Comparisons:
Looking at differences between 2 means or 2 proportions

- Answers questions about comparisons between two groups or two measurements for the same group

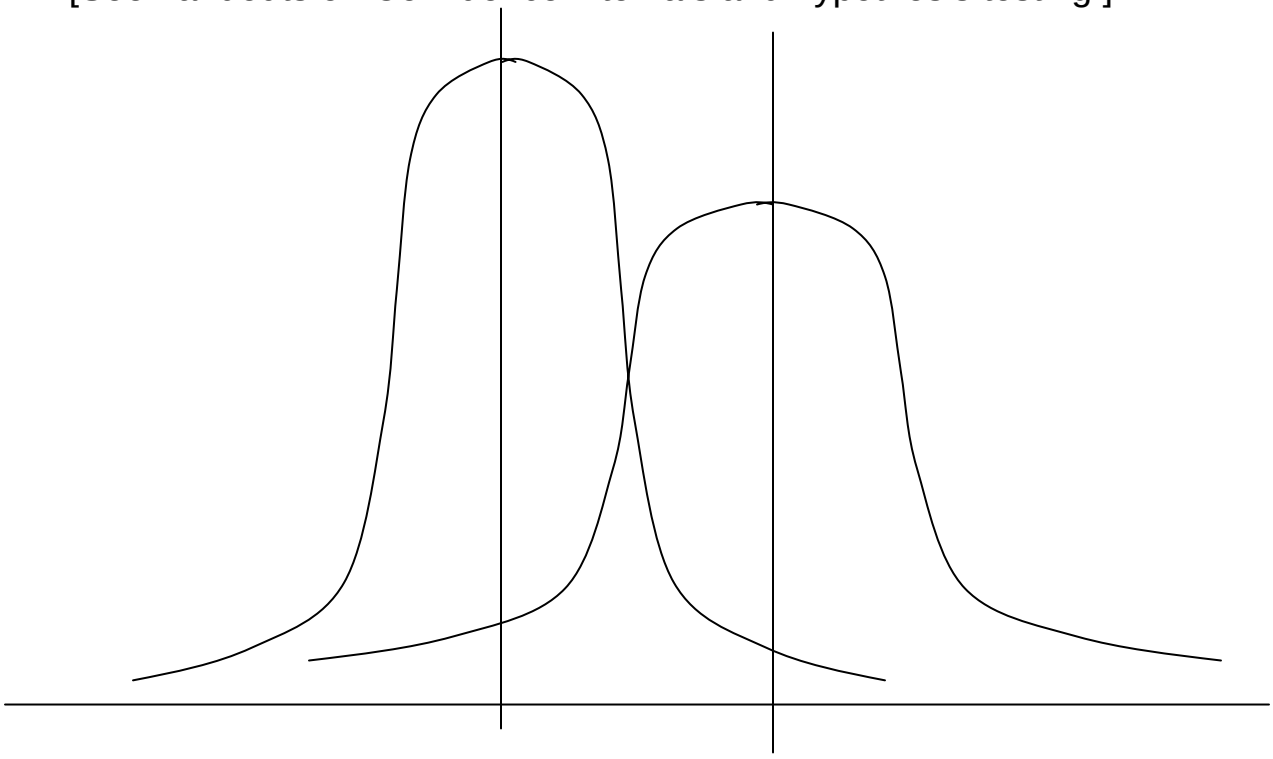
- Focuses on Difference between two outcomes rather than level of one outcome

- Uses same tools (confidence intervals, hypothesis tests, p values)

- Variation in outcomes is due to randomness of which members of the sub-populations end up in the two sub-samples

- Outcomes can be for Independent sub-samples or paired sub-samples (these have 2 different sets of tools)

[See handouts on Confidence intervals and hypothesis testing.]



King County is interested in the implications of applying sales tax to out of state purchases for their revenue. Do men and women, on average, make different numbers of out of state purchases?

Group Statistics

	PERSON*S SEX	N	Mean	Std. Deviation	Std. Error Mean
# TIMES MADE INTERNET PURCHASE LAST YR	1.MALE	394	10.9340	24.95952	1.25744
	2.FEMALE	577	10.1473	44.24820	1.84208

Confidence Interval:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = (10.93 - 10.15) \pm 1.96\sqrt{\frac{(24.96)^2}{394} + \frac{(44.24)^2}{577}} = .79 \pm 4.37$$

So, we estimate that, in King County, men make somewhere between 3.6 fewer and 5.2 more purchases than do women, on average. [A better way to put this is that our survey suggests that men and women make a similar number of purchases.]

Hypothesis test:

1) H_0: $\mu_1-\mu_2=0$ H_a: $\mu_1-\mu_2\neq 0$

2) Find critical value (cut-off)
If use 5% significance, Zc=1.96

$$z = \frac{\left(\overline{x_1} - \overline{x_2}\right) - \left(\mu_1 - \mu_2\right)_{null}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\left(10.93 - 10.15\right) - \left(0\right)}{\sqrt{\frac{(24.96)^2}{394} + \frac{(44.24)^2}{577}}} = .353$$

3) |Z| < Zc so can't reject Ho p= 2x(.5-.1368)=.726

So survey can't rule out that men and women in King Co. average similar levels of out of state purchases.

```
T-TEST
  GROUPS=q2p6(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=q8p8
  /CRITERIA=CIN(.95) .
```

T-Test

Group Statistics

	PERSON*S SEX	N	Mean	Std. Deviation	Std. Error Mean
# TIMES MADE INTERNET PURCHASE LAST YR	1.MALE	394	10.9340	24.95952	1.25744
	2.FEMALE	577	10.1473	44.24820	1.84208

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
# TIMES MADE INTERNET PURCHASE LAST YR	Equal variances assumed	.131	.718	.320	969	.749	.78670	2.45969	-4.04024	5.61364
	Equal variances not assumed			.353	939.033	.724	.78670	2.23034	-3.59033	5.16372

```
RECODE
  q8p8
  (0=Copy)  (SYSMIS=SYSMIS)  (1 thru Highest=1)  INTO  netpurch .
VARIABLE LABELS netpurch 'dummy for made intern purch'.
EXECUTE .

RECODE
  region
  (3=1)  (SYSMIS=SYSMIS)  (ELSE=0)  INTO  kingco .
VARIABLE LABELS kingco 'dummy for king co'.
EXECUTE .

T-TEST
  GROUPS=kingco(0 1)
  /MISSING=ANALYSIS
  /VARIABLES=netpurch
  /CRITERIA=CIN(.95) .
```

T-Test

Group Statistics

	dummy for king co	N	Mean	Std. Deviation	Std. Error Mean
dummy for made intern purch	.00	3387	.5648	.49586	.00852
	1.00	971	.7024	.45745	.01468

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
dummy for made intern purch	Equal variances assumed	376.009	.000	-7.751	4356	.000	-.13756	.01775	-.17236	-.10277
	Equal variances not assumed			-8.104	1678.954	.000	-.13756	.01697	-.17085	-.10427

Are people in King County more likely to make out of state purchases than people in other regions?

Confidence Interval:

$$(\hat{p}_1-\hat{p}_2)\pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$=(.56-.70)\pm 1.96\sqrt{\frac{.56(1-.56)}{3387}+\frac{.70(1-.70)}{971}}$$

$$=-.14\pm(1.96)(.017)=-.17\,to-.10$$

Hypothesis test:

1) H_0: $p_1-p_2=0$ H_a: $p_1-p_2\neq 0$

2) Find critical value (cut-off)
If use 5% significance, Zc=1.96

$$\hat{p}=\frac{x_1+x_2}{n_1+n_2}=.595$$

$$z=\frac{(\hat{p}_1-\hat{p}_2)-D_{null}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{(.56-.70)-0}{\sqrt{.595(1-.595)\left(\frac{1}{3387}+\frac{1}{971}\right)}}=-7.75$$

3) |Z| > Zc so can reject Ho p<.001