# Non-parametric tests

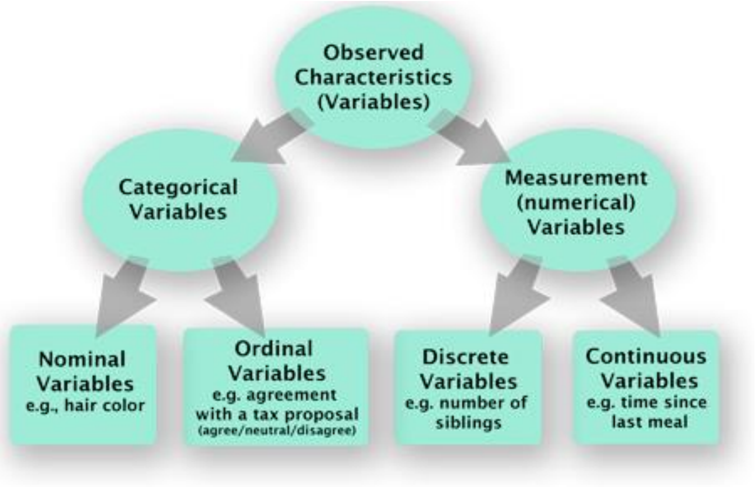
How to deal with Categorial data?

How to deal with cases where parametric assumptions are violated?

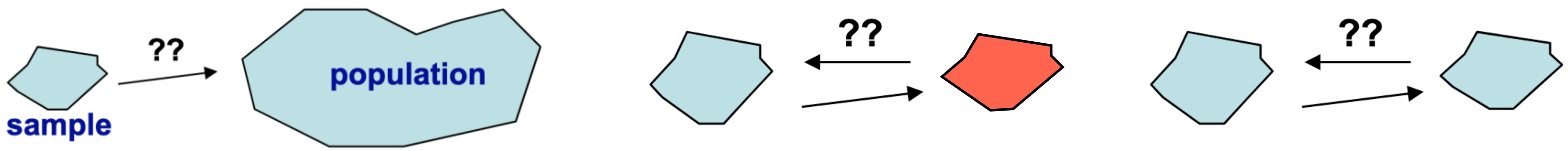
# Selecting a statistical test

- Different tests are used according to the level of measurement:
  - Interval
  - Ordinal
  - Categorical

- Parametric vs non-parametric (makes no assumption on the population distribution or sample size) assumptions

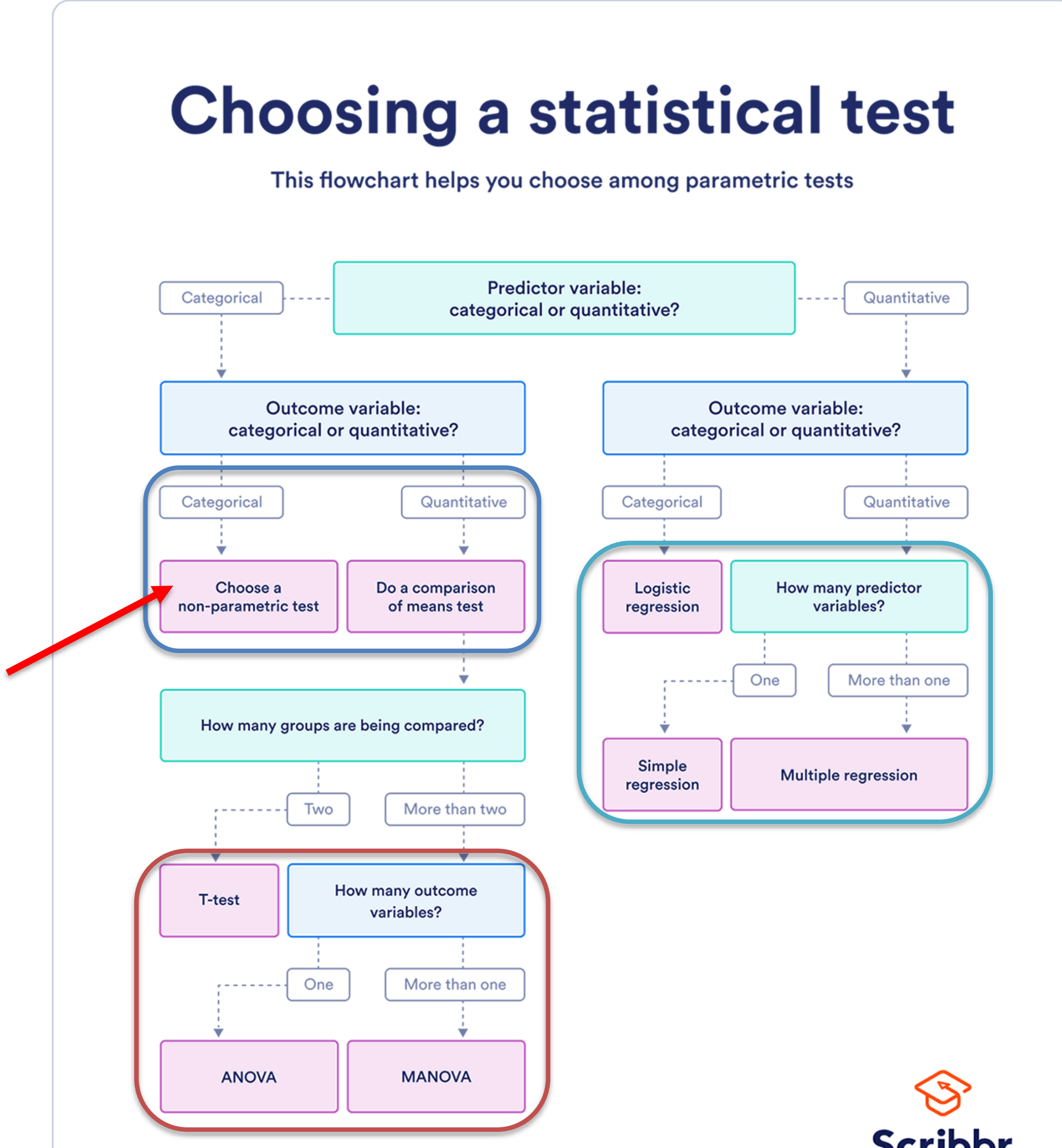


# Selecting a statistical test




- Different tests are used for varying amount of groups/conditions:
  - two samples
  - > two samples

- Different tests are used for related versus unrelated designs:
  - unrelated samples = between subjects designs
  - related samples = within subjects designs & matched pairs




https://www.scribbr.com/statistics/statistical-tests/

| | Predictor variable | Outcome variable | Research question example |
|---|---|---|---|
| **Simple linear regression** | • Continuous<br>• 1 predictor | • Continuous<br>• 1 outcome | What is the effect of income on longevity? |
| **Multiple linear regression** | • Continuous<br>• 2 or more predictors | • Continuous<br>• 1 outcome | What is the effect of income and minutes of exercise per day on longevity? |
| **Logistic regression** | • Continuous | • Binary | What is the effect of drug dosage on the survival of a test subject? |

Regression tests

Regression tests look for **cause-and-effect** relationships. They can be used to estimate the effect of one or more continuous variables on another variable.

https://www.scribbr.com/statistics/statistical-tests/

| | Predictor variable | Outcome variable | Research question example |
|---|---|---|---|
| **Paired t-test** | • Categorical<br>• 1 predictor | • Quantitative<br>• groups come from the same population | What is the effect of two different test prep programs on the average exam scores for students from the same class? |
| **Independent t-test** | • Categorical<br>• 1 predictor | • Quantitative<br>• groups come from different populations | What is the difference in average exam scores for students from two different schools? |
| **ANOVA** | • Categorical<br>• 1 or more predictor | • Quantitative<br>• 1 outcome | What is the difference in average pain levels among post-surgical patients given three different painkillers? |
| **MANOVA** | • Categorical<br>• 1 or more predictor | • Quantitative<br>• 2 or more outcome | What is the effect of flower species on petal length, petal width, and stem length? |

## Comparison tests

Comparison tests look for **differences among group means**. They can be used to test the effect of a categorical variable on the mean value of some other characteristic.

T-tests are used when comparing the means of precisely two groups (e.g., the average heights of men and women). ANOVA and MANOVA tests are used when comparing the means of more than two groups (e.g., the average heights of children, teenagers, and adults).

https://www.scribbr.com/statistics/statistical-tests/

## Correlation tests

Correlation tests **check whether variables are related** without hypothesizing a cause-and-effect relationship.

These can be used to test whether two variables you want to use in (for example) a multiple regression test are autocorrelated.

| | Variables | Research question example |
|---|---|---|
| **Pearson's *r*** | • 2 continuous variables | How are latitude and temperature related? |

# Selecting a statistical test

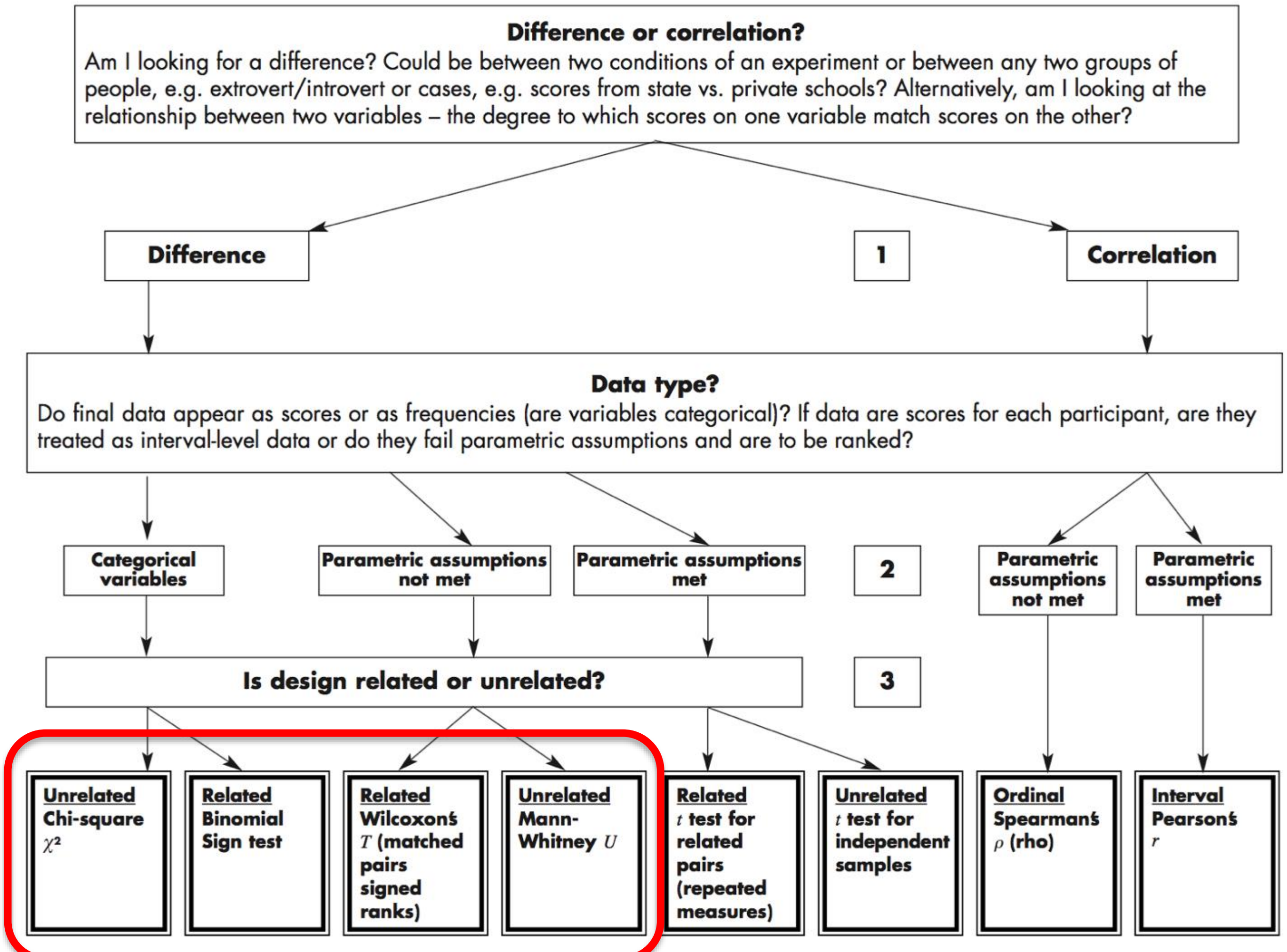



**Figure 23.1** Choosing an appropriate two-sample test.

| | Predictor variable | Outcome variable | Use in place of... |
|---|---|---|---|
| **Spearman's *r*** | • Quantitative | • Quantitative | Pearson's *r* |
| **Chi square test of independence** | • Categorical | • Categorical | Pearson's *r* |
| **Sign test** | • Categorical | • Quantitative | One-sample *t*-test |
| **Kruskal–Wallis *H*** | • Categorical<br>• 3 or more groups | • Quantitative | ANOVA |
| **ANOSIM** | • Categorical<br>• 3 or more groups | • Quantitative<br>• 2 or more outcome variables | MANOVA |
| **Wilcoxon Rank-Sum test** | • Categorical<br>• 2 groups | • Quantitative<br>• groups come from | Independent t-test |
| **Wilcoxon Signed-rank test** | • Categorical<br>• 2 groups | • Quantitative<br>• groups come from the same population | Paired t-test |

## Choosing a nonparametric test

Non-parametric tests don't make as many assumptions about the data, and are useful when one or more of the common statistical assumptions are violated. However, the inferences they make aren't as strong as with parametric tests.

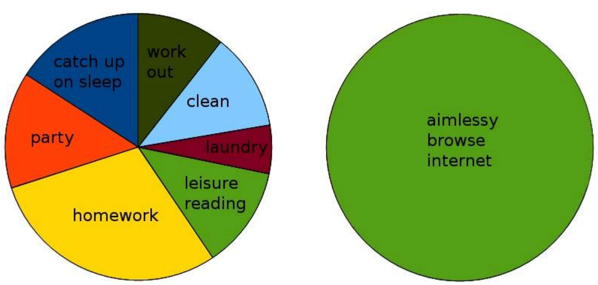
# Different types of tests

| Test type | Between subjects designs (Independent samples) | Within subject designs (repeated measures/matched pairs) |
|---|---|---|
| Non-parametric (for categorical data) | $\boxed{\textit{Chi-square test}}$ | *The binomial sign test* |
| Non-parametric (for ordinal data) | *Mann-Whitney U* | *Wilcoxon Signed-Rank Test*<br>*The binomial sign test* |
| Parametric | *Unrelated t-test (level of data: interval)* | *Related t-test (level of data: interval)* |

# Chi-Square Test

**Theoretical**
**categorical disribution**
**vs**
**Observed**
**categorical distribution**

Weekend in college




Expectation Reality

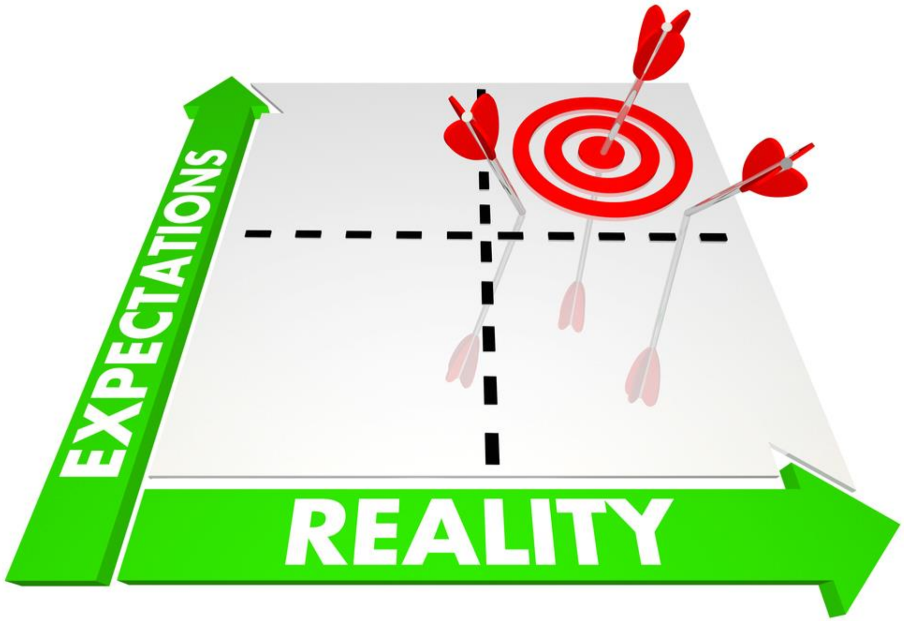
preference for one brand

# Chi-Square test

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

# $H_0$

# $H_A$




Data collected

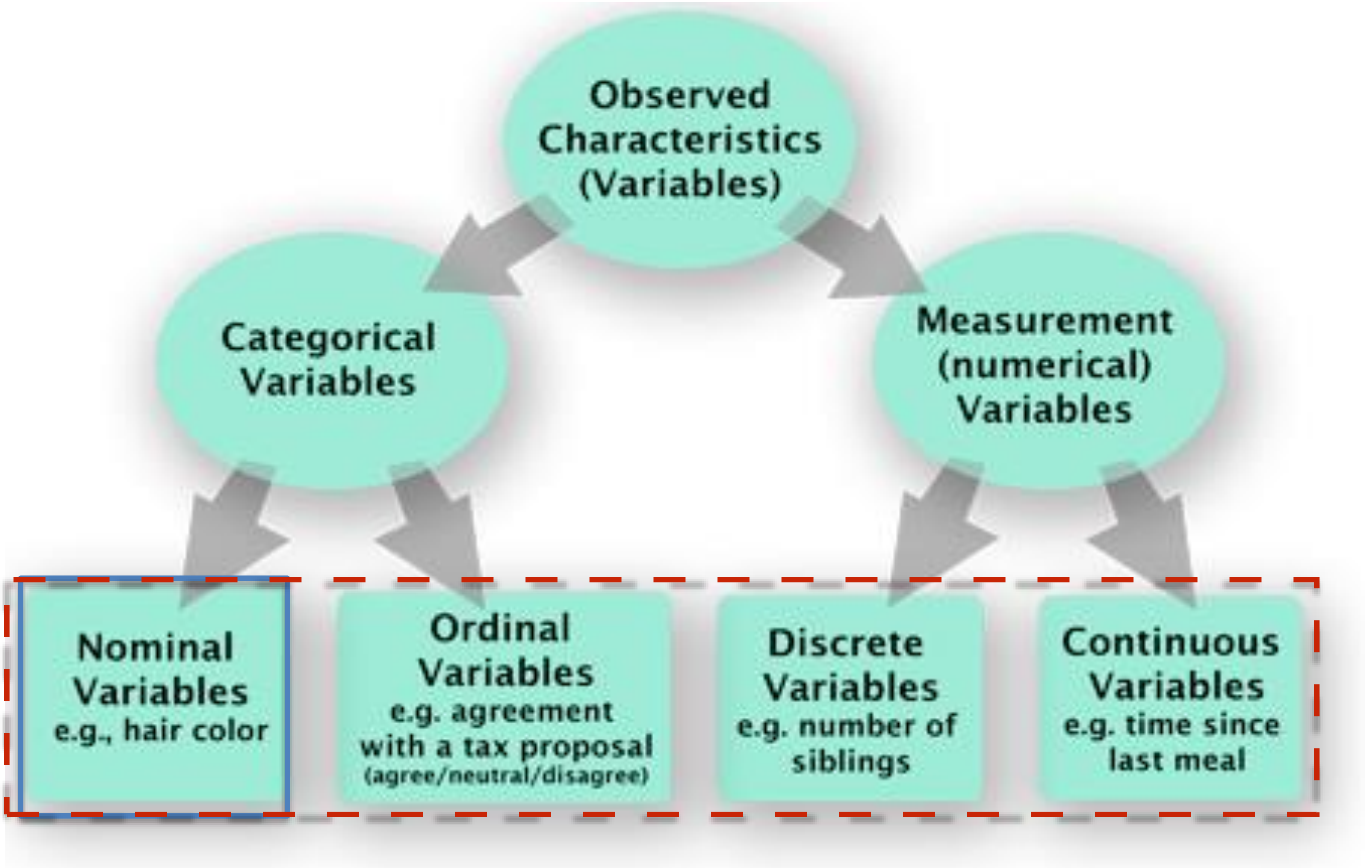
# Chi-Square Test - Applications

- Goodness-of-fit:
  - compare the observed sample distribution with the expected probability distribution
  - $H_0$ = no difference from a known population
- Chi-Square fit test:
  - determines how well theoretical distribution fits the empirical distribution
  - $H_0$ = no difference, equal proportions

# Chi-Square Test - Applications

- Test for Independence (for two variables):
  - test *relationship* between two separate variables
  - $H_0$ = there is no relationship between the variables
    - eg: females prefer pepsi more than males

# Chi-Square Test



# Applications of Chi-square test

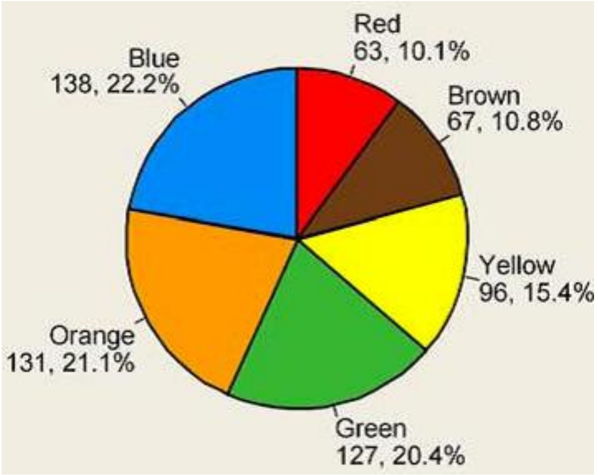
- Chi-square test is used to compare categorical variables. There are two type of chi-square test
  - 1. **Goodness of fit test** which determines if a sample matches the population.
  - 2. A chi-square fit **test for two independent variables** is used to compare two variables in a contingency table to check if the data fits.
  - The hypothesis being tested for chi-square is
    - Null: Variable A and Variable B are independent
    - Alternate: Variable A and Variable B are not independent.

- What to expect?
  - A small chi-square value means that data fits / variables independent
  - A high chi-square value means that data doesn’t fit / variables not independent


# Chi-Square test/goodness-of-fit


# $H_0$

M&Ms Color Distribution % according to their website



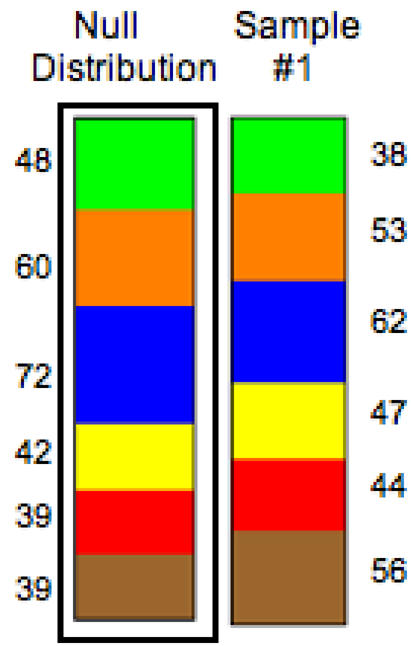

**$H_0$:** The color distribution is equal

**revised $H_0$:**

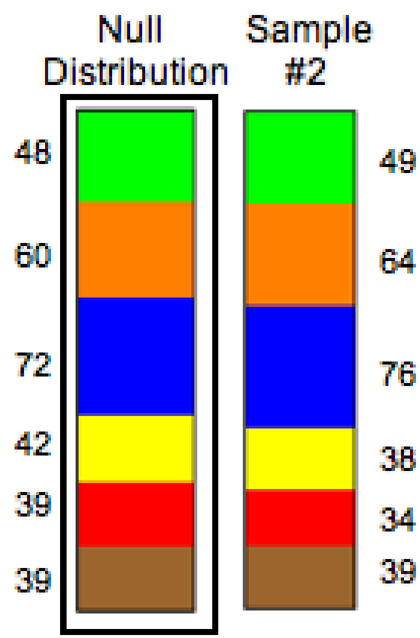
The color distribution is 13% brown, 13% red, 14% yellow, 24% blue, 20% orange, 16% green

**$H_A$:** The color distribution is different from 13% brown, 13% red, 14% yellow, 24% blue, 20% orange, 16% green


$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$




$\chi^2 = 12.94$



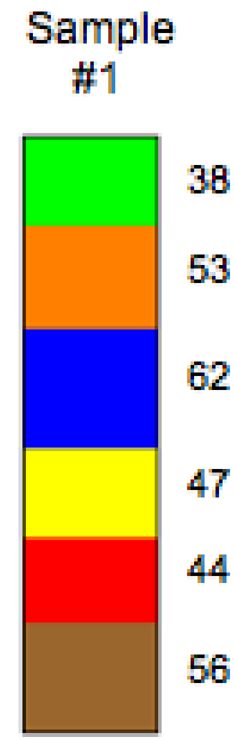

$\chi^2 = 1.53$

df = ?

$H_0$ ?

# Chi-Square distribution & df

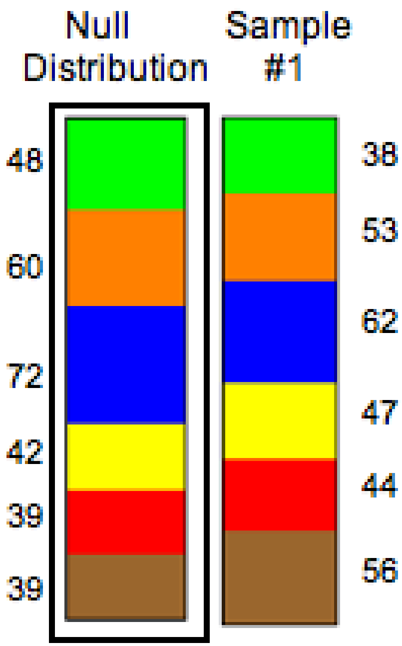
- Degrees of freedom for goodness-of-fit
  - number of cells you would need to calculate all other cell values, assuming we know marginal values
- df = *C*-1, *C* = no. of categories

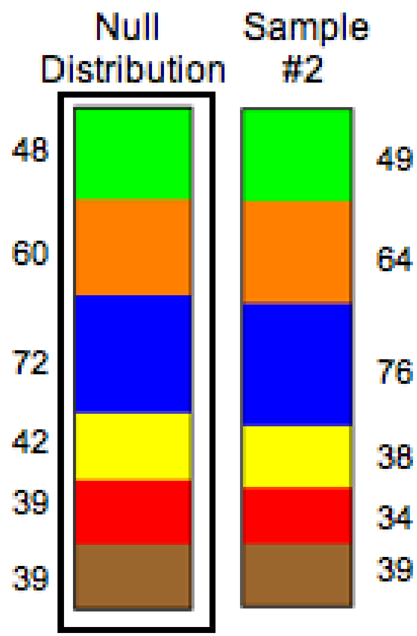



**df = ?**

https://www.khanacademy.org/math/statistics-probability/inference-categorical-data-chi-square-tests/chi-square-goodness-of-fit-tests/v/chi-square-distribution-introduction


$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$




**df = 5**

Critical values of the Chi-square distributio
with *d* degrees of freedom

| | Probability of exceeding the critical value | | | | | |
|---|---|---|---|---|---|---|
| *d* | 0.05 | 0.01 | 0.001 | *d* | 0.05 | 0.01 |
| 1 | 3.841 | 6.635 | 10.828 | 11 | 19.675 | 24.725 |
| 2 | 5.991 | 9.210 | 13.816 | 12 | 21.026 | 26.217 |
| 3 | 7.815 | 11.345 | 16.266 | 13 | 22.362 | 27.688 |
| 4 | 9.488 | 13.277 | 18.467 | 14 | 23.685 | 29.141 |
| 5 | 11.070 | 15.086 | 20.515 | 15 | 24.996 | 30.578 |
| 6 | 12.592 | 16.812 | 22.458 | 16 | 26.296 | 32.000 |
| 7 | 14.067 | 18.475 | 24.322 | 17 | 27.587 | 33.409 |
| 8 | 15.507 | 20.090 | 26.125 | 18 | 28.869 | 34.805 |
| 9 | 16.919 | 21.666 | 27.877 | 19 | 30.144 | 36.191 |
| 10 | 18.307 | 23.209 | 29.588 | 20 | 31.410 | 37.566 |

$\chi^2 = 12.94$

$\chi^2 = 1.53$


$H_0$

?

# Degrees of Freedom (*df*)

- number of independent pieces of information that go into the estimate of a parameter
- df depends on
  - particular calculation you will be performing
  - what you already know before making calculation

https://www.youtube.com/watch?v=rATNoxKg1yA

https://www.youtube.com/watch?v=rATNoxKg1yA&ab_channel=JamesGilbert

EXAMPLE

**H$_A$:** artists typically tend to be Aries or Cancer

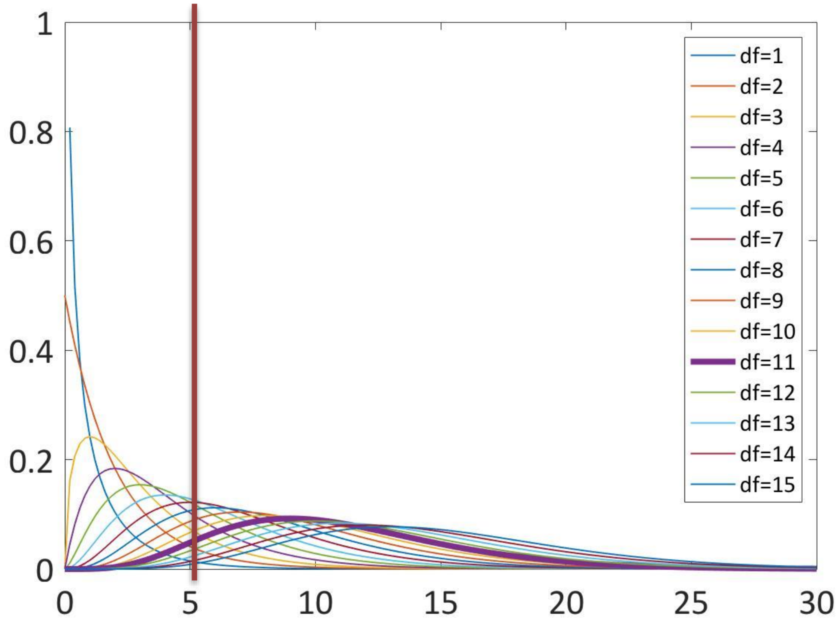
**H**

| Category | Observed | Expected |
|---|---|---|
| Aries | 29 | 21.333 |
| Taurus | 24 | 21.333 |
| Gemini | 22 | 21.333 |
| Cancer | 19 | 21.333 |
| Leo | 21 | 21.333 |
| Virgo | 18 | 21.333 |
| Libra | 19 | 21.333 |
| Scorpio | 20 | 21.333 |
| Sagittarius | 23 | 21.333 |
| Capricorn | 18 | 21.333 |
| Aquarius | 20 | 21.333 |
| Pisces | 23 | 21.333 |

**256 artists**

Please submit on Moodle [07-Mar-2025]:
Is the Null Hypothesis Accepted or Rejected based on Chi-Square Test

**df = ?**




**Chi-square Distribution Table**

| d.f. | .995 | .99 | .975 | .95 | .9 | .1 | .05 | .025 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 2.71 | 3.84 | 5.02 |
| 2 | 0.01 | 0.02 | 0.05 | 0.10 | 0.21 | 4.61 | 5.99 | 7.38 |
| 3 | 0.07 | 0.11 | 0.22 | 0.35 | 0.58 | 6.25 | 7.81 | 9.35 |
| 4 | 0.21 | 0.30 | 0.48 | 0.71 | 1.06 | 7.78 | 9.49 | 11.14 |
| 5 | 0.41 | 0.55 | 0.83 | 1.15 | 1.61 | 9.24 | 11.07 | 12.83 |
| 6 | 0.68 | 0.87 | 1.24 | 1.64 | 2.20 | 10.64 | 12.59 | 14.45 |
| 7 | 0.99 | 1.24 | 1.69 | 2.17 | 2.83 | 12.02 | 14.07 | 16.01 |
| 8 | 1.34 | 1.65 | 2.18 | 2.73 | 3.49 | 13.36 | 15.51 | 17.53 |
| 9 | 1.73 | 2.09 | 2.70 | 3.33 | 4.17 | 14.68 | 16.92 | 19.02 |
| 10 | 2.16 | 2.56 | 3.25 | 3.94 | 4.87 | 15.99 | 18.31 | 20.48 |
| 11 | 2.60 | 3.05 | 3.82 | 4.57 | 5.58 | 17.28 | 19.68 | 21.92 |
| 12 | 3.07 | 3.57 | 4.40 | 5.23 | 6.30 | 18.55 | 21.03 | 23.34 |
| 13 | 3.57 | 4.11 | 5.01 | 5.89 | 7.04 | 19.81 | 22.36 | 24.74 |
| 14 | 4.07 | 4.66 | 5.63 | 6.57 | 7.79 | 21.06 | 23.68 | 26.12 |


zodiac signs are evenly distributed across artists

$H_0$

# Chi-Square test for independence

test ***relationship*** between two separate variables

$H_{01}$ = there is no relationship between extroversion and comfort level of dancing in public


test ***difference*** between two conditions

$H_{02}$ = there is no difference in comfort level of dancing in public between introverts and extraverts

# Chi-Square (example)

$$\chi^2_c = \sum \frac{(O_i - E_i)^2}{E_i}$$

| | Extroverts | Introverts | TOTAL |
|---|---|---|---|
| **Not comfortable** | 10 | 40 | 50 |
| **comfortable** | 40 | 10 | 50 |
| **TOTAL** | 50 | 50 | 100 |

**observed** frequencies of Introverts and Extroverts who say they would or would not feel comfortable dancing in public

# Chi-Square (example)

| | Extroverts | Introverts | TOTAL |
|---|---|---|---|
| **Not comfortable** | 10 | 40 | 50 |
| **comfortable** | 40 | 10 | 50 |
| **TOTAL** | 50 | 50 | 100 |

**expected** frequencies if the null hypothesis were true?

# Chi-Square (example)

$$\frac{\text{row total} \times \text{column total}}{\text{total } n \text{ for table}}$$

| | Extroverts | Introverts | TOTAL |
|---|---|---|---|
| Not comfortable | **25**<br>10 | **25**<br>40 | 50 |
| comfortable | **25**<br>40 | **25**<br>10 | 50 |
| TOTAL | 50 | 50 | 100 |

**observed** and **expected** frequencies Introverts and Extroverts who say they would or would not feel comfortable dancing in public

# Chi-Square (example)

| | Extroverts | Introverts | TOTAL |
|---|---|---|---|
| Comfortable | 10 (25) | 40 (25) | 50 |
| Not Comfortable | 40 (25) | 10 (25) | 50 |
| TOTAL | 50 | 50 | 100 |

$$\chi^2 = \sum \frac{(\text{observed - expected})^2}{\text{expected}}$$

$$\frac{(40-25)^2}{25} + \frac{(10-25)^2}{25} + \frac{(10-25)^2}{25} + \frac{(40-25)^2}{25}$$

$$9 + 9 + 9 + 9$$

$$\chi^2 = 36$$

**df=?**

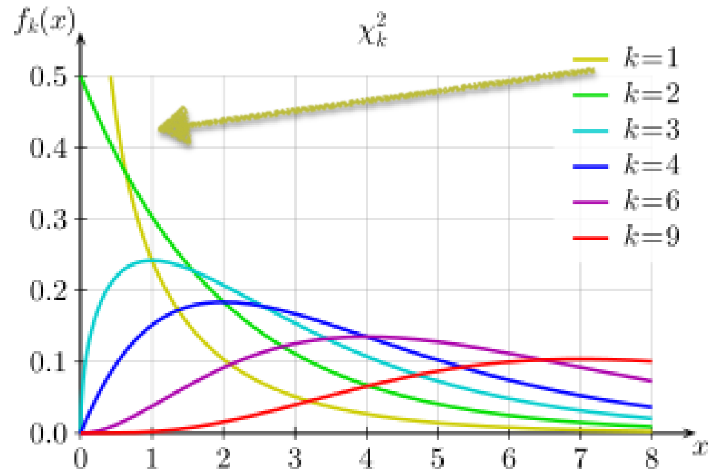
# Chi-Square (example)

- Degrees of freedom for independence
  - number of cells you would need to calculate all other cell values, assuming we know marginal values

$$df = (R-1)(C-1)$$

$$df = (2-1)(2-1) = 1$$

- Our chi-square is significant
  - Introverts tend to feel more comfortable dancing in public compared to Extroverts (surprise!)
  - Cooked-up data!!

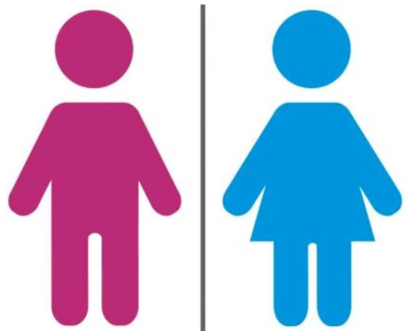



$$\chi^2 = 36$$

$H_0$

EXAMPLE

# Gender Study



MENTAL HEALTH FIRST AID®

**$H_0$:** There is no relationship between gender and willingness to use metal health services

**$H_{A1}$:** The distribution of reported willingness to use mental health services for males has proportions that are different from those in females

**$H_0$:** The distribution of reported willingness to use mental health services has the same proportions for males and females

**$H_A$:** The distribution of reported willingness to use mental health services for males has proportions that are different from those in females

# Contingency Table

$$\frac{\text{row total} \times \text{column total}}{\text{total } n \text{ for table}}$$

Willingness to Use Mental Health Services (n=150)

| | No | Maybe | Yes | **Total** |
|---|---|---|---|---|
| Males | 17 **12** | 32 **30** | 11 **18** | 60 |
| Female | 13 **18** | 43 **45** | 34 **27** | 90 |
| **Total** | 30 | 75 | 45 | 150 |

df = (R-1)(C-1) **df = 2**

Note: How did we estimate "expected" values?

30 said NO - so we would expect equal distribution of 15 and 15.

BUT only 40% are males so 40% of 30 makes it 12.

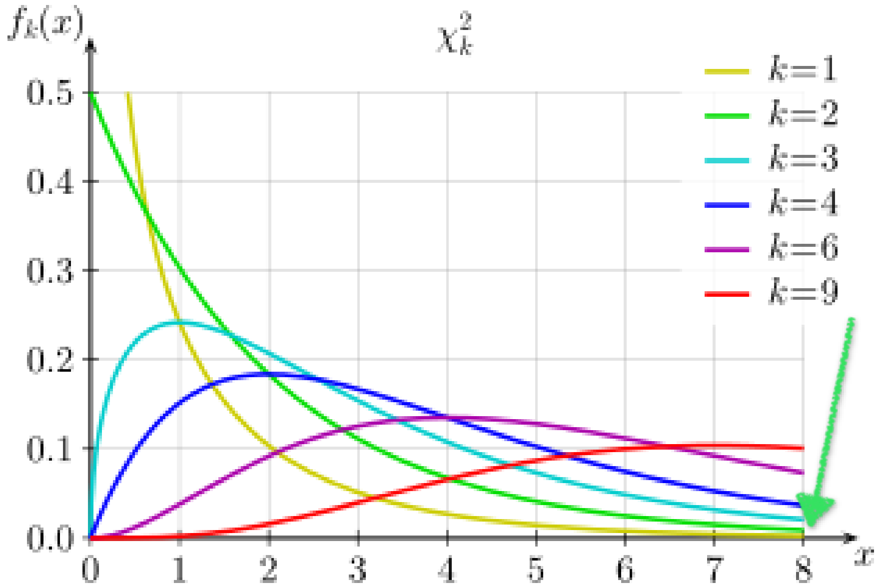
EXAMPLE

# Gender Study

Willingness to Use Mental Health Services

$\chi^2$ **= 8.23**

**df = 2**




$H_0$ REJECTED

Males are less willing to use Mental Health Services

# Effect Size

## Effect size in Chi square

- For a 2 x 2 table -> Phi Coefficient

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

Correlation between two categorical variables

Phi of 0.1 small, 0.3 medium, 0.5 large

- For larger tables -> Cramer's V coeffiecient (> 2 x 2)

$$V = \sqrt{\frac{\chi^2}{n \times df^*}}$$

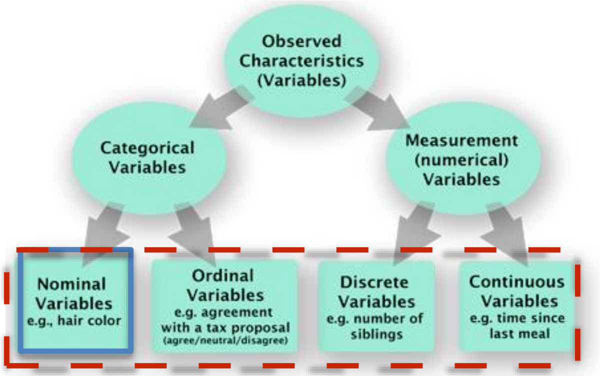
Df* is the smallest of C-1, R-1

Phi = sqrt(8.23/150)
= 0.23

Results showed a significant difference between males' and females' attitude toward using metal health services, $\chi^2$ (2, *n = 150*) = 8.23, *p* < .05, *V* = 0.23

# Chi-Square Test and Correlation

| Participant | Self-Esteem $X$ | Academic Performance $Y$ |
|---|---|---|
| A | 13 | 73 |
| B | 19 | 88 |
| C | 10 | 71 |
| D | 22 | 96 |
| E | 20 | 90 |
| F | 15 | 82 |
| . | . | . |
| . | . | . |
| . | . | . |




| | | Level of Self-Esteem | | | |
|---|---|---|---|---|---|
| | | High | Medium | Low | |
| Academic Performance | High | 17 | 32 | 11 | 60 |
| | Low | 13 | 43 | 34 | 90 |
| | | 30 | 75 | 45 | $n$ = 150 |

# Chi-square and independent measures *t* and ANOVA

| Participant | Self-Esteem $X$ | Academic Performance $Y$ |
|---|---|---|
| A | 13 | 73 |
| B | 19 | 88 |
| C | 10 | 71 |
| D | 22 | 96 |
| E | 20 | 90 |
| F | 15 | 82 |
| . | . | . |
| . | . | . |
| . | . | . |

| | | Level of Self-Esteem | | | |
|---|---|---|---|---|---|
| | | High | Medium | Low | |
| Academic Performance | High | 17 | 32 | 11 | 60 |
| | Low | 13 | 43 | 34 | 90 |
| | | 30 | 75 | 45 | $n = 150$ |

# Median Test for Independent Samples

- non-parametric alternative to independent measures *t*-test (or ANOVA) to determine significant group differences
- $H_0$ = different samples come from population that share a common median

Self-Esteem Scores for Children at Three Levels of Academic Performance

| High | | Medium | | | | Low | |
|---|---|---|---|---|---|---|---|
| 22 | 14 | 22 | 13 | 24 | 20 | 11 | 19 |
| 19 | 18 | 18 | 22 | 10 | 16 | 13 | 15 |
| 12 | 21 | 19 | 15 | 14 | 19 | 20 | 16 |
| 20 | 18 | 11 | 18 | 11 | 10 | 10 | 18 |
| 23 | 20 | 12 | 19 | 15 | 12 | 15 | 11 |

# Median Test for Independent Samples

- calculate median for combined group (n = 40)
- within each group, perform median (17) split and fill contingency table

Self-Esteem Scores for Children
at Three Levels of Academic Performance

| High | | Medium | | | | Low | |
|---|---|---|---|---|---|---|---|
| 22 | 14 | 22 | 13 | 24 | 20 | 11 | 19 |
| 19 | 18 | 18 | 22 | 10 | 16 | 13 | 15 |
| 12 | 21 | 19 | 15 | 14 | 19 | 20 | 16 |
| 20 | 18 | 11 | 18 | 11 | 10 | 10 | 18 |
| 23 | 20 | 12 | 19 | 15 | 12 | 15 | 11 |

| | Academic Performance | | |
|---|---|---|---|
| | High | Medium | Low |
| Above Median | 8 | 9 | 3 |
| Below Median | 2 | 11 | 7 |

# Median Test for Independent Samples

| | Academic Performance | | |
|---|---|---|---|
| | High | Medium | Low |
| Above Median | 8 **5** | 9 **10** | 3 **5** |
| Below Median | 2 **5** | 11 **10** | 7 **5** |

$\chi^2 = 5.4$ **df = 2** $\chi^2 = 5.99$ **(p < .05)**

—> not sufficient evidence to conclude that there are significant differences among the self-esteem for these three groups of students

# Chi-Square test

**Limitations**

- Observations must be unique to one cell (Between subjects)
  - each person must fall into only one cell
  - not valid for within subject designs (repeated measures/matched pairs)
- Only frequencies can be studied, not means, percentages, ratios, etc.
- Low **expected** frequencies cause problems (should be $\geq 5$)
  - loss of statistical power
- No group should contain less than 10 (or 5) (try to regroup instead)
- Not apt for low sample size.
- Informs of presence or absence (probability of occurrence) of association but doesn't measure strength of association

Can J Psychiatry. 1998 Oct;43(8):837-42.

# Life after chi-squared: an introduction to log-linear analysis.

Streiner DL[1], Lin E.

Author information

1 Kunin-Lunenfeld Applied Research Unit, Baycrest Centre for Geriatric Care, North York, Ontario. dstreiner@rotman-baycrest.on.ca

**Abstract**

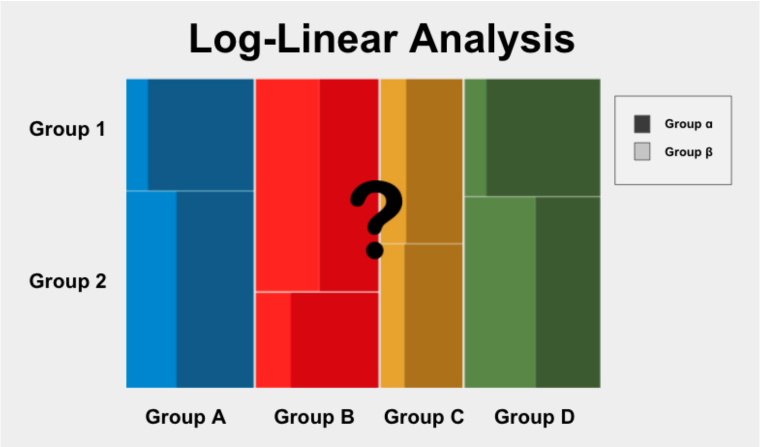
Chi-squared tests are used to examine the relationships among categorical variables. However, they are difficult to use and interpret when more than 2 variables are involved. In such cases, it is better to use a related statistic, called log-linear analysis. This article is an introduction to log-linear models, illustrating how they can be used to tease apart relationships among several variables in looking at the factors associated with photonumerophobia.

| | Age category of car | | |
|---|---|---|---|
| | New | Old | Total |
| **Male drivers** | | | |
| **Behaviour at amber light** | | | |
| Stopped | 79 | 63 | 142 |
| Did not stop | 87 | 95 | 182 |
| Total | 166 | 158 | 324 |
| **Female drivers** | | | |
| **Behaviour at amber light** | | | |
| Stopped | 95 | 83 | 178 |
| Did not stop | 51 | 94 | 145 |
| Total | 146 | 177 | 323 |
| Total old/new cars: | 312 | 335 | 647 |

**Table 18.15** Stopping behaviour of male and female drivers in old and new cars.

# Log-Linear Analysis

- variable of interest is proportional or categorical
- have two or more options
- no assumptions of IV or DV
- used for both hypothesis testing and model building



# Different types of tests (Summary)

| **Test type** | **Between subjects designs (Independent samples)** | **Within subject designs (repeated measures/matched pairs)** |
|---|---|---|
| Non-parametric (for categorical data) | Chi-square | *The binomial sign test* |
| Non-parametric (for ordinal data) | *Mann-Whitney U* | *Wilcoxon Signed-Rank test*<br>*The binomial sign test* |
| Parametric | *Unrelated t-test (level of data: interval)* | *Related t-test (level of data: interval)* |