Statistiek I

Nonparametric Tests

John Nerbonne

CLCG, Rijks*universiteit* Groningen

http://www.let.rug.nl/nerbonne/teach/Statistiek-I/

Nonparametric Tests

NONPARAMETRIC, DISTRIBUTION-FREE Tests —aren't summarized as parameters to distributions, i.e. $N(0,1)$, $t(18)$, $F(3,36)$ or $B(10,0.3)$

- applied when distribution unknown
  & when dist. violates condition of parametric test
- often best option for nonnumeric data
- less sensitive than parametric tests!
- χ^2 is also non-parametric
- several popular tests
  - Mann-Whitney (U-Test)—like *t*-test
    Kruskal-Wallis ($>$ 2 groups) —where dist. not normal (but still symm.)
  - Wilcoxon Signed-Rank Test—like paired t-test where dist. not normal (but still symm.)
  - Sign Test—where asymmetry possible

Mann-Whitney U-Test

alternative to t-test (independent samples)

- applicable to ordinal data
- compares two samples
- tests H_0 : samples from same population
  vs. H_a : samples from diff. populations
- alternative to independent sample t-test
- example: SSHA (Survey of Study Habits & Attitudes) compares men, women on motivation, study habits and attitudes

Women's Scores: $154, 109, 137, 115, 140, 154, \ldots$
Men's Scores: $108, 140, 114, 91, 180, 115, \ldots$
(see exercises)

Mann-Whitney U-Test: Example

Women's Scores: $154, 109, 137, 115, 140, 154, \ldots$
Men's Scores: $108, 140, 114, 91, 180, 115, \ldots$
Take the combined set $W \cup M$, order it from lowest to highest rank

1	2	3	...
91	108	109	...
M	M	W	

Sum the ranks for both groups, $\Sigma M, \Sigma F$

$$U_M = n_M n_W + \frac{n_M(n_M+1)}{2} - \Sigma M$$
$$U_W = n_M n_W + \frac{n_W(n_W+1)}{2} - \Sigma W$$

Mann-Whitney U-Test: Definition

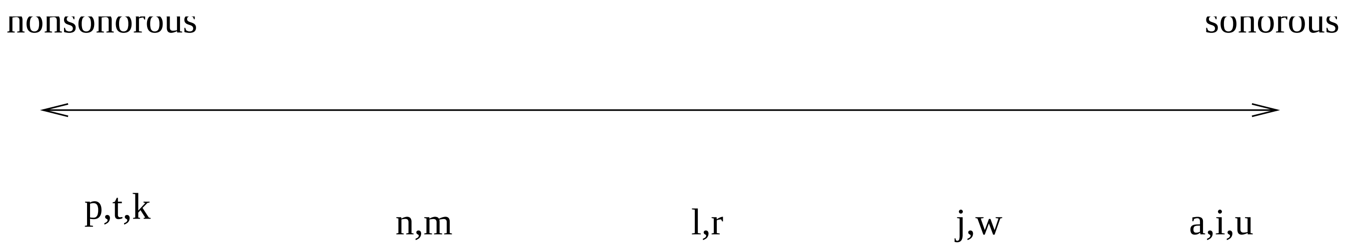
Sum the ranks for both groups, $\Sigma M, \Sigma W$
Use smaller of U_1, U_2 (here U_M, U_W), call it U
Note: if distribution is skewed, this will tend to be small (sum of ranks will be large)

Test often applied to Likkert data, i.e. of the form
`On a scale of` 1 `(easiest)` −7 `(hardest), the difficulty of this sheet is` ......

generalization to several groups: **Kruskal-Wallis**

Mann-Whitney — Example

Bastiaanse, Gilbers, v/d Linde 'Sonority Substitutions in Broca's & Conduction Aphasia' *J.Neurolinguistics* 8(4), '94
sonority scale: phonological **not** phonetic notion



sonority substitution: one that replaces a segment, changing the sonority, e.g. /pIn/ → /pIt/

Bastiaanse et al.'s Use of Mann-Whitney

- background hypothesis: conduction aphasia has more to do with higher levels of linguistics organization
- expectation: errors involving change in sonority indicate phonological problems
- therefore we expect more sonority errors in conduction aphasia than Broca's aphasia
- H_0: about the same proportion in both aphasia's
- looks like *t*-test, but distribution not normal, therefore Mann-Whitney test
- result: confirmation of alternative hypothesis (more sonority substituions in conduction aphasia)

Mann-Whitney

- useful fallback for *t*-test for independent samples
- no applicability to single-sample situations, paired data

Wilcoxon

Wilcoxon's Signed Rank Test

- like t-test, applicable to single sample, paired samples!
- normally applied to numeric data outside normal dist.
- numeric data is translated into ranked, signed data
- distribution should be roughly symmetric, not skewed
  —since hypothesis is about mean μ
- potentially applicable to pure rankings
  —need to rank differences

Wilcoxon Applied to Single Sample

Translation into ranked, signed data

Example: test reports claim $\mu = 92$ (for dyslexics) on test of dyslexia. You suspect that 92 is too high and arrange to have it administered to 10 randomly chosen dyslexics.

$$H_0\colon \mu = 92$$
$$H_a\colon \mu < 92$$

Results:

78	104	84	70	96
73	87	85	76	94

Wilcoxon Calculations

Convert the data

- col 2 convert to $\pm$ diff. to μ
- col 4 rank **unsigned** data
- col 5 add signs to ranks

Score	**Diff.**		**Rank**	**Signed Rank**
x	$\delta = x - \mu$	$\lvert\delta\rvert$	**of** $\lvert\delta\rvert$	r
78	-14	14	7	-7
95	3	3	2	2
84	-8	8	6	-6
70	-22	22	10	-10
96	4	4	3	3
73	-19	19	9	-9
87	-5	5	4	-4
85	-7	7	5	-5
76	-16	16	8	-8
94	2	2	1	1

W, the test statistic, is the sum of **positive** ranks (here, $W = 6$)

Wilcoxon p-Values

$$H_0\colon \mu = 92 \qquad \textbf{and} \qquad H_a\colon \mu < 92$$

If H_0 is true, then positive and negative magnitudes should be roughly the same, i.e.

$$\frac{1}{2}\sum_{i=1}^{n} i, \text{ where } n \text{ is the size of the sample}$$

Refer to tables (or SPSS or S+) for critical values of W

$$P(W_{10} \leq 8) = 0.025$$

W_{10} since the prob. of W depends on sample size n
This is **one-tailed** prob. —since hypothesis is one-tailed.

Wilcoxon in Two-Sided Hypotheses

$p = 0.025$ extreme enough to reject $H_0 : \mu = 92$ in favor of one-tailed $H_a : \mu < 92$
If we'd examined two-sided $H'_a : \mu \neq 92$, then we should have obtained:

$$P(W_{10} \leq 8) = 0.05$$

naturally, less strong **against** H_0.

Probabilities of W

	2-tailed Signif.	
	0.05	0.01
	1-tailed Signif.	
N	0.025	0.005
6	0	–
7	2	–
8	9	0
9	5	1
10	8	3
11	10	5
12	13	7
13	17	9
14	21	12
15	25	15
⋮	⋮	⋮
20	52	37
⋮	⋮	⋮
25	89	68

i.e. $P(W_{10} < 8) = 0.05$ in 2-sided hypothesis

Wilcoxon vs. t-test

Sometimes, tables list only small positive values, but right skewing results in large positive value
—To test hypothesis of right skew, use magnitude of sum of negative ranks

To compare mean in single sample of unknown σ (to some hypothesis), use the t-test

- **unless** population symmetric but not normal, e.g., some bimodal distributions
  then use Wicoxon
- what if the population is non-normal **and** asymmetric?
  —z test with large sample (> 100)

Paired Samples in Wilcoxon

Wilcoxon also used to as substitute for paired-sample *t*-test
Example: S+ exercise (French Listening Test before and after course)

Person	Before	After
1	32	34
2	31	31
⋮	⋮	⋮
20	23	26

(Assumption: dist. nonnormal)

$H_0 : \mu_b = \mu_a$ **(no diff.)**; $H_a : \mu_b < \mu_a$ **(improvement)**

1. Calculate $\delta_i = t_{ia} - t_{ib}$, convert this to signed ranks (as above), etc.
2. Use $\mu_{\delta_i} = 0$ as H_0, $\mu_{\delta_i} > 0$ as H_a, treat as single sample.
3. See laboratory exercise.

Sign Test

When all else fails ... **sign test** (use PROPORTIONS, M&M, § 5.1, earlier sheets)

- divides data into classes $+$, $-$ and 0 (only)
  e.g. positive, negative, and no change
- use: when dist. nonnormal, asymmetric
- compares proportion of positive to negative
- tests whether division is roughly chance-like
  H_0 : no weighting toward $+$ (or $-$), no change

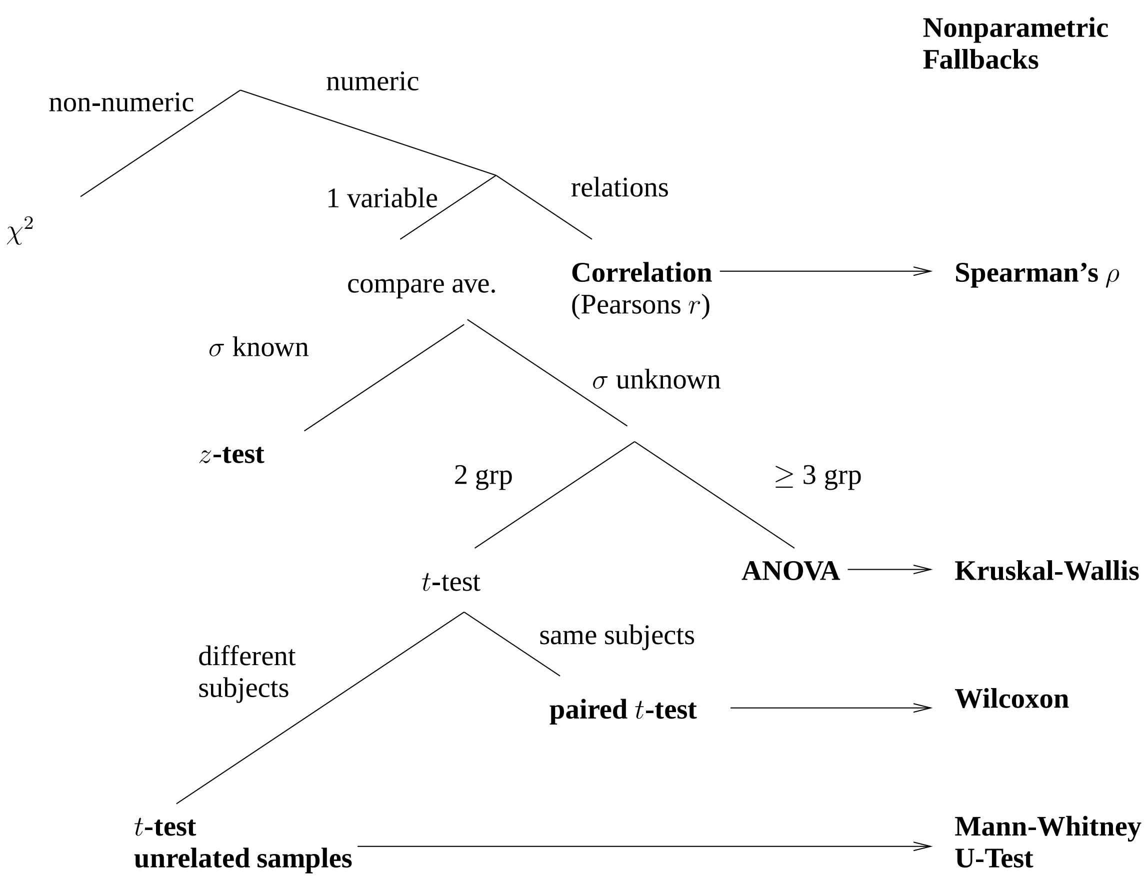
example

1. 22 aphasics judged subjectively (as belonging to one of two categories)
   question: are the judgements roughly similar?
   method: count same as $+$, different as $-$

Nonparametric Tests

NONPARAMETRIC, DISTRIBUTION-FREE Tests

- applied when distribution unknown
  & when dist. violates condition of parametric test
- often best option for nonnumeric data
- less sensitive than parametric tests!
- several easy, useful tests
  - Mann-Whitney (U-Test)—for indep. sample t-test
    Kruskal-Wallis—allows > 2 groups
    —assumes symmetry, but not normal dist.
  - Wilcoxon Signed-Rank Test—for paired t-test
    —assumes symmetry, not normal dist.
  - Sign Test—when all else fails



Statistics in Research

Research Article/ Honor's Thesis
Background

- explain background theory clearly, consistently
  minimal wrt deriving testable prediction
- explain novelty
  - genuine novelty
- derive testable predictions
  identify auxiliary assumptions
- if another theory is contrasted
  - be fair
  - show contrast in testable predictions
- summarize relevant earlier studies

Population/Sample

Design

- be clear on how theory is related to test
- describe population, relation to sample, size of sample
- note use of volunteers, drop outs
- use a control group (if possible)
  - assign subjects to control randomly

Reporting Statistical Analysis

Analysis

- make data available (ftp server)
- examine data w. descriptive statistics, tables, graphics
- justify choice of test
- show that requirements met, e.g., normal dist.
- note significance level

Discuss Results

Conclusions

- interpret results esp. vis-a-vis theory
- discuss “failed” hypotheses, too
- be sensitive to size of result vs. significance
- discuss alternative explanations
- sketch further questions

Next Week

Choice of Tests, Review for Exam