Hypothesis Test for a Single Proportion (The Math Sorcerer)

Luis Cárceles

2024-04-01

See the YouTube play list by The Math Sorcerer.

Introduction

The statistic to be used in a single proportion test is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where $\hat{p}$ is the sample proportion, p_0 is the expected proportion and n is the sample size.

It turns out that z is distributed N(0,1)

Hypothesis test using R: Let n be the sample size. If $n \leq 30$, we use ***binom.test***. If $n > 30$ we will use ***prop.test***

Problems

Problem 1

Suppose 222 subjects are treated with a drug that is used to treat pain and 53 of them developed nausea. Use a 0.1 significance level to test the claim that more than 20% of users develop nausea.

Solution:

1. State null and alternate Hypothesis:

- H_0: $p = 0.20$
- H_1: $p > 0.20$ (Right tailed test)

2. Compute the Test Statistic

```
p10<-0.20
X1<-53
n1<-222
p1bar<-X1/n1
z1<-(p1bar-p10)/sqrt(p10*(1-p10)/n1)
sprintf("The value of the z statistic is: %f", z1)
```

```
## [1] "The value of the z statistic is: 1.442986"
```

3. Compute the p-value

```
p1_value<-1-pnorm(z1, 0, 1)
sprintf("The p-value of the test is: %f", p1_value)
```

```
## [1] "The p-value of the test is: 0.074512"
```

4. Make a Decision

p-value is greater than the level of significance (0.1) then we FAIL to reject H_0.

5. Interpretation of the results

At the 1% of significance level there's not sufficient evidence to support H_1 (our claim)

Extra. Using R for Hypothesis Test

```
prop.test(x=53, n=222, p=0.20, alternative="greater")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  53 out of 222, null probability 0.2
## X-squared = 1.8471, df = 1, p-value = 0.08706
## alternative hypothesis: true p is greater than 0.2
## 95 percent confidence interval:
##  0.1929251 1.0000000
## sample estimates:
##         p
## 0.2387387
```

Problem 2

In a study of cell phone usage and brain hemispheric dominance, an internet survey was e-mailed to 6983 subjects randomly selected from an online group involved with ears. There were 1317 surveys returned. Use a 0.01 significance level to test the claim that the return rate is less than 20%. Use the p-value method and use the normal distribution as an approximation to the binomial distribution.

Solution:

1. State null and alternate Hypothesis:

- H_0: $p = 0.20$
- H_1: $p < 0.20$ (Left tailed test)

2. Compute the Test Statistic

```
p20<-0.20
X2<-1317
n2<-6983
p2bar<-X2/n2
z2<-(p2bar-p20)/sqrt(p20*(1-p20)/n2)
sprintf("The value of the z statistic is: %f", z2)
```

```
## [1] "The value of the z statistic is: -2.381398"
```

3. Compute the p-value

```
p2_value<-pnorm(z2, 0, 1)
sprintf("The p-value of the test is: %f", p2_value)
```

```
## [1] "The p-value of the test is: 0.008624"
```

4. Make a Decision

p-value is lower than the level of significance (0.01), then we reject H_0.

5. Interpretation of the results

At the 1% of significance level there's sufficient evidence to support H_1 (our claim)

Extra. Using R for Hypothesis Test

```
prop.test(x=1317, n=6983, p=0.20, alternative="less")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  1317 out of 6983, null probability 0.2
## X-squared = 5.6, df = 1, p-value = 0.00898
## alternative hypothesis: true p is less than 0.2
## 95 percent confidence interval:
##  0.0000000 0.1964937
## sample estimates:
##         p
## 0.1886009
```

Problem 3

Consider a drug testing company that provides a test for marijuana usage. Among 321 tested subjects, results from 29 subjects were wrong (either a false positive or a false negative). Use a 0.10 significance level to test the claim that less than 10 percent of the test results are wrong.

Solution:

1. State null and alternate Hypothesis:

- H_0: $p = 0.10$
- H_1: $p < 0.10$ (Left tailed test)

2. Compute the Test Statistic

```
p30<-0.10
X3<-29
n3<-321
p3bar<-X3/n3
z3<-(p3bar-p30)/sqrt(p30*(1-p30)/n3)
sprintf("The value of the z statistic is: %f", z3)
```

```
## [1] "The value of the z statistic is: -0.576750"
```

3. Compute the p-value

```
p3_value<-pnorm(z3, 0, 1)
sprintf("The p-value of the test is: %f", p3_value)
```

```
## [1] "The p-value of the test is: 0.282054"
```

4. Make a Decision

p-value is lower than the level of significance (0.10), then we FAIL to reject H_0.

5. Interpretation of the results

At the 10% of significance level there's NO sufficient evidence to support H_1 (our claim)

Extra. Using R for Hypothesis Test

```
prop.test(x=29, n=321, p=0.10, alternative="less")
```

```
##
##  1-sample proportions test with continuity correction
```

```
##
## data:  29 out of 321, null probability 0.1
## X-squared = 0.23399, df = 1, p-value = 0.3143
## alternative hypothesis: true p is less than 0.1
## 95 percent confidence interval:
##  0.0000000 0.1219411
## sample estimates:
##          p
## 0.09034268
```

Problem 4

Trials in an experiment with a polygraph include 97 results that include 23 cases of wrong results and 74 cases of correct result. Use a 0.05 significance level to test the claim that such polygraph results are correct less than 80% of the time. Identify the null hypothesis, alternative hypothesis, tests statistic, p-value, conclusion about the null hypothesis and final conclusion that addresses the original claim. Use the p-value method and the normal distribution as an approximation of the binomial distribution.

Solution:

1. State null and alternate Hypothesis:

- H_0: $p = 0.80$
- H_1: $p < 0.80$ (Left tailed test)

2. Compute the Test Statistic

```r
p40<-0.80
X4<-74
n4<-97
p4bar<-X4/n4
z4<-(p4bar-p40)/sqrt(p40*(1-p40)/n4)
sprintf("The value of the z statistic is: %f", z4)
```

```
## [1] "The value of the z statistic is: -0.913812"
```

3. Compute the p-value

```r
p4_value<-pnorm(z4, 0, 1)
sprintf("The p-value of the test is: %f", p4_value)
```

```
## [1] "The p-value of the test is: 0.180408"
```

4. Make a Decision

p-value is lower than the level of significance (0.05), then we FAIL to reject H_0.

5. Interpretation of the results

At the 5% of significance level there's NO sufficient evidence to support H_1 (our claim)

Extra. Using R for Hypothesis Test

```r
prop.test(x=74, n=97, p=0.80, alternative="less")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  74 out of 97, null probability 0.8
## X-squared = 0.6192, df = 1, p-value = 0.2157
## alternative hypothesis: true p is less than 0.8
```

```
## 95 percent confidence interval:
##  0.0000000 0.8306761
## sample estimates:
##         p
## 0.7628866
```

Problem 5

Consider a sample of 52 football games, where 31 of them were won by the home team. Use a 0.10 significance level to test the claim that the probability that he home team wins is greater than one-half

Solution:

1. State null and alternate Hypothesis:

- H_0: $p = 0.50$
- H_1: $p > 0.50$ (Right tailed test)

2. Compute the Test Statistic

```r
p50<-0.50
X5<-31
n5<-52
p5bar<-X5/n5
z5<-(p5bar-p50)/sqrt(p50*(1-p50)/n5)
sprintf("The value of the z statistic is: %f", z5)
```

```
## [1] "The value of the z statistic is: 1.386750"
```

3. Compute the p-value

```r
p5_value<-1-pnorm(z5, 0, 1)
sprintf("The p-value of the test is: %f", p5_value)
```

```
## [1] "The p-value of the test is: 0.082759"
```

4. Make a Decision

p-value is lower than the level of significance (0.1) then we must reject H_0.

5. Interpretation of the results

At the 10% of significance level there are sufficient evidence to support H_1 (our claim)

Extra. Using R for Hypothesis Test

```r
prop.test(x=31, n=52, p=0.50, alternative="greater")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  31 out of 52, null probability 0.5
## X-squared = 1.5577, df = 1, p-value = 0.106
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.472659 1.000000
## sample estimates:
##         p
## 0.5961538
```

In this case the p-value supplied by prop.test function is slightly greater than the significance level required. This is because it is using a different test statistic (x-squared). The same happens with the binom.test function

```r
binom.test(x=31, n=52, p=0.50, alternative="greater")
```

```
##
##  Exact binomial test
##
## data:  31 and 52
## number of successes = 31, number of trials = 52, p-value = 0.1058
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.4726687 1.0000000
## sample estimates:
## probability of success
##              0.5961538
```