ACADGILD

SESSION 9: Statistical Inference

Assignment 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 UCBAdmissions from 1973. Assuming these observations constituted a simple random sample, are they consistent with the officerâ..s claim, or do they provide evidence that the acceptance

rate was significantly less than 40%? Use an Î± = 0.01 significance level.

SOLUTION :

1. **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 - P(-1.39 < X < 1.39)
1 - (pnorm(1.39, mean = 0, sd=1) - pnorm(-1.39, mean = 0, sd=1))
```

Right-sided area: P (Z ≥ z-score) = 1 – Left-sided area

P(Z > 2.64) = +2.64 is 0.99585 & –2.64 is 0.00415 (right side= 1-leftside)

P(| Z | > 1.39) = +1.39 Is 0.91774 & -1.39 is 0.08226 (right side= 1-leftside)

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

```
Console ~/
> pnorm(2.64, mean = 0, sd = 1, lower.tail = FALSE)
[1] 0.004145301
> #  Find P(|Z| > 1.39)
> #  = 1 - P(-1.39 < X < 1.39)
> 1 - (pnorm(1.39, mean = 0, sd=1) - pnorm(-1.39, mean = 0, sd=1))
[1] 0.1645289
>
```

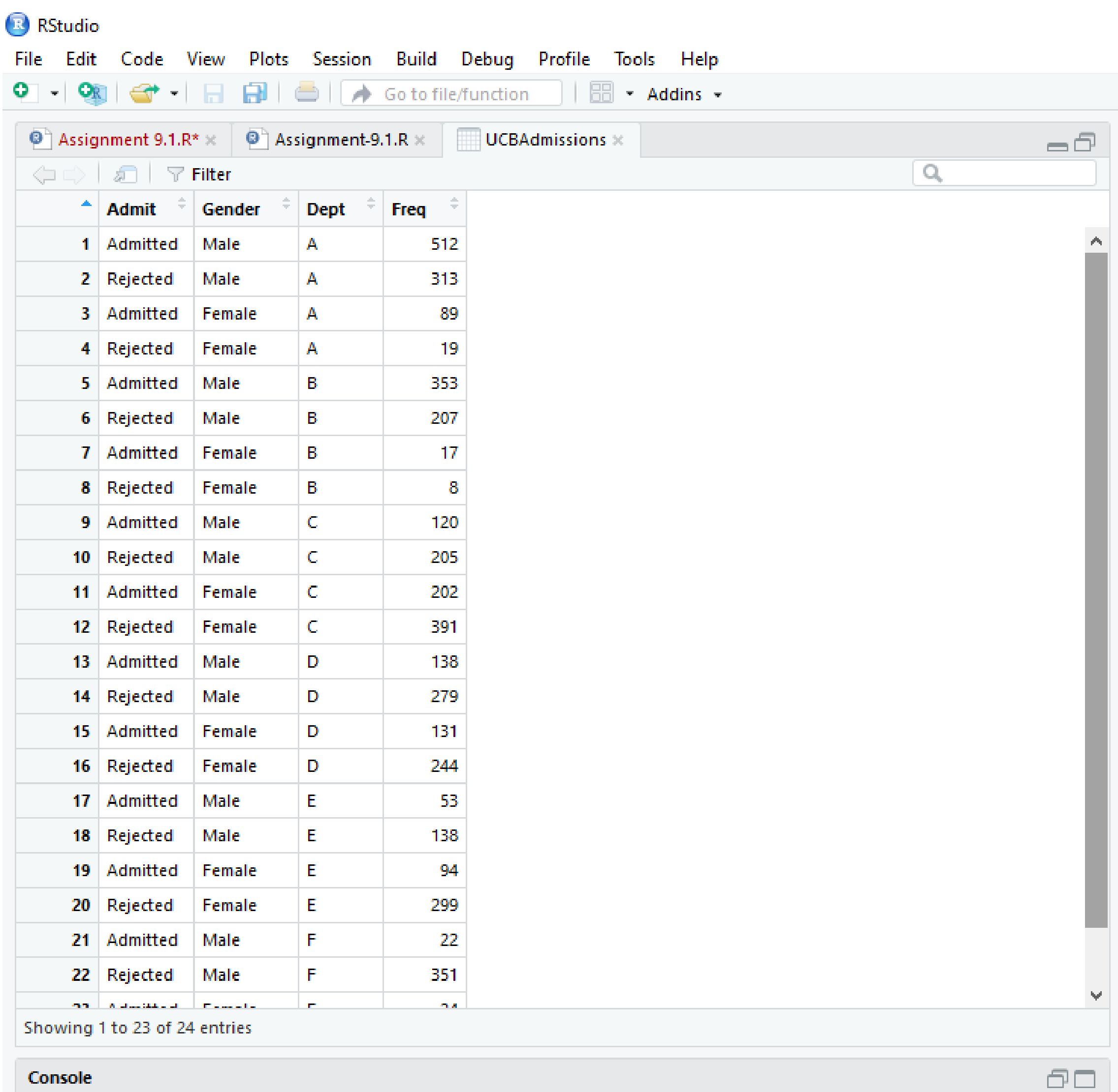
2. The R-script for the given problem is as follows:

```
View(UCBAdmissions)
class(UCBAdmissions)


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

    -qnorm(0.99)   # to find z alpha

    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))
    -qnorm(0.95)
    -qnorm(0.99)
```

```
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:

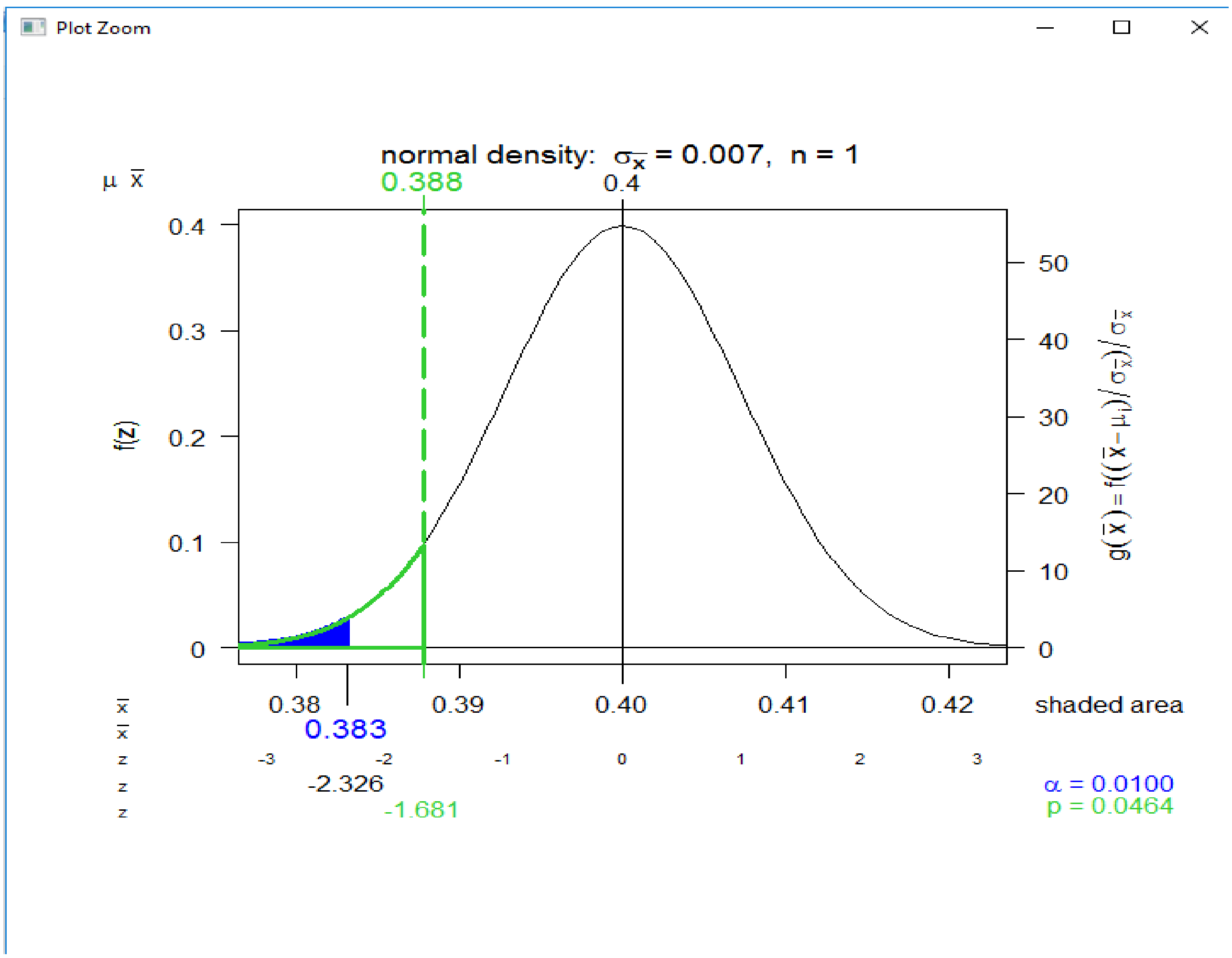
```
> View(UCBAdmissions)
```



	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

```
> class(UCBAdmissions)
[1] "table"

> -qnorm(0.99)   # to find z alpha
[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

> # calculate the value of the test statistic.
> phat <- 1755/(1755 + 2771)
> (phat - 0.4)/sqrt(0.4 * 0.6/(1755 + 2771))
[1] -1.680919
> -qnorm(0.95)
[1] -1.644854
> -qnorm(0.99)
[1] -2.326348

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



INTERPRETATIONS / CONCLUSIONS

- Our null hypothesis in this problem is H0 : p = 0.4 &
- The alternative hypothesis is H1 : p < 0.4.
- We reject the null hypothesis if ˆp is too small, that is, if pˆ − 0.4 √ 0.4(1 − 0.4)/n < −zα, where α = 0.01 and −z0.01 is
- z alpha = -2.326348 is found
- t-statistics is -1.680919.
- p- value i.e. 0.046 is greater than alpha i.e. 0.01
- -qnorm(0.99) [1] -2.326348

There is no mathematical difference between the errors, however. The bottom line is that we choose one type of error to control with an iron fist, and we try to minimize the probability of making the other type. That being said, null hypotheses are often by design to correspond to the "simpler" model, so it is often easier to analyze (and thereby control) the probabilities associated with Type I Errors.

- Now we calculate the value of the test statistic.

```
> phat <- 1755/(1755 + 2771)
> (phat - 0.4)/sqrt(0.4 * 0.6/(1755 + 2771))
[1] -1.680919
```

- Our test statistic is not less than −2.32,so it does not fall into the critical region.
- Therefore, 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 α = 0.01 significance level.

- We are going to do the example all over again. Everything will be exactly the same except for one change. Suppose we choose significance level α = 0.05 instead of α = 0.01. Are the 1973 data consistent with the officer's claim? Our null and alternative hypotheses are the same. Our observed test statistic is the same: it was approximately -1.68. But notice that our critical value has changed: $\alpha = 0.05$ and $-z_{0.05}$ is

```
> -qnorm(0.95)
[1] -1.644854
```

- Our test statistic is less than -1.64 so it now falls into the critical region! We now reject the null hypothesis and conclude that the 1973 data provide evidence that the true proportion of students admitted to the graduate school of UCB in 1973 was significantly less than 40%. The data are not consistent with the officer's claim at the $\alpha = 0.05$ significance level.
- If we choose $\alpha = 0.05$ then we reject the null hypothesis, but if we choose $\alpha = 0.01$ then we fail to reject the null hypothesis. Our final conclusion seems to depend on our selection of the significance level. This is bad; for a particular test, we never know whether our conclusion would have been different if we had chosen a different significance level.