Computer Project 3: Random Sampling

Your Name

2020-06-30

Box with Known Contents

Mean and SD of the Box

Here is the box model from Exercise 11 on page 329.

```
box <- c(1, 2, 2, 5)
box
```

```
## [1] 1 2 2 5
```

Compute the mean of the box:

```
m <- mean(box)
m
```

```
## [1] 2.5
```

and the SD of the box:

```
s <- sd(box) * sqrt(3/4)
s
```

```
## [1] 1.5
```

EV (expected value) and SE (standard error) of the Sum Sample of Size 100

If we take a random sample (with replacement) of size 100, the EV of the sum is:

```
n <- 100
ev <- n * m
ev
```

```
## [1] 250
```

and the SE of the sum is:

```
se <- sqrt(n) * s
se
```

```
## [1] 15
```

Results of a Random Sample of Size 100

Randomly select 100 tickets without replacement.

```
x <- sample(box, n, replace=TRUE)
x
```

```
##   [1] 5 1 5 2 1 2 1 1 1 1 2 5 2 2 5 5 5 1 2 1 1 5 2 2 5 1 2 1 1 2 1 2 1 2 1
##  [36] 2 2 2 2 2 2 1 1 5 2 5 2 1 1 2 1 5 1 2 2 5 5 2 2 1 2 2 1 2 5 5 1 2 2 2
##  [71] 5 1 2 2 5 2 2 2 5 5 1 2 2 2 5 2 5 2 1 2 1 5 1 5 2 2 2 5 2 2
```

Now find the sum of the sample:

```
sum(x)
```

```
## [1] 243
```

Compute the z-score for the observed sum:

```
z <- (sum(x) - ev) / se
z
```

```
## [1] -0.4666667
```

Hi < hi

How likely is it to get a sum as extreme or more extreme than the value we observed? Use either pnorm(z) or 1 - pnorm(z) to answer:

```
# Type your R command here in the Rmd file (either pnorm(z) or 1 - pnorm(z))
```

Is the sum that you got very low, about average or very high? Use the z-score or pnorm output to answer.

Results of a Random Sample of Size 10,000

Recalculate the EV for the sum:

```
n <- 10000
ev <- n * m
ev
```

```
## [1] 25000
```

and the SE for the sum:

```
se <- sqrt(n) * s
se
```

```
## [1] 150
```

Randomly select 10^4 tickets without replacement.

```
x <- sample(box, n, replace=TRUE)
```

Now find the sum of the sample:

```
sum(x)
```

```
## [1] 24886
```

Compute the z-score for the sum:

```
z <- (sum(x) - ev) / se
z
```

```
## [1] -0.76
```

How likely is it to get a sum as extreme or more extreme than the value we observed? Use either pnorm(z) or 1 - pnorm(z) to answer:

```
# Type your R command here in the Rmd file (either pnorm(z) or 1 - pnorm(z))
```

Is the sum that you got very low, about average or very high? How did your answers for n=100 and n=10,000 differ? Is that what you would expect?

Confidence Intervals

Let's create a mystery box. It contains 25,000 zeros and 1s but we don't know how many of each:

```
n <- 5
box <- rep(sample(0:1, n, replace=T), 25000/n)
```

Sample of Size 100

We don't have time to count all the 0s and 1s so let's take a small sample from the box:

```
n <- 100
x <- sample(box, n, replace = TRUE)
x
```

```
##   [1] 0 1 1 1 1 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1 1 1 1 0 0 0 0 0 1
##  [36] 0 1 0 0 1 0 1 1 0 0 0 1 1 1 0 0 1 0 1 0 0 1 0 1 0 1 0 1 0 0 0 0 0 0 0
##  [71] 0 1 0 0 1 0 1 0 0 1 1 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1
```

The fraction of 1s in our sample is:

```
m <- mean(x)
m
```

```
## [1] 0.35
```

We want a ± on this estimate. Use the observed fraction to estimate the SD of the box:

```
s <- (1 - 0) * sqrt(m * (1-m))  # Using the shortcut method.
s
```

```
## [1] 0.4769696
```

and the SE of the percent:

```
se <- s / sqrt(n)
se
```

```
## [1] 0.04769696
```

A 95% confidence interval on the fraction of 1s is:

```
c(m - 2*se, m + 2*se)
```

```
## [1] 0.2546061 0.4453939
```

What are we 95% confident about? Could the true fraction of 1s be outside this range?

Sample of Size 1000

Let's try a bigger sample:

```r
n <- 1000
x <- sample(box, n, replace = TRUE)
```

The fraction of 1s in our sample is:

```r
m <- mean(x)
m
```

```
## [1] 0.406
```

We want a $\pm$ on this estimate. Use the observed fraction to estimate the SD of the box:

```r
s <- (1 - 0) * sqrt(m * (1-m))  # Using the shortcut method.
s
```

```
## [1] 0.4910845
```

and the SE for the percent:

```r
se <- s / sqrt(n)
se
```

```
## [1] 0.01552946
```

A 95% confidence interval on the fraction of 1s is:

```r
c(m - 2*se, m + 2*se)
```

```
## [1] 0.3749411 0.4370589
```

What are we 95% confident about? Could the true fraction of 1s be outside this range? Is this confidence interval wider or narrower than the first one we computed? Why?

The true fraction of 1s in the box is:

```r
mean(box)
```

```
## [1] 0.4
```

Is this value in your confidence intervals?