Two-Sample Methods

PSY 5101: Advanced Statistics for Psychological and Behavioral Research 1

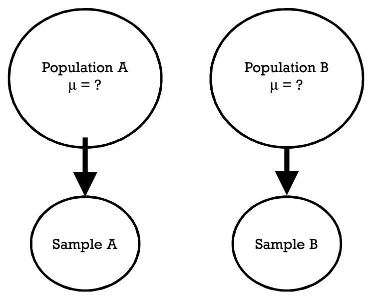
Two Samples

- The one-sample t-test and test of correlation are realistic, useful statistical tests
- The tests that we will learn next are even more useful because they do not need a known value of μ
- They both use two samples
- For example, these statistics can be used to evaluate research concerning two groups of people who saw a brief film of a car wreck
  - Is there any difference in estimates of speed between those who were asked, "How fast were the cars going when they hit into each other?" vs "How fast were the cars going when they smashed into each other?"

Two Sample Methods

- Common characteristics of two sample studies
  - Utilize two groups (usually with equal sample sizes)
  - Interested in the differences in the means from the two groups
  - Focused on a comparison of the two groups and their means rather than a comparison with some known standard value of a population parameter
- Important Note: two-sample t-tests can be used for either true experiments or non-experimental designs
  - Simply using a t-test does NOT allow you to make inferences about causation
  - The capacity to make causal inferences is determined by your research design…not by your choice of test statistic

Two Sample Methods



Two Sample Methods

- We are interested in the difference between two samples
  - We are comparing two populations by evaluating the mean difference
- In order to evaluate the mean difference between the two populations, we sample from both of the populations and we compare the sample means on a given variable
- We need two samples (or groups) and we compare them using a continuous dependent variable
  - Example: We could compare men and women on their levels of math anxiety

Difference in Means and Mean Difference

- Consider two groups of children with the following spelling scores
  - Group 1: 5, 2, 8, 9, 10
  - Group 2: 6, 7, 5, 8, 9
- $\overline{X}_{\text{Group 1}} = 6.8$
- $\overline{X}_{\text{Group 2}} = 7.0$
- Difference in means = $\overline{X}_{\text{Group 1}} - \overline{X}_{\text{Group 2}} = 6.8 - 7.0 = -0.2$
- Mean difference = $\frac{\sum(XG_{roup\ 1}\quad X_{Group\ 2})}{Number\ of\ pairs\ of\ scores} = \frac{(-1)+(-5)+3+1+1}{5} = -0.2$
- Both approaches will yield the same answer
  - We will use the "mean difference" approach when we have dependent samples

Two-Independent-Samples *t*-Test

		Two-Independent-Samples *t*-test
1.	Situation/hypotheses	Two samples Independent samples H_o: $\mu_1=\mu_2$ σ^2 unknown
2.	Test statistic	$t = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\left[\frac{s^2_1(n_1-1)+s^2_2(n_2-1)}{n_1+n_2-2}\right]\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$
3.	Distribution	$t_{df=n_1+n_2-2}$
4.	Assumptions	1. Populations are normal 2. $\sigma^2_1=\sigma^2_2$ 3. Observations are independent

More about these later

Two-Independent-Samples *t*-Test

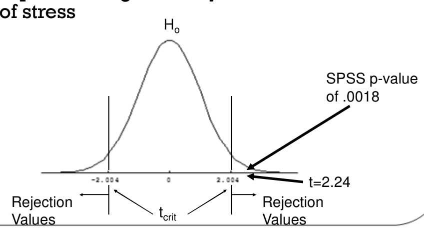
- We have independent samples whenever there is not any obvious dependency present
  - Example: randomly assigning 50 participants to take an experimental drug and 50 participants to take a placebo would result in two independent samples
  - When we cover the two-dependent-samples t-test, we will see some of these obvious ways that samples can be dependent (e.g., sibling pairs)
- Why is $df=n_1+n_2-2$?
  - The denominator for the two-independent-samples t-test has both s^2_1 and s^2_2
  - In $s^2_1 = \frac{\sum(X_1 - \overline{X}_1)^2}{n_1 - 1}$, there are n_1 independent X_1 scores and one statistic ($\overline{X}_1$)
  - So, for s^2_1, the df equals n_1-1
  - Similarly for s^2_2, we have the same df (i.e., df = n_2-1)
  - Adding the two df together gives df = n_1-1+ n_2-1 = n_1+ n_2-2

Two-Independent-Samples *t*-Test

- One group (PC=perceived control) of 20 students thought items they submitted might be selected for the test. The other group (NC=no control) of 20 was told that writing the items was a study aid. Students were randomly assigned to groups. Exam stress was measured by self-reported number of stress symptoms
  - H_o: $\mu_{PC} = \mu_{NC}$
  - H_1: $\mu_{PC} \neq \mu_{NC}$
- Results:
  - Group NC: $\overline{X}_1$ = 15, s^2_1 = 60
  - Group PC: $\overline{X}_2$ = 10, s^2_2 = 40
  - df = n_1+n_2-2 = 30+30-2 = 58
  - Critical values for df = 58 are ±2.004
  - The computed value of t = 2.24, so we reject H_o: $\mu_{PC} = \mu_{NC}$ because 2.24>2.004

Two-Independent-Samples *t*-Test

- The sampling distribution for t for the exam stress example is shown below
- We reject H_o: $\mu_{PC}=\mu_{NC}$ because t=2.24>2.004 <u>OR</u> because p=.0018<α=.05
- The two groups differ significantly in number of symptoms of stress



Two-Independent-Samples *t*-Test: Robustness

- What happens to t when its assumptions are not met? Is t still a "good" statistic? Is α still equal to .05?
- The topic of robustness of test statistics examines their quality or validity when an assumption is not met (when the assumption is ***violated***)
- A statistic is robust to violation of an assumption if
  - Its sampling distribution is well-fit by its theoretical distribution
  - $\alpha_{true} \approx \alpha_{set}$
  - Note that α_{true} is from the sampling distribution and α_{set} is from the theoretical distribution
- When α_{set}=.05, "approximately equals" is generally defined as .04 to .06
  - We get this information from research on statistics

Two-Independent-Samples *t*-Test: Robustness

	Is it met?	**Is the t-test robust?**	**Health Analogy**
Normality	Rarely	Yes, except for mixed distributions where 5-10% of the population are lumped as outliers	Cold or flu bug: Robust with some exceptions
$\sigma^2_1=\sigma^2_2$	Rarely	Yes, on the condition that $n_1=n_2\geq15$. The samples must be both equal and large	Measles: Robust on the condition you have had a measles shot
Independence	Usually…because researchers use appropriate designs and avoid obvious dependency	No, $\alpha_{true}>\alpha_{set}$, like .60 instead of .05, or $\alpha_{true}<\alpha_{set}$, like .001 instead of .05. But independence is usually met	Reactor meltdown: Not robust but you can avoid this issue

Aspin-Welch-Satterthwaite (AWS) Test: AWS *t'*

		AWS t'
1.	Situation/hypotheses	Two samples Independent samples Samples are small (<15), samples are unequal in size, or sample variances are unequal H_o: $\mu_1=\mu_2$ σ^2 unknown
2.	Test statistic	$t' = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}}$
3.	Distribution	$t_{df'} = \frac{\left(\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}\right)^2}{\frac{\left(\frac{s^2_1}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s^2_2}{n_2}\right)^2}{n_2 - 1}}$
4.	Assumptions	1. Populations are normal 2. Observations are independent

More about these later

SPSS Output

```
T-TEST GROUPS=sex(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=mathanx
  /CRITERIA=CI(.95).
```

T-Test

[DataSet0]

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
mathanx	male	20	10.2500	2.76967	.61932
	female	20	12.4000	3.05045	.68210

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
mathanx	Equal variances assumed	.134	.716	-2.334	38	.025	-2.15000	.92131
	Equal variances not assumed			-2.334	37.651	.025	-2.15000	.92131

Sample sizes are equal

Two-independent-samples t-test

dfs are different for the two t-tests

Levene's test examines the assumption that the variances are equal. If p < .05, then the variances are not equal and you should use the AWS t'

AWS t'

Effect Size for *t*-Test

- How strong is the actual effect? That is, what proportion of variability in the dependent variable is accounted for by the independent variable?
- What is needed is an estimate of the magnitude that is relatively independent of sample size
  - Estimates of magnitude or effect size tell us how strongly two or more variables are related or how large the difference is between groups
- Eta squared: $\eta^2 = \frac{t^2}{t^2+df}$

Two-Dependent-Samples *t*-Test

		Two-Dependent-Samples *t*-test
1.	Situation/hypotheses	Two samples Dependent samples: X,X pairs H_o: $\mu_1 = \mu_2$ σ^2 unknown
2.	Test statistic	$t = \frac{\bar{d} - \mu_d}{\sqrt{\frac{s^2_d}{N}}}$ (Note: μ_d is usually 0) N = # of pairs
3.	Distribution	$t_{df=N-1}$
4.	Assumptions	1. Population of ds is normal 2. ds are independent

Two-Dependent-Samples *t*-Test: X,X Pairs

- We have dependent samples whenever we have X,X pairs of scores
  - This dependency is created because of the research design
- Such pairs can happen in at least three different ways:
  - Researcher-produced pairs
    - If students in the exam stress study had been matched on GPA, the researcher would have produced the pairs
    - The X scores on number of symptoms in the PC group would be dependent on the X scores in the NC group
  - Naturally occurring pairs
    - For example, husband-wife pairs, siblings, roommates, etc.
  - Repeated measures
    - This could be the pre-test and post-test scores when people are measured before and after a treatment

Two-Dependent-Samples *t*-Test: Test Statistic

- This test statistic is based on first getting difference scores ($d = X_1 - X_2$)
- Then the statistics in *t* can be computed:
  - $\bar{d} = \frac{\sum d}{N}$
  - $s^2_d = \frac{\sum(d - \bar{d})^2}{N - 1}$

	X_1	X_2	$d = X_1 - X_2$
1	9	6	9-6 = 3
2	8	4	8-4 = 4
3	5	4	5-4 = 1
4	7	8	7-8 = -1
5	etc.	…	…

- Why is df = N-1?
  - The denominator for the two-dependent-samples t-test has s^2_d
  - In s^2_d, there are N independent ds and one statistic ($\bar{d}$)
  - So, for s^2_d and the two-dependent-samples t, the df equals N-1

Two-Dependent-Samples *t*-Test: Example

- Does a new drug (Flexx) increase flexibility in 6 physical therapy patients?
  - H_o: $\mu_{Post} \leq \mu_{Pre}$
  - H_1: $\mu_{Post} > \mu_{Pre}$
- Results: for pre-Flexx scores, $\overline{X}$ =14.3, and for post-Flexx scores, $\overline{X}$ = 38.6
- Computations found $\overline{d}$ =24.3, $s^2{}_d$ =347.87, and t = 3.20
- With N=6 patients, df = N-1 = 6-1 = 5
- The critical value for df = 5 is 2.015
- So we reject H_o: $\mu_{Post} \leq \mu_{Pre}$ because 3.20>2.015

	Pre	Post	d = Post-Pre
1	1	6	6-1 = 5
2	1	8	8-1 = 7
3	1	13	13-1 = 12
4	13	45	45-13 = 32
5	30	75	75-30 = 45
6	40	85	85-40 = 45

SPSS Output

```
T-TEST PAIRS=post WITH pre (PAIRED)
  /CRITERIA=CI(.9500)
  /MISSING=ANALYSIS.
```

T-Test

[DataSet0]

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	post	38.6667	6	35.13782	14.34496
	pre	14.3333	6	16.96663	6.92660

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	post & pre	6	.985	.000

This is the value of the two-dependent-samples t-test

We reject the null hypothesis because ½ of the p-value is less than .05

Paired Samples Test

		Paired Differences				
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper
Pair 1	post - pre	24.33333	18.65118	7.61431	4.76012	43.90655

Paired Samples Test

		t	df	Sig. (2-tailed)
Pair 1	post - pre	3.196	5	.024

Two-Dependent-Samples *t*-Test: Example

- The sampling distribution for t for the Flexx example is shown below
- We reject H_o: $\mu_{Post} \leq \mu_{Pre}$ because t=3.20>2.015
- OR we reject H_o: $\mu_{Post} \leq \mu_{Pre}$ because ½ p = .012< α=.05 and t is positive