PS1 Solutions

Teaching Staff

February 2023

Question 1

1

```
set.seed(8675309)

### Create a vector
vec1 = 1:1000
avec1 = seq(1,1000,1)
```

2

```
### Sampling
vec2 = sample(vec1, replace = F)
avec2 = sample(avec1)
```

3

```
### Creating a data frame
dat = data.frame(vec1, vec2)
```

4

```
### compute correlation
cor(dat$vec1, dat$vec2)
```

```
## [1] -0.02283693
```

5

The correlation will be close to 0 because randomization has mechanically broken the relationship between the two variables. The actual correlation might not be 0 due to small sample size.

Question 2

1

```
hdat = read.csv("../data/data_health_synth_small.csv")
```

2

```r
dim(hdat)
```

```
## [1] 48784     4
```

```r
## alternatively
nrow(hdat)
```

```
## [1] 48784
```

```r
ncol(hdat)
```

```
## [1] 4
```

3

```r
summary(hdat)
```

```
##       cost            race               female          bps_mean
##  Min.   :     0   Length:48784       Min.   :0.0000   Min.   :   0.0
##  1st Qu.:  1200   Class :character   1st Qu.:0.0000   1st Qu.: 118.0
##  Median :  2800   Mode  :character   Median :1.0000   Median : 127.0
##  Mean   :  7660                      Mean   :0.6306   Mean   : 127.3
##  3rd Qu.:  6600                      3rd Qu.:1.0000   3rd Qu.: 136.0
##  Max.   :550500                      Max.   :1.0000   Max.   :1323.0
##                                                       NA's   :10668
```

4

```r
### Mean difference
cost_b = mean(hdat$cost[hdat$race == "black"])
cost_w = mean(hdat$cost[hdat$race == "white"])

ate = cost_b - cost_w
ate
```

```
## [1] 1782.368
```