

# Project 4 Solutions

STAT 401: Fall 2006

Due Wednesday, October 11

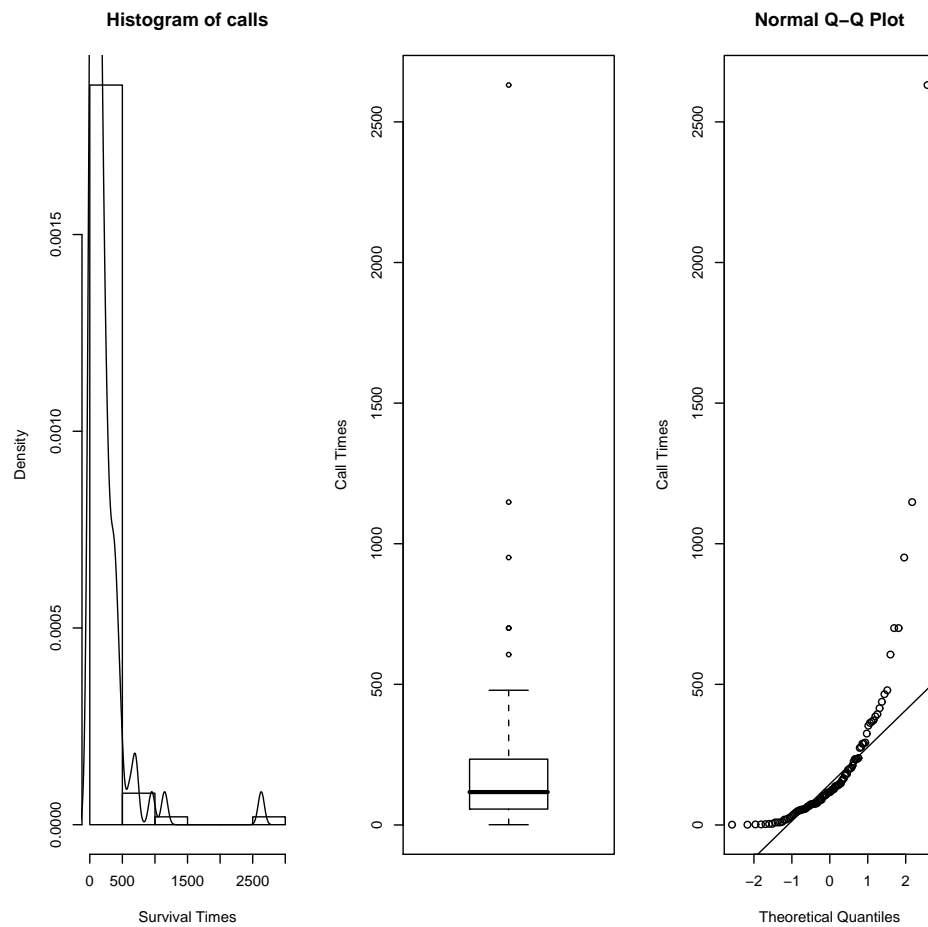
1. (a) (2 pts) To be a completely randomized design, the subjects must have been **randomly** assigned to the two treatment groups, either control or oxytocin whiff.  
(b) (2 pts) We can conclude that oxytocin **caused** the subjects in this experiment to transfer their entire bankrolls to a trustee more than twice as often as non-oxytocin whiffing subjects. Whether or not oxytocin would cause similar behavior in the general populace is dubious since the subjects are probably from a biased sample of volunteers, probably hungry college students.  
(c) (2 pts)
  - i. If none of these four adults whiffs oxytocin, the probability that exactly one adult decides to transfer their entire bankroll to a trustee is  $4(.21)(.79)^3 \approx .4142$ . This is because there are four ways to choose exactly one person,  $P(\text{transfer}|\text{no oxytocin}) = .21$  and  $P(\text{no transfer}|\text{no oxytocin}) = 1 - .21$ .
  - ii. If two of the adults whiff oxytocin and two do not, then the probability that exactly one of the adults transfers their entire bankroll to a trustee is  $2(.45)(.55)(.79)^2 + 2(.55)^2(.21)(.79) \approx 0.4093$ . This is because there are two possibilities: one of the oxytocin whiffing adults could be the one with to transfer, or one of the two other adults could be the one to transfer. The probability that it is one of the oxytocin whiffing adults who transfer is  $P(\text{one oxytocin whiffer transfers}) = 2(.45)(.55)(.79)^2$ . Similarly, we have that  $P(\text{one non - oxytocin whiffer transfers}) = 2(.55)^2(.21)(.79)$ .
2. (5 pts)
  - (a) Since the article states that “Twice as many mosquitoes gravitated toward kids carrying gametocytes as flew toward the other two groups,” then  $P(\text{mosquito flies towards a kid with gametocytes}) = \frac{2}{3}$ . Algebraically, if  $x$  is equal to number of mosquitos which fly towards the other groups, then one can solve  $2x = 100 - x$  and see that  $x = 33$ .
  - (b) When randomly selecting two of the mosquitos from out of 100 (without replacement), the probability that BOTH mosquitos go to the tent with kids carrying gametocytes is  $\frac{67}{100} \left(\frac{66}{99}\right) \approx .446$ . This is because  $P(\text{first mosquito flies to gametocyte tent}) = \frac{67}{100}$  and  $P(\text{second mosquito flies to gametocyte tent}|\text{first mosquito flies to gametocyte tent}) = \frac{66}{99}$ .
  - (c) By independence, the probability that both of the mosquitos fly towards the infected kid is  $\frac{2}{3} \left(\frac{2}{3}\right) = \frac{4}{9} = \bar{4}$ .
  - (d) Since the sample size in problem #2b of  $2 < .05(100)$ , then the conditional probability .4377 is very close to  $\bar{4}$ , the independent probability calculated in problem #2c.
  - (e)
    - i. The individuals are the mosquitos.
    - ii. The response is which of the three tents the mosquito flies towards.
3. (3 pts)
  - (a)  $P(X \leq 14.8) = P(Z \leq \frac{14.8-15}{.1}) \approx 0.02275$ . See the Appendix for R-code.
  - (b)  $P(14.7 < X < 15.1) = P(\frac{14.7-15}{.1} < Z < \frac{15.1-15}{.1}) \approx 0.84$ . See the Appendix for R-code.
  - (c) The probability that one tank holds at most 15 gallons is  $P(X \leq 15) = \frac{1}{2}$ . Assuming independence, then the probability that two tanks hold at most 15 gallons is  $\frac{1}{2}^2 = \frac{1}{4}$ .
4. (2 pts)

- (a) The probability that the sample proportion of mosquitos who fly towards the infected tent is over 75% is  $P(p > .75) = P\left(Z > \frac{.75 - \frac{2}{3}}{\frac{\sqrt{2}}{30}}\right) \approx .03855$ . See Appendix for the R code.
- (b) The probability the sample proportion of mosquitos who fly towards the infected tent is exactly  $\frac{2}{3}$  is 0. This is because for any continuous variable, the probability of attaining any single value is always 0.

5. (12 pts, 2 pts for (d) and (j), 1 pt otherwise)

- (a) Figure 1 contains a density plot, a boxplot, and a normal probability plot of the service calls data.

**Figure 1: Checking for Normality**



- (b) Table 1 gives the sample mean, sample standard deviation, and the five number summary for the service calls data.

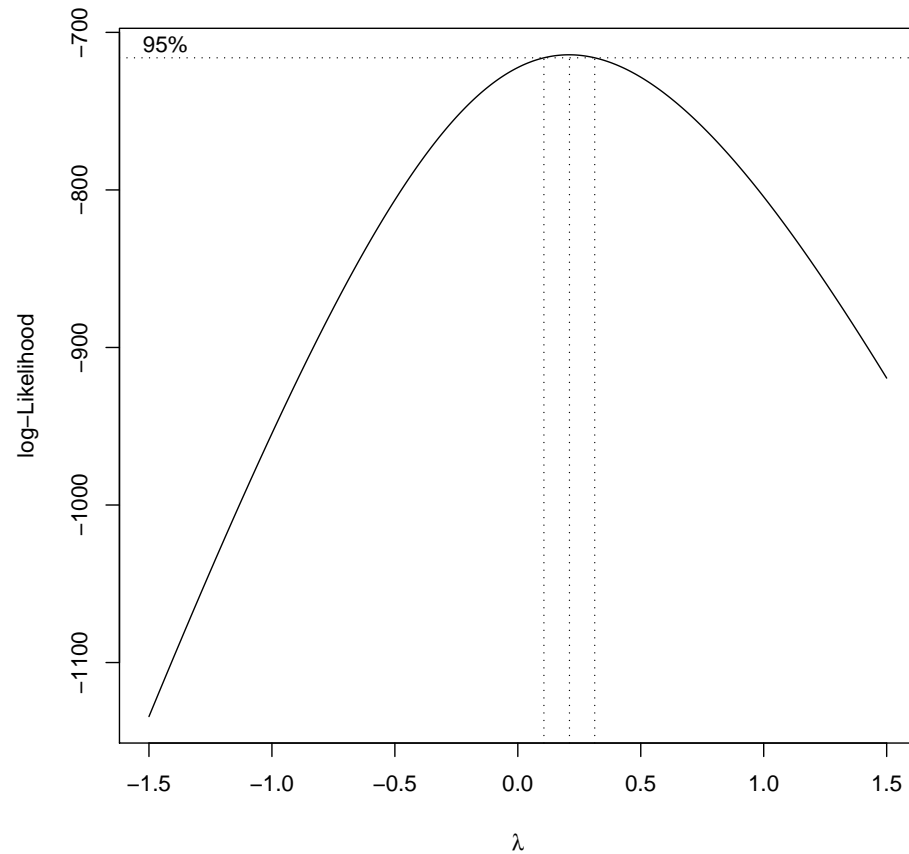
**Table 1: Statistics for Service Calls data**

$\bar{x}$	$s$	min	$Q_1$	$\bar{X}$	$Q_3$	max
200.79	313.03	1.00	56.75	117.00	234.00	2631.00

- (c) The correlation between the normal scores and the calls data shown in Figure 1 is 0.712 (see the Appendix for details). Since  $.712 < r_{\text{critical}} = .98$ , then the evidence suggests that there is a deviation from normality.

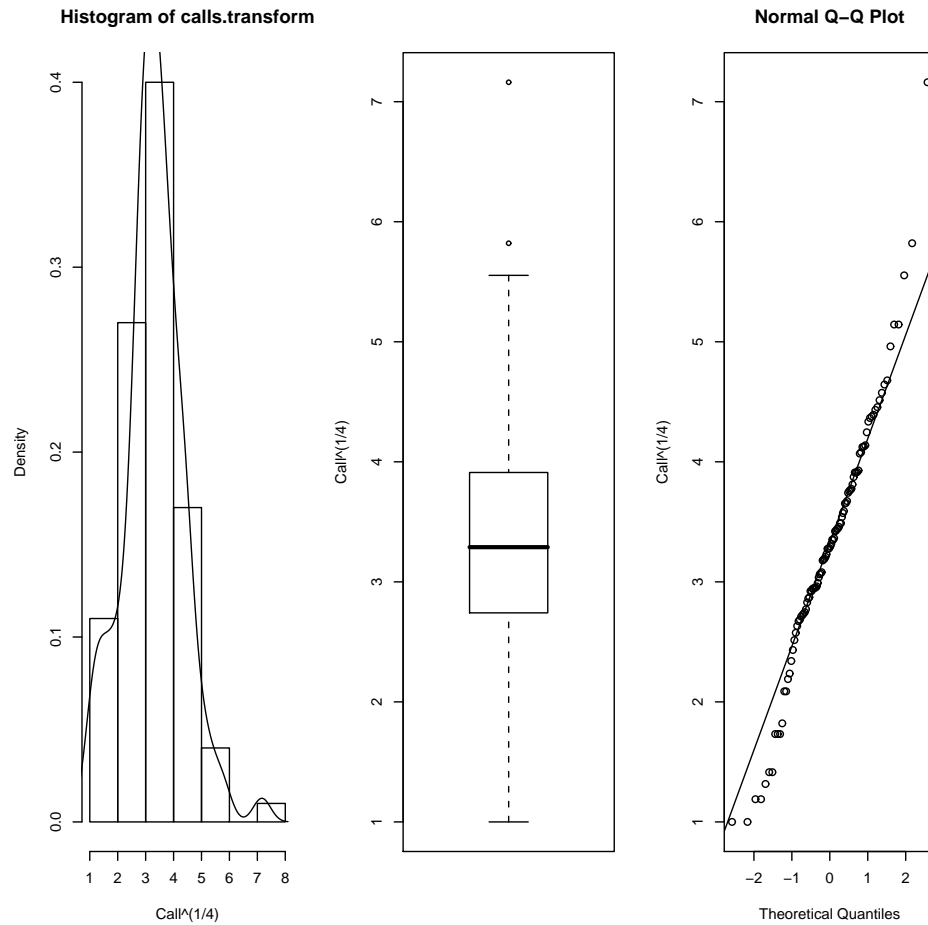
- (d) The severe right skew in the histogram and boxplot, the severe deviation from linearity in the normal probability plot, as well as the test of correlation all indicate that the service calls data are not normal.
- (e) Figure 2 shows the confidence interval for the appropriate  $\lambda$  for the Box-Cox transform.

**Figure 2: Confidence interval for  $\lambda$  for the Box-Cox transform**



- (f) From the confidence interval, it appears that the transform  $Y = X^{\frac{1}{4}}$  is appropriate for the service calls data.
- (g) Figure 3 contains a density plot, a boxplot, and a normal probability plot of the transformed calls data.

**Figure 3: Checking for normality of the transformed data**



(h) Table 2 gives the sample mean, sample standard deviation, and and five number summary for the transformed service calls data.

**Table 2: Statistics for  $(\text{Service Calls})^{1/4}$**

$\bar{x}$	$s$	min	$Q_1$	$\tilde{X}$	$Q_3$	max
3.3016	1.0573	1.000	2.745	3.289	3.911	7.162

- (i) The correlation between the normal scores and the transformed calls is 0.9865378 (see the Appendix for details). Since this value is larger than the critical  $r$  value of .98 then the evidence fails to suggest that the transformed data deviates from normality.
- (j) The histogram of the transformed data is less skewed than the original service calls data. The boxplot looks more symmetric and the normal probability plot looks better. Also considering, the test of correlation, we conclude that the evidence fails to suggest that the transformed data is not normal. It appears that our transformation to normality was effective.

## Appendix

```
> #PROBLEM 3
> pnorm(14.8,mean=15,sd=.1)
[1] 0.02275013
>
> pnorm(15.1,mean=15,sd=.1)-pnorm(14.7,mean=15,sd=.1)
[1] 0.8399948
>
>
> # PROBLEM 4
> pnorm(.75,mean=2/3,sd=sqrt(2)/30,lower.tail=F)
[1] 0.03854994
>
> # PROBLEM 5
> service=read.table("project4Calls.txt",header=T)
> attach(service)
>
> #5(a)
> par(mfrow=c(1,3))
> hist(calls,freq=F,xlab="Survival Times")
> lines(density(calls))
> boxplot(calls,ylab="Call Times")
> qqnorm(calls,ylab="Call Times")
> qqline(calls)
>
>
> #5(b)
> mean(calls);sd(calls);summary(calls)
[1] 200.79
[1] 313.0268
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.00  56.75  117.00  200.80  234.00 2631.00
>
>
> # 5(d)
> xy=qqnorm(calls)
> cor(xy$y,xy$x)
[1] 0.710987
>
>
> #5(e)
> library(MASS)
> par(mfrow=c(1,1))
> boxcox(calls~1,plotit=T,lambda=seq(-1.5,1.5,.01))
>
> #5(f)
> calls.transform=calls^(1/4)
>
>
> #5(g)
```

```

> par(mfrow=c(1,3))
> hist(calls.transform,freq=F,xlab="Call^(1/4)")
> lines(density(calls.transform))
> boxplot(calls.transform,ylab="Call^(1/4)")
> qqnorm(calls.transform,ylab="Call^(1/4)")
> qqline(calls.transform)
>
> # 5(h)
> mean(calls.transform);sd(calls.transform);summary(calls.transform)
[1] 3.301633
[1] 1.057268
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  2.745   3.289   3.302   3.911   7.162
>
>
> # 5(i)
> xy=qqnorm(calls.transform)
> cor(xy$y,xy$x)
[1] 0.9865378

```