

Nonparametric Statistics

1

WHAT ABOUT NON-PARAMETRICS?

- No clear theoretical probability distribution, so empirical distributions needed.
- Hence, less knowledge of form of data e.g. ranks instead of values.
- Need not focus on parameter estimation or testing; when do frequently based on “less-good” parameters, e.g. Medians; otherwise test “properties, e.g. randomness, symmetry, quality etc.
- weaker assumptions, implicit in
 - smaller sample size
 - different data - also implicit from other points. **Levels of Measurement-** Nominal, Ordinal.

2

ADVANTAGES/DISADVANTAGES

Advantages

- Power may be better if assumptions weaker
- Smaller samples and less work etc. as before.

Disadvantages - also implicit from earlier points

- loss of information /power etc. when do know more on data /when assumptions do apply
- Separate tables each test

General bases/principles: Binomial - cumulative tables, Ordinal data, Normal - large samples, Kolmogorov-Smirnov for Empirical Distributions - shift in Median/Shape, Confidence Intervals- more work to establish. Use Confidence Regions and Tolerance Intervals

Errors – Type I, Type II . **Power** from significance level, true value, size of sample, actual test.

3

Main Nonparametric Statistical Tests

- **Sign test.**
- **Wilcoxon signed rank test.**
- **Median Test.**
- **Mann Whitney U Test.**
- **Kolmogorov-Smirnov Test.**
- **Kruskal-Wallis One-Way ANOVA.**
- **Friedman Test**

4

Sign Test

- Used in situations with limited ability to assess ranking of differences
 - can only assess if score for a subject is less than, greater than, or equal to the paired score
 - Test statistic depends only on the sign of the differences
 - Special case of one-sample binomial test with $p=0.5$
- Assumptions
 - Random sample
 - Ordinal measurement with continuous values

5

Sign Test

- **Example:** Suppose want to test if weights of a certain item likely to be more or less than 220 g.
From 12 measurements, selected at random, count how many above, how many below. Obtain 9(+), 3(-)
- **Hypothesis :** H_0 : Median = 220. “Test” on basis of counts of signs.
- Binomial situation, $n=12$, $p=0.5$.
For this distribution
 $P\{3 \leq X \leq 9\} = 0.9614$ while $P\{X \leq 2 \text{ or } X \geq 10\} = 1 - 0.9614 = 0.0386$
Result is not strongly significant.
- **Notes:** Need not be Median as “Location of test” (Describe distributions by Location, dispersion, shape). Location = median, “quartile” or other percentile.
Many variants of sign test - including e.g. runs of + and - signs for “randomness”.

6

Wilcoxon Signed-Rank test

- Widely used alternative to t-test
 - t-test robust against deviations from normality, but some situations are extremely non-normal.
 - Uses relative rankings, but does not depend on interval assumptions
- Use sums of ranks of differences
 - More 'powerful' than the sign test which ignores the magnitude of the differences.
 - Tests the number of positive differences in pair
- Assumptions
 - Random sample
 - Ordinal data from a continuous distribution
 - Symmetric population distributed around mean

7

Example: Wilcoxon Signed-Rank test

- Direction and Magnitude** : $H_0: \theta = 220$ (?Symmetry)
- Arrange all sample deviations from median in order of magnitude and replace by ranks (1 = smallest deviation, n largest). High value of Sum positive (or negative) ranks, relative to the other $\Rightarrow H_0$ unlikely, e.g.

Weights 126 142 156 228 245 246 370 419 433 454 478 503
Diffs. -94 -78 -64 8 25 26 150 199 213 234 258 283
Rearrange 8 25 26 -64 -78 -94 150 199 213 234 258 383
Signed ranks 1 2 3 -4 -5 -6 7 8 9 10 11 12

Clearly $S_{\text{negative}} = 15$ and $< S_{\text{positive}}$

Tables of form: Reject H_0 if $\text{Min}(S_{\text{negative}}, S_{\text{positive}}) \leq$ tabled value
 For example here, $n=12$ at $\alpha = 5\%$ level, tabled value =13, so do not reject H_0

8

Median test

- Used to test the hypothesis that two samples are from populations with equal medians
- Calculates the proportions in each group above/below the common median of the two groups
- Uses a chi-square test to test the differences between these frequencies
- Assumptions
 - Independent, random samples
 - Ordinal scale and continuous values
 - Any difference reflected in median

9

Wilcoxon-Mann-Whitney U Test

- Test of ranks between two samples
- Ranks the pooled observations in the two samples and then total the ranks in each sample
- If the medians are the same, the ranks will be similar
- Assumptions
 - Independent, random samples
 - Ordinal scale and continuous values
 - Any difference reflected in the medians

10

Wilcoxon-Mann-Whitney test

- Parallel** with parametric again H_0 : Samples from same populations (Medians same) vs H_1 : Medians are not the same
- For two samples, size m, n , calculate joint ranking and Sum for each sample, giving S_m and S_n . Should be similar if populations sampled are.
- $S_m + S_n =$ sum of all ranks = $\frac{1}{2}(m+n)(m+n+1)$ and result tabulated for

$$U_m = S_m - \frac{1}{2}m(m+1), \quad U_n = S_n - \frac{1}{2}n(n+1)$$
- Clearly, $U_m = mn - U_n$ so need only calculate one.
- Tables typically give, for various m, n , the value to exceed for smallest U in order to reject H_0 . 1-tailed/2-tailed. Easier: use sum of smaller ranks or fewer values.
- Example in brief:** If sum of ranks =12 say, probability based on no. possible ways of obtaining 12 out of Total no. possible sums

11

Example: Wilcoxon-Mann-Whitney test

- Take example on weights earlier. Assume we now have a second set from another sample
 29 39 60 78 82 112 125 170 192 224 263 275 276 286 369 756
- Combined ranks for the two samples are:
Value 29 39 60 78 82 112 125 126 142 156 170 192 224 228 245
Rank 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Value 246 263 275 276 286 369 370 419 433 454 478 503 756
Rank 16 17 18 19 20 21 22 23 24 25 26 27 28

Here $m = 16, n=12$ and $S_m = 1+2+3+\dots+21+28 = 187$
 So $U_m = 51$, and thus $U_n = 141$. Clearly, can check by calculating U_n directly
 For a 2-tailed test at 5% level $U=53$ from tables and our value is greater than the lower of the two, so reject H_0 . **Medians are different here**

12

Kolmogorov-Smirnoff Goodness-of-Fit Test

- Alternative to the chi-square goodness-of-fit test to test for difference between a sample cumulative distribution and a theoretical cumulative distribution
- Test statistic is the greatest difference at any point between the sample distribution and the theoretical distribution
- It is often used when the data have not met either the assumption of normality or the assumption of equal variances.
- The null and alternative hypotheses for the K-S test relate to the equality of the two distribution functions [usually noted as F(X) or F(Y)].

Thus, the typical two-tailed hypothesis becomes:

$$H_0: F(X) = F(Y) \quad \text{vs.} \quad H_1: F(X) \neq F(Y)$$

13

General Advantages/Disadvantages K-S

- Easy to apply
- Relatively easy to obtain C.I.
- Generally deals well with continuous data.
- Discrete possible, but test criteria not exact, so can be inefficient.
- For two groups, need same number of observations
- Distinction between location/shape differences not established.
- Advantages
 - Does not require grouped observations
 - Can be used with continuous measures
 - Can be used with any sample size.

14

KOLMOGOROV-SMIRNOV and EMPIRICAL DISTRIBUTIONS

- **Purpose** - to compare set of measurements (two groups with each other) or one group with expected - to analyse differences.
- **Cannot** assume Normality of underlying distribution, shape form, so need enough sample values to base comparison on (e.g. ≥ 4 , 2 groups)
- **Major features** - sensitivity to differences in both shape and location of Medians (do not distinguish which is different)
- **Empirical** c.d.f., not p.d.f. - looks for **consistency** by comparing population curve (expected case) with empirical curve (sample values)

Step function
$$S(x) = E.D.F. = \frac{\text{No. sample values} \leq x}{n}$$

→value at each step from data

- S(x) should never be too far from F(x) = "expected" form
- **Test Basis** is $\text{Max. Diff. } |F(x_i) - S(x_i)|$

15

Kruskal-Wallis One-Way ANOVA

- An alternative to parametric ANOVA using ranks
 - Uses more information than the k-sample median test
- Assign ranks to grouped observations and then sum ranks in each group
- Test statistics is distributed as a chi-square statistics for larger samples
 - Distribution for small samples tabled
- Assumptions
 - Independent, random samples
 - Ordinal measurements
 - Distributions differ by location parameter

16

MANY SAMPLES- Kruskal-Wallis

- Direct extension of W-M-W. Tests: H_0 : Medians are the same.
- Rank total number of observations for all samples from smallest (rank 1) to highest (rank N for N values). Tied observations given mid-rank.
- r_{ij} is rank of observation x_{ij} and s_i = sum of ranks in i^{th} sample (group)
- Compute treatment and total SSQ ranks - **uncorrected** given as

$$S_t^2 = \sum_i \frac{s_i^2}{n_i}, \quad S_r^2 = \sum_{i,j} r_{ij}^2$$

- For **no ties**, this simplifies $S_r^2 = N(N+1)(2N+1)/6$
- Subtract off **correction for mean** for each, given by $C = \frac{1}{4} N(N+1)^2$
- **Test Statistic**
$$T = \frac{(N-1)[S_t^2 - C]}{(S_r^2 - C)} \approx \chi_{i-1}^2$$

i.e. approx. χ^2 for moderate/large N. Simplifies further if no ties.

17

LARGE SAMPLES and CONFIDENCE INTERVALS

- **Normal Approximation** for S the smaller in magnitude of rank sums

$$Z(\text{or } U = S.N.D.) = \frac{S + \frac{1}{2} - \frac{1}{4}n(n+1)}{\sqrt{[n(n+1)(2n+1)/24]}} = \frac{\text{Obs} - \text{Exp.}}{SE} \sim N(0,1)$$

so C.I. as usual

- **General for C.I.** Basic idea to take pairs of observations, calculate mean and omit largest / smallest of $(1/2)(n)(n+1)$ pairs in similar way to rank sums for H.T. Usually, computer-based - re-sampling or graphical techniques.
- **Alternative Forms** -common for non-parametrics
e.g. for Wilcoxon Signed Ranks. Use $W = |S_p - S_n|$ = magnitude of differences between positive /negative rank sums. Different table.
- **Ties** - complicate distributions and significance. Assign mid-ranks

18

PAIRING/RANDOMIZED BLOCKS - Friedman

- Blocks of units, **two** treatments allocated at random within block = matched pairs; can use a variant of **sign test** (on differences)
- **Many samples or units = Friedman** (simplest case of R.B. design)
- **Recall** comparisons **within** pairs/blocks more precise than **between**, so including Blocks term, "removes" block effect as source of variation.
- Friedman's test- replaces observations by ranks (within blocks) to achieve this. (Thus, ranked data can also be used directly).

Have x_{ij} = response. Treatment i , ($i=1, \dots, t$) in each block j , ($j=1, \dots, b$)

Ranked within blocks

Sum of ranks obtained **each treatment** s_i , $i=1, \dots, t$

If r_{ij} rank (mid-rank if ties), raw (uncorrected) rank SSQ $S_i^2 = \sum_{j=1}^b r_{ij}^2$

19

Friedman contd.

- With no ties simplifies $S_i^2 = bt(t+1)(2t+1)/6$
- **Need also** SSQ(All ranks) = Uncorrected **Total** SSQ in ANOVA

$$S_i^2 = \sum_j \frac{s_{ij}^2}{b}$$

- Again, the correction factor analogous to that for K-W

$$C = \frac{1}{4}bt(t+1)^2$$

- and common form of Friedman Test Statistic

$$T_1 = \frac{b(t-1)(S_i^2 - C)}{(S_i^2 - C)} \approx \chi_{t-1}^2$$

t, b not very small, otherwise need exact tables.

20

NON-PARAMETRIC C.I. in GENOMICS. BOOTSTRAP

- **Bootstrapping**= re-sampling technique used to obtain Empirical distribution for estimator in construction of non-parametric C.I.
- **Effective** when distribution unknown or complex
- **More computation** than parametric approaches and may fail when sample size of original experiment is small
- **Re-sampling** implies sampling from a sample - usually to estimate empirical properties, (such as variance, distribution, C.I. of an estimator) and to obtain EDF of a test statistic- common methods are **Bootstrap, Jackknife, shuffling**
- **Aim** = approximate numerical solutions (like confidence regions). Can handle bias in this way - e.g. MLE of variance σ^2 , mean unknown
- both Bootstrap and Jackknife used, Bootstrap more often for C.I.

21

Bootstrap/Non-parametric C.I. contd.

- **Basis** - both Bootstrap and others rely on fact that sample cumulative distn fn. (CDF or just DF) = MLE of a population Distribution Fn. $F(x)$
- **Define** Bootstrap sample as a r.s. size n , drawn with replacement from a sample of n objects

For S the original sample, $S = (x_1, x_2, \dots, x_n)$

$P\{\text{drawing each item, object or group}\} = 1/n$

Bootstrap sample S^b obtained from original, s.t. sampling n times with replacement gives $S_b = (x_1^b, x_2^b, \dots, x_n^b)$

Power relies on the fact that large number of re-sampling samples can be obtained from a single original sample, so if repeat process b times

obtain S_j^b , $j=1, 2, \dots, b$, with each of these a **bootstrap replication**

22

Contd.

- **Estimator** - obtained from each sample. If $\hat{\theta}_j^b = F(S_j^b)$ is the estimate for the j th replication, then bootstrap mean and variance

$$\bar{\theta}^b = \frac{1}{b} \sum_{j=1}^b \hat{\theta}_j^b, \quad \hat{V}^b = \frac{1}{b-1} \sum_{j=1}^b (\hat{\theta}_j^b - \bar{\theta}^b)^2$$

while $\text{Bias}^b = \bar{\theta}^b - \theta$

- **CDF of Estimator** is $\text{CDF}(x) = P\{\hat{\theta}_b^b \leq x\}$ for b replications so **C.I.** with confidence coefficient γ is then - for a % ile **C.I.**

$$\{CDF^{-1}[0.5(1-\gamma)], CDF^{-1}[0.5(1+\gamma)]\}$$

- **Normal Approx. for mean:** Large $b \rightarrow \bar{\theta}^b \pm Z_{(1-\gamma)/2} \sqrt{\hat{V}^b}$ $Z \sim N(0,1)$ or t_{b-1} -distribution if No. bootstrap reps. smaller

23

Example:

- **Recall** rust resistant gene problem
- Suppose MLE, 1000 bootstrapping replications gave results:

	R.F.	Escapes
	$\hat{\theta}$	\hat{p}
Parametric Var	0.0001357	0.00099
95% C.I.	(0, 0.0455)	(0.162, 0.286)
95% Interval (Likelihood)	(0.06, 0.056)	(0.17, 0.288)
Bootstrap Var	0.0001666	0.0009025
Bias	0.0000800	0.0020600
95% C.I. Normal	(0, 0.048)	(0.1675, 0.2853)
95% C.I. (Percentile)	(0, 0.054)	(0.1815, 0.2826)

24

SUMMARIZING NON-PARAMETRIC USAGE

- **Sign Tests**- wide number of variants. Simple basis
- **Wilcoxon Signed Rank**-Compare medians paired data(measurements)
-Conditions/Assumptions-No. pairs ≥ 6 ; Distns. Same shape.
- **Mann-Whitney U** - Compare medians 2 groups
- Conditions/Assumptions- $(N \geq 4)$; Distributions same shape
- **Kolmogorov-Smirnov** - Compare either medians, shapes (distributions).
Conditions etc.-
 $(N \geq 4)$, two features not distinguished. 1 or 2 groups(equal numbers if 2)
- **Kruskal-Wallis**- Many group comparisons. Conditions etc.- Groups can be unequal size. Distributions same shape.
- **Friedman** - Many group comparison of medians. Conditions etc.-
Data in R.B. design. Distributions same shape.

25

Comparison between Two Samples

Independent Samples.

- Normal Data, Equal Variances or Unequal Variances, t-test.
- Non-normal Data, Equal Variances, Wilcoxon -Mann-Whitney U-test.
- Non-normal Data, Unequal Variances, Kolmogorov-Smirnov test.

Paired Samples.

- Normal Data, continuous variable, t-test.
- Non-normal Data, ranked variable, Wilcoxon Signed-Rank test.

26