Test of
Difference in Proportions

Other than testing the ratio of two probabilities of an event (proportions), we can also test the difference of probabilities (proportions) between the exposed and unexposed groups.

Suppose:

- π_E is the probability of the event in the exposed group
- $\pi_{\bar{E}}$ is the probability of the event in the unexposed group
- n_E is the sample size for the exposed group
- $n_{\bar{E}}$ is the sample size for the unexposed group

We are testing:

$$H_0: \pi_E = \pi_{\bar{E}}$$
$$H_1: \pi_E \neq \pi_{\bar{E}}$$

- The test statistic follows a normal distribution when the sample size is large:

$$\hat{\pi}_E - \hat{\pi}_{\bar{E}} \sim N\left(\pi_E - \pi_{\bar{E}}, \sqrt{\frac{\pi_E(1-\pi_E)}{n_E} + \frac{\pi_{\bar{E}}(1-\pi_{\bar{E}})}{n_{\bar{E}}}}\right)$$

- The estimated difference in proportions is : $\hat{\pi}_E - \hat{\pi}_{\bar{E}} = \frac{n_{11}}{n_1} - \frac{n_{21}}{n_2}$
- The estimated variance is: $v = \frac{\hat{\pi}_E(1-\hat{\pi}_E)}{n_1} + \frac{\hat{\pi}_{\bar{E}}(1-\hat{\pi}_{\bar{E}})}{n_2}$
- We can also compute a (1 - α) % confidence interval:
- $(\hat{\pi}_E - \hat{\pi}_{\bar{E}}) \pm z_{\frac{\alpha}{2}}\sqrt{v}$

Example:

- British TB patients were randomized to receive either streptomycin and bed rest (streptomycin) or simply bed rest (control). The outcome of interest is whether or not the patient showed improvement. The data are given in the table below.

		Improvement in Condition		
		Yes	No	
Treatment Group	Streptomycin (E)	38	17	55
	Control ($\overline{E}$)	17	35	52
		55	52	107

Solution

- Let $\hat{\pi}_E$ be the proportion of all TB patients who were given the streptomycin and showed improvement at 6 months, $\hat{\pi}_{\bar{E}}$ is the proportion of patients who show improvement but only receiving bed rest.
- $\hat{\pi}_E = \frac{38}{55} = 0.691$ $\qquad$ $\hat{\pi}_{\bar{E}} = \frac{17}{52} = 0.327$
- $v = \frac{0.691 * 0.309}{55} + \frac{0.327 * 0.673}{52} = 0.0081$
- **95% confidence interval :** $(0.691 - 0.327) \pm 1.96\sqrt{0.0081} \equiv (0.187, 0.540)$
- Interpretation: We are 95% confident that this interval covers the difference of the true proportions of improvement between those receiving streptomycin and those receiving control. Since the 95% confidence interval does not include 0, the null value, we can reject the null hypothesis at the 0.05 level and conclude that the streptomycin effect is different than the effect of bed rest alone. Moreover, as the interval exceeds 0, it appears to provide a better effect.

STATA example

- Recall the first example where we calculate the difference and confidence interval by hand, STATA can help us do the work:

```
. prtesti 55 38 52 17, count

Two-sample test of proportions                    x: Number of obs =        55
                                                  y: Number of obs =        52
------------------------------------------------------------------------------
             |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .6909091   .0623121                      .5687797    .8130385
           y |   .3269231   .0650509                      .1994256    .4544205
-------------+----------------------------------------------------------------
        diff |    .363986   .0900801                      .1874323    .5405397
             |  under Ho:   .0966736     3.77   0.000
------------------------------------------------------------------------------
        diff = prop(x) - prop(y)                                  z =   3.7651
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.9999         Pr(|Z| > |z|) = 0.0002          Pr(Z > z) = 0.0001
```