Lecture 21

1

Topics:

I) Hypothesis Formalism

II) approximate p-values
translate approx z-scores
to probabilities

III) exact p-values

---

e.g. Suppose your data gives

$$\hat{\theta} = 29.2$$

where the sampling variability in $\hat{\theta}$ is described by

$$\hat{\theta} \sim N(\theta, 100).$$

Question: Test the null hypothesis

$H_0: \theta = 0.$

Fancy syntax for the words "is the sample $\hat{\theta}$ consistent with the population value $\theta = 0$?"

Answer:

Find the z-score of $\hat{\theta} = 29.2$ under the assumption $\theta = 0$.

$$z = \frac{\hat{\theta} - E(\hat{\theta})}{sd(\hat{\theta})} = \frac{29.2 - \theta}{10}$$

$$\overset{\text{by } H_0}{=} \frac{29.2 - 0}{10}$$

$$= 2.92$$

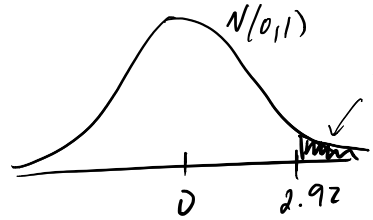
2

Since a z-score of 2.92 is kinda unlikely we are inclined to reject H_0 in favor for a different value of θ.

... But how unlikely is "kinda unlikely"?

Here is the picture:

If $\theta = 0$ then $z = \frac{\hat{\theta} - 0}{10}$ is a draw from



This area is 0.0018 and is called a p-value for testing $H_0: \theta = 0$.

This is what we observed

* p-values quantifies how hard it is to get $\hat{\theta} = 29.2$ (or even more rare) assuming H_0 is true.
* small p-values $\Rightarrow$ more rare
  $\Rightarrow$ H_0 is probably false

  large p-values $\Rightarrow$ less rare
  $\Rightarrow$ H_0 is consistent with the data $\hat{\theta} = 2.92$
* p-values effectively translate a z-score to a rarity measure that others can understand. i.e. since

  the p-value $= 0.0018 \approx \frac{1.05}{600}$

  we would need to sample from $N(0, 100)$ about 600 times before getting $\hat{\theta} = 2.92$ or bigger.

Using the CLT for approximating p-values

The CLT is the main tool for translating z-scores to p-values when testing a population mean μ.

e.g. Huge list of numbers

$$y_1, y_2, \ldots, y_N \quad \text{where } N = 5\times 10^{22}$$

You want to estimate

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i = \bar{y}$$

i.e. the list average.

Suppose the list s.d. is given by

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2} = 7.2$$

Sometimes called the population s.d.

To construct $\hat{\mu}$ you randomly sample (w/rep) $n=45$ numbers from the list & get $Y_1, Y_2, \ldots, Y_n$ with

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 22.1.$$

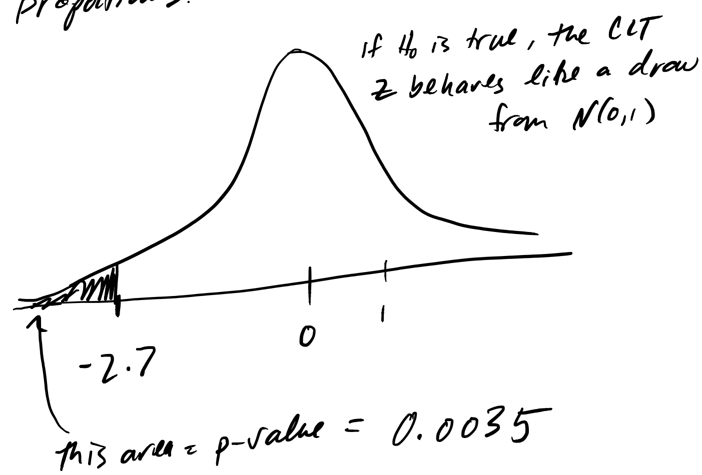
Question: Test the null hypothesis

$$H_0: \mu = 25$$

Answer: First find the z-score of $\hat{\mu} = 22.1$ given $\mu = 25$ & $\sigma = 7.2$.

$$z = \frac{\hat{\mu} - E(\hat{\mu})}{sd(\hat{\mu})} = \frac{\hat{\mu}-\mu}{(\sigma/\sqrt{n})} = \frac{22.1-25}{7.2/\sqrt{45}} = -2.7$$

Now find the p-value of this z-score to quantify rarity in terms of proportions.



So, we would see a z-score of -2.7 or more extreme ~3 out of 1000 tries.

This is evidence that H_0 is wrong.

Remember: Hypothesis testing is only a technique for arguing against someone who believes H_0.

Sometimes there is an agreed upon p-value cutoff (also called "α-level" or "the level of the test") where if the (p-value) < (p-value cutoff) then the observed z-score is rare enough that H_0 is rejected.

In the last example if the p-value cut off is 0.01, $\boxed{H_0 \text{ is rejected}}$. Alternatively if the p-value cut off is 0.001 then $\boxed{H_0 \text{ can not be rejected}}$.

Note: Hypothesis testing only works when H_0 is rejected.

What happens when σ is not known

e.g. same example as before but suppose we are not given that $\sigma = 7.2$
To handle this case just estimate σ

The Data:

$$Y_1, Y_2, \ldots, Y_n \quad \text{where } n = 45$$

where

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = 22.1$$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2} = 6.9$$

Sometimes called the sample standard deviation s.

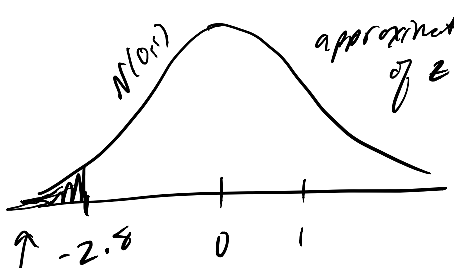
Question: Test the null

$$H_0: \mu = 25$$

Answer: the z-score of the estimate:

$$z = \frac{\hat{\mu} - E(\hat{\mu})}{sd(\hat{\mu})} = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} = \frac{22.1 - 25}{\sigma/\sqrt{45}} \approx \frac{22.1-25}{\hat{\sigma}/\sqrt{45}} = -2.8$$

Now, if we just ignore the fact that we have $\hat{\sigma}$ instead of σ we can quickly find an approximate p-value



approximate behavior of z if H_0 is true

$\boxed{\text{approximate p-value} = 0.0025}$

which suggests H_0 is wrong.

Note: if n is small this approximation could be wrong.

Question: Can one compute an exact p-value for the observations $\hat{\mu} = 22.1$, $\hat{\sigma} = 6.9$?

The approximate p-value 0.0025 is based on two approximations

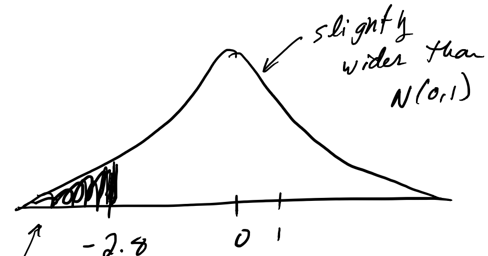
I) the CLT that $\hat{\mu}$ fluctuates like a Normal distribution

II) using $\hat{\sigma}$ instead of σ for computing the z-score.

If you assume the population is normally distributed then the approximated z-score

$$\frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{n}} \text{ is a } \underbrace{t^{(n-1)} \text{ random variable}}_{\text{"a t distribution on } n-1 \text{ degrees of freedom"}}$$

with pdf



This area is the exact p-value under the assumption that the population is Normal.
using a calculator gives

$\boxed{\text{the exact p-value} = 0.0037}$

Comparing two population averages

e.g. Two huge lists of numbers

$x_1, x_2, \ldots,$

$y_1, y_2, \ldots$

Let

μ_x = the Ave of the x_i's

μ_y = the Ave of the y_i's.

σ_x = the "population s.d." of the x_i's

σ_y = the "population s.d." of the y_i's.

You want to estimate

$$\mu_x - \mu_y = ?$$

So you sample 25 numbers from the first list & get

$$X_1, X_2, \ldots, X_{n_x}, \quad n_x = 25$$

also sample 20 numbers from the second list & get

$$Y_1, Y_2, \ldots, Y_{n_y}, \quad n_y = 20$$

Suppose this data results in

$$\hat\mu_x = \bar X = -2.5$$

$$\hat\mu_y = \bar Y = 1.23$$

$$\left.\begin{array}{l}\hat\sigma_x = 10.2\\ \hat\sigma_y = 5.1\end{array}\right\} \text{sample standard deviations of the two samples.}$$

Question: Test the null hypothesis

$H_0: \mu_x = \mu_y$ (i.e. $\mu_x - \mu_y = 0$)

with a p-value cut off of 0.001.

Answer: Start by approximating the z-score of the estimate

$$\hat\mu_x - \hat\mu_y = -2.5 - 1.23$$

using the null hypothesis H_0 that $\mu_x - \mu_y = 0$

Note: MF1 & MF2 give

$$E(\hat\mu_x - \hat\mu_y) = \mu_x - \mu_y$$

$$sd(\hat\mu_x - \hat\mu_y) = \sqrt{\frac{\sigma_x^2}{25} + \frac{\sigma_y^2}{20}}$$

Therefore

$$z\text{-score} = \frac{(\hat\mu_x - \hat\mu_y) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{25} + \frac{\sigma_y^2}{20}}}$$

$$= \frac{(-2.5 - 1.23) - (0)}{\sqrt{\frac{\sigma_x^2}{25} + \frac{\sigma_y^2}{20}}}$$

hypothesized value of $\mu_x - \mu_y$

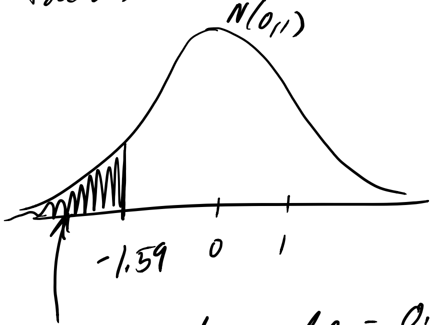
$$\approx \frac{(-2.5 - 1.23) - (0)}{\sqrt{\frac{10.2^2}{25} + \frac{5.1^2}{20}}}$$

plug in $\hat\sigma_x = 10.2$ & $\hat\sigma_y = 5.1$ for σ_x & σ_y

$$= -1.59$$

9

The approximate Z-score -1.59 translates to an approximate p-value as follows



approximate p-value $= 0.056$

This is much bigger than the cutoff $\Rightarrow$ can't reject H_0.

If you want an exact p-value in this case you must assume both populations are exactly Normal & $\sigma_x = \sigma_y = \sigma$

when $\sigma_x = \sigma_y$ Notice

$$sd(\hat{\mu}_x - \hat{\mu}_y) = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

$$= \sqrt{\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}}$$

$$= \sqrt{\sigma^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}$$

$$= \sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

Now approximate σ with a "pooled estimate of s.d." denoted $\hat{\sigma}_p$ or S_p & given by

$$\hat{\sigma}_p = \sqrt{\underbrace{\frac{n_x - 1}{n_x + n_y - 2}\hat{\sigma}_x^2 + \frac{n_y - 1}{n_x + n_y - 2}\hat{\sigma}_y^2}_{\text{a weighted ave of } \hat{\sigma}_x^2 \,\&\, \hat{\sigma}_y^2.}}$$

10

Now under the Normal population assumption & $\sigma_x = \sigma_y = \sigma$ the approximate z-score

$$\frac{(\hat{\mu}_x - \hat{\mu}_y) - (\mu_x - \mu_y)}{\hat{\sigma}_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$ is a t random variable on $n_x + n_y - 2$ degrees of freedom.

In the above example

$$\hat{\sigma}_p = \sqrt{\frac{25-1}{25+20-2}(10.2)^2 + \frac{20-1}{25+20-2}(5.1)^2}$$

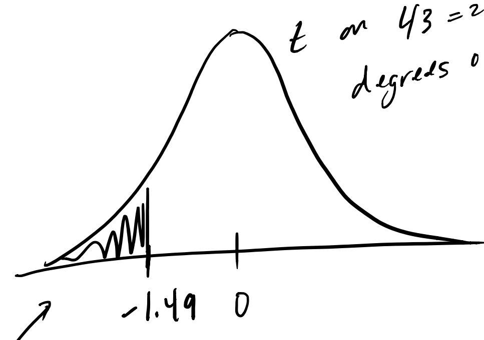
$$= 8.34$$

And therefore the approximate z-score of the estimates, assuming $H_0 : \mu_x - \mu_y = 0$, is given by

$$\frac{(\hat{\mu}_x - \hat{\mu}_y) - (\mu_x - \mu_y)}{\hat{\sigma}_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{-2.5 - 1.23 - 0}{8.34\sqrt{\frac{1}{25} + \frac{1}{20}}}$$

$$= -1.49$$

which behave like a draw from t on $43 = 25 + 20 - 2$ degrees of freedom



This is the exact p-value $= 0.07$

Note: this exact p-value is different from the "quick guess" method but not different enough to change the conclusion

$\therefore$ do not reject H_0 since $(\underset{\text{p-value}}{0.07} > \underset{\text{cut-off}}{0.001})$.