MATH 1342

Section 8.4

Comparing Two Proportions

When comparing two population proportions in an inference test, we use a **two-sample *z* test** for the proportions.

The null and alternate hypotheses would be:

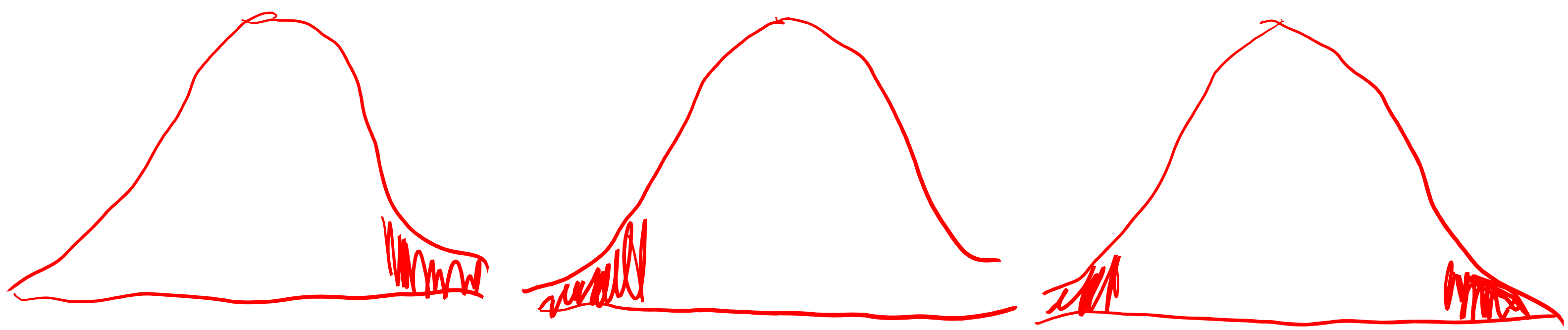
$$H_0: p_1 = p_2 \quad H_a: p_1 > p_2$$

or

$$H_0: p_1 = p_2 \quad H_a: p_1 < p_2$$

or

$$H_0: p_1 = p_2 \quad H_a: p_1 \neq p_2$$



qnorm(1-a)
1-pnorm(z)

qnorm(a)
pnorm(z)

qnorm(a/2)
2*pnorm(z)

The assumptions are the same as for a confidence interval for the difference of two proportions:

1. Both samples must be independent SRSs from the populations of interest.
2. The population sizes are both at least ten times the sizes of the samples.
3. The number of successes and failures in both samples must all be ≥ 10.

#3: $n_1\hat{p}_1 \geq 10$

$n_1(1-\hat{p}_1) \geq 10$

$n_2(1-\hat{p}_2) \geq 10$

$n_2\hat{p}_2 \geq 10$

And the test statistic is:

$=0$

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$$

If p_1 and p_2 are unknown, we will use $\hat{p}_1$ and $\hat{p}_2$ to approximate standard deviation. When we substitute $\hat{p}_1$ and $\hat{p}_2$ into standard deviation "formula," this gives us the standard error of

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

Example:

Group 1: (Honors): n = 100, x = 18
Group 2: (Academic): n =200, x = 32

alpha = 5% = 0.05

Is the proportion of left-handed students higher in honors classes than in academic classes? Two hundred academic and one hundred honors students from grades 6-12 were selected throughout a school district and their left or right handedness was recorded. The sample information is:

	Honors	Academic
Sample size	100	200
Number of left-handed students	18	32

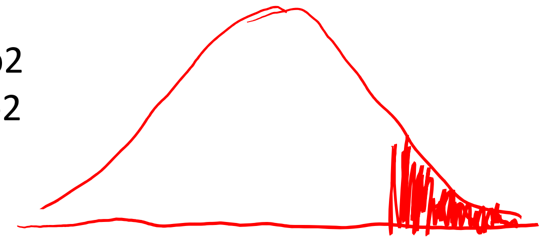
Is there sufficient evidence at the 5% significance level to conclude that the proportion of left-handed students is greater in honors classes?

```
> phat1=18/100
> phat1
[1] 0.18
```

```
> phat2=32/200
> phat2
[1] 0.16
```

Workspace

Ho: p1 = p2
Ha: p1 > p2



Rejection Region: z > 1.645

```
 100*phat1
1] 18
 100*(1-phat1)
1] 82
 200*(phat2)
1] 32
 200*(1-phat2)
1] 168
```

All are greater than 10.

```
qnorm(1-0.05)
1] 1.644854
```

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Test statistic does not fall within the rejection region. (FRHo)

Z =
```
(phat1-phat2)/sqrt(phat1*(1-phat1)/100+phat2*(1-phat2)/200)
1] 0.4315319
```

(If you get an error statement, count your left and right parenthesis. You probably made a mistake there)

p-value:

```
1-pnorm(0.4315319)
1] 0.3330408
```

= 33.3%

$\alpha = 5\%$

$p > \alpha$

FRH_0

Another Example:

North Carolina State University looked at the factors that affect the success of students in a required chemical engineering course. Students must get a C or better in the course in order to continue as chemical engineering majors. There were 65 students from urban or suburban backgrounds, and 52 of these students succeeded. Another 55 students were from rural or small-town backgrounds; 30 of these students succeeded in the course. Test the claim to see if there is a difference between the urban and suburban success rates at the 5% level.

Group 1: (Urban)
n = 65
x = 52

Group 2: (Rural)
n = 55
x = 30

Ho: p1 = p2
Ha: p1 ≠ p2

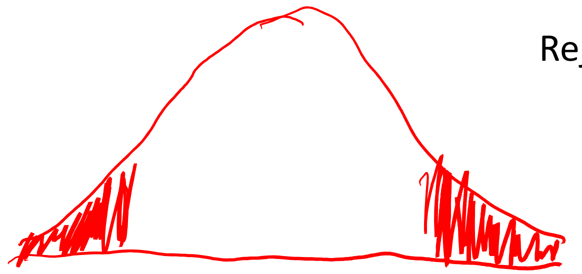
```
> phat1=52/65
> phat1
[1] 0.8
> phat2=30/55
> phat2
[1] 0.5454545
```

alpha = 5% = 0.05

Workspace

```
> 65*phat1
[1] 52
> 65*(1-phat1)
[1] 13
> 55*phat2
[1] 30
> 55*(1-phat2)
[1] 25
```

All are greater than 10



Rejection Region: z < -1.960 or z > 1.960

```
qnorm(0.05/2)
.] -1.959964
```

```
-
(phat1-phat2)/sqrt(phat1*(1-phat1)/65+phat2*(1-phat2)/55)
.] 3.049066
```

z = 3.049 (z falls within our rejection region)

p-value test:

```
2*pnorm(-3.049066)
1] 0.00229554
```

P = 0.23% $\alpha = 5\%$

$P < \alpha$

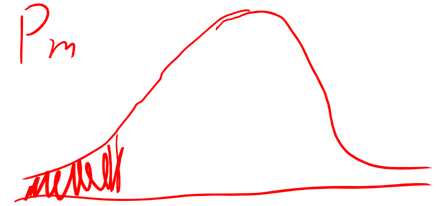
R H_0

Popper 30:

phat1 = 0.12, n1 = 243 phat2 = 0.15, n2 = 322 alpha = 0.05

Insurance companies claim there is no difference between accidents in first year drivers that are male or female. You believe that females are less accident prone. Based on car insurance companies, 12% of women drivers are in an accident (sample size of 243) in their first year of driving and 15% of men (sample size of 322) are in an accident in their first year of driving. You wish to test this at a significance level of 5%.

$H_o: P_f = P_m$



1. Determine the alternate hypothesis:

a. $p_f < p_m$ b. $p_f > p_m$ c. $p_f = p_m$ d. $p_f \neq p_m$

2. Determine the rejection region.

a. $z < -1.645$ b. $z > 1.645$ c. $z < -1.960$ and $z > 1.960$

```
qnorm(0.05)
1] -1.644854
```

3. Determine the value of the test statistic.

a. -1.023 b. -102.3 c. -4.828 d. -1.041

4. Based on this, what is the conclusion?

a. RH_o b. FRH_o

```
(0.12-0.15)/sqrt(0.12*(1-0.12)/243+0.15*(1-0.15)/322)
] -1.040982
```

5. Determine the p-value:

a. 1.49% b. .149% c. 14.9% d. 1.041%

6. Based on this, what is our conclusion?

a. RH_o b. FRH_o

```
pnorm(-1.040982)
] 0.148942
```

$P = 14.89\%$

$p > \alpha$