Homework 05

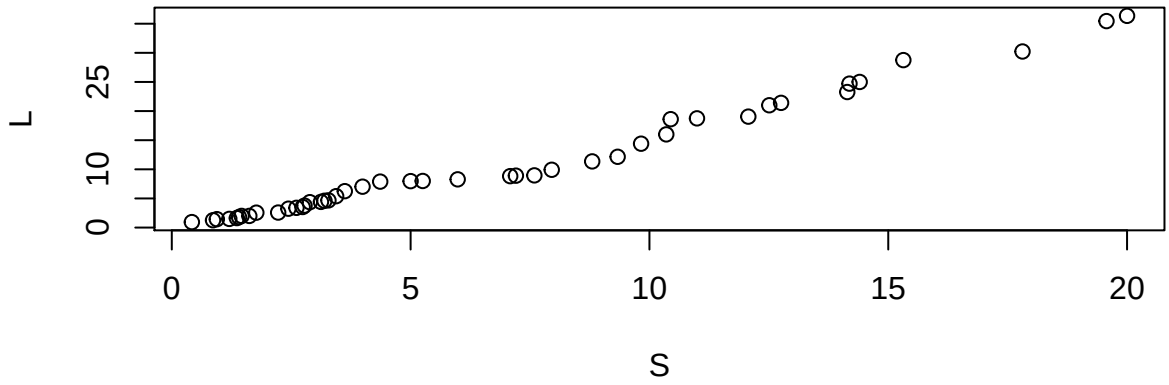
Joseph Blubaugh
jblubau1@tamu.edu

STAT 641-720

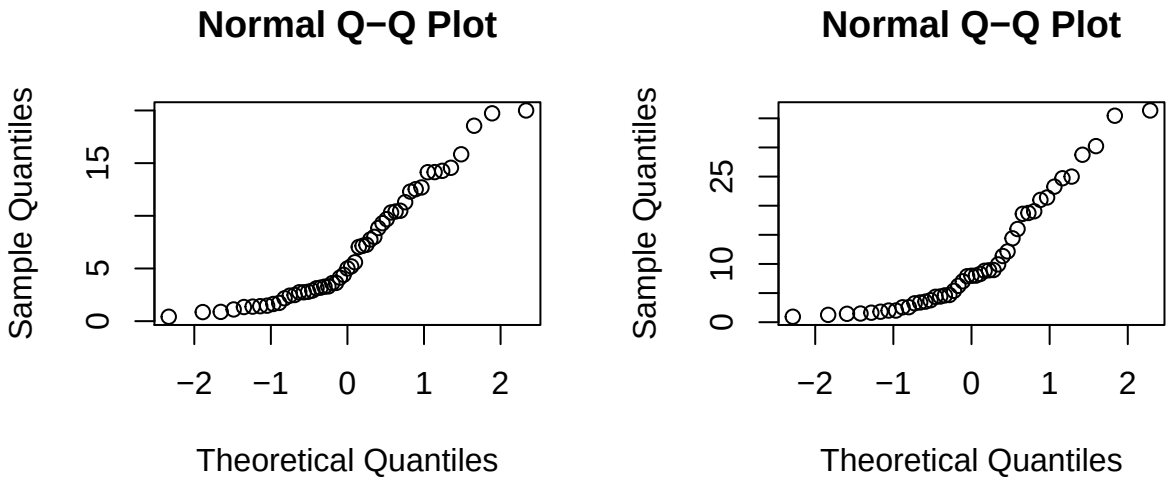
I. This distribution would have most of mass centralized and looks symmetrical except for some extreme values on both the left and right side.

II.

1.

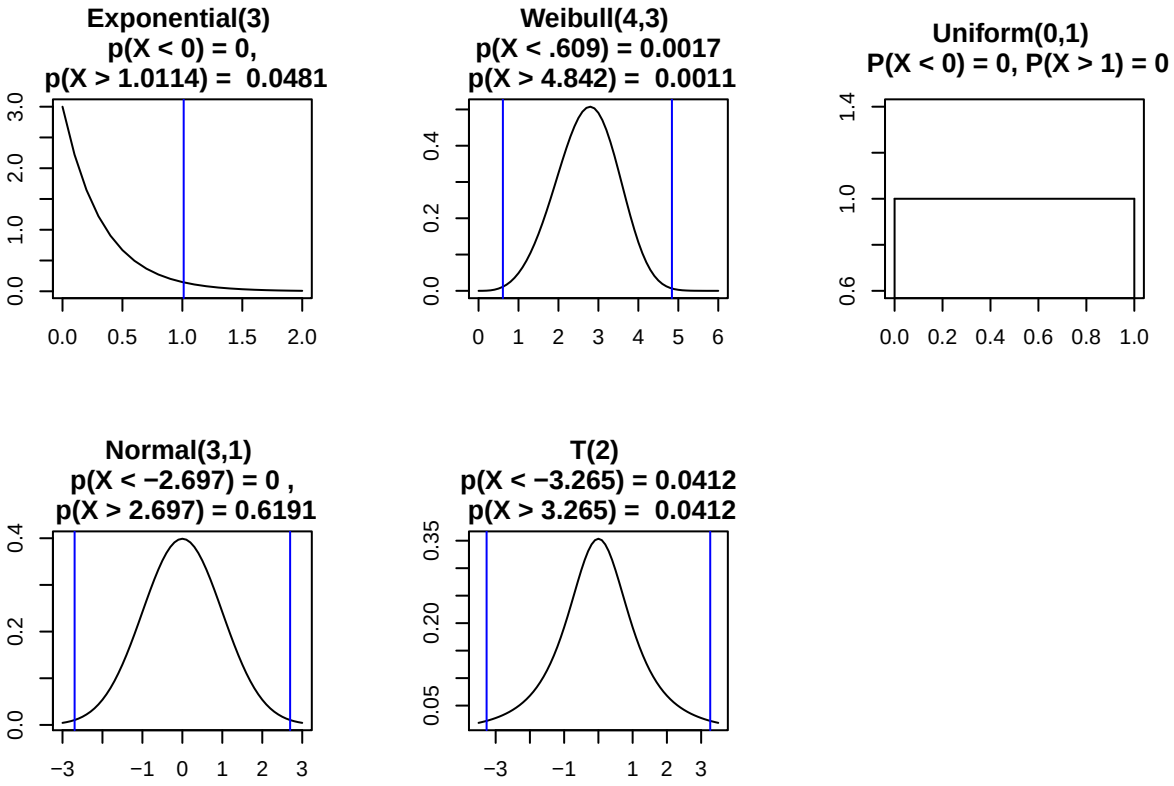


2.



3. Larger litters tend to have larger brain weight relative to body weight. There is also larger variance in relative brain weight for larger litters.

III.



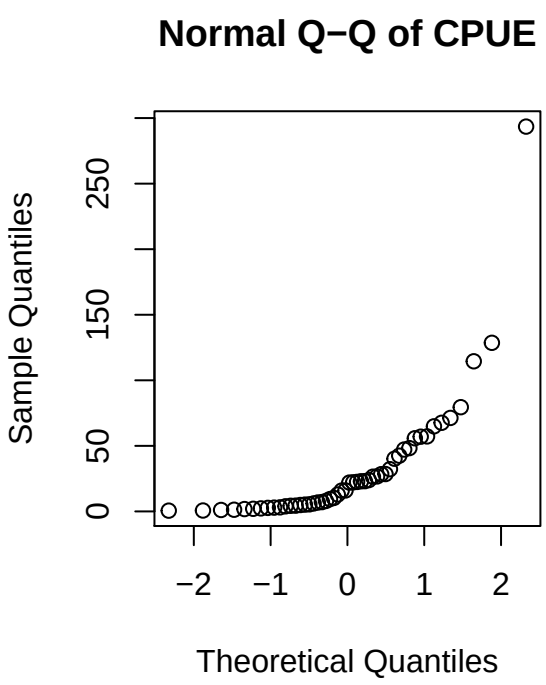
IV. Since all E_i must be larger than 1, defects 3 or more must all be combined. Also we need to use MLE as an estimator for the unknown probability of failure which in this case is .26. The Chi-Squared GOF results in a very low p-value so we can conclude that the model is a poor fit.

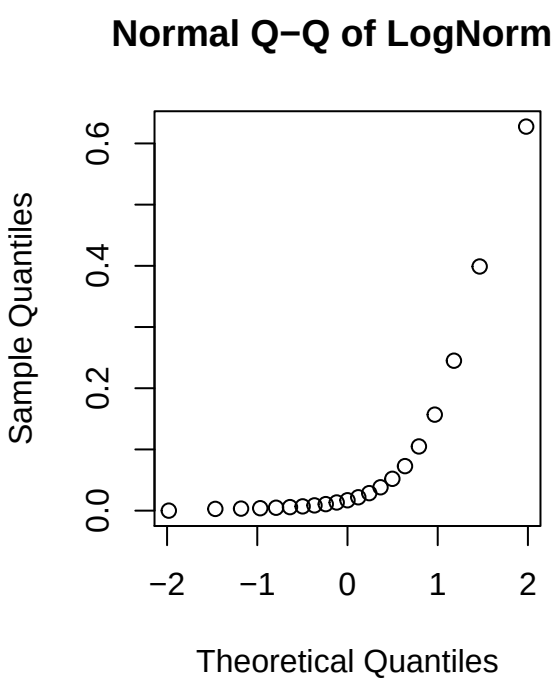
```
##       pi    Ei  Oi     Qi
## 1 0.4052 113.5 121 0.4956
## 2 0.4271 119.6 110 0.7706
## 3 0.1501  42.0  38 0.3810
## 4 0.0176   4.9  11 7.5939
```

```
## [1] "Chi-Squared GOF: 0.0263"
```

V.

1. QQ Plots of CPUE and Log-Normal are similarly shaped. The goodness of fit test tells us that the model fits well when using a Log-Normal model with paramters specified by the MLE.






```
##     meanlog      sdlog
##   2.5906275   1.4388286
##  (0.2034811) (0.1438829)
```

```
##       pi   Ei Oi     Qi
## 1 0.4179 20.9 21 0.0005
## 2 0.2928 14.6 14 0.0247
## 3 0.1620  8.1 10 0.4457
## 4 0.1088  5.4  5 0.0296
```

```
## [1] "Log-Normal GOF: 0.919"
```

2. The best estimate for θ is .073. Both $\theta = log(x)$ and $\theta = x^{.073}$ pass the shapiro test for normality, and $\theta = .073$ performs better, but log is probabily the better choice because its easier to interpret.

```
##
##  Shapiro-Wilk normality test
##
## data:  x1
## W = 0.9815, p-value = 0.6166
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x2
## W = 0.9773, p-value = 0.4451
```

VI. Plot1: Mixture of .90 Normal(10, 1) & .010 Normal(30, 3)
Plot2: Uniform(0, .7)
Plot3: Gamma(1.2, 25)
Plot4: Exponential(80)

VII.

1. (E): AD does a better job a detecting departures in the tails of the distribution

2. (A): AD is best because it looks at the differences in the entire model and has greater sensitivity

3. (D): AD Distribution free models are completely specified so their parameters do not have to be estimated

4. (E): Chi-Square test requires transforming continuous distributions into discrete distributions and has the some problem as relative frequency histograms

5. (C): Dealing with count data in a distrete distribution so GOF would work

6. (D):