# Parametric & Nonparametric Models for Tests of Association

- Models we will consider
- $X^2$ Tests for qualitative variables
- Parametric tests
  - Pearson's correlation
- Nonparametric tests
  - Spearman's rank order correlation (Rho)
  - Kendal's Tau

## Statistics We Will Consider

| DV ➔ | Categorical | Parametric Interval/ND | Nonparametric Ordinal/~ND |
|---|---|---|---|
| univariate stats | mode, #cats | mean, std | median, IQR |
| univariate tests | gof $X^2$ | 1-grp t-test | 1-grp Mdn test |
| association | $X^2$ | Pearson's r | Spearman's r |
| 2 bg | $X^2$ | t- / F-test | M-W K-W Mdn |
| k bg | $X^2$ | F-test | K-W Mdn |
| 2wg | McNem Crn's | t- / F-test | Wil's Fried's |
| kwg | Crn's | F-test | Fried's |

M-W -- Mann-Whitney U-Test  Wil's -- Wilcoxin's Test  Fried's -- Friedman's F-test
K-W -- Kruskal-Wallis Test
Mdn -- Median Test  McNem -- McNemar's $X^2$  Crn's – Cochran's Test

Statistical Tests of Association w/ qualitative variables

Pearson's $X^2$

$$X^2 = \sum \frac{(of - ef)^2}{ef}$$

Can be 2x2, 2xk or kxk – depending upon the number of categories of each qualitative variable

- H0: There is no pattern of relationship between the two qualitative variables.
- degrees of freedom df = (#colums - 1) * (#rows - 1)
- Range of values 0 to ∞
- Reject Ho: If $X^2_{obtained} > X^2_{critical}$

$$ef = \frac{\text{Row total} * \text{Column total}}{N}$$

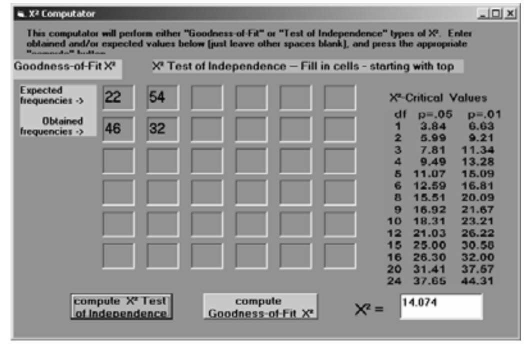
| | Col 1 | Col 2 | |
|---|---|---|---|
| Row 1 | 22 | 54 | 76 |
| Row 2 | 46 | 32 | 78 |
| | 68 | 86 | 154 |

The expected frequency for each cell is computed assuming that the H0: is true – that there is no relationship between the row and column variables.

If so, the frequency of each cell can be computed from the frequency of the associated rows & columns.

| | Col 1 | Col 2 | |
|---|---|---|---|
| Row 1 | (76*68)/154 | (76*86)/154 | 76 |
| Row 2 | (78*68)/154 | (78*86)/154 | 78 |
| | 68 | 86 | 154 |

$$X^2 = \sum \frac{(of - ef)^2}{ef}$$




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

$X^2_{1, .05} = 3.84$

$X^2_{1, .01} = 6.63$

p = .0002 using online p-value calculator

So, we would reject H0: and conclude that there is a pattern of relationship between the variables.

Parametric tests of Association using ND/Int variables

## Pearson's correlation

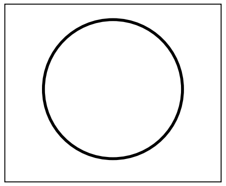
- H0: No linear relationship between the variables, in the population represented by the sample.
- degrees of freedom df = N - 2
- range of values - 1.00 to 1.00
- reject Ho: If $|r_{obtained}| > r_{critical}$

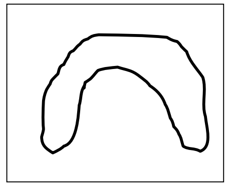
Pearson's correlation is an index of the direction and extent of the linear relationship between the variables.

It is important to separate the statements…

- there is no linear relationship between the variables
- there is no relationship between the variables
- correlation only addresses the former!

Correlation can not differentiate between the two bivariate distributions shown below – both have no linear relationship





One of many formulas for r is shown on the right.

$$r = \frac{\Sigma\ Z_X * Z_Y}{N}$$

- each person's "X" & "Y" scores are converted to Z-scores (M=0 & Std=1).
- r is calculated as the average Z-score cross product.

+r results when most of the cross products are positive (both Zs + or both Zs -)

-r results when most of the cross products are negative (one Z + & other Z-)

Nonparametric tests of Association using ~ND/~Int variables

## Spearman's Correlation

- H0: No rank order relationship between the variables, in the population represented by the sample.
- degrees of freedom df = N - 2
- range of values - 1.00 to 1.00
- reject Ho: If $| r_{obtained} | > r_{critical}$

## Computing Spearman's r

One way to compute Spearman's correlation is to convert X & Z values to ranks, and then correlate the ranks using Pearson's correlation formula, applying it to the ranked data. This demonstrates…

- rank data are "better behaved" (i.e., more interval & more ND) than value data
- Spearman's looks at whether or not there is a linear relationship between the ranks of the two variables

The most common formula for Spearman's Rho is shown on the right.

$$r = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

To apply the formula, first convert values to ranks.

| | # practices | # correct | rank # practices | rank # correct | d | $d^2$ |
|---|---|---|---|---|---|---|
| S1 | 6 | 21 | 4 | 5 | -1 | 1 |
| S2 | 2 | 18 | 1 | 4 | -3 | 9 |
| S3 | 4 | 7 | 2 | 1 | 1 | 1 |
| S4 | 9 | 15 | 5 | 3 | 2 | 1 |
| S5 | 5 | 10 | 3 | 2 | 1 | 1 |

$\Sigma d^2 = 13$

$$r = 1 - \frac{6 * 13}{5 * 24} = 1 - .65 = .35$$

For small samples (n < 20) r is compared to r-critical from tables.
For larger samples, r is transformed into t for NHSTesting.

Remember to express results in terms of the direction and extent of rank order relationship !

So, how does this strange-looking formula work? Especially the "6" ???

$$r = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

Remember that we're working with "rank order agreement" across variable – a much simpler thing than "linear relationship" because there are a finite number of rank order pairings possible!

If there is complete rank order agreement between the variables …
→then, d = 0 for each case & $\Sigma d^2 = 0$
→so, r = 1-0
→r = 1 → indicating a perfect rank-order correlation

If the rank order of the two variables is exactly reversed…
→ $\Sigma d^2$ can be shown to be $n(n^2-1)/3$
→ the equation numerator becomes $6 * n(n^2 – 1)/3 = 2 * n(n^2 – 1)$
→ so, r = 1 – 2
→ r = -1 → indicating a perfect reverse rank order correlation

If there is no rank order agreement of the two variables …
→$\Sigma d^2$ can be shown to be $n(n^2-1)/6$
→the equation numerator becomes $6 * n(n^2 – 1)/6 = n(n^2 – 1)$
→so, r = 1 – 1
→ r = 0 → indicating no rank order correlation

Nonparametric tests of Association using ~ND/~Int variables

## Kendall's Tau

- H0: No rank order concordance between the variables, in the population represented by the sample.
- degrees of freedom df = N - 2
- range of values - 1.00 to 1.00
- reject Ho: If | robtained | > rcritical

All three correlations have the same mathematical range (-1, 1).

But each has an importantly different interpretation.

Pearson's correlation
- direction and extent of the linear relationship between the variables

Spearman's correlation
- direction and extent of the rank order relationship between the variables

Kendall's tau
- direction and proportion of concordant & discordant pairs

The most common formula for Kendall's Tau is shown on the right.**

$$tau = \frac{2(C-D)}{n(n - 1)}$$

| | # practices | # correct | rank # practices X | rank # correct Y |
|---|---|---|---|---|
| S1 | 6 | 21 | 4 | 5 |
| S2 | 2 | 18 | 1 | 4 |
| S3 | 4 | 7 | 2 | 1 |
| S4 | 9 | 15 | 5 | 3 |
| S5 | 5 | 10 | 3 | 2 |

To apply the formula, first convert values to ranks.

| | # practices | # correct | rank # practices X | rank # correct Y |
|---|---|---|---|---|
| S2 | 2 | 18 | 1 | 4 |
| S3 | 4 | 7 | 2 | 1 |
| S5 | 5 | 10 | 3 | 2 |
| S1 | 6 | 21 | 4 | 5 |
| S4 | 9 | 15 | 5 | 3 |

Then, reorder the cases so they are in rank order for X.

**There are other forumlas for tau that are used when there are tied ranks.

| | # practices ` X | # correct Y | rank # practices X | rank # correct Y | **C** | **D** |
|---|---|---|---|---|---|---|
| S2 | 2 | 18 | 1 | 4 | 1 | 3 |
| S3 | 4 | 7 | 2 | 1 | 3 | 0 |
| S5 | 5 | 10 | 3 | 2 | 2 | 0 |
| S1 | 6 | 21 | 4 | 5 | 0 | 1 |
| S4 | 9 | 15 | 5 | 3 | | |
| | | | | sum | 6 | 4 |

For each case…

C = the number of cases listed below it that have a larger Y rank

(e.g., for S2, C=1 → there is one case below it with a higher rank - S1 )

D = the number of cases listed below it that have a smaller Y rank

(e.g., for S2, D=3 → there are 3 cases below it with a lower rank - S3 S5 S4)

$$\text{tau} = \frac{2(C-D)}{n(n-1)} = \frac{2(6-4)}{5(5-1)} = \frac{4}{20} = .20$$

For small samples (n < 20) tau is compared to tau-critical from tables.
For larger samples, tau is transformed into Z for NHSTesting.