Categorical and discrete data. Non-parametric tests

Dr. Hemal Pandya

Learning Objectives

- Develop the need for inferential techniques that require fewer, or less stringent, assumptions than the methods of earlier chapters
- Introduce nonparametric tests that are based on ranks (i.e., on an ordering of the sample measurements according to their relative magnitudes)

Non-parametric tests:

estimate sample differences when the known distribution shapes cannot help, or even confuse

Parametric Vs Non-parametric Tests

Parametric Tests:

- The population mean (μ), standard deviation (s) and proportion (p) are called the parameters of a distribution.
- Tests of hypotheses concerning the mean and proportion are based on the assumption that the population(s) from where the sample is drawn is normally distributed.
- Tests based on the above parameters are called parametric tests.

Parametric Vs. Non-parametric Tests

Non-Parametric Tests:-

- There are situations where the populations under study are not normally distributed. The data collected from these populations is extremely skewed. Therefore, the parametric tests are not valid.
- The option is to use a non-parametric test. These tests are called the distribution-free tests as they do not require any assumption regarding the shape of the population distribution from where the sample is drawn.
- These tests could also be used for the small sample sizes where the normality assumption does not hold true.

Advantages of Non-Parametric Tests

- They can be applied to many situations as they do not have the rigid requirements of their parametric counterparts, like the sample having been drawn from the population following a normal distribution.

- There can be applications where a numeric observation is difficult to obtain but a rank value is not. By using ranks, it is possible to relax the assumptions regarding the underlying populations.

- Non-parametric tests can often be applied to the nominal and ordinal data that lack exact or comparable numerical values.

- Non-parametric tests involve very simple computations compared to the corresponding parametric tests.

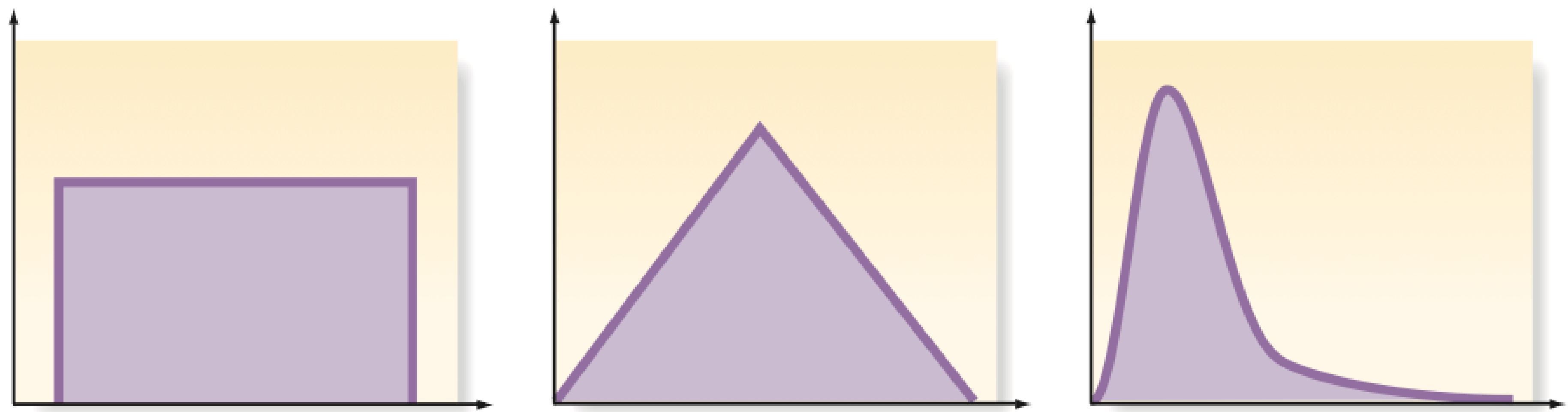
Disadvantages of Non-Parametric Tests

- A lot of information is wasted because the exact numerical data is reduced to a qualitative form. The increase or the gain is denoted by a plus sign whereas a decrease or loss is denoted by a negative sign. No consideration is given to the quantity of the gain or loss.
- Non-parametric methods are less powerful than parametric tests when the basic assumptions of parametric tests are valid.
- Null hypothesis in a non-parametric test is loosely defined as compared to the parametric tests. Therefore, whenever the null hypothesis is rejected, a non-parametric test yields a less precise conclusion as compared to the parametric test.

Difference between Parametric & Non-parametric Tests

	Parametric Tests	Non-Parametric Tests
Assumptions:	Normality assumption is required.	Normality assumption is not required.
	Uses the metric data.	Ordinal or interval scale data is used.
	Can be applied for both small and large samples.	Can be applied for small samples.
Applications:	One sample using Z or t statistics.	One sample using the sign test.
	Two independent samples using a t or z test.	Two independent samples using the Mann-Whitney U statistics.
	Two paired samples using a t or z test.	Two paired samples using the sign test and Wilcoxon matched pair rank test.
	Randomness – no test in parametric is available.	Randomness – using runs test.
	Several independent samples using F test in ANOVA.	Several independent samples using Kruskal–Wallis test.

Nonnormal Distributions - *t*-Statistic is Invalid



a. Flat distribution b. Peaked distribution c. Skewed distribution

Distribution-Free Tests

Distribution-free tests are statistical tests that do not rely on any underlying assumptions about the probability distribution of the sampled population.

The branch of inferential statistics devoted to distribution-free tests is called **nonparametrics**.

Nonparametric statistics (or tests) based on the ranks of measurements are called **rank statistics** (or **rank tests**).

Parametric and nonparametric tests of significance

	Nonparametric tests		Parametric tests
	Nominal data	*Ordinal data*	*Ordinal, interval, ratio data*
One group	Chi square goodness of fit	Wilcoxon signed rank test	**One group t-test**
Two unrelated groups	Chi square	Wilcoxon rank sum test, Mann-Whitney test	**Student's t-test**
Two related groups	McNemar's test	Wilcoxon signed rank test	**Paired Student's t-test**
K-unrelated groups	Chi square test	Kruskal -Wallis one way analysis of variance	**ANOVA**
K-related groups		Friedman matched samples	**ANOVA with repeated measurements**

14 Nonparametric Methods and Chi-Square Tests (1)

- Using Statistics
- The Sign Test
- The Runs Test - A Test for Randomness
- The Mann-Whitney *U* Test
- The Wilcoxon Signed-Rank Test
- The Kruskal-Wallis Test - A Nonparametric Alternative to One-Way ANOVA

14 Nonparametric Methods and Chi-Square Tests (2)

- The Friedman Test for a Randomized Block Design
- The Spearman Rank Correlation Coefficient
- A Chi-Square Test for Goodness of Fit
- Contingency Table Analysis - A Chi-Square Test for Independence
- A Chi-Square Test for Equality of Proportions

14 LEARNING OBJECTIVES

After reading this chapter you should be able to:

- Differentiate between parametric and nonparametric tests
- Conduct a sign test to compare population means
- Conduct a runs test to detect abnormal sequences
- Conduct a Mann-Whitney test for comparing population distributions
- Conduct a Wilkinson’s test for paired differences

14 LEARNING OBJECTIVES (2)

After reading this chapter you should be able to:

- Conduct a Friedman's test for randomized block designs
- Compute Spearman's Rank Correlation Coefficient for ordinal data
- Conduct a chi-square test for goodness-of-fit
- Conduct a chi-square test for independence
- Conduct a chi-square test for equality of proportions

14-1 Using Statistics (Parametric Tests)

- Parametric Methods
  - ✓Inferences based on assumptions about the nature of the population distribution
    - Usually: population is normal
  - ✓Types of tests
    - z-test or t-test
      - » Comparing two population means or proportions
      - » Testing value of population mean or proportion
    - ANOVA
      - » Testing equality of several population means

Nonparametric Tests

- Nonparametric Tests
  - ✓ Distribution-free methods making no assumptions about the population distribution
  - ✓ Types of tests
    - Sign tests
      - » Sign Test: Comparing paired observations
      - » McNemar Test: Comparing qualitative variables
      - » Cox and Stuart Test: Detecting trend
    - Runs tests
      - » Runs Test: Detecting randomness
      - » Wald-Wolfowitz Test: Comparing two distributions

Nonparametric Tests (Continued)

- Nonparametric Tests
  - ✓ Ranks tests
    - Mann-Whitney U Test: Comparing two populations
    - Wilcoxon Signed-Rank Test: Paired comparisons
    - Comparing several populations: ANOVA with ranks
      - » Kruskal-Wallis Test
      - » Friedman Test: Repeated measures
  - ✓ Spearman Rank Correlation Coefficient
  - ✓ Chi-Square Tests
    - Goodness of Fit
    - Testing for independence: Contingency Table Analysis
    - Equality of Proportions

Nonparametric Tests (Continued)

- Deal with ***enumerative*** (frequency counts) data.
- Do not deal with specific population parameters, such as the mean or standard deviation.
- Do not require assumptions about specific population distributions (in particular, the normality assumption).

Key Ideas

Distribution-Free Tests

Do not rely on assumptions about the probability distribution of the sampled population

Key Ideas

Nonparametrics

Distribution-free tests that are based on **rank statistics**

One-sample nonparametric test for the population median – **sign test**

Nonparametric test for *two independent samples* – **Wilcoxon rank sum test**

Nonparametric test for *matched pairs* – **Wilcoxon signed rank test**

Key Ideas

Nonparametrics

Nonparametric test for a completely randomized design – **Kruskal-Wallis test**

Nonparametric test for a randomized block design – **Friedman test**

Nonparametric test for rank correlation – **Spearman's test**

Nonparametric Methods

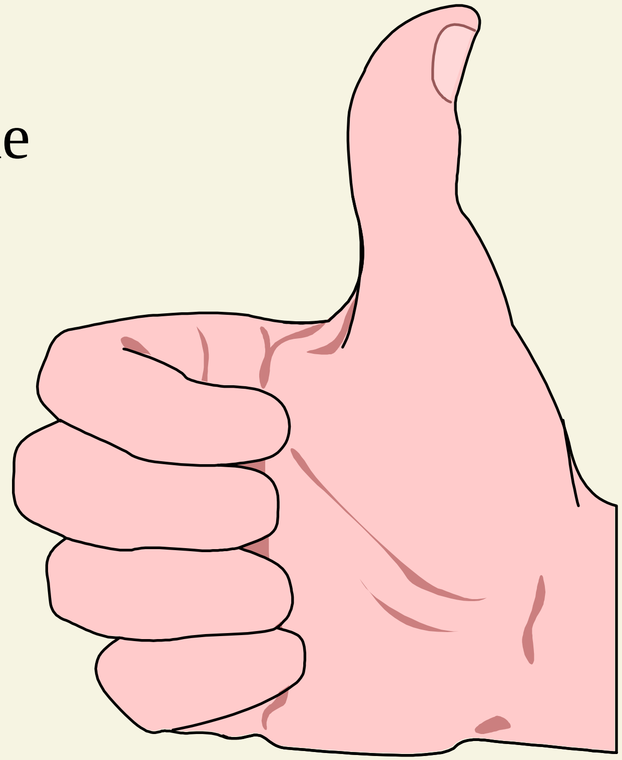
- Most of the statistical methods referred to as parametric require the use of interval- or ratio-scaled data.
- Nonparametric methods are often the only way to analyze nominal or ordinal data and draw statistical conclusions.
- Nonparametric methods require no assumptions about the population probability distributions.
- Nonparametric methods are often called distribution-free methods.

Nonparametric Methods

- In general, for a statistical method to be classified as nonparametric, it must satisfy at least one of the following conditions.
  - The method can be used with nominal data.
  - The method can be used with ordinal data.
  - The method can be used with interval or ratio data when no assumption can be made about the population probability distribution.

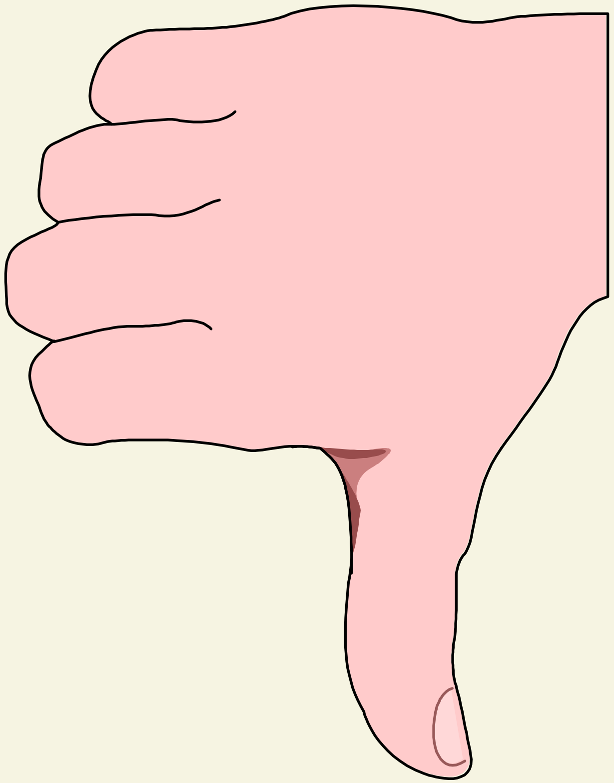
Advantages of Nonparametric Tests

- Used with all scales
- Easier to compute
  - Developed originally before wide computer use
- Make fewer assumptions
- Need not involve population parameters
- Results may be as exact as parametric procedures



Disadvantages of Nonparametric Tests

- May waste information
  - If data permit using parametric procedures
  - Example: converting data from ratio to ordinal scale
- Difficult to compute by hand for large samples
- Tables not widely available

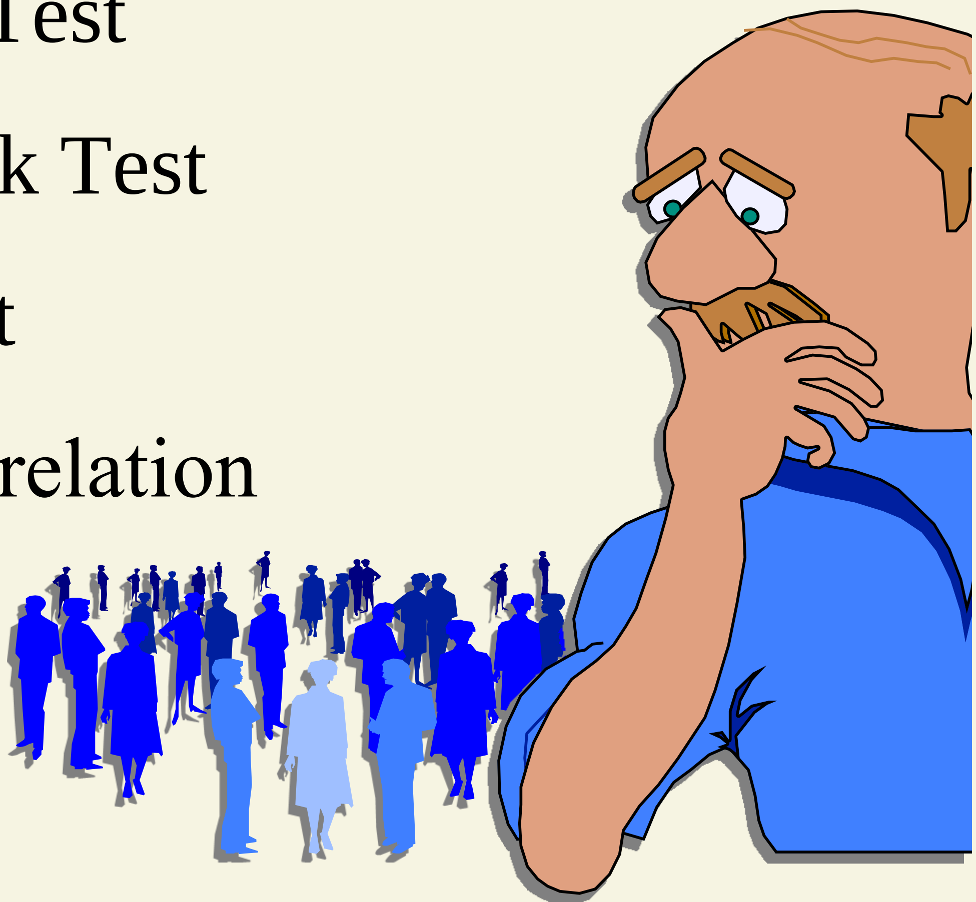


Non-parametric Pros and Cons

- Advantages of non-parametric tests
  - Shape of the underlying distribution is irrelevant - does not have to be normal
  - Large outliers have no effect
  - Can be used with data of ordinal quality
- Disadvantages
  - Less Power - less likely to reject H_0
  - Reduced analytical sophistication. With nonparametric tests there are not as many options available for analysing your data
  - Inappropriate to use with lots of tied ranks

Frequently Used Nonparametric Tests

- Sign Test
- Wilcoxon Rank Sum Test
- Wilcoxon Signed Rank Test
- Kruskal Wallis *H*-Test
- Spearman's Rank Correlation Coefficient



Single Population Inferences

Sign Test

- A common application of the sign test involves using a sample of *n* potential customers to identify a preference for one of two brands of a product.
- The objective is to determine whether there is a difference in preference between the two items being compared.
- To record the preference data, we use a plus sign if the individual prefers one brand and a minus sign if the individual prefers the other brand.
- Because the data are recorded as plus and minus signs, this test is called the sign test.

Sign Test

- Tests one population median, η (eta)
- Corresponds to t-test for one mean
- Assumes population is continuous
- Small sample test statistic: Number of sample values above (or below) median
    - Alternative hypothesis determines
- Can use normal approximation if $n \geq 30$

One-Sample Sign Test

- The test on mean discussed in Chapter 12 is based upon the assumption that the samples are drawn from a population having roughly the shape of a normal distribution.
- This assumption gets violated, especially while using the non-metric data (ordinal or nominal).
- In such situations, the standard tests can be replaced by a non-parametric test.
- One such test is called one-sample sign test.

One-Sample Sign Test

- Suppose the interest is in testing the null hypothesis $H_0 : \mu = \mu_0$ against a suitable alternative hypothesis.
- Let n denote the size of sample for any problem. To conduct a sign test, each sample observation greater than μ_0 is replaced by a plus sign, whereas each value less than μ_0 is replaced by a minus sign.
- In case a sample observation equals μ_0, it is omitted and the size of the sample gets reduced accordingly.

One Sample Sign Test

- Testing the given null hypothesis is equivalent to testing that these plus and minus signs are the values of a random variable having a binomial distribution with p = ½. For a large sample, z test as given below is used:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{npq}}$$

Where, μ = mean of binomial distribution = np
σ = standard deviation of binomial distribution = $\sqrt{npq}$

As the binomial distribution is a discrete one whereas the normal distribution is a continuous distribution, a correction for continuity is to be made. For this, X is decreased by 0.5 if X > np and increased by 0.5 if X < np.

One-Sample Sign Test

As under the null hypothesis, p = ½, therefore

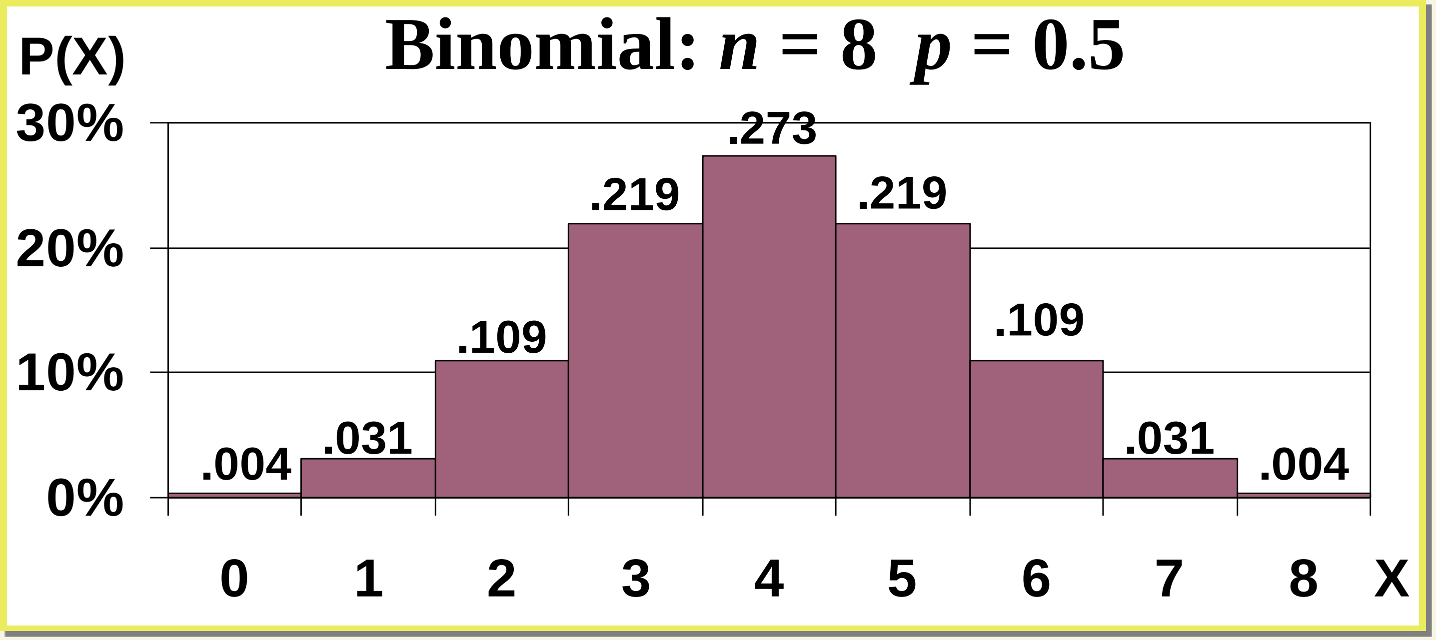
$$\mu = np = \frac{n}{2} = 0.5\,n \text{ and } \sigma = \sqrt{npq} = \sqrt{n \times \frac{1}{2} \times \frac{1}{2}} = \frac{\sqrt{n}}{2} = 0.5\sqrt{n}\,.$$

For a given level of significance, the absolute value of computed Z is compared with absolute value of tabulated Z to accept or reject the null hypothesis.

Two-Sample Sign Test

- This test is a non-parametric version of paired-sample t-test.
- It is based upon the sign of a pair of observations.
- Suppose a sample of respondents is selected and their views on the image of a company are sought.
- After some time, these respondents are shown an advertisement, and thereafter, the data is again collected on the image of the company.
- For those respondents, where the image has improved, there is a positive and for those where the image has declined there is a negative sign assigned and for the one where there is no change, the corresponding observation is dropped from the analysis and the sample size reduced accordingly.
- The key concept underlying the test is that if the advertisement is not effective in improving the image of the company, the number of positive signs should be approximately equal to the number of negative signs.
- For small samples, a binomial distribution could be used, whereas for a large sample, the normal approximation to the binomial distribution could be used, as already explained in the one-sample sign test.

Sign Test Uses *p*-Value to Make Decision



P-value is the probability of getting an observation at least as extreme as we got. If 7 of 8 observations 'favor' H_a, then p-value = $p(x \geq 7)$ = .031 + .004 = .035.

If $\alpha = .05$, then reject H_0 since p-value $\leq \alpha$.

Sign Test for a Population Median η

One-Tailed Test

H_0: $\eta = \eta_0$

H_a: $\eta > \eta_0$ [or H_a: $\eta < \eta_0$]

Test statistic:

S = Number of sample measurements greater than η_0 [or S = number of measurements less than η_0]

Sign Test for a Population Median η

Observed significance level:

p-value = $P(x \geq S)$

where x has a binomial distribution with parameters n and $p = .5$

(Use Table II, Appendix B)

Rejection region: Reject H_0 if p-value $\leq .05$

Sign Test for a Population Median η

Two-Tailed Test

H_0: $\eta = \eta_0$

H_a: $\eta \neq \eta_0$

Test statistic:

S = Larger of S_1 and S_2, where S_1 is the number of sample measurements less than η_0 and S_2 is the number of measurements greater than η_0

Sign Test for a Population Median η

Observed significance level:

p-value = $2P(x \geq S)$

where x has a binomial distribution with parameters n and $p = .5$

(Use Table II, Appendix B)

Rejection region: Reject H_0 if p-value $\leq .05$

Conditions Required for Valid Application of the Sign Test

The sample is selected randomly from a continuous probability distribution.

[*Note*: No assumptions need to be made about the shape of the probability distribution.]

Large Sample Sign Test for a Population Median η

One-Tailed Test

H_0: $\eta = \eta_0$

H_a: $\eta > \eta_0$ [or H_a: $\eta < \eta_0$]

Test statistic: $$z = \frac{(S - .5) - .5n}{.5\sqrt{n}}$$

S = Number of sample measurements greater than η_0 [or S = number of measurements less than η_0] The "– .5" is the "correction for continuity."

Large Sample Sign Test for a Population Median η

The null hypothesized mean value is $np = .5n$, and the standard deviation is

$$\sqrt{npq} = \sqrt{n(.5)(.5)} = .5\sqrt{n}$$

Rejection region: $z > z_\alpha$

where tabulated z-values can be found in Table IV of Appendix B.

Large Sample Sign Test for a Population Median η

Two-Tailed Test

H_0: $\eta = \eta_0$

H_a: $\eta \neq \eta_0$

Test statistic: $$z = \frac{(S - .5) - .5n}{.5\sqrt{n}}$$

S = Larger of S_1 and S_2, where S_1 is the number of sample measurements less than η_0 and S_2 is the number of measurements greater than η_0

Large Sample Sign Test for a Population Median η

The null hypothesized mean value is $np = .5n$, and the standard deviation is

$$\sqrt{npq} = \sqrt{n(.5)(.5)} = .5\sqrt{n}$$

Rejection region: $z > z_{\alpha/2}$

where tabulated *z*-values can be found in Table IV of Appendix B.

Sign Test Example

You're an analyst for Chef-Boy-R-Dee. You've asked 7 people to rate a new ravioli on a 5-point Likert scale (1 = terrible to 5 = excellent). The ratings are:

2 5 3 4 1 4 5

At the **.05** level of significance, is there evidence that the **median** rating is **less than 3**?

Sign Test Solution

- H_0: $\eta = 3$
- H_a: $\eta < 3$
- $\alpha = .05$
- Test Statistic:

$S = 2$

(Ratings 1 & 2 are less than $\eta = 3$: 2, 5, 3, 4, 1, 4, 5)

p-value:

$P(x \geq 2) = 1 - P(x \leq 1)$

$= .937$

(Binomial Table, $n = 7$, $p = 0.50$)

Decision:

Do not reject at $\alpha = .05$

Conclusion:

There is no evidence median is less than 3

Sign Test

- Comparing paired observations
  - ✓ Paired observations: X and Y
  - ✓ **$p = P(X > Y)$**

• Two-tailed test	H_0: $p = 0.50$ H_1: $p \neq 0.50$
• Right-tailed test	H_0: $p \leq 0.50$ H_1: $p > 0.50$
• Left-tailed test	H_0: $p \geq 0.50$ H_1: $p < 0.50$
• Test statistic:	T = Number of + signs

Sign Test Decision Rule

- Small Sample: Binomial Test
  - ✓ For a two-tailed test, find a critical point corresponding as closely as possible to $\alpha/2$ (C_1) and define C_2 as $n-C_1$. Reject null hypothesis if $T \leq C_1$ or $T \geq C_2$.
  - ✓ For a right-tailed test, reject H_0 if $T \geq C$, where C is the value of the binomial distribution with parameters n and $p = 0.50$ such that the sum of the probabilities of all values less than or equal to C is as close as possible to the chosen level of significance, α.
  - ✓ For a left-tailed test, reject H_0 if $T \geq C$, where C is defined as above.

Sign Test

- Comparing paired observations
  - ✓ Paired observations: X and Y
  - ✓ **$p = P(X > Y)$**

•	Two-tailed test	$H_0: p = 0.50$ $H_1: p \neq 0.50$
•	Right-tailed test	$H_0: p \leq 0.50$ $H_1: p > 0.50$
•	Left-tailed test	$H_0: p \geq 0.50$ $H_1: p < 0.50$
•	Test statistic:	T = Number of + signs

Sign Test Decision Rule

- Small Sample: Binomial Test
  - ✓ For a two-tailed test, find a critical point corresponding as closely as possible to $\alpha/2$ (C_1) and define C_2 as n-C_1. Reject null hypothesis if $T \leq C_1$ or $T \geq C_2$.
  - ✓ For a right-tailed test, reject H_0 if $T \geq C$, where C is the value of the binomial distribution with parameters n and p = 0.50 such that the sum of the probabilities of all values less than or equal to C is as close as possible to the chosen level of significance, α.
  - ✓ For a left-tailed test, reject H_0 if $T \geq C$, where C is defined as above.

Example

CEO	Before	After	Sign	
1	3	4	1	+
2	5	5	0	
3	2	3	1	+
4	2	4	1	+
5	4	4	0	
6	2	3	1	+
7	1	2	1	+
8	5	4	-1	-
9	4	5	1	+
10	5	4	-1	-
11	3	4	1	+
12	2	5	1	+
13	2	5	1	+
14	2	3	1	+
15	1	2	1	+
16	3	2	-1	-
17	4	5	1	+

$n = 15$ $T = 12$

$\alpha \approx 0.025$

$C1=3$ $C2 = 15-3 = 12$

H0 rejected, since

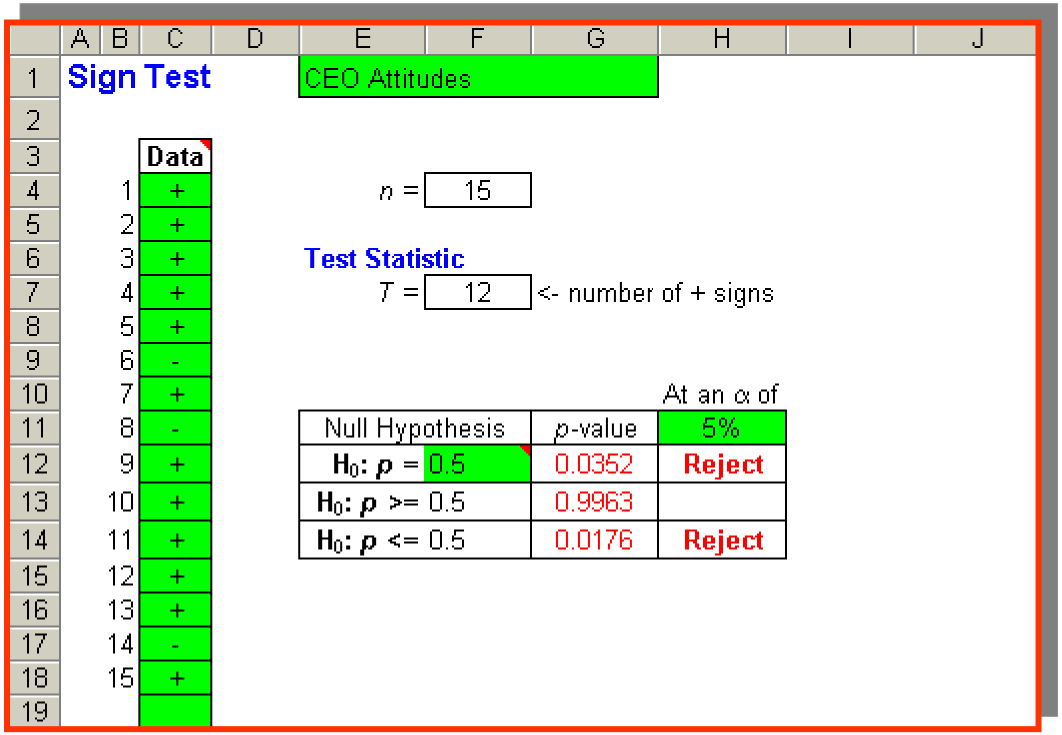
$T \geq C2$

Cumulative Binomial Probabilities (n=15, p=0.5)

x	F(x)
0	0.00003
1	0.00049
2	0.00369
3	0.01758
4	0.05923
5	0.15088
6	0.30362
7	0.50000
8	0.69638
9	0.84912
10	0.94077
11	0.98242
12	0.99631
13	0.99951
14	0.99997
15	1.00000

C1

Example 14-1- Using the Template



H_0: p = 0.5
H_1: p ≠ 0.5
Test Statistic: T = 12
p-value = 0.0352.
For α = 0.05, the null hypothesis is rejected since 0.0352 < 0.05.

Thus one can conclude that there is a change in attitude toward a CEO following the award of an MBA degree.

Run Test for Randomness

Run test is used to test the randomness of a sample.

Run: A run is defined as a sequence of like elements that are preceded and followed by different elements or no elements at all.

Let

n = Total size of the sample
n_1 = Size of sample in group 1
n_2 = Size of sample in group 2
r = Number of runs

For large samples, either $n_1 > 20$ or $n_2 > 20$, the distribution of runs (r) is normally distributed with:

Mean $$\mu_r = 1 + \frac{2n_1n_2}{n_1+n_2}$$

Standard Deviation $$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$$

Run Test for Randomness

The hypothesis is to be tested is:

H_0 : The pattern of sequence is random.

H_1 : The pattern of sequence is not random.

For a large sample the test statistic is given by:

$$Z = \frac{r - \mu_r}{\sigma_r}$$

For a given level of significance, if the absolute value of computed z is greater than the absolute value of tabulated z, null hypothesis is rejected.

In case of numerical data, the original data are grouped into two categories, one above and second below median.

14-3 The Runs Test - A Test for Randomness

A **run** is a sequence of like elements that are preceded and followed by different elements or no element at all.

Case 1: S|E|S|E|S|E|S|E|S|E|S|E|S|E|S|E|S|E|S|E : R = 20 Apparently nonrandom
Case 2: SSSSSSSSSS|EEEEEEEEEE : R = 2 Apparently nonrandom
Case 3: S|EE|SS|EEE|S|E|SS|E|S|EE|SSS|E : R = 12 Perhaps random

A **two-tailed hypothesis test for randomness:**

H0: Observations are generated randomly

H1: Observations are not generated randomly

Test Statistic:

R=Number of Runs

Reject H0 at level α if R $\leq$ C1 or R $\geq$ C2, as given in Table 8, with total tail probability $P(R \leq C1) + P(R \geq C2) = \alpha$.

Runs Test: Examples

Table 8:	Number of Runs (r)									
(n1,n2)	11	12	13	14	15	16	17	18	19	20
•										
•										
•										
(10,10)	0.586	0.758	0.872	0.949	0.981	0.996	0.999	1.000	1.000	1.000

Case 1: n1 = 10 n2 = 10 R= 20 p-value≈0

Case 2: n1 = 10 n2 = 10 R = 2 p-value ≈0

Case 3: n1 = 10 n2 = 10 R= 12

$$\text{p-value} = 2[P(R \geq 12)] = 2[1-F(11)]$$

$$= (2)(1-0.586) = (2)(0.414) = 0.828$$

H0 not rejected

Large-Sample Runs Test: Using the Normal Approximation

The mean of the normal distribution of the number of runs:

$$E(R) = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

The standard deviation:

$$\sigma_R = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

The *standard normal test statistic:*

$$z = \frac{R - E(R)}{\sigma_R}$$

Large-Sample Runs Test: Example 14-2

Example 14-2: n1 = 27 n2 = 26 R = 15

$$E(R) = \frac{2n_1 n_2}{n_1 + n_2} + 1 = \frac{(2)(27)(26)}{(27+26)} + 1 = 26.49 + 1 = 27.49$$

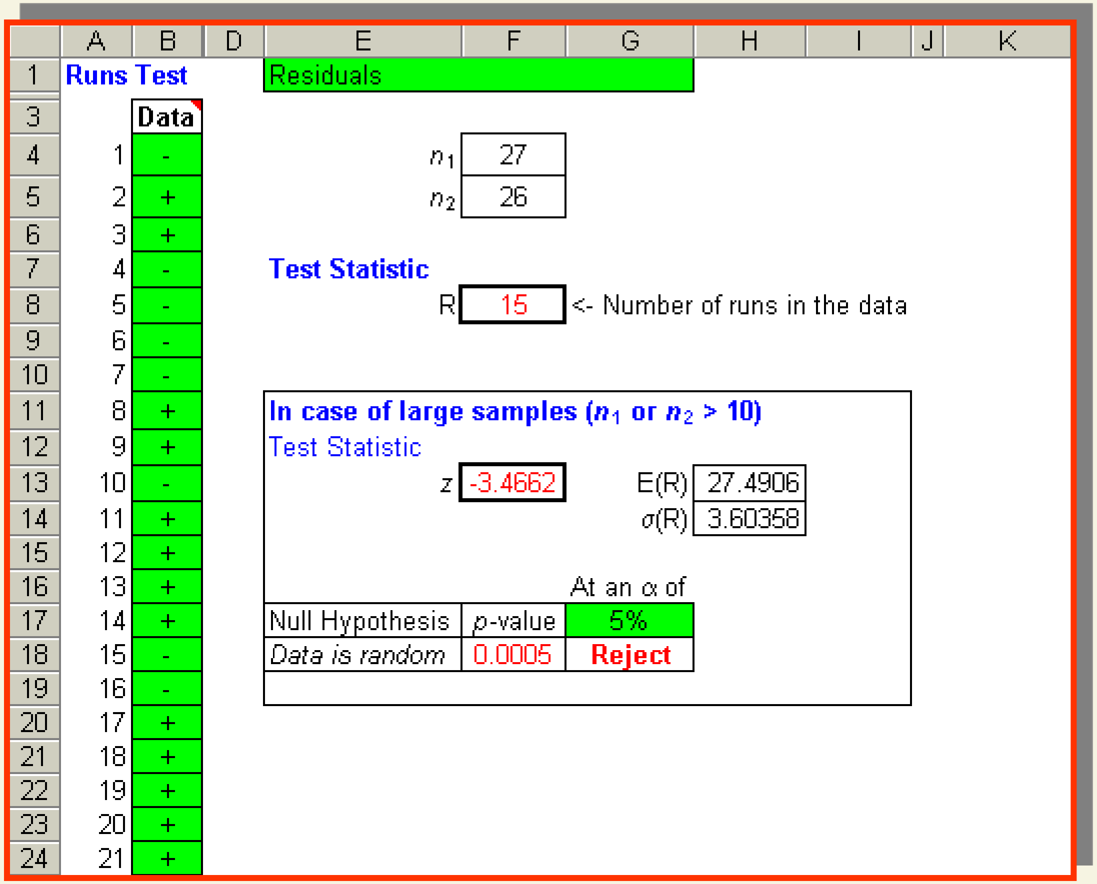
$$\sigma_R = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}} = \sqrt{\frac{(2)(27)(26)((2)(27)(26) - 27 - 26))}{(27+26)^2 (27+26-1)}}$$

$$= \sqrt{\frac{1896804}{146068}} = \sqrt{12.986} = 3.604$$

$$z = \frac{R - E(R)}{\sigma_R} = \frac{15 - 27.49}{3.604} = -3.47 \qquad \text{p - value} = 2(1 - .9997) = 0.0006$$

H0 should be rejected at any common level of significance.

Large-Sample Runs Test: Example 14-2 – Using the Template



Note:

The computed p-value using the template is 0.0005 as compared to the manually computed value of 0.0006. The value of 0.0005 is more accurate.

Reject the null hypothesis that the residuals are random.

Using the Runs Test to Compare Two Population Distributions (Means): the Wald-Wolfowitz Test

The null and alternative hypotheses for the Wald-Wolfowitz test:

H0: The two populations have the same distribution

H1: The two populations have different distributions

The test statistic:

R = Number of Runs in the sequence of samples, when the data from both samples have been sorted

The Wald-Wolfowitz Test: Example

Sales	**Sales Person**	**Sales (Sorted)**	**Sales Person (Sorted)**	**Runs**
35	A	13	B	
44	A	16	B	
39	A	17	B	
48	A	21	B	
60	A	24	B	1
75	A	29	A	2
49	A	32	B	
66	A	33	B	3
17	B	35	A	
23	B	39	A	
13	B	44	A	
24	B	48	A	
33	B	49	A	
21	B	50	A	
18	B	60	A	
16	B	66	A	
32	B	75	A	4

n1 = 10 n2 = 9 R= 4

p-value = $2[P(R \leq 4)] = 0.002$

H0 may be rejected

Table (n1,n2)	Number of Runs (r) 2	3	4	5
•				
•				
•				
(9,10)	0.000	0.000	0.002	0.004 ...

Ranks Tests

- Ranks tests
  - ✓ Mann-Whitney U Test: Comparing two populations
  - ✓ Wilcoxon Signed-Rank Test: Paired comparisons
  - ✓ Comparing several populations: ANOVA with ranks
    - Kruskal-Wallis Test
    - Friedman Test: Repeated measures

Mann-Whitney U Test for Independent Samples

- This test is an alternative to a t test for testing the equality of means of two independent samples
- The application of a t test involves the assumption that the samples are drawn from the normal population.
- If the normality assumption is violated, this test can be used as an alternative to a t test.

Mann-Whitney U Test for Independent Samples

- A two-tailed hypothesis for a Mann-Whitney test could be written as:

H0 : Two samples come from identical populations

or

Two populations have identical probability distribution.

H1 : Two samples come from different populations

or

Two populations differ in locations.

Mann-Whitney U Test for Independent Samples

The following steps are used in conducting this test:

(i) The two samples are combined (pooled) into one large sample and then we determine the rank of each observation in the pooled sample. If two or more sample values in the pooled samples are identical, i.e., if there are ties, the sample values are each assigned a rank equal to the mean of the ranks that would otherwise be assigned.

(ii) We determine the sum of the ranks of each sample. Let R_1 and R_2 represent the sum of the ranks of the first and the second sample whereas n_1 and n_2 are the respective sample sizes of the first and the second sample. For convenience, choose n_1 as a small size if they are unequal so that $n_1 \leq n_2$. A significant difference between R_1 and R_2 implies a significant difference between the samples.

Mann-Whitney U Test for Independent Samples

(iii) Define $U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$

and $U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$

Please note that the following expression will hold true:

$$U_1 + U_2 = n_1 n_2$$

If n_1 or $n_2 > 10$, a Z test would be appropriate. For this purpose, either of U_1 or U_2 could be used for testing a one-tailed or a two-tailed test. In this test, U_2 will be used for the purpose.

Mann-Whitney U Test for Independent Samples

Under the assumption that the null hypothesis is true, the U_2 statistic follows an approximately normal distribution with mean:

$$\mu_{u_2} = \frac{n_1 n_2}{2}$$

and standard deviation:

$$\sigma_{u_2} = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

The test statistic is:

$$Z = \frac{U_2 - \mu_{u_2}}{\sigma_{u_2}}$$

- Assuming the level of significance as equal to α, if the absolute sample value of Z is greater than the absolute critical value of Z, i.e., $Z_{\alpha/2}$, the null hypothesis is rejected.
- A similar procedure is used for a one-tailed test.

14-4 The Mann-Whitney U Test (Comparing Two Populations)

The null and alternative hypotheses:

H0: The distributions of two populations are identical

H1: The two population distributions are not identical

The Mann-Whitney *U* statistic:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \qquad \mathrm{R}_1 = \sum \text{Ranks from sample 1}$$

where n1 is the sample size from population 1 and n2 is the sample size from population 2.

$$E[U] = \frac{n_1 n_2}{2} \qquad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$\text{The large-sample test statistic: } z = \frac{U - E[U]}{\sigma_U}$$

The Mann-Whitney U Test: Example 14-4

Model	Time	Rank	Rank Sum
A	35	5	
A	38	8	
A	40	10	
A	42	12	
A	41	11	
A	36	6	52
B	29	2	
B	27	1	
B	30	3	
B	33	4	
B	39	9	
B	37	7	26

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$
$$= (6)(6) + \frac{(6)(6+1)}{2} - 52$$
$$= 5$$

Cumulative Distribution Function of the Mann-Whitney U Statistic

u	n2=6, n1=6	
.		
.		
.		
4	0.0130	
5	0.0206	$P(u \leq 5)$
6	0.0325	
.		
.		
.		

Example 14-5: Large-Sample Mann-Whitney U Test

Score	Program	Score Rank	Rank Sum
85	1	20.0	20.0
87	1	21.0	41.0
92	1	27.0	68.0
98	1	30.0	98.0
90	1	26.0	124.0
88	1	23.0	147.0
75	1	17.0	164.0
72	1	13.5	177.5
60	1	6.5	184.0
93	1	28.0	212.0
88	1	23.0	235.0
89	1	25.0	260.0
96	1	29.0	289.0
73	1	15.0	304.0
62	1	8.5	312.5

Score	Program	Score Rank	Rank Sum
65	2	10.0	10.0
57	2	4.0	14.0
74	2	16.0	30.0
43	2	2.0	32.0
39	2	1.0	33.0
88	2	23.0	56.0
62	2	8.5	64.5
69	2	11.0	75.5
70	2	12.0	87.5
72	2	13.5	101.0
59	2	5.0	106.0
60	2	6.5	112.5
80	2	18.0	130.5
83	2	19.0	149.5
50	2	3.0	152.5

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= (15)(15) + \frac{(15)(15+1)}{2} - 312.5 = 32.5$$

$$E[U] = \frac{n_1 n_2}{2} = \frac{(15)(15)}{2} = 112.5$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$= \sqrt{\frac{(15)(15)(15+15+1)}{12}} = 24.109$$

$$z = \frac{U - E[U]}{\sigma_U} = \frac{32.5 - 112.5}{24.109} = -3.32$$

Since the test statistic is z = -3.32, the p-value ≈ 0.0005, and H0 is rejected.

Example 14-5: Large-Sample Mann-Whitney U Test – Using the Template

Mann-Whitney U Test

The Learning Curve method

	Group	Data	Rank
1	2	65	10
2	2	57	4
3	2	74	16
4	2	43	2
5	2	39	1
6	2	88	23
7	2	62	8.5
8	2	69	11
9	2	70	12
10	2	72	13.5
11	2	59	5
12	2	60	6.5
13	2	80	18
14	2	83	19
15	2	50	3
16	1	85	20
17	1	87	21
18	1	92	27
19	1	98	30
20	1	90	26

Counts

1	15
2	15

Test Statistic

U 32.5

Rank Sums

1	312.5
2	152.5

In case of large samples (n_1 or n_2 > 10)

Test Statistic

z -3.3182

E[U] 112.5

σ(U) 24.1091

Null Hypothesis	p-value	At an α of 5%
$H_0: \mu_1 = \mu_2$	0.0009	Reject
$H_0: \mu_1 >= \mu_2$	0.9995	
$H_0: \mu_1 <= \mu_2$	0.0005	Reject

Since the test statistic is z = -3.32, the p-value ≈ 0.0005, and H0 is rejected.

That is, the LC (Learning Curve) program is more effective.

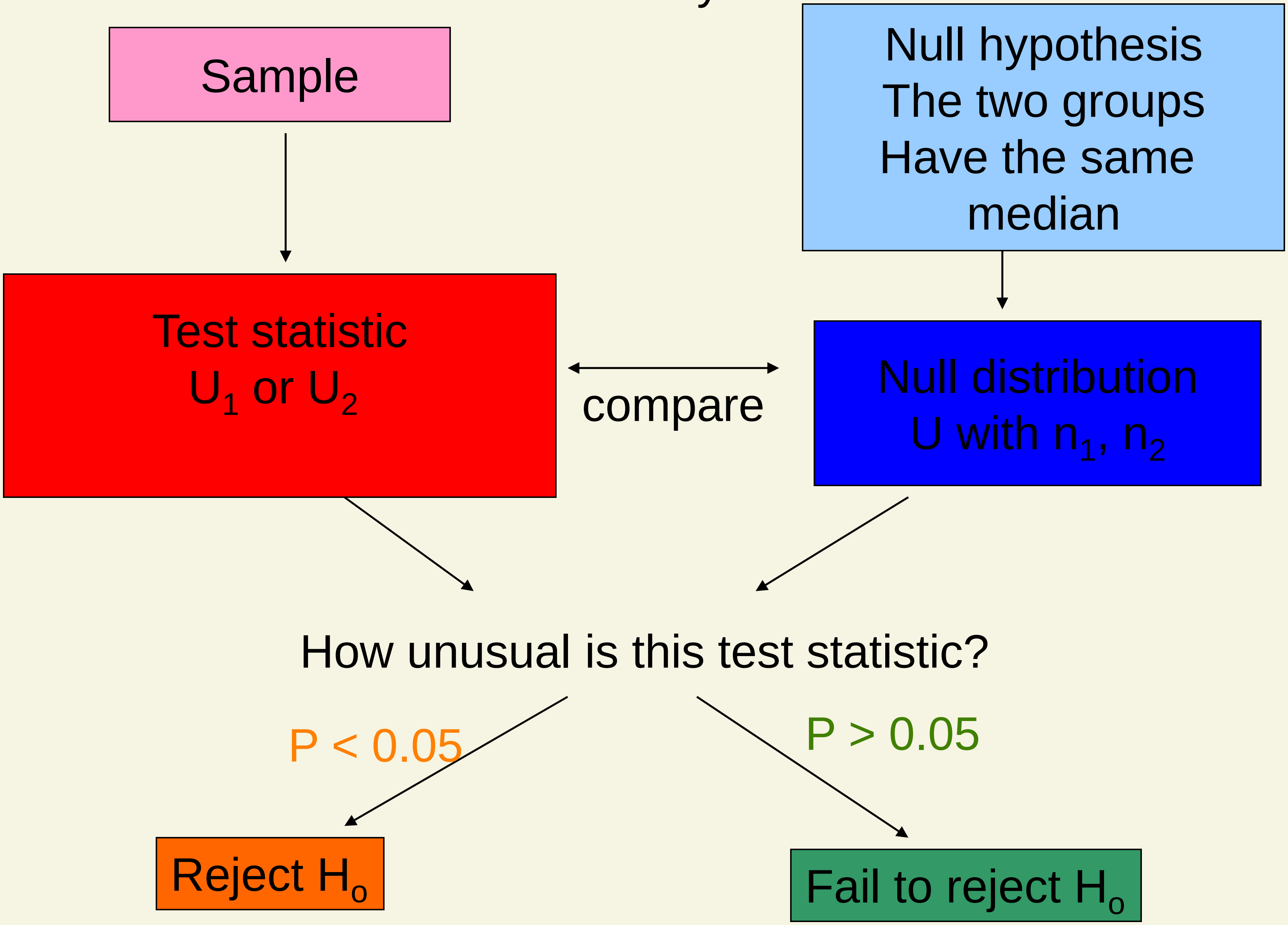
When Should I Use the Mann-Whitney U Test?

- Non-parametric distribution
- Independent Measures
- Ratio or interval data, but must be reduced to ordinal data because it is non-parametric
- Different variances for two groups
- It assumes Random assignments to groups
- Test statistic: U

Mann-Whitney U Test

- What is it for? A non-parametric test to compare the central tendencies of two groups
- It assumes Random assignments to groups
- Test statistic: U

Mann-Whitney U test



Null Hypothesis

- Involves the creation of two competing explanations for the data recorded.
  - Idea 1:These are pattern-less random data. Any observed patterns are due to chance. This is the null hypothesis **H0**
  - Idea 2: There is a defined pattern in the data. This is the alternative hypothesis **H1**
- Without the statement of the competing hypotheses, no meaning test can be run.

Step One

- Arrange all the observations into a single ranked series. That is, rank all the observations without regard to which sample they are in.
- In other words, combine all of the data from both groups into a single column, in order, but keep track of what group they came from.

Mann-Whitney U Test

- If you have ties:
  - Rank them anyway, pretending they were slightly different
  - Find the average of the ranks for the identical values, and give them all that rank
  - Carry on as if all the whole-number ranks have been used up

Example

Data

14
2
5
4
2
14
18
14

Example

Data	Sorted Data	
14	2	G1
2	2	G2
5	4	G2
4	5	G1
2	14	G1
14	14	G2
18	14	G1
14	18	G2

Example

Data	Sorted Data	
14	2	G1
2	2	G2
5	4	G2
4	5	G1
2	14	G1
14	14	G2
18	14	G1
14	18	G2

TIES

Example

Data	Sorted Data	
14	2	G1
2	2	G2
5	4	G2
4	5	G1
2	14	G1
14	14	G2
18	14	G1
14	18	G2

TIES

Rank them anyway, pretending they were slightly different

Example

Data	Sorted Data		Rank A
14	2	G1	1
2	2	G2	2
5	4	G2	3
4	5	G1	4
2	14	G1	5
14	14	G2	6
18	14	G1	7
14	18	G2	8

Example

Data	Sorted Data		Rank A
14	2	G1	1
2	2	G2	2
5	4	G2	3
4	5	G1	4
2	14	G1	5
14	14	G2	6
18	14	G1	7
14	18	G2	8

Find the average of the ranks for the identical values, and give them all that rank

Example

Data	Sorted Data		Rank A	
14	2	G1	1	Average = 1.5
2	2	G2	2	
5	4	G2	3	
4	5	G1	4	
2	14	G1	5	
14	14	G2	6	Average = 6
18	14	G1	7	
14	18	G2	8	

Example

Data	Sorted Data		Rank A	Rank
14	2	G1	1	1.5
2	2	G2	2	1.5
5	4	G2	3	3
4	5	G1	4	4
2	14	G1	5	6
14	14	G2	6	6
18	14	G1	7	6
14	18	G2	8	8

Example

Data	Sorted Data		Rank A	Rank
14	2	G1	1	1.5
2	2	G2	2	1.5
5	4	G2	3	3
4	5	G1	4	4
2	14	G1	5	6
14	14	G2	6	6
18	14	G1	7	6
14	18	G2	8	8

First, sort them back into the two groups,
Then these can now be used for the Mann-Whitney U test

Step Two

- Add up the ranks for the observations which came from sample 1(the smaller group, fewer participants).
- Then add up the sum of ranks in sample 2 (Larger group)

Step Three

- *U* is then given by:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

where *n*1 is the sample size for sample 1(smaller group), and *R*1 is the sum of the ranks in sample 1

Step Four

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}.$$

Step Five

- The smaller value of $U1$ and $U2$ is the one used when consulting significance tables.

1-tail test at α = 0.05 *or 2-tail test at α = 0.10*

N_2 \ N_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2					0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
3			0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
4			0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
5		0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
6		0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
7		0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
8		1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
9		1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
10		1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
11		1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
12		2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
13		2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
14		2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
15		3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
16		3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
17		3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
18		4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

$N_1 < N_2$

Compare

- If the smaller value of U1 or U2 is smaller than the critical value in the chart, then the probability that the differences in the groups is obtained by chance is less than 0.05, and you may reject the null hypothesis.

14-4 The Mann-Whitney U Test (Comparing Two Populations)

The null and alternative hypotheses:

H0: The distributions of two populations are identical

H1: The two population distributions are not identical

The Mann-Whitney U statistic:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \qquad \mathrm{R}_1 = \sum \text{Ranks from sample 1}$$

where n1 is the sample size from population 1 and n2 is the sample size from population 2.

$$E[U] = \frac{n_1 n_2}{2} \qquad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$\text{The large-sample test statistic: } z = \frac{U - E[U]}{\sigma_U}$$

The Mann-Whitney U Test: Example 14-4

Model	Time	Rank	Rank Sum
A	35	5	
A	38	8	
A	40	10	
A	42	12	
A	41	11	
A	36	6	52
B	29	2	
B	27	1	
B	30	3	
B	33	4	
B	39	9	
B	37	7	26

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$
$$= (6)(6) + \frac{(6)(6 + 1)}{2} - 52$$
$$= 5$$

Cumulative Distribution Function of the Mann-Whitney U Statistic

n2=6

n1=6

u		
.		
.		
.		
4	0.0130	P(u≤5)
5	0.0206	
6	0.0325	
.		
.		
.		

Example 14-5: Large-Sample Mann-Whitney U Test

Score	Program	Score Rank	Rank Sum
85	1	20.0	20.0
87	1	21.0	41.0
92	1	27.0	68.0
98	1	30.0	98.0
90	1	26.0	124.0
88	1	23.0	147.0
75	1	17.0	164.0
72	1	13.5	177.5
60	1	6.5	184.0
93	1	28.0	212.0
88	1	23.0	235.0
89	1	25.0	260.0
96	1	29.0	289.0
73	1	15.0	304.0
62	1	8.5	312.5

Score	Program	Score Rank	Rank Sum
65	2	10.0	10.0
57	2	4.0	14.0
74	2	16.0	30.0
43	2	2.0	32.0
39	2	1.0	33.0
88	2	23.0	56.0
62	2	8.5	64.5
69	2	11.0	75.5
70	2	12.0	87.5
72	2	13.5	101.0
59	2	5.0	106.0
60	2	6.5	112.5
80	2	18.0	130.5
83	2	19.0	149.5
50	2	3.0	152.5

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = (15)(15) + \frac{(15)(15+1)}{2} - 312.5 = 32.5$$

$$E[U] = \frac{n_1 n_2}{2} = \frac{(15)(15)}{2} = 112.5$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(15)(15)(15+15+1)}{12}} = 24.109$$

$$z = \frac{U - E[U]}{\sigma_U} = \frac{32.5 - 112.5}{24.109} = -3.32$$

Since the test statistic is z = -3.32,
the p-value ≈ 0.0005, and H0 is rejected.

Mann-Whitney *U*-test

- Based on ordinal data
- If differences exist scores in one group should be larger than in the other

Group A Scores	**Group B Scores**
3, 4, 4, 9	7, 10, 10, 12

Rank ordering the data

- Scores must be combined and rank ordered to carry out the analysis e.g.,

Original scores:	3	4	4	7	9	10	10	12
Ordinal scores:	1	2	3	4	5	6	7	8
Final Ranks:	1	2.5	2.5	4	5	6.5	6.5	8

- If there is a difference, scores for one group should be concentrated at one end (e.g., end which represents a high score) while the scores for the second group are concentrated at the other end

Null hypothesis

- H_0: There is no tendency for ranks in one treatment condition to be systematically higher or lower than the ranks in the other treatment condition.
- Could also be thought of as
  - Mean rank for inds in the first treatment is the same as the mean rank for the inds in the second treatment
    - Less accurate since average rank is not calculated

Calculation

- For each data point, need to identify how many data points in the **other group** have a **larger rank order**
- Sum these for each group - referred to as *U* scores
- As difference between two Gs increases so the difference between these two sum scores (*U* values) increases

Calculating *U* scores

Rank	Score	No of data points in alternative G with larger rank
1	3	4
2.5	4	4
2.5	4	4
4	7	1
5	9	3
6.5	10	0
6.5	10	0
8	12	0
U score for both Gs	U_A	15
	U_B	1

Determining significance

- Mann-Whitney *U* value = the smaller of the two *U* values calculated - here it is 1
- With the specified *n* for each group you can look up a value of *U* which your result should be **equal to or lower than** to be considered sig

Mann-Whitney *U* table
(2 groups of 4 two-tailed),

Table A5.07: Critical Values for the Wilcoxon/Mann-W

Nondirectional α=.05 (Directional α=.025)

	n_2															
n_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1
3	-	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6
4	-	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11
5	-	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15
6	-	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21
7	-	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26
8	-	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31
9	-	0	2	4	7	10	12	15	17	21	23	26	28	31	34	37
10	-	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42
11	-	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47
12	-	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53

Note extremes…

- At the extreme there should be no overlap and therefore the Mann-Whitney U value should be = 0
- As the two groups become more alike then the ranks begin to intermix and U becomes larger

Reporting the result

- Critical $U = 0$
- Critical value is dependent on n for each group
- U=1 (n=4,4), p>.05, two tailed

Formula for calculation

- Previous process can be tedious and therefore using a formula is more ‘straight forward’

$$U_A = n_A n_B + \frac{n_A(n_A + 1)}{2} - \sum R_A$$

Where $\sum R_A$ is the sum of ranks for Group A

Comparing Two Populations: Independent Samples

Wilcoxon Signed-Rank Test for Paired Samples

- The case of paired sample (dependent sample) was discussed in Chapter 12 using a t distribution.
- The use of t distribution is based on the normality assumption.
- However, there are instances when the normality assumption is not satisfied and one has to resort to a non-parametric test. One such test earlier discussed was the two-sample sign test.
- In two-sample sign test, only the sign of the difference (positive or negative) was taken into account and no weightage was assigned to the magnitude of the difference.
- The Wilcoxon matched-pair signed rank test takes care of this limitation and attaches a greater weightage to the matched pair with a larger difference.
- The test, therefore, incorporates and makes use of more information than the sign test.
- This is, therefore, a more powerful test than the sign test.

Wilcoxon Signed-Rank Test for Paired Samples

The test procedure is outlined in the following steps:

(i) Let d_i denote the difference in the score for the i^{th} matched pair. Retain signs, but discard any pair for which d = 0.

(ii) Ignoring the signs of difference, rank all the d_i's from the lowest to highest. In case the differences have the same numerical values, assign to them the mean of the ranks involved in the tie.

(iii) To each rank, prefix the sign of the difference.

(iv) Compute the sum of the absolute value of the negative and the positive ranks to be denoted as T– and T+ respectively.

(v) Let T be the smaller of the two sums found in step iv.

Wilcoxon Signed-Rank Test for Paired Samples

- When the number of the pairs of observation (n) for which the difference is not zero is greater than 15, the T statistic follows an approximate normal distribution under the null hypothesis, that the population differences are centered at 0.
- The mean μ_T and standard deviation σ_T of T are given by:

$$\mu_T = \frac{n(n+1)}{4} \quad \text{and} \quad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

- The test statistics is given by:

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

Wilcoxon Signed-Rank Test for Paired Samples

- For a given level of significance α, the absolute sample Z should be greater than the absolute $Z_{\alpha/2}$ to reject the null hypothesis.
- For a one-sided upper tail test, the null hypothesis is rejected if the sample Z is greater than Z_{α} and for a one-sided lower tail test, the null hypothesis is rejected if sample Z is less than $-Z_{\alpha}$.

Conditions Required for Valid Wilcoxon Rank Sum Test

1. The two samples are random and independent.
2. The two probability distributions from which the samples are drawn are continuous.

Wilcoxon Rank Sum Test Procedure

1. Assign ranks, R_i, to the $n_1 + n_2$ sample observations
   - If unequal sample sizes, let n_1 refer to smaller-sized sample
   - Smallest value = 1
   - Average ties
2. Sum the ranks, T_i, for each sample
3. Test statistic is T_i (smallest sample)

Wilcoxon Rank Sum Test for Large Samples ($n_1 \geq 10$ and $n_2 \geq 10$)

Let D_1 and D_2 represent the probability distributions for populations 1 and 2, respectively.

One-Tailed Test

H_0: D_1 and D_2 are identical

H_a: D_1 is shifted to the right of D_2

[or D_1 is shifted to the left of D_2]

Wilcoxon Rank Sum Test for Large Samples ($n_1 \geq 10$ and $n_2 \geq 10$)

Test statistic: $$z = \frac{T_1 - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Rejection region:

$z > z_\alpha$ (or $z < -z_\alpha$)

Wilcoxon Rank Sum Test for Large Samples ($n_1 \geq 10$ and $n_2 \geq 10$)

Let D_1 and D_2 represent the probability distributions for populations 1 and 2, respectively.

Two-Tailed Test

H_0: D_1 and D_2 are identical

H_a: D_1 is shifted to the right or to the left of D_2

Wilcoxon Rank Sum Test for Large Samples ($n_1 \geq 10$ and $n_2 \geq 10$)

Test statistic: $$z = \frac{T_1 - \frac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Rejection region:

$|z| > z_{\alpha/2}$

Signed Rank Test Procedure

1. Obtain difference scores, $D_i = X_{1i} - X_{2i}$
2. Take absolute value of differences, $|D_i|$
3. Delete differences with 0 value
4. Assign ranks, R_i, where smallest = 1
5. Assign ranks same signs as D_i
6. Sum '+' ranks (T_+) and '–' ranks (T_-)
   - Test statistic is T_- or T_+ (one-tail test)
   - Test statistic is T = smaller of T_- or T_+ (two-tail test)

14-5 The Wilcoxon Signed-Ranks Test (Paired Ranks)

The null and alternative hypotheses:

H0: The median difference between populations are 1 and 2 is zero

H1: The median difference between populations are 1 and 2 is not zero

Find the difference between the ranks for each pair, *D = x1 -x2*, and then rank the absolute values of the differences.

The Wilcoxon T statistic is the smaller of the sums of the positive ranks and the sum of the negative ranks:

$$T = \min\left(\sum(+), \sum(-)\right)$$

For small samples, a left-tailed test is used, using the values in Appendix C, Table 10.

$$E[T] = \frac{n(n+1)}{4} \qquad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

The large-sample test statistic:

$$z = \frac{T - E[T]}{\sigma_T}$$

Example 14-6

Sold (1)	Sold (2)	D=x1-x2	ABS(D)	Rank ABS(D)	Rank (D>0)	Rank (D<0)
56	40	16	16	9.0	9.0	0
48	70	-22	22	12.0	0.0	12
100	60	40	40	15.0	15.0	0
85	70	15	15	8.0	8.0	0
22	8	14	14	7.0	7.0	0
44	40	4	4	2.0	2.0	0
35	45	-10	10	6.0	0.0	6
28	7	21	21	11.0	11.0	0
52	60	-8	8	5.0	0.0	5
77	70	7	7	3.5	3.5	0
89	90	-1	1	1.0	0.0	1
10	10	0	*	*	*	*
65	85	-20	20	10.0	0.0	10
90	61	29	29	13.0	13.0	0
70	40	30	30	14.0	14.0	0
33	26	7	7	3.5	3.5	0
				Sum:	**86**	**34**

T=34

	n=15
P=0.05	30
P=0.025	25
P=0.01	20
P=0.005	16

H0 is not rejected (Note the arithmetic error in the text for store 13)

Example 14-7

Hourly Messages	Md0	D=x1-x2	ABS(D)	Rank ABS(D)	Rank (D>0)	Rank (D<0)
151	149	2	2	1.0	1.0	0.0
144	149	-5	5	2.0	0.0	2.0
123	149	-26	26	13.0	0.0	13.0
178	149	29	29	15.0	15.0	0.0
105	149	-44	44	23.0	0.0	23.0
112	149	-37	37	20.0	0.0	20.0
140	149	-9	9	4.0	0.0	4.0
167	149	18	18	10.0	10.0	0.0
177	149	28	28	14.0	14.0	0.0
185	149	36	36	19.0	19.0	0.0
129	149	-20	20	11.0	0.0	11.0
160	149	11	11	6.0	6.0	0.0
110	149	-39	39	21.0	0.0	21.0
170	149	21	21	12.0	12.0	0.0
198	149	49	49	25.0	25.0	0.0
165	149	16	16	8.0	8.0	0.0
109	149	-40	40	22.0	0.0	22.0
118	149	-31	31	16.5	0.0	16.5
155	149	6	6	3.0	3.0	0.0
102	149	-47	47	24.0	0.0	24.0
164	149	15	15	7.0	7.0	0.0
180	149	31	31	16.5	16.5	0.0
139	149	-10	10	5.0	0.0	5.0
166	149	17	17	9.0	9.0	0.0
82	149	33	33	18.0	18.0	0.0
				Sum:	**163.5**	**161.5**

$$E[T] = \frac{n(n+1)}{4} = \frac{(25)(25+1)}{4} = 162.5$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{25(25+1)((2)(25)+1)}{24}} = \sqrt{\frac{33150}{24}} = 37.165$$

The large - sample test statistic:

$$z = \frac{T - E[T]}{\sigma_T} = \frac{163.5 - 162.5}{37.165} = 0.027$$

H_0 cannot be rejected

The Kruskal-Wallis Test

- The Kruskal-Wallis test is in fact a non-parametric counterpart to the one-way ANOVA.
- The test is an extension of the Mann-Whitney U test.
- Both of them require that the scale of the measurement of a sample value should be at least ordinal.
- The hypothesis to be tested in-Kruskal-Wallis test is:

H_0 : The k populations have identical probability distribution.

H_1 : At least two of the populations differ in locations.

The Kruskal-Wallis Test

The procedure for the test is listed below:

(i) Obtain random samples of size n_1, ..., n_k from each of the k populations. Therefore, the total sample size is

$$n = n_1 + n_2 + ... + n_k$$

(ii) Pool all the samples and rank them, with the lowest score receiving a rank of 1. Ties are to be treated in the usual fashion by assigning an average rank to the tied positions.

(iii) Let r_i = the total of the ranks from the i^{th} sample.

The Kruskal-Wallis Test

- The Kruskal-Wallis test uses the $\chi 2$ to test the null hypothesis. The test statistic is given by:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{r_i^2}{n_i} - 3(n+1),$$

which follows a $\chi 2$ distribution with the k–1 degrees of freedom.

Where, k = Number of samples
n = Total number of elements in k samples.

- The null hypothesis is rejected, if the computed $\chi 2$ is greater than the critical value of $\chi 2$ at the level of significance α.

Kruskal-Wallis *H*-Test

- Tests the equality of more than two (p) population probability distributions
- Corresponds to ANOVA for more than two means
- Used to analyze completely randomized experimental designs
- Uses χ^2 distribution with $p - 1$ df
  - if sample size $n_j \geq 5$

Kruskal-Wallis *H*-Test for Comparing *k* Probability Distributions

Rejection region:

$H > \chi^2_\alpha$ with $(k - 1)$ degrees of freedom

Ties: Assign tied measurements the average of the ranks they would receive if they were unequal but occurred in successive order. For example, if the third-ranked and fourth-ranked measurements are tied, assign each a rank of (3 + 4)/2 = 3.5. The number should be small relative to the total number of observations.

Kruskal-Wallis *H*-Test Procedure

1. Assign ranks, R_i, to the n combined observations
   - Smallest value = 1; largest value = n
   - Average ties
2. Sum ranks for each group
3. Compute test statistic

Squared total of each group

$$H = \left(\frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} \right) - 3(n+1)$$

14-6 The Kruskal-Wallis Test - A Nonparametric Alternative to One-Way ANOVA

The Kruskal-Wallis hypothesis test:

H0: All k populations have the same distribution

H1: Not all k populations have the same distribution

The Kruskal-Wallis test statistic:

$$H = \frac{12}{n(n+1)}\left(\sum_{j=1}^{k}\frac{R_j^2}{n_j}\right) - 3(n+1)$$

If each nj > 5, then H is approximately distributed as a χ^2.

Example 14-8: The Kruskal-Wallis Test

Software	Time	Rank	Group	RankSum
1	45	14	1	90
1	38	10	2	56
1	56	16	3	25
1	60	17		
1	47	15		
1	65	18		
2	30	8		
2	40	11		
2	28	7		
2	44	13		
2	25	5		
2	42	12		
3	22	4		
3	19	3		
3	15	1		
3	31	9		
3	27	6		
3	17	2		

$$H = \frac{12}{n(n+1)}\left(\sum_{j=1}^{k}\frac{R_j^2}{n_j}\right) - 3(n+1)$$

$$= \frac{12}{18(18+1)}\left(\frac{90^2}{6} + \frac{56^2}{6} + \frac{25^2}{6}\right) - 3(18+1)$$

$$= \left(\frac{12}{342}\right)\left(\frac{11861}{6}\right) - 57$$

$$= 12.3625$$

$\chi^2(2,0.005)=10.5966$, so H0 is rejected.

Further Analysis (Pairwise Comparisons of Average Ranks)

If the null hypothesis in the Kruskal-Wallis test is rejected, then we may wish, in addition, compare each pair of populations to determine which are different and which are the same.

The pairwise comparison test statistic:

$$D = \left| \bar{R}_i - \bar{R}_j \right|$$

where $\bar{R}_i$ is the mean of the ranks of the observations from population i.

The critical point for the paired comparisons:

$$C_{KW} = \sqrt{(\chi^2_{\alpha,k-1})\left[\frac{n(n+1)}{12}\right]\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

Reject if $D > C_{KW}$

Pairwise Comparisons: Example 14-8

Critical Point:

$$C_{KW} = \sqrt{(\chi^2_{\alpha,k-1})\left[\frac{n(n+1)}{12}\right]\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

$$= \sqrt{(9.21034)\frac{18(18+1)}{12}\left(\frac{1}{6}+\frac{1}{6}\right)}$$

$$= \sqrt{87.49823} = 9.35$$

$$\overline{R}_1 = \frac{90}{6} = 15 \qquad D_{1,2} = |15 - 9.33| = 5.67$$

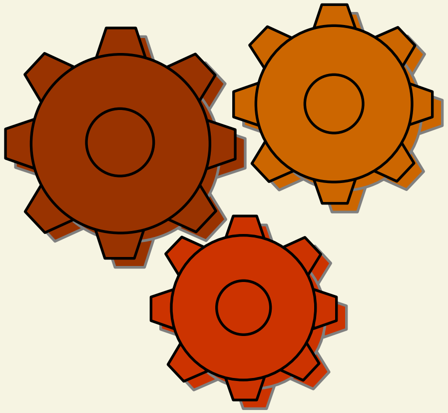
$$\overline{R}_2 = \frac{56}{6} = 9.33 \qquad D_{1,3} = |15 - 4.17| = 10.83***$$

$$\overline{R}_3 = \frac{25}{6} = 4.17 \qquad D_{2,3} = |9.33 - 4.17| = 5.16$$

Kruskal-Wallis *H*-Test Example

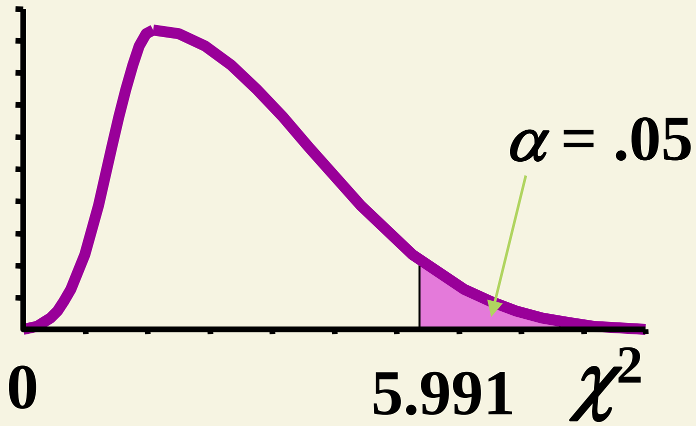
As production manager, you want to see if three filling machines have different filling times. You assign 15 similarly trained and experienced workers, 5 per machine, to the machines. At the **.05** level of significance, is there a difference in the **distribution** of filling times?

Mach1	Mach2	Mach3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40



Kruskal-Wallis *H*-Test Solution

- H_0: Identical Distrib.
- H_a: At Least 2 Differ
- $\alpha = .05$
- $df = p - 1 = 3 - 1 = 2$
- Critical Value(s):



Kruskal-Wallis *H*-Test Solution

Raw Data

Mach1	Mach2	Mach3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

Ranks

Mach1	Mach2	Mach3

Kruskal-Wallis *H*-Test Solution

Raw Data

Mach1	Mach2	Mach3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

Ranks

Mach1	Mach2	Mach3
		1

Kruskal-Wallis *H*-Test Solution

Raw Data			Ranks		
Mach1	Mach2	Mach3	Mach1	Mach2	Mach3
25.40	23.40	20.00			2
26.31	21.80	22.20			
24.10	23.50	19.75			1
23.74	22.75	20.60			
25.10	21.60	20.40			

Kruskal-Wallis *H*-Test Solution

Raw Data

Mach1	Mach2	Mach3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

Ranks

Mach1	Mach2	Mach3
		2
		1
		3

Kruskal-Wallis *H*-Test Solution

Raw Data

Mach1	Mach2	Mach3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

Ranks

Mach1	Mach2	Mach3
14	9	2
15	6	7
12	10	1
11	8	4
13	5	3

Kruskal-Wallis *H*-Test Solution

Raw Data				Ranks		
Mach1	Mach2	Mach3		Mach1	Mach2	Mach3
25.40	23.40	20.00		14	9	2
26.31	21.80	22.20		15	6	7
24.10	23.50	19.75		12	10	1
23.74	22.75	20.60		11	8	4
25.10	21.60	20.40		13	5	3
			Total	65	38	17

Kruskal-Wallis *H*-Test Solution

$$H = \left(\frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} \right) - 3(n+1)$$

$$= \left(\frac{12}{(15)(16)} \left(\frac{(65)^2}{5} + \frac{(38)^2}{5} + \frac{(17)^2}{5} \right) \right) - 3(16)$$

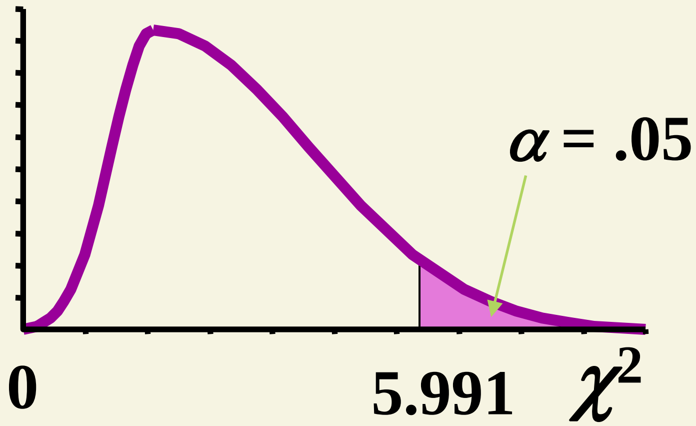
$$= \left(\frac{12}{240} \right) (191.6) - 48$$

$$= 11.58$$

$$H = \frac{12}{n(n+1)} \sum n_j \left(\bar{R}_j - \bar{R} \right)^2$$

Kruskal-Wallis H-Test Solution

- H_0: Identical Distrib.
- H_a: At Least 2 Differ
- $\alpha = .05$
- $df = p - 1 = 3 - 1 = 2$
- Critical Value(s):



Test Statistic:

$H = 11.58$

Decision:

Reject at $\alpha = .05$

Conclusion:

There is evidence population distrib. are different

14-5 The Wilcoxon Signed-Ranks Test (Paired Ranks)

The null and alternative hypotheses:

H0: The median difference between populations are 1 and 2 is zero

H1: The median difference between populations are 1 and 2 is not zero

Find the difference between the ranks for each pair, *D = x1 -x2*, and then rank the absolute values of the differences.

The Wilcoxon T statistic is the smaller of the sums of the positive ranks and the sum of the negative ranks:

$$T = \min\left(\sum(+), \sum(-)\right)$$

For small samples, a left-tailed test is used, using the values in Appendix C, Table 10.

$$E[T] = \frac{n(n+1)}{4} \qquad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

The large-sample test statistic:

$$z = \frac{T - E[T]}{\sigma_T}$$

Example 14-6

Sold (1)	Sold (2)	D=x1-x2	ABS(D)	Rank ABS(D)	Rank (D>0)	Rank (D<0)
56	40	16	16	9.0	9.0	0
48	70	-22	22	12.0	0.0	12
100	60	40	40	15.0	15.0	0
85	70	15	15	8.0	8.0	0
22	8	14	14	7.0	7.0	0
44	40	4	4	2.0	2.0	0
35	45	-10	10	6.0	0.0	6
28	7	21	21	11.0	11.0	0
52	60	-8	8	5.0	0.0	5
77	70	7	7	3.5	3.5	0
89	90	-1	1	1.0	0.0	1
10	10	0	*	*	*	*
65	85	-20	20	10.0	0.0	10
90	61	29	29	13.0	13.0	0
70	40	30	30	14.0	14.0	0
33	26	7	7	3.5	3.5	0
				Sum:	**86**	**34**

T=34

	n=15
P=0.05	30
P=0.025	25
P=0.01	20
P=0.005	16

H0 is not rejected (Note the arithmetic error in the text for store 13)

Example 14-7

Hourly Messages	Md0	D=x1-x2	ABS(D)	Rank ABS(D)	Rank (D>0)	Rank (D<0)
151	149	2	2	1.0	1.0	0.0
144	149	-5	5	2.0	0.0	2.0
123	149	-26	26	13.0	0.0	13.0
178	149	29	29	15.0	15.0	0.0
105	149	-44	44	23.0	0.0	23.0
112	149	-37	37	20.0	0.0	20.0
140	149	-9	9	4.0	0.0	4.0
167	149	18	18	10.0	10.0	0.0
177	149	28	28	14.0	14.0	0.0
185	149	36	36	19.0	19.0	0.0
129	149	-20	20	11.0	0.0	11.0
160	149	11	11	6.0	6.0	0.0
110	149	-39	39	21.0	0.0	21.0
170	149	21	21	12.0	12.0	0.0
198	149	49	49	25.0	25.0	0.0
165	149	16	16	8.0	8.0	0.0
109	149	-40	40	22.0	0.0	22.0
118	149	-31	31	16.5	0.0	16.5
155	149	6	6	3.0	3.0	0.0
102	149	-47	47	24.0	0.0	24.0
164	149	15	15	7.0	7.0	0.0
180	149	31	31	16.5	16.5	0.0
139	149	-10	10	5.0	0.0	5.0
166	149	17	17	9.0	9.0	0.0
82	149	33	33	18.0	18.0	0.0
				Sum:	**163.5**	**161.5**

$$E[T] = \frac{n(n+1)}{4} = \frac{(25)(25+1)}{4} = 162.5$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{25(25+1)((2)(25)+1)}{24}} = \sqrt{\frac{33150}{24}} = 37.165$$

The large - sample test statistic:

$$z = \frac{T - E[T]}{\sigma_T} = \frac{163.5 - 162.5}{37.165} = 0.027$$

H_0 cannot be rejected

14-6 The Kruskal-Wallis Test - A Nonparametric Alternative to One-Way ANOVA

The Kruskal-Wallis hypothesis test:

H0: All k populations have the same distribution

H1: Not all k populations have the same distribution

The Kruskal-Wallis test statistic:

$$H = \frac{12}{n(n+1)}\left(\sum_{j=1}^{k}\frac{R_j^2}{n_j}\right) - 3(n+1)$$

If each nj > 5, then H is approximately distributed as a χ^2.

Example 14-8: The Kruskal-Wallis Test

Software	Time	Rank	Group	RankSum
1	45	14	1	90
1	38	10	2	56
1	56	16	3	25
1	60	17		
1	47	15		
1	65	18		
2	30	8		
2	40	11		
2	28	7		
2	44	13		
2	25	5		
2	42	12		
3	22	4		
3	19	3		
3	15	1		
3	31	9		
3	27	6		
3	17	2		

$$H = \frac{12}{n(n+1)}\left(\sum_{j=1}^{k}\frac{R_j^2}{n_j}\right) - 3(n+1)$$

$$= \frac{12}{18(18+1)}\left(\frac{90^2}{6} + \frac{56^2}{6} + \frac{25^2}{6}\right) - 3(18+1)$$

$$= \left(\frac{12}{342}\right)\left(\frac{11861}{6}\right) - 57$$

$$= 12.3625$$

$\chi^2(2,0.005)=10.5966$, so H0 is rejected.

Further Analysis (Pairwise Comparisons of Average Ranks)

If the null hypothesis in the Kruskal-Wallis test is rejected, then we may wish, in addition, compare each pair of populations to determine which are different and which are the same.

The pairwise comparison test statistic:

$$D = \left| \bar{R}_i - \bar{R}_j \right|$$

where $\bar{R}_i$ is the mean of the ranks of the observations from population i.

The critical point for the paired comparisons:

$$C_{KW} = \sqrt{(\chi^2_{\alpha,k-1})\left[\frac{n(n+1)}{12}\right]\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Reject if $D > C_{KW}$

Pairwise Comparisons: Example 14-8

Critical Point:

$$C_{KW} = \sqrt{(\chi^2_{\alpha,k-1})\left[\frac{n(n+1)}{12}\right]\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

$$= \sqrt{(9.21034)\frac{18(18+1)}{12}\left(\frac{1}{6}+\frac{1}{6}\right)}$$

$$= \sqrt{87.49823} = 9.35$$

$$\bar{R}_1 = \frac{90}{6} = 15 \qquad D_{1,2} = |15 - 9.33| = 5.67$$

$$\bar{R}_2 = \frac{56}{6} = 9.33 \qquad D_{1,3} = |15 - 4.17| = 10.83***$$

$$\bar{R}_3 = \frac{25}{6} = 4.17 \qquad D_{2,3} = |9.33 - 4.17| = 5.16$$

Rank Correlation

Spearman's Rank Correlation Coefficient

- Measures correlation between **ranks**
- Corresponds to Pearson product moment correlation coefficient
- Values range from –1 to +1
- Formula (shortcut)

$$r_s = 1 - \frac{6\sum d_i^{\,2}}{n\left(n^2 - 1\right)}$$

$d_i = u_i - v_i$ (difference in ranks of *i*th observation for samples 1 and 2)

n = number of pairs of observations

Spearman's Rank Correlation Procedure

1. Assign ranks, R_i , to the observations of each variable separately
2. Calculate differences, d_i , between each pair of ranks
3. Square differences, d_i^2, between ranks
4. Sum squared differences for each variable
5. Use shortcut approximation formula

Spearman's Rank Correlation Example

You're a research assistant for the FBI. You're investigating the relationship between a person's attempts at deception and percent changes in their pupil size. You ask subjects a series of questions, some of which they must answer dishonestly. At the **.05** level of significance, what is the **correlation coefficient**?

Subj.	Deception	Pupil
1	87	10
2	63	6
3	95	11
4	50	7
5	43	0

Spearman's Rank Correlation Table

Subj.	Decep.	R_{1i}	Pupil	R_{2i}	d_i	d_i^2
1	87	4	10	4	0	0
2	63	3	6	2	1	1
3	95	5	11	5	0	0
4	50	2	7	3	-1	1
5	43	1	0	1	0	0
					$\Sigma d_i^2 = 2$	

Spearman's Rank Correlation Solution

$$r_s = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)}$$

$$= 1 - \frac{6\left(2\right)}{5\left(5^2 - 1\right)}$$

$$= 1 - 0.10$$

$$= 0.90$$

Spearman's Nonparametric Test for Rank Correlation

One-Tailed Test

H_0: $\rho = 0$

H_a: $\rho > 0$ (or H_a: $\rho < 0$)

Test statistic: $r_s = 1 - \dfrac{6\sum d_i^{\,2}}{n\left(n^2 - 1\right)}$

Rejection region: $r_s > r_{s,\alpha}$ or $r_s < -r_{s,\alpha}$ when $\rho_s < 0$ where $r_{s,\alpha}$ is the value from Table XVI corresponding to the upper-tail area α and n pairs of observations

Spearman's Nonparametric Test for Rank Correlation

Ties: Assign tied measurements the average of the ranks they would receive if they were unequal but occurred in successive order. For example, if the third-ranked and fourth-ranked measurements are tied, assign each a rank of $(3 + 4)/2 = 3.5$. The number should be small relative to the total number of observations.

Spearman's Nonparametric Test for Rank Correlation

Two-Tailed Test

H_0: $\rho = 0$

H_a: $\rho \neq 0$

Test statistic:

$$r_s = 1 - \frac{6\sum d_i^{\,2}}{n\left(n^2 - 1\right)}$$

Rejection region: $|r_s| > r_{s,\alpha/2}$

where $r_{s,\alpha/2}$ is the value from Table XVI corresponding to the upper-tail area $\alpha/2$ and n pairs of observations

Spearman's Nonparametric Test for Rank Correlation

Ties: Assign tied measurements the average of the ranks they would receive if they were unequal but occurred in successive order. For example, if the third-ranked and fourth-ranked measurements are tied, assign each a rank of (3 + 4)/2 = 3.5. The number should be small relative to the total number of observations.

Conditions Required for a Valid Spearman's Test

1. The sample of experimental units on which the two variables are measured is randomly selected.
2. The probability distributions of the two variables are continuous.

14-8 The Spearman Rank Correlation Coefficient

The **Spearman Rank Correlation Coefficient** is the simple correlation coefficient calculated from variables converted to ranks from their original values.

The Spearman Rank Correlation Coefficient (assuming no ties):

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \qquad \text{where } d_i = R(x_i) - R(y_i)$$

Null and alternative hypotheses:

$$H_0: \rho_s = 0$$
$$H_1: \rho_s \neq 0$$

Critical values for small sample tests from Appendix C, Table 11

Large sample test statistic:

$$z = r_s\sqrt{(n-1)}$$

Spearman Rank Correlation Coefficient: Example 14-11

MMI	S&P100	R-MMI	R-S&P	Diff	Diffsq
220	151	7	6	1	1
218	150	5	5	0	0
216	148	3	3	0	0
217	149	4	4	0	0
215	147	2	2	0	0
213	146	1	1	0	0
219	152	6	7	-1	1
236	165	9	10	-1	1
237	162	10	9	1	1
235	161	8	8	0	0
				Sum:	**4**

Table 11: α=0.005

n	
⋮	
7	------
8	0.881
9	0.833
10	0.794
11	0.818
⋮	

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{(6)(4)}{(10)(10^2-1)} = 1 - \frac{24}{990} = 0.9758 > 0.794 \quad \mathbf{H_0 \ rejected}$$