

Assignment 9.1

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

Find $P(Z > 2.64)$

Find $P(|Z| > 1.39)$

Answer

Right-sided area: $P(Z \geq z\text{-score}) = 1 - \text{Left-sided area}$

$P(z > 2.64)$ = +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)

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 $\hat{\alpha} = 0.01$ significance level.

Our null hypothesis in this problem is $H_0: p = 0.4$ and the alternative hypothesis is $H_1: p < 0.4$.

We reject the null hypothesis if \hat{p} is too small, that is, if $\hat{p} - 0.4 \leq -z_{\alpha} \sqrt{0.4(1 - 0.4)/n}$, where $\alpha = 0.01$ and $-z_{0.01}$ is

`> -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.

TESTS FOR PROPORTIONS 233 Our only remaining task is to find the value of the test statistic and see where it falls relative to the critical value. We can find the number of people admitted and not admitted to the UCB graduate school with the following.

```
> A <- as.data.frame(UCBAdmissions)
```

```
> head(A)
```

	Admit <fctr>	Gender <fctr>	Dept <fctr>	Freq <dbl>
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
6 rows				

```
> xtabs(Freq ~ Admit, data = A) Admit
```

```
Admit
Admitted Rejected
1755      2771
1755      2771
```

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 $\alpha = 0.01$ significance level.

We are going to do Example all over again. Everything will be exactly the same except for one change. Suppose we choose significance level $\alpha = 0.05$ instead of $\alpha = 0.01$. 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. What is going on, here? 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.

R script

```
A <- as.data.frame(UCBAdmissions)
head(A)
xtabs(Freq ~ Admit, data = A)
phat <- 1755/(1755 + 2771)
(phat - 0.4)/sqrt(0.4 * 0.6/(1755 + 2771))
-qnorm(0.95)
-qnorm(0.99)
```

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

```
Admit
Admitted Rejected
      1755      2771
[1] -1.680919
[1] -1.644854
[1] -2.326348
```