

# b201a final code

Lillian Chen

12/5/2021

```
#1a
d<- 0.68
pnorm(d, mean=0, sd=1)

## [1] 0.7517478

options(knitr.kable.NA = '')
k <- 3
N <- 12
df_b <- k-1
df_w <- N-k
df_tot <- N-1
grandmean <- (4*28.60 + 4*31.38 + 4*19.08)/N
mean11 <- 28.60
mean12 <- 31.38
mean13 <- 19.08
ssb <- 4*(28.60-grandmean)^2 + 4*(31.38-grandmean)^2 + 4*(19.08-grandmean)^2
ssw <- 3*(3.22)^2 + 3*(3.17)^2 + 3*(1.81)^2
msb <- ssb/df_b
msw <- ssw/df_w
Fstat <- msb/msw
p <- 1 - pf(Fstat, k-1, N-k)

q7a <- data.frame(Source = c("Between groups (Treatment)",
                             "Within groups (Error)",
                             "Total"),
                  df = c(k-1, N-k, N-1),
                  SS = c(ssb, ssw, ssb+ssw),
                  MS = c(msb, msw, (ssb+ssw)/(N-1)),
                  F = c(Fstat, NA, NA),
                  p = c(p, NA, NA))

kbl(q7a, booktabs = T,
     caption = "ANOVA Table 2A") %>%
  kable_styling(latex_options = c("hold_position"))

abdata <- data.frame(speed = c("penicillin", "tetracycline", "erythromycin"),
                     n = c(4,4,4),
                     means = c(28.60,31.38,19.08),
                     sd = c(3.22,3.17,1.81))
```

Table 1: ANOVA Table 2A

| Source                     | df | SS       | MS       | F        | p         |
|----------------------------|----|----------|----------|----------|-----------|
| Between groups (Treatment) | 2  | 332.8651 | 166.4325 | 21.07328 | 0.0004022 |
| Within groups (Error)      | 9  | 71.0802  | 7.8978   |          |           |
| Total                      | 11 | 403.9453 | 36.7223  |          |           |

```
#using anovaMean function from package HH for one-way ANOVA with summary statistics
anova <- HH::anovaMean(abdata$speed, abdata$n, abdata$means, abdata$sd, ylabel=c("antibiotic"))
```

```
## Warning in ybar - (ybar %*% n)/sum(n): Recycling array of length 1 in vector-array arithmetic is deprecated
## Use c() or as.vector() instead.
```

```
anova
```

```
## Analysis of Variance Table
##
## Response: antibiotic
##
## Terms added sequentially (first to last)
##           Df Sum of Sq Mean Sq F value    Pr(F)
## antibiotic  2      332.87  166.433   21.073 0.0004022 ***
## Residuals   9       71.08    7.898
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 2c
(Rsq <- 1 - anova$`Sum of Sq`[2]/sum(anova$`Sum of Sq`))
```

```
## [1] 0.8240351
```

```
# 2d: for bonferroni correction, divide by # pairwise comparisons for
# new alpha for each individual test to maintain exp-wise error rate
(astar <- 0.05/3)
```

```
## [1] 0.01666667
```

```
#2e
SSEfull <- 24.21
SST <- 403.97 # same total sum of squares
(SSmodfull <- SST - SSEfull)
```

```
## [1] 379.76
```

```
SSmodreduced <- 332.95 # SSmod from ANOVA table
dfmodfull <- 3 # 4 - 1 (k=4 comes from 3 groups + additional age predictor)
dfmodreduced <- 2 # dfmod from ANOVA table
dfefull <- 8 # 12 - 4 (N=12, k=4)
```

```
# calculate partial F ratio
# numerator (MS additional explained)
(num <- (SSmodfull - SSmodreduced)/(dfmodfull-dfmodreduced))
```

```
## [1] 46.81
```

```
#denominator (MS still unexplained)
(denom <- SSEfull/dfefull)
```

```
## [1] 3.02625
```

```
# partial F ratio
(Fpartial <- num/denom)
```

```
## [1] 15.46799
```

```
# critical value of F with degrees of freedom 1, 8 at alpha=0.05
(Fstar <- qf(0.95,1,8))
```

```
## [1] 5.317655
```

```
# p-value for observed Fpartial with degrees of freedom 1, 8
(pvalue <- 1 - pf(Fpartial, 1, 8))
```

```
## [1] 0.00433874
```

```
# 3a
df_f <- 110
df_m <- 125
N_f <- df_f + 1
N_m <- df_m + 1
(Ntot <- N_f + N_m)
```

```
## [1] 237
```

```
# 3b
regf <- function(x){
  y <- -153.12891 + 4.16361*x
  return(y)}

regm <- function(x){
  y <- -125.69807 + 3.68977*x
  return(y)}
```

```
regf(60) # predicted weight for girl with height = 60 in
```

```
## [1] 96.68769
```

```
regm(60) # predicted weight for boy with height = 60 in
```

```
## [1] 95.68813
```

```
# 3c
SST_f <- 38121.10811
df_f <- 110
SST_m <- 49840.41468
df_m <- 125
(svar_f <- SST_f/df_f) # sample variance of weight in girls
```

```
## [1] 346.5555
```

```
(svar_m <- SST_m/df_m) # sample variance of weight in boys
```

```
## [1] 398.7233
```

```
# 3f
# ref group = girls
B0 <- -153.13 # intercept when indicator is 0
B1 <- 4.16 # slope when indicator is 0
B2 <- -125.70 - B0 # additional contribution to intercept when indicator is 1
B3 <- 3.69 - B1 # additional contribution to intercept when indicator is 1
SSR <- 16614.59 + 18714.35 # SSR from both female and male models

(MSE <- SSR/(237-3)) # residual variance or MSE; 237 observations - 3 parameters
```

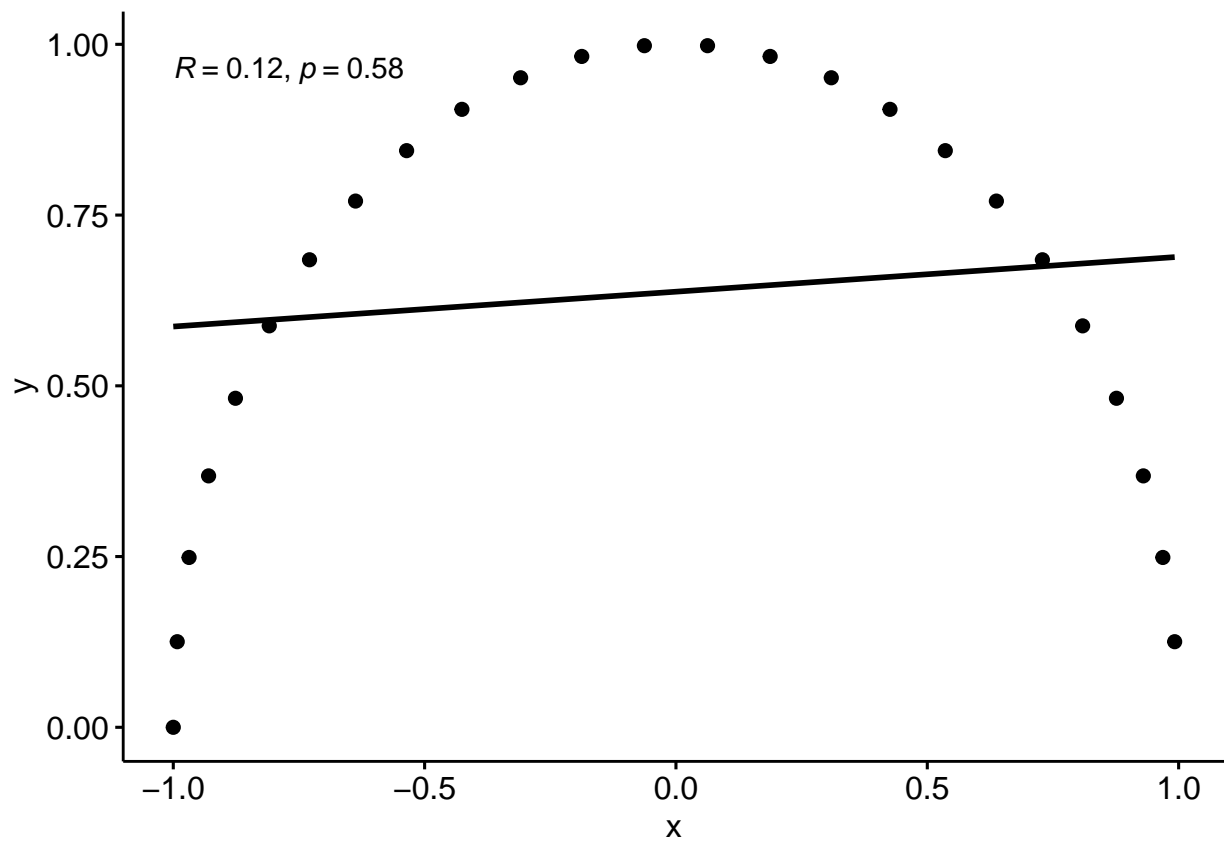
```
## [1] 150.9784
```

```
u <- seq(0.04, 1, 0.04)
x <- cos(pi*u)
e <- dnorm(x, mean = 0, sd = 0.01)
y <- sin(pi*u) + e

data <- data.frame(u=u, x = x, x2 = x^2, y = y)

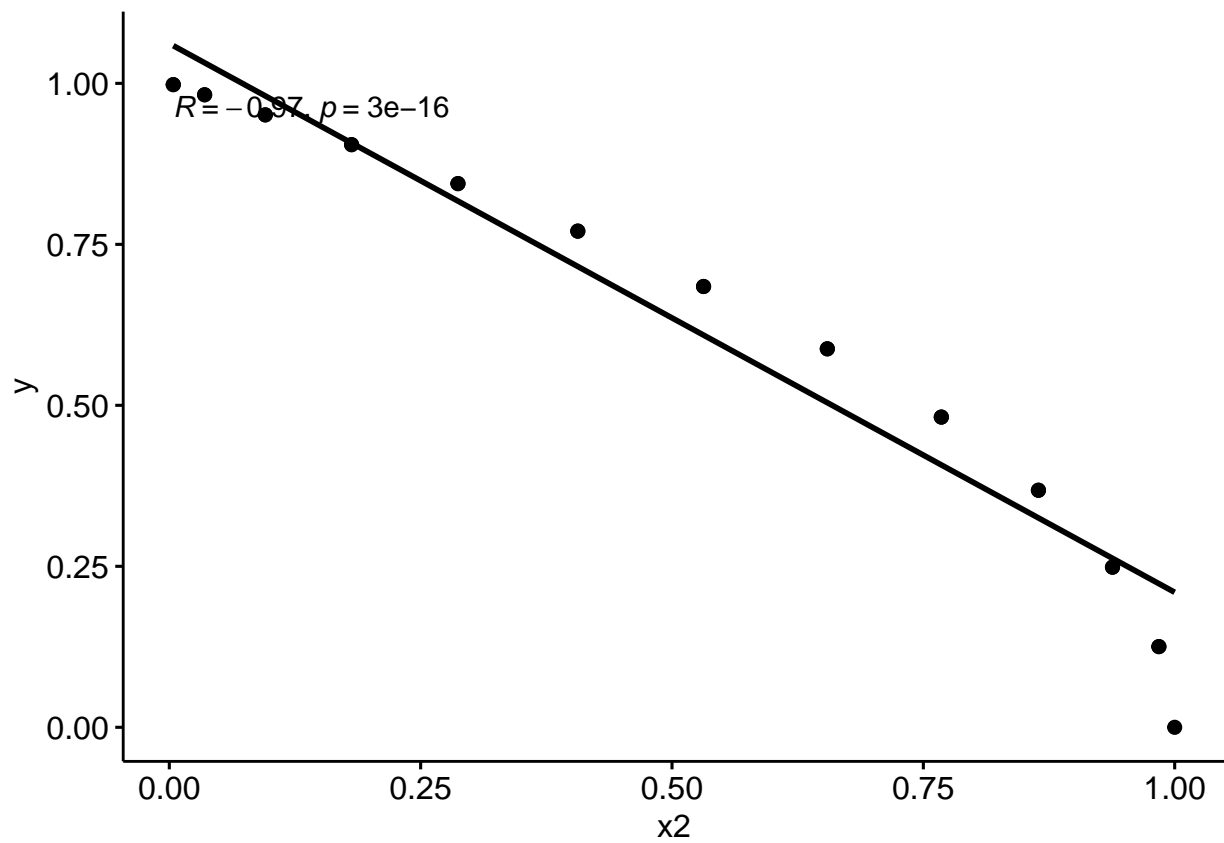
ggscatter(data = data, x = "x", y = "y", add = "reg.line", cor.coef = T)
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

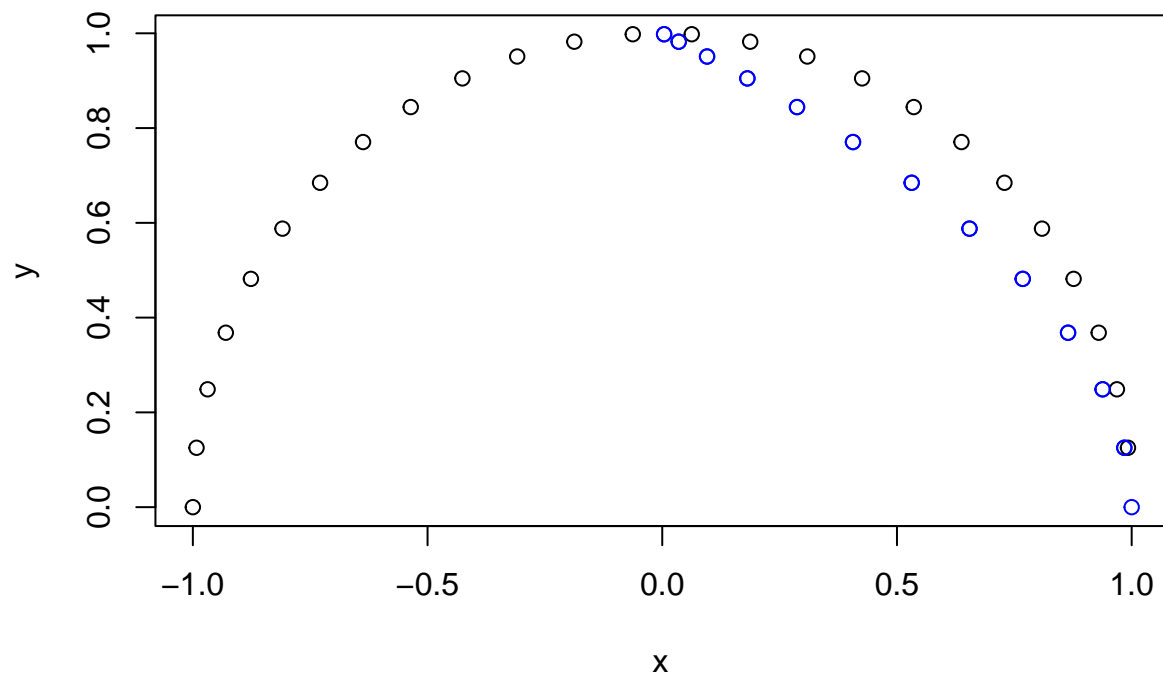


```
ggscatter(data = data, x = "x2", y = "y", add = "reg.line", cor.coef = T)
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



```
plot(x,y)
points(x^2,y, col="blue")
```



```
lm(y~x, data = data)
```

```
##  
## Call:  
## lm(formula = y ~ x, data = data)  
##  
## Coefficients:  
## (Intercept)          x  
##    0.63782      0.05103
```

```
lm(y~x+x2, data=data)
```

```
##  
## Call:  
## lm(formula = y ~ x + x2, data = data)  
##  
## Coefficients:  
## (Intercept)          x          x2  
##    1.06120      0.01694     -0.84948
```