Lecture 09. Nonparametric Statistics (Chapter 14)

Ping Yu

HKU Business School
The University of Hong Kong

Plan of This Lecture

- Nonparametric tests are appropriate when the data used are qualitative data (nominal or ordinal data) or numerical data without normality assumption.

- Goodness-of-Fit Tests: Specified Probabilities
- Goodness-of-Fit Tests: Population Parameters Unknown
- Contingency Tables
  - The above three sections are about good-of-fit tests, and the following three are nonparametric counterparts of those in Lectures 6 and 8.
  - Different from the usual test, we hope the null rather than the alternative is correct, so we view accepting the null as an indication that using the null distribution is not unreasonable, without interpreting it as sufficient proof that the null distribution is the truth.
- Nonparametric Tests for Paired or Matched Samples
- Nonparametric Tests for Independent Random Samples
- The Kruskal-Wallis Test (one-way ANOVA)
- Spearman Rank Correlation

Goodness-of-Fit Tests: Specified Probabilities

Testing Procedure

- This test is also called (Karl) Pearson's chi-squared test.
- Assume the population can be partitioned into K categories.
- $H_0 : P\left(\text{Category } k\right) = p_k$, $k = 1, \cdots, K$, where p_k is known and $\sum_{k=1}^{K} p_k = 1$.

Table 14.4 Observed and Expected Numbers for n Observations and K Categories

Category	1	2	...	K	Total
Observed number	O_1	O_2	...	O_K	n
Probability (under H_0)	P_1	P_2	...	P_K	1
Expected number (under H_0)	$E_1 = nP_1$	$E_2 = nP_2$	...	$E_K = nP_K$	n

- Test Statistic: the chi-square r.v.

$$\chi^2 = \sum_{i=1}^{K} \frac{\left(O_i - E_i\right)^2}{E_i}$$

measures the goodness-of-fit of sample numbers observed to what would be expected under H_0, and follows the χ^2_{K-1} distribution under H_0 when $\min\left(\{E_i\}_{i=1}^{K}\right) \geq 5$.

- Decision Rule: reject H_0 if $\chi^2 > \chi^2_{K-1,\alpha}$.

Why the Test Statistic Takes This Form?

- Because $\sqrt{n}\left(\hat{p}_i - p_i\right) \to N\left(0, p_i\left(1-p_i\right)\right)$, the test statistic should be

$$\sum_{i=1}^{K-1}\left(\frac{\sqrt{n}\left(\hat{p}_i - p_i\right)}{\sqrt{p_i\left(1-p_i\right)}}\right)^2 = \sum_{i=1}^{K-1}\frac{n\left(\hat{p}_i - p_i\right)^2}{p_i\left(1-p_i\right)} = \sum_{i=1}^{K-1}\frac{\left(O_i - E_i\right)^2}{np_i\left(1-p_i\right)},$$

which follows the χ^2_{K-1} distribution under H_0 as n gets large, where the upper limit of summation is $K-1$ because $\hat{p}_K$ is implied by $\{\hat{p}_i\}_{i=1}^{K-1}$.

- Take $K = 2$. We will show this form of test statistic will reduce to the χ^2 r.v.:

$$\begin{aligned}\frac{\left(O_1 - E_1\right)^2}{np_1\left(1-p_1\right)} &= \frac{\left(O_1 - E_1\right)^2\left(1-p_1\right)}{np_1\left(1-p_1\right)} + \frac{\left(O_1 - E_1\right)^2 p_1}{np_1\left(1-p_1\right)} \\ &= \frac{\left(O_1 - E_1\right)^2}{E_1} + \frac{\left(O_2 - E_2\right)^2}{E_2},\end{aligned}$$

where note that $O_1 - E_1 = n - O_2 - \left(n - E_2\right) = E_2 - O_2$.

Example 14.2: Is There a Change in Customer Preferences?

- $H_0 : p_A = 0.30$, $p_B = 0.50$, $p_C = 0.15$ and $p_D = 0.05$, where the four probabilities are derived from historical preference data.
- $n = 200$, so $E_A = np_A = 60$, etc.

Table 14.6 Have Customer Preferences Changed?

Type of Candy Bar	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O - E_i)^2/E_i$
A. Mr. Goodbar	50	60	−10	100	100/60 = 1.67
B. Hershey's Milk Chocolate	93	100	−7	49	49/100 = 0.49
C. Hershey's Special Dark	45	30	15	225	225/30 = 7.50
D. Krackel	12	10	2	4	4/10 = 0.40
					$\chi^2 = 10.06$

- The p-value is between 0.01 and 0.025, so we reject the null at level 5%.

Goodness-of-Fit Tests: Population Parameters Unknown

Testing Procedure

- H_0 : the population follows a distribution (e.g., the binomial, the Poisson, or the normal distributions) with unknown parameters.
- Test Statistic:

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i},$$

  which follows the χ^2_{K-m-1} distribution under H_0, where E_i is the expected number of category i under H_0 with the unknown parameter estimated, and m is the number of unknown parameters.
- Decision Rule: reject H_0 if $\chi^2 > \chi^2_{K-m-1,\alpha}$.

A Test for the Poisson Distribution

- One application of this test is to resolve disputed authorship by counting the numbers of occurrences of particular words in blocks of text and then comparing these numbers with those whose authorship is known, where the numbers of occurrences are assumed to follow a Poisson distribution.
- Example 14.4: Federalist Papers

Table 14.7 Occurrences of the Word *may* in 262 Blocks of Text in *The Federalist Papers*

Number of Occurrences	0	1	2	3 or More
Observed frequency	156	63	29	14

- We want to test whether the population distribution of occurrences of the word *may* is Poisson.

continue

- Recall that the pmf of Poisson distribution is $p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \cdots$.
- Since λ is unknown, we estimate it by the sample mean $\hat{\lambda} = 0.66$, where recall that the mean of a Poisson distribution is λ.
- Then $p(0), p(1), p(2)$ and $p(>2) = 1 - \sum_{i=0}^{2} p(i)$ can be estimated by $p\left(0|\hat{\lambda}\right)$, $p\left(1|\hat{\lambda}\right)$, $p\left(2|\hat{\lambda}\right)$ and $p\left(>2|\hat{\lambda}\right) = 1 - \sum_{i=0}^{2} p(i|\hat{\lambda})$, and the chi-square r.v. can be constructed, where E_i's need not be rounded to integers.

Table 14.8 Observed and Expected Frequencies for *The Federalist Papers*

Number of Occurences	0	1	2	3 or More	Total
Observed frequencies	156	63	29	14	262
Probabilities	0.5169	0.3411	0.1126	0.0294	1
Expected frequencies under H_0	135.4	89.4	29.5	7.7	262

- $\chi^2 = 16.08 > 13.816 = \chi^2_{4-1-1,0.001}$, so the null is overwhelmingly rejected.

Jarque-Bera Test for Normality

- The normality assumption is very important for the tests in Lectures 6 and 8.
- We can use normal probability plots in Lecture 4 to check for evidence of nonnormality (by visually determining if the dots were "close" to the straight line).
- Here we provide a more rigorous procedure, the Jarque-Bera test [figures here], which is easy to carry out and likely to be more powerful.
- Test Statistic:

$$\text{JB} = n\left[\frac{\text{skewness}^2}{6} + \frac{(\text{kurtosis} - 3)^2}{24}\right],$$

  which follows the χ^2_2 distribution under H_0 as n gets large, where skewness and kurtosis are sample skewness and sample kurtosis, respectively.
- Decision Rule: reject H_0 if $\chi^2 > \chi^2_{2,\alpha}$.

History of the Jarque-Bera Test

Carlos M. Jarque (1954-), Mexican

Anil K. Bera (1955-), UIUC

(**) Other Tests for Normality

- Besides the Jarque-Bera test, there are many other tests for normality, such as the Kolmogorov-Smirnov test, Anderson-Darling test, Rya-Joiner test, Shapiro-Wilk test, and the Lilliefors test.
- For example, the Kolmogorov-Smirnov test is based on the empirical distribution function F_n of $\{x_i\}_{i=1}^n$, which is defined as

$$F_n(x) = \frac{1}{n}\#(x_i \le x) = \frac{1}{n}\sum_{i=1}^n 1(x_i \le x).$$ [figures here]

- By the LLN, $F_n(x)$ converges to $E[1(x_i \le x)] = F(x) = \Phi(x)$ under H_0 as $n \to \infty$.
- The Kolmogorov-Smirnov statistic is the maximal distance between F_n and Φ,

$$KS = \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)|.$$

- We reject the null when KS is large; it is good to know that the distribution of KS is the same for every continuous null distribution, e.g. Φ here, and is available in $\mathbb{R}$.
- For large n, under H_0,

$$\lim_{n\to\infty} P\left(\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| > z/\sqrt{n}\right) = 2\sum_{j=1}^{\infty}(-1)^{j+1}e^{-j^2z^2},$$

which can be used to obtain the critical value or p-value.

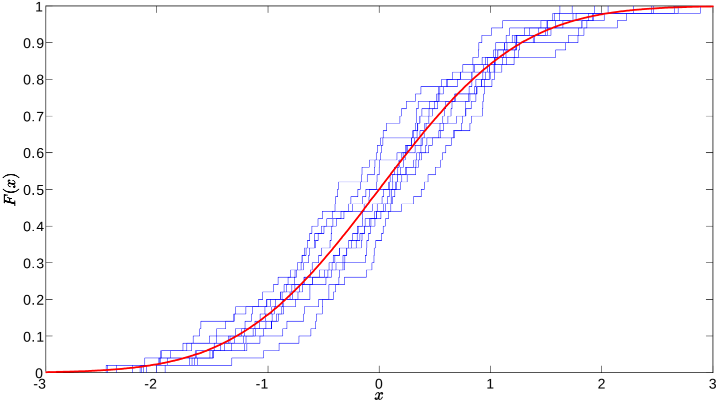


Figure: Empirical Distribution Functions: 10 samples from $N(0,1)$ with sample size $n = 50$

continue

- As in the last section, if the null distribution of F depends on some unknown parameters, e.g., μ and σ^2 in the normal distribution, we can extend the Kolmogorov-Smirnov statistic to

$$KS^* = \sup_{x \in \mathbb{R}} \left| F_n(x) - \Phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right) \right|$$

  for an estimator of $\left(\hat{\mu}, \hat{\sigma}^2\right)$ for $\left(\mu, \sigma^2\right)$, e.g., $\hat{\mu} = \bar{x}$, and $\hat{\sigma}^2 = s^2$.
- The null distribution of KS^* is different from that of KS, depending on the null distribution, the estimator $\left(\hat{\mu}, \hat{\sigma}^2\right)$ used, and even on the true $\left(\mu, \sigma^2\right)$.
- The null distribution of KS^* has been tabulated for a few special cases including the normality case, but it is more convenient to simulate the critical value or p-value.
- Specifically, we simulate n samples from $N\left(\hat{\mu}, \hat{\sigma}^2\right)$, and then construct $F_n(x)$ and KS^*; such a procedure can be repeated for (a large) B times to approximate the null distribution of KS^*.

Contingency Tables

Testing Association in Contingency Tables[1]

Table 14.10 Cross-Classification of n Observations in an $r \times c$ Contingency Table

	CHARACTERISTIC B				
Characteristic A	1	2	...	c	Total
1	O_{11}	O_{12}	...	O_{1c}	R_1
2	O_{21}	O_{22}	...	O_{2c}	R_2
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
r	O_{r1}	O_{r2}	...	O_{rc}	R_r
Total	C_1	C_2	...	C_c	n

- H_0: there is no association between characteristics A and B, or A an B are independent.
- Under H_0, the distribution of R_i among c columns should be same as the total number n among the c columns, so the estimated expected number of observations at cell (i,j) is $E_{ij} = R_i \frac{C_j}{n}$ for $i = 1, \cdots, r$ and $j = 1, \cdots, c$, where $\frac{C_j}{n}$ is the proportion of column i in n.

[1] The term contingency table was first used by Karl Pearson.

continue

- Test Statistic:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}},$$

which follows the $\chi^2_{(r-1)(c-1)}$ distribution under H_0 if no more than 20% of E_{ij} is less than 5,[2] where

$$E_{ij} = \frac{R_i C_i}{n} \text{ for } i = 1, \cdots, r \text{ and } j = 1, \cdots, c$$

are the estimated expected numbers in the cross-classification.
- The df of χ^2 should be $rc - 1 - (r-1) - (c-1) = (r-1)(c-1)$, where $(r-1) + (c-1)$ parameters are implicitly estimated.

- Decision Rule: reject H_0 if $\chi^2 > \chi^2_{(r-1)(c-1),\alpha}$.

[2] Sometimes, adjacent classes can be combined to meet this assumption.

Example 14.6: Market Differentiation

- Consumers are exposed to different products ($r = 3$) and asked what comes to their mind ($c = 2$) when they see or hear of this product.

Table 14.11 Automobile by Consumer Perception

AUTOMOBILE	SPORTY	SAFETY	TOTAL
BMW	256	74	330
Mercedes	41	42	83
Lexus	66	34	100
Total	**363**	**150**	**513**

Table 14.12 Observed (and Expected) Number of Customers in Each Cross-Classification

AUTOMOBILE	SPORTY	SAFETY	TOTAL
BMW	256 (233.5)	74 (96.5)	330
Mercedes	41 (58.7)	42 (24.3)	83
Lexus	66 (70.8)	34 (29.2)	100

- $\chi^2 = \frac{(256-233.5)^2}{233.5} + \cdots = 26.8 > \chi^2_{2,0.001} = 13.816$, so we can conclude that the market is indeed differentiated.

Nonparametric Tests for Paired or Matched Samples

Sign Test

- Data: $\{(x_i, y_i)\}_{i=1}^n$ with $x_i = y_i$ discarded or $\{\text{sign}(d_i)\}_{i=1}^n$, where x_i and y_i may be nominal, and $d_i = x_i - y_i$, e.g., preferences of a product, and d_i is only known to be positive or negative.
- $H_0 : p = 0.5$, where $p = P(\text{sign}(D) = +) = P(X > Y)$, where D, X and Y are the r.v.'s generating d_i, x_i and y_i, respectively.
- Test Statistic:

$$S = \text{the number of pairs wih a positive difference},$$

  which follows the Binomial$(n, 0.5)$ distribution under H_0.
- Decision Rule: let $B \sim$Binomial$(n, 0.5)$.
  - If $H_1 : p > 0.5$, reject H_0 if the p-value = $P(B \geq S) < \alpha$.
  - If $H_1 : p < 0.5$, reject H_0 if the p-value = $P(B \leq S) < \alpha$.
  - If $H_1 : p \neq 0.5$, let $S = \max(S_+, S_-)$, and reject H_0 if the p-value = $2P(B \geq S) < \alpha$, where $S_\pm = \#(\text{sign}(d_i) = \pm)$.
- This test can also be used to test whether the median (not mean) of a r.v. X (i.e., one sample) is equal to a specified value μ, where $x_i - \mu$ plays the role of d_i.

Example 14.8: Product Preference

- A pizza store wants to know whether a new pizza sauce should be adopted. Eight students are asked to rate the two sauces on a scale of 1 to 10 (a higher number indicating a greater liking).

Table 14.14 Student Ratings for Pizza Sauce

	Rating			
Student	Original Pizza Sauce	New Pizza Sauce	Difference (Original − New)	Sign of Difference
A	6	8	−2	−
B	4	9	−5	−
C	5	4	1	+
D	8	7	1	+
E	3	9	−6	−
F	6	9	−3	−
G	7	7	0	0
H	5	9	−4	−

- H_1: $p < 0.5$. The p-value$= P(B \leq 2) = 0.227$, where $B \sim$Binomial$(8-1, 0.5)$. So a change in the pizza sauce is not recommended.

Wilcoxon Signed Rank Test

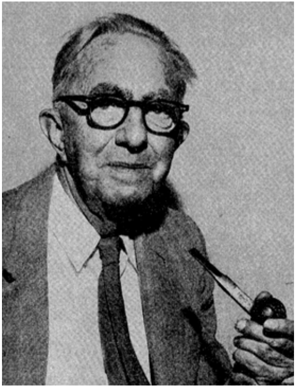
- The sign test uses only a very limited amount of information, namely, which product is preferred, but ignores the strengths of the preferences, so may not be powerful when n is small.
- The Wilcoxon signed rank test [figures here] uses also the magnitude of the differences (ranks) besides their signs (so d_i is at least ordinal).
- Assumption: d_i is symmetrically distributed.
- H_0: $\text{median}(D) = 0$.[3]
- Test Statistic:

$$W = \sum_{i=1}^{n} \text{sign}(d_i) r_i = T_+ - T_-,$$

where as in the sign test, d_i's with $d_i = 0$ are discarded, r_i is the rank of $|d_i|$ when $|d_i|$, $i = 1, \cdots, n$, are sorted in ascending order, with ties assigned the average of the ranks they occupy, $T_+ = \sum_{i=1}^{n} r_i 1(d_i > 0)$ and $T_- = \sum_{i=1}^{n} r_i 1(d_i < 0)$.
- An equivalent test statistic in the textbook is $T = \min(T_+, T_-)$.

[3] When the matched pair x_i and y_i are drawn from populations having the same distribution, then d_i is symmetrically distributed with center at 0, so this test can also be used to test whether x_i and y_i are drawn from the same distribution.

History of Wilcoxon Tests



Frank Wilcoxon (1892-1965), Irish American[4]

- He contributed two tests, the Wilcoxon signed-rank test and the Wilcoxon rank-sum test, which are nonparametric alternatives to the paired and unpaired Student's t-tests, respectively.

[4] He is also a chemist besides a statistician.

Example 14.9: Product Preference

- Decision Rule: $T < c_{n,\alpha}$, where the critical value depends on n and α is available from Appendix Table 10.
  - An advantage of T over W is that the decision rule of W depends on H_1 being one-sided or two-sided but T does not [if two sided, $\alpha \to \alpha/2$].

Table 14.15 Calculation of Wilcoxon Test Statistic for Taste Preference Data

Taster	Difference	Rank (+)	Rank (−)
A	−2		3
B	−5		6
C	1	1.5	
D	1	1.5	
E	−6		7
F	−3		4
G	0		
H	−4		5
		Rank sum 3	**25**

Wilcoxon signed rank statistic T = minimum (3, 25) = 3

- $T = 3 < 4 = c_{7,0.05}$ (G is discarded, so $n = 7$; implicitly, H_1: median$(D) < 0$), so we reject the null and draw a different conclusion from the sign test that the new sauce should be adopted.

Normal Approximation to the Sign Test

- When $n > 20$,

$$Z = \frac{S^* - \mu}{\sigma} = \frac{S^* - 0.5n}{0.5\sqrt{n}}$$

approximately follows $N(0,1)$, where under H_0, $\mu = np = 0.5n$, $\sigma = \sqrt{np(1-p)} = 0.5\sqrt{n}$, and S^* is the test statistic corrected for continuity:

$$\begin{aligned}
\text{if } H_1 &: \quad p > 0.5, \text{ then } S^* = S - 0.5; \\
\text{if } H_1 &: \quad p < 0.5, \text{ then } S^* = S + 0.5; \\
\text{if } H_1 &: \quad p \neq 0.5, \text{ then } S^* = S + 0.5 \text{ if } S < \mu \text{ or } S^* = S - 0.5 \text{ if } S > \mu.
\end{aligned}$$

- The continuity correct factor in S^* compensates for estimating discrete data with a continuous distribution and provide a closer approximation to the p-value.

Normal Approximation to the Wilcoxon Signed Rank Test

- When $n > 20$,

$$Z = \frac{T - \mu_T}{\sigma_T}$$

approximately follows $N(0,1)$, where under H_0,

$$\begin{aligned} \mu_T &= E[T] = \frac{n(n+1)}{4}, \\ \sigma_T^2 &= Var(T) = \frac{n(n+1)(2n+1)}{24}. \end{aligned}$$

- It can be shown that $\mu_W = 0$ and $\sigma_W^2 = \frac{n(n+1)(2n+1)}{6}$ under H_0, and the corresponding normal approximation for W can be applied.

- Decision Rule:
  - If H_1 is one-sided, no matter median$(D) > 0$ or < 0, reject H_0 if $Z < -z_\alpha$.
  - If H_1 is two-sided, i.e., median$(D) \neq 0$, reject H_0 if $Z < -z_{\alpha/2}$.

Nonparametric Tests for Independent Random Samples

Mann-Whitney U Test

- The Mann-Whitney test [figures here] is a nonparametric test based on two independent samples $\{x_i\}_{i=1}^{n_1}$ and $\{y_j\}_{j=1}^{n_2}$.
- H_0: X and Y have the same distribution vs. H_1: X and Y do not have the same distribution, where X and Y are the population r.v.'s generating x_i and y_j.
  - When the scale and shape of the distributions of X and Y are the same, H_0 reduces to test whether the medians of X and Y are the same or not, which corresponds to $\mu_x - \mu_y = 0$ in Lecture 6 where $x_i \sim N\left(\mu_x, \sigma^2\right)$, $y_j \sim N\left(\mu_y, \sigma^2\right)$.
- The basic idea of the Mann-Whitney test is that under H_0, $P(X < Y) = P(Y < X)$. So define

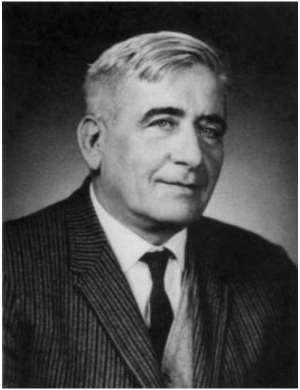
$$S(X,Y) = \begin{cases} 1, & \text{if } Y < X, \\ 1/2, & \text{if } Y = X, \\ 0, & \text{if } Y > X, \end{cases}$$

and the test statistic is

$$U_1 = \sum_{i=1}^{n_1} \sum_{i=1}^{n_2} S\left(x_i, y_j\right),$$

which is a U-statistic, so the name "U test".

History of the Mann-Whitney Test



Henry B. Mann (1905-2000), OSU

Donald R. Whitney (1915-2007), OSU[5]

[5]He was a student of Mann at OSU.

Two Alternative Forms of the Mann-Whitney Test

- The form in the textbook: $U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$, where we pool the two samples and rank the observations in ascending order with ties assigned the average of the next available ranks (or adjacent ranks), and R_1 is the sum of the ranks of $\{x_i\}_{i=1}^{n_1}$.
  - It can be shown that $U_1 = R_1 - \frac{n_1(n_1+1)}{2}$, where $\frac{n_1(n_1+1)}{2}$ is R_1 in the worst scenario.
  - We can parallelly define $U_2 = R_2 - \frac{n_2(n_2+1)}{2}$, where R_2 is the sum of the ranks of $\{y_j\}_{j=1}^{n_2}$.
  - Because $R_1 + R_2 = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$, where $n = n_1 + n_2$, we have $U_1 + U_2 = n_1 n_2$, so $U = U_2$.
  - It can be shown that for both U_1 and U_2, the mean is $\mu_U = \frac{n_1 n_2}{2} (> 0)$ and the variance is $\sigma_U^2 = \frac{n_1 n_2 (n+1)}{12}$, so
  $$z = \frac{U - \mu_U}{\sigma_U}$$
  is approximately $N(0,1)$ when $\min(n_1, n_2) \geq 10$, and the p-value is $P(|Z| > |z|)$ with $Z \sim N(0,1)$.
  - In practice, the smaller of U_1 and U_2 is used to compare with the critical value.

continue

- Wilcoxon Rank Sum Test: $T = R_1$.
  - From the mean and variance of U, we can see

$$\begin{aligned}\mu_T &= \frac{n_1(n_1+1)}{2} + \frac{n_1 n_2}{2} = \frac{n_1(n+1)}{2}, \\ \sigma_T^2 &= \sigma_U^2 = \frac{n_1 n_2 (n+1)}{12};\end{aligned}$$

  therefore,

$$z = \frac{U - \mu_T}{\sigma_T}$$

  is approximately $N(0,1)$ when $\min(n_1, n_2) \geq 10$.
  - The formulae of σ_T^2 and σ_U^2 need to be corrected when there are a large number of ties.
- (*) Denote r_i, $i = 1, \cdots, n_1$, as the ranks of x_i, then $R_1 = \sum_{i=1}^{n_1} r_i$. Under H_0, $x_1, \cdots, x_{n_1}, y_1, \cdots, y_{n_2}$ can be viewed as a sample of size $n_1 + n_2$ from a fixed (unknown) distribution. The ranks $r_1, \cdots, r_{n_1}$, can then be viewed as an arbitrary selection of n_1 numbers out of the numbers $\{1, 2, \cdots, n_1 + n_2\}$. The distribution of R_1 under H_0 is therefore independent of this unknown distribution, and can be determined using combinatorial arguments. This distribution has been tabulated and is available through $\mathbb{R}$.

The Kruskal-Wallis Test

(Section 15.3)

The Kruskal-Wallis H Test

- This Kruskal-Wallis H test [figure here] is a nonparametric counterpart of the one-way ANOVA where normality is not assumed; it is also an extension of the Mann-Whitney test to the $K > 2$ case.
- The null is that all subgroups have the same distribution which will reduce to the same median when all other aspects of the K distributions except the central location are the same.
- Like the Mann-Whitney test, we pool all samples and rank them in ascending order with the rank of x_{ij} being r_{ij}, and define

$$R_i = \sum_{j=1}^{n_i} r_{ij}, i = 1, \cdots, K,$$

as the sum of the ranks for subgroup i.
- Also define

$$\bar{r}_i = \frac{R_i}{n_i}$$

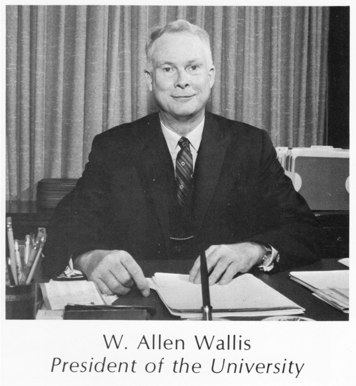
and

$$\bar{\bar{r}} = \frac{1}{n} \sum_{i=1}^{K} \sum_{j=1}^{n_i} r_{ij},$$

which are the counterparts of $\bar{x}_i$ and $\bar{\bar{x}}$.

History of the Kruskal-Wallis Test

William H. Kruskal (1919-2005), Chicago



Wilson A. Wallis (1912-1998), Rochester

continue

- The test statistic is

$$H=(n-1)\frac{\sum_{i=1}^{K}n_i\left(\bar{r}_i-\bar{\bar{r}}\right)^2}{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(r_{ij}-\bar{\bar{r}}\right)^2}:=(n-1)\frac{SSG}{SST}.$$

  - Since $SST=SSG+SSW$, H is an increasing function of SSG and a large H will induce rejection of H_0,
- When there are no ties, it is not hard to show that the denominator of H is equal to $\frac{(n-1)n(n+1)}{2}$ and $\bar{\bar{r}}=\frac{n+1}{2}$, which implies the W on Page 663:

$$H=\frac{12}{n(n+1)}\sum_{i=1}^{K}\frac{R_i^2}{n_i}-3(n+1).$$

- Under H_0, $\frac{SST}{n-1}\to\sigma^2$ and $\frac{SSG}{\sigma^2}\to\chi^2_{K-1}$, where σ^2 is the variance of r_{ij}, so H approximately follow the χ^2_{K-1} distribution. As a result, the decision rule is to reject H_0 if $H>\chi^2_{K-1,\alpha}$.
- (**) The counterpart of MSD is Dunn's test which can be used to detect which of the sample pairs are different.

Example Continued

Table 15.6 Fuel-Consumption Figures (in Miles per Gallon) and Ranks from Three Independent Random Samples

A-Cars	Rank	B-Cars	Rank	C-Cars	Rank
22.2	11	24.6	20	22.7	12
19.9	1	23.1	13	21.9	7
20.3	2.5	22.0	8	23.2	14
21.4	6	23.5	16.5	24.1	19
21.2	5	23.6	18	22.1	9.5
21.0	4	22.1	9.5	23.4	15
20.3	2.5	23.5	16.5		
Rank sum	**32**		**101.5**		**76.5**

- $R_1 = 32, R_2 = 101.5$ and $R_3 = 76.5$, so

$$H = \frac{12}{20 \times 21}\left[\frac{32^2}{7} + \frac{101.5^2}{7} + \frac{76.5^2}{6}\right] - 3 \times 21 = 11.10.$$

Since $\chi^2_{2,0.01} = 9.210$, we reject H_0 at the 1% level, same conclusion as the one-way ANOVA.

Spearman Rank Correlation

Testing No Correlation Between X and Y [Section 11.7]

- Assumption: X and Y are jointly normally distributed, i.e.,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\mu, \Sigma),$$

where

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

- $H_0 : \rho = 0$.
- Test Statistic:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

which follows the t_{n-2} distribution under H_0.

- If $H_1 : \rho > 0$, then the decision rule: reject H_0 if $t > t_{n-2,\alpha}$.
- If $H_1 : \rho < 0$, then the decision rule: reject H_0 if $t < t_{n-2,\alpha}$.
- If $H_1 : \rho \neq 0$, then the decision rule: reject H_0 if $|t| > t_{n-2,\alpha/2}$.
- Rule of Thumb: set $t_{n-2,\alpha/2} = 2$, then $|t| > t_{n-2,\alpha/2}$ is approximately $|r| > \frac{2}{\sqrt{n}}$.

Spearman Rank Correlation

- The above test suffers from two problems: (i) it is a parametric test depending on the normality assumption; (ii) it is based on r so can be seriously affected by odd extreme observations.
- The test based on the Spearman rank correlation coefficient [figures here] can avoid these two problems.
- For a random sample $\{(x_i, y_i)\}_{i=1}^n$, define the ranks of $\{x_i\}_{i=1}^n$ as $\{r_{xi}\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ as $\{r_{yi}\}_{i=1}^n$. Then the Spearman rank correlation coefficient r_s is the sample correlation between $\{r_{xi}\}_{i=1}^n$ and $\{r_{yi}\}_{i=1}^n$. If neither $\{r_{xi}\}_{i=1}^n$ nor $\{r_{yi}\}_{i=1}^n$ contain tied ranks, then
$$r_s = 1 - \frac{6\sum_{i=1}^n d_i^2}{n\left(n^2-1\right)},$$
where $d_i = r_{xi} - r_{yi}$.
- It assesses how well the relationship between two random variables can be described using a monotone function. [figures here]

History of the Spearman Rank Correlation

Charles E. Spearman (1863-1945), UCL[6]

[6]He is a pioneer of factor analysis. His statistical work was not appreciated by his University College colleague Karl Pearson and there was a long feud between them.

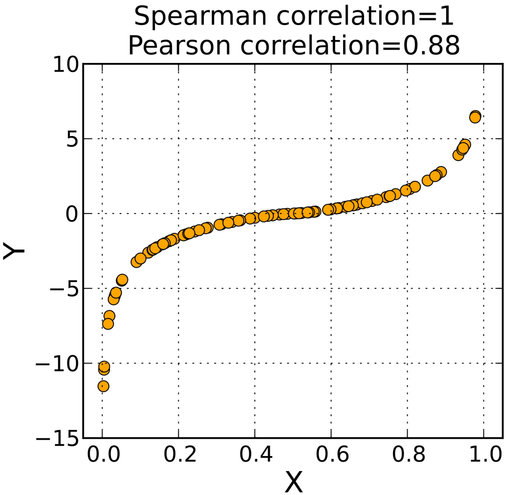
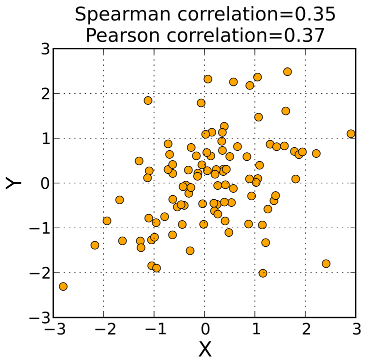
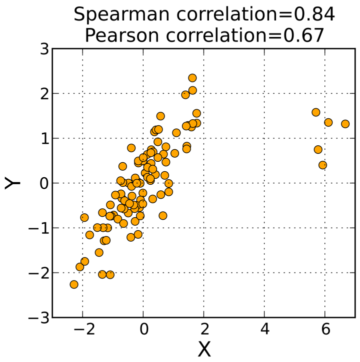


Figure: A Spearman correlation of 1 results when the two variables being compared are monotonically related, even if their relationship is not linear.





When the data are roughly elliptically distributed and there are no prominent outliers, the Spearman correlation and Pearson correlation give similar values.

The Spearman correlation is less sensitive than the Pearson correlation to strong outliers that are in the tails of both samples.

- The robustness of the Spearman correlation to outliers in the right figure is because it limits the outlier to the value of its rank.

Testing No Association Between X and Y

- If the population Spearman rank correlation coefficient is denoted as ρ_s, then

$$H_0 : \rho_s = 0.$$

- Decision Rule:
  - If $H_1 : \rho_s > 0$, reject H_0 if $r_s > r_{s,\alpha}$.
  - If $H_1 : \rho_s < 0$, reject H_0 if $r_s < r_{s,\alpha}$.
  - If $H_1 : \rho_s \neq 0$, reject H_0 if $|r_s| > r_{s,\alpha/2}$.
- The critical values can be found from Appendix Table 11.

Summary: Correspondence Between Parametric Tests and Nonparametric Tests

Parametric Tests	Nonparametric Tests
one normal mean : $\mu = \mu_0$ with $x_i \sim N(\mu, \sigma^2)$ and σ^2 known or unknown	goodness-of-fit test: $F = F_0$ with F_0 known or $F_0 = F_\theta$ - Jarque-Bera test when F_0 is normality
matched pair: $\mu_d = 0$ with $d_i \sim N(\mu_d, \sigma_d^2)$ and σ_d^2 unknown	1. sign test: $P(\text{sign}(D) = +) = 0.5$ or $\text{med}(D) = 0$ 2. Wilcoxon signed rank test: $\text{med}(D) = 0$ with D being symmetric
independent samples: $\mu_d = 0$ with $x_i \sim N(\mu_x, \sigma^2)$, $y_j \sim N(\mu_y, \sigma^2)$ and σ^2 unknown	Mann-Whitney U test or Wilcoxon rank sum test: $F_X = F_Y$
one-way ANOVA: $\mu_1 = \cdots = \mu_K$ with $x_{ij} \sim N(\mu_i, \sigma^2)$ and σ^2 unknown	Kruskal-Wallis H test: $F_1 = \cdots = F_K$
no correlation test: $\rho = 0$ with (x_i, y_i) jointly normal	1. contigency table: discrete X and Y are independent 2. no Spearman rank correlation: $\rho_s = 0$