Statistical Inference

Inference for Numerical Variables

Behnam Bahrak
Spring 2020

Experiment

PLAYING A COMPUTER GAME DURING LUNCH AFFECTS FULLNESS, MEMORY FOR LUNCH, AND LATER SNACK INTAKE

distraction and recall of food consumed and snacking

- **Sample:** 44 patients: 22 men and 22 women
- **Study design:**
  - randomized into two groups:
    (1) play solitaire while eating – "win as many games as possible"
    (2) eat lunch without distractions
  - both groups provided same amount of lunch
  - offered biscuits to snack on after lunch

Biscuit intake	$\bar{x}$	s	n
solitaire	52.1g	45.1g	22
No distraction	27.1g	26.4g	22

Estimating the difference between independent means

point estimate ± margin of error

$$(\bar{x}_1 - \bar{x}_2) \pm t^\star_{df} SE_{(\bar{x}_1 - \bar{x}_2)}$$

Standard error of difference between two independent means:	$SE_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

DF for t statistic for inference on difference of two means:	$df = min(n_1 - 1, n_2 - 1)$

Exact Degrees of Freedom

$$df = \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)C^2 + (1 - C)^2(n_1 - 1)}$$

where

$$C = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Conditions for inference for comparing two independent means

1. ***Independence:***

✓**within groups:** sampled observations must be independent

- random sample/assignment
- if sampling without replacement, n < 10% of population

✓**between groups:** the two groups must be independent of each other (non-paired)

2. ***Sample size/skew:*** The more skew in the population distributions, the higher the sample size needed.

Example

- Estimate the difference between the average post-meal snack consumption between those who eat with and without distractions.

biscuit intake	$\bar{x}$	s	n
solitaire	52.1	45.1	22
no distraction	27.1	26.4	22

$$(\bar{x}_{wd} - \bar{x}_{wod}) \pm t^*_{df} SE = (52.1 - 27.1) \pm 2.08 \times \sqrt{\frac{45.1^2}{22} + \frac{26.4^2}{22}}$$

$$= 25 \pm 2.08 \times 11.14$$

$$= 25 \pm 23.17$$

$$= (1.83, 48.17)$$

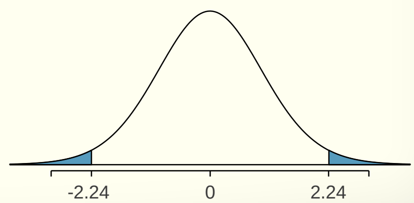
Example

- Do these data provide convincing evidence of a difference between the average post-meal snack consumption between those who eat with and without distractions?

biscuit intake	$\bar{x}$	s	n
solitaire	52.1	45.1	22
no distraction	27.1	26.4	22

$H_0: \mu_{wd} - \mu_{wod} = 0$
$H_A: \mu_{wd} - \mu_{wod} \neq 0$

$T_{21} = \frac{25-0}{11.14} = 2.24$



Decision Making

biscuit intake	$\bar{x}$	s	n
solitaire	52.1	45.1	22
no distraction	27.1	26.4	22

95% confidence interval: (1.83g, 48.17g)

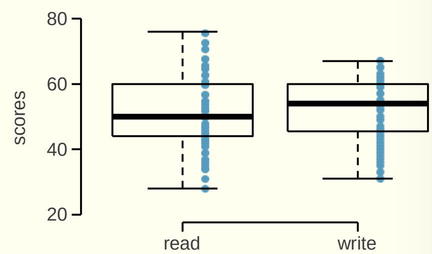
$H_0: \mu_{wd} - \mu_{wod} = 0$
$H_A: \mu_{wd} - \mu_{wod} \neq 0$

agree

p-value ≈ 0.04 Reject H_0

High school and beyond

- 200 observations were randomly sampled from the High School and Beyond survey.
- The same students took a reading and writing test.
- At a first glance, how are the distributions of reading and writing scores similar?
- How are they different?



Example

- Can the reading and writing scores for a given student student assumed to be independent of each other?

	ID	read	write
1	70	57	52
2	86	44	33
3	141	63	44
4	172	47	52
…	…	…	…
200	137	63	65

Analyzing paired data

- When two sets of observations have this special correspondence (not independent), they are said to be paired.
- To analyze paired data, it is often useful to look at the difference in outcomes of each pair of observations:

  diff = read – write

- It is important that we always subtract using a consistent order.

	ID	read	write	diff
1	70	57	52	5
2	86	44	33	11
3	141	63	44	19
4	172	47	52	-5
…	…	…	…	…
200	137	63	65	-2

Analyzing paired data

Parameter of interest

- Average difference between the reading and writing scores of **all** high school students.

μ_{diff}

Point estimate

- Average difference between the reading and writing scores of **sampled** high school students.

$\bar{x}_{\text{diff}}$

Analyzing paired data

- If in fact there was no difference between the scores on the reading and writing exams, what would you expect the average difference to be?

	ID	read	write	diff
1	70	57	52	5
2	86	44	33	11
3	141	63	44	19
4	172	47	52	-5
…	…	…	…	…
200	137	63	65	-2

$\bar{x}_{\text{diff}} = -0.545$

$s_{\text{diff}} = 8.887$

$n_{\text{diff}} = 200$

Hypotheses for paired means

H_0: $\mu_{\text{diff}} = 0$

- There is no difference between the average reading and writing scores.

H_A: $\mu_{\text{diff}} \neq 0$

- There is a difference between the average reading and writing scores.

Nothing New!

one numerical variable

diff
5
11
19
-5
…
-2

hypothesis about the mean

H_0: $\mu_{\text{diff}} = 0$

H_A: $\mu_{\text{diff}} \neq 0$

Example

- Calculate the test statistic and the p-value for this hypothesis test.

H_0: $\mu_{\text{diff}} = 0$

H_A: $\mu_{\text{diff}} \neq 0$

$\bar{x}_{\text{diff}} = -0.545$

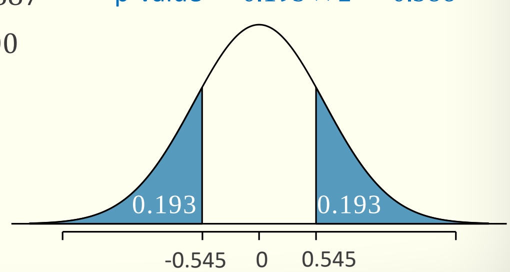
$s_{\text{diff}} = 8.887$

$n_{\text{diff}} = 200$

p-value $= 0.193 \times 2 = 0.386$

$$T = \frac{-0.545 - 0}{\frac{8.887}{\sqrt{200}}} = -0.87$$

$df = 200 - 1 = 199$



Question

- Which of the following is the correct interpretation of the p-value?

(a) Probability that the average scores on the reading and writing exams are equal.

(b) Probability that the average scores on the reading and writing exams are different.

✓(c) Probability of obtaining a random sample of 200 students where the average difference between the reading and writing scores is at least 0.545 (in either direction), if in fact the true average difference between the scores is 0. P(observed or more extreme outcome | H_0 is true)

(d) Probability of incorrectly rejecting the null hypothesis if in fact the null hypothesis is true. P(reject | H_0 is true) = P(Type 1 error)

Summary

- Paired data (2 vars.) → differences (1 var.)
- Most often H_0: $\mu_{\text{diff}} = 0$
- Same individuals: pre-post studies, repeated measures, etc.
- Different (but dependent) individuals: twins, partners, etc.

Truth vs. Decision Table

goal: keep α and β low

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1-\alpha$	Type 1 error, α
	H_A true	Type 2 error, β	$1-\beta$

- Type 1 error is rejecting H_0 when you shouldn't have, and the probability of doing so is α (significance level).
- Type 2 error is failing to reject H_0 when you should have, and the probability of doing so is β.
- **Power** of a test is the probability of correctly rejecting H_0, and the probability of doing so is $1-\beta$.

Question

- Suppose a pharmaceutical company has developed a new drug for lowering blood pressure, and they are preparing a clinical trial to test the drug's effectiveness.
- They recruit people who are taking a particular standard blood pressure medication, and half of the subjects are given the new drug (treatment) and the other half continue to take their current medication through generic-looking pills to ensure blinding (control).
- What are the hypotheses for a two-sided hypothesis test in this context?

$$H_0: \mu_{\text{trmt}} - \mu_{\text{ctrl}} = 0$$
$$H_A: \mu_{\text{trmt}} - \mu_{\text{ctrl}} \neq 0$$

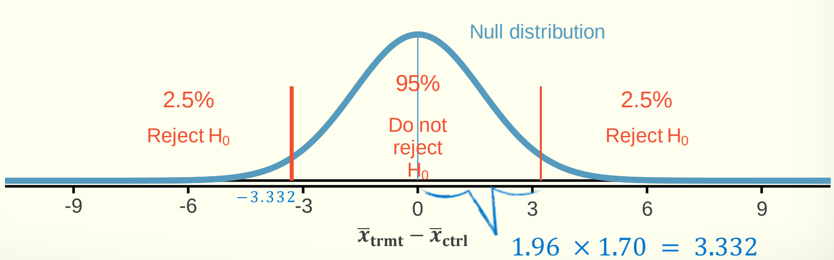
Question

- Suppose researchers would like to run the clinical trial on patients with systolic blood pressures between 140 and 180 mmHg.
- Suppose previously published studies suggest that the standard deviation of the patients' blood pressures will be about 12 mmHg and the distribution of patient blood pressures will be approximately symmetric.
- If we had 100 patients per group, what would be the approximate standard error for difference in sample means of the treatment and control groups?

$$SE = \sqrt{\frac{12^2}{100} + \frac{12^2}{100}} = 1.70$$

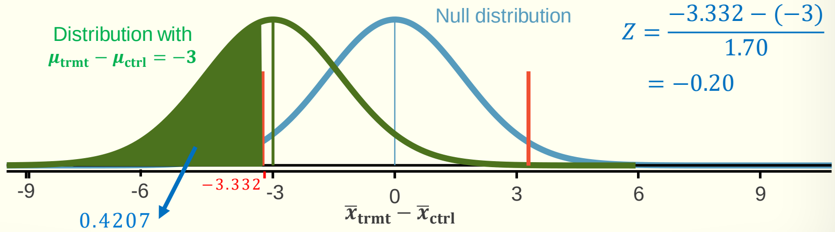
Question

- For what values of the difference between the observed averages of blood pressure in treatment and control groups (effect size) would we reject the null hypothesis at the 5% significance level?



Power

➢ Suppose that the company researchers care about finding any effect on blood pressure that is 3 mmHg or larger vs the standard medication. What is the power of the test that can detect this effect?



$$Z = \frac{-3.332 - (-3)}{1.70} = -0.20$$

Power = $P(\bar{x}_{trmt} - \bar{x}_{ctrl} < -3.332 \mid \mu_{trmt} - \mu_{ctrl} = -3) = P(Z < -0.20) = 0.4207$

Power

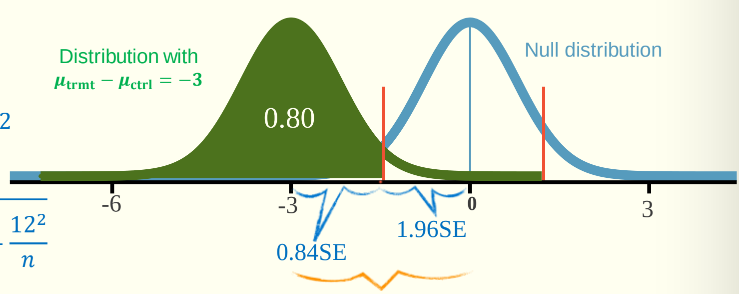
What sample size will lead to a power of 80% for this test?

$$SE = \frac{3}{2.8} = 1.07142$$

$$1.07142 = \sqrt{\frac{12^2}{n} + \frac{12^2}{n}}$$

$$n = 250.88$$

$$n \geq 251$$



$0.84SE + 1.96SE = 2.8SE = 3$ mmHg

Because $P(Z < 0.84) = 0.80$

Summary

- Calculate required sample size for a desired level of power
- Calculate power for a range of sample sizes, and choose target power