MATH 3307
Lesson 28

Confidence Interval for the Difference of Two Proportions

This would involve the percent of difference between two groups.

For example, you want to determine the difference in success rates between students from high-income families, versus students with low-income families.

As an additional example, you want to determine the difference in heights between people of two different races.

Confidence Interval for the Difference of Two Proportions

The assumptions that need to be satisfied for a two-sample proportion are slightly different than those for a one-sample.

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.

$n_1*\hat{p}_1$
$n_2*\hat{p}_2$
$n_1*(1-\hat{p}_1)$
$n_2*(1-\hat{p}_2)$

To make the comparison, we will need to find the difference of the two proportions, $\hat{p}_1 - \hat{p}_2$. The standard error for this difference is $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$. So our formula for the confidence interval is:

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

In research, p1 as the larger of the two proportions.
For our quizzes, the first sample mentioned, will be considered first group. (You may end up with negative values)

Confidence Intervals

General formula: *statistic* ± *margin of error*

One-sample z-test: $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$

Two-proportion z-test: $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

One-sample t-test: $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

One-proportion z-test: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Two-sample z-test: $(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Two-sample t-test: $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Example:

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Example:

The National Research Council of the Philippines reported that 210 of 361 members in biology are women, but only 34 of 86 members in mathematics are women. Establish a 96% confidence interval estimate of the difference in proportions of women in biology and mathematics in the Philippines. Interpret your results.

Biology:
x = 210
n = 361

Math:
x = 34
n = 86

CL = 96%

```
> phat1=210/361
> 361*phat1
[1] 210
> 361*(1-phat1)
[1] 151
> phat2=34/86
> 86*phat2
[1] 34
> 86*(1-phat2)
[1] 52
```

All assumptions are met (all values >10)

```
> (phat1-phat2)-qnorm(1.96/2)*sqrt(phat1*(1-phat1)/361+phat2*(1-phat2)/86)
[1] 0.06567434
> (phat1-phat2)+qnorm(1.96/2)*sqrt(phat1*(1-phat1)/361+phat2*(1-phat2)/86)
[1] 0.3070629
```

[0.0657, 0.3071]

We are 96% confident that the difference in proportions between women in biology and women in math in the Philippines is between 6.6% and 30.7%

Popper 24:

Urban: x = 20, n = 75

Rural: x = 18, n = 36

```
> phat1=20/75
> phat1
[1] 0.2666667
> 75*phat1
[1] 20
> 75*(1-phat1)
[1] 55
```

```
> phat2=18/36
> paht2
Error: object
> phat2
[1] 0.5
> phat2*36
[1] 18
> 36*(1-phat2)
[1] 18
```

A study was conducted to observe the difference in probability of having large families in urban versus rural settings. A large family was defined as a family having more than 3 children. Out of 75 urban families surveyed, 20 of them had large families. Of the 36 rural families surveyed, 18 of them had large families.

1. What is the proportion of large families for urban environments?

   a. 0.267 b. 0.5 c. 0.24 d. 0.556

2. Does the urban environment match the needed criteria for this process?

   a. Yes b. No

3. What is the proportion of large families for the rural environments?

   a. 0.267 b. 0.5 c. 0.24 d. 0.556

4. Does the rural environment match the needed criteria for this procedure?

   a. Yes b. No

Popper 24: Continued

A study was conducted to observe the difference in probability of having large families in urban versus rural settings. A large family was defined as a family having more than 3 children. Out of 75 urban families surveyed, 20 of them had large families. Of the 36 rural families surveyed, 18 of them had large families.

5. What is the difference in the proportion of large families between the two environments?

   a. 0.55 b. 2.3 c. 0.233 d. 0.556

```
phat2-phat1
1] 0.2333333
```

6. What is the standard error of the difference in these two environments?

   a. 0.5665 b. 0.0977 c. 0.0882 d. 0.7333

```
sqrt(phat1*(1-phat1)/75+phat2*(1-phat2)/36)
1] 0.09773358
```

7. What is the margin of error that will give a 95% confidence level?

   a. 0.95 b. 0.164 c. 0.1916 d. -1.6448

```
qnorm(1.95/2)*sqrt(phat1*(1-phat1)/75+phat2*(1-phat2)/36)
1] 0.1915543
```

8. What is the 95% confidence interval for the difference between these environments?

   a. [0.23, 0.34] b. [0.34, 0.95] c. [0.04, 0.42] d. [0.13, 0.95]

```
(phat2-phat1)-qnorm(1.95/2)*sqrt(phat1*(1-phat1)/75+phat2*(1-phat2)/36)
1] 0.04177904
(phat2-phat1)+qnorm(1.95/2)*sqrt(phat1*(1-phat1)/75+phat2*(1-phat2)/36)
1] 0.4248876
```