Computer Project 4: Hypothesis Testing

Your Name

2020-06-30

z-Test Examples

We think the average of the contents of a box is 50. Let's take some samples then test the hypothesis.

Sample Size = 10

Here are the results of a sample of size 10.

```
n = 10
x = rnorm(n, mean=secret.mean, sd=secret.sd)
x
```

```
##  [1] 58.03581 54.22438 44.67735 41.21284 43.03266 49.97621 48.53443
##  [8] 54.28705 59.17319 49.68922
```

Compute the mean of our sample:

```
m = mean(x)
m
```

```
## [1] 50.28431
```

and the SD of the sample:

```
s = sd(x) * sqrt((length(x)-1)/length(x))
s
```

```
## [1] 5.834993
```

Let's run a z-test to see if 50.2843139 is close enough to 50. First we need the SE for the average:

```
se = s / sqrt(n)
se
```

```
## [1] 1.845187
```

Now compute the z-score:

```
z = (m - 50)/se
z
```

```
## [1] 0.1540841
```

The p-value is the chance of getting a z-score that is as extreme or more extreme than the observed value. *Use either pnorm(z) or 1 - pnorm(z) below to compute the p-value:*

```
# Use pnorm(z) or 1 - pnorm(z) here
```

The null hypothesis is that the difference between 50.2843139 and 50 is due to chance. *What is the alternative hypothesis? Do the data support the null hypothesis or the alternative?*

Sample Size = 100

Here are the results of a sample of size 100

```
n = 100
x = rnorm(n, mean=secret.mean, sd=secret.sd)
x
```

```
##   [1] 42.07436 48.67414 56.35148 49.29005 44.77100 43.15599 40.36281
##   [8] 43.15709 48.37080 50.09788 45.62559 42.74050 48.93534 48.14910
##  [15] 33.50848 42.38003 43.51463 50.17857 54.97672 50.13880 47.73977
##  [22] 43.70527 57.79110 40.62618 46.50926 53.34343 51.85595 51.89446
##  [29] 47.63019 52.44515 38.43207 47.15525 55.60813 59.96824 41.01267
##  [36] 43.09975 39.68737 50.68895 50.22176 46.47982 46.11044 49.37807
##  [43] 52.45944 42.20659 56.96548 52.80743 50.17537 54.01557 50.59173
##  [50] 53.61426 42.85504 50.76119 47.20148 61.82925 44.97424 41.06085
##  [57] 50.27884 47.81138 45.15449 51.30295 46.68604 52.17305 43.83999
##  [64] 52.44442 51.51922 47.24622 45.80142 46.44106 51.70981 57.95443
##  [71] 59.99385 51.65491 46.62763 50.50940 45.50532 52.75492 48.24551
##  [78] 53.61093 39.70719 40.41735 46.18796 45.65025 55.55246 48.49341
##  [85] 45.26322 51.35335 47.06690 54.60386 52.50411 46.00547 50.70987
##  [92] 56.66938 47.84312 49.34030 53.58468 49.06553 48.82679 49.80746
##  [99] 51.06480 47.09299
```

Compute the mean of our sample:

```
m = mean(x)
m
```

```
## [1] 48.65431
```

and the SD of the sample:

```
s = sd(x) * sqrt((length(x)-1)/length(x))
s
```

```
## [1] 5.11107
```

Let's run a z-test to see if 48.6543079 is close enough to 50. First we need the SE for the average:

```
se = s / sqrt(n)
se
```

```
## [1] 0.511107
```

Now compute the z-score:

```
z = (m - 50)/se
z
```

```
## [1] -2.632897
```

The p-value is the chance of getting a z-score that is as extreme or more extreme than the observed value. *Use either pnorm(z) or 1 - pnorm(z) below to compute the p-value:*

```
# Use pnorm(z) or 1 - pnorm(z) here
```

The null hypothesis is that the difference between 48.6543079 and 50 is due to chance. *What is the alternative hypothesis? Do the data support the null hypothesis or the alternative?*

t-Test Example

R has a built-in t-test command. Let's use it to run the $n = 100$ hypothesis test that we did above.

```
t.test(x, mu=50)
```

```
##
##  One Sample t-test
##
## data:  x
## t = -2.6197, df = 99, p-value = 0.01019
## alternative hypothesis: true mean is not equal to 50
## 95 percent confidence interval:
##  47.63505 49.67356
## sample estimates:
## mean of x
##  48.65431
```

The output gives us the t-score, the degrees of freedom, the p-value and a 95% confidence interval.

Use the p-value to explain if the data supports the null or alternative hypothesis. Is 50 included in the confidence interval?