoker
t = 0.4162, df = 8.444, p-value = 0.6876
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2712908  0.3921241
sample estimates:
mean of x mean of y
0.9762500 0.9158333


> # High p-value. Conclude that there is no significant evidence that the two
> # plasma ascorbic acid levels are different in smokers and non-smokers.
>
> ######################## SECOND FIX OF DATA FILE ##############################
> # Alternative is to put NA where data doesn't match
```

Data File Used in this Analysis:

```
# M 3070 - 1     Plasma Ascorbic Acid Data 2       6-16-11
# Treibergs
#
# From Walpole, Myers, Myers, Ye, "Probability and Statitics for Engineers
# and Scientists, 7th ed.," Prentice Hall, 2002.
#
# data from Virginia PolyYtudy of plasma ascorbic acid levels of pregnant
# women who areYmokers  versus nonsmokers. 33 healthy women in last
# trimester of age 15 to 32 had plasma ascorbic acid (milligrams per 100
# milliliters) measured in blood samples. Is there sufficient evidence to
# conclude that there is a difference of ascorbic acid levels inYmokers
# and nonsmokers? Assume dataYets from normal pops with unequal variances.
NonSmoke Smoke
0.97    .48
0.72    .71
1.00    .98
0.81    .68
0.62   1.18
1.32   1.36
1.24    .78
0.99   1.64
0.90    NA
0.74    NA
0.88    NA
0.94    NA
1.16    NA
0.86    NA
```

```
0.85    NA
0.58    NA
0.57    NA
0.64    NA
0.98    NA
1.09    NA
0.92    NA
0.78    NA
1.24    NA
1.18    NA
```

Resume R Session:

```
> t2 <- read.table("M3074PlasmaAcidData2.txt",header=TRUE)
> attach(t2)
> names(t2)
[1] "NonSmoke" "Smoke"
> t2
   NonSmoke Smoke
1      0.97  0.48
2      0.72  0.71
3      1.00  0.98
4      0.81  0.68
5      0.62  1.18
6      1.32  1.36
7      1.24  0.78
8      0.99  1.64
9      0.90    NA
10     0.74    NA
11     0.88    NA
12     0.94    NA
13     1.16    NA
14     0.86    NA
15     0.85    NA
16     0.58    NA
17     0.57    NA
18     0.64    NA
19     0.98    NA
20     1.09    NA
21     0.92    NA
22     0.78    NA
23     1.24    NA
24     1.18    NA
> # Pick out from the Smoke vector those entries that are not NA
> Smok <- Smoke[!is.na(Smoke)]
> Smok
[1] 0.48 0.71 0.98 0.68 1.18 1.36 0.78 1.64
> NonSmoke
 [1] 0.97 0.72 1.00 0.81 0.62 1.32 1.24 0.99 0.90 0.74 0.88 0.94 1.16 0.86 0.85
[17] 0.58 0.57 0.64 0.98 1.09 0.92 0.78 1.24 1.18
> Smok-Smoker
```

```
 [1] 0 0 0 0 0 0 0
> NonSmoke-NonSmoker
 [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
> ######################### PLOTS TO CHECK NORMALITY #########################
> layout(matrix(1:4,ncol=2,byrow=TRUE))
> # Standardize
> sNonSmok <- (NonSmoke - mean(nonSmoke)) / sd(NonSmoke)
> sSmok <- (Smok - mean(Smok)) / sd(Smok)
> qqnorm(sNonSmok,ylab="Standardized NonSmoker Plasma",main="Non-Smoker Normal QQ Plot")
> abline(0,1,col=2)
> qqnorm(sSmok,ylab="Standardized Smoker Plasma",main="Smoker Normal QQ Plot")
> abline(0,1,col=4)
> # M3074PlasmaAcid1.pdf
>
> # Both data sets look pretty normal.
```


Non-Smoker Normal QQ Plot
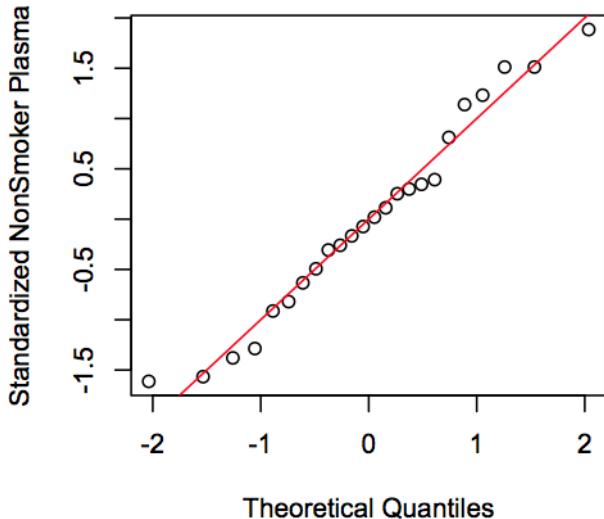
Smoker Normal QQ Plot
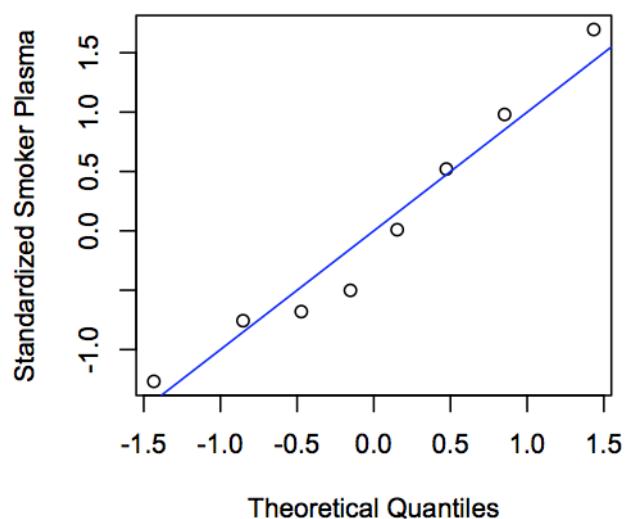
```
> #################### t-TEST BY HAND ####################
>
> v1 = var(Smok)
> v2 <- var(NonSmoke)
> n1 <- length(Smok)
> n2 <- length(NonSmoke)
> x1bar <- mean(Smok)
> x2bar <- mean(NonSmoke)
> nu <- (v1/n1 + v2/n2)^2/((v1/n1)^2/(n1-1)+(v2/n2)^2/(n2-1)); nu
[1] 8.44419
> t <-(x1bar-x2bar)/sqrt(v1/n1 + v2/n2);t
[1] 0.4161971
> pvalue <- 2*pt(t,nu,lower.tail=FALSE);pvalue
[1] 0.6876436
> alpha<- .05
> ta2 <- qt(alpha/2,nu,lower.tail=FALSE); ta2
[1] 2.285059
> se <- sqrt(v1/n1 + v2/n2); se
[1] 0.1451636
> # 2-sided 95% CI on mu1-mu2
> c(x1bar-x2bar-ta2*se,x1bar-x2bar+ta2*se)
[1] -0.2712908  0.3921241
> # ests of mu1, mu2
> c(x1bar,x2bar)
[1] 0.9762500 0.9158333
>
> # All same as from canned t.test()
```