Causality and Hypothesis Testing

Week 09 Day 01

DS 3000 – Foundations of Data Science

1

Reminders

HW 5

Optional

Due Thursday, October 31

FP3: Project Summary

Available on BB

Due Monday, November 11

2

Outline

Causality

Experiments and Experimental Designs

Hypothesis Testing

3

Causality

4

5

Establishing Causality

X → Y

Facebook Usage → Happiness

6

Experiments

Goal:

To establish the existence of a cause-and-effect relationship between two variables

A true experiment

Must demonstrate that changes in one variable are directly responsible for causing changes in the second variable

Four Basic Elements:

Manipulation, Measurement, Comparison, and Control

7

Four Basic Elements of an Experiment

Manipulation

Researcher manipulates one variable by changing its value to create a set of two or more treatment conditions

Measurement

A second variable is measured for a group of participants resulting in a set of scores in each treatment condition

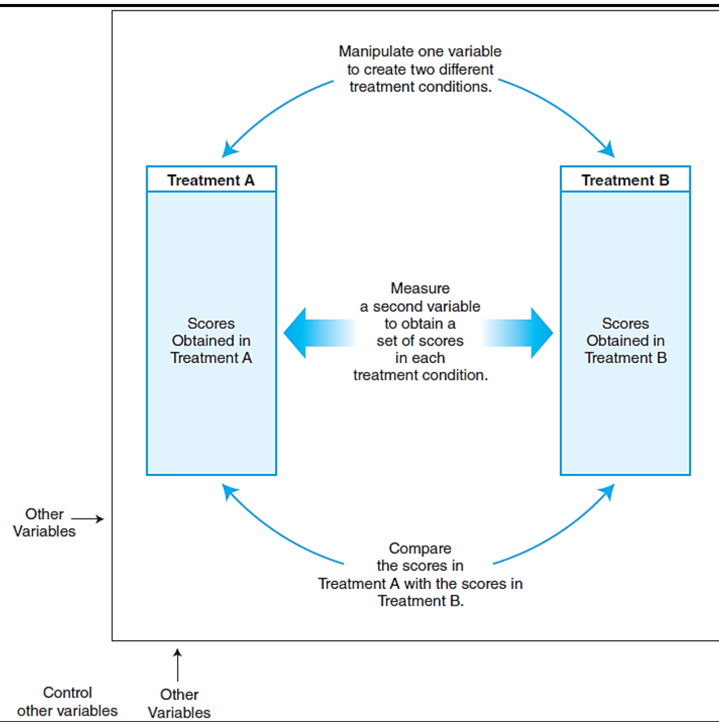
Comparison

The scores in one treatment condition are compared with the scores in another treatment condition

Control

All other variables are controlled to be sure that they do not influence the two variables being examined

8



9

Terminology for Experiments

Independent variable (IV)

The variable that is manipulated by the researcher

By generating a set of treatment conditions or levels

Dependent variable (DV)

The variable that is measured in each of the treatment conditions

Observed for changes to assess the effects of manipulating the IV

Extraneous variables

All other variables other than the IVs and DVs

10

IVs and DVs: Example 1

In an experiment examining immersion in movies, two groups of participants were used.

One group watched Harry Potter and Deathly Hallows Part 1 for 30 minutes on a 4.5-inch phone and the second group watched the same movie for 30 minutes on a 13-inch screen.

Then, both groups provided immersion ratings using the Immersive Experiences Questionnaire (to indicate how immersed they were in the movie).

For this experiment, identify the independent variable and the dependent variable

11

IVs and DVs: Example 2

In an experiment examining immersion in movies, two groups of participants were used.

One group watched Harry Potter and Deathly Hallows Part 1 for 30 minutes on a 13-inch screen and the second group watched the same movie for 60 minutes on a 13-inch screen.

Then, both groups provided immersion ratings using the Immersive Experiences Questionnaire (to indicate how immersed they were in the movie).

For this experiment, identify the independent variable and the dependent variable

12

IV:
Gaming
Platform

VR Gaming ← Levels → Desktop Gaming

14

IV:
Gaming
Platform

DV:
Game enjoyment

16

Inferential Questions

Does **screen size** affect **viewer immersion** when watching movies on Netflix?

Does **the new UI layout** enable users to **complete their tasks** faster than the current layout?

Does **gaming platform** affect **game enjoyment**?

17

Experimental Designs

18

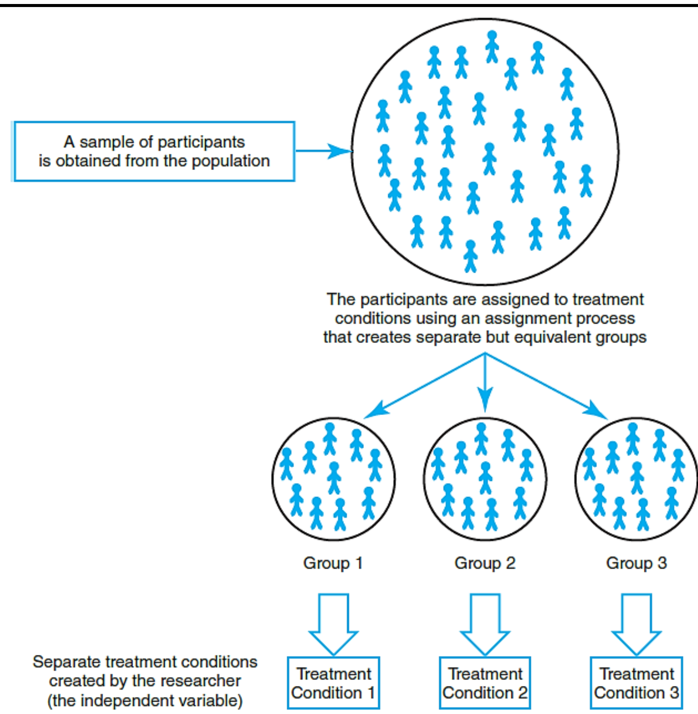
Between-Subjects Designs

Experimental design wherein a researcher manipulates an IV to generate different treatment conditions

A **separate** group of participants is assigned to each of the different conditions

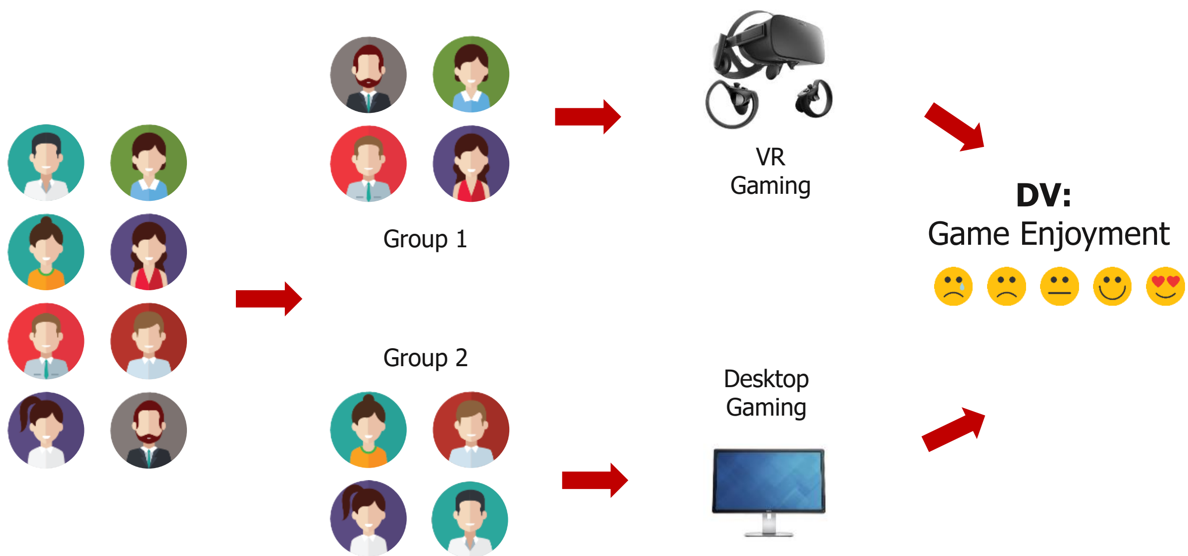
The DV for each participant is measured

19



20

Between-Subjects Design: Example



22

Within-Subjects Designs

Uses a **single group** of participants and tests each individual in all of the different treatments being compared

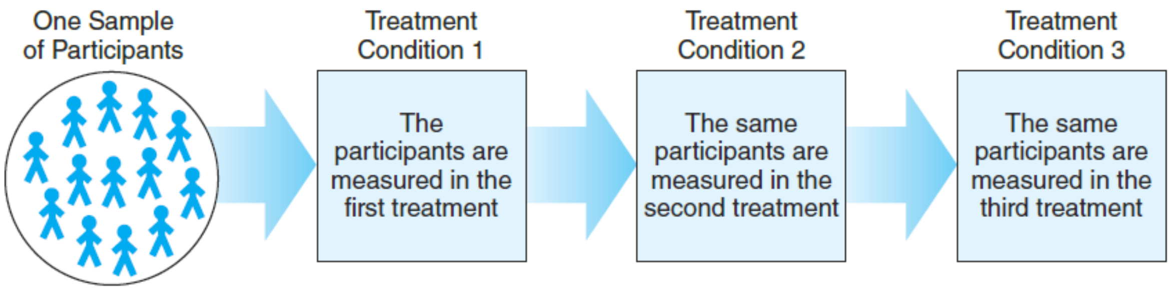
Same group of individuals participates in every level of the IV

- Study repeats measurements of the same individuals under different conditions (Often called a repeated-measures design)

Goal: to look for differences between treatment conditions within the same group of participants

23

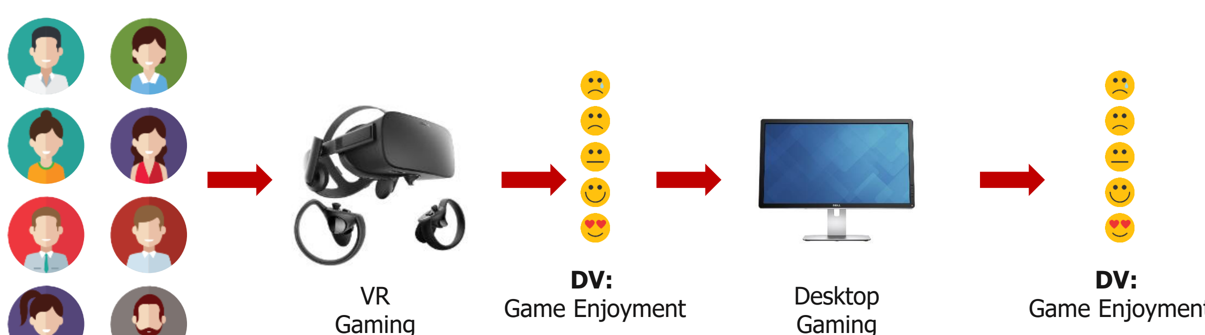
The Structure of a Within-Subjects Design



All participants go through the entire series of treatments but not necessarily in the same order.

24

Within-Subjects Design: Example



26

Hypothesis Testing

27

Hypotheses

Hypothesis:

A statement that describes or explains a relationship between variables; a conjectural explanation

Identifies the specific variables involved and describes how they are related

28

Hypotheses

Theory: People like jelly donuts best

Hypothesis: Given cream-filled donuts and jelly-filled donuts, people will rate jelly donuts as more enjoyable

Compared to cream-filled donuts, jelly-filled donuts will be rated as more enjoyable

29

Hypothesis Testing

A hypothesis test

A systematic procedure that determines whether the sample data provide convincing evidence to support the original research hypothesis

Goal

Rule out chance as a plausible explanation for the results

30

Hypothesis Testing

Attempts to distinguish between two explanations for the sample data:

1. Patterns in the data represent systematic relationships among variables in the population
   - IV is causing the changes in the DV
2. Patterns in the data were produced by random variation from chance or sampling error
   - Changes in the DV are caused by chance

31

Logic of Hypothesis Testing

State hypothesis about a population

Predict the expected characteristics of the sample based on the hypothesis

Obtain a random sample from the population

Compare the obtained sample data with the prediction made from the hypothesis

- If consistent, hypothesis is reasonable
- If discrepant, hypothesis is rejected

32

The Five Basic Elements of a Hypothesis Test

Null Hypothesis

Sample statistic

Standard error

Test statistic

Alpha level (level of significance)

33

Hypotheses

Null hypothesis (H_0)

States that, in the general population, there is no effect/difference/change/ or relationship

Alternative hypothesis (H_1)

States that there is an effect, change, a difference, or a relationship in the general population (your research hypothesis)

The null hypothesis is exactly the opposite of the research, or alternative, hypothesis.

34

Null Hypothesis: Example

Hypothesis:

VR gaming will lead to a better gaming user experience

Null hypothesis:

VR gaming will lead to a gaming user experience that is similar to desktop gaming experience

VR
Gaming

Desktop
Gaming

35

More on the Null Hypothesis

According to the null hypothesis, any patterns in the sample are nothing more than chance (sampling error)

According to the alternative hypothesis, the patterns observed in the data are caused by the manipulation of the IV

The hypothesis test determines whether the results are sufficient to reject the null hypothesis

Therefore aka Null Hypothesis Significance Testing (NHST)

36

The Five Basic Elements of a Hypothesis Test

Null Hypothesis

Sample statistic

Standard error

Test statistic

Alpha level (level of significance)

37

Sample Statistic

The data collected from the sample

The mean score for the dependent variable

Hypothesis:
VR gaming will lead to a better gaming user experience

Sample statistic:
The average of game enjoyment ratings

38

The Five Basic Elements of a Hypothesis Test

Null Hypothesis

Sample statistic

Standard error

Test statistic

Alpha level (level of significance)

39

Standard Error

A measure of the average, or standard, **distance** between a **sample statistic** and the corresponding **population parameter**.

Typically, there is some discrepancy between a sample statistic (what you measure in the sample) and the population parameter (the actual value in the population)

SE tells you how much discrepancy to expect

40

Sample Mean and Population Mean

We use samples as a way of estimating the value of the parameter in a population

The sample mean is an estimate of the population mean

Sampling Error

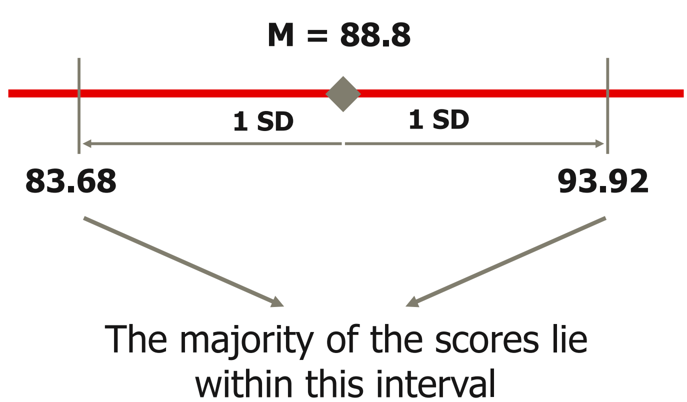
The discrepancy, or amount of error, that exists between a sample statistic and the corresponding population parameter

41

The Mean and Standard Deviation

Describe an entire distribution with just two numbers (M and SD)

The mean DADA scores was 88.8 (SD = 5.12)



42

DADA – Riddiculus Spell

Suppose the entire student population who performed this spell is composed of 100 students only

All students in the population performed the spell and Professor Lupin graded their performance

43

Population Scores

P01	83	**P21**	92	**P41**	92	**P61**	96	**P81**	81
P02	74	**P22**	87	**P42**	94	**P62**	84	**P82**	78
P03	84	**P23**	99	**P43**	95	**P63**	94	**P83**	96
P04	96	**P24**	83	**P44**	84	**P64**	90	**P84**	96
P05	84	**P25**	86	**P45**	80	**P65**	87	**P85**	81
P06	95	**P26**	85	**P46**	88	**P66**	90	**P86**	90
P07	100	**P27**	82	**P47**	87	**P67**	95	**P87**	100
P08	87	**P28**	95	**P48**	95	**P68**	94	**P88**	89
P09	89	**P29**	82	**P49**	83	**P69**	84	**P89**	92
P10	86	**P30**	91	**P50**	81	**P70**	94	**P90**	78
P11	84	**P31**	85	**P51**	99	**P71**	84	**P91**	88
P12	85	**P32**	78	**P52**	88	**P72**	100	**P92**	80
P13	85	**P33**	95	**P53**	82	**P73**	92	**P93**	92
P14	88	**P34**	84	**P54**	94	**P74**	89	**P94**	90
P15	85	**P35**	89	**P55**	78	**P75**	85	**P95**	89
P16	93	**P36**	82	**P56**	86	**P76**	81	**P96**	79
P17	85	**P37**	85	**P57**	94	**P77**	81	**P97**	85
P18	86	**P38**	85	**P58**	88	**P78**	89	**P98**	88
P19	91	**P39**	75	**P59**	84	**P79**	76	**P99**	82
P20	100	**P40**	87	**P60**	89	**P80**	88	**P100**	85

44

DADA – Riddiculus Spell

Suppose the entire student population who performed this spell is composed of 100 students only

All students in the population (N = 100) performed the spell

Population DADA Scores: $\mu = 86.8$ and $\sigma = 5.98$

Our sample was composed of 10 students

Sample DADA Scores: M = 88.8 and SD = 5.12

This sample was one of many possible samples that could be taken from the entire population of 100 students who performed the spell

45

Sample 1		Sample 2		Sample 3		Sample 4		Sample 5	
P26	90	P14	93	P31	93	P55	84	P97	93
P15	87	P85	91	P66	91	P77	89	P73	83
P74	92	P77	89	P51	83	P02	88	P74	92
P05	89	P70	95	P26	90	P99	93	P86	92
P96	94	P67	90	P61	85	P66	91	P20	91
P97	93	P05	89	P01	92	P14	93	P32	87
P27	79	P44	85	P06	93	P80	88	P55	84
P84	81	P49	88	P37	87	P78	83	P70	95
P46	84	P61	85	P77	89	P54	90	P65	93
P80	88	P42	78	P82	72	P12	90	P92	82
***M* = 87.7**		***M* = 88.3**		***M* = 87.5**		***M* = 88.9**		***M* = 89.2**	

Sample 6		Sample 7		Sample 8		Sample 9		Sample 10	
P89	87	P26	90	P10	85	P02	88	P31	93
P63	80	P86	92	P03	82	P18	80	P13	86
P23	90	P47	93	P28	95	P73	83	P33	87
P27	79	P06	93	P60	81	P06	93	P30	92
P09	86	P87	83	P44	85	P40	80	P47	93
P06	93	P82	72	P55	84	P14	93	P26	90
P16	93	P41	94	P80	88	P55	84	P86	92
P66	91	P74	92	P99	93	P48	89	P92	82
P69	89	P55	84	P38	75	P20	91	P21	95
P02	88	P72	83	P67	90	P89	87	P05	89
***M* = 87.6**		***M* = 87.6**		***M* = 85.8**		***M* = 86.8**		***M* = 89.9**	

46

Sample Means

Samples are variable; two samples are very, very rarely identical

Samples differ from each other

- Given a random sample it is unlikely that sample means would always be the same

We can present these sample means as a distribution

47

Distribution of Sample Means

The distribution of sample means is the collection of sample means for **all** the possible random samples of a particular size (*n*) that can be obtained from a population

In theory you need to imagine that we're taking hundreds or thousands of samples to construct a sampling distribution

This means that if we took the average of **all** sample means of the same size (n) we'd get the value of the population mean

Sample	M
1	87.7
2	88.3
3	87.5
4	88.9
5	89.2
6	87.6
7	87.6
8	85.8
9	86.8
10	89.9

M = 87.9

This is not equal to population mean because n = 10

48

Variability of Sample Means

Variability of a distribution of *scores* is measured by the standard deviation of the scores

Variability of a distribution of *sample means* is measured by the **standard deviation** of the *sample means*

This is called the **standard error of sample means**

Sample	**M**
1	87.7
2	88.3
3	87.5
4	88.9
5	89.2
6	87.6
7	87.6
8	85.8
9	86.8
10	89.9

M = 87.9

SD = 1.13

49

Standard Error of the Mean

The standard error of the mean is the standard deviation of sample means

Provides a measure of how much distance is expected on average between a sample mean and a population mean

- A measure of how representative a sample is likely to be of the population

Referred to as standard error of sample means or standard error of the mean, or standard error

Denoted by SE or SEM $$SE_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

50

Calculating Standard Error of a Sample Mean

The distribution of DADA scores:

80, 82, 86, 87, 89, 90, 92, 92, 94, 96

M = 88.8, *SD* = 5.12

n = 10

Standard Error = $SE_{\bar{x}} = \frac{s_x}{\sqrt{n}}$ = 5.12 / $\sqrt{10}$

SE = 1.62 → **The amount of average distance between the mean DADA score of 88.8, obtained from the sample of 10 students, and the mean DADA score all students in the population**

51

Standard Error of the Mean

A large standard error (relative to the sample mean) means that there is a lot of variability between the means of different samples

the sample we have might not be representative of the population

A small standard error indicates that most sample means are similar to the population mean

our sample is likely to be an accurate reflection of the population

52

Standard Deviation vs. Standard Error

The standard deviation describes scores in terms of average distance from the mean

- Tells us how spread out the individual scores are from the mean

The standard error describes the standard deviation of sample means

- Tells us how close our sample mean is to the true mean of the overall population

53

The Five Basic Elements of a Hypothesis Test

Null Hypothesis

Sample statistic

Standard error

Test statistic

Alpha level (level of significance)

54

Test Statistic

A summary value that measures the degree to which the sample data are in accordance with the null hypothesis (no difference)

Compares the sample statistic with the null hypothesis, using the standard error as a baseline

$$\text{Test statistic} = \frac{\text{Sample statistic} - \text{Parameter from the null hypothesis}}{\text{Standard error}}$$

$$= \frac{\text{Actual difference between the data and the hypothesis}}{\text{Difference expected by chance}}$$

55

Test Statistic

A large value for the test statistic (much larger than 1) indicates a large discrepancy between the sample statistic and the parameter specified by **the null hypothesis**, and leads to rejecting the null hypothesis

$$\text{Test statistic} = \frac{\text{Sample statistic} - \text{Parameter from the null hypothesis}}{\text{Standard error}}$$

$$= \frac{\text{Actual difference between the data and the hypothesis}}{\text{Difference expected by chance}}$$

56

The Five Basic Elements of a Hypothesis Test

Null Hypothesis

Sample statistic

Standard error

Test statistic

Alpha level (level of significance)

57

Alpha Level (Level of Significance)

Provides a criterion for interpreting the test statistic

The alpha level, or level of significance, for a hypothesis test is the maximum probability that the research result was obtained simply by chance

An alpha level of .05 means that the test demands that there is less than a 5% (.05) probability that the results are caused only by chance

58

Alpha Level (Level of Significance)

The goal of a hypothesis test is to rule out chance as a plausible explanation for the results

To achieve this, we need to determine

- which results are reasonable to expect just by chance (without any effect of the IV)
- which results are extremely unlikely to be obtained by chance alone

The alpha level

- a probability value that defines what is extremely unlikely
- the probability that the sample results would be obtained even if the null hypothesis were true

59

Alpha Level (Level of Significance)

In the literature, significance levels are reported as *p* values

By convention, alpha levels are set at .05, .01, or .001

- If your *p* value is smaller than your alpha level, your result is significant

p < .05:

There is less than a .05 probability that the result is caused by chance (or there is a 5% risk that the result is just a result of chance)

Since the probability is so small, you can reject the null hypothesis

60

Alpha Level (Level of Significance)

When the results of a research study satisfy the criterion imposed by the alpha level, the results are said to be **significant**, or **statistically significant**

A significant result, or a statistically significant result, means that it is extremely unlikely that the research result was obtained simply by chance.

The smaller the level of significance, the more confidence you have in the result (that the difference did not occur by chance)

61

Hypothesis Testing: A Magical Example

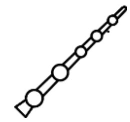
https://www.youtube.com/watch?v=aJSh1zkPEvc

63

Hypothesis Testing: A Magical Example

Theory:

"Elder wand is more powerful than any in existence."



Experimental task:

Light candles using the *Incendio* charm

Experimental hypothesis:

Wizards will light more candles when using the Elder Wand than any other wand

64

Hypothesis Testing: A Magical Example

100 wizards perform the *Incendio* charm using both the Elder Wand and a regular wand

We count the number of candles they were able to light using each wand

65

Hypothesis Testing: A Magical Example

Imagine that we found that 75% of wizards could light more candles using the Elder Wand than using the regular wand

Based on the null hypothesis, assuming that the Elder Wand is no different whatsoever, is it likely that 75% of wizards would be able to light more candles?

- Chances are very low
- If H_0 is true, they should light the same number of candles using both wands

Thus, it is very unlikely that we would have obtained these results if were H_0 true, increasing our confidence in H_1

66

Hypothesis Testing: A Magical Example

What if we found that only 1 wizard (1%) could light more candles using the Elder Wand than using the regular wand

Based on the null hypothesis, assuming that the Elder Wand is no different whatsoever, is it likely that 1% of wizards would be able to light more candles?

- Chances are higher
- If H_0 is true, no wizards should be able to light more candles using the Elder Wand

Thus, the chances of getting these data if H_0 were true are higher than before, supporting $H_{0.}$

67

Errors in Hypothesis Testing

Hypothesis testing is an inferential process

- Uses limited information from a sample to make a statistical decision, and then from it a general conclusion
- Sample data used to make the statistical decision allows us to make an inference and draw a conclusion about a population

Errors are possible

- Type I Error
- Type II Error

68

Errors in Hypothesis Testing: Type I Error

The mistake of rejecting a null hypothesis that is actually true and should not be rejected

Implication:
Researcher concludes that a difference exists when there is none

Alpha level is the probability that a test will lead to a Type I error

A false report: a researcher finds evidence for a significant result when there is no effect in the population

- An unusual or extreme sample leads to this error

69

Errors in Hypothesis Testing: Type II Error

The mistake of not rejecting a null hypothesis that is really false and should be rejected

Implication: Researcher has failed to detect a real difference

Type II error probability is not easily identified

- Occurs when the effect is so small that it does not show up in the sample

The key to controlling Type II error is the sample size

- Large sample sizes address the power issue and reduce Type II error rates

70

Five Steps in Hypothesis Testing

Step 1: State the hypotheses

- Avoid HARKing

Step 2: Set the criteria for a decision

- Determine alpha level (generally alpha = .05 in most sciences)

Step 3: Collect/retrieve data

- Always after stating your hypothesis and setting the criteria (ensures objectivity)

Step 4: Compute sample statistics

- Choose and conduct your statistical test

Step 5: Make a decision

- If p < alpha level, you reject the null hypothesis
  If not, you fail to reject the null hypothesis

71

Choosing Statistical Tests for NHST

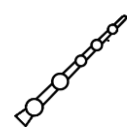
Experiment Design	# of IVs	# of Conditions for each IV	Statistical Test
	1	2	Independent-samples *t* test
Between-subjects	1	3 or more	One-way ANOVA
	2 or more	2 or more	Factorial ANOVA
	1	2	Paired-samples *t* test
Within-subjects	1	3 or more	Repeated measures ANOVA
	2 or more	2 or more	Repeated measures ANOVA
Between- and within-subjects	2 or more	2 or more	Mixed (split-plot) ANOVA

72

Hypothesis Testing: A Magical Example

Theory:

"Elder wand is more powerful than any in existence."



Experimental task:

Light candles using the *Incendio* charm

Experimental hypothesis:

Wizards will light more candles when using the Elder Wand than any other wand

74

Hypothesis Testing: A Magical Example

100 wizards perform the *Incendio* charm using both the Elder Wand and a regular wand

We count the number of candles they were able to light using each wand

75

Hypothesis Testing: A Magical Example

Imagine that we found that 75% of wizards could light more candles using the Elder Wand than using the regular wand

Based on the null hypothesis, assuming that the Elder Wand is no different whatsoever, is it likely that 75% of wizards would be able to light more candles?

- Chances are very low
- If H_0 is true, they should light the same number of candles using both wands

Thus, it is very unlikely that we would have obtained these results if were H_0 true, increasing our confidence in H_1

76

Hypothesis Testing: A Magical Example

What if we found that only 1 wizard (1%) could light more candles using the Elder Wand than using the regular wand

Based on the null hypothesis, assuming that the Elder Wand is no different whatsoever, is it likely that 1% of wizards would be able to light more candles?

- Chances are higher
- If H_0 is true, no wizards should be able to light more candles using the Elder Wand

Thus, the chances of getting these data if H_0 were true are higher than before, supporting H_0.

77

Five Steps in Hypothesis Testing

Step 1: State the hypotheses

- Avoid HARKing

Step 2: Set the criteria for a decision

- Determine alpha level (generally alpha = .05 in most sciences)

Step 3: Collect/retrieve data

- Always after stating your hypothesis and setting the criteria (ensures objectivity)

Step 4: Compute sample statistics

- Choose and conduct your statistical test

Step 5: Make a decision

- If p < alpha level, you reject the null hypothesis
- If not, you fail to reject the null hypothesis

78

t Tests

Used to compare two means to see if they are **significantly** different from each other

Hypotheses

H_0: No difference between the two means ($\mu_A = \mu_B$)

H_A: The two means are different ($\mu_A \neq \mu_B$)

Two common *t* tests

Independent-samples *t* test

Paired-samples (or dependent-samples) *t* test

79

t Tests

Used to compare two means to see if they are **significantly** different from each other

Hypotheses

H_0: No difference between the two means ($\mu_A = \mu_B$)

H_A: The two means are different ($\mu_A \neq \mu_B$)

Two common *t* tests

Independent-samples *t* test

Paired-samples (or dependent-samples) *t* test

80

Independent-Samples *t* Test

81

Independent-Samples *t* Test

Used for between-subjects designs
- One categorical IV with two levels
- One quantitative DV measured in each condition for different groups of participants

The mean value of DV is calculated for each group

The sample mean difference is used to test a hypothesis about the corresponding population mean difference

The null hypothesis states that the population mean difference is zero

82

Independent-Samples *t* Test

The sample mean difference is used to test a hypothesis about the corresponding population mean difference

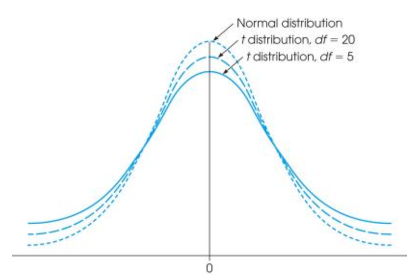
The null hypothesis states that the population mean difference is zero

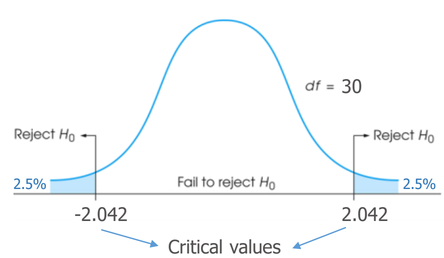
$$t = \frac{\overline{X}_1 - \overline{X}_2}{\text{estimate of the standard error}} \qquad t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)}}$$

83

***t* Table**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										





84

Hypothesis Testing: A Magical Example

Research question:

Is the Elder Wand more powerful than any in existence?

IV: Wand

Dependent variable:

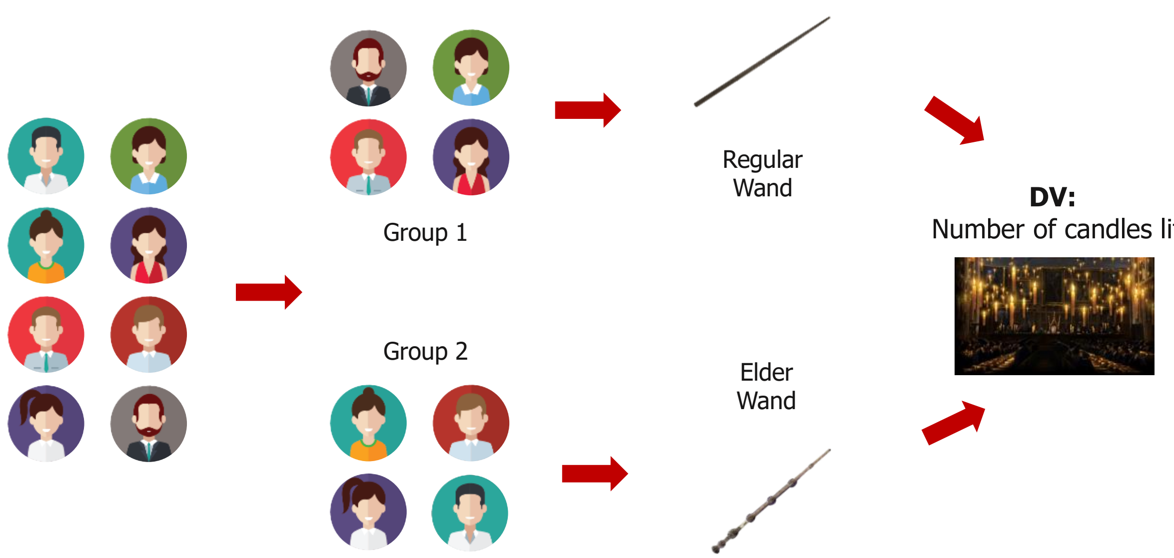
Number of candles lit

Experimental hypothesis:

Wizards will light more candles when using the Elder Wand compared to when using the regular wand

85

Experimental Procedure



86

Choosing Statistical Tests for NHST

Experiment Design	# of IVs	# of Conditions for each IV	Statistical Test
Between-subjects	**1**	**2**	**Independent-samples *t* test**
	1	3 or more	One-way ANOVA
	2 or more	2 or more	Factorial ANOVA
Within-subjects	1	2	Paired-samples *t* test
	1	3 or more	Repeated measures ANOVA
	2 or more	2 or more	Repeated measures ANOVA
Between- and within-subjects	2 or more	2 or more	Mixed (split-plot) ANOVA

87

88

Independent-Samples *t* Test

Assumptions

1. The variances of the dependent variable in the two groups are roughly equal
2. The dependent variable is approximately normally distributed

SciPy allows you to check for these assumptions

Assumption 1: Levene's Test of Equality of Variances

Assumption 2: Shapiro-Wilk Test of Normality

You want **non-significant** results from these tests ($p > .05$)!

89

Reporting An Independent-Samples *t* Test

Report the results of checking for assumptions

Report *t* statistic, degrees of freedom (*df*), *p*-value

$t(df) = x$, $p = y$ or $p < .05$

Report the mean and standard error of the DV for each group

Group 1: ($n = a$, $M = b$, $SE = c$)

Group 2: ($n = x$, $M = y$, $SE = z$)

90

Reporting Assumptions

For an independent-samples *t* test

Assumptions of normality, as assessed by Shapiro-Wilk's test ($p > .05$) and homogeneity (equality) of variances, as assessed by Levene's test ($p > .05$) were met.

* Usually precedes the results of the independent-samples *t* test

91

Reporting An Independent-Samples *t* Test

Test & Purpose

An independent sample *t* test was conducted to compare the number of candles lit by wizards using the Elder Wand and using the regular wand.

Actual results

Results showed a statistically significant difference between the wands, $t(38) = 8.74$, $p < .001$. An examination of the average number of candles lit in each wand condition revealed that the wizards who used the Elder Wand could light a greater number of candles ($n = 20$, $M = 21.1$, $SE = .49$) compared to those who used the regular wand ($n = 20$, $M = 15.6$, $SE = .40$).

Meaning

These results indicate that the Elder Wand is more powerful than a regular wand in terms of enabling wizards to perform the *Incendio* spell.

92

Experimental Procedure

Elder
Wand

DV:
Number of
candles lit

Regular
Wand

DV:
Number of
candles lit

93