HUDM4122
Probability and Statistical Inference

April 15, 2015

HW8

Which of these do we use to refer to a result where p=0.08?

- Statistically Significant
- Marginally Significant
- Not Significant
- Highly Significant

How do we refer to a result where $p=0.5$?

- Statistically Significant
- Marginally Significant
- Not Significant
- Highly Significant

P7: Common wrong answer = 0.99
Does that make sense?

- 36 students use a curriculum, and take a pre-test and post-test.
- The average learning gain in your sample is 5 points, with a
  standard deviation of 12 points.
- You want to know if the average learning gain is statistically
  significantly greater than 0 (i.e. is our children learning?)
- What is the p value for a two-tailed Z test? (Give two digits after the decimal)

P8: Common wrong answer = 0
Does that make sense?

- 64 students with a specific behavioral disorder participate in an intervention designed for their needs, and are observed afterwards.
- The clinical observation scale goes from 0 to 10, with any score below 3 considered evidence for appropriate behavior.
- The average clincal observation score in your sample is 3.2 points, with a standard deviation of 4 points.
- What is the p value for a two-tailed Z test? (Give two digits after the decimal)

P7 and P8

- A lot of students gave p values over 1 for these problems
- Does that make sense?

Try one by yourselves

- Students in Yonkers are being evaluated for whether they are ready to go in Yonkers's fabled gifted and talented program
- The test scale goes from 0 to 10, with any Yonker scoring above 7 considered ready for the gifted program
- I claim to have a prep course that will enable any kid to pass this test
- 100 randomly selected Yonkers take my prep course, and average a test score of 7.3, with a standard deviation of 3 points
- Is my program actually working?
- What is the p value for a two-tailed Z test?
  (Give two digits after the decimal)

P10

- You’re comparing the difference between Bob's Discount Math Curriculum and SaxonMath
- Bob's: average grade = 58, standard deviation = 7, sample size = 49
- SaxonMath: average grade = 62, standard deviation = 10.5, sample size = 49

Compute a two-tailed Z test to find out whether the difference between curricula is statistically significant.

Try it by yourselves

- You're comparing the difference between Ryan's Gifted Prep Course and Bob's Discount Gifted Prep Course
- Bob's: average score = 7.1, standard deviation = 5, sample size = 100
- Ryan's: average score = 7.3, standard deviation = 3, sample size = 100

Compute a two-tailed Z test to find out whether the difference between curricula is statistically significant.

Statistical Power

- From last class you may remember

In the traditional statistical significance paradigm

- You control α
- You are unable to control β

“Type I Error”

- False Positive
- Rejecting the Null Hypothesis when the Null Hypothesis is true
- Saying the result is statistically significant when there’s nothing there
- α

“Type II Error”

- False Negative
- Accepting the Null Hypothesis when the Null Hypothesis is false
- Saying the result is not statistically significant when there’s actually something there
- β

But there is a way to try to control β

Statistical Power Analysis

Statistical Power Analysis

- You can reduce the probability of a false negative, by increasing your sample size

- You can estimate the sample size needed to avoid false negatives

Power

- We refer to $1-\beta$ as ***power***

Power

- We refer to $1-\beta$ as ***power***

- $1-\beta = P(\text{reject } H_0 \text{ when } H_a \text{ is true})$

Examples

- If there is a 20% chance we reject the null hypothesis when it is false, power = 20%

- If there is a 80% chance we reject the null hypothesis when it is false, power = 80%

Typical desired value (magic number)

- **If there is a 80% chance we reject the null hypothesis when it is false, power = 80%**

Questions? Comments?

Estimating Power

- For One-Sample Z-test

Estimating power

- You need to know
  - the value you're comparing to, μ for H_0
  - the sample size n

- You need to guesstimate
  - The sample mean $\bar{x}$ you expect for the actual data
  - The standard deviation σ you believe the data has

Estimating power for $\alpha = 0.05$

- First find

- $\bar{x}_{1.96} = \mu + (1.96)(\frac{\sigma}{\sqrt{n}})$
- $\bar{x}_{-1.96} = \mu - (1.96)(\frac{\sigma}{\sqrt{n}})$

- These values represent the left and right bounds of the acceptance region
  - Reverse of the rejection region

Estimating power for α = 0.05

- The corresponding Z values for the left and right bounds of the acceptance region are

- $\dfrac{\bar{x}_{-1.96}-\bar{x}}{\frac{\sigma}{\sqrt{n}}}$

- $\dfrac{\bar{x}_{1.96}-\bar{x}}{\frac{\sigma}{\sqrt{n}}}$

Estimating power for α = 0.05

- $\beta = P(Z_{\beta left} < Z < Z_{\beta right})$

- Power = 1 - β

Estimating power for $\alpha \neq 0.05$

- Substitute different critical values of Z_α in place of 1.96

Example

- You are trying to find out the statistical power of the following test:

- You study the pre-post learning gains of Brunei Math
- You want to know if the learning gain is, on average, greater than 0
- You give Brunei Math to 36 students

- You guesstimate that Brunei Math will lead to learning gain of 10, with standard deviation of 30

Example

- You are trying to find out the statistical power of the following test:

- You study the pre-post learning gains of Brunei Math
- You want to know if the learning gain is, on average, greater than 0 (μ)
- You give Brunei Math to 36 students (n)

- You guesstimate that Brunei Math will lead to learning gain of 10 ($\bar{x}$), with standard deviation of 30 (σ)

Example

- $\bar{x}_{1.96} = \mu + (1.96)(\frac{\sigma}{\sqrt{n}})$
- $\bar{x}_{-1.96} = \mu - (1.96)(\frac{\sigma}{\sqrt{n}})$

- $\bar{x}_{1.96} = 0 + (1.96)(\frac{30}{\sqrt{36}}) = +9.8$
- $\bar{x}_{-1.96} = 0 - (1.96)(\frac{30}{\sqrt{36}}) = -9.8$

Example

- The corresponding Z values for the left and right bounds of the acceptance region are

- $Z_{\beta right} = \frac{\bar{x}_{-1.96} - \bar{x}}{\frac{\sigma}{\sqrt{n}}} = \frac{-9.8 - 10}{\frac{30}{\sqrt{36}}}$ =-0.04

- $Z_{\beta left} = \frac{\bar{x}_{1.96} - \bar{x}}{\frac{\sigma}{\sqrt{n}}} = \frac{9.8 - 10}{\frac{30}{\sqrt{36}}}$ = -3.96

Example

- $\beta = P(-0.04 > Z > -3.96) = 0.484$

- Power $= 1 - \beta = 0.516$

- Which, BTW, is awful; 48.4% probability of saying the result is not statistically significant when there's actually something there

Questions? Comments?

You Try It

- You are trying to find out the statistical power of the following test:

- You study the pre-post learning gains of Johor Bahru Math
- You want to know if the learning gain is, on average, greater than 0
- You give Johor Bahru Math to 49 students

- You guesstimate that Johor Bahru Math will lead to learning gain of 12, with standard deviation of 35

What if learning gain was 15?

- You are trying to find out the statistical power of the following test:

- You study the pre-post learning gains of Johor Bahru Math
- You want to know if the learning gain is, on average, greater than 0
- You give Johor Bahru Math to 49 students

- You guesstimate that Johor Bahru Math will lead to learning gain of ***15***, with standard deviation of 35

What if standard deviation was 20?

- You are trying to find out the statistical power of the following test:

- You study the pre-post learning gains of Johor Bahru Math
- You want to know if the learning gain is, on average, greater than 0
- You give Johor Bahru Math to 49 students

- You guesstimate that Johor Bahru Math will lead to learning gain of 12, with standard deviation of ***20***

What if n was 100?

- You are trying to find out the statistical power of the following test:

- You study the pre-post learning gains of Johor Bahru Math
- You want to know if the learning gain is, on average, greater than 0
- You give Johor Bahru Math to ***100*** students

- You guesstimate that Johor Bahru Math will lead to learning gain of 12, with standard deviation of 35

You can improve statistical power

- By going for a bigger sample size
- By finding a more powerful intervention
- By reducing your standard deviation
  - The ways to do this are usually dodgy; for example, by sampling from a very homogenous sub-population rather than randomly

Note

- Statistical power estimation is based on guesstimation
- How do you really know the sample mean and standard deviation until you’ve run the study?
- You could estimate values based on prior research, but you don’t really know…

Questions? Comments?

Statistical power of a two-sample test

- Same basic concept
- Plugging in the different way of calculating Z and SE that we saw for two-sample tests

- Considered out of scope both by the book and other 4122 instructors

- And I'm happy to go along with that practice to spare you lots of calculating…

Final questions for the day?

Upcoming Classes

- 4/20 Independent-samples t-test
  - ***HW9 due***
- 4/22 Paired-samples t-test
- 4/23 Special session on SPSS