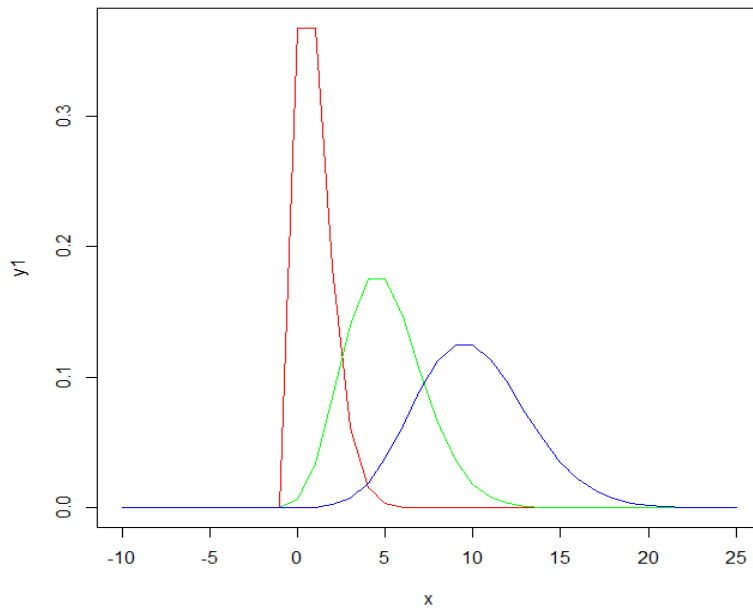**Josh Rhoades**
HW2
BIOSTAT

1.(a)
x <- seq(-10,25,1)
y1 <- dpois(x,1)
y2 <- dpois(x,5)
y3 <- dpois(x,10)

plot(x,y1,type="l",col="red")
lines(x,y2,col="green")
lines(x,y3,col="blue")

1.(b)
```
curve( dt(x,1), xlim=c(-3,3), ylim=c(0,.4), col='red', lwd=2 )
curve( dt(x,5), add=T, col='green', lwd=2 )
curve( dt(x,50), add=T, col='orange', lwd=2 )
curve( dnorm(x), add=T, lwd=3, lty=3 )
title(main="Student T distributions")
legend(par('usr')[2], par('usr')[4], xjust=1,
     c('df=1','df=5', 'df=50', 'Normal N(0,1)'),
     lwd=c(2,2,2,2),
     lty=c(1,1,1,3),
     col=c('red', 'green', 'orange', par("fg")))
```

Student T distributions
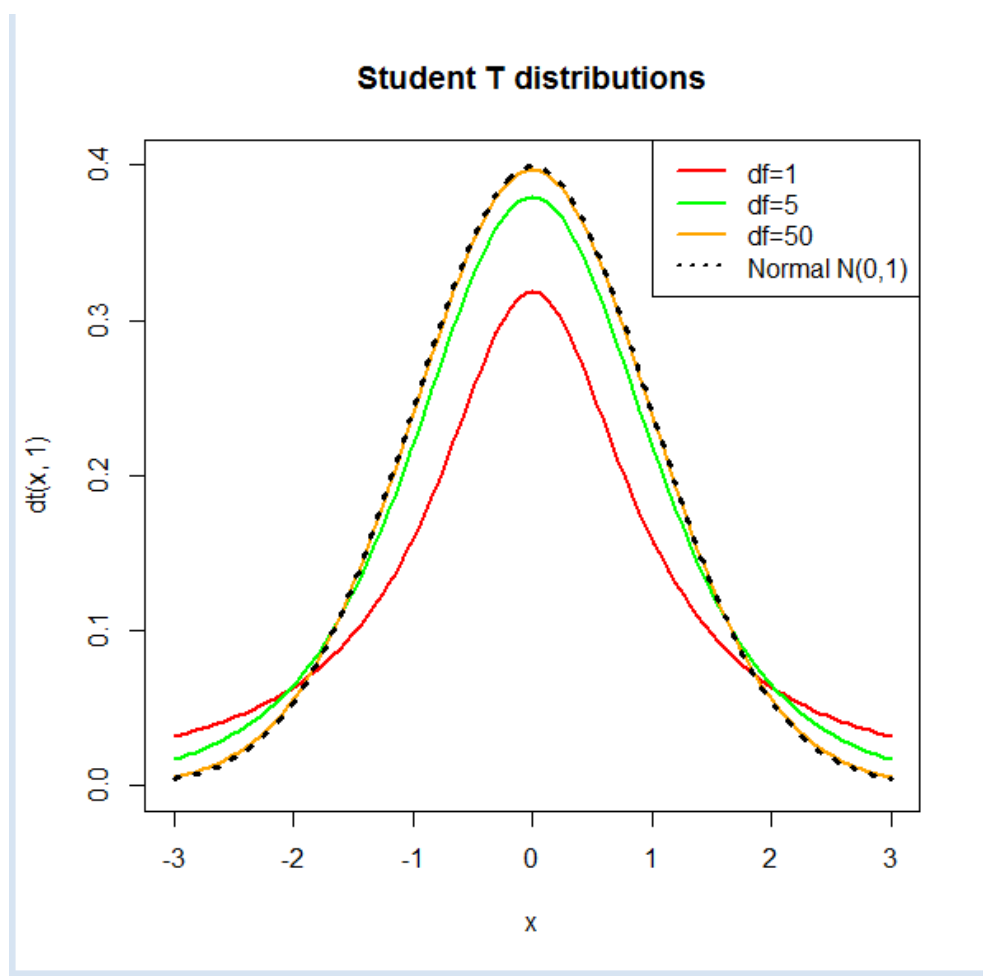
(taken and adapted from (http://zoonek2.free.fr/UNIX/48_R/07.html)

1.(c)
```
curve(dchisq(x,1), xlim=c(0,10), ylim=c(0,.6), col='red', lwd=3)
curve(dchisq(x,5), add=T, col='green', lwd=3)
curve(dchisq(x,15), add=T, col='blue', lwd=3)
abline(h=0,lty=3)
abline(v=0,lty=3)
legend(par('usr')[2], par('usr')[3], xjust=1,
     c('df=1', 'df=5', 'df=15',
     lwd=3,
     lty=1,
     col=c('red', 'green', 'blue')
     )
title(main='Chi^2 Distributions')
```
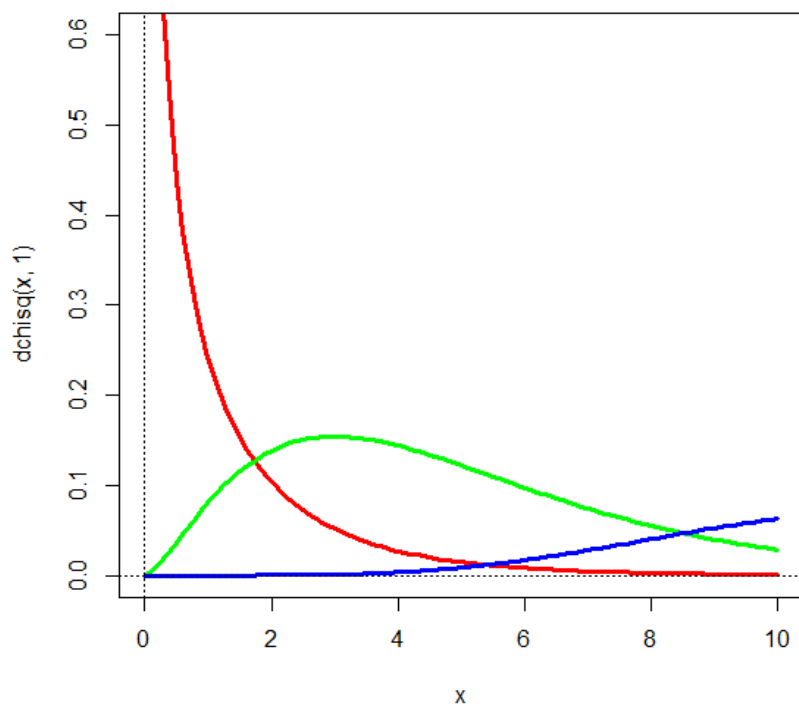
2.

2800
qqnorm(c(0.025,0.975)
mean=2800,sd=800/sqrt(78))
[1] 2622.462 2977.538

qnorm(c(0.025,0.975))
[1] -1.959964 1.959964

(2994-2800)/(800/sqrt(78))
[1] 2.141702
z=0.0324 power = .9838
1-.9838 = 0.0162

3100
qnorm(c(0.025,0.975),mean=3100,sd=800/sqrt(78))
[1] 2922.462 3277.538

(2994-3100)/(800/sqrt(78))
[1] -1.170208
z=0.2420 power =.8790 probability of type 2 error is 2.42%
qnorm(c(0.025,0.975))
[1] -1.959964 1.959964

3200
qnorm(c(0.025,0.975),mean=3200,sd=800/sqrt(78))
[1] 3022.462 3377.538
(2994-3200)/(800/sqrt(78))
[1] -2.274178
z=0.0232 power = .9884 probability of type 2 error is 2.32%
qnorm(c(0.025,0.975))
[1] -1.959964 1.959964

3400
qnorm(c(0.025,0.975),mean=3400,sd=800/sqrt(78))
[1] 3222.462 3577.538
(2994-3400)/(800/sqrt(78))
[1] -4.482119
z is off chart A1
qnorm(c(0.025,0.975))
[1] -1.959964 1.959964

3500
 qnorm(c(0.025,0.975),mean=3500,sd=800/sqrt(78))
[1] 3322.462 3677.538
(2994-3500)/(800/sqrt(78))

[1] -5.586089
z is off chart A1
qnorm(c(0.025,0.975))
[1] -1.959964 1.959964

3600
qnorm(c(0.025,0.975),mean=3600,sd=800/sqrt(78))
[1] 3422.462 3777.538
(2994-3600)/(800/sqrt(78))
[1] -6.690059
z is off chart A1
qnorm(c(0.025,0.975))
[1] -1.959964 1.959964

3(a)
summary(days)
     days
 Min.   : 22.00
 1st Qu.: 61.75
 Median : 87.50
 Mean   : 99.29
 3rd Qu.:119.50
 Max.   :307.00

n=78
df=n-1=77
One Tailed
qt(0.95, 77)
[1] 1.664885

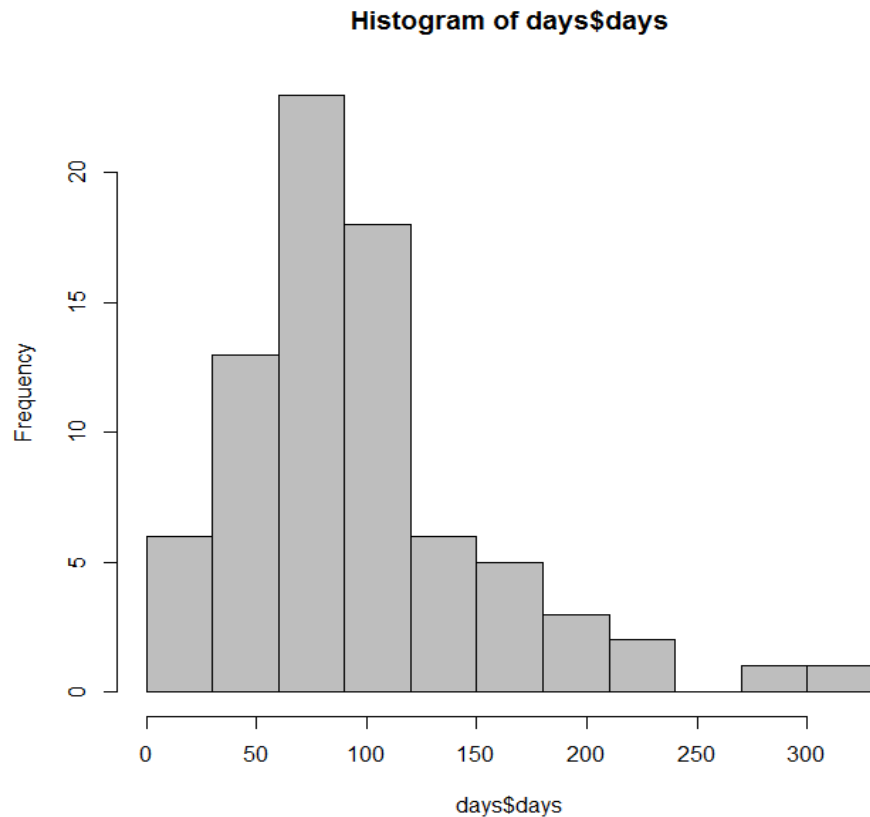99.29487+(1.664885)(57.66836 /8.83176086633)
110.165996267

Accept Ho

Two Tailed
> qt(0.975, 77)
[1] 1.991254

sd(days$days)
[1] 57.66836

99.29487+(1.991254)(57.66836 /8.83176086633)
112.297073554
99.29487-(1.991254)(57.66836 /8.83176086633)
86.2926664461
Acceptance region if from 86.2926664461 to 112.297073554

Accept Ho
3.b

Histogram of days$days

The distribution is skewed, thus not centered normally around the meal. Also there are several outliers to the right. Therefore the assumptions of a normal distribution are not met.

4.(a)
Null Hypothesis: The rate of Legionaires disease will be equal among the control populations and the treatment group from potential outbreak area.

leg <- c(65, 24, 52, 86, 120, 82, 399, 87, 139)
cont <- c(12, 10, 31, 6, 5, 5, 29, 9, 12)

t.test(leg, cont)

    Welch Two Sample t-test

data: leg and cont

t = 2.7951, df = 8.127, p-value = 0.02302
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 18.41227 189.36551
sample estimates:
mean of x mean of y
117.11111 13.22222

4(b)
There clearly is something going on, the results are significant. Perhaps there is another relationship that has yet to be explored. Does nickel effect immune system capability? Are high levels of nickel advantageous to *Legionella* growth?

5.a
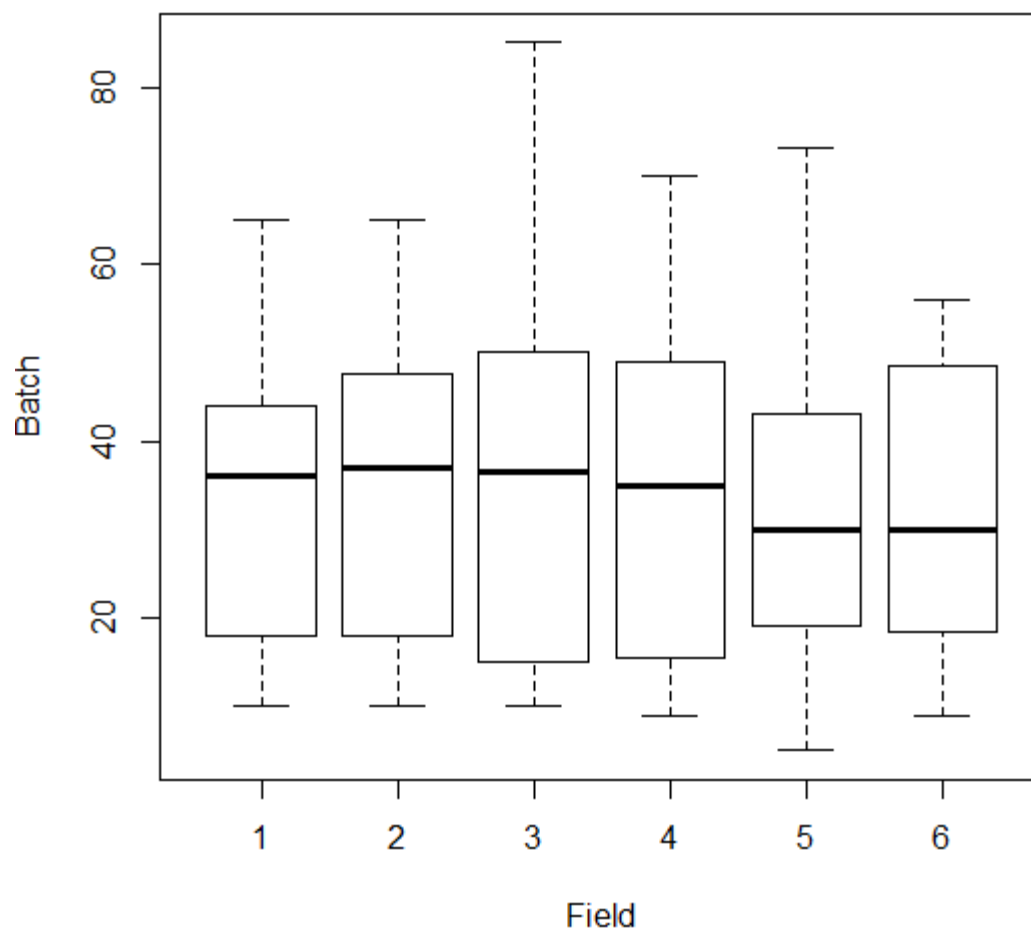
boxplot(Field~Batch, data=pipeline, xlab = 'Field', ylab = 'Batch')

boxplot(Lab~Batch, data=pipeline, xlab = 'Lab', ylab = 'Batch')
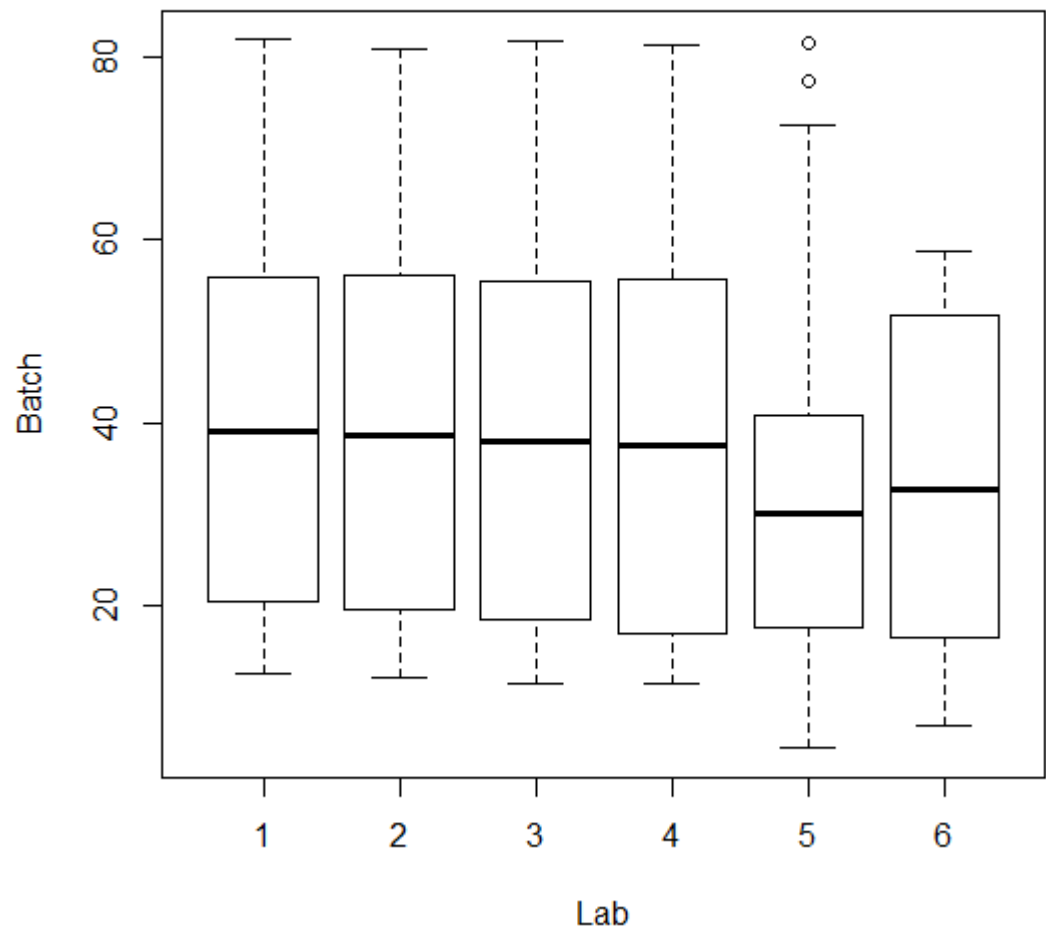
5,(b)
```
> mean(pipeline$V1)
[1] 33.57944
> mean(pipeline$V2)
[1] 39.09907
```

p values of batch vs batch t test of Field values

	1	2	3	4	5	6
1x						
2	0.8721x					
3	0.9397	0.934x				
4	0.9415	0.9318	0.998x			
5	0.4221	0.3361	0.3878	0.3874x		
6	0.4976	0.4241	0.4658	0.4661	0.939x	

This shows that 1 2 3 and 4 are similar, 5 and 6 are similar, but these two groupings are different from each other. This corresponds with the Histograms above, the means of 5 and 6 are different from the rest.

5.c
I would use a two way t test.
 t.test(pipeline$V1, pipeline$V2, conf.level = 0.9)

       Welch Two Sample t-test

data: pipeline$V1 and pipeline$V2
t = -1.8776, df = 199.369, p-value = 0.06189
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 -10.3775674 -0.6616849
sample estimates:
mean of x mean of y
 33.57944 39.09907