Experimental Design and Analysis

Instructor: Mark Hancock

What is an experiment?

You will be able to describe some of the key elements of an experiment that can be analyzed with a statistical test.

Elements of an Experiment

- People
- Data
  - Measurement
- Hypothesis

- (There are more… we'll learn about them later)

People

- Sample (participants)
  - People in your study

- Population
  - E.g., Canadians, computer scientists, artists
  - People we want to *generalize* to

Data

- Variable
  - E.g., technique, task time, number of errors
- Statistic
  - Mean, median, mode, standard deviation, etc.
  - Taken from the *sample*
- Parameter
  - Taken from the *population*

Hypothesis (examples)

- The average temperature in Calgary is less than -20°C.
- A pair of dice will result in a roll of 7 more than it will result in a roll of 10.
- Canadians prefer Hockey to Baseball.

Hypotheses

- Carman is a great foosball player.
- Carman is better at foosball than Mark.
- Carman wins more foosball games than Mark.
- Mark scores more points in foosball than Carman.

Hypotheses

- TreeMaps are easy to use.
- TreeMaps are better than Phylotrees.
- People find leaf nodes faster with TreeMaps than with Phylotrees.
- People find sibling nodes faster with Phylotrees than with TreeMaps.

Null Hypothesis

"… the null hypothesis is a *pinpoint* statement as to the unknown quantitative value of the *parameter* in the *population[s]* of interest."

Huck, S.W. *Reading Statistics and Research*

Null Hypothesis

- Calgary Temperature:
  - $\mu_{Calgary}$ = -20˚C
- Dice Rolling:
  - $\mu_7 = \mu_{10}$
  - $\mu_7 - \mu_{10} = 0$
- Foosball:
  - $\mu_{Carman} = \mu_{Mark}$
- Tree Vis:
  - $\mu_{TreeMap} = \mu_{Phylotrees}$

What is the null hypothesis?

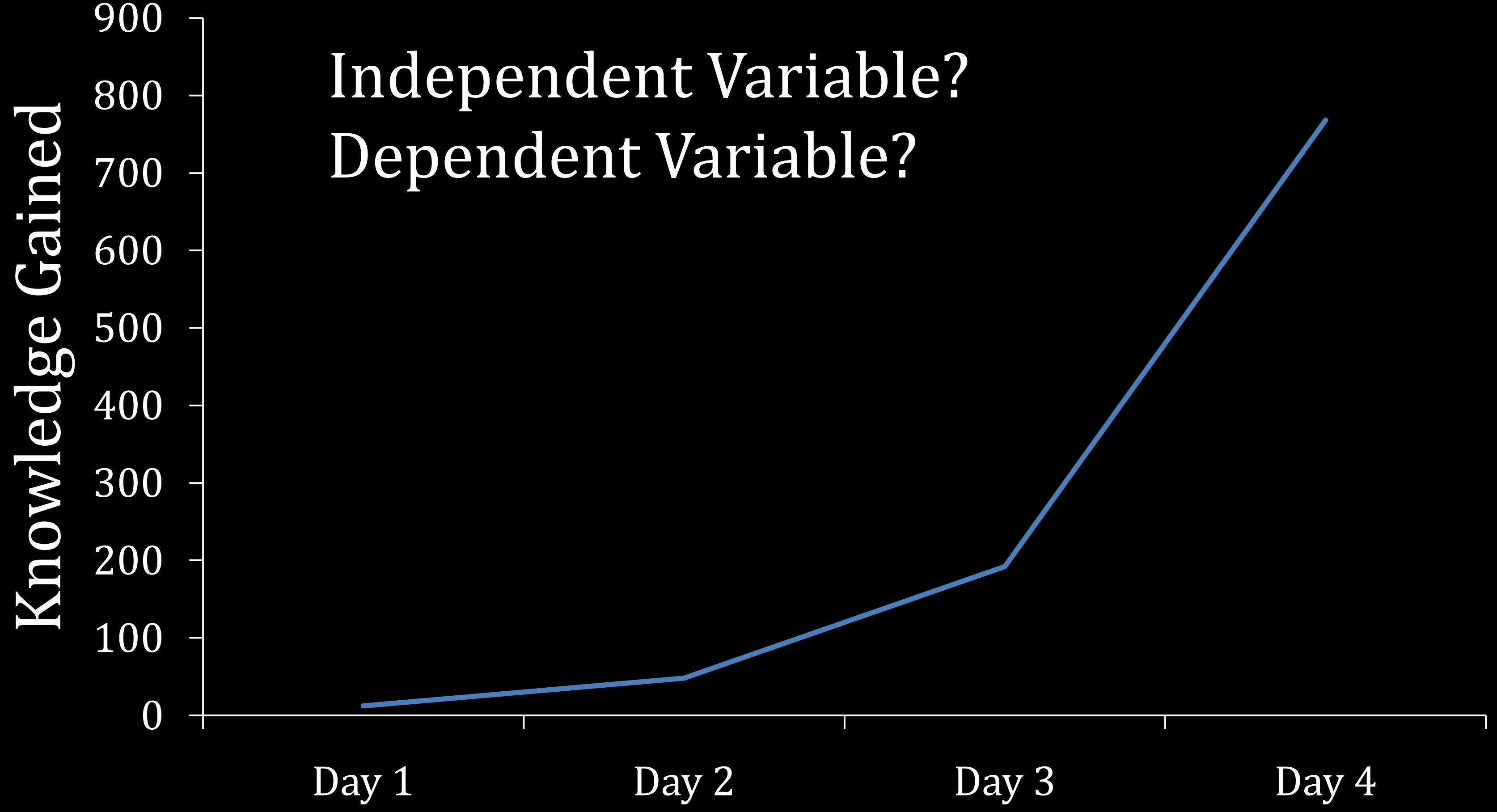
- Hypothesis: The average temperature in Vancouver is higher than the average temperature in Calgary

$$H_0: \mu_{Calgary} = \mu_{Vancouver}$$

Elements of an Experiment

- People
  - Sample/Participants
  - Population
- Data
  - Variables
  - Statistics
  - Parameters
- Hypotheses
  - Null Hypothesis

Label this graph:



Variables

- Independent Variables (Factors)
  – What you (the experimenter) are changing during the experiment
- Dependent Variables (Measures)
  – What is being measured
- Constants
  – What is the same for all participants

Variables (Exercise)

Questionnaire: ask computer science students to rate their favourite teacher.

- What independent variables would you use?
- What dependent variables would you use?
- What would you keep constant?

Problem: hypotheses are about *population*, but we only have access to data from a *sample*.

You will be able to describe why the *Law of Large Numbers* and the *Central Limit Theorem* allow us to make general statements about a population based on information about a sample.

Slides by Mark Hancock

Dice Rolling

- Roll one die
  - Predictions?

- Roll one die 5 times and take the average?
  - Predictions?

- Roll one die 100 times and take the average?
  - Predictions?

Dice Rolling

- Roll one die n times:
  - Possible outcomes: 1, 2, 3, 4, 5, 6
  - Probability of rolling X: $P(X) = 1/6 = 16.7\%$
  - Expected Value:

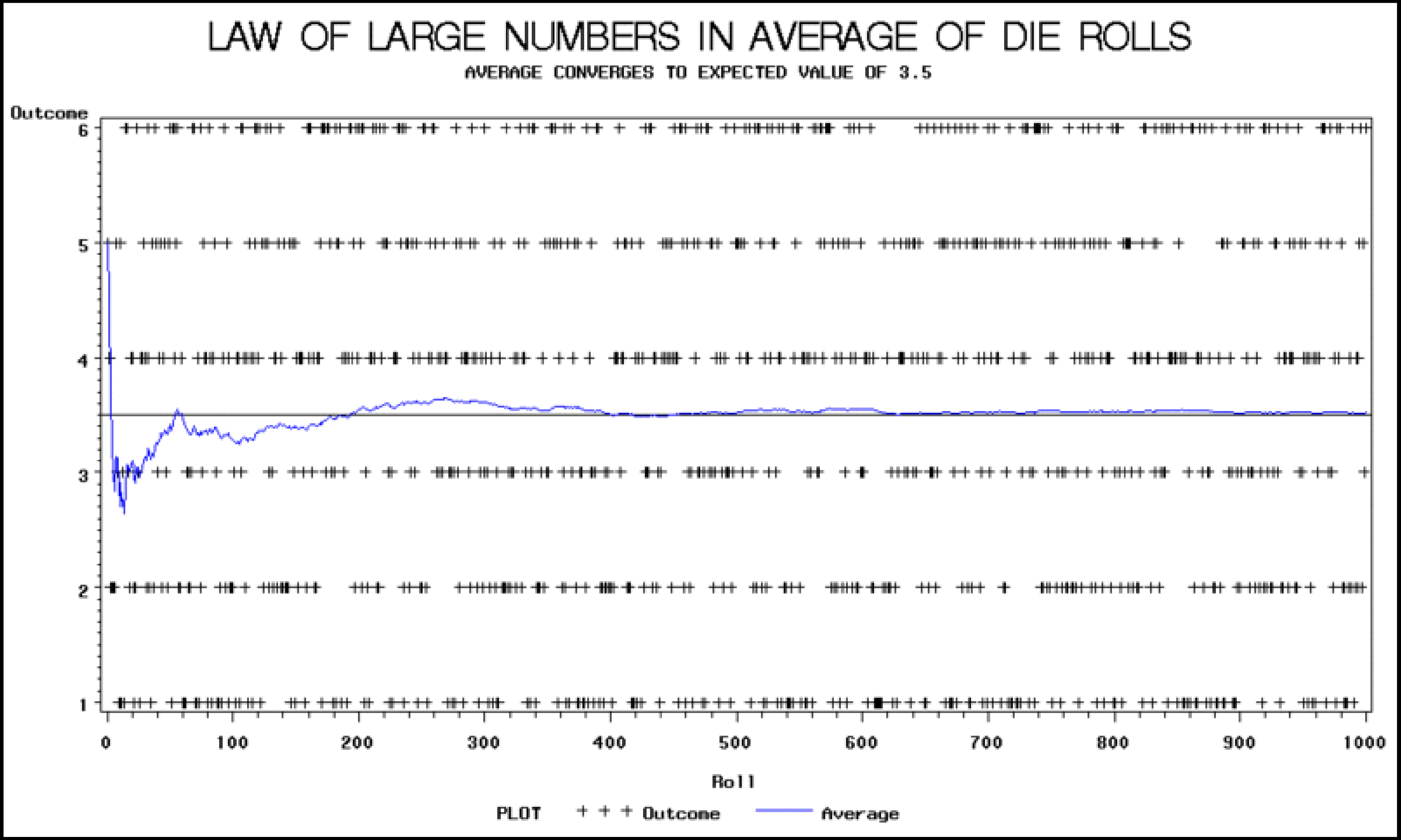
$$E(X) = 1(1/6) + 2(1/6) + \ldots + 6(1/6) = 3.5$$

Law of Large Numbers

"Given a sample of *independent* and identically distributed random variables with a finite *expected value*, the *average* of these observations will *eventually approach* and stay close to the expected value."

"Law of large numbers." *Wikipedia*

Experiment



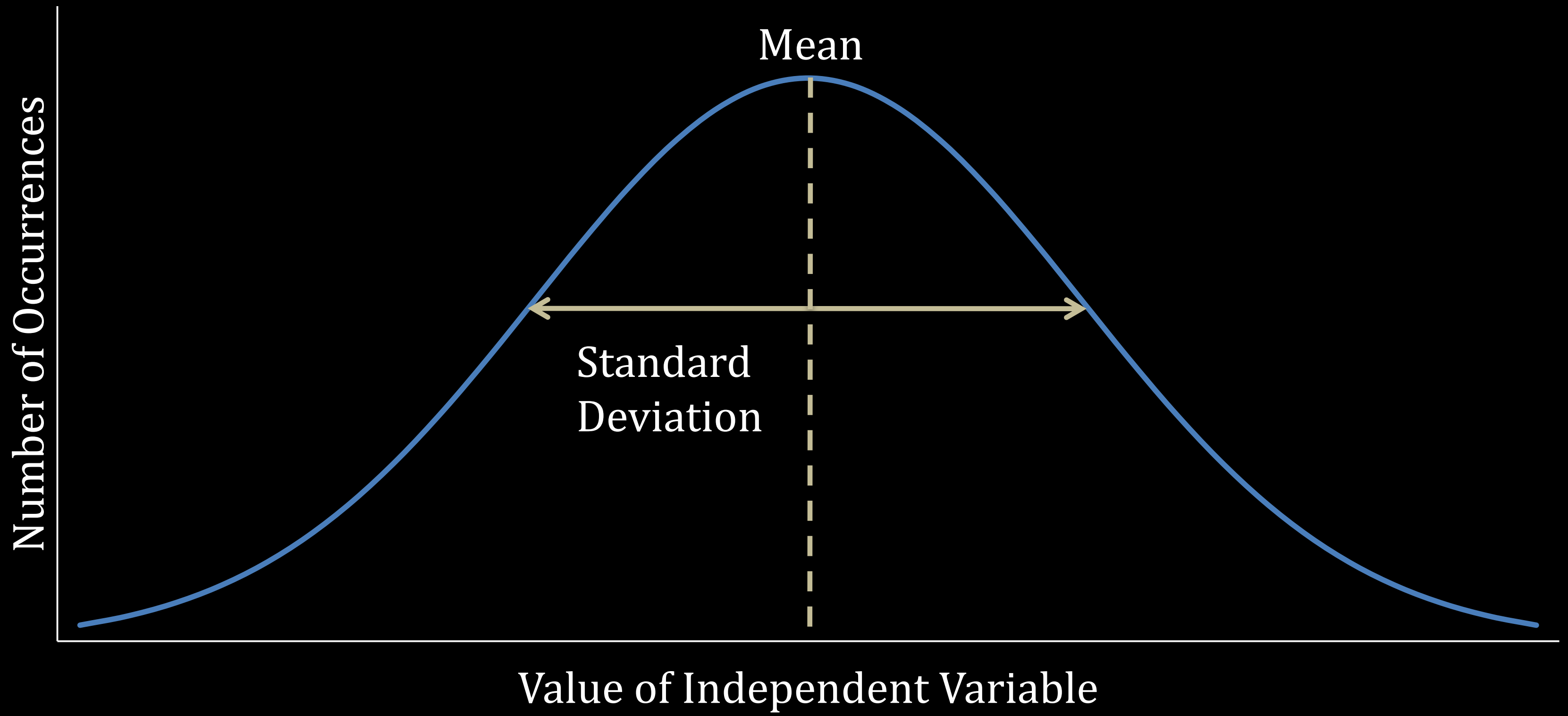
"Law of Large Numbers", *Wikipedia*

Central Limit Theorem

"…if the sum of *independent* identically distributed random variables has a finite *variance*, then it will be approximately *normally distributed*."

"Central Limit Theorem." *Wikipedia*

Gaussian/Normal Distribution



LLN vs. CLT

Law of Large Numbers

"Given a sample of *independent* and identically distributed random variables with a finite *expected value*, the average of these observations will *eventually approach* and stay close to the expected value."

Central Limit Theorem

- "…if the sum of *independent* identically distributed random variables has a finite *variance*, then it will be approximately *normally distributed*."

Generalize from μ_{sample} to $\mu_{population}$

- What information do we have?
  - Sample mean
  - Sample variance

- What information do we seek?
  - Population mean
  - Population variance

Generalize from μ_{sample} to $\mu_{population}$

- Assumptions:
  - population has expected value of $\mu_{population}$
  - population has finite variance $\sigma_{population}$

- Conclude:
  - provided we have enough people (N is large):

$$\mu_{sample} \to \mu_{population} \text{ (by LLN)}$$

$$\sigma_{sample} \to \sigma_{population} \text{ (by CLT)}$$

Summary

- Dependent/independent variables
- Constants
- Law of Large Numbers:
  - eventually data tends to the *expected value*
- Central Limit Theorem:
  - most data tends toward a *normal distribution*

Break: 15 Minutes

Significance and Power

Slides by Mark Hancock

You will be able to identify two types of errors and be able to avoid these errors when running a study.

Types of mistakes

1. Find a difference when there isn't one

2. Find no difference when there is one

Rejecting the Null Hypothesis (H_0)

Decision \ Reality	H_0 true	H_0 false
H_0 true	✓	Type II
H_0 false	Type I	✓

Slides by Mark Hancock

Rejecting the Null Hypothesis (H_0)

Decision \ Reality	H_0 true	H_0 false
H_0 true	✓	β
H_0 false	α or p	✓

- Significance (α)
  - calculated *after* the experiment
- Power ($1 - \beta$)
  - calculated *before* (a priori) or *after* (post hoc)
  - depends on *effect size* and *sample size*

How do we avoid these errors?

1. Decide before the analysis how acceptable this would be (e.g., $p < .05$).

2. The smaller the effect size you expect, the larger sample size you need.

(Student's) T-Test

Who is attributed with the discovery of the Student's T-Test?

- A student!
  - William Sealy Gosset
- Guinness Brewery employee
- Monitored beer quality

You will be able to formulate the appropriate *null hypothesis* and calculate the *t-value* for data from a sample.

Null Hypotheses

- $\mu = \mu_0$ (constant value)

- $\mu_A = \mu_B$

Assumptions

- Data is distributed *normally*
- Equal variance: $\sigma_A = \sigma_B$ (for second H_0)

Example

- Independent Variable:
  – TreeMap vs. Phylotrees
- Dependent Variable:
  – Time to find a leaf node
- Data:
  – 30 people used TreeMap, 30 used Phylotrees
  – Found one leaf node each

Check the Null Hypothesis

- Null Hypothesis (for *population*):

$$\mu_{\mathrm{T}} = \mu_{\mathrm{P}}$$

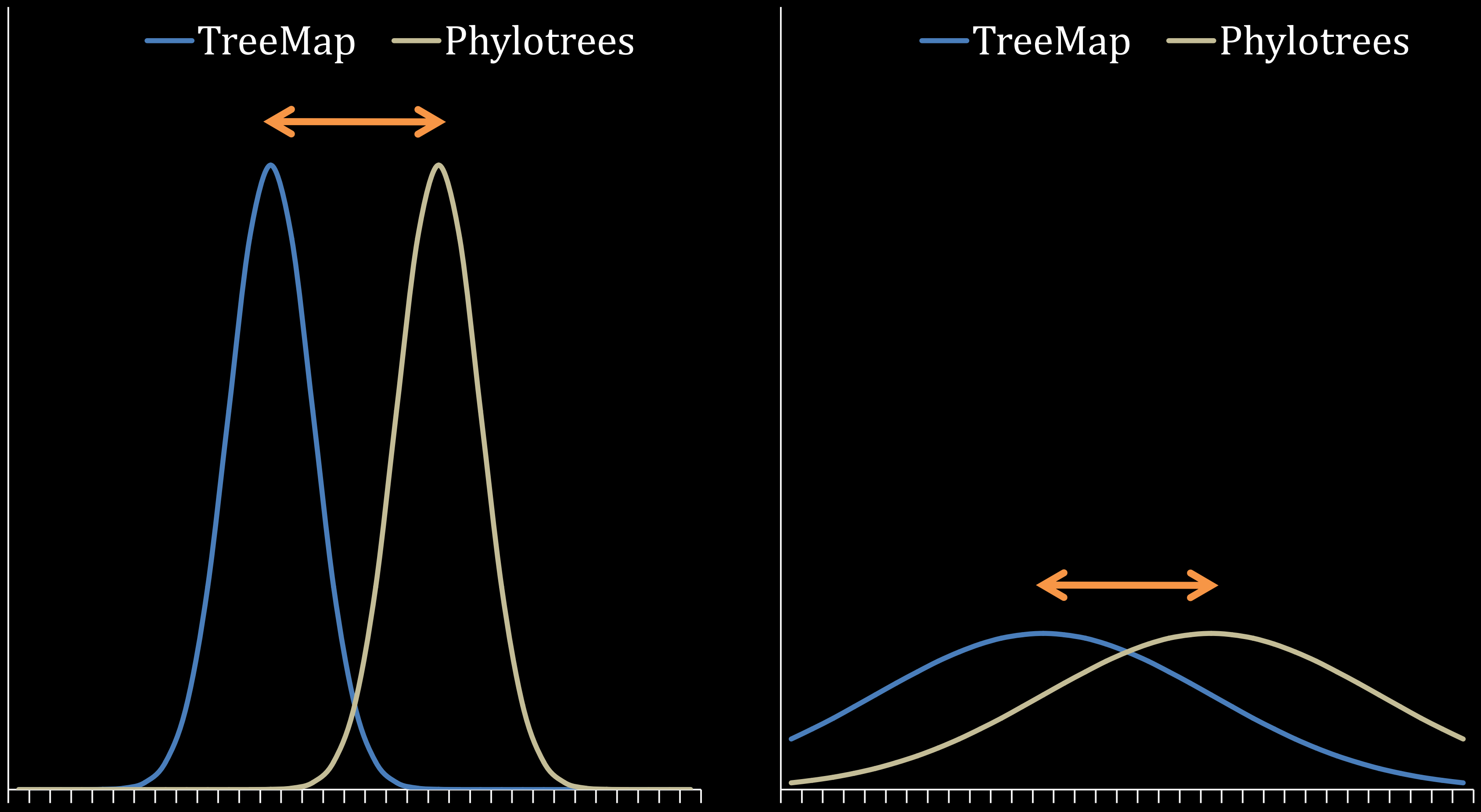
or

$$\mu_{\mathrm{T}} - \mu_{\mathrm{P}} = 0$$

- Test A (for *sample*):

check value of $\mu_{\mathrm{T}} - \mu_{\mathrm{P}}$

How is the data distributed?



Sample Distribution

How do you account for the differences in variance?

$$t = \frac{\text{mean difference}}{\text{variability}}$$

$$t = \frac{\mu_T - \mu_P}{\text{variability}}$$

$$\text{variability} = \sqrt{\frac{\sigma_T^2}{n_T} + \frac{\sigma_P^2}{n_P}}$$

a.k.a. Standard Error of difference between means

$$\text{important ratio} = \frac{\text{difference in experiment variables}}{\text{differenc in error}}$$

Interpreting the T Ratio

- What makes the ratio large?
  1. Larger difference
  2. Smaller variance

- Large t => more likely to be a real difference

How do we find significance?

- Look up in a table (the math is too hard for humans to do)

- Pick a level of significance (e.g., α = .05) and find the row corresponding to your sample size ($df = n - 1$).
- If t > (value in that cell), then $p < \alpha$

df	0.2	0.1	0.05	0.025	0.02	0.01	0.005
1	1.376	3.078	6.314	12.706	15.894	31.821	63.656
2	1.061	1.886	2.920	4.303	4.849	6.965	9.925
3	0.978	1.638	2.353	3.182	3.482	4.541	5.841
4	0.941	1.533	2.132	2.776	2.999	3.747	4.604
5	0.920	1.476	2.015	2.571	2.757	3.365	4.032
				...			
95	0.845	1.291	1.661	1.985	2.082	2.366	2.629
96	0.845	1.290	1.661	1.985	2.082	2.366	2.628
97	0.845	1.290	1.661	1.985	2.082	2.365	2.627
98	0.845	1.290	1.661	1.984	2.081	2.365	2.627
99	0.845	1.290	1.660	1.984	2.081	2.365	2.626
100	0.845	1.290	1.660	1.984	2.081	2.364	2.626
				...			

Break: 20 Minutes

Analysis of Variance (ANOVA)

Slides by Mark Hancock

You will be able to formulate the appropriate *null hypothesis* and calculate the *F-score* for data from a sample.

Null Hypotheses

- $\mu_A = \mu_B = \mu_c = \ldots$
- Remember: "the null hypothesis is a *pinpoint* statement "
- $\sigma_\mu = 0$

Assumptions

- Data is distributed *normally*
- *Homogeneity* of variance: $\sigma_A = \sigma_B = \sigma_C = \ldots$
- A, B, C, … are *independent* from one another

Example

- Independent Variable:
  - TreeMap vs. Phylotrees vs. ArcTrees
- Dependent Variable:
  - Time to find a leaf node
- Data:
  - 30 people used TreeMap, 30 used Phylotrees, 30 used ArcTrees
  - Found one leaf node each

Check the Null Hypothesis

- Null Hypothesis (for *population*):

$$\mu_{\mathrm{T}} = \mu_{\mathrm{P}} = \mu_{\mathrm{A}}$$

- Test A (for *sample*):

check H_0 for sample

we know this is not enough!

$$\text{important ratio} = \frac{\text{difference in experiment variables}}{\text{differenc in error}}$$

$$F = \frac{\text{variance of the means between each group}}{\text{variance of the means within each group}}$$

Degrees of Freedom (*df*)

- How many more pieces of data you need
- e.g., if you have μ, you need n-1 pieces of data to find the missing piece of data

Sum of Squares (*SS*)

- Measure of variance

$$\sum_{i=1}^{n}(x_i - \mu)^2$$

Mean Square

- "Mean" of the sum of squares

$$MS = \frac{SS}{df}$$

F-Score (Fisher's Test)

$$F = \frac{MS_{\text{between groups}}}{MS_{\text{within groups}}}$$

F-Score (Fisher's Test)

$$F = \frac{\text{variance of the means between each group}}{\text{variance of the means within each group}}$$

Example

	Degrees of Freedom	Sum of Squares	Mean Square	F
Between Groups	3	988.19	329.40	4.5
Within Groups	146	10679.72	73.15	
Total	149	11667.91		

- How many groups (i.e., how many means are we comparing)?
- How many total participants?
- Report as: $F(3,146) = 4.5$

Example

- "…The results of a one-way ANOVA indicated that UFOV [useful field of view] reduction increased with dementia severity, $F(2,52) = 15.36$, $MSe = 5371.5$, $p < .0001$.

- How many groups of participants were there?
- How many total participants were there?
- Fill in the table from before…

Example

	Degrees of Freedom	**Sum of Squares**	**Mean Square**	**F**
Between Groups	2	165,012.48	82,506.24	15.36
Within Groups	52	279,318	5,371.5	
Total	54	444,330.48		

Example

- Independent Variable:
  - TreeMap vs. Phylotrees vs. ArcTrees
- Data:
  - 30 people used TreeMap, 30 used Phylotrees, 30 used ArcTrees
  - Found one leaf node each

- Fill in the degrees of freedom column

Example

	Degrees of Freedom	Sum of Squares	Mean Square	F
Between Groups	2			
Within Groups	87			
Total	89			

What does large F mean?

- Remember:

$$F = \frac{MS_{\text{between groups}}}{MS_{\text{within groups}}}$$

- Consider the null hypothesis
- What does each value estimate?

F-table (α = .05)

		$df_{between}$									
		1	2	3	4	5	6	7	8	9	10
df_{within}	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98

Summary

- Analysis of Variance (ANOVA) is used to compare 2 or more means
- The F-score and *df* indicate the probability of a Type I error in rejecting the null hypothesis

Summary of First Day

- Elements of an experiment
- Null Hypothesis
- Variables (independent/dependent)
- Law of Large Numbers/Central Limit Theorem
- Significance and Power
- T-Test
- One-way ANOVA

Next Week

- Two-way & three-way ANOVA
- Non-parametric tests