CH. 9 – 10 *t*-tests

Paired Samples (Repeated Measures)
Independent Samples

PY 221 Statistics & Research Methods I
Dr. Valenti

Outline for Ch. 9-10 – t-tests

1. Overview of when to use t-tests
2. What is a *t*-test, conceptually?
   - Null and alternative hypotheses
3. Experimental designs & types of t-tests
   - Paired samples *t*-test (aka, repeated measures) - what it's used for
   - Independent samples *t*-test - what it's used for
4. Confidence intervals for difference scores
5. Effect sizes

IMPORTANT NOTE about T-TESTS

Predictor variable *is always* ***qualitative/categorical*** *w/only* ***2 categories/levels.***
Outcome variable *is always* ***quantitative.***

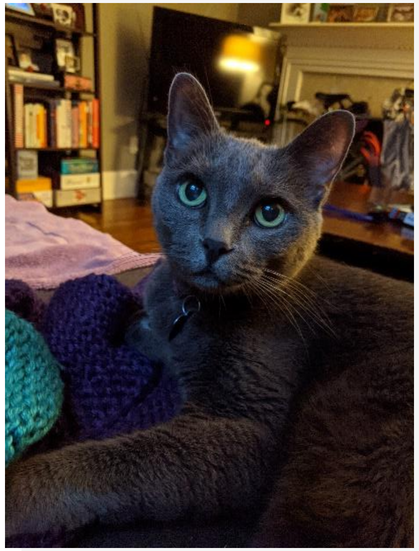
- What do t-tests allow you to do?
  - examine whether the apparent *differences between two groups* on an outcome variable are real differences (vs. produced by chance)
  - examine whether the apparent *changes* in people's scores *across two time* points are real changes (vs. produced by chance)
  - predict the value of an outcome variable based on the **category** (group) it came from, when there are *two* categories (groups)

Recall: Predictor variable - the assumed cause; in experiments, the predictor is called the independent variable (IV)

Recall: Outcome variable - the assumed effect; in experiments, the outcome is called the dependent variable (DV)

Research questions that would involve a *t*-test - examples

1. Does watching *Scream* or watching the news produce greater anxiety?
2. Does taking notes by hand versus taking notes with a laptop lead to better memory for those notes?
3. Do vegan diets or non-vegan diets lead to greater muscle strength?
4. Does wearing an invisibility cloak versus no invisibility cloak lead to more mischievous behavior?



*Now generate your own **research question**, about cats, that would require a t-test to analyze the data. Also list the **predictor** & **outcome variables**, and the **levels** of the predictor variable.*




Practice your Understanding

1. **Which of these research questions can be answered using a t-test?**
   A. Is outdoor temperature (in Fahrenheit) related to ice-cream sales at Big Spoon Creamery?
   B. Is $200 a good estimate of ice-cream sales at Big Spoon Creamery on a typical October day?
   C. Does Big Spoon Creamery sell more ice-cream in the Spring or in the Fall?
   D. Do kids who play soccer, kids who play lacrosse, or kids who play neither sport eat more Big Spoon Creamery ice-cream?

Dr. Valenti wants to compare the average final course grade (% grade) of freshmen vs. non-freshmen in her PY 101 class.

2. What is the outcome variable?

A. final course grade in %
B. students' class (freshmen vs. non-fr)

3. What is the predictor variable?

A. final course grade
B. students' class (freshmen vs. non-fr)

Q1) C
Q2) A
Q3) B

Outline for Ch. 9-10 – t-tests

1. Overview of when to use t-tests
2. **What is a *t*-test, conceptually?**
   - Null and alternative hypotheses
3. Experimental designs & types of t-tests
   - Paired samples *t*-test (aka, repeated measures) - what it's used for
   - Independent samples *t*-test - what it's used for
4. Confidence intervals for difference scores
5. Effect sizes

What is a *t*-test, conceptually?

all test statistics = $\frac{effect}{error}$ =

(including *t*)

What is a *t*-test, conceptually?

- EX: Suppose we run a study to compare memory for class content when taking notes with a laptop vs. by hand, & suppose there's *variation* (there are differences) in memory scores across the two groups…

- **Systematic Variation:**
  - differences in the outcome variable created by the predictor (often, an experimental manip.)
  - EX: *differences in memory scores created by method of note-taking (hand vs. laptop)*
- **Unsystematic Variation ("noise")**
  - differences in the outcome variable created by other, possibly unknown factors that differ across conditions/groups
    - ***examples?***

$$\text{all test statistics} = \frac{effect}{error} = \frac{systematic\ variation}{unsystematic\ variation}$$

(including *t*)

What is a *t*-test, conceptually?

$$\text{all test statistics} = \frac{effect}{error} = \frac{systematic\ variation}{unsystematic\ variation}$$

Recall: t-tests allow us to examine whether the apparent differences between two groups on an outcome variable are *real* differences (vs. produced by chance)

What is a *t*-test, conceptually?

- EX: Suppose we run the invisibility cloak study, & there's *variation* (there are differences) in how mischievous different Ps' behavior is …

- **Systematic Variation:**
  - EX: *differences in <u>mischievousness</u> created by …*

    <u>wearing vs. not wearing the cloak</u>

- **Unsystematic Variation ("noise") --** EX: ***examples?***

$$\text{all test statistics} = \frac{effect}{error} = \frac{systematic\ variation}{unsystematic\ variation}$$

$$t = \frac{size\ of\ the\ difference/change}{error\ in\ the\ measurement\ of\ the\ size\ of\ the\ difference/change}$$

Recall: t-tests allow us to examine whether the apparent differences between two groups on an outcome variable are *real* differences (vs. produced by chance)

Practice with systematic and unsystematic variation

Research Q: Do vegan vs. non-vegan diets lead to more muscle strength?

Suppose that diet was manipulated and muscle strength was later measured, and suppose that amount of muscle strength varied across all participants in the study…

1. Provide an example of what might produce *systematic variation* for this study.
2. Provide an example of what might produce *unsystematic variation* for this study.
3. Holding all else constant, as the *difference* in muscle strength between the vegan and non-vegan groups *increased*, would the value of the t-statistic:

   increase, decrease, or stay the same?

1. The type of diet – vegan vs. non-vegan (this is the only answer)
2. Ps’ motivation to gain muscle strength, Ps’ muscle strength at start of diet, Ps’ age, biol. sex (many possible answers)
3. Increase. If the difference between the groups increases, the numerator (effect, systematic variation), increases, meaning the overall t-statistic will increase.

Outline for Ch. 9-10 – t-tests

1. Overview of when to use t-tests
2. **What is a *t*-test, conceptually?**
   - **Null and alternative hypotheses**
3. Experimental designs & types of t-tests
   - Paired samples *t*-test (aka, repeated measures) - what it's used for
   - Independent samples *t*-test - what it's used for
4. Confidence intervals for difference scores
5. Effect sizes

$$t = \frac{size\ of\ the\ difference/change}{error\ in\ the\ measurement\ of\ the\ size\ of\ the\ difference/change}$$

What type of hypothesis are we testing?

- a hypothesis about:
  - the <u>size of the difference</u> between two separate groups, or
  - the <u>size of the change</u> across time among people who had measurements taken twice.

- Given what you learned in Ch. 7, **what size** difference/change (small vs. large) will the ***null*** *hypothesis* predict?

Null & Alternative Hypotheses for t-tests

<u>null</u> hypothesis:

- the *<u>difference</u>* between the outcome means for each group = 0.
- the *<u>change</u>* across time for the outcome means for our Ps = 0.
- $\mu_1 - \mu_2 = 0$ *or* $\mu_D = 0$ *or* $\mu_1 = \mu_2$
- *In words:* There is no effect of the [IV] on the [DV].
- *In words:* [Our outcome variable] *does not* systematically vary based on [groups of the predictor, i.e., conditions of the IV].
- *In words:* There are no differences in [outcome variable] between [level 1 of predictor] and [level 2 of predictor].

<u>alternative</u> hypothesis:

- the *<u>difference</u>* between the outcome means for each group ≠ 0.
- the *<u>change</u>* across time for the outcome means for our Ps ≠ 0.
- $\mu_1 - \mu_2 \neq 0$ *or* $\mu_D \neq 0$ *or* $\mu_1 \neq \mu_2$
- *In words:* There *is* an effect of the [IV] on the [DV].
- *In words:* [Our outcome variable] *does* systematically vary based on [groups of the predictor, i.e., conditions of the IV].
- *In words:* There *are* differences in [outcome variable] between [level 1 of predictor] and [level 2 of predictor].

Null & Alternative Hypotheses for Cloak Example

- IV = whether or not cloak is worn / DV = amount of mischievous behavior
  - IV condition 1: wears invisibility cloak
  - IV condition 2: no cloak worn

μ_1 indicates average mischievousness of people wearing cloak
μ_2 indicates average mischievousness of people not wearing cloak

- Hypotheses
  - H_0: $\mu_1 = \mu_2$
    - The cloak has no effect on mischievous behavior.

    OR: There are no differences in mischievous behavior between cloak wearers and non-wearers.
  - H_1: $\mu_1 \neq \mu_2$
    - The cloak does produce an effect on mischievous behavior.

    OR: There is a difference in mischievousness between cloak wearers and non-wearers.

PRACTICE: Generate the Null and Alternative Hypotheses, in Symbols and Words

- Research question: Does watching *Scream* or watching the news produce greater anxiety? (condition 1: *It*, condition 2: news)

- Null:

  H_0: $\mu_1 = \mu_2$

  - The program watched has no effect on anxiety; there is no difference in anxiety between *Scream* watchers and news watchers.

- Alternative:

  H_1: $\mu_1 \neq \mu_2$

  - The program watched affects anxiety; there is a difference in anxiety between *Scream* watchers and news watchers.

Practice your Understanding

The owners of Big Spoon Creamery are interested in whether Big Spoon sells more ice-cream in the Spring or in the Fall.

Q1) What would be the null hypothesis?

A. There is not more ice-cream sold in the Spring.
B. There is a significant difference in amount of ice-cream sold in the Spring vs. the Fall.
C. There is an effect of season (Spring vs. Fall) on ice-cream sales.
D. The amount of ice-cream sold in Spring is no different from the amount of ice-cream sold in Fall.

Q2) Which of these relates to unsystematic variation in this example?

A. Some people like ice-cream more than other people.
B. Some people live closer to Big Spoon Creamery locations than other people.
C. Fall typically has hotter weather than Spring, and causes people to want ice-cream more.
D. All of the above.
E. A and B above.

Q1) D
Q2) E

Outline for Ch. 9-10 – t-tests

1. Overview of when to use t-tests
2. What is a *t*-test, conceptually?
   - Null and alternative hypotheses
3. **Experimental designs & types of t-tests**
   - **Paired samples *t*-test (aka, repeated measures) - what it's used for**
   - **Independent samples *t*-test - what it's used for**
4. Confidence intervals for difference scores
5. Effect sizes

Ch. 1 review: How do researchers run *experiments*?

- Manipulate a predictor (called an independent variable)
  - Manipulated vars are always qualitative/categorical. ***T*-tests involve binary IVs, specifically.**
- Randomly assign conditions of the IV to each participant.
- Standardize the procedures across conditions of the IV.
- Measure a DV among all Ps

What are the two types of *experimental designs* we learned about in Ch. 1?

- Within-subjects design
  - The *same* entities (Ps) take part in *all* experimental conditions
- Between-subjects design
  - *Different* entities (Ps) in each experimental condition

Two types of t-tests

The two ways of assigning IV conditions for experiments are associated with two different types of t-tests.

- Within-subjects design → Paired samples t-test (repeated measures)
  - *Same* entities are in *all* conditions (i.e., at each level)
    - or entities are naturally *paired* & complete the same measures
- Between-subjects design → Independent samples t-test
  - *Different* entities in each condition (i.e., at each level)
    - or different entities in each naturally-occurring group

Paired samples t-test: What it is used for

- Tests for differences between two groups/conditions when the *same entities* are in each group/condition (or when we have *pairs* of scores for each unit).
  - Ex 1: For the first ½ of semester, students took notes by hand and for the second ½, students took notes using laptops. Each student experienced both note-taking methods. Midterm exam scores were compared with final exam scores.

  - Ex 2: 20 sets of heterosexual romantic couples were asked how satisfied they were with their relationships, on a scale from 1- not at all *to* 10 – an extreme amount. Scores were compared across males and females to see if they tended to be equally satisfied.