0. Load our first set of subjects:

```r
cntrfact <- read.table("http://www.stat.washington.edu/tsr/s566/labs/cntrfact1.dat")
```

Total number of subjects:

```r
(n <- nrow(cntrfact))
```

```r
## [1] 500
```

See the true potential outcomes of the first 10 patients. *The researcher typically will not know these!*

- ydrug \(\equiv Y(x = 1)\)
- yplac \(\equiv Y(x = 0)\)

```r
head(cntrfact, n = 10)
```

```r
##  yplac ydrug
## 1   0   1
## 2   0   1
## 3   1   0
## 4   0   1
## 5   0   1
## 6   1   0
## 7   0   1
## 8   0   0
## 9   0   1
##10   0   1
```

Contingency table of response types for all subjects.

```r
table(cntrfact)
```

```r
## ydrug
## yplac  0  1
## 0  46 313
## 1  98  43
```
1. Randomly assign some treatments:

```r
# make vectors for the counterfactual outcomes
ydrug <- cntrfact[, "ydrug"]
yplac <- cntrfact[, "yplac"]

pdrug <- 0.5  # true probability to receive drug
set.seed(566)
tmp <- rbinom(n, 1, pdrug)  # Bernoulli random variable
trt <- ifelse(test=tmp, yes="Drug", no="Plac")
rm(tmp)

table(trt) / n  # empirical probability to receive drug

## trt
## Drug  Plac
## 0.494 0.506
```

2. Get the observed outcomes.

```r
yobs <- ydrug*(trt == "Drug") + yplac*(trt == "Plac")

# try to work out what this did
# Hint: "TRUE" corresponds to 1; "False" to 0

Let’s see what the researcher actually gets to see after an experiment is performed.

```r
noquote(cbind("trt"=trt[1:10], "yobs"=yobs[1:10]))

##     trt yobs
##  [1,]  Plac  0
##  [2,]  Plac  0
##  [3,]  Drug  0
##  [4,]  Plac  0
##  [5,]  Plac  0
##  [6,]  Drug  0
##  [7,]  Drug  1
##  [8,]  Drug  0
##  [9,]  Drug  1
## [10,]  Plac  0
```

```r
table(yobs)

## yobs
##  0  1
## 242 258
```

```r
table(trt)
```

```r
```
3. Find proportion of successes in the drug group:

```r
mean(yobs[trt=="Drug"])
```

## [1] 0.737

Compare this to the true proportion of successes if everyone got the drug. *This is not something the researcher can see!*

```r
mean(ydrug)
```

## [1] 0.712

4. Find proportion of successes in the placebo/control group:

```r
mean(yobs[trt=="Plac"])
```

## [1] 0.3

Compare this to the true proportion of successes if everyone got the placebo. *Again, not something the researcher can see!*

```r
mean(yplac)
```

## [1] 0.282

5. Compute estimate of the average causal effect (ACE):

```r
mean(yobs[trt=="Drug"]) - mean(yobs[trt=="Plac"])
```

## [1] 0.436

Compare this to the true average causal effect:
\texttt{mean(ydrug) - mean(yplac)}

## [1] 0.43

Any difference between the estimate and the true value is due to sampling variability i.e. it would disappear as $n \to \infty$.

### 6. Response types

How many individuals are of type Never Recover? Helped? Hurt? Always Recover? \textit{Note that we could have done this before we assigned treatment.}

\texttt{table(ydrug,yplac)}

<table>
<thead>
<tr>
<th>yplac</th>
<th>ydrug 0</th>
<th>ydrug 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>46</td>
<td>98</td>
</tr>
<tr>
<td>1</td>
<td>313</td>
<td>43</td>
</tr>
</tbody>
</table>

\begin{verbatim}
helped <- sum(ydrug*(1-yplac))
hurt <- sum((1-ydrug)*yplac)
alwaysrec <- sum(ydrug*yplac)
doomed <- sum((1-ydrug)*(1-yplac))

# how about a nice table:
noquote( cbind(c("helped","hurt","alwaysrec","doomed"),
   c(helped,hurt,alwaysrec,doomed)/n) )
\end{verbatim}

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>helped</td>
<td>313</td>
<td>0.626</td>
<td></td>
</tr>
<tr>
<td>[2]</td>
<td>hurt</td>
<td>98</td>
<td>0.196</td>
<td></td>
</tr>
<tr>
<td>[3]</td>
<td>alwaysrec</td>
<td>43</td>
<td>0.086</td>
<td></td>
</tr>
<tr>
<td>[4]</td>
<td>doomed</td>
<td>46</td>
<td>0.092</td>
<td></td>
</tr>
</tbody>
</table>

Could this have been seen by the researcher?
Exercises:

(I)
Now repeat from step 1, but replace `set.seed(566)` with `set.seed(D)`, where D is your favourite integer. When you repeat the subsequent steps, what is the same? What is (slightly) different?

(II)
Now repeat from step 1, but this time change `pdrug <- 0.5` to `pdrug <- P`, where P is your favourite probability (strictly) between 0 and 1. When you repeat the subsequent steps, what is the same? What has changed? Is our estimate of the average causal effect very different?

(III)
Now repeat from step 0 but with a different data set (cntrfact2.dat).

```r
rm(cntrfact)
cntrfact <- read.table("http://www.stat.washington.edu/tsr/s566/labs/cntrfact2.dat")
```

Repeat all steps and pay particular attention to:

- 5. the true and estimated ACE
- 6. the matrix of patient types

(IV)
Now repeat from step 0 but with another data set (cntrfact3.dat):

```r
rm(cntrfact)
cntrfact <- read.table("http://www.stat.washington.edu/tsr/s566/labs/cntrfact3.dat")
```

Repeat all steps and pay particular attention to:

- 5. the true and estimated ACE
- 6. the matrix of patient types

Contrast these to those obtained from (III).

(V)
Looking at your answers to (III) and (IV), is the following statement correct?

- “If the average causal effect is (close to) zero, then the drug is (close to) doing nothing.”

Why? **Hint:** think about the perspective of an individual (or his lawyer) vs. the perspective of a policy analyst.