



ACADGILD

SESSION 9: Statistical Inference

Assignment 1

Submitted by: Munmun Ghosal

Login Id: munmun55@gmail.com

(M):+91-8007178659

Data Analytics

Table of Contents

1. Problem Statement.....	3
2. Solution.....	3

1. Problem Statement

1. If Z is norm (mean = 0, sd = 1)
find $P(Z > 2.64)$
find $P(|Z| > 1.39)$
2. Suppose p = the proportion of students who are admitted to the graduate school of the University of California at Berkeley, and suppose that a public relation officer boasts that UCB has historically had a 40% acceptance rate for its graduate school. Consider the data stored in the table UCBA admissions from 1973. Assuming these observations constituted a simple random sample, are they consistent with the officer's claim, or do they provide evidence that the acceptance rate was significantly less than 40%? Use an $\alpha = 0.01$ significance level.

2. Solution

The R-script for the given problem is as follows:

```
# 1. If Z is norm (mean = 0, sd = 1)

# Find P(Z > 2.64)
pnorm(2.64, mean = 0, sd = 1, lower.tail = FALSE)

# Find P(|Z| > 1.39)
1 - (pnorm(1.39, mean = 0, sd=1) - pnorm(-1.39, mean = 0, sd=1))
```

The output of the R-Script (from Console window) is given as follows:

```
> pnorm(2.64, mean = 0, sd = 1, lower.tail = FALSE)
[1] 0.004145301
> 1 - (pnorm(1.39, mean = 0, sd=1) - pnorm(-1.39, mean = 0, sd=1))
[1] 0.1645289
```

Conclusion/Interpretation:

- $P(Z > 2.64)$ 0.004145301
- $P(|Z| > 1.39)$ is 0.1645289

The R-script for the given problem is as follows:

```
View(UCBAdmissions)
class(UCBAdmissions)

# Null hypothesis, H0 is p= 0.40
# Alternative Hypothesis , Ha is p < 0.4

-qnorm(0.99)

A <- as.data.frame(UCBAdmissions)
head(A)

xtabs(Freq ~ Admit, data = A)

# calculate the value of the test statistic.
phat <- 1755/(1755 + 2771)
(phat - 0.4)/sqrt(0.4 * 0.6/(1755 + 2771))

prop.test(1755, 1755 + 2771, p = 0.4, alternative = "less",
          conf.level = 0.99, correct = FALSE)

library(IPSUR)
library(HH)
library(ggplot2)
temp <- prop.test(1755, 1755 + 2771, p = 0.4, alternative = "less", conf.level = 0.99,
                  correct = FALSE)
plot(temp, "Hypoth")
```

The output of the R-Script (from Console window) is given as follows:

```
> View(UCBAdmissions)
> class(UCBAdmissions)
[1] "table"
> -qnorm(0.99)
[1] -2.326348
> A <- as.data.frame(UCBAdmissions)
> head(A)
  Admit Gender Dept Freq
1 Admitted   Male   A  512
2 Rejected   Male   A  313
3 Admitted Female   A   89
4 Rejected Female   A   19
5 Admitted   Male   B  353
6 Rejected   Male   B  207
> xtabs(Freq ~ Admit, data = A)
Admit
Admitted Rejected
```

```

      1755      2771
> phat <- 1755/(1755 + 2771)
> (phat - 0.4)/sqrt(0.4 * 0.6/(1755 + 2771))
[1] -1.680919
> prop.test(1755, 1755 + 2771, p = 0.4, alternative = "less",
+           conf.level = 0.99, correct = FALSE)

```

1-sample proportions test without continuity correction

```

data: 1755 out of 1755 + 2771, null probability 0.4
X-squared = 2.8255, df = 1, p-value = 0.04639
alternative hypothesis: true p is less than 0.4
99 percent confidence interval:
 0.0000000 0.4047326
sample estimates:
      p
0.3877596

```

```

> library(IPSUR)
> library(HH)
Loading required package: lattice
Loading required package: grid
Loading required package: latticeExtra
Loading required package: RColorBrewer
Loading required package: multcomp
Loading required package: mvtnorm
Loading required package: survival
Loading required package: TH.data
Loading required package: MASS

```

Attaching package: 'TH.data'

The following object is masked from 'package:MASS':

geyser

```

Loading required package: gridExtra
> library(ggplot2)

```

Attaching package: 'ggplot2'

The following object is masked from 'package:latticeExtra':

layer

```

> temp <- prop.test(1755, 1755 + 2771, p = 0.4, alternative =
"less", conf.level = 0.99, correct = FALSE)
> #par(mfrow = c(1,1))
> plot(temp, "Hypoth")

```

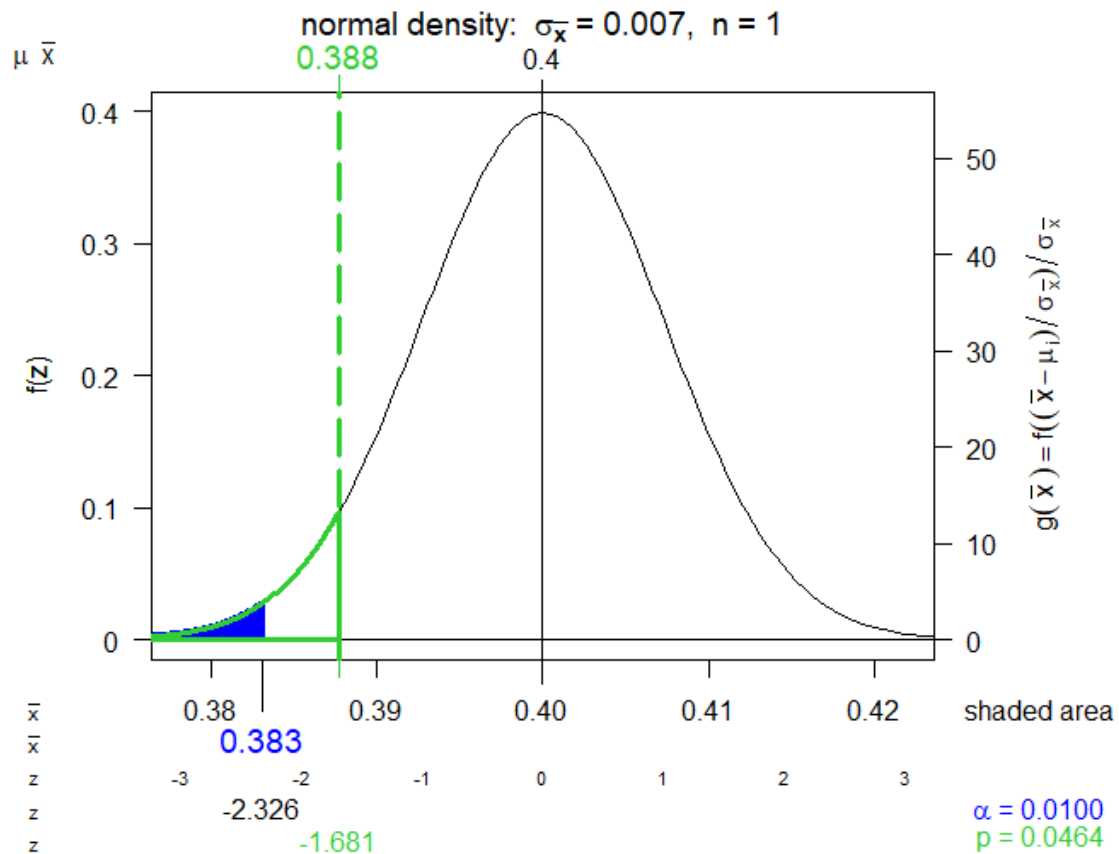
The data UCBAmissions is viewed as follows:

Assignment 9_1.R × UCBAmissions ×

← → ↺ ↻ Filter

	Admit	Gender	Dept	Freq
1	Admitted	Male	A	512
2	Rejected	Male	A	313
3	Admitted	Female	A	89
4	Rejected	Female	A	19
5	Admitted	Male	B	353
6	Rejected	Male	B	207
7	Admitted	Female	B	17
8	Rejected	Female	B	8
9	Admitted	Male	C	120
10	Rejected	Male	C	205
11	Admitted	Female	C	202
12	Rejected	Female	C	391
13	Admitted	Male	D	138
14	Rejected	Male	D	279
15	Admitted	Female	D	131
16	Rejected	Female	D	244
17	Admitted	Male	E	53
18	Rejected	Male	E	138
19	Admitted	Female	E	94
20	Rejected	Female	E	299
21	Admitted	Male	F	22
22	Rejected	Male	F	351
23	Admitted	Female	F	24

Showing 1 to 23 of 24 entries



Conclusion/Interpretation:

- Null hypothesis, H_0 is $p = 0.40$
- Alternative Hypothesis, H_a is $p < 0.4$
- $z_{\alpha} = -2.326348$ is found
- t-statistics is -1.680919 .
- p-value i.e. 0.046 is greater than α i.e. 0.01
- The p-value does not fall into the critical region. We fail to reject the null hypothesis that "the true proportion of students admitted to graduate school is less than 40% and say that the observed data are consistent with the officer's claim at the $\alpha = 0.01$ significance level.