STAT461 HW3

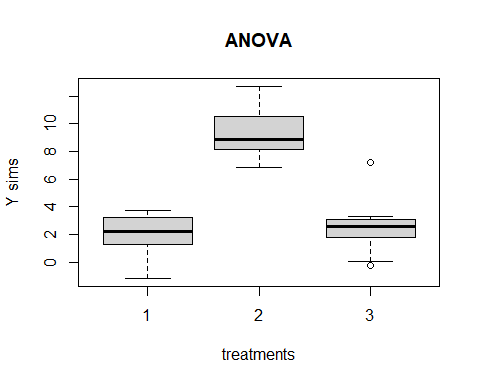
Xiangyu Ren

9/18/2020

#### Problem 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:

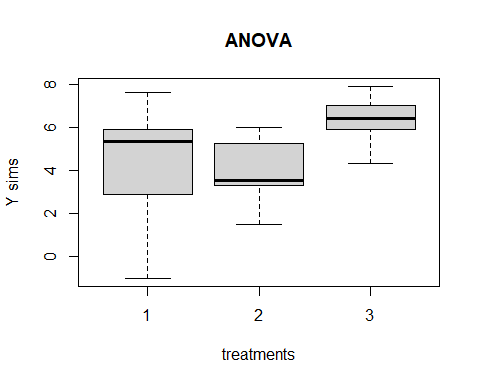
#### with , and treatment effects . Your solution should include your R code and a plot of the simulated values.

treatment = c(rep(1, 10), rep(2, 10), rep(3, 10))  
mu = 4.7  
sig\_s = 4  
t1 = -3  
t2 = 5  
t3 = -2  
n = length(treatment)  
exp.units = 1:n  
treatment.random = sample(treatment)  
mean1 = mu + t1  
mean2 = mu + t2  
mean3 = mu + t3  
l1 = treatment.random == 1  
l2 = treatment.random == 2  
l3 = treatment.random == 3  
means = rep(NA, n)  
means[l1] = mean1  
means[l2] = mean2  
means[l3] = mean3  
ep = rnorm(n, mean = 0, sd = sqrt(sig\_s))  
Y = means + ep  
boxplot(Y~treatment.random, xlab = "treatments", ylab = "Y sims", main = "ANOVA")



#### Problem 2. Consider the situation in Problem 1. The experimenter wants to consider a reduced model where . 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).

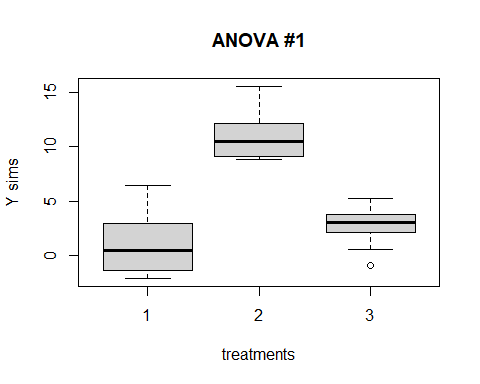
treatment = c(rep(1, 10), rep(2, 10), rep(3, 10))  
mu = 4.7  
sig\_s = 4  
t1 = 0  
t2 = 0  
t3 = 0  
n = length(treatment)  
exp.units = 1:n  
treatment.random = sample(treatment)  
mean1 = mu + t1  
mean2 = mu + t2  
mean3 = mu + t3  
l1 = treatment.random == 1  
l2 = treatment.random == 2  
l3 = treatment.random == 3  
means = rep(NA, n)  
means[l1] = mean1  
means[l2] = mean2  
means[l3] = mean3  
ep = rnorm(n, mean = 0, sd = sqrt(sig\_s))  
Y = means + ep  
boxplot(Y~treatment.random, xlab = "treatments", ylab = "Y sims", main = "ANOVA")



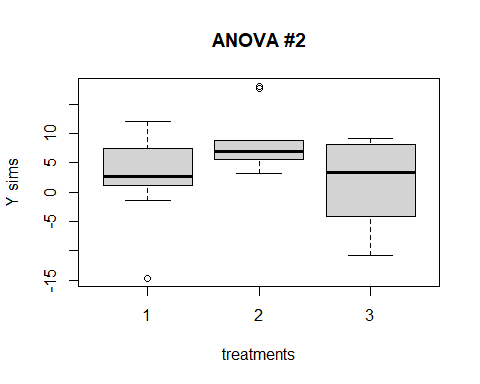
Comparing to problem 1, the means of different treatments in problem 2 are about the same. There is no treatment effect on each experiement.

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

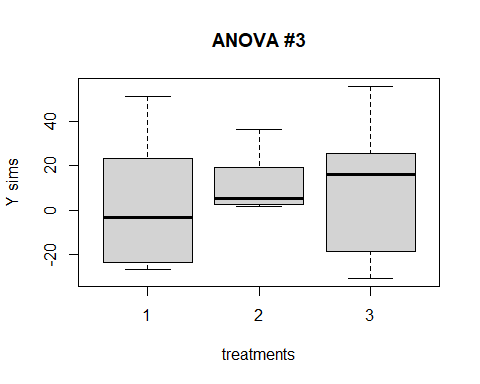
treatment = c(rep(1, 10), rep(2, 10), rep(3, 10))  
mu = 4.7  
t1 = -3  
t2 = 5  
t3 = -2  
n = length(treatment)  
exp.units = 1:n  
treatment.random = sample(treatment)  
mean1 = mu + t1  
mean2 = mu + t2  
mean3 = mu + t3  
l1 = treatment.random == 1  
l2 = treatment.random == 2  
l3 = treatment.random == 3  
means = rep(NA, n)  
means[l1] = mean1  
means[l2] = mean2  
means[l3] = mean3  
Y = means + rnorm(n, mean = 0, sd = sqrt(4))  
boxplot(Y~treatment.random, xlab = "treatments", ylab = "Y sims", main = "ANOVA #1")



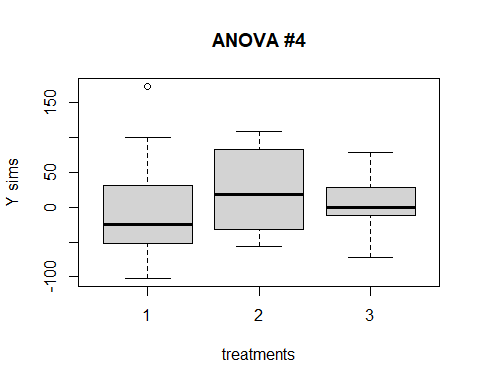
Y = means + rnorm(n, mean = 0, sd = sqrt(40))  
boxplot(Y~treatment.random, xlab = "treatments", ylab = "Y sims", main = "ANOVA #2")



Y = means + rnorm(n, mean = 0, sd = sqrt(400))  
boxplot(Y~treatment.random, xlab = "treatments", ylab = "Y sims", main = "ANOVA #3")



Y = means + rnorm(n, mean = 0, sd = sqrt(4000))  
boxplot(Y~treatment.random, xlab = "treatments", ylab = "Y sims", main = "ANOVA #4")



After setting , don’t have an obvious difference of Y from the anova boxplot, comparing to other three boxplots. So, when doing an experiment, need to control irrelevant variables in order to reduce errors or residuals.

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

Under the model in problem 1, the distribution of is normal distribution.

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

Under the model in problem 1, the distribution is normal distribution.

#### Problem 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?

Under the model in problem 1, the distribution of the difference is normal distribution.