Stanford MEDICINE

Department of Psychiatry and Behavioral Sciences

Machine Learning for Neuroimaging

Autumn 2023

Session 5 – 10/10/2023

Basics for statistical analysis

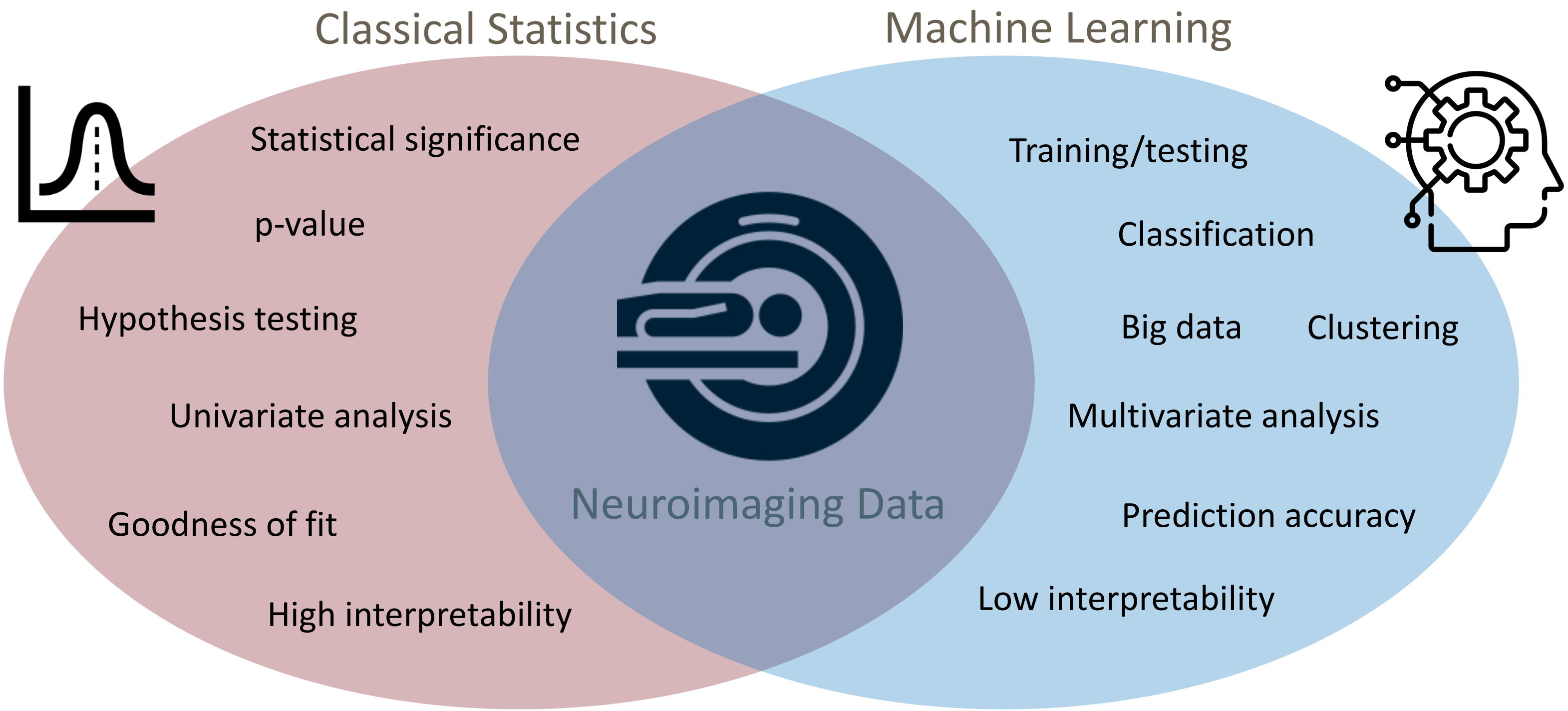
Today…

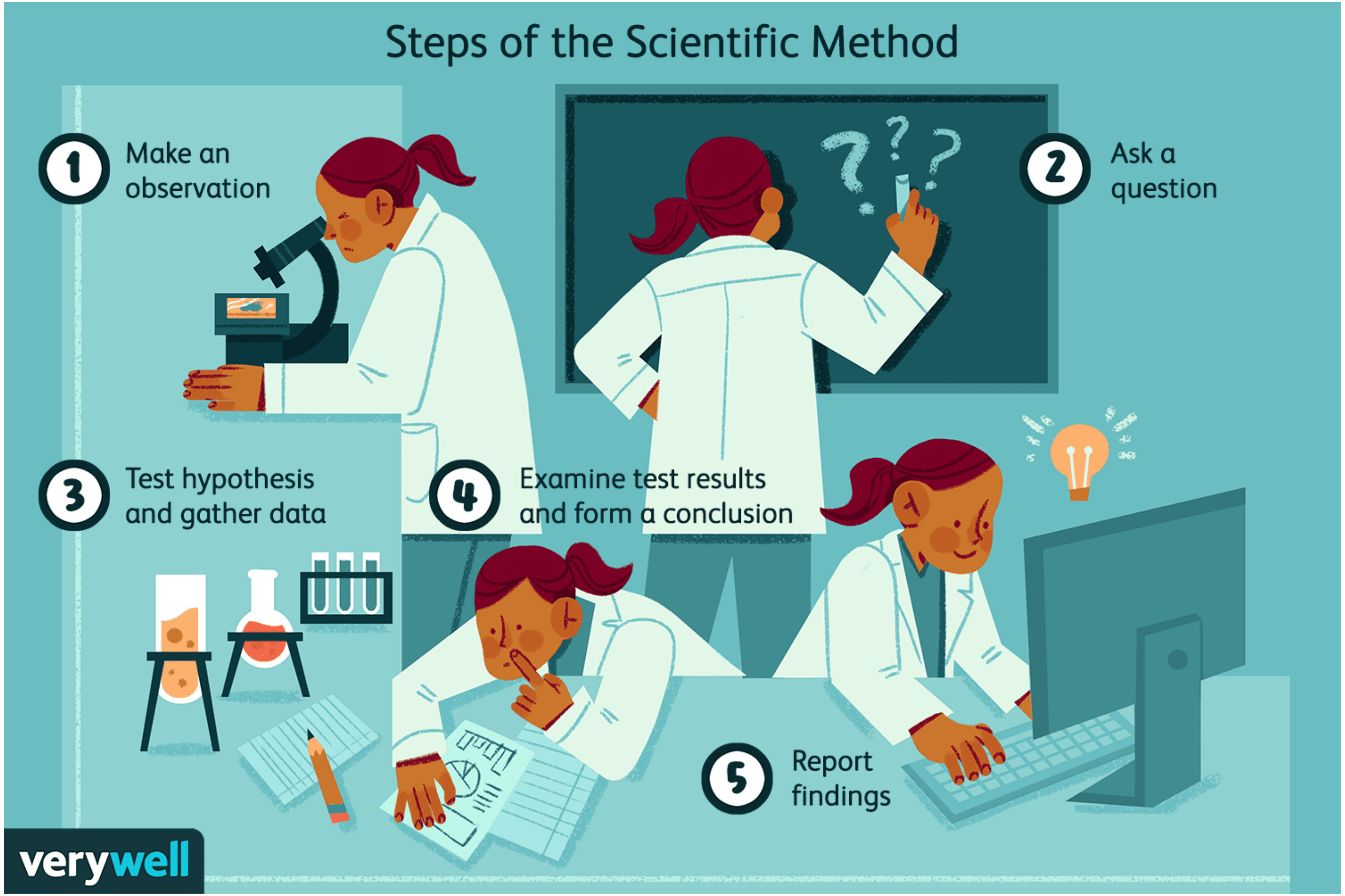
- Hypothesis Testing
- *t* Test
- Confounding Effects
- Chi-squared Test
- Multiple Comparisons

Assignment

- Writing assignment (due on 10/24)
- Reading assignment
- Office hours (please send emails to schedule place/time)

Data Science in Neuroimaging




Steps of the Scientific Method
1 Make an observation
2 Ask a question
3 Test hypothesis and gather data
4 Examine test results and form a conclusion
5 Report findings
verywell

History of Hypothesis Testing

- Early 20th century

Karl Pearson

William Sealy Gosset
(aka *Student*)

Ronald Fisher

Pearson:
- Chi-squared test
- Standard deviation
- Correlation

Student:
- t-distribution

Fisher:
- Null hypothesis
- Significant test
- Exact test
- ANOVA

Hypothesis Testing – A Formal Recipe

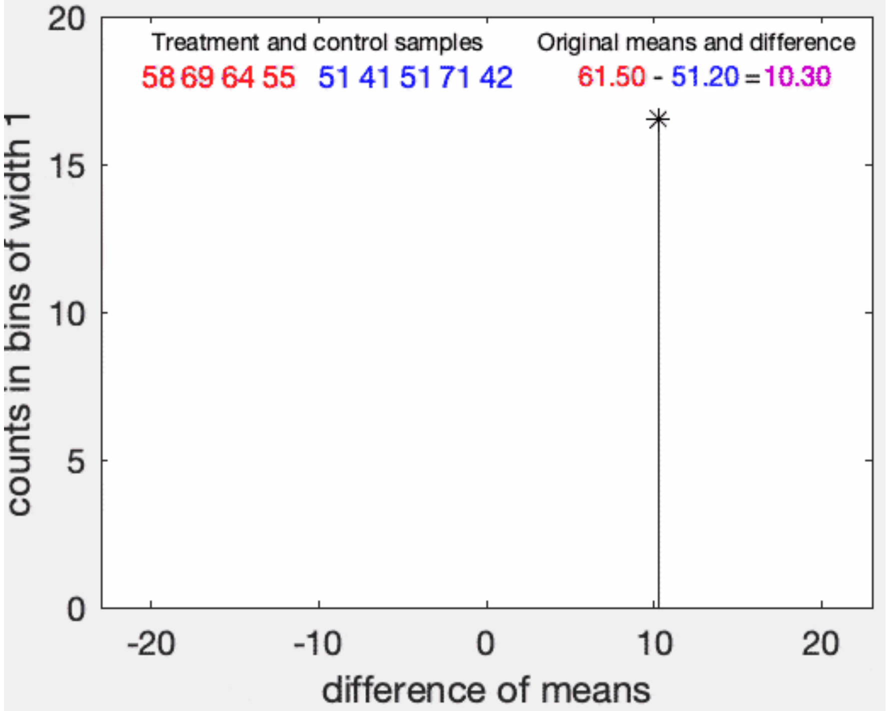
- A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis.
  – Choose a test statistic and set a null hypothesis
    - *The effect investigated by the analysis does not occur*
  – Derive the null distribution
    - *Distribution of possible results under the null hypothesis*
  – Select a significance level
  – Decide to either reject the null hypothesis or not to reject it
    - *Compare the actual observed result to the null distribution*

An "Engineering" Solution: Permutation Test

- Two groups of samples
  - 5 control samples: (51, 41, 51, 71, 42)
  - 4 treatment samples: (58, 69, 64, 55)
  - Question: is there a difference between groups?
    - (58+69+64+55)/4 – (51+41+51+71+42)/5 = 10.3 <--- Observed mean difference!
- If there is no group difference (null hypothesis), can I observe 10.3 just by chance?
- Random Permutation:
  - Control (51, 41, 55, 51, 64), Treatment (58, 42, 69, 71)
    - (58+42+69+71)/4 – (51+41+55+51+64)/5 = 7.6
  - Control (69, 64, 51, 71, 42), Treatment (58, 55, 51, 41)
    - (58+55+51+41)/4 – (69+64+51+71+42)/5 = -8.15
  - ……

Non-parametric test

- Permutation test
  - Build null distribution via permutation
    1. Randomly re-group the data into n and m samples
    2. Compute the mean difference from regrouped data
    3. Repeat K times
  - Compare actual mean difference with null distribution (p= the # permuted mean difference is larger than the original mean difference / K)
  - Reject null hypothesis: $p=0.001 < 0.05$
  - Cannot reject: $p=0.2 > 0.05$



Hypothesis Testing

- Assuming population follows a <u>parametric distribution</u>
- Testing hypotheses about <u>population parameters</u>

Yes

Parametric tests

t-test

F-test

Z-test

…

No

Nonparametric tests

Chi-squared test

U-test

Permutation test

…

Hypothesis Testing

- Assuming population follows a parametric distribution
- Testing hypotheses about population parameters

Yes

Parametric tests

t-test

F-test

Z-test

…

Stanford MEDICINE | Department of Psychiatry and Behavioral Sciences

Today…

- Hypothesis Testing
- *t* Test
- Confounding Effects
- Chi-squared Test
- Multiple Comparisons

Two-Sample t-test: Student's *t*-distribution

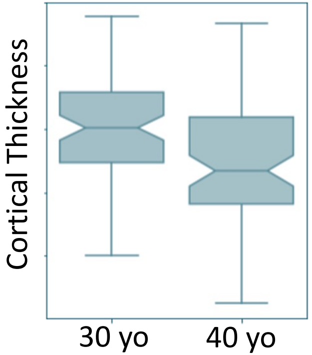
- *t*-distribution plays a critical role in estimating the mean of a normally distributed population in situations where the the population's standard deviation is unknown.
  - $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_m\}$ are independently and identically drawn from a **Gaussian distribution**
  - Null hypothesis: the mean difference $\bar{X} - \bar{Y} = 0$
  - Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$, $S^2 = \frac{(\mathrm{n}-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$ be the sample mean and the unbiased sample variance ($s_X^2 = \frac{\sum_i (X_i - \bar{X})^2}{n-1}$, $s_Y^2 = \frac{\sum_i (Y_i - \bar{Y})^2}{m-1}$)
  - Construct t-statistic $t = \frac{\bar{X} - \bar{Y}}{S * \sqrt{\frac{1}{n} + \frac{1}{m}}}$
  - t follows a Student's t-distribution with $n + m - 2$ degrees of freedom
  - Compute p–value based on observed *t*-value

Two-sample *t*-test

Test whether cortical thickness grows with aging?

Group 1	Age 30
Participant 1	3
Participant 2	2.7
…	…
Participant n	2.5

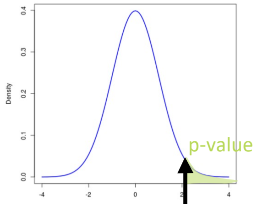
Group 2	Age 40
Participant 1	3
Participant 2	2.7
Participant 3	2.2
…	…
…	…
Participant m	2.5



Null hypothesis: $\overline{X_1} - \overline{X_2} = 0$

$$t = \frac{\overline{X_1} - \overline{X_2}}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim$$

Null distribution of DOF n+m-2



Observed $\overline{X_1}$ $\overline{X_2}$ S_p

Stanford MEDICINE | Department of Psychiatry and Behavioral Sciences

Example: One sample *t*-test

Test whether cortical thickness grows with aging?

	Age 30
Participant 1	3
Participant 2	2.7
Participant 3	2.2
…	…
Participant n	2.5

Example: One sample *t*-test

Test whether cortical thickness grows with aging?

	Age 30	Age 40
Participant 1	3	2.6
Participant 2	2.7	2.5
Participant 3	2.2	2.3
…	…	…
Participant n	2.5	2.4

Growth *X*
-0.4
-0.2
0.1
…
-0.1

Null hypothesis: the growth (temporal change) is 0

Null hypothesis: $X \sim N(0, \sigma^2)$

One-Sample t-test: Student's *t*-distribution

- $X_1, \ldots, X_n$ are independently and identically drawn from a **Gaussian distribution**
- Null hypothesis: the mean of that Gaussian is μ
- Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ be the sample mean and the unbiased sample variance
- Construct t-statistic $t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$
- t follows a Student's t-distribution with $n-1$ degrees of freedom
- Compute p–value based on observed *t*-value

Example: One sample *t*-test

Test whether cortical thickness grows with aging?

	Age 30	Age 40
Participant 1	3	2.6
Participant 2	2.7	2.5
Participant 3	2.2	2.3
…	…	…
Participant n	2.5	2.4

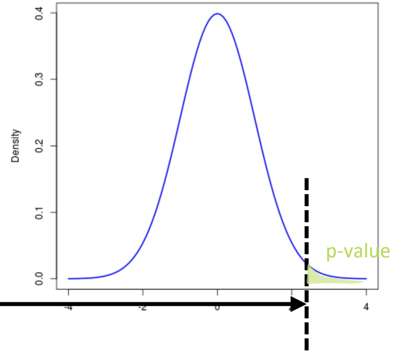
Growth *X*
-0.4
-0.2
0.1
…
-0.1

Null hypothesis: the growth (temporal change) is 0

Null hypothesis: $X \sim N(0, \sigma^2)$

$$t = \frac{\bar{X}-0}{S/\sqrt{n}} \sim$$

Null distribution of DOF n-1



Observed t−value $t = \frac{\bar{X}-0}{S/\sqrt{n}}$

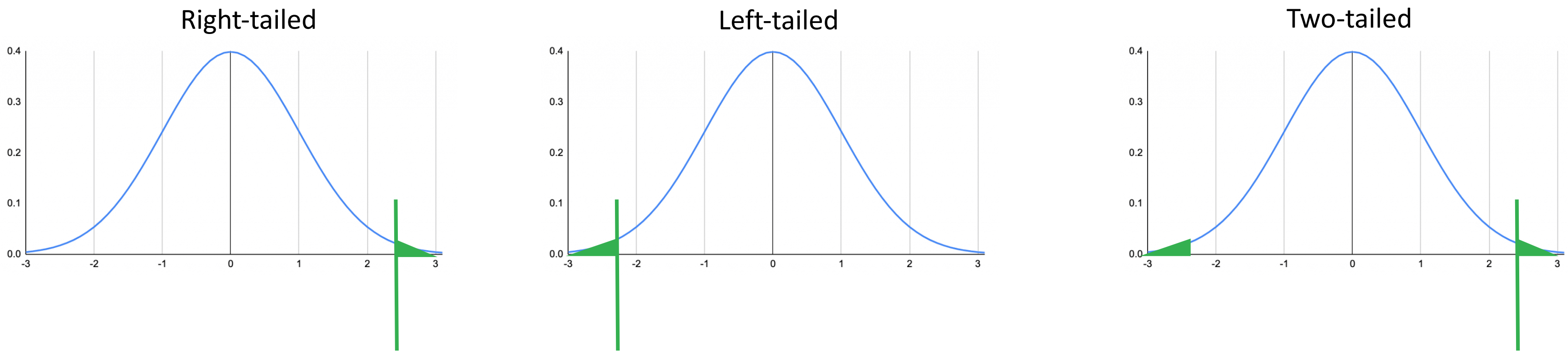
Pitfall of *p*-values

- Correct: probability of obtaining the observed results under the null hypothesis
- Incorrect: probability of the null hypothesis being wrong

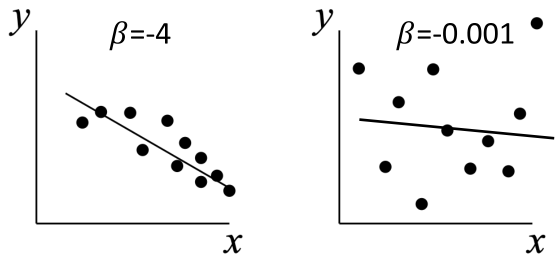
Stanford MEDICINE | Department of Psychiatry and Behavioral Sciences

What are paired vs unpaired *t*-tests?

- Reading assignment 1
  - Find out what paired and unpaired *t*-tests are, and how they are related to one-sample and two-sample *t*-tests.
  - What are one-tailed and two-tailed p-values (you have to use this concept in the assignment)?

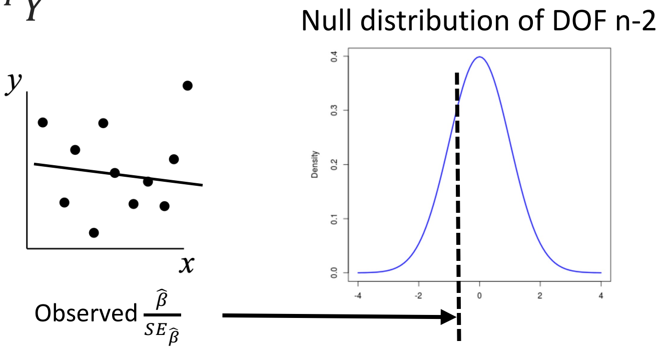


Slope of a regression line



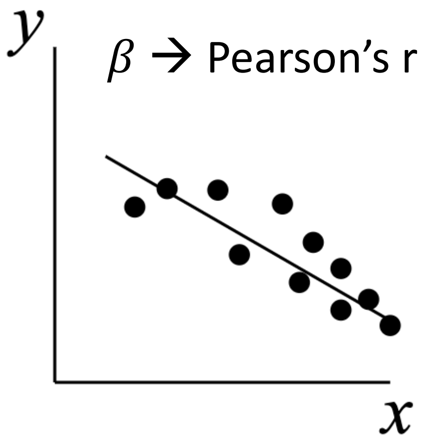
- Fitting the linear regression model $Y = \alpha + \beta X + \epsilon$
  Question: whether there is a linear relationship?

- Null hypothesis: $\beta = 0$ (there is no correlation between X, Y)
- Null distribution:
  - Let $\hat{\beta}$ be the least-squares estimator $\hat{\beta} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}X^{T}Y$
  - $\frac{\hat{\beta}}{SE_{\hat{\beta}}} \sim$ t-distribution with DOF n-2



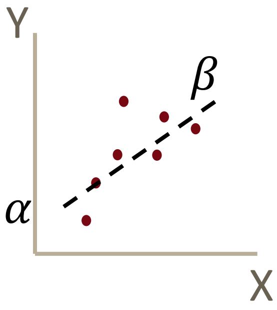
Pearson Correlation

- Very similar concept/computation as the slope test
- β of the standardized X, Y equals the correlation coefficient



All these tests are General Linear Models (GLM)!

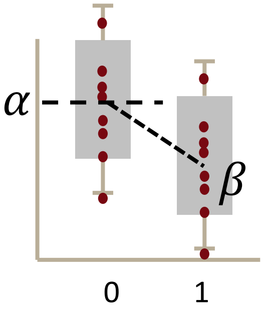
Slope



$$Y = \alpha + \beta X + \epsilon$$

Continuous

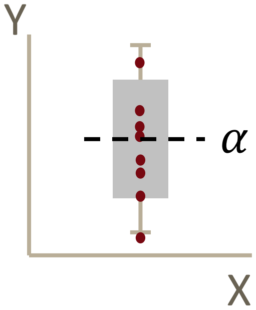
Two-sample



$$Y = \alpha + \beta X + \epsilon$$

{0,1}

One-sample



$$Y = \alpha + \beta 0 + \epsilon$$

Test this linear coefficient!

Multivariate Linear Regression $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \ldots + \epsilon$

age, sex, race, race

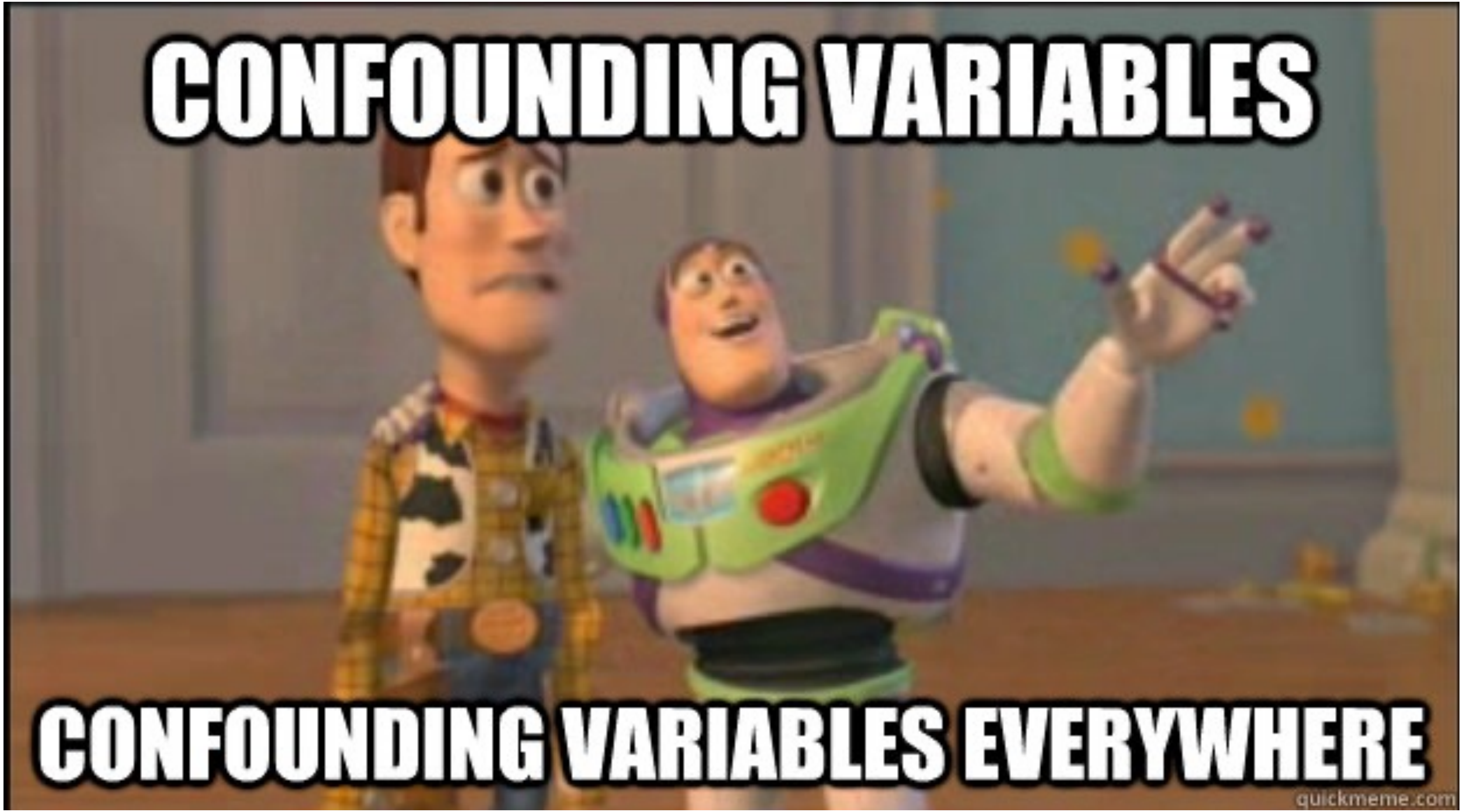
p_1 p_2 p_3 p_4

Other Popular Test Statistics

- Reading assignment 2
  - Find out what Z-test and F-test are and when those tests are preferred over *t*-tests

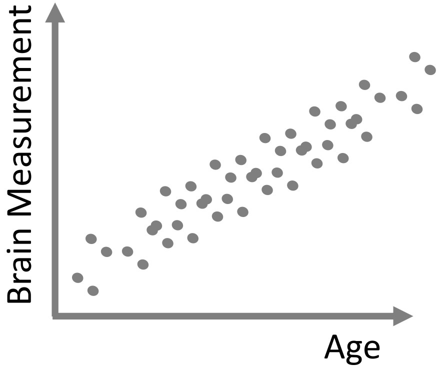
Today…

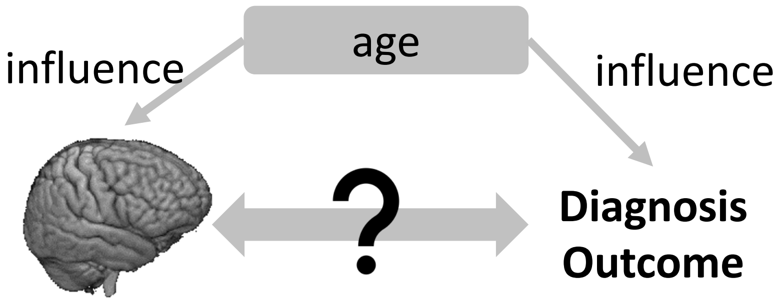
- Hypothesis Testing
- *t* Test
- **Confounding Effects**
- Chi-squared Test
- Multiple Comparisons


CONFOUNDING VARIABLES
CONFOUNDING VARIABLES EVERYWHERE
quickmeme.com

Confounding Variables

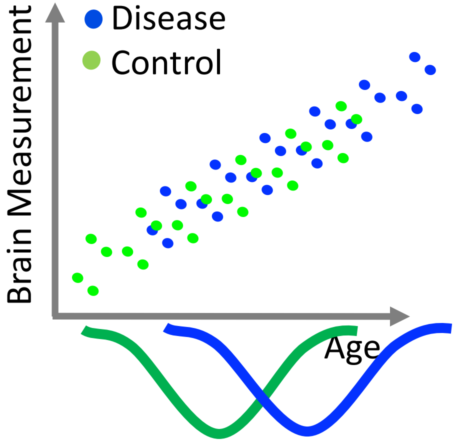
- A confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association.

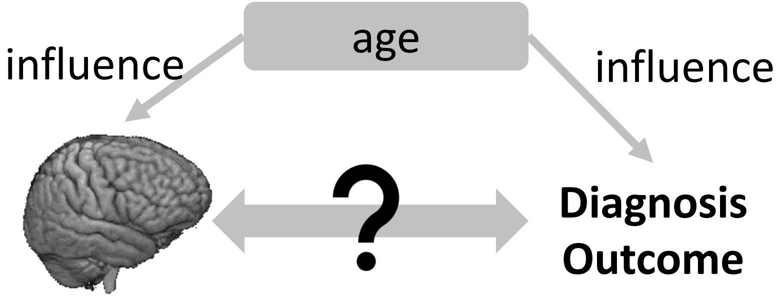




Scenario 1

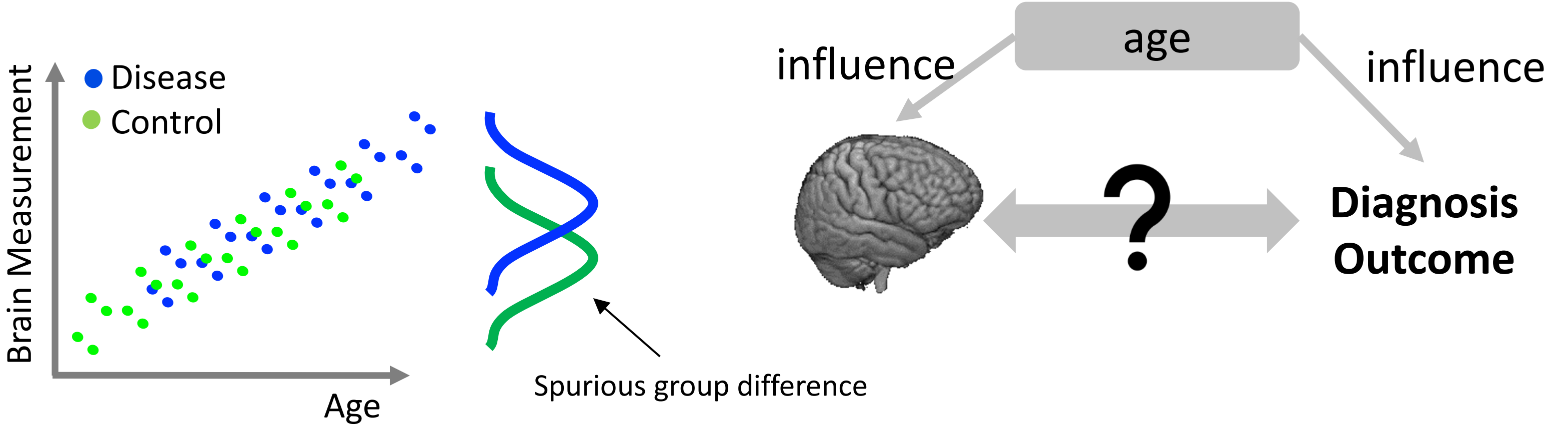
- A confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association.





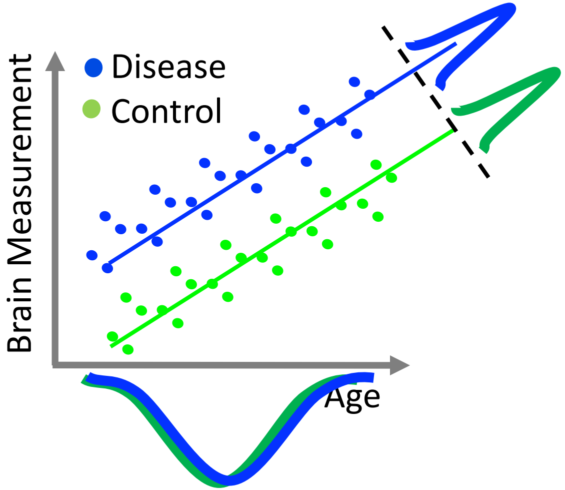
Scenario 1

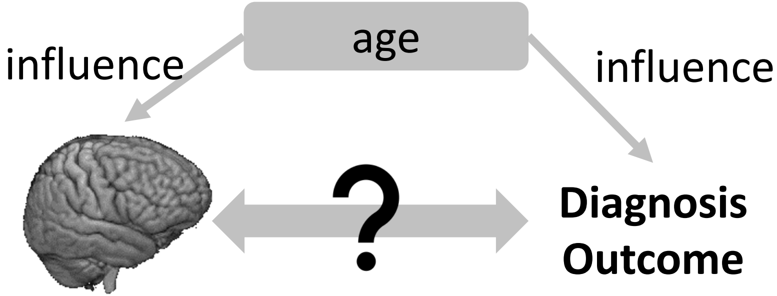
- A confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association.



Scenario 2

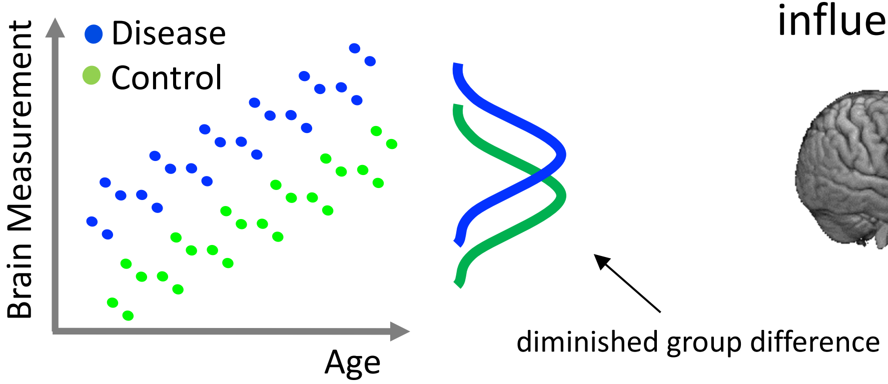
- A confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association.

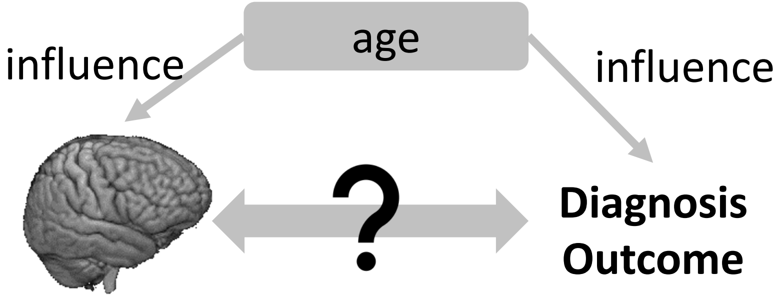




Scenario 2

- A confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association.

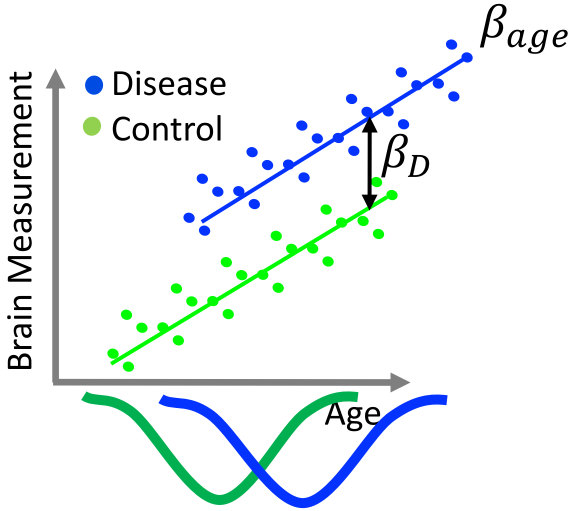




Confounder as a Covariate in GLM

- A covariate is an independent variable that can influence the outcome of a given statistical trial, but which is not of direct interest.

$$Y = \beta_0 + \beta_{age} age + \underset{\substack{\uparrow \\ t\text{-test on } \beta_D}}{\beta_D} \overset{\substack{\{0,1\} \\ \downarrow}}{D} + \epsilon$$

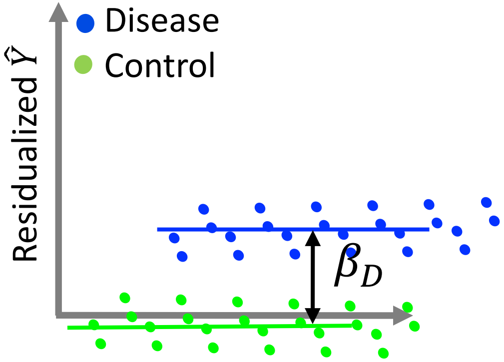


Compute Correct Group Differences by Residualization

- Compute disease effect by removing the aging effect.
  - Subtract $\beta_{age} age$ from raw measurements

$$Y = \beta_0 + \beta_{age} age + \beta_D D + \epsilon$$

$$\hat{Y} = Y - \beta_{age} age$$

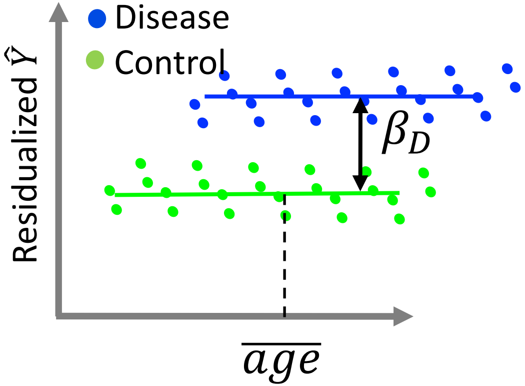


Compute Correct Group Differences by Residualization

- Compute disease effect by removing the aging effect.
  - Subtract $\beta_{age} age$ from raw measurements
  - Adjust 'reference point' to $\overline{age}$

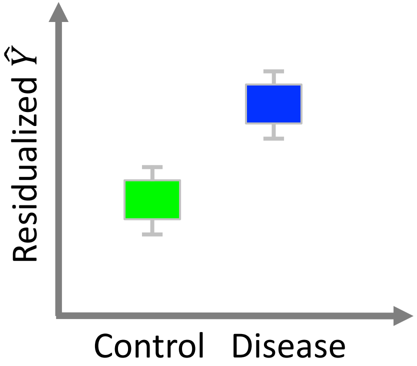
$$Y = \beta_0 + \beta_{age} age + \beta_D D + \epsilon$$

$$\hat{Y} = Y - \beta_{age} age + \beta_{age} \overline{age}$$



Compute Correct Group Differences by Residualization

- Compute disease effect by removing the aging effect.
  - Subtract $\beta_{age} age$ from raw measurements
  - Adjust 'reference point' to $\overline{age}$
  - Use boxplot to visualize $\hat{Y}$



Adeli et al., Chained Regularization for Identifying Brain Patterns Specific to HIV Infection, NeuroImage, 2018

Interaction in GLM

- An interaction effect occurs when the effect of one variable depends on the value of another variable

$$Y = \beta_0 + \beta_{age} age + \beta_D D + \epsilon$$

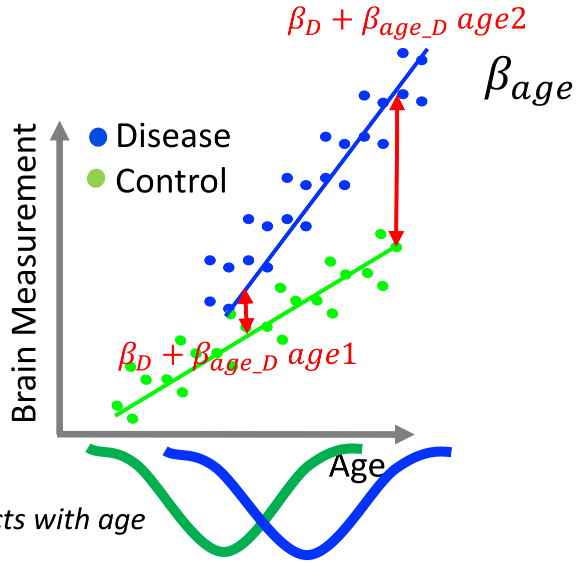
$$Y = \beta_0 + \beta_{age} age + (\beta_D + \beta_{age\_D}\ age) * D + \epsilon$$

The disease effect varies with age

$$Y = \beta_0 + \beta_{age} age + \beta_D D + \beta_{age\_D}\ age * D + \epsilon$$

Whether there is an overall disease effect

Whether there is the disease effect interacts with age



Today…

- Hypothesis Testing
- *t* Test
- Confounding Effects
- Chi-squared Test
- Multiple Comparisons

Hypothesis Testing

- Assuming population follows a parametric distribution
- Testing hypotheses about population parameters

What if the samples do not follow normal distribution?

Nonparametric tests

Chi-squared test

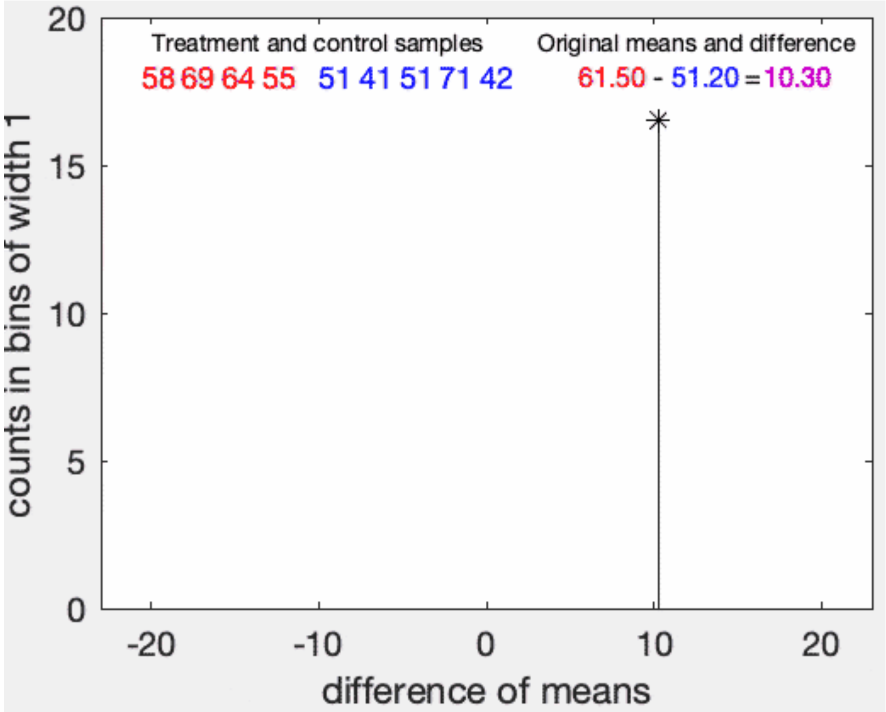
U-test

Permutation test

…

Non-parametric test

- Permutation test
  - Build null distribution via permutation
    1. Randomly re-group the data into n and m samples
    2. Compute the mean difference from regrouped data
    3. Repeat K times
  - Compare actual mean difference with null distribution (p= the # permuted mean difference is larger than the original mean difference / K)
- Advantage
  - Application to any test statistic
  - No assumptions on data distribution
- Disadvantage
  - Time consuming
  - Minimum p = 1/K



Non-parametric test

- Reading assignment 3
  - Find out what Mann–Whitney U test and Spearman's Correlation are.

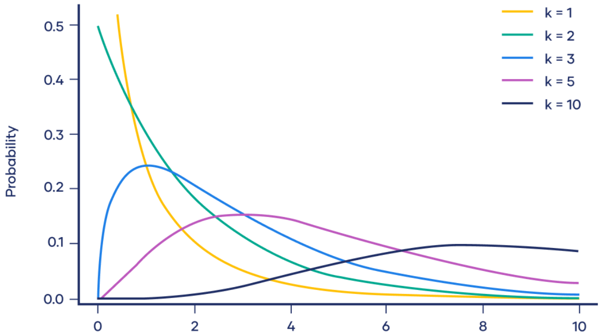
Chi-squared Test

- Example: Test a whether a coin is fair

	Head	Tail	Total
Observed	65	35	100
Expected	50	50	

Contingency table

- Chi-squared test is used to determine whether the expected frequencies in a contingency table differ from observed frequencies.
  - N samples classified into K classes with the ith class having X_i samples (observed frequency)
  - Null hypothesis: observed frequency follows expected frequency $\{M_i\}$
  - Null Distribution:
    - $X^2 = \sum_{i=1}^{K} \frac{(X_i - M_i)^2}{M_i}$
    - Follows a Chi-squared distribution of DOF k
  - Compute p-value from observed X^2



Chi-squared Test

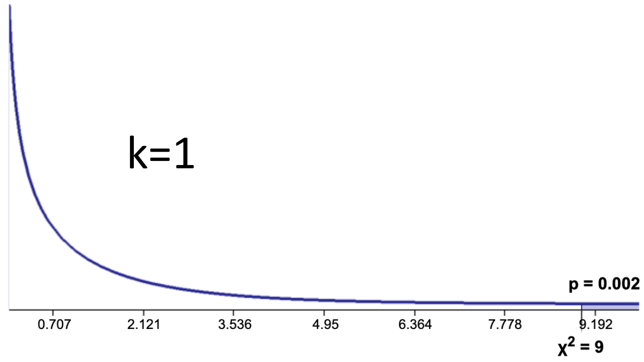
- Example: Test a whether a coin is fair

	Head	Tail
Expected	50	50

	Head	Tail	Total
Observed	65	35	$\boxed{100}$

Marginal frequency

- X^2 follows a Chi–squared distribution of DOF k
  - k = # of independent cells given marginal frequency
- Observed $X^2 = \frac{(65-50)^2}{50} + \frac{(35-50)^2}{50} = 9$



Stanford MEDICINE | Department of Psychiatry and Behavioral Sciences

Chi-squared Test: Another Example

- Frequencies of male and female being right-handed or left-handed
- Null Hypothesis: the frequency of handedness is the same across sexes
  - Translation: $\{X_{ij}\}$ follow expected frequency $\{M_{ij}\}$
  - How to compute $\{M_{ij}\}$?

Contingency Table

	Right-handed	Left-handed
Male	43	9
Female	44	4

Chi-squared Test: Another Example

- Null Hypothesis: the frequency of handedness is the same across sexes
  - Compute expected frequencies from marginal frequencies!
  - There are 100 samples, with 52% males and 87% overall right-handed rate.
  - Expected $m_{11} = 100 * \frac{52}{100} * \frac{87}{100} = 45.24$
  - DOF = # of independent cells given marginal frequency

Expected Frequency

	Right-handed	Left-handed	Total
Male	?		52
Female			48
Total	87	13	

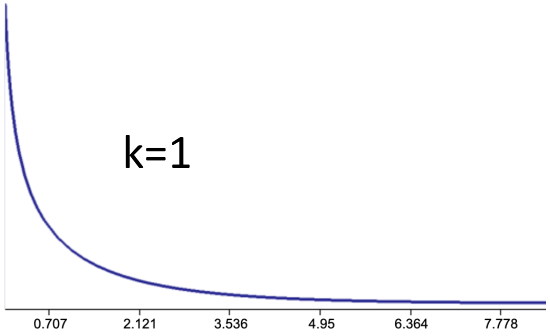
Contingency Table

	Right-handed	Left-handed	Total
Male	43	9	52
Female	44	4	48
Total	87	13	100

Marginal frequencies

Chi-squared Test: Another Example

- Null Hypothesis: the frequency of handedness is the same across sexes
  - Compute observed $X^2 = \sum\sum\frac{(X_{ij}-M_{ij})^2}{M_{ij}}$
  - DOF = # of independent cells given marginal frequency = (2-1) * (2-1) = 1



Contingency Table

	Right-handed	Left-handed	Total
Male	43	9	52
Female	44	4	48
Total	87	13	

Marginal frequencies

Other statistics related to Chi-squared tests

- Reading assignment 4
  - Find out what Binomial test and Fisher's Exact test are

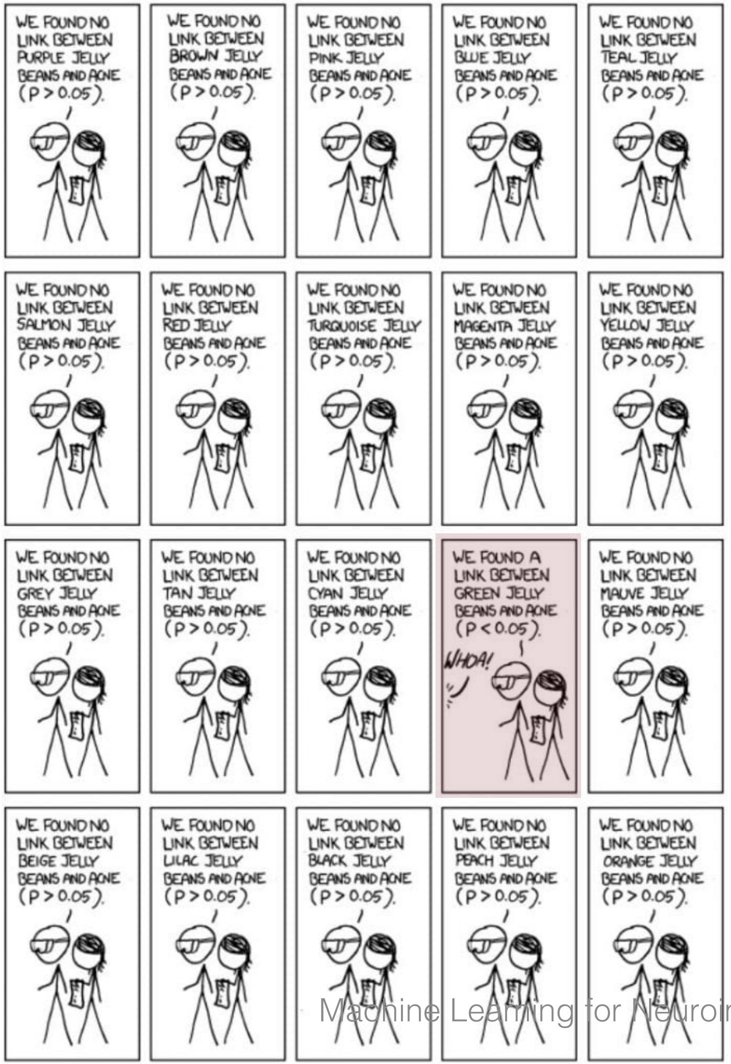
Today…

- Hypothesis Testing
- *t* Test
- Confounding Effects
- Chi-squared Test
- Multiple Comparisons

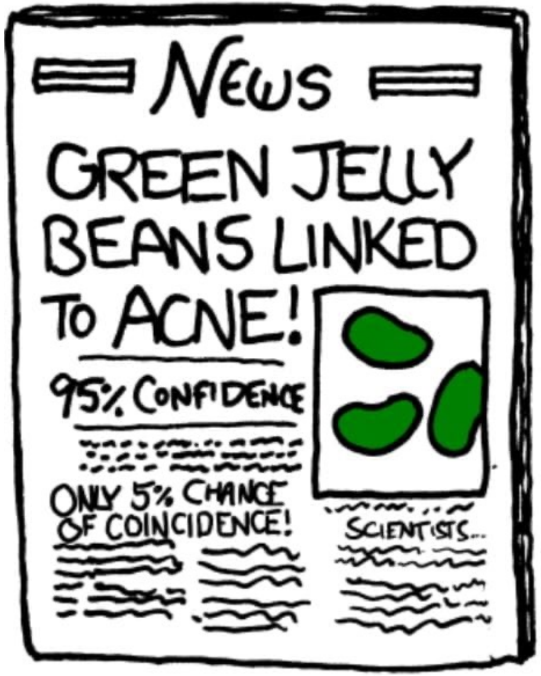
Pitfall of *p*-values

- *p*-hacking and reproducibility crisis

M. Baker, 1500 scientists lift the lid on reproducibility, Nature, 2016

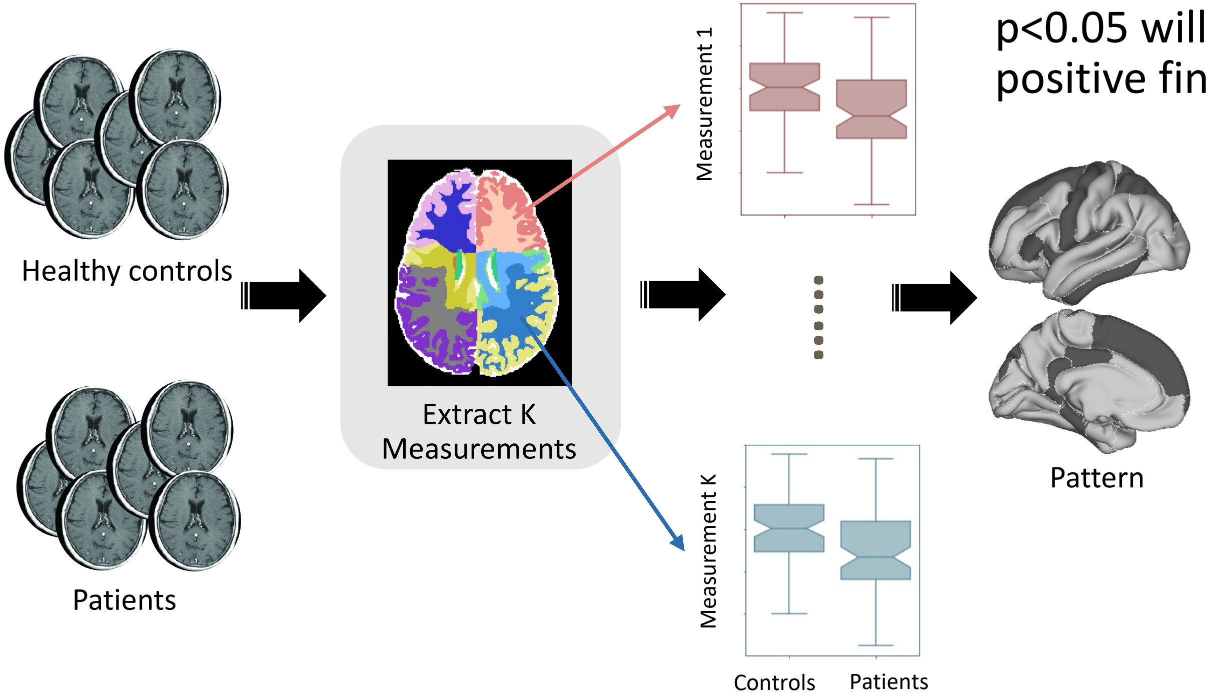


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).



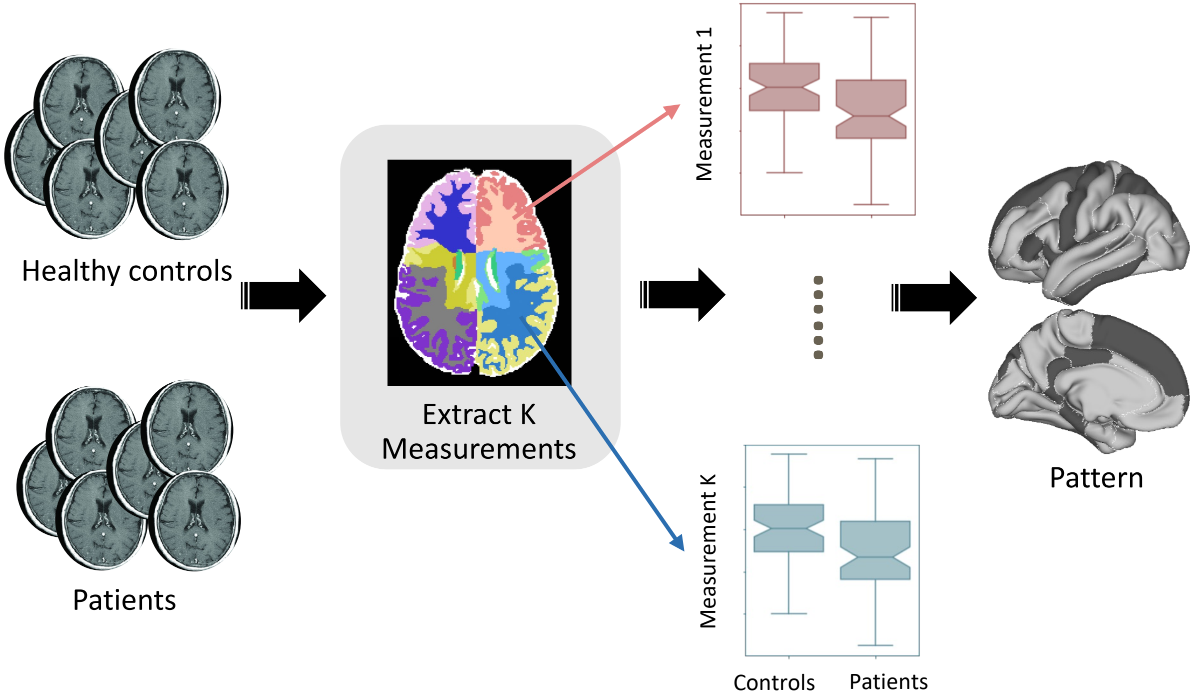
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Multiple Comparisons

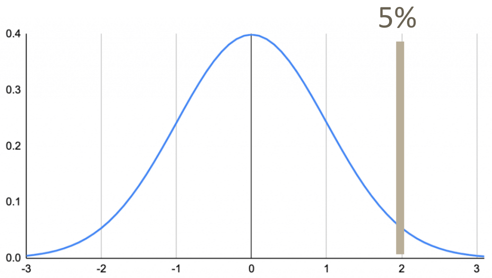


p<0.05 will introduce one false positive finding for every 20 regions!

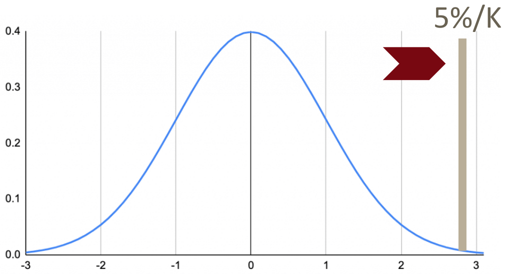
Multiple Comparisons



One **false positive** every 20 tests



Push to a more stringent threshold



Bonferroni Correction

- α of a single test: probability of falsely rejecting a true null hypothesis
- FWER (familywise error rate): probability of rejecting at least one true null hypothesis among a family of true hypotheses.
- Assume we have m null hypotheses to be tested, with m_0 null hypotheses are true (m_0 unknown)
  - Rejects the null hypothesis for each $p_{\mathrm{i}} \leq \frac{\alpha}{m}$

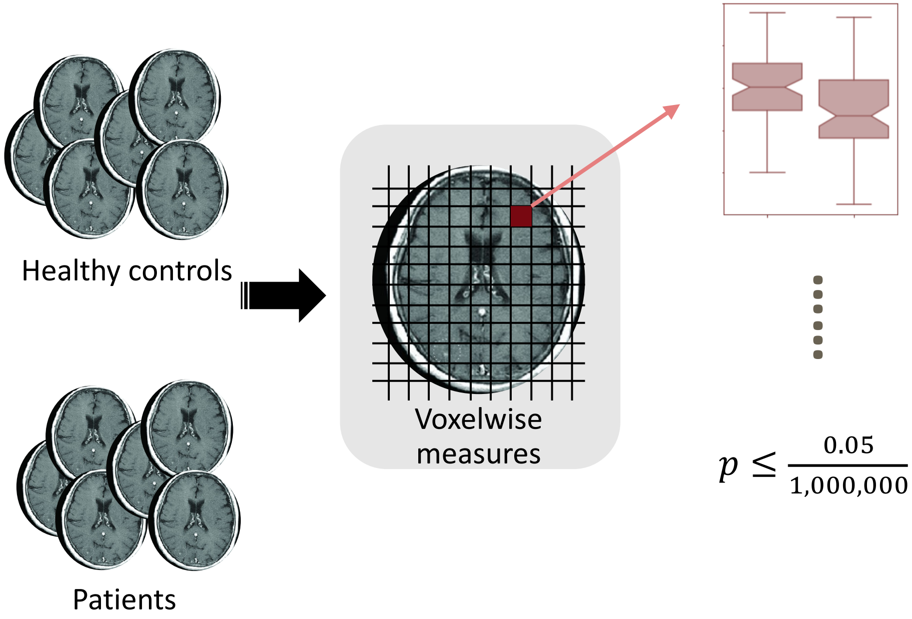
$$\mathrm{FWER} = P\left\{\bigcup_{i=1}^{m_0}\left(p_i \leq \frac{\alpha}{m}\right)\right\} \leq \sum_{i=1}^{m_0}\left\{P\left(p_i \leq \frac{\alpha}{m}\right)\right\} = m_0\frac{\alpha}{m} \leq \alpha.$$

  - Disadvantage: too conservative (high Type-II error rate) when m is large

Other Correction Procedures

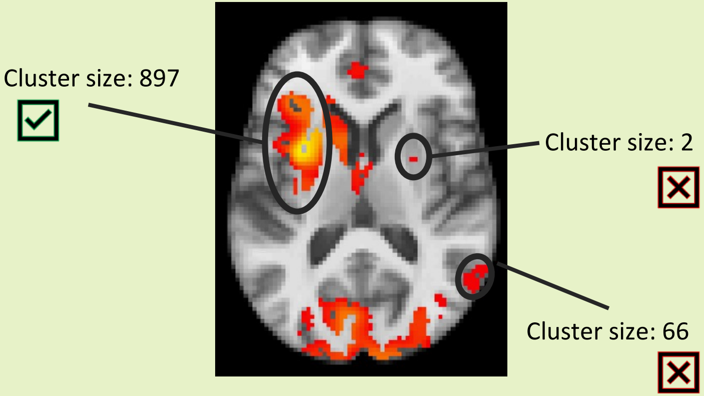
- Reading assignment 5
  – Find out what false discovery rate (FDR) and Benjamini–Hochberg procedure are

Voxel-wise Correction



$$p \leq \frac{0.05}{1{,}000{,}000}\ ?$$

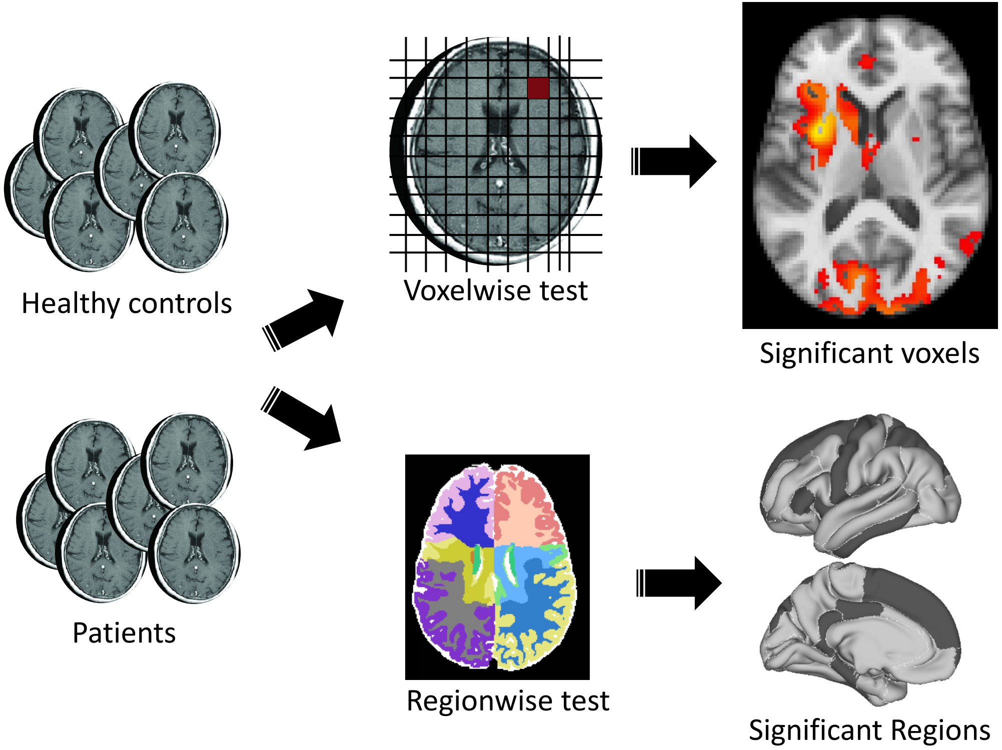
Voxels passing a certain significance level



Build a null distribution of cluster size via random field theory or permutation testing

Eklund et al., Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates, PNAS, 2016

Limitations of Hypothesis Testing



- Population-level conclusion
- Require a hypothesis beforehand
- Large # of “independent” tests

Assignment

- Due on 10/24
  - P1, P2: write calculation steps, equations, etc.
  - P3: source code with results, comments, visualization, description of what you did (easiest in Jupyter notebook)

Thank you!