

May 4, 2022

Austin G. Oswald

Comparing Categorical Variables (Chi-Square)



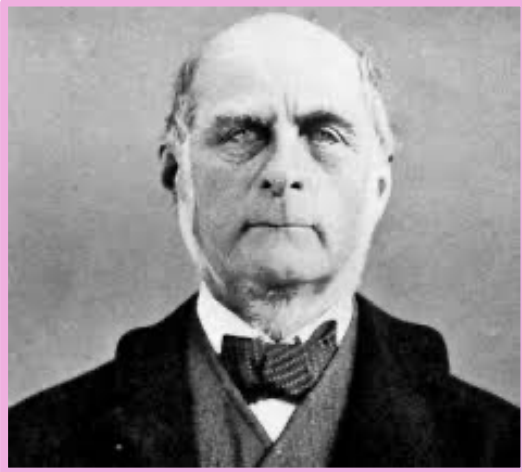
Agenda

- Check in
- History of statistics
- Last week's lab review
- Overview of chi-square
- Computer lab



[Image](#) by Artem Bryzgalov via Unsplash, shared under the Unsplash license.

Violent History of Statistical Reasoning



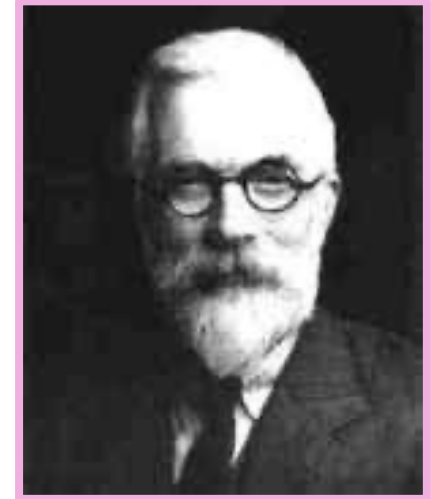
Francis Galton (1822-1911)

- Coined the term eugenics
- Developed regression and correlation techniques
- Motivations: racial improvement



Karl Pearson (1857-1936)

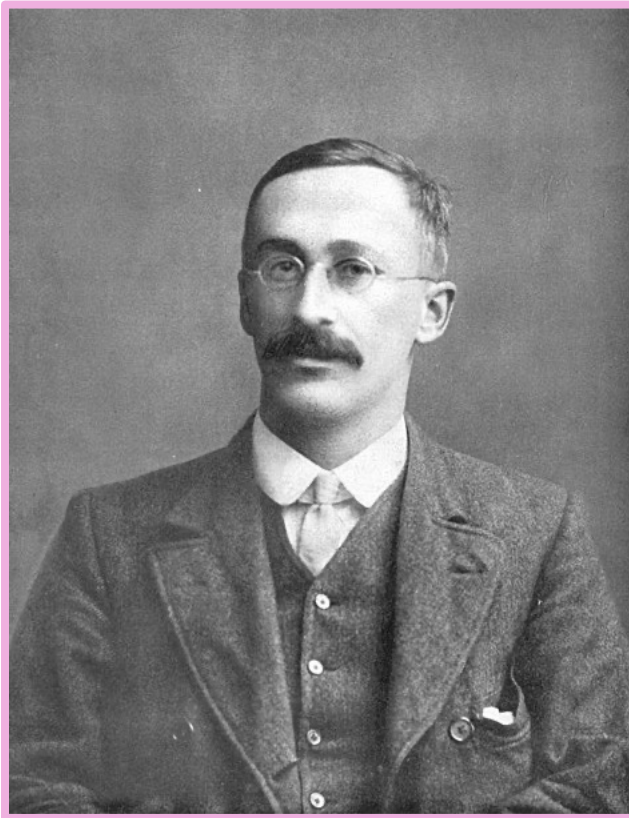
- Improved upon concepts of regression and correlation
- Developed Chi-Square test
- Motivations: elitism, racial improvement



RA Fisher (1890-1962)

- Later generation of eugenics
- Developed Fisher's exact test and Analysis of Variance (ANOVA)
- Motivation: insisting on racial difference

Another History?



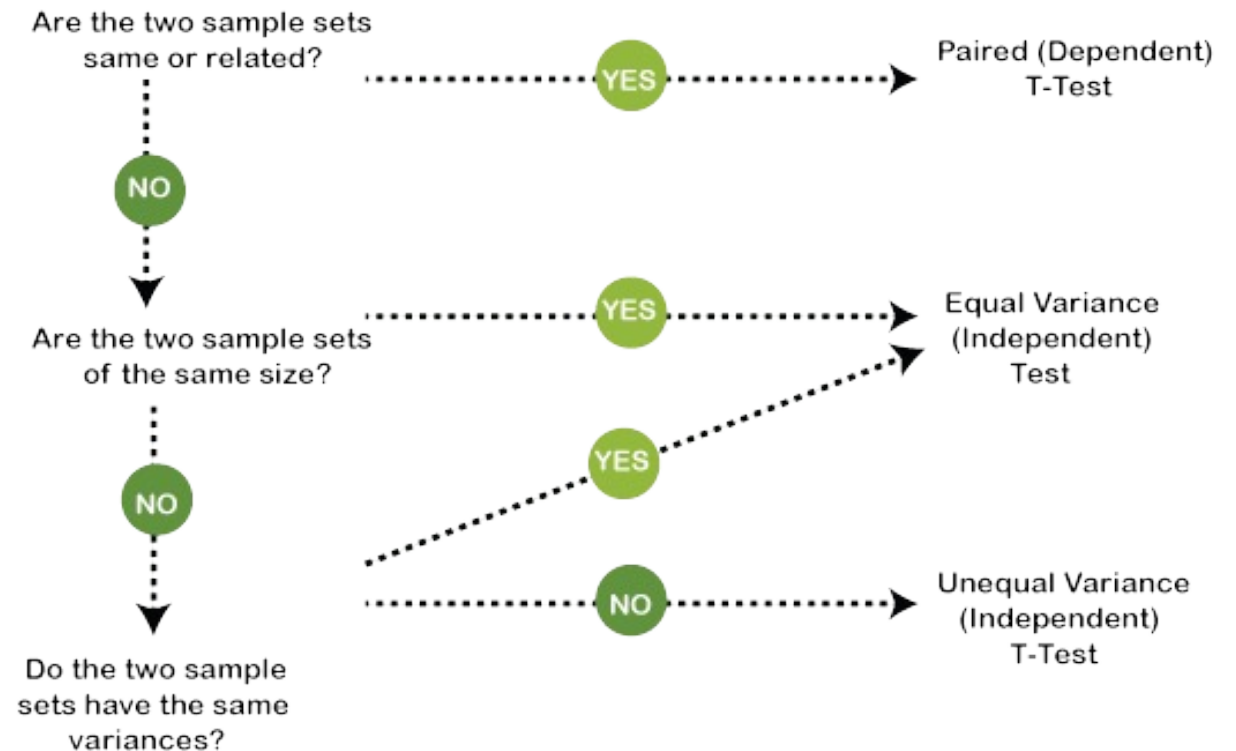
William Sealy Gosset (1876-1937)

- Chemist and Head Experimental Brewer of Guinness
- Published under the pseudonym "Student"
- Developed the t-test
- Motivation: Better beer



T-tests (re)Explained

- Comparing the mean (average) between two groups
- Works with continuous variables
 - Comparing two groups on mean scores on a scale; mean ages; mean hours spent in unpaid labor



Overview of p-values

A p value is used in hypothesis testing to help you support or reject the null hypothesis (no difference between the two groups). The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

There's a significant (not due to chance) difference between the groups if....

- $P < 0.05$ *
- $P < 0.01$ **
- $P < 0.001$ ***

**Chi-square (χ^2)
allows us to
determine whether
the differences in
proportions across
groups are
significantly
different.**



Pearson's Chi-Square Test

- Use to see whether there's a relationship between two categorical variables
 - Compares the frequencies/proportions you observe in certain categories to the frequencies you might expect to get in those categories by chance
- Testing whether the observed counts are significantly different from the expected counts OR if the proportions are significantly different between two groups

Potential Research Questions

- Suppose we had two groups of people: younger adults and older adults
- We also have a variable called depression diagnosis
 - Yes=1
 - No=0

Do the older adults have a significantly larger proportion of depression than the younger ones?

- Suppose we had two groups of people: cisgender and trans/gender expansive
- We also have a variable called poverty level
 - Living below the poverty threshold=0
 - Living at or above the poverty threshold=1

Is there a significantly higher proportion of poverty among trans and gender expansive people than cisgender people?

Assumptions of Pearson's Chi-Square

Fisher's exact test



Non-parametric test

Independence of data (no repeated measures)

No less than 5 in any cell (expected frequency) – chi-square distribution

Corrections for Pearson's Chi-Square

- Fisher's Exact
 - Use when a cell has less than five frequencies
- Yates' Correction
 - Use when working with a 2x2 table (two independent/dependent variables)

Class Activity

Is the subjective experience of being an MSW intern different for community organizing and clinical students regarding perceived level of support and overall experience?

Independent Variable

- Practice Method
 - Clinical=1
 - CO=2

Dependent Variable(s)

- PREP; JOB_OFF;
INT_JOB_
ALT_INT_

Working in RStudio



Studio[®]

[Check out results from t-tests from our class project last year](#)