

# Lab 5

Kristina Arevalo

Fri Mar 19 2021

## Contents

<b>Problem 1</b>	<b>2</b>
simulated data . . . . .	2
ggplot . . . . .	2
<b>Problem 2</b>	<b>3</b>
simulated data . . . . .	3
anova . . . . .	4
ggplot . . . . .	4
<b>Bonus Question</b>	<b>5</b>

## Problem 1

In the context of making decisions about whether or not observed data is consistent or inconsistent with a Null hypothesis, it is possible to make errors. The questions below ask you to create simulated data containing patterns that could lead to correct and incorrect decisions about the role of the null hypothesis.

Consider a design with 3 groups, and 10 people per group. Assume that the dependent variable is assumed to be normally distributed, and use unit normal distributions with mean = 0, and sd = 1 in your simulations.

Create simulated data for the above design that could be produced by the null hypothesis, and that results in a  $F$  value that is smaller than the critical value for  $F$  in this design (assume  $\alpha = .05$ ). Report the ANOVA, and show a ggplot of the means in the simulated data. Furthermore, display the individual data points on top of the means. Would you reject the null hypothesis in this situation, and would you be incorrect or correct in rejecting the null? (3 points)

### simulated data

```
levels <- 3
n_per_level <- 10
crit_F <- qf(.95,2,27)

for(i in 1:1000){
  sim_data <- tibble(participants = 1:(levels*n_per_level),
                     IV = as.factor(rep(1:levels, each = n_per_level)),
                     DV = rnorm(levels*n_per_level,0, 1))

  my.aov <- aov(DV~IV, data = sim_data) %>% summary()
  simulated_F <- my.aov[[1]]$`F value`[1]

  if(simulated_F < crit_F) break
}
```

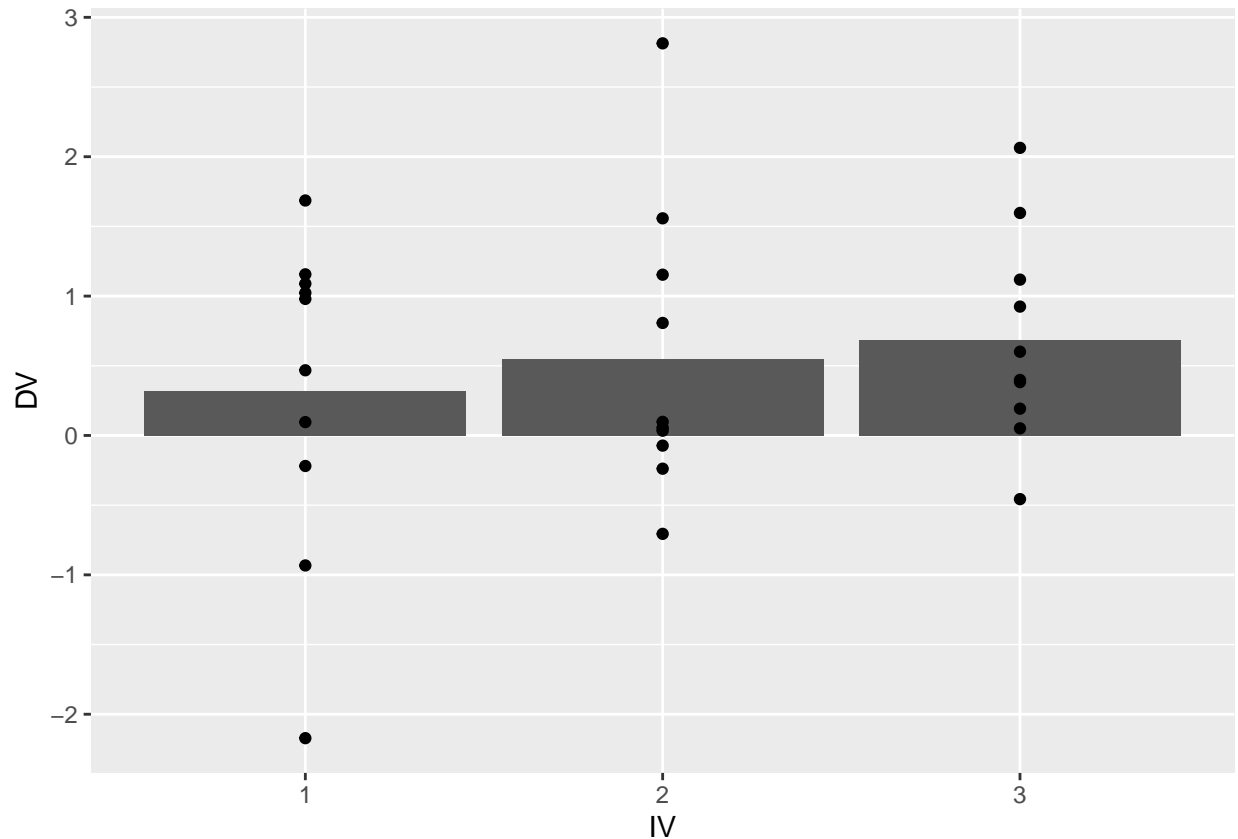
```
#anova
```

```
my.aov
```

```
##          Df Sum Sq Mean Sq F value Pr(>F)
## IV          2  0.699  0.3494   0.347  0.71
## Residuals  27 27.202  1.0075
```

### ggplot

```
ggplot(sim_data, aes(x = IV, y= DV))+
  geom_bar(stat= "summary", fun = "mean")+
  geom_point()
```



I would fail to reject the null and I believe I would be correct in doing so because the p-value of obtaining this F-value (or smaller) is not less than .05 and we assumed an alpha of .05. These results are not significant.

## Problem 2

Create simulated data for the above design that could be produced by the null hypothesis, and that results in a value that is larger than the critical value for in this design (assume alpha = .05). Report the ANOVA, and show a ggplot of the means in the simulated data. Furthermore, display the individual data points on top of the means. Would you reject the null hypothesis in this situation, and would you be incorrect or correct in rejecting the null? (3 points)

### simulated data

```
levels <- 3
n_per_level <- 10
crit_F <- qf(.95,2,27)

for(i in 1:1000){
  sim_data <- tibble(participants = 1:(levels*n_per_level),
                     IV = as.factor(rep(1:levels, each = n_per_level)),
                     DV = rnorm(levels*n_per_level,0, 1))

  my.aov <- aov(DV~IV, data = sim_data) %>% summary()
```

```
simulated_F <- my.aov[[1]]$`F value`[1]
```

```
if(simulated_F > crit_F) break  
}
```

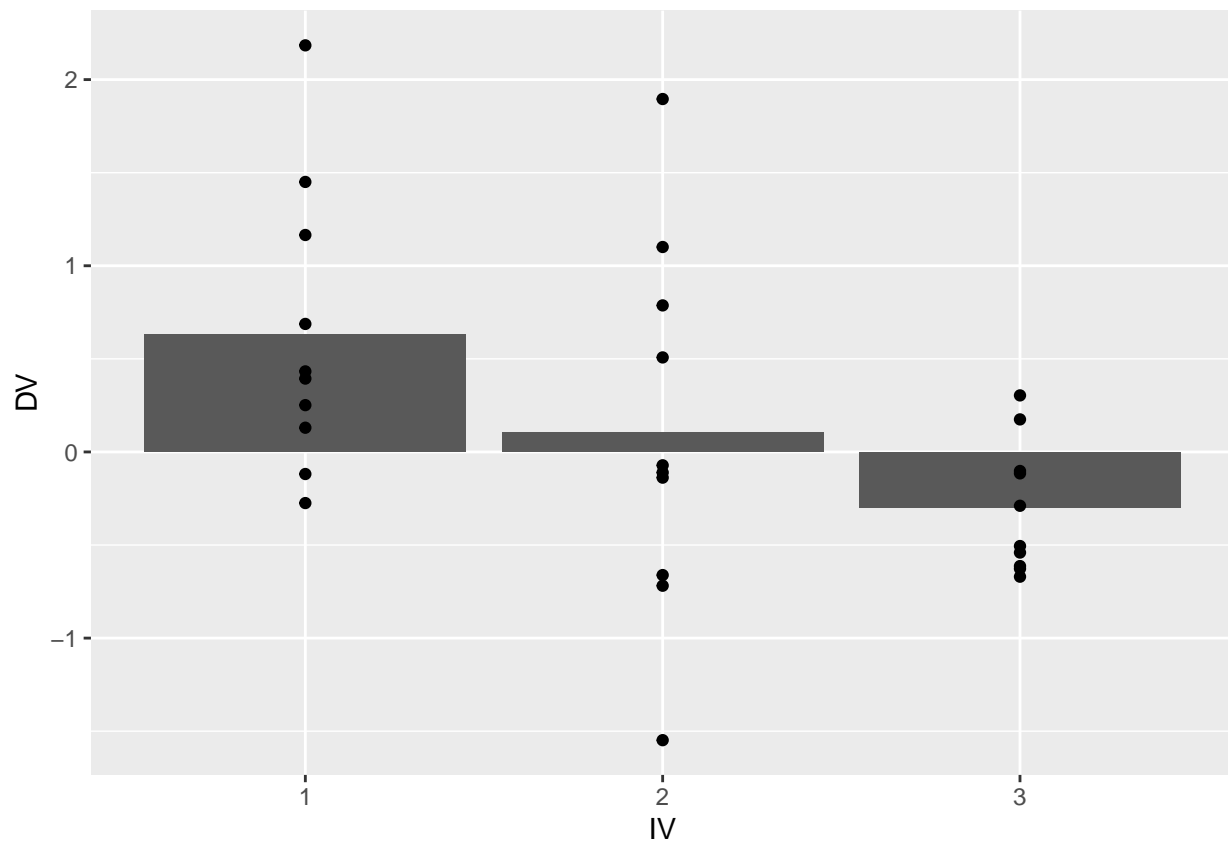
## anova

```
my.aov
```

```
##      Df Sum Sq Mean Sq F value Pr(>F)  
## IV      2  4.337   2.1687   3.824 0.0345 *  
## Residuals 27 15.311   0.5671  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## ggplot

```
ggplot(sim_data, aes(x = IV, y = DV)) +  
  geom_bar(stat = "summary", fun = "mean") +  
  geom_point()
```



I would reject the null because my p-value is less than .05. I would be correct in doing so because my p-value is less than my alpha level. These results would be considered significant.

## Bonus Question

In the lab we saw that F-distribution is robust to violations of the assumptions of ANOVA. For example, the simulation of the null based on a bi-modal distribution was very similar to the true F distribution. For this bonus question, show that you can “break” the F-distribution. Specifically, can you run a simulation that samples numbers from a non-normal distribution that does produce a very different looking F-distribution? (3 points)

```
levels <- 3
n_per_level <- 10

save_F_values <- length(1000)
for(i in 1:1000){
  random_data <- tibble(subjects = 1:(levels*n_per_level),
                        IV = as.factor(rep(1:levels, each = n_per_level)),
                        DV = rnorm(levels*n_per_level, 0, 1)
                      )
  aov.out <- aov(DV ~ IV, data = random_data)
  simulated_F <- summary(aov.out)[[1]]$`F value`[1]
  save_F_values[i] <- simulated_F
}

F_comparison <- tibble(type = rep(c("analytic", "simulated"), each = 1000),
                      F_value = c(rf(1000, levels-1, levels*n_per_level), save_F_values))

ggplot(F_comparison, aes(x=F_value, color = type))+
  geom_freqpoly(bins = 50)
```

