

# ANOVA vs Regression

## PART 1 REGRESSION VS ANOVA

**An ANOVA-design and a regression-design experiment**

**Now to do the actual assingment:**

```

# set up your data:

N <- length(antibiotics$growth)
y <- antibiotics$growth
x<- antibiotics$trt
antibiotics$x1 <- ifelse(x=="ab1", 1, 0) #dummy variables
antibiotics$x2 <- ifelse(x=="ab2", 1, 0) #dummy variables
antibiotics$x3 <- ifelse(x=="ab3", 1, 0) #dummy variables

#Build null model
nullmod<-function(p,x,y){
  B0=p[1]
  sigma=exp(p[2])

  expected=B0

  null=-sum(dnorm(x=y,mean=expected,sd=sigma, log = TRUE))
  return(null)
}

#Build full model
fullmod<-function(p,x,y){
  B0=p[1]
  B1=p[2]
  B2=p[3]
  B3=p[4]
  sigma=exp(p[5])

  expected=B0+B1*x[,3]+B2*x[,4]+B3*x[,5]

  full=-sum(dnorm(x=y,mean=expected,sd=sigma,log = TRUE))
  return(full)
}

# Check fit
nullguess <- c(1, 2)
fullguess <- c(18, -12, -3, -4, 1)
fitnull=optim(par=nullguess,fn=nullmod,y=antibiotics$growth) # converges
fitfull=optim(par=fullguess,fn=fullmod,x=antibiotics,y=antibiotics$growth) # converges

# Get t.statistic .
t.stat <- 2*(fitnull$value - fitfull$value)
t.stat # 37.90134

```

```
## [1] 37.90134
```

```

# get degrees of freedom
df <- length(fitfull$par) - length(fitnull$par)

# and p value
1-pchisq(t.stat, df=df) # 2.965739e-08

```

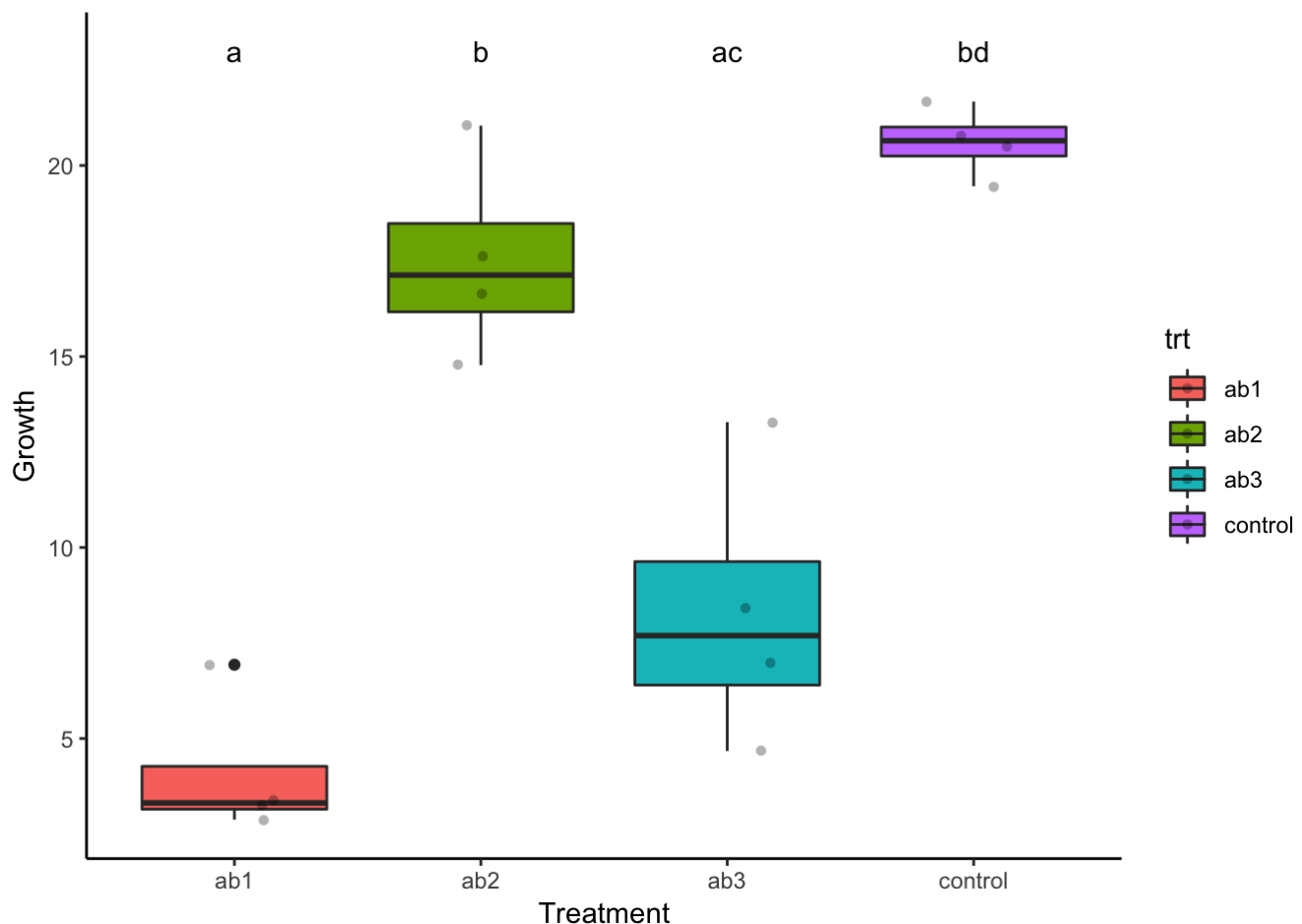
```
## [1] 2.965739e-08
```

Looks like there's a significant difference between the null and full anova model ( $t = 25.69$ ,  $p < .001$ ), suggesting that there are differences between treatments.

## Graphical visualization!

Using a posthoc Tukey test (see hidden code) we can find significant differences between groups, and add these differences as letters to our visualization (see graph below)

```
p <- ggplot(antibiotics, aes(x=trt, y=growth, fill=trt)) +
  geom_boxplot()
p + geom_jitter(shape=16, position=position_jitter(0.2), alpha=0.3) + theme_classic() +
  labs(x = "Treatment", y = "Growth") + annotate("text", x = c(1, 2, 3, 4), y = c(23, 23,
  23, 23), label = c("a", "b", "ac", "bd"))
```



## Regression

Build the null model:

```
lmnullmodreg<-function(p,x,y){
  B0=p[1]
  sigma=exp(p[2])

  expected=B0

  null=-sum(dnorm(x=y,mean=expected,sd=sigma, log = TRUE))
  return(null)
}
```

Now build the extension:

```
lmfullmodreg<-function(p,x,y){
  B0=p[1]
  B1=p[2]
  sigma=exp(p[3])

  expected=B0+B1*x

  null=-sum(dnorm(x=y,mean=expected,sd=sigma, log = TRUE))
  return(null)
}
```

Give it a guess:

```
lmnullregguess <- c(1, 1)
lmfullregguess <- c(1, 2, 3)
lmfitregnull=optim(par=lmnullregguess,fn=lmnullmodreg,x=sugar$sugar,y=sugar$growth) # converges
lmfitregfull=optim(par=lmfullregguess,fn=lmfullmodreg,x=sugar$sugar,y=sugar$growth) # converges
```

Now to compare:

```
# Get t statistic
t.statreg <- 2*(lmfitregnull$value - lmfitregfull$value)
t.statreg # 39.92512
```

```
## [1] 39.92512
```

```
# get degrees of freedom
dfreg <- length(lmfitregfull$par) - length(lmfitregnull$par)

#and p value
1-pchisq(t.stat, df=dfreg) # 7.441429e-10, woot
```

```
## [1] 7.441429e-10
```

Comparing between the null model of no effect of sugar concentration on growth to the model taking this affect into account, we can see that sugar concentration does have a significant affect on growth ( $t = 39.92512$ ,  $p < .0001$ ). Looking at the graph below, we can tell that sugar concentration has a positive affect, with higher concentrations of sugar increasing growth.

## Visualization!

```
a <- ggplot(data=sugar,aes(x=sugar,y=growth))  
a + geom_point() + coord_cartesian() + labs( x = "sugar concentration", y = "growth") +  
  theme_classic() + geom_smooth(method = "lm")
```

