

# STAT 113

## Tests for One and Two Multi-category Variables

Colin Reimer Dawson

Oberlin College

November 30, 2017

# Outline

Testing "Goodness of Fit" for a Categorical Distribution

Testing for Association of Categorical Variables

## Non-Binary Categorical Variables

- The distribution of a binary variable is defined entirely by the proportion of “successes”,  $p$ .
- We have randomization and CLT-based methods to do tests and construct intervals about a single proportion
- We can also compare two proportions *about the same thing* from two different groups.
- Categorical variables with more than two levels need more than one proportion to define their distribution.

## Birth Month and ADHD Diagnosis

A new study examines whether the youngest children in a school grade are more likely to be diagnosed with attention-deficit/hyperactivity disorder (ADHD) than their older peers in the same grade. The study involved almost a million children between the ages of 6 and 12 in British Columbia, Canada. The cutoff date for entering school in Canada is December 31, so those born in January are the oldest in any given class and those born in December are the youngest. The table below shows all ADHD diagnoses in the study, classified by the quarter of the year in which the child was born.

Birth Months	ADHD Diagnoses	Proportion of Births
Jan.-Mar.	6880	0.244
Apr.-Jun.	7982	0.258
Jul.-Sept.	9161	0.257
Oct.-Dec.	8945	0.241
Total	32698	1.000

## R Code for Goodness of Fit Test

Demo (Mostly Uses R as a Calculator)

## Handout