Lecture 13 - Tests of One Proportions

Sta102 / BME 102

Colin Rundel

October 19th, 2015

Testing in Context

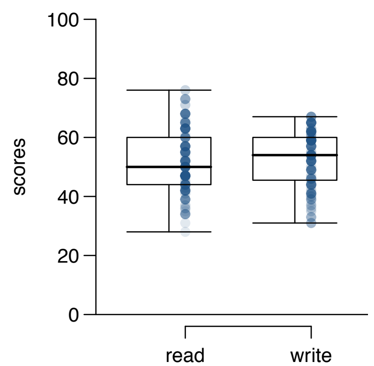
Independent Variable

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?

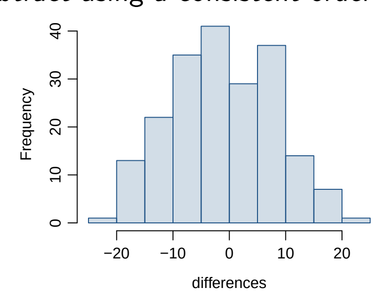
Analyzing paired data

- 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* variable - the difference.
- We are testing to see if the average difference is or is not 0.

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

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

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

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

Tidying the data

We prefer data where each row represents a unit of observation - in this case a well. What does that look like?

well	zinc bottom	zinc top	diff
1	0.43	0.415	0.015
2	0.266	0.238	0.028
3	0.567	0.39	0.177
4	0.531	0.41	0.121
5	0.707	0.605	0.102
6	0.716	0.609	0.107
7	0.651	0.632	0.019
8	0.589	0.523	0.066
9	0.469	0.411	0.058
10	0.723	0.612	0.111

Inference

Lets use a confidence interval to evaluate the difference in zinc concentration between the bottom and top of a well.

$$\bar{x}_{diff} = 0.08, \quad s = 0.052, \quad n = 10$$

95% Confidence Interval:

$$PE \pm CV \times SE$$

$$\bar{x}_{diff} \pm t^{\star}_{df=9} \times \frac{s}{\sqrt{n}}$$

$$0.08 \pm 2.26 \times \frac{0.052}{\sqrt{10}}$$

$$(0.043, 0.118)$$

Calculating power - Step 0 and 1

If we were to conduct a hypothesis test to evaluate if the difference is between the bottom and top is statistically significant, what is the power of these hypotheses to detect a difference of 0.08?
Step 0: What do we know?

$H_0 : \mu_{diff} = 0, \quad H_A : \mu_{diff} \neq 0, \quad \alpha = 0.05, \quad n = 10, \quad SE = 0.0164, \quad \delta = 0.08, \quad 1 - \beta = ?$

Step 1: What values of $\bar{x}$ would let us reject H_0?

$$P(T > t \text{ or } T < -t) < 0.05 \quad \Rightarrow \quad t_{df=9} > 2.26$$

$$P\left(\frac{\bar{x} - 0}{0.0164} > 2.26 \text{ or } \frac{\bar{x} - 0}{0.0164} < -2.26\right) = 0.05$$

$$\bar{x} > 0.0164 \times 2.26 \text{ or } \bar{x} > 0.0164 \times -2.26$$

$$\bar{x} > 0.037 \text{ or } \bar{x} < -0.037$$

Calculating power - Step 2

What is the power of our hypotheses and data to detect a difference of 0.05 in p?
Step 0: What do we know?

$H_0 : p = 0.8, \quad H_A : p > 0.8, \quad \alpha = 0.05, \quad n = 670, \quad SE = 0.0154, \quad \delta = 0.05, \quad 1 - \beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$\bar{x} > 0.037 \text{ or } \bar{x} < -0.037$$

Step 2: Assume $p = 0 + \delta = 0.08$, what is the probability we reject H_0?

$$
\begin{aligned}
&P(\bar{x} > 0.037 \text{ or } \bar{x} < -0.037 | \mu_{diff} = 0.08) \\
&= P\left(T > \frac{0.037 - 0.08}{0.0168}\right) + P\left(T < \frac{-0.037 - 0.08}{0.0168}\right) + \\
&= P(T > -2.56) + P(T < -6.96) \\
&= 0.985
\end{aligned}
$$

Example - Experimental Design

Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?

(a) All 1000 get the drug

(b) 500 get the drug, 500 don't

Results from the GSS

The GSS asks the same question, below is the distribution of responses from the 2010 survey:

All 1000 get the drug	99
500 get the drug 500 don't	571
Total	670

Parameter and point estimate

We would like to estimate the proportion of all Americans who have good intuition about experimental design, i.e. would answer "500 get the drug 500 don't".

What are the parameter of interest and the point estimate?

- *Parameter of interest:* Proportion of *all* Americans who have good intuition about experimental design.

$$p \text{ (a population proportion)}$$

- *Point estimate:* Proportion of *sampled* Americans who have good intuition about experimental design.

$$\hat{p} \text{ (a sample proportion)}$$

Inference on a proportion

What percent of all Americans have a good intuition about experimental design, i.e. would answer "500 get the drug 500 don't"?

- We can answer this research question using a confidence interval, which we know is has the form

$$point\ estimate \pm critical\ value \times standard\ error$$

- What we need to know then is

$$SE_{\hat{p}} = ? \qquad CV = ?$$

Proportions and the CLT

What kind of probability model can we use for $\hat{p}$?

It may be useful to instead think about $n\hat{p}$, what distribution will that have?

$$n\hat{p} \sim Binom(n,\, p)$$

$$n\hat{p} \approx X \sim N(\mu = np,\, \sigma^2 = np(1-p))$$

We can then find the distribution of $\hat{p}$ by dividing by n,

$$\hat{p} \approx X/n \sim N(\mu = p,\, \sigma^2 = p(1-p)/n)$$

Central limit theorem (as applied to proportions)

A sample proportion will have a sampling distribution that is nearly normal with mean equal to the population proportion, p, and standard error equal to $\sqrt{p\,(1-p)/n}$.

$$\hat{p} \sim N\left(\mu = p,\ \sigma^2 = SE^2 = \frac{p\,(1-p)}{n}\right)$$

But of course this is true only under certain conditions ... any guesses?

Assumptions/conditions:

1. *Independence*:
   - *Random sample*
   - *10% condition*: If sampling without replacement, $n <$ 10% of the population.
2. *Normality*: At least 10 successes ($np \geq 10$) and 10 failures ($n(1-p) \geq 10$).

Back to experimental design...

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Estimate (using a 95% confidence interval) the proportion of all Americans who have the correct intuition about experimental design?

Given: $n = 670$, $\hat{p} = \frac{571}{670} = 0.85$.

Are CLT conditions met?

1. *Independence*: The sample is random, and $670 <$ 10% of all Americans, therefore we can assume that one respondent's response is independent of another.
2. *Success-failure*: 571 people answered correctly (successes) and 99 answered incorrectly (failures), both are greater than 10.

Calculating the Confidence Interval

We are given that $n = 670$, $\hat{p} = 0.85$, we also just learned that the standard error of the sample proportion is $SE = \sqrt{\frac{p(1-p)}{n}}$. What is the 95% confidence interval for this proportion?

$$\begin{aligned} CI &= \text{point estimate} \pm \text{margin of error} \\ &= \text{point estimate} \pm \text{critical value} \times SE \\ &= \hat{p} \pm z^\star \times SE \\ &= 0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}} = (0.82, 0.88) \end{aligned}$$

Choosing a sample size

How many people should you sample in order to reduce the margin of error of a 95% confidence interval down to 1%.

$$ME = z^\star \times SE$$

$$0.01 \geq 1.96 \times \sqrt{\frac{p \times (1-p)}{n}}$$

$$0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \textit{Using } \hat{p} \textit{ from previous study}$$

$$0.01^2 \geq 1.96^2 \times \frac{0.85 \times 0.15}{n}$$

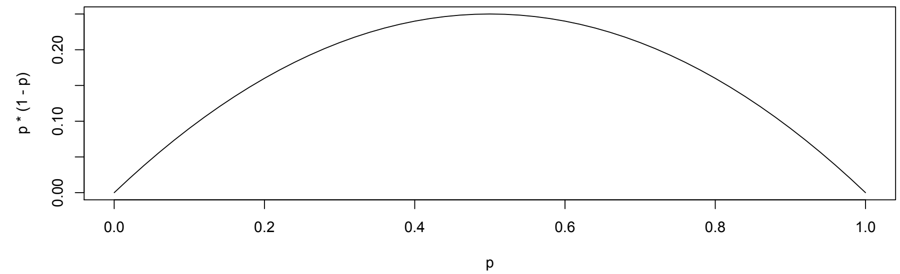
$$n \geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2}$$

$$n \geq 4898.04 \rightarrow n \textit{ should be at least 4,899}$$

What if there isn't a previous study?

... use $\hat{p} = 0.5$. Why?

- if you don't know any better, 50-50 is a good guess
- $\hat{p} = 0.5$ gives the most conservative estimate – largest standard error and thus the largest possible sample size.



HT for proportions

Given what we know so far, how should we set up a hypothesis test for evaluating if more than 80% of all Americans have good intuition about experimental design?

H_A is what we are interested in and H_0 represents the status quo, both must be about the population parameter of interest.

Parameter of interest: p, point estimate: $\hat{p}$

Hypotheses:

$$H_0 : p = 0.8$$
$$H_A : p > 0.8$$

CI vs. HT for proportions

For a test of one proportion our null and alternative hypotheses are about p, therefore when we assume H_0 is true we fix $p = p_0$. Hence,

- Standard error:
  - CI: calculate using observed sample proportion:

$$SE = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

  - HT: calculate using the null value:

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$

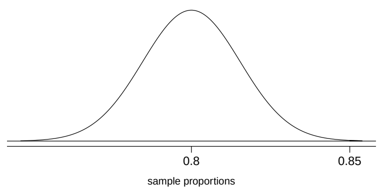
- Success-failure condition:
  - CI: At least 10 *observed* successes and failures, calculated using the sample proportion, $\hat{p}$
  - HT: At least 10 *expected* successes and failures, calculated using the null value, p_0

Back to the GSS

The GSS found that 571 out of 670 (85%) of Americans answered the question on experimental design correctly. Do these data provide convincing evidence that more than 80% of Americans have a good intuition about experimental design?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

$$\begin{aligned} SE &= \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154 \\ Z &= \frac{0.85 - 0.80}{0.0154} = 3.25 \\ p - value &= 1 - 0.9994 = 0.0006 \end{aligned}$$



Since p-value is small we reject H_0. The data provide convincing evidence that more than 80% of Americans have a good intuition on experimental design.

Calculating power - Step 0 and 1

What is the power of our hypotheses and data to detect a difference of 0.05 in p?
Step 0: What do we know?

$H_0 : p = 0.8, \quad H_A : p > 0.8, \quad \alpha = 0.05, \quad n = 670, \quad SE = 0.0154, \quad \delta = 0.05, \quad 1 - \beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$P(Z > z) < 0.05 \quad \Rightarrow \quad z > 1.645$$

$$P\left(\frac{\hat{p} - 0.8}{0.0154} > 1.645\right) = 0.05$$

$$\hat{p} > 0.8 + 0.0154 \times 1.645$$

$$\hat{p} > 0.825$$

Calculating power - Step 2

What is the power of our hypotheses and data to detect a difference of 0.05 in p?
Step 0: What do we know?

$H_0 : p = 0.8, \quad H_A : p > 0.8, \quad \alpha = 0.05, \quad n = 670, \quad SE = 0.0154, \quad \delta = 0.05, \quad 1 - \beta = ?$

Step 1: What values of $\hat{p}$ would let us reject H_0?

$$\hat{p} > 0.825$$

Step 2: Assume $p = 0.8 + \delta = 0.85$, what is the probability we reject H_0? Since p changed, so does $SE = \sqrt{0.85(1 - 0.85)/670} = 0.0138$.

$$\begin{aligned} & P(\hat{p} > 0.825 | p = 0.85) \\ &= P\left(Z > \frac{0.825 - 0.85}{0.0138}\right) \\ &= P(Z > -1.811) \\ &= 0.965 \end{aligned}$$

Common Misinterpretations

11% of 1,001 Americans responding to a 2006 Gallup survey stated that they have objections to celebrating Halloween on religious grounds. At 95% confidence level, the margin of error for this survey a is $\pm 3\%$. A news piece on this study's findings states: "More than 10% of all Americans have objections on religious grounds to celebrating Halloween."

Is this statement justified?