

week 2 Inferential Statistics

[hypothesis testing, confidence intervals, regression analysis]

Objectives:-

- Descriptive Statistics vs Inferential Statistics
- Qualitative and Quantitative
- Scales of Measurement - Nominal, Ordinal, Interval, & Ratio Scale Data
- Hypothesis testing and the Null Hypothesis, clearly explained
- Alternative Hypotheses: Main Ideas
- p-values: What they are and how to interpret them
- How to calculate p-values
- Confidence Intervals, clearly explained
- Regression analysis

Video 1 Inferential Statistics

Sample data to make an inference
or draw a conclusion of the population

Confidence Intervals & Margins of error

Video 2 Quantitative Data

Numerical Data (Numbers)

Discrete Continuous

(Counting) (Measurement)

Whole number Float (any num)

0 1 2 3 2.5

7 dogs 6 ft

Qualitative Data

Descriptive data (words)

see, feel, taste, hear, smell

black Rough

video ③ Scales of Measurement

Nominal

Qualitative / categorical

Names, Colors, labels, Gender, etc

[Order doesn't matter]

Ordinal

Ranking / placement

[Differences cannot be measured]

[Order matters]

Interval

[Order matters]

[Differences can be measured (except ratios)]

No True "0" Starting point

Ratio

[Order matters]

[Differences are measurable (including ratios)]

contains a "0" Starting point

	Nominal	Ordinal	Interval	Ratio
labeled	✓	✓	✓	✓
Order	X	✓	✓	✓
Measure Diff.	X	X	✓	✓
True 0 start	X	X	X	✓

video ④ Hypothesis & Null Hypothesis

We can make a hypothesis and if data give us strong evidence that the hypothesis is wrong, then we can reject the hypothesis

but when we have data that is similar to the hypothesis, but not exactly the same, then the best we can do is fail to reject the hypothesis

Null Hypothesis does not require preliminary data because the only value that represents no diff. is 0

video ⑤ Alternative Hypotheses

Statistical test of decision whether reject or fail to reject the Null Hypothesis



data

[Null Hypotheses
Alternative Hypotheses]

two group of data

The Alternative is the opposite of the Null Hypothesis

video ① p-values

are numbers between 0 and 1, that, quantify how confident we should be

the closer p-value is to 0 the more confidence we have

how small does a p-value have to be before we are sufficiently confident
threshold is 0.05

it means that there is no diff if we do this exact same experiment a bunch of time

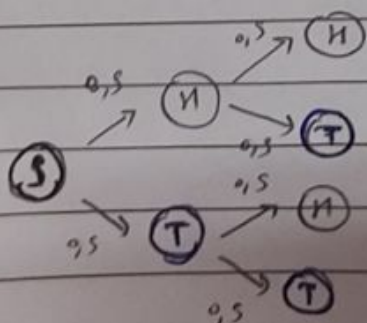
Getting a small p-value when there is no difference is called a False positive

p-value however it tells if there are difference or not, it doesn't tell how difference it is

video ② how to calculate p-values

One-Sided and Two-Sided
avoided

p-values are determined by adding up probabilities



$$= \begin{pmatrix} H & H \end{pmatrix} \frac{1}{4}$$

$$\left[\begin{pmatrix} H & T \end{pmatrix} \right. \quad \frac{1}{2}$$

$$\left. \begin{pmatrix} T & T \end{pmatrix} \right] \frac{1}{4}$$

$\frac{\text{number of favorable outcomes}}{\text{total outcomes}}$

p-value

- the probability random chance
- the probability of observing something else that is equally rare
- the probability of observing something rarer or more extreme

$$p\text{-value for 2 Heads} = 0,25 + 0,25 + 0 = 0,5$$

$0,5 > 0,05$ Failed to reject the hypothesis
my coin is no different from a normal coin

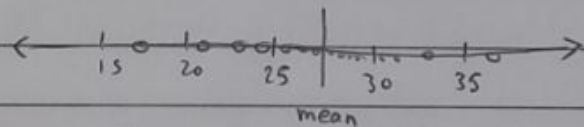
to calculate p-values with distribution, you add up the percentages of area under the curve.

One-sided p-value Which direction we want to see change in should be avoided

video 8 Confidence Intervals

Bootstrap Refresh

bootstrapping



- * Randomly select weights from the original sample (duplicates are ok)
- * calculate the mean
- * Repeat step 1, 2 until calculated a lot of means ($> 10,000$)

95% confidence intervals

is just an interval that covers 95% of the means

video ② Regression analysis

analysis relationships between variables

Categorical continuous

Regression line

simple model can be extended to include more
independent variables
