Distribution-Free Tests

Dr.Mamatha.H.R

Professor

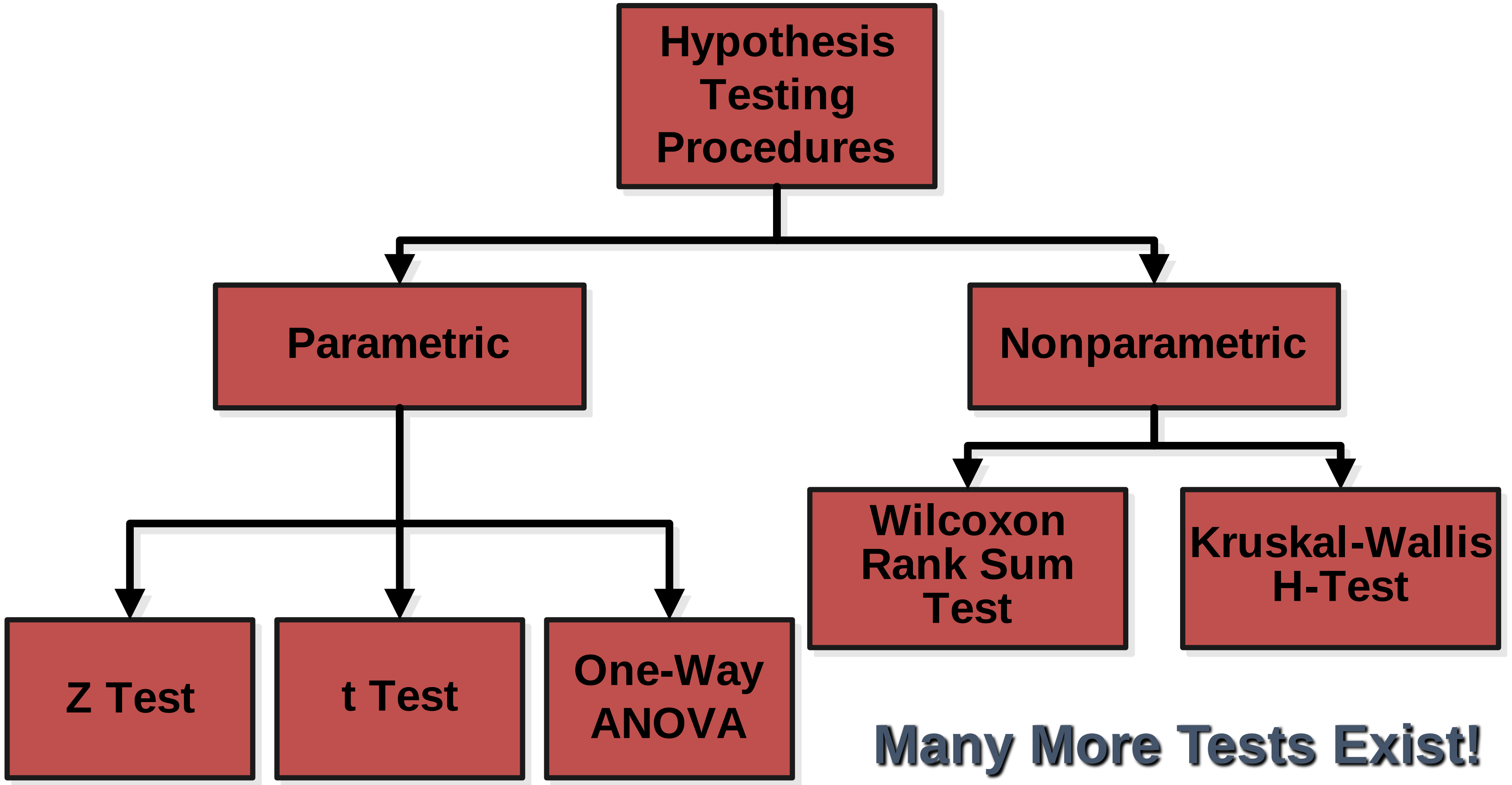
Department of Computer Science and Engineering

PES University

Bangalore

Course material created using various Internet resources and text book

Hypothesis Testing Procedures



Parametric vs Nonparametric Statistics

- Parametric tests usually based upon certain assumptions about the population from which the samples were drawn or picked.
  - Very famous assumption is the normality assumption that is data being analyzed are randomly selected from a normally distributed population.
  - Important restriction is that parametric test usually requires quantitative measurement that yield interval or ratio level data.

- Nonparametric Statistics *are based on fewer assumptions about the population and the parameters.*
  - Sometimes called "distribution-free" statistics.
  - A variety of nonparametric statistics are available for use with nominal or ordinal data.

Definition

- Nonparametric methods : rank-based methods are used when we have no idea about the population distribution from which the data is sampled.
- Used for small sample sizes.
- Used when the data are measured on an ordinal scale and only their ranks are meaningful.

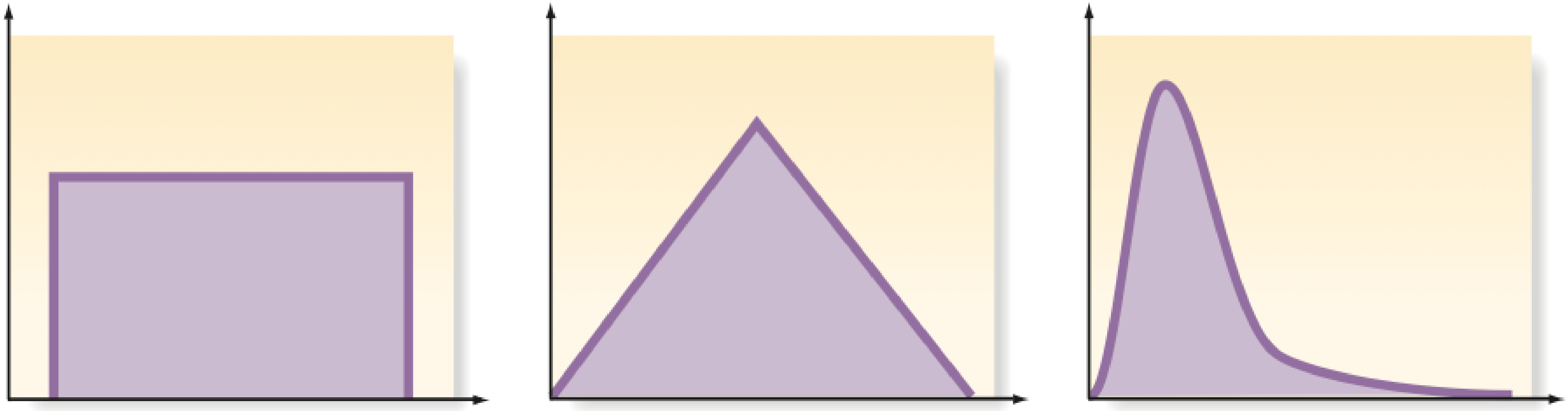
Parametric Test Procedures

- Involve population parameters
  - Example: population mean
- Require interval scale or ratio scale
  - Whole numbers or fractions
  - Example: height in inches (72, 60.5, 54.7)
- Have stringent assumptions
  - Example: normal distribution
- Examples: *z*-test, *t*-test, *F*-test, χ^2-test

Nonparametric Test Procedures

- Do not involve population parameters
  - Example: probability distributions, independence
- Data measured on any scale
  - Ratio or interval
  - Ordinal
    - Example: good-better-best
  - Nominal
    - Example: male-female
- Example: Wilcoxon rank sum test

Nonnormal Distributions - *t*-Statistic is Invalid



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

Distribution-Free Tests

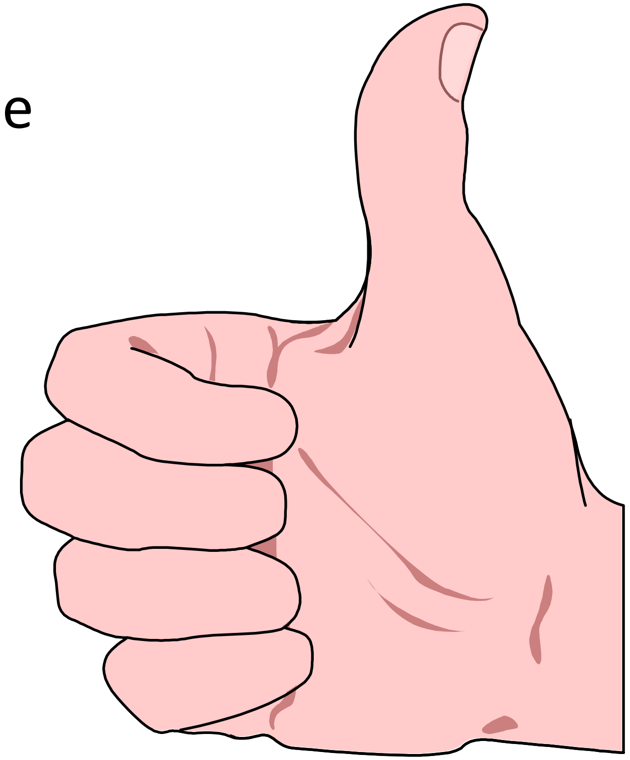
Distribution-free tests are statistical tests that do not rely on any underlying assumptions about the probability distribution of the sampled population.

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**).

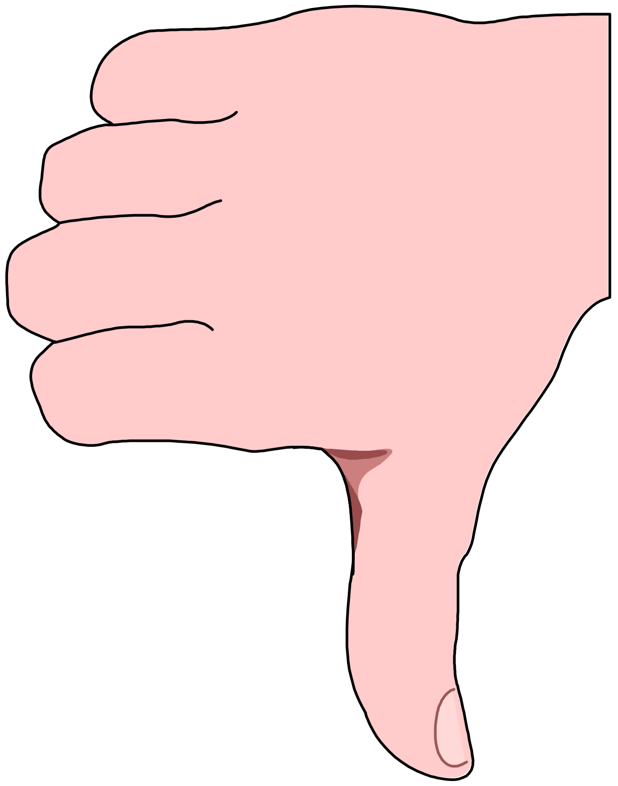
Advantages of Nonparametric Tests

- 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



Frequently Used Nonparametric Tests

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

Some test (parametric and non parametric)

Testing pattern	Parametric	Non-parametric
One sample / single population	T-test ✓	Wilcoxon signed rank test
Two samples (Two Pop.)	Two sample t-test	Wilcoxon rank sum test
Matched pairs	Apply one sample t test $t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}}$ with $\nu = n-1$	Apply one sample test for difference between pairs
Several independent samples Imp	One way ANOVA F-test	Kruskal Wallis Test

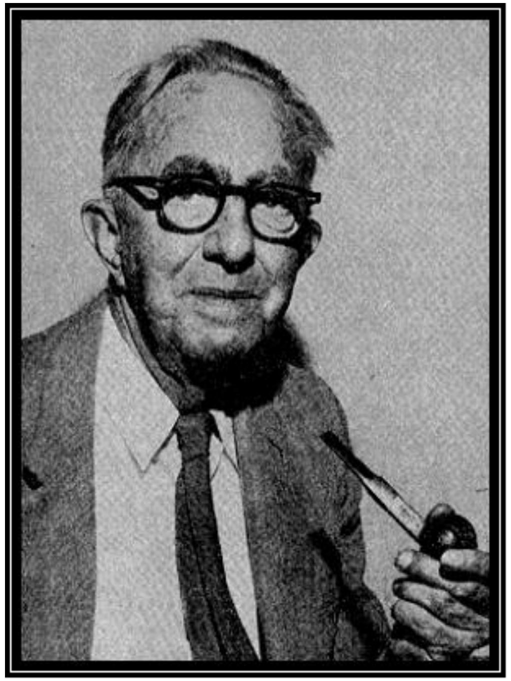
Types of non-parametric Tests

- Sign test for paired data and one sample sign test.
  - Positive or negative values are substituted for quantitative values.
- Mann-Whitney U Test or a Rank Sum Test.
  - Where two independent samples have been drawn from the same population.
- Kruskal Wallis Test.
  - Rank sum test which generalizes the ANOVA.
- One Sample Run Test.
  - Determine the randomness of samples.
- Rank correlation test.
- Kolmogorov-Simrnov Test.
- Kendal Test of Concordance.
- Median Test for Two independent samples.
- Wilcoxon signed rank test.

Wilcoxon signed rank test

Inventor

Frank Wilcoxon (2 September 1892 in County Cork, Ireland – 18 November 1965, Tallahassee, Florida, USA) was a chemist and statistician, known for development of several statistical tests.



What is it used for?

- Two related samples
- Matched samples
- Repeated measurements on a single sample

Example

- The nickel content, in parts per thousand by weight, is measured for six welds. The results are 9.3, 0.9, 9.0, 21.7, 11.5, and 13.9. Let
- μ represent the mean nickel content for this type of weld. It is desired to test $H0 : \mu \geq 12$ versus $H1 : \mu < 12$.

- The Student's t test is not appropriate, because there are two outliers, 0.9 and 21.7, which indicate that the population is not normal.

- The Wilcoxon signed-rank test can be used in this situation.

- This test does not require the population to be normal.
- It does, however, require that the population be continuous (rather

  than discrete), and that the probability density function be symmetric. (The normal is a special case of a continuous symmetric population.)

- The given sample clearly comes from a continuous population, and the presence of outliers on either side make it reasonable to assume that the population is approximately symmetric as well.

- Under $H0$, the population mean is $\mu = 12$. Since the population is assumed to be symmetric, the population median is 12 as well.

- To compute the rank-sum statistic,

- subtract 12 from each sample observation to obtain differences.
- The difference closest to 0, ignoring sign, is assigned a rank of 1.
- The difference next closest to 0, again ignoring sign, is assigned a rank of 2, and so on.
- Finally, the ranks corresponding to negative differences are given negative signs.
- Denote the sum of the positive ranks $S+$ and the sum of the absolute values of the negative ranks $S-$.
- Either $S+$ or $S-$ may be used as a test statistic;
- we shall use $S+$

x	$x - 12$	Signed Rank
11.5	–0.5	–1
13.9	1.9	2
9.3	–2.7	–3
9.0	–3.0	–4
21.7	9.7	5
0.9	–11.1	–6

- In this example
- $S+ = 2+5 = 7$, and $S- = 1+3+4+6 = 14$.

- Note that since the sample size is 6,
- $S+ + S- = 1 + 2 + 3 + 4 + 5 + 6 = 21$.

- For any sample, it is the case that
- $S+ + S- = 1+2+\cdots+n = n(n+1)/2$.

- In some cases, where there are many more positive ranks than negative ranks, it is easiest to first compute $S-$ by summing the negative ranks and then computing
- $S+ = n(n + 1)/2 - S-$.

- how *S*+ can be used as a test statistic.

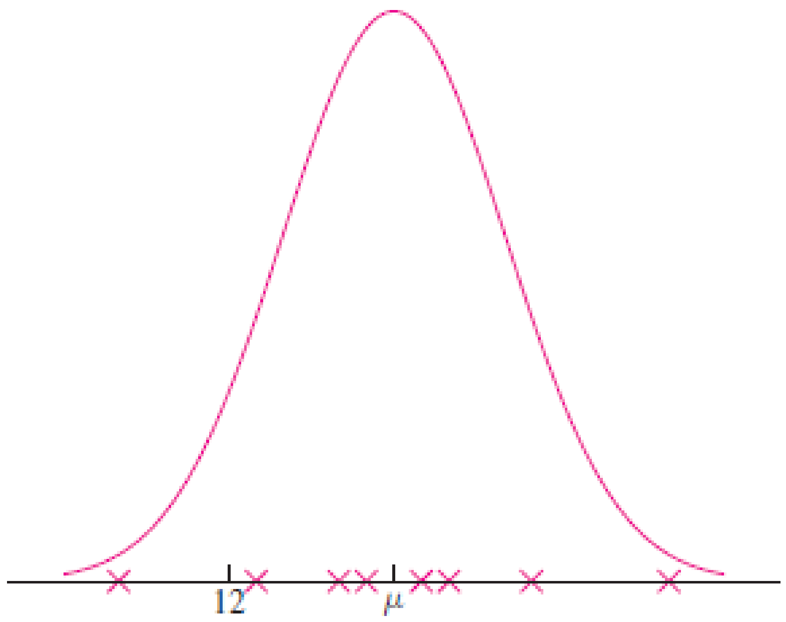


FIGURE 6.17 The true median is greater than 12. Sample observations are more likely to be above 12 than below 12. Furthermore, the observations above 12 will tend to have larger differences from 12 than the observations below 12. Therefore S_+ is likely to be large.

- In Figure 6.17, $\mu > 12$. For this distribution, positive differences are more probable than negative differences and tend to be larger in magnitude as well.
- Therefore it is likely that the positive ranks will be greater both in number and in magnitude than the negative ranks, so *S*+ is likely to be large.

- In Figure 6.18, $\mu < 12$, and the situation is reversed.
- Here the positive ranks are likely to be fewer in number and smaller in magnitude, so $S+$ is likely to be small.

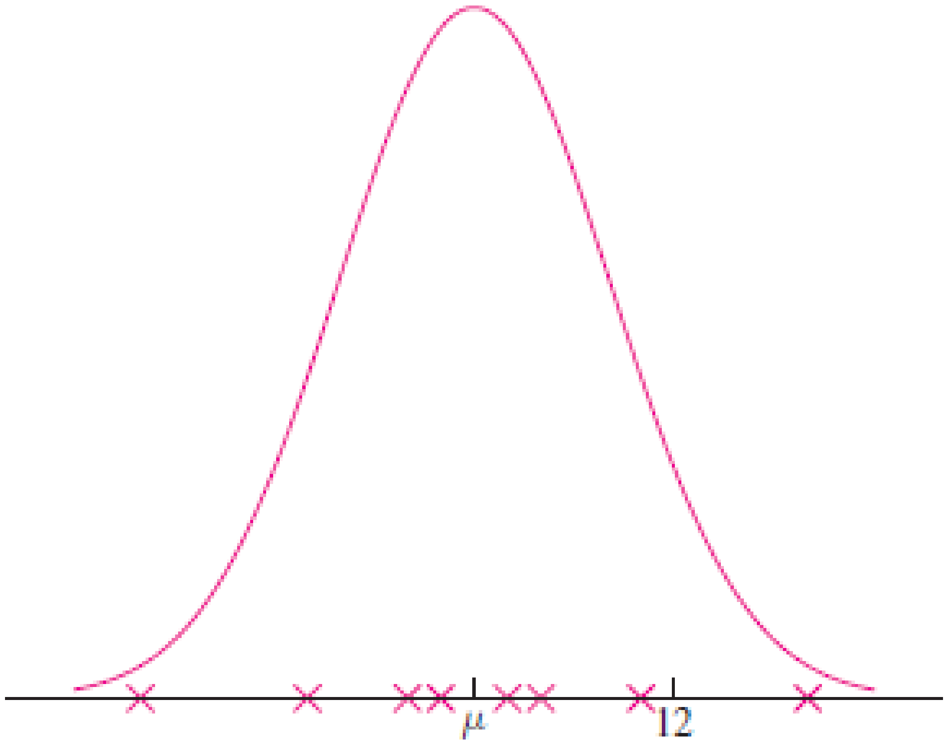


FIGURE 6.18 The true median is less than 12. Sample observations are more likely to be below 12 than above 12. Furthermore, the observations below 12 will tend to have larger differences from 12 than the observations above 12. Therefore S_+ is likely to be small.

- in general,
- large values of $S+$ will provide evidence against a null
- hypothesis of the form $H0 : \mu \leq \mu 0$,

- while small values of $S+$ will provide evidence against a null hypothesis of the form $H0 : \mu \geq \mu 0$.

- the null hypothesis is $H0 : \mu \geq 12$, so a small value of $S+$ will
- provide evidence against $H0$.
- We observe $S+ = 7$.
- The P-value is the probability of observing a value of $S+$ that is less than or equal to 7 when $H0$ is true.
- Table presents certain probabilities for the null distribution of $S+$.
- Consulting this table with the sample size $n = 6$, we find that the probability of observing a value of 4 or less is 0.1094.
- The probability of observing a value of 7 or less must be greater

  than this, so we conclude that $P > 0.1094$, and thus do not reject $H0$

Hypothesis

- Assumption: Population must be symmetric
- Reason: under the null hypothesis the right skewed population tends to have higher value of w_+ and the left skewed population tends to have higher value of w_-.
- Under this assumption, mean equals median.
- H_0: $\mu = \mu_0$
- H_a: $\mu > \mu_0$ ($\mu \neq \mu_0$ or $\mu < \mu_0$)

Testing procedure

- Compute the differences d_i
- Rank d_i in terms of absolute values
- Statistics:

  $w_+ =$ sum of the ranks of positive differences

  $w_- =$ sum of the ranks of negative differences
- Rejection region:

  Reject H_0 if w_+ is large or if w_- is small,

  or if $\mathrm{P} - value = P(W > w_+) < \alpha$

Examples

- In the example discussed previously, the nickel content for six welds was measured to be 9.3, 0.9, 9.0, 21.7, 11.5, and 13.9. Use these data to test $H0 : \mu \leq 5$ versus $H1 : \mu > 5$.

x	$x - 5$	Signed Rank
9.0	4.0	1
0.9	−4.1	−2
9.3	4.3	3
11.5	6.5	4
13.9	8.9	5
21.7	16.7	6

- The observed value of the test statistic is $S+ = 19$.
- Since the null hypothesis is of the form $\mu \le \mu 0$, large values of $S+$ provide evidence against $H0$.
- Therefore the P-value is the area in the right-hand tail of the null distribution, corresponding to values greater than or equal to 19.

- Consulting Table A.5 shows that the P-value is 0.0469

- Use the data in previous Example to test $H0 : \mu = 16$ versus $H1 : \mu \neq 16$.

x	$x - 16$	Signed Rank
13.9	−2.1	−1
11.5	−4.5	−2
21.7	5.7	3
9.3	−6.7	−4
9.0	−7.0	−5
0.9	−15.1	−6

- null hypothesis is of the form $H0 : \mu = \mu 0$, this is a two-tailed test.
- The observed value of the test statistic is $S+ = 3$.
- Consulting Table A.5, we find that the area in the left-hand tail, corresponding to values less than or equal to 3, is 0.0781.
- The *P*-value is twice this amount, since it is the sum of areas in two equal tails.
- Thus the *P*-value is $2(0.0781) = 0.1562$.

Ties

- Sometimes two or more of the quantities to be ranked have exactly the same value.
- Such quantities are said to be tied.
- The standard method for dealing with ties is to assign to each tied observation the average of the ranks they would have received if they had differed slightly.
- For example, the quantities 3, 4, 4, 5, 7 would receive the ranks 1, 2.5, 2.5, 4, 5,
- and the quantities 12, 15, 16, 16, 16, 20 would receive the ranks 1, 2, 4, 4, 4, 6.

Differences of Zero

- If the mean under $H0$ is $\mu 0$, and one of the observations is equal to $\mu 0$, then its difference is 0, which is neither positive nor negative.

- An observation that is equal to $\mu 0$ cannot receive a signed rank.
- The appropriate procedure is to drop such observations from the
- sample altogether, and to consider the sample size to be reduced by the number of these observations.

- Use the data in previous Example to test $H0 : \mu = 9$ versus $H1 : \mu \neq 9$.

x	$x - 9$	Signed Rank
9.0	0.0	—
9.3	0.3	1
11.5	2.5	2
13.9	4.9	3
0.9	−8.1	−4
21.7	12.7	5

- The value of the test statistic is $S+ = 11$.
- The sample size for the purposes of the test is 5, since the value 9.0 is not ranked.
- Entering Table A.5 with sample size 5, we find that if $S+ = 12$, the P-value would be $2(0.1562) = 0.3124$.
- We conclude that for $S+ = 11$, $P > 0.3124$.

Example

- Test the thermostat setting data if the median setting differs from 200˚F

Table 1.1 Thermostat setting data

x	202.2	203.4	200.5	202.5	206.3	198.0	203.7	200.8	201.3	199.0
diff.	2.2	3.4	0.5	2.5	6.3	-2.0	3.7	0.8	1.3	-1.0
rank	6	8	1	7	10	5	9	2	4	3

- H_0: $\mu = 200$ vs H_a: $\mu \neq 200$
- $w_{-} = 5 + 3 = 8$
- $w_{+} = 6 + 8 + 1 + 7 + 10 + 9 + 2 + 4 = 47$
- $\mathrm{P} - value = 2P(W > 47) < 2 \times 0.024 = 0.048$
- Conclusion: The population median differs from the design setting of 200˚F at $\alpha = 0.05$

Large sample approximation

When the sample size *n* is large, the test statistic *S*+ is approximately normally distributed.

A rule of thumb is that the normal approximation is good if $n > 20$. It can be shown by advanced methods that under *H*0

- *S*+ has mean $=E(S+)=n(n+1)/4$
- Variance$=n(n+1)(2n+1)/24$.

$$z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

The Wilcoxon signed-rank test is performed by computing the *z*-score of *S*+, and then using the normal table to find the *P*-value

- The article “Exact Evaluation of Batch-Ordering Inventory Policies in Two-Echelon Supply Chains with Periodic Review” (G. Chacon, *Operations Research*, 2001:79–98) presents an evaluation of a reorder point policy, which is a rule for determining when to restock an inventory. Costs for 32 scenarios are estimated.

- Let μ represent the mean cost. Test $H0 : \mu \geq 70$ versus $H1 : \mu < 70$.
- The data, along with the differences and signed ranks, are presented in Table

TABLE 6.1 Data for Example 6.18

x	$x - 70$	Signed Rank	x	$x - 70$	Signed Rank	x	$x - 70$	Signed Rank
79.26	9.26	1	30.27	–39.73	–12	11.48	–58.52	–23
80.79	10.79	2	22.39	–47.61	–13	11.28	–58.72	–24
82.07	12.07	3	118.39	48.39	14	10.08	–59.92	–25
82.14	12.14	4	118.46	48.46	15	7.28	–62.72	–26
57.19	–12.81	–5	20.32	–49.68	–16	6.87	–63.13	–27
55.86	–14.14	–6	16.69	–53.31	–17	6.23	–63.77	–28
42.08	–27.92	–7	16.50	–53.50	–18	4.57	–65.43	–29
41.78	–28.22	–8	15.95	–54.05	–19	4.09	–65.91	–30
100.01	30.01	9	15.16	–54.84	–20	140.09	70.09	31
100.36	30.36	10	14.22	–55.78	–21	140.77	70.77	32
30.46	–39.54	–11	11.64	–58.36	–22			

- **Solution**
- The sample size is $n = 32$, so
- the mean is $n(n + 1)/4 = 264$ and
- the variance is
- $n(n+1)(2n+1)/24 = 2860$.

- The sum of the positive ranks is $S+ = 121$.
- We compute

- $z = 121 - 264/ \sqrt{2860} = -2.67$

- Since the null hypothesis is of the form $H0 : \mu \geq \mu 0$, small values of $S+$ provide
- evidence against $H0$.
- Thus the P-value is the area under the normal curve to the left of $z = -2.67$. This area, and thus the P-value, is 0.0038.

The Wilcoxon Rank-Sum Test

- The Wilcoxon rank-sum test, also called the Mann–Whitney test, can be used to test the difference in population means in certain cases where the populations are not normal.
- Two assumptions are necessary.
- First the populations must be continuous.
- Second, their probability density functions must be identical in shape and size; the only possible difference between them being their location.

- To describe the test, let $X_1, \ldots, X_m$ be a random sample from one population and let $Y_1, \ldots, Y_n$ be a random sample from the

  other.
- We adopt the notational convention that when the sample sizes are unequal, the smaller sample will be denoted $X_1, \ldots, X_m$.
- Thus the sample sizes are m and n, with $m \leq n$.
- Denote the population means by μX and μY , respectively.

- The test is performed by ordering the $m + n$ values obtained by combining the two samples, and assigning ranks $1, 2, \ldots, m + n$ to them.
- The test statistic, denoted by W, is the sum of the ranks corresponding to $X_1, \ldots, X_m$.
- Since the populations are identical with the possible exception of location, it follows that if $\mu_X < \mu_Y$, the values in the X sample will tend to be smaller than those in the Y sample, so the rank sum W will tend

  to be smaller as well.
- By similar reasoning, if $\mu_X > \mu_Y$, W will tend to be larger.

- more generally, that we have samples of observations from each of two populations A and B containing nA and nB observations respectively.
- We wish to test the hypothesis that the distribution of X-measurements in population A is the same as that in B, which we will write symbolically as H0 : A = B.
- The departures from H0 that the Wilcoxon test tries to detect are location shifts.

- If we expect to detect that the distribution of A is shifted to the right of distribution B as in Fig. , we will write this as H1 : A > B.
- The other two possibilities are H1 : A< B. (A is shifted to the left of B),
- the two sided-alternative, which we will write as H1 : A ≠ B, for situations in which we have no strong prior reason for expecting a shift in a particular direction.

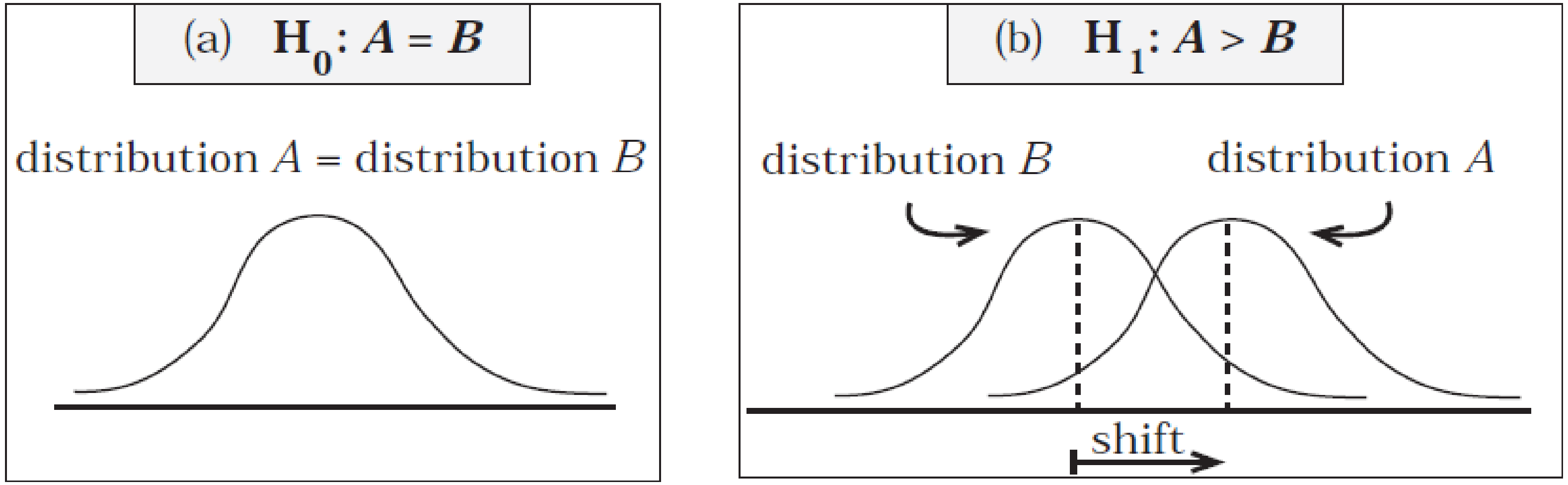


Figure 2 : Illustration of $H_0 : A = B$ versus $H_1 : A > B$.

- Resistances, in mΩ, are measured for five wires of one type and six wires of another type. The results are as follows:
- X: 36 28 29 20 38
- Y: 34 41 35 47 49 46
- Use the Wilcoxon rank-sum test to test $H0 : \mu X \geq \mu Y$ versus $H1 : \mu X < \mu Y$.

Value	Rank	Sample	Value	Rank	Sample
20	1	X	38	7	X
28	2	X	41	8	Y
29	3	X	46	9	Y
34	4	Y	47	10	Y
35	5	Y	49	11	Y
36	6	X			

- The test statistic W is the sum of the ranks corresponding to the X values,
- so $W = 1+2+3+6+7 = 19$.
- To determine the P-value,we consult Table A.6 (in Appendix A).
- We note that small values of W provide evidence against $H0 : \mu X \geq \mu Y$, so the P value is the area in the left-hand tail of the null distribution.
- Entering the table with $m = 5$ and $n = 6$ we find that the area to the left of $W = 19$ is 0.0260. This is the P-value.

Large-Sample Approximation

- When both sample sizes *m* and *n* are greater than 8, it can be shown by advanced methods that the null distribution of the test statistic *W* is approximately normal with
- Mean $= m(m+n+1)/2$ and
- variance $= mn(m+n+1)/12$.
- In these cases the test is performed by computing the *z*-score of *W*, and then using the normal table to find the *P*-value.

z-score is

$$z = \frac{W - m(m+n+1)/2}{\sqrt{mn(m+n+1)/12}}$$

- The article “Cost Analysis Between SABER and Design Bid Build Contracting Methods” (E. Henry and H. Brothers, *Journal of Construction Engineering and Management*, 2001:359–366) presents data on construction costs for 10 jobs bid by the traditional method (denoted X) and 19 jobs bid by an experimental system (denoted Y). The data, in units of dollars per square meter, and their ranks, are presented in Table. Test $H0 : \mu X \leq \mu Y$ versus $H1 : \mu X > \mu Y$.

Value	Rank	Sample	Value	Rank	Sample
57	1	X	613	16	X
95	2	Y	622	17	Y
101	3	Y	708	18	X
118	4	Y	726	19	Y
149	5	Y	843	20	Y
196	6	Y	908	21	Y
200	7	Y	926	22	X
233	8	Y	943	23	Y
243	9	Y	1048	24	Y
341	10	Y	1165	25	X
419	11	Y	1293	26	X
457	12	X	1593	27	X
584	13	X	1952	28	X
592	14	Y	2424	29	Y
594	15	Y			

- **Solution**
- The sum of the X ranks is $W = 1+12+13+16+18+22+25+26+27+28 = 188$.
- The sample sizes are $m = 10$ and $n = 19$.
- We use the normal approximation and compute
- $z = 188 - 10(10 + 19 + 1)/2 \sqrt{10(19)(10 + 19 + 1)/12} = 1.74$
- Large values of W provide evidence against the null hypothesis.
- Therefore the P-value is the area under the normal curve to the right of
- $z = 1.74$.
- From the z table
- we find that the P-value is 0.0409.

Distribution-Free Methods Are Not Assumption-Free

- The necessary assumptions are actually rather restrictive.

- symmetry for the signed-rank test
- identical shapes and spreads for the rank-sum test