Uses and Abuses of Non-parametric Statistics in Medical Research

John P Klein, PhD

Sponsored by the Clinical and Translational Science Institute (CTSI)
and the Department of Population Health / Division of Biostatistics

Speaker Disclosure

In accordance with the ACCME policy on speaker disclosure, the speaker and planners who are in a position to control the educational activity of this program were asked to disclose all relevant financial relationships with any commercial interest to the audience. The speaker and program planners have no relationships to disclose.

Outline

- Introduction
  - What is a nonparametric test?
  - Motivating example
- Why do we use nonparametric tests
- Why not to use nonparametric tests
- Specific tests for some common situations
  - One sample, paired sample
  - Two independent samples
  - One way layout
  - Two way layout
  - Measures of Association
- Concluding remarks

Parametric Statistics

- Body of Statistical Methods Based on an Assumed Model for the Underlying Population from which the Data was sampled
- Inference is about some parameter (mean, variance, correlation) of the population
- If model is incorrect the inference may be misleading

Example of Parametric Procedures

- Classical t-test assumes normal populations
- Analysis of Variance assume normal populations
- Logistic regress assumes binomial population
- Pearson's correlation coefficient assumes bivariate normal

Non Parametric Statistics or Distribution Free statistics

- Body of statistical methods that relax the assumptions about the underlying population model
- Typically statistics based on ranks or simple counts
- Can be used for both ordinal and nominal data

Simple Example

- Study of the effects of tranquilizer on Hamilton Depression Score

X-Pre trial depression score

Y-Second visit depression score

H_0 : No change in depression score

H_A : Score is lower after treatment

Patient	Pre Score X	Post Score Y	Change in Score X-Y
1	1.83	0.80	1.03
2	0.50	0.65	-0.15
3	1.62	0.60	1.02
4	2.48	2.05	0.43
5	1.68	1.06	0.62
6	1.88	1.29	0.59
7	1.55	1.06	0.49
8	3.16	3.14	0.02
9	1.30	1.29	0.01

Data

1.03 -0.15 1.02 0.43 0.62 0.59 0.49 0.02 0.01

Sign Test

- Count number of positive differences s=8
- If null hypothesis is true the number of positive differences is like flipping a fair coin n times so

P-value =Pr[S≥s|n,1/2] (one sided)

= 2xSmaller of {Pr[S≥s|n,1/2], Pr[S≤s|n,1/2]}

- P value= Pr[S≥8|9,1/2]=0.019
- Hence Hamilton score decreased by tranquilizer

Wilcoxon Sign Rank Test

Patient	d=X-Y	Rank |d|
1	1.03	9
2	-0.15	**3**
3	1.02	8
4	0.43	4
5	0.62	7
6	0.59	6
7	0.49	5
8	0.02	2
9	0.01	1

Wilcoxon Sign Rank Test

- Add up ranks associated with the positive (or negative) differences T^+ (T^-)
- Note sum of ranks =n(n+1)/2
- Under H_o + and – values should be mixed so T^+ should be close to its average value n(n+1)/4
- Exact p-values can be found in tables which were found by numeration of all possible samples or using a large sample approximation

$$Z=(T^+-n(n+1)/4)/\{n(n+1)/(2n+1)/24\}^{1/2}$$

Calculations

- $n(n+1)/2=9x10/2=45$
- $T^-=3$ $T^+=45-3=42$
- P-value from table $p=0.006$
- Normal Approximation
- $Mean=n(n+1)/4=22.5$
- $Var=n(n+1)(2n+1)/24=71.25$
- $Z=(42-22.5)/71.25^{1/2}=2.31$ $p=0.0104$
- Hence Hamilton score decreased by tranquilizer

Why Nonparametrics?

1. Fewer Assumptions
2. Exact p-values for small sample sizes
3. Usually easier to apply. Involves counts and ranks
4. Often easier to understand why the methods work
5. Since it works on ranks does not require numerical data but can be performed on ordinal data

Why Nonparametrics?

6. Provides simple tests for complicated hypotheses
   - Example   k sample

H_o: $\mu_1=\mu_2=. . . =\mu_k$ versus

H_o: at least one μ_j is different

Parametric: Analysis of Variance (ANOVA)

Non-Parametric: Kruskal-Wallis Test

H_o: $\mu_1=\mu_2=. . . =\mu_k$ versus

H_o: $\mu_1<\mu_2<. . . <\mu_k$

Parametric: Isotonic regression ??

Non-Parametric: Jonkheere Test

Why Nonparametrics?

7. Exact confidence intervals are available
8. Gives protection against outliers
9. Software for nonparametric methods is available in most packages
10. When software is not available replace data by ranks in normal theory software

Why not Nonparametrics?

1. Loss of power when the parametric models is chosen correctly.

i.e. More likely to not reject the null hypothesis when it is false

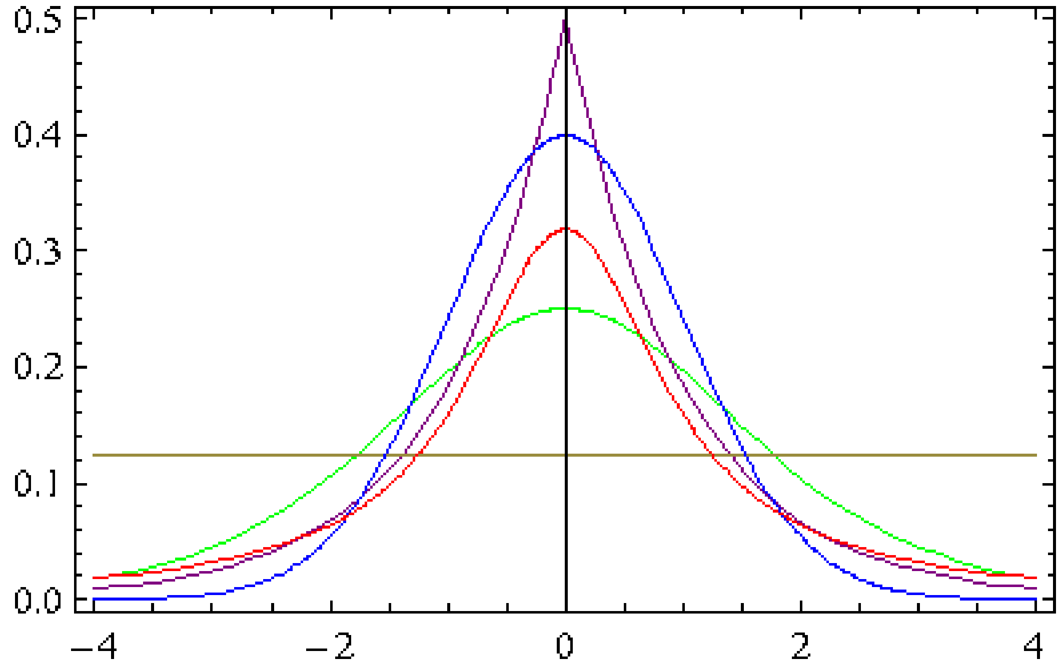
- Loss of power depends on the tails of the underlying distribution.

- Measured by the “Asymptotic relative efficiency” e(1, 2).

e(1, 2)=sample size for test 2/sample size for test 1

To achieve equal power

If e(1, 2) <1 then test 2 is more powerful $n_2=e(1,2)n_1$



	Normal	Uniform	Logistic	Double Exponential	Cauchy
e(sign test,paired t test)	.637	.333	.822	2.00	∞
e(Sign Rank, paired t)	.955	1.000	1.097	1.50	∞

Why Not Nonparametrics?

2. For large samples the Central Limit Theorem says the mean is approximately normally distributed so normal theory tests such as t-test, ANOVA which compare means are appropriate for any underlying distribution.

One or Paired sample tests

- Location
  - Parametric test: t-test
  - Nonparametric tests
    - Sign Test (No Assumptions)
    - Wilcoxon Sign Rank Test
      - Works best for symmetric populations
    - Confidence intervals based on Hodges-Lehman statistic

Two Independent Samples Location Tests

- Parametric Tests (assumes normality)
  - t-test (Assumes equal variances)
  - t-test (Scatterhwaite's approximation)
- Nonparametric tests
  - Mann-Whitney-Wilcoxon test
    - Looks at the sum of the ranks of the observations from the 1st sample in the pooled sample
  - Linear Rank tests $\Sigma\ \phi(R_i)$ where R_i is rank of the i^{th} observation in sample 1 in pooled sample. (e.g.. Log rank test, normal scores test, etc)

Two Independent Samples Tests for equal variance

- Parametric test (assumes normality)
  - F test
- Non parametric tests
  - Ansari-Bradley test (assumes equal medians)
    1. Rank Data in pooled sample
    2. Assign scores 1 2 3 4 5 5 4 3 2 1 (n=10)
    3. Add up scores for 1st sample and reject if too small or too large
  - Miller's Jackknife test if unequal medians

Two Independent Samples Omnibus tests

- Empirical distribution function

  $F_n(x) = [\# \text{ data points} \leq x]/n$
- Kolmogorov-Smirnov test

  $\max[F_n(x) - G_n(x)]$
- Cramer-von-Mises test

  Area between $F_n(x) - G_n(x)$

One way layout
k=3 or more independent groups

- Parametric—(assumes normality, equal variances) Analysis of Variance for case 1
- Nonparametric Model $X_{ij}=\theta+\tau_j+E_{ij}$, $i=1,\ldots,n_j$, $j=1,\ldots,k$
  - Case 1 H_A:[$\tau_1, \ldots, \tau_k$ not all equal]
    - Kruskal-Wallis test
  - Case 2 H_A:[$\tau_1 \leq \tau_2 \leq \ldots \leq \tau_k$ with at least one strict inequality]
    - Jonckheere-Terpstra Test

One way layout
k=3 or more independent groups

- Nonparametric Model $X_{ij}=\theta+\tau_j+E_{ij}$, $i=1,\ldots,n_j$, $j=1,\ldots,k$
  - Case 3
  - H_A:[$\tau_1 \le \tau_2 \le \ldots \le \tau_{p-1} \le \tau_p \le \tau_{p+1} \ge \ldots \ge \tau_k$ with at least one strict inequality]
    - Mack-Wolfe test

Two-Way Layout

$X_{ijt}=\theta+\beta_i+\tau_j+E_{ijt}$ $i=1,\ldots,n$ Subjects (Blocks)

$j=1,\ldots,k$ Treatments

$t=1,\ldots,c_{ij}$ replicates

- Parametric test –ANOVA
- Nonparametric test
  - Friedman's Test (H_A: τ's not equal)

Measures of Association

- Parametric—Pearson's Product moment correlation coefficient, R. $-1 \leq R \leq +1$
  - Measures strength of linear association when data is bivariate normal $-1 \leq R \leq +1$
  - When the marginals are not normal you can have perfect association and an $R \neq \pm 1$

Measures of Association Nonparametric

- Spearman's rho—Usual correlation coefficient using ranks
- Kendall's tau

  Pairs (X_i,Y_i), (X_j,Y_j) are

  concordant if $(X_i-Y_i)(X_j-Y_j) > 0$;

  discordant if $(X_i-Y_i)(X_j-Y_j) < 0$

  Tau =(#concordant-#discordant pairs)/(n(n-1)/2)

- Both measures are
  - Between -1,+1
  - Equal to 0 if X and Y are independent
  - Equal to +/-1 if Y=a+bX

Remarks

- Most of these methods are in most software packages. One particularly good package is MINITAB
- Good Reference is Hollander and Wolfe *Nonparametric Statistical Methods.* Wiley1999

Summary

- Nonparametric statistics provide a viable alternative to normal theory statistics
- Nonparametric statistics are particularly useful when
  - The sample size is small
  - You want protection from outliers
  - You have ordinal data

Summary

Nonparametric statistics are particularly useful when

- You are modeling association between variables
- You are doing tests with ordered or umbrella alternative
- You have survival data

Resources

- The Clinical and Translation Science Institute (CTSI) supports education, collaboration, and research in clinical and translational science: www.ctsi.mcw.edu
- The Biostatistics Consulting Service provides comprehensive statistical support

  http://www.mcw.edu/biostatsconsult.htm

Free drop-in consulting

- MCW/Froedtert/CHW:
  - Monday, Wednesday, Friday 1 – 3 PM @ CTSI Administrative offices (LL772A)
  - Tuesday, Thursday 1 – 3 PM @ Health Research Center, H2400
- VA: 1st and 3rd Monday, 8:30-11:30 am
  - VA Medical Center, Building 70, Room D-21
- Marquette: 2nd and 4th Monday, 8:30-11:30 am
  - Olin Engineering Building, Room 338D