Pairs Assignment 02/27

Kerem Tuncer

Question 16.14

Let's simulate a canvassing study where 100 neighborhoods are chosen. Half of these neighborhoods are canvassed and the other half is not. We assume that a 5 percentage point increase in election turnout is brought on to the neighborhood because of the canvassing. However, we have used the rnorm() function with a mean 5 and a standard deviation of 1 to add some stochastic differences in effect magnitude.

```
set.seed(19)
library(rstanarm)
```

```
## Warning: package 'rstanarm' was built under R version 4.0.5
```

```
## Warning: package 'Rcpp' was built under R version 4.0.5
```

```
library("retrodesign")
```

```
## Warning: package 'retrodesign' was built under R version 4.0.5
```

```
n <- 100
turnout_control <- rnorm(n, mean = 45, sd = 15)
turnout_treatment <- turnout_control + rnorm(n, mean = 5, sd = 1)
z <- sample(0:1, n, replace = TRUE)
turnout <- ifelse(z==1, turnout_treatment, turnout_control)
dat <- as.data.frame(cbind(z, turnout))

fit_1 <- stan_glm(turnout ~ z, data = dat, refresh = 0)
fit_1
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      turnout ~ z
##  observations: 100
##  predictors:   2
## ------
##             Median MAD_SD
## (Intercept) 45.0    2.2
## z            5.4    3.3
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 15.8    1.1
```

```
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

```r
pnorm(5.4/3.3, lower.tail=FALSE)
```

```
## [1] 0.05088175
```

The estimated average treatment effect is 5.4. This means that being canvassed increases the turnout in elections by 5.4 percentage points in a given neighborhood. The standard error of the estimate is 3.3. The standard error is too big, which makes our estimate not precise. Now, let's look at power calculations.

```r
set.seed(1907)

retrodesign(A = 5.4, s = 3.3, n.sims = 100000)
```

```
## $power
## [1] 0.3732817
##
## $typeS
## [1] 0.0004323026
##
## $exaggeration
## [1] 1.619161
```

The power of our study was 0.37, which means that there is a 37% chance that we will correctly reject the null hypothesis when the alternative hypothesis is true. This power isn't considered good research in most fields. In contrast, the type s error rate is 0.0004, which is very small. So, we are pretty confident that the sign of our estimate is correct. This makes sense given that we have a simulated dataset. Finally, our type M rate is 1.619161.

To increase the power of study, let's assume that we recruited 900 more neighborhoods, so we have 1000 neighborhoods in total now. Let's redo the analysis.

```r
set.seed(1907)
n <- 1000
turnout_control <- rnorm(n, mean = 45, sd = 15)
turnout_treatment <- turnout_control + rnorm(n, mean = 5, sd = 1)
z <- sample(0:1, n, replace = TRUE)
turnout <- ifelse(z==1, turnout_treatment, turnout_control)
dat <- as.data.frame(cbind(z, turnout))

fit_2 <- stan_glm(turnout ~ z, data = dat, refresh = 0)
fit_2
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      turnout ~ z
##  observations: 1000
##  predictors:   2
## ------
```

```
##              Median MAD_SD
## (Intercept) 44.7    0.6
## z            5.1    0.9
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 15.2    0.3
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

```r
retrodesign(A = 5.1, s = 0.9, n.sims = 100000)
```

```
## $power
## [1] 0.999895
##
## $typeS
## [1] 1.20497e-14
##
## $exaggeration
## [1] 0.9993903
```

Now, as expected, our power is very high. This means that there is an extremely high chance that we will correctly reject the null hypothesis when the alternative hypothesis is true. Likewise, our type S and type M error rates are much smaller now.