STAT 461 Lab 3 - The ANOVA MODEL

1.Simulating from the ANOVA Model

Recall from lectures that the one-way ANOVA model with effects coding is written

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, \quad i = 1, 2, \ldots, v \quad t = 1, 2, \ldots, r_i$$

$$\epsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

To better understand this model, we will simulate experimental data from the model by

1. Randomly assigning treatments to all experimental units.
2. Specifying all parameters.

   Note that neither μ or τ_i are estimable all by themselves, so it is sufficient to specify $\mu + \tau_i$ for $i = 1, 2, \ldots, v$ and to specify σ^2.
3. Simulating $\epsilon_{it} \sim N(\mu, \sigma^2)$
4. Adding $\mu + \tau_i + \epsilon_{it}$ to obtain each response Y_{it}.

For an example, consider the case in which we have two treatments: A and B, and have $n = 20$ experimental units, which will be evenly divided with $r_A = 10$ receiving treatment A and $r_B = 10$ receiving treatment B. Further, we assume that

$$\mu = 2$$

$$\tau_A = 1$$

$$\tau_B = -4$$

$$\sigma^2 = 0.6$$

Then the response of any experimental unit given treatment "A" is normally distributed with mean $\mu + \tau_A = 3$ and variance 0.6.

$$Y_{At} \sim N(3, 0.6)$$

Similarly, the response of any experimental unit given treatment "B" is normally distributed with mean $\mu + \tau_B = -2$ and variance 0.6.

$$Y_{Bt} \sim N(-2, 0.6)$$

The following R code will simulate response data $\{Y_{it}\}$ from this setup.

We first randomly assign the treatments to the experimental units.

```r
## Make an unsorted list of the treatment labels, one for each experimental unit
treatments.not.random=c(rep("A",10), rep("B",10))
## randomly re-order the treatments
treatment=sample(treatments.not.random)
## match this order to the list of experimental units.
n=length(treatment)
Exp.Unit=1:n
## make a table that shows the random treatment assignment
CRD.table=data.frame(Exp.Unit,treatment,row.names=NULL)
CRD.table
```

```
##    Exp.Unit treatment
## 1         1         A
## 2         2         B
## 3         3         A
## 4         4         A
## 5         5         B
## 6         6         A
## 7         7         A
## 8         8         B
## 9         9         A
## 10       10         B
## 11       11         A
## 12       12         B
## 13       13         A
## 14       14         A
## 15       15         B
## 16       16         B
## 17       17         B
## 18       18         B
## 19       19         B
## 20       20         A
```

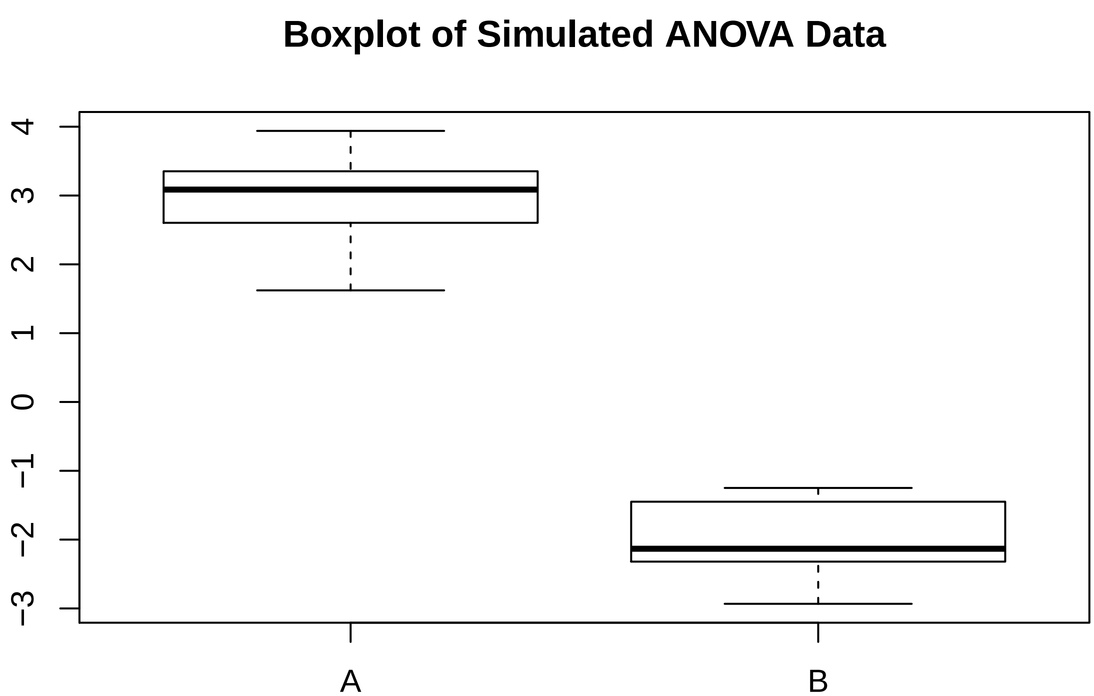
We then simulate from the ANOVA model by creating a "mean" which is different for each treatment. So if the experimental unit received treatment A, then the mean is $\mu + \tau_A = 3$, and if the experimental unit received treatment B, then the mean is $\mu + \tau_B = -2$. This mean is then added to a normally-distributed random variable with zero mean and variance equal to $\sigma^2 = 0.6$. Note that all experimental units share this variance, regardless of treatment.

```
## Set parameters
mu=2
tau.A=1
tau.B=-4
s2=0.6
## Make a vector of means
n=length(treatment)
means=rep(NA,n)
## assign different means for each treatment
means[treatment=="A"]=mu+tau.A
means[treatment=="B"]=mu+tau.B
## Simulate Y_it = mu + tau_i + e_it, e_it~N(0,s2)
Y.sim=means+rnorm(n,mean=0,sd=sqrt(s2))
## Display results
SimData=data.frame(Exp.Unit,treatment,Y.sim)
SimData
```

```
##    Exp.Unit treatment     Y.sim
## 1         1         A  2.982243
## 2         2         B -1.599507
## 3         3         A  2.713517
## 4         4         A  3.352519
## 5         5         B -2.199526
## 6         6         A  2.377362
## 7         7         A  3.190226
## 8         8         B -2.066091
## 9         9         A  3.213393
```

```
## 10        10         B -2.319861
## 11        11         A  1.621918
## 12        12         B -1.393993
## 13        13         A  3.881251
## 14        14         A  3.939147
## 15        15         B -1.448991
## 16        16         B -2.252716
## 17        17         B -2.932835
## 18        18         B -2.756484
## 19        19         B -1.249085
## 20        20         A  2.603414
```

```
## boxplot of simulations
boxplot(Y.sim~treatment,main="Boxplot of Simulated ANOVA Data")
```



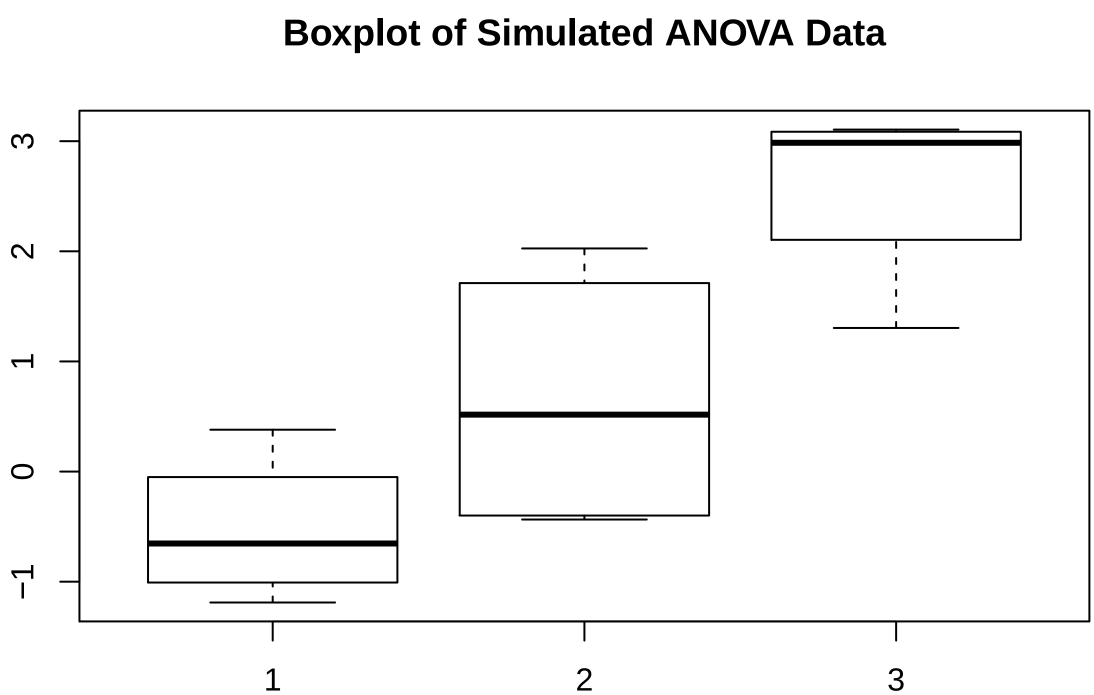
In this simulation code, you will need to adjust (1) the lines defining the treatment means and (2) the error variance. For example, the following code will simulate responses from the ANOVA model if there are 12 experimental units, 4 of which will receive each of the three treatments: "1", "2", and "3", with $\mu = -1$, $\tau_1 = 1$, $\tau_2 = 2$, $\tau_3 = 3$ and error variance $\sigma^2 = 2$.

```
## Set parameters
mu=-1
tau.1=1
tau.2=2
tau.3=3
s2=2
## set treatments
treatment=c(rep(1,4),rep(2,4),rep(3,4))
n=length(treatment)
## Make a vector of means
n=length(treatment)
means=rep(NA,n)
## assign different means for each treatment
```

```
means[treatment==1]=mu+tau.1
means[treatment==2]=mu+tau.2
means[treatment==3]=mu+tau.3
## Simulate Y_it = mu + tau_i + e_it, e_it~N(0,s2)
Y.sim=means+rnorm(n,mean=0,sd=sqrt(s2))
## Display results
SimData=data.frame(Exp.Unit=1:n,treatment,Y.sim)
SimData
```

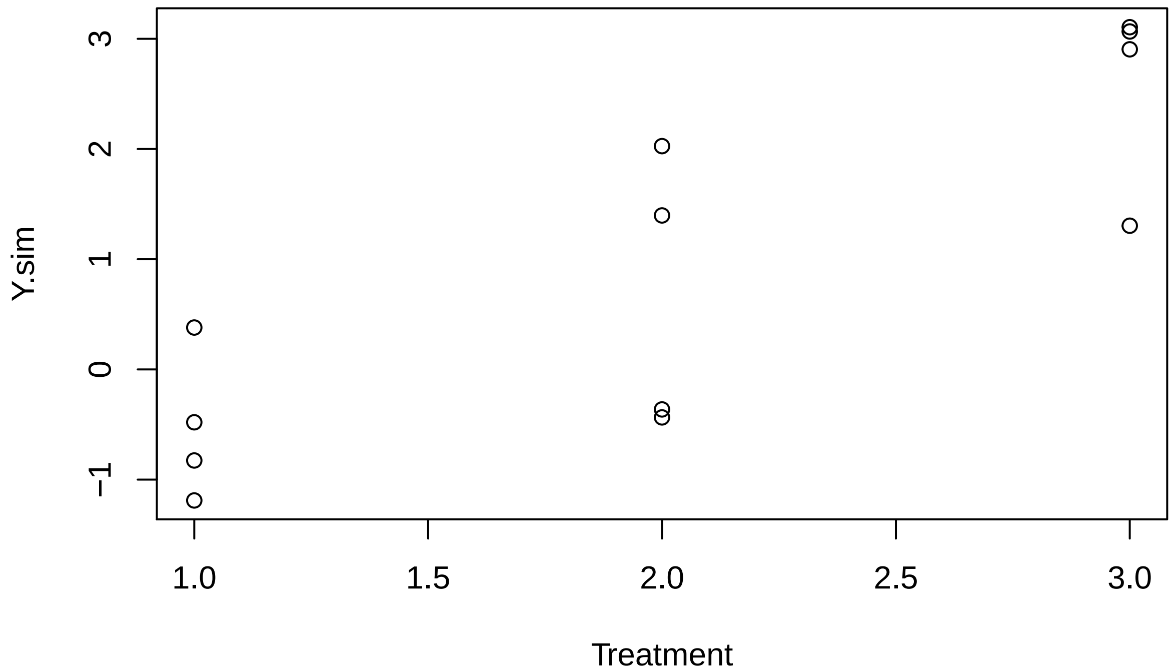
```
##    Exp.Unit treatment      Y.sim
## 1         1         1  0.3800987
## 2         2         1 -1.1889369
## 3         3         1 -0.8263061
## 4         4         1 -0.4797851
## 5         5         2 -0.4355677
## 6         6         2 -0.3631684
## 7         7         2  2.0258682
## 8         8         2  1.3971870
## 9         9         3  1.3038072
## 10       10         3  3.1051620
## 11       11         3  3.0662718
## 12       12         3  2.9044630
```

```
## boxplot of simulations
boxplot(Y.sim~treatment,main="Boxplot of Simulated ANOVA Data")
```



If we have numeric names for the treatments like 1,2,3. . .), then we can plot the resulting values using the plot command:

```
plot(x=treatment,y=Y.sim,xlab="Treatment",ylab="Y.sim")
```



Y.sim
3
2
1
0
−1
1.0
1.5
2.0
2.5
3.0
Treatment

Homework Assignment (Use R for Q1-Q3)

1. Use R to randomly assign 10 experimental units to each of three treatments (1, 2, and 3). Then simulate responses for the 30 experimental units satisfying the one-way ANOVA model:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, \quad i = 1, 2, \ldots, v \quad t = 1, 2, \ldots, r_i$$

$$\epsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

   with $\mu = 4.7$, $\sigma^2 = 4$, and treatment effects $\tau_1 = -3$, $\tau_2 = 5$, and $\tau_3 = -2$. Your solution should include your R code and a plot of the simulated values.

2. Consider the situation in Problem 1. The experimenter wants to consider a reduced model where $\tau_1 = \tau_2 = \tau_3 = 0$. Simulate responses for the 30 experimental units satisfying this reduced model. Compare boxplots of simulated responses under this reduced model with boxplots of simulated responses under the full model described in Problem 1 (where there are differences in the treatment effects).

3. Now explore what happens to data simulated from the model in Problem 1 when the error variance increases. Try multiple values for σ^2 and find a value of σ^2 for which you cannot see any noticable difference in the boxplots of response values from the three treatments.

4. Under the model in Problem 1, what is the distribution of Y_{23}, the response from the 3rd experimental unit to receive treatment 2?

5. Under the model in Problem 1, what is the distribution of

$$\bar{Y}_{2\cdot} = \frac{1}{r_2} \sum_{t=1}^{r_2} Y_{2t}?$$

6. Under the model in Problem 1, what is the distribution of the difference between an experimental unit receiving treatment 1 and an experimental unit receiving treatment 2?