Week 5 Lab

Donald Elrod

April 5, 2017

```
set.seed(111396)
```

Question 1

```
f = c(157, 69, 35, 17, 1, 1)
x = 0:5
```

Question 2

```
sample_mean = sum(f*x)/sum(f)
sample_mean
```

```
## [1] 0.7107143
```

```
sample_var = sum(f*x*x)/sum(f)
sample_var
```

```
## [1] 1.439286
```

Question 3

```
p.hat = sample_mean/5
p.hat
```

```
## [1] 0.1421429
```

Question 4

```
var_fit = 5*p.hat*(1-p.hat)
var_fit
```

```
## [1] 0.6096913
```

The fitted variance is much lower than the sample variance, and therefore the binomial model is not a good fit for the data

Question 5

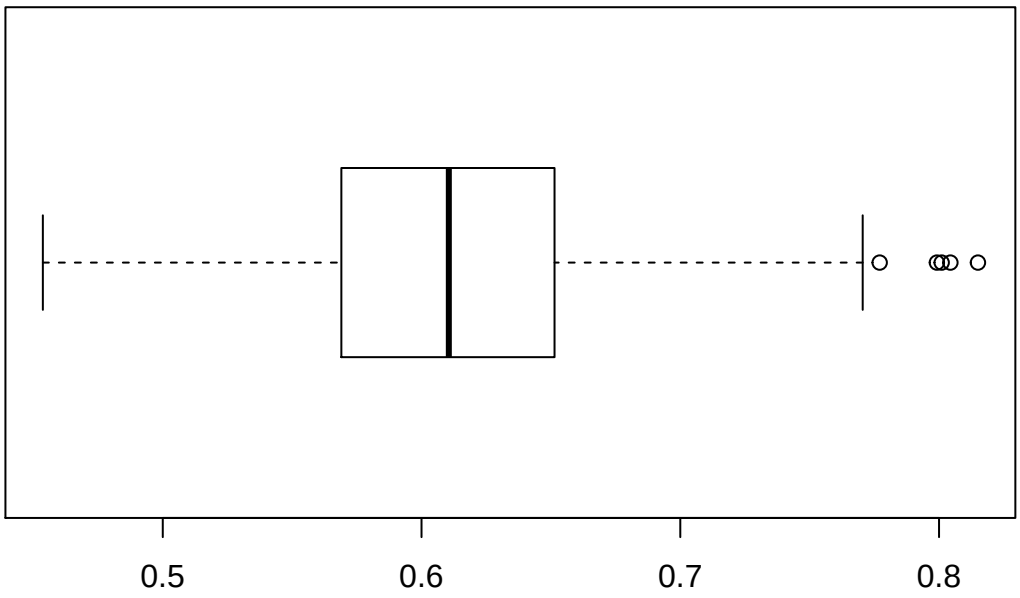
```
sim.vars = 0


for (i in 1:1000) {
  samp = rbinom(280, 5, p.hat)
  sim.vars[i] = var(samp)
}

summary(sim.vars)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.4536  0.5691  0.6105  0.6111  0.6513  0.8150
```

```
boxplot(sim.vars, horizontal = T)
```



The range of the variances vary greatly, with the lowest being .4436 and the highest being .7956 which is almost double. Therefore the polynomial model is not a good fit for this data

Question 6

```
E.p = sample_mean
E.p
```

```
## [1] 0.7107143
```

```
Var.p = (sample_var - E.p*(1-(E.p/5)))/20
Var.p
```

```
## [1] 0.04147972
```

Question 7

```
z1 = (E.p + sqrt(Var.p)) * 1000
z1
```

```
## [1] 914.38
```

```
z2 = (E.p - sqrt(Var.p)) * 1000
z2
```

```
## [1] 507.0486
```

Question 8

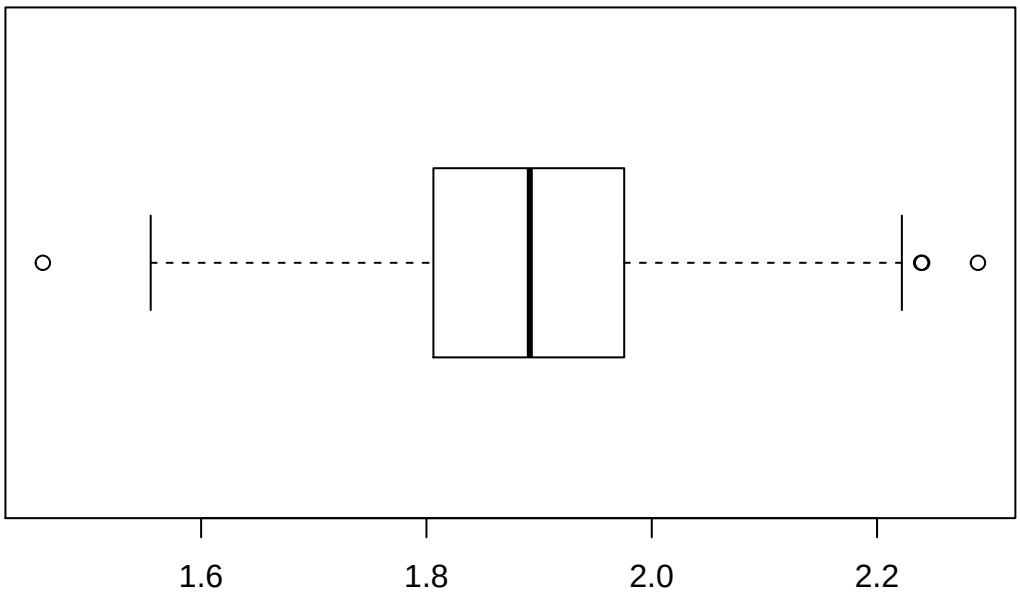
```
Z = sample(x=c(z1,z2), size=280, replace=T)
random.p = Z/1000

final.vars = 0

for (i in 1:1000) {
  samp_8 = rbinom(280, 5, random.p)
  final.vars[i] = var(samp_8)
}
summary(final.vars)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.459   1.806   1.892   1.894   1.975   2.290
```

```
boxplot(final.vars, horizontal = T)
```



There is a larger variance in the two-point distribution than there is in the random distribution, which makes sense as the values are always between one of two numbers