Parametric & Nonparametric Models for Within-Groups Comparisons

- overview
- X^2 tests
- parametric & nonparametric stats
- Mann-Whitney U-test
- Kruskal-Wallis test
- Median test

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 for BG Designs w/ qualitative variables

Pearson's X^2

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

Can be 2x2 or kxk – depending upon the number of categories of the qualitative outcome variable

- H0: Populations represented by the design conditions have the same distribution across conditions/categories of the outcome variable
- 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

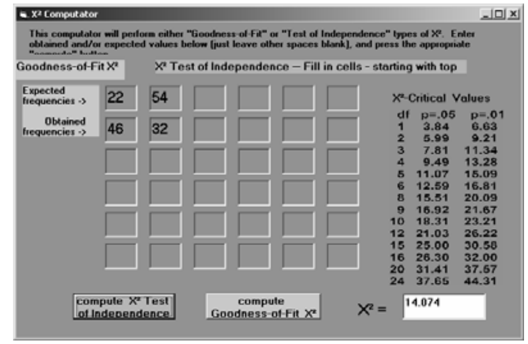
The expected frequency for each cell is computed assuming that the H0: is true – that there is no relationship between the row and column variables.

Usually the column variable is the grouping variable and the row variable is the DV.

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 the two groups have different distributions of responses of the qualitative DV.

Parametric tests for BG Designs using ND/Int variables

t-tests

- H0: Populations represented by the IV conditions have the same mean DV.
- degrees of freedom df = N - 2
- Range of values $-\infty$ to ∞
- Reject Ho: If $|t_{obtained}| > t_{critical}$
- Assumptions
  - data are measured on an interval scale
  - DV values from both groups come from ND with equal STD

ANOVA

- H0: Populations represented by the IV conditions have the same mean DV.
- degrees of freedom df numerator = k-1, denominator = N - k
- Range of values 0 to ∞
- Reject Ho: If $F_{obtained} > F_{critical}$
- Assumptions
  - data are measured on an interval scale
  - DV values from both groups come from ND with equal STD

Nonparametric tests for BG Designs using ~ND/~Int variables

The nonparametric BG models we will examine, and the parametric BG models with which they are most similar…

2-BG Comparisons

Mann-Whitney U test	between groups t-test

2- or k-BG Comparisons

Kruskal-Wallis test	between groups ANOVA
Median test	between groups ANOVA

As with parametric tests, the k-group nonparametric tests can be used with 2 or k-groups.

Let's start with a review of applying a between groups t-test

Here are the data from such a design :

Qual variable is whether or not subject has a 2-5 year old

Quant variable is "liking rating of Barney" (1-10 scale)

No Toddlertoddler		1+ Toddlers	
s1	2	s3	6
s2	4	s5	8
s4	6	s6	9
s8	7	s7	10
M =	4.75	M =	8.25

The BG t-test would be used to compare these group means.

When we perform this t-test …

As you know, the H0: is that the two groups have the same mean on the quantitative DV, but we also …

1. Assume that the quantitative variable is measured on a interval scale -- that the difference between the ratings of "2" and "4" mean the same thing as the difference between the ratings of "8" and "6".
2. Assume that the quant variable is normally distributed.
3. Assume that the two samples have the same variability (homogeneity of variance assumption)

Given these assumptions, we can use a t-test tp assess the

H0: $M_1 = M_2$

Nonparametric tests for BG Designs using ~ND/~Int variables

If we want to "avoid" these first two assumptions, we can apply the Mann-Whitney U-test

The test does not depend upon the interval properties of the data, only their ordinal properties -- and so we will convert the values to ranks

• lower scores have lower ranks, and *vice versa*

  • e.g. #1 values 10 11 13 14 16
    ranks 1 2 3 4 5

• Tied values given the "average rank" of all scores with that value

  • e.g. #2 values 10 12 12 13 16
    ranks 1 2.5 2.5 4 5

  • e.g., #3 values 9 12 13 13 13
    ranks 1 2 4 4 4

Preparing these data for analysis as ranks...

No Toddlestoddler	rating	ranks	1+ Toddlers	rating	ranks
s1	2	1	s3	6	3.5
s2	4	2	s5	8	6
s4	6	3.5	s6	9	7
s8	7	5	s7	10	8
	Σ = 11.5			Σ = 24.5	

All the values are ranked at once -- ignoring which condition each "S" was in.

Notice the group with the higher values has the higher summed ranks

The "U" statistic is computed from the summed ranks. U=0 when the summed ranks for the two groups are the same (H0:)

There are two different "versions" of the H0: for the Mann-Whitney U-test, depending upon which text you read.

The "older" version reads:

H0: The samples represent populations with the same distributions of scores.

Under this H0:, we might find a significant U because the samples from the two populations differ in terms of their:

• centers (medians - with rank data)

• variability or spread

• shape or skewness

This is a very "general" H0: and rejecting it provides little info.

Also, this H0: is not strongly parallel to that of the t-test (which is specifically about mean differences)

Over time, "another" H0: has emerged, and is more commonly seen in textbooks today:

H0: The two samples represent populations with the same median (assuming these populations have distributions with identical variability and shape).

You can see that this H0:

- increases the specificity of the H0: by making assumptions (That's how it works - another one of those "trade-offs")
- is more parallel to the H0: of the t-test (both are about "centers")
- has essentially the same distribution assumptions as the t-test (equal variability and shape)

Finally, there are two "forms" of the Mann-Whitney U-test:

With smaller samples (n < 20 for both groups)

- compare the summed ranks fo the two groups to compute the test statistic -- U
- Compare the $W_{obtained}$ with a $W_{critical}$ that is determined based on the sample size

With larger samples (n > 20)

- with these larger samples the distribution of U-obtained values approximates a normal distribution
- a Z-test is used to compare the Uobtained with the Ucritical
- the $Z_{obtained}$ is compared to a critical value of 1.96 (p = .05)

Nonparametric tests for BG Designs using ~ND/~Int variables

The Kruskal- Wallis test

- applies this same basic idea as the Mann-Whitney U-test (comparing summed ranks)
- can be used to compare any number of groups.

- DV values are converted to rankings
  - ignoring group membership
  - assigning average rank values to tied scores
- Score ranks are summed within each group and used to compute a summary statistic "H", which is compared to a critical value obtained from a X^2 distribution to test H0:
  - groups with higher values will have higher summed ranks
  - if the groups have about the same values, they will have about the same summed ranks

H0: has same two “versions” as Mann-Whitney U-test

- groups represent populations with same score distributions
- groups represent pops with same median (assuming these populations have distributions with identical variability and shape).

- Rejecting H0: tells only that there is some pattern of distribution/median difference among the groups
  - specifying this pattern requires pairwise K-W follow-up analyses
  - Bonferroni correction -- $p_{critical} = (.05 / \text{\# pairwise comps})$

Nonparametric tests for BG Designs using ~ND/~Int variables

Median Test -- also for comparing 2 or multiple groups

The intent of this test was to compare the medians of the groups, without the “distributions are equivalent” assumptions of the Mann-Whitney and Kruskal-Wallis tests

This was done in a very creative way

- compute the grand median (ignoring group membership)
- for each group, determine which members have scores above the grand median, and which have scores below the grand median

Assemble the information into a contingency table

- Perform a Pearson’s (contingency table) X^2 to test for a pattern of median differences (pairwise follow-ups)
- Please note: The median test has substantially less power than the Kruskal-Wallis test for the same sample size

e.g., $Mdn_1 = Mdn_2 = Mdn_3$

	G_1	G_2	G_3
⇧	12	13	21
⇩	13	11	19

$X^2 = 0$

e.g., $Mdn_1 > Mdn_2 < Mdn_3$

	G_1	G_2	G_3
⇧	20	8	22
⇩	5	16	18

$X^2 > 0$