Section 12: Hypothesis testing and p values

STOR 155.02, Spring ’21

updated 2021-04-13

What you will learn

- is in the title

Resources

- Textbook: 5.3

Eyes on the schedule

Four lectures remain

- April 13, 20, 27, May 4

Final project due May 11, 3p.m. sharp

- **no excuses for lateness**
- **except** by approved university exam excuses or exceptional circumstances
- assigned one week prior

Formula: pct (decimal) CI

a. Find the low quantile z^*

$$P\left(Z \leq -z^*\right) = \frac{1 - pct}{2}$$

Excel: `NORM.S.INV(.5*(1-pct))` and Python: `stats.norm.ppf(.5*(1-pct))`

with pct in decimals, e.g. 0.025 for 95% CI

b. or find the high quantile (same z^*)

$$P(Z \geq z^*) = 1 - \frac{1 - pct}{2}$$

c. Calculate the sample point estimates

$$\text{low } = \hat{m}_n - z^* \frac{\hat{s}_n}{\sqrt{n}}, \quad \text{high } = \hat{m}_n + z^* \frac{\hat{s}_n}{\sqrt{n}}$$

Shortcut: `NORM.INV(.5*(1-pct), mhat, shat/sqrtn)` for low

Shortcut: `stats.norm.interval(pct, mhat, shat/sqrtn)` both

need spear *and* net!

thank your classmate for making everything right

Hypothesis testing

Starter example

Adults in Turkey

- suppose you know the **true parameter values**
- mean caloric intake of **3500 calories / day**
- with s.d. of 500 calories

(Source: FAO, 2006-08 via Wikipedia)

Sample of 100 Ankara University students

- **random sample**
- average intake of **3450 calories / day**

Conclude Ankara U. student average calories is different from Turkish population average?

- could happen if **mean for Ankara U. students is in reality smaller**
- or if we got a different average **just because of randomness in sampling**

Assume averages are the same

What is the probability we would get the sample average 3450 or farther away from 3500 from a random sample size 100?

We have point estimates for the true mean (3500) and standard deviation (500).

By the CLT we know the distribution of the sample mean $\bar{X}_{100}$ is approximately Normal $(3500, \frac{500}{\sqrt{100}})$,

$$P\left(\frac{|\bar{X}_{100} - 3500|}{50} > \frac{|3450 - 3500|}{50}\right) \approx P\left(Z > 1 \text{ or } Z < -1\right)$$

$$= 2P\left(Z > 1\right) \approx 0.32$$

That's a pretty high **chance a random sample of size 100 would have average like the one we saw, or smaller**!

Why the 2 above? Area under bell curve for $z < -1$ is the same as for $z > 1$.

Hypothesis test recipe

Model

Data $x_1 \ldots x_n$ sampled randomly from an unknown population r.v. X, with true mean m and s.d. s.

Research question

Do our data give strong evidence that

$$m \neq m_0$$

for some number m_0?

Could also consider asking a one-sided question, $m < m_0$ or $m > m_0$.

Point estimates

$\hat{s}_n$, $\hat{m}_n$ are point estimates for s, m based on the data, e.g. $\hat{m}_n$ is the sample mean of $x_1 \ldots x_n$.

H_0 'null hypothesis', H_1 'alternative'

- H_0 is the statement you want to reject (if possible)
- using the hypothesis test

in this class **will always be about true mean m being equal to some other number m_0**

$$H_0 : m = m_0, \qquad H_1 : m \neq m_0$$

or can consider $H_1 : m < m_0$, $H_1 : m > m_0$.

Will try to **'reject H_0 in favor of H_1'**

Make confidence interval for $\bar{X}_n$ assuming H_0 is true

$$\text{low } = \hat{m}_n - z^* \frac{\hat{s}_n}{\sqrt{n}}, \quad \text{high } = \hat{m}_n + z^* \frac{\hat{s}_n}{\sqrt{n}}$$

If $m_0 >$ high or $m_0 <$ low, reject H_0 in favor of H_1

Which pct to choose for CI?

Rejection with higher pct confidence interval gives stronger evidence

'Significance level' of hypothesis test

$\alpha = \text{ test significance level } = 1 - \text{ pct used in CI}$

intuitively

Rejecting a hypothesis test with α level means if we redid this experiment we would **incorrectly reject H_0 only α percent of the time**

'incorrectly reject' means our test tells us to reject H_0 when in fact it is true

Example

$$\alpha = 0.05 \text{ level corresponds to } 0.95 \text{ CI}$$

Revisit example

True mean and s.d. daily caloric intake for Ankara students are the unknown numbers m, s.

Hypotheses

$m_0 = 3500$

$$H_0 : m = 3500, \qquad H_1 : m \neq 3500$$

Confidence interval pct = 0.95

Assume we get the same standard deviation for Ankara students, $\hat{s}_n = 500$.

$$\text{low } = 3450 - z^* \frac{500}{\sqrt{100}}, \quad \text{high } = 3450 + z^* \frac{500}{\sqrt{100}}$$

rounding to 3 decimals

$$(\text{low}, \text{high}) = (2470.018, 4429.982)$$

Q: Reject H_0?

Q: Significance level α?

Q: If you consider a CI with pct $= 0.98$ would your answer change?

p-values

Definition

> > "The p-value is the probability of observing data at least as favorable to the alternative hypothesis as our current data set, **if the null hypothesis is true.**"
> >
> > -- textbook ch 5.3.4

Translation

$$H_0 : m = m_0, \quad H_1 : m \neq m_0$$

> > p-value is the probability of observing data that is at least $|\hat{m}_n - m_0|$ away from m_0 as what we saw...
> >
> > meaning chance that if we redrew our sample we would see a sample mean at least as different from m_0 as our current estimate $\hat{m}_n$

Calculation

'two-sided test'

$$H_1 : m \neq m_0$$

$$p = P\left(\frac{|\bar{X}_n - m_0|}{\hat{s}_n/\sqrt{n}} > \frac{|\hat{m}_n - m_0|}{\hat{s}_n/\sqrt{n}}\right) = 2P\left(Z < -\frac{|\hat{m}_n - m_0|}{\hat{s}_n/\sqrt{n}}\right)$$

'one-sided test'

$$H_1 : m < m_0$$

$$p = P\left(\frac{\bar{X}_n - m_0}{\hat{s}_n/\sqrt{n}} < \frac{\hat{m}_n - m_0}{\hat{s}_n/\sqrt{n}}\right) = P\left(Z < \frac{\hat{m}_n - m_0}{\hat{s}_n/\sqrt{n}}\right)$$

Often you'll see cutoff value referred to as the **'z-statistic'**. These are for the two-sided or one-sided tests:

$$\frac{|\hat{m}_n - m_0|}{\hat{s}_n/\sqrt{n}} \text{ or } \frac{\hat{m}_n - m_0}{\hat{s}_n/\sqrt{n}}$$

Accept-reject hypothesis test using p values

$$H_0 : m = m_0, \quad H_1 : m \neq m_0$$

$$\alpha = \text{ test significance level}$$

reject H_0 if

$$p \leq \alpha$$

This setup holds for one-sided tests too.

Flip it: Significance in terms of p-value

p-value is the largest significance level at which you could reject the null

How small a significance level / p-value is good enough?

Very much depends on the context

choose smaller α when you really don't want to reject incorrectly

Higgs boson discovery:

$$\alpha = 3 \times 10^{-7} \approx 1/3,500,000$$

Scientific American blog post

generic standards you will see

$$\alpha = 0.05, 0.1$$

But why? Should instead think carefully about what makes sense for the question.

Example for poll

> > It is suspected that patients treated with a certain drug have systolic blood pressures (SBP) that are lower on average than 115, which is the mean for people not treated.
> >
> > A random sample of 20 patients is treated with the drug. Their sample average SBP is 112 and the sample s.d. is 10.

PollEv.com/brendanbrown849

poll closes at _________


five more minutes