Lecture 13 - Tests of Two Means (cont.)

Sta102 / BME 102

Colin Rundel

February 27th, 2015

Testing in Context

Independent Variable

Dependent Variable

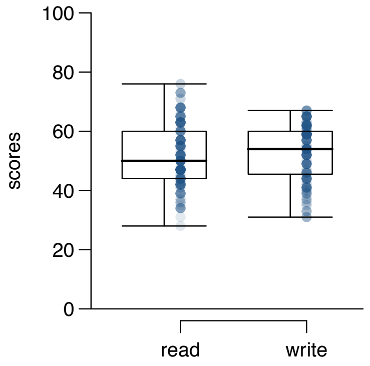
	None	Numerical	Categorical (2 levels)	Categorical (>2 levels)
Numerical	Test of One Mean	Regression	Test of Two Means	ANOVA
Categorical (2 levels)	Test of One Proportion	Logistic Regression	Test of Two Proportions	χ^2 - Test of Independence
Categorical (>2 levels)	χ^2 - GoF	~~Multinomial Regression~~	χ^2 - Test of Independence	χ^2 - Test of Independence

---

Example - Reading and Writing

200 randomly selected high school students took a reading and writing test and their scores are shown below. At a first glance, does there appear to be a difference between the average reading and writing test score?

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



Do you think reading and writing scores are independent?

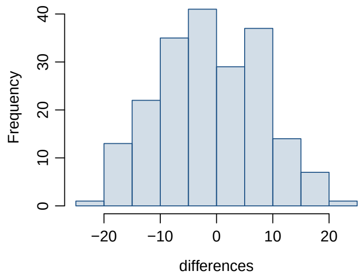
Analyzing paired data

- 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.

$$\text{diff} = \text{read} - \text{write}$$

- 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



Parameter and point estimate

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

$$\mu_{diff}$$

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

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

Setting the hypotheses

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?

What are the hypotheses for testing if there is a difference between the average reading and writing scores?

H_0: There is no difference between the average reading and writing score.

$$\mu_{diff} = 0$$

H_A: There is a difference between the average reading and writing score.

$$\mu_{diff} \neq 0$$

Nothing new here

- The analysis is no different than what we have done before.
- We have data from *one* sample: differences.
- We are testing to see if the average difference is different than 0.

	diff
$\bar{x}$	-0.545
s	8.89
n	200

$$T = \frac{\bar{X} - \mu}{SE} = \frac{-0.545 - 0}{8.89/\sqrt{200}} = -0.877$$

$$\begin{aligned} \text{p-value} &= P(T < -0.877 \text{ or } T > 0.877) \\ &= 2 \times P(T < -0.877) = 2 \times 0.19 = 0.38 \end{aligned}$$

Example - Zinc

Trace metals in drinking water affect the flavor and unusually high concentrations can pose a health hazard. Data were collected by measuring zinc concentration at the bottom and at the surface of 10 randomly sampled wells in Wake country.

We would like to evaluate whether the true average concentration of zinc at the bottom of the well water exceeds that of the surface water. Data are given below.

well	zinc	location
1	0.43	bottom
2	0.266	bottom
3	0.567	bottom
4	0.531	bottom
5	0.707	bottom
6	0.716	bottom
7	0.651	bottom

well	zinc	location
8	0.589	bottom
9	0.469	bottom
10	0.723	bottom
1	0.415	surface
2	0.238	surface
3	0.39	surface
4	0.41	surface

well	zinc	location
5	0.605	surface
6	0.609	surface
7	0.632	surface
8	0.523	surface
9	0.411	surface
10	0.612	surface

Calculating Power - Two Means

- *Step 0:* Pick a meaningful effect size δ and a significance level α
- *Step 1:* Calculate the range of values for the point estimate beyond which you would reject H_0 at the chosen α level.
- *Step 2:* Calculate the probability of observing a value from preceding step if the sample was derived from a population where $\mu = \mu_{H_0} + \delta$

Example - Salmon contamination

Mirex is a carcinogenic insecticide that is known to enter certain watersheds in runoff from agricultural fields. The EPA recommends that any fish caught in these areas be screened for mirex, with any values above 0.08 ppm being considering unsafe. Researchers want to test 150 randomly caught salmon from a river in the pacific northwest for potential organic contaminants. Based on previous research they expect the sample standard deviation of mirex to be around 0.05 ppm.

What would the power of a test be to detect an average concentration of mirex that is 0.01 ppm above the EPA's limit?

Salmon - Details

- cases: 150 salmon
- variable(s): only 1 - concentration of contaminant (numerical)
- test: test of population mean against a null value, t test since σ unknown
  - independence: random sample, 150 $<$ 10% of all salmon → concentration of contaminant in one salmon in the sample is independent of another
  - sample size/skew: the distribution isn't extremely skewed, and $n > 30$
- parameter of interest: μ, true mean concentration of contaminant in all farmed salmon
- point estimate: $\bar{x}$, mean concentration of contaminant in sampled farmed salmon
- hypotheses: $H_0 : \mu = 0.08$; $H_A : \mu > 0.08$
- effect size: $\delta = 0.01$

Example - Reading

An educator believes that new reading activities for elementary school children will improve reading comprehension scores. She plans to randomly assign third graders to an 8-week program in which some will use these activities (18 students) and others will experience traditional teaching methods (20 students). At the end of the experiment, both groups will take a reading comprehension exam.

Previous studies using this exam have resulted in sample standard deviations of 8 points. What would the power of a test be to detect a 5 points improvement in the treatment group?

Reading - Details

- cases: 38 students
- variable(s): (Dep) score - numerical, (Indep) treatment - categorical
- test: Test of Two Means (unpaired)
  - independence w/in groups: random assignment, 18, 20 both $<$ 10% of all students
  - independence btw groups: random assignment, different students in each condition
  - sample size/skew: Assume neither distribution is extremely skewed
- parameter of interest: $\mu_{new} - \mu_{trad}$
- point estimate: $\bar{x}_{new} - \bar{x}_{trad}$
- hypotheses: (two-tailed)

  $H_0 : \mu_{new} = \mu_{trad}$
  $H_A : \mu_{new} \neq \mu_{trad}$
- effect size: $\delta = 5$