

Bios 6301: Assignment 3

Wooyeol Lee

October 7, 2015

Question 1

10 points

1. Use GitHub to turn in the first three homework assignments. Make sure the teacher (couthcommander) and TA (trippcm) are collaborators. (5 points)
2. Commit each assignment individually. This means your repository should have at least three commits. (5 points)

Question 2

15 points

Write a simulation to calculate the power for the following study design. The study has two variables, treatment group and outcome. There are two treatment groups (0, 1) and they should be assigned randomly with equal probability. The outcome should be a random normal variable with a mean of 60 and standard deviation of 20. If a patient is in the treatment group, add 5 to the outcome. 5 is the true treatment effect. Create a linear model for the outcome by the treatment group, and extract the p-value (hint: see assignment1). Test if the p-value is less than or equal to the alpha level, which should be set to 0.05.

Repeat this procedure 1000 times. The power is calculated by finding the percentage of times the p-value is less than or equal to the alpha level. Use the `set.seed` command so that the professor can reproduce your results.

1. Find the power when the sample size is 100 patients. (10 points)

```
#set seed
set.seed(1000)
nperson=100
j<-c(1:nperson)
nsim <-1000
pvals<-numeric(nsim)
for(i in seq_along(pvals)) {
  # make variables n=100
  group <- rbinom(nperson, size=1, prob=0.5)
  outcome <- rnorm(nperson, mean=60, sd=20)
  #add 5
  outcome<-ifelse(group[j] ==1, outcome[j]+5, outcome[j])
  #extract p-value
  pvals[i]<- summary(lm(outcome ~ group))$coefficients[2,4]
}
mean(pvals < 0.05)
```

```
## [1] 0.223
```

1. Find the power when the sample size is 1000 patients. (5 points)

```

set.seed(1000)
nperson=1000
j<-c(1:nperson)
nsim <-1000
pvals<-numeric(nsim)
for(i in seq_along(pvals)) {
  # make variables n=1000
  group <- rbinom(nperson, size=1, prob=0.5)
  outcome <- rnorm(nperson, mean=60, sd=20)
  #add 5
  outcome<-ifelse(group[j] ==1, outcome[j]+5, outcome[j])
  #extract p-value
  pvals[i]<- summary(lm(outcome ~ group))$coefficients[2,4]
}
mean(pvals < 0.05)

```

```
## [1] 0.968
```

Question 3

15 points

Obtain a copy of the [football-values lecture](#). Save the 2015/proj_rb15.csv file in your working directory. Read in the data set and remove the first two columns.

```

fb <- read.csv("C:/Users/Wooyeol/Dropbox/me/coursework/fall2015/statistical computing/homework/proj_rb15.csv")
fb <- fb[,3:10]
vcov.fb <- var(fb)
means.fb <- colMeans(fb)

```

1. Show the correlation matrix of this data set. (3 points)

```
cor(fb)
```

```

##          rush_att rush_yds rush_tds rec_att rec_yds rec_tds
## rush_att 1.0000000 0.9975511 0.9723599 0.7694384 0.7402687 0.5969159
## rush_yds 0.9975511 1.0000000 0.9774974 0.7645768 0.7345496 0.6020994
## rush_tds 0.9723599 0.9774974 1.0000000 0.7263519 0.6984860 0.5908348
## rec_att  0.7694384 0.7645768 0.7263519 1.0000000 0.9944243 0.8384359
## rec_yds  0.7402687 0.7345496 0.6984860 0.9944243 1.0000000 0.8518924
## rec_tds  0.5969159 0.6020994 0.5908348 0.8384359 0.8518924 1.0000000
## fumbles  0.8589364 0.8583243 0.8526904 0.7459076 0.7224865 0.6055598
## fpts      0.9824135 0.9843044 0.9689472 0.8556928 0.8340195 0.7133908
##          fumbles      fpts
## rush_att 0.8589364 0.9824135
## rush_yds 0.8583243 0.9843044
## rush_tds 0.8526904 0.9689472
## rec_att  0.7459076 0.8556928
## rec_yds  0.7224865 0.8340195
## rec_tds  0.6055598 0.7133908
## fumbles  1.0000000 0.8635550
## fpts      0.8635550 1.0000000

```

1. Generate a data set with 30 rows that has a similar correlation structure. Repeat the procedure 10,000 times and return the mean correlation matrix. (10 points)

```
library(MASS)
keep.1 <- 0
loops <- 10000
for(i in seq(loops)) {
  fb.sim <- mvrnorm(30, mu=means.fb, Sigma= vcov.fb)
  keep.1 <- keep.1 + cor(fb.sim)/loops
}
keep.1
```

```
##          rush_att rush_yds rush_tds  rec_att  rec_yds  rec_tds
## rush_att 1.0000000 0.9974623 0.9715725 0.7643399 0.7348569 0.5911449
## rush_yds 0.9974623 1.0000000 0.9768402 0.7593619 0.7290266 0.5961409
## rush_tds 0.9715725 0.9768402 1.0000000 0.7210311 0.6929974 0.5845830
## rec_att  0.7643399 0.7593619 0.7210311 1.0000000 0.9941970 0.8340757
## rec_yds  0.7348569 0.7290266 0.6929974 0.9941970 1.0000000 0.8477972
## rec_tds  0.5911449 0.5961409 0.5845830 0.8340757 0.8477972 1.0000000
## fumbles  0.8548889 0.8541098 0.8481087 0.7397548 0.7160245 0.5980223
## fpts      0.9818098 0.9837315 0.9679681 0.8519648 0.8299098 0.7077617
##          fumbles      fpts
## rush_att 0.8548889 0.9818098
## rush_yds 0.8541098 0.9837315
## rush_tds 0.8481087 0.9679681
## rec_att  0.7397548 0.8519648
## rec_yds  0.7160245 0.8299098
## rec_tds  0.5980223 0.7077617
## fumbles  1.0000000 0.8591717
## fpts      0.8591717 1.0000000
```

1. Generate a data set with 30 rows that has the exact correlation structure as the original data set. (2 points)

```
fb.sim <- mvrnorm(30, mu = means.fb, Sigma = vcov.fb, empirical=T)
fb.sim <- as.data.frame(fb.sim)
cor(fb.sim)
```

```
##          rush_att rush_yds rush_tds  rec_att  rec_yds  rec_tds
## rush_att 1.0000000 0.9975511 0.9723599 0.7694384 0.7402687 0.5969159
## rush_yds 0.9975511 1.0000000 0.9774974 0.7645768 0.7345496 0.6020994
## rush_tds 0.9723599 0.9774974 1.0000000 0.7263519 0.6984860 0.5908348
## rec_att  0.7694384 0.7645768 0.7263519 1.0000000 0.9944243 0.8384359
## rec_yds  0.7402687 0.7345496 0.6984860 0.9944243 1.0000000 0.8518924
## rec_tds  0.5969159 0.6020994 0.5908348 0.8384359 0.8518924 1.0000000
## fumbles  0.8589364 0.8583243 0.8526904 0.7459076 0.7224865 0.6055598
## fpts      0.9824135 0.9843044 0.9689472 0.8556928 0.8340195 0.7133908
##          fumbles      fpts
## rush_att 0.8589364 0.9824135
## rush_yds 0.8583243 0.9843044
## rush_tds 0.8526904 0.9689472
## rec_att  0.7459076 0.8556928
```

```
## rec_yds 0.7224865 0.8340195
## rec_tds 0.6055598 0.7133908
## fumbles 1.0000000 0.8635550
## fpts    0.8635550 1.0000000
```

Question 4

10 points

Use \LaTeX to create the following expressions.

1. Hint: `\Rightarrow` (4 points)

$$P(B) = \sum_j P(B|A_j)P(A_j),$$

$$\Rightarrow P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

1. Hint: `\zeta` (3 points)

$$\hat{f}(\zeta) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x \zeta} dx$$

2. Hint: `\partial` (3 points)

$$\mathbf{J} = \frac{d\mathbf{f}}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$