Slides 14: Non-parametric tests

- Sign test
- Confidence interval for the median
- Rank sum test
- Signed rank test
- Kruskal-Wallis test
- Fridman test

1. Herb Alpert
2. Madonna
3. Celine Dion
4. Julio Iglesias
5. Beyonce
6. Dolly Parton
7. Gloria Estefan
8. Bruce Springsteen
9. Victoria Beckham
10. Barbara Streisand

Example of ranking: 10 richest singers

The list of frequentist tests

One-sample tests

- One sample t-test: normal population distribution
- Large sample test for mean
- Large sample test for proportion: categorical data
- Small sample test for proportion: categorical data
- Chi-squared test of goodness of fit: categorical data, large sample
- Chi-squared test of independence: categorical data, large sample
- Model utility test: linear model, several explanatory variables, normal noise, homoscedasticity

Two-sample tests

- Two sample t-test: normal populations, equal variances, independent samples
- Fisher's exact test: categorical data, independent samples
- McNemar: categorical data, matched samples, large samples

Several samples

- ANOVA 1: normal population distributions, equal variances, independent samples
- ANOVA 2: normal population distributions, equal variances, matched samples
- Chi-squared test of homogeneity: categorical data, independent samples, large samples

Non-parametric tests

- Sign test: one sample
- Signed rank test: two matched samples, symmetric distribution of differences
- Rank sum test: two independent samples
- Kruskal-Wallis: several independent samples
- Fridman: several matched samples

Consider a random sample $(x_1, \ldots, x_n)$ without assuming any parametric model for the unknown population distribution. The population median m is defined by the equality

$$\mathrm{P}(X < m) = \mathrm{P}(X > m).$$

If it is known that the population distribution is continuous, we get $\mathrm{P}(X \le m) = \frac{1}{2}$.

The sign test is a non-parametric test of H_0: $m = m_0$ against the two-sided alternative H_1: $m \neq m_0$. The sign test statistic

$$y_0 = \sum_{i=1}^{n} 1_{\{x_i \le m_0\}}$$

counts the number of observations below the null hypothesis value. It has a simple null distribution $Y_0 \overset{H_0}{\sim} \mathrm{Bin}(n, 0.5)$.

Question. 11 subjects exhibited values (y_i, x_i) before and after smoking a cigarette. Only one had $y_i > x_i$. How can we apply the sign test?

Consider a random sample $(x_1, \ldots, x_n)$ from a population distribution with unknown median m. Let $(x_{(1)}, \ldots, x_{(n)})$ be the ordered sample. For any given $k < \frac{n}{2}$ we treat

$$I_m = (x_{(k)}, x_{(n-k+1)})$$

as a confidence interval for m. The number of observations below m

$$y = \sum_{i=1}^{n} 1_{\{x_i \le m\}}$$

has the symmetric binomial distribution $Y \sim \text{Bin}(n, 0.5)$. The confidence level of I_m is computed from

$$p_k = \mathrm{P}(X_{(k)} < m < X_{(n-k+1)}) = \mathrm{P}(k \le Y \le n-k) = \sum_{i=k}^{n-k} \binom{n}{i} 2^{-n}$$

For example, if $n = 25$,

k	6	7	8	9	10	11	12
p_k	99.6%	98.6%	95.7%	89.2%	77.0%	57.6%	31.0%

The rank sum test is a non-parametric test for two independent samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$, which does not assume normality of population distributions.

Assume continuous population distributions F_1 and F_2, and consider

$$H_0\colon F_1 = F_2 \text{ against } H_1\colon F_1 \neq F_2.$$

The rank sum test procedure:

1. pool the samples and replace the data values by their ranks $1, 2, \ldots, n+m$, from the smallest value to the largest,
2. compute two test statistics $r_1 =$ sum of the ranks of x-observations, and $r_2 =$ sum of y-ranks, where $r_1 + r_2 = \frac{(n+m)(n+m+1)}{2}$,
3. use the null distribution table, which depend only on the sample sizes n and m, to find a p-value.

For $n \geq 10$, $m \geq 10$ apply approximate null distributions $\mathrm{N}(\mu_1, \sigma)$ and $\mathrm{N}(\mu_2, \sigma)$

$$\mu_1 = \frac{n(n+m+1)}{2}, \quad \mu_2 = \frac{m(n+m+1)}{2}, \quad \sigma^2 = \frac{mn(n+m+1)}{12}.$$

Two friends A and B competing on 100 m distance made independent runnings. Athlete A ran $n = 3$ times and B $m = 4$ times. All three best times were shown by A.

What is the one-sided p-value?

The rank sum test statistics

$$r_1 = 1 + 2 + 3 = 6, \quad r_2 = 4 + 5 + 6 + 7 = 22.$$

Under the null hypothesis of no difference, the event $r_1 = 6$ corresponds to drawing from an urn containing 3 black and 4 white balls.

If 3 balls are drawn without replacement, the probability that all 3 are black balls is

$$\tfrac{3}{7}\tfrac{2}{6}\tfrac{1}{5} = \tfrac{1}{35} = 0.03$$

The one-sided p-value is

$$\mathrm{P}(R_1 \le 6) = \mathrm{P}(R_1 = 6) = \mathrm{P}(\text{all three balls are black}) = 0.03$$

less than 5%.

Return to smoking and platelet aggregation example.

Before y_i	After x_i	$d_i = x_i - y_i$	Rank of $\lvert d_i \rvert$	Signed rank
25	27	2	2	+2
25	29	4	3.5	+3.5
27	37	10	6	+6
28	43	15	8.5	+8.5
30	46	16	10	+10
44	56	12	7	+7
52	61	9	5	+5
53	57	4	3.5	+3.5
53	80	27	11	+11
60	59	–1	1	–1
67	82	15	8.5	+8.5

Without assumption of normality, we can apply the sign test to

$$H_0\colon m = 0 \text{ against } H_1\colon m \neq 0$$

where m is the median of the differences. A two-sided p-value of the sign test is $2[(0.5)^{11} + 11(0.5)^{11}] = 0.012$.

Signed rank test

The sign test takes into account only the sign of the differences. The signed rank test is another non-parametric test of $H_0 : m = 0$. It requires an extra assumption: the distribution of differences is symmetric about its median m. Test statistics: either w_+ or w_-

$$w_+ = \sum_{i:d_i>0} \mathrm{rank}(|d_i|), \qquad w_- = \sum_{i:d_i<0} \mathrm{rank}(|d_i|)$$

The null distributions of W_+ and W_- are the same and tabulated for smaller values of n. For $n \geq 20$, one can use the $\mathrm{N}(\mu, \sigma)$ approximation of the null distribution with

$$\mu = \tfrac{n(n+1)}{4}, \qquad \sigma^2 = \tfrac{n(n+1)(2n+1)}{24}.$$

Example: platelet aggregation

Observed value of the test statistic $w_- = 1$. Two-sided p-value $= 0.002$. It is important to verify the extra assumption and check the symmetry of differences (d_i) around its sample median.

A nonparametric test for $I \geq 3$ independent samples each of size n, without assuming normality. Null hypothesis of no treatment effect

$$H_0: \text{ all } N = I \cdot n \text{ observations are equal in distribution.}$$

Extending the idea of the rank-sum test dealing with $I = 2$ samples, consider the pooled sample of size N.

Let r_{ik} be the pooled ranks of the sample values y_{ik}, so that

$$\sum_i \sum_k r_{ik} = 1 + 2 + \ldots + N = \tfrac{N(N+1)}{2},$$

implying that the mean rank is $\bar{r}_{..} = \frac{N+1}{2}$.

$$\boxed{\text{Kruskal-Wallis test statistic } W = \tfrac{12n}{N(N+1)} \textstyle\sum_{i=1}^{I} (\bar{r}_{i.} - \tfrac{N+1}{2})^2}$$

Reject H_0 for large W using the exact null distribution table.

An approximate null distribution $W \overset{H_0}{\approx} \chi^2_{I-1}$ is used if $N \geq 15$.

In the table below the actual measurements are replaced by their ranks $1 \div 70$. The observed Kruskal-Wallis test statistic $W = 28.17$. Using χ^2_6-distribution table we get a p-value of approximately 0.0001.

Labs	1	2	3	4	5	6	7
	70	4	35	6	46	48	38
	63	3	45	7	21	5	50
	53	65	40	13	47	22	52
	64	69	41	20	8	28	58
	59	66	57	16	14	37	68
	54	39	32	26	42	2	1
	43	44	51	17	9	31	15
	61	56	25	11	10	34	23
	67	24	29	27	33	49	60
	55	19	30	12	36	18	62
Means	58.9	38.9	38.5	15.5	26.6	27.4	42.7

Question. What is $\bar{r}_{..}$ in this case? Why df $= 6$?

Data y_{ij} for I treatments and J blocks. Friedman test is a nonparametric test for testing H_0: no treatment effect.

The Friedman test is based on within block ranking:

$$(r_{1j}, \ldots, r_{Ij}) = \text{ranks of } (y_{1j}, \ldots, y_{Ij}),$$

so that for each $j = 1, \ldots, J$,

$$r_{1j} + \ldots + r_{Ij} = 1 + 2 + \ldots + I = \tfrac{I(I+1)}{2}.$$

For these ranks, we have $\frac{1}{I}(r_{1j} + \ldots + r_{Ij}) = \frac{I+1}{2}$ and therefore $\bar{r}_{..} = \frac{I+1}{2}$.

$$\boxed{\text{Friedman test statistic } Q = \tfrac{12J}{I(I+1)} \textstyle\sum_{i=1}^{I} (\bar{r}_{i.} - \tfrac{I+1}{2})^2}$$

Test statistic Q is a measure of agreement between J rankings, so we reject H_0 for large values of Q.

An approximate null distribution $Q \overset{H_0}{\approx} \chi^2_{I-1}$.

Question. Why the Kruskal-Wallis test cannot be applied here?

Example: itching

From the rank values r_{ij} and $\bar{r}_{i.}$ given in the next table and $\frac{I+1}{2} = 4$, we find the Friedman test statistic value to be $Q = 14.86$.

Using the chi-squared distribution table with df $= 6$ we obtain the p-value is approximately 2.14%.

Subject	No Drug	Placebo	Papa	Morphine	Amin	Pent	Trip
BG	5	7	1	6	3.5	2	3.5
JF	6	5	1	2	3	7	4
BS	7	6	4	2	1	5	3
SI	6	7	1	4	3	2	5
BW	3	4	2	5	1	7	6
TS	7	3	1	5	2	4	6
GM	7	5	1	2	3	6	4
SS	1	2	5	3	7	6	4
MU	5	3	2	4	6	7	1
OS	4	7	5	2	1	3	6
$\bar{r}_{i.}$	5.10	4.90	2.30	3.50	3.05	4.90	4.25

Question. What is $\bar{r}_{..}$ in this case?