“We cannot solve problems

by using the same kind of thinking

that we used when we created them.”

– Albert Einstein

COHEN CHAP 7. T-TEST FOR 2 INDEPENDENT SAMPLE MEANS

For EDUC/PSY 6600

INTRO

Same continuous DV compared across 2 independent (random) samples

Is there a significant difference between the 2 group means?

- Do 2 samples come from *different* normal distributions with the same mean?

aka…

- Independent-groups design
- Between-subjects design
- Between-groups design
- Randomized-groups design

STEPS OF A HYPOTHESIS TEST

$$H_0: Diff_\mu = 0 \quad \mu_1 - \mu_2 = 0 \quad \mu_1 = \mu_2$$
$$H_1: Diff_\mu \neq 0 \quad \mu_1 - \mu_2 \neq 0 \quad \mu_1 \neq \mu_2$$

"t-Test Statistic": $t = \dfrac{est - hyp}{SE_{est}}$

1) State the Hypotheses (Null & Alternative)
2) Select the Statistical Test & Significance Level
   - α level
   - One vs. Two tails
3) Select random samples and collect data
4) Find the region of Rejection
   - Based on α & # of tails
5) Calculate the Test Statistic
   - Examples include: z, t, F, χ^2
6) Make the Statistical Decision

*even use z, if N > 100'ish

Separate Variance t-Test

(need HOV)

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

$$\boldsymbol{df} = n_1 + n_2 - 2$$

Pooled-Variance t-Test

(different sample sizes)

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{{s_p}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$min(n_1, n_2) - 1 < \boldsymbol{df} < n_1 + n_2 - 2$$

$$s^2 = \frac{ss}{n-1}$$
$${s_p}^2 = \frac{SS_1 + SS_2}{n_1 + n_2 - 2}$$

EXAMPLE 1

Group 1 - Drug	Group 2 - Placebo
11	11
1	11
0	5
2	8
0	4

depression

1 Drug	N	Valid	5
		Missing	0
	Mean		2.80
	Std. Deviation		4.658
2 Placebo	N	Valid	5
		Missing	0
	Mean		7.80
	Std. Deviation		3.271

In order to assess the efficacy of a new antidepressant drug, 10 clinically depressed participants are randomly assigned to one of two groups. Five participants are assigned to Group 1, which is administered the antidepressant drug for 6 months. The other 5 participants are assigned to Group 2, which is administered a placebo during the same 6 month period

Assume that prior to introducing the treatments, the experimenter confirmed that the level of depression in the 2 groups was equal

After 6 months, all participants are rated by a psychiatrist (blind to participant assignment) on their level of depression

EXAMPLE 1

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
depression	Equal variances assumed	.200	.667	-1.964	8	.085	-5.000	2.546	-10.870	.870
	Equal variances not assumed			-1.964	7.173	.089	-5.000	2.546	-10.990	.990

After 6 months, the five participants taking the drug scored numerically lower on the depression scale (M = 2.80, SD = 4.66), compared their five counter parts taking placebo (M = 7.80, SD = 3.27). To test the effectiveness of the drug at reducing depression, an independent samples *t*-test was performed. The distribution of depression scores were sufficiently normal for the purposes of conducting a t-test (i.e. skew < |2.0| and kertosis < |9.0|; Schmider, Sigler, Danay, Beyer, & Buhner, 2010). Additionally, the assumption of homogeneity of variances was tested and satisfied via Levene's *F*-test, *F*(4, 4) = .20, p = .667. The independent samples *t*-test did not find a statistically significant effect, *t*(8) = -1.964, p = .085. Thus, there is no evidence this drug reduces depression.

ASSUMPTIONS (SIMILAR TO 1-SAMPLE T-TESTS)

1. BOTH Samples were drawn **INDEPENDENTLY** at **random** (at least as representative as possible)

   Nothing can be done to fix NON-representative samples!

   Can not statistically test…violation: paired-samples t-test

2. The variable has a **NORMAL** distribution, for BOTH population

   Not as important if the sample is large (Central Limit Theorem)

   IF the sample is far from normal &/or small n, might want to transform variables

   Look at plots: histogram, boxplot, & QQ plot (straight 45° line) ← sensitive to outliers!!!

   Skewness & Kurtosis: Divided value by its *SE* & > ± 2 indicates issues

   Shapiro-Wilks test (small N): $p < .05$ → not normal

   Kolmogorov-Smirnov test (large N):

3. HOV = **Homogeneity of Variance**: BOTH populations have the sample spread

   use Levene's F-test (null= HOV)

RANDOM ASSIGNMENT

Random assignment to groups ↓ experimenter biases

- Cases are enumerated
- Numbers drawn and assigned to group in any of several ways

Does not ensure equality of group characteristics

Experiment: Random assignment & manipulation of IV

- Treatment vs. control or 2 treatment groups

Quasi-experiment: Either randomization or manipulation

Non-experiment: Neither randomization or manipulation

- Participants self-select or form naturally occurring groups

VIOLATIONS OF ASSUMPTIONS

Equal groups: Violations 'hurt' less

- Heterogeneity of variance
  - Small effects, p-value inaccurate $\pm$ .02
- Non-normality
  - Small effects, p-value inaccurate $\pm$ .02
  - However: If samples are highly skewed or are skewed in opposite directions p-values can be *very* inaccurate
- Both
  - Moderate effects if N is large, p-value can be inaccurate
  - Large effects if N is small, p-value can be *very* inaccurate

ALTERNATIVES (ASSUMPTIONS VIOLATED)

Violation of normality or ordinal DV

- Two Sample Wilcoxon test (aka, Mann-Whitney *U* Test)

Sample re-use methods

- Rely on empirical, rather than theoretical, probability distributions
  - Exact statistical methods
  - Permutation and randomization tests
  - Bootstrapping

VIOLATIONS OF ASSUMPTIONS

<u>Unequal groups: Violations 'hurt' more</u>

- Heterogeneity of variance
  - Large effects
- Non-normality
  - Large effects
- Both
  - <u>Huge</u> effects

- p-values can be **very** inaccurate with unequal ns and violations of assumptions, especially when N is small

CONFIDENCE INTERVALS

95% *CI* for <u>difference</u> between means: μ_1 - μ_2

Rearrange independent-samples *t*-test formula

$$CI_{1-\alpha} = \left(\bar{X}_1 - \bar{X}_2\right) \pm t_{\alpha/2} * (s_{\bar{X}_1 - \bar{X}_2})$$

Estimation as NHST

- If H_0: $\mu_1 = \mu_2$ and if *CI* does NOT contain 0, reject H_0
- If H_0: $\mu_1 = \mu_2$ and if *CI* does contain 0, fail to reject H_0

Compute for in-class example

EXAMPLE 2

An industrial psychologist is investigation the effects of different types of motivation on the performance of simulated clerical tasks. The 10 participants in the "individual motivation" sample are told that they will be rewarded according to how many tasks they successfully complete. The 12 participants in the "group motivation" sample are told that they will be rewarded according to the average number of tasks completed by all the participants in their sample. The number of tasks completed by each participant are as follows:

- Individual Motivation: 11, 17, 14, 12, 11, 15, 13, 12, 15, 16
- Group Motivation: 10, 15, 4, 8, 9, 14, 6, 15, 7, 11, 13, 5

Group Statistics

	Motivation	N	Mean	Std. Deviation	Std. Error Mean
Tasks	1 Individual	10	13.60	2.119	.670
	2 Group	12	9.75	3.888	1.122

EXAMPLE 2

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
Tasks	Equal variances assumed	4.829	.040	2.797	20	.011	3.850	1.376	.979	6.721
	Equal variances not assumed			2.946	17.518	.009	3.850	1.307	1.099	6.601

The number of tasks completed was numerically higher among the individually motivated group (n = 10, M = 13.60, SD = 2.12), compared to the individuals being motivated by the groups results (n = 12, M = 9.75, SD = 3.89). To test the difference in mean productivity, an independent samples *t*-test was performed. The distribution of depression scores were sufficiently normal for the purposes of conducting a t-test (i.e. skew < |2.0| and kertosis < |9.0|; Schmider, Sigler, Danay, Beyer, & Buhner, 2010). Additionally, the assumption of homogeneity of variances was tested and rejected via Levene's *F*-test, *F*(9, 11) = 4.829, p = .040. The independent samples, separate variances *t*-test found a statistically significant effect, *t*(17.52) = 2.946, p = .009. Thus, individual motivation does result in a mean 3.85 additional tasks completed than group motivation (95% CI: [1.01, 6.60]).