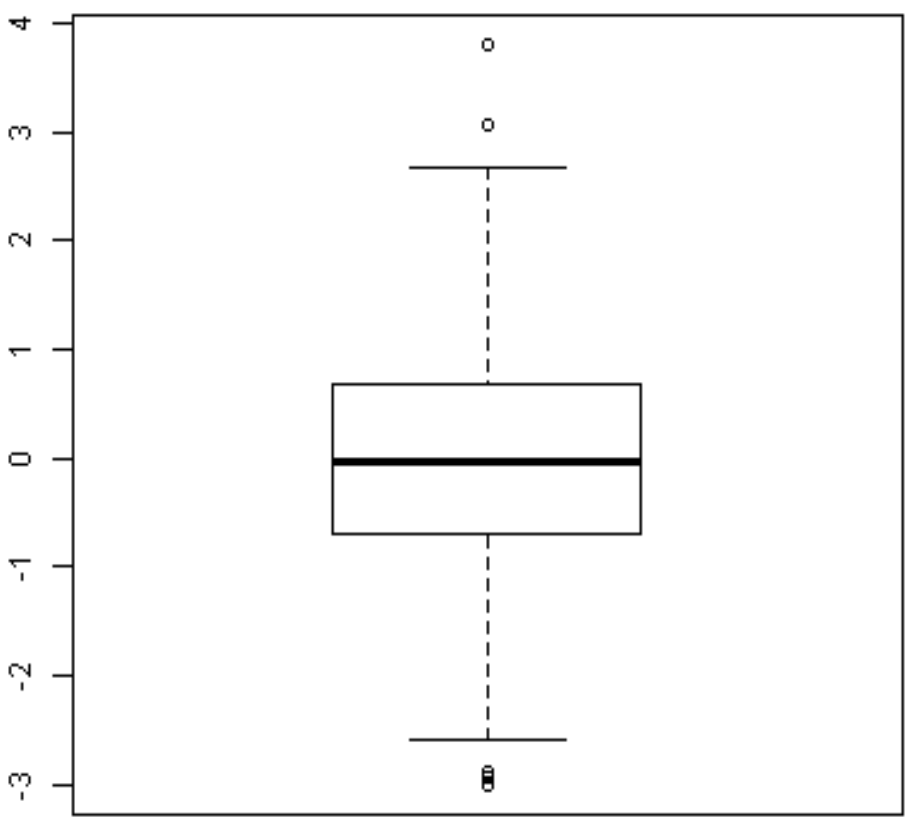
Distribution-Free Hypothesis Tests

Keegan Korthauer
Department of Statistics
UW Madison

Motivation



We want to perform a HT using a small sample

Problems:

- Boxplot shows that the sample does not come from a normal distribution (outliers)
- We have not yet learned any methods that will apply in this situation

Distribution-Free Methods

- When we can't assume normality, we can rely on a class of methods that are **distribution-free**
  - Also called **nonparametric**
  - They do not require that the population distribution has a specific form (i.e. normal) nor do they rely on the CLT
  - They do require some basic assumptions about the population distribution, however
- We will learn about two such methods:
  - **Wilcoxon signed-rank test**: One-sample test for a population mean
  - **Wilcoxon rank-sum test**: Two-sample test for the difference in population means

Wilcoxon Signed-Rank Test

- Analogous to the one-sample t-test in Sec. 6.4

- Use when sample cannot be assumed to come from a normal distribution

- Two assumptions are necessary:
  1. Population is continuous
  2. Probability density function is symmetric

Wilcoxon Signed-Rank Test: The Idea

- **Parametric t-test**: test statistic follows a *t* distribution under the null and measures the deviation of the *sample mean* from the hypothesized value μ_0

- **Nonparametric Wilcoxon Signed-Rank Test**:
  - If the distribution is symmetric, then the mean will be equal to the median
  - Test statistic measures the deviation of the *median* from the hypothesized value μ_0
  - If most of the observed sample values fall on one side of μ_0 then reject the null hypothesis; median is likely not equal to μ_0
  - If about half of the observed sample values fall on each side of μ_0 then do not reject the null hypothesis; it is plausible that the median is equal to μ_0

Wilcoxon Signed-Rank Test: The Idea

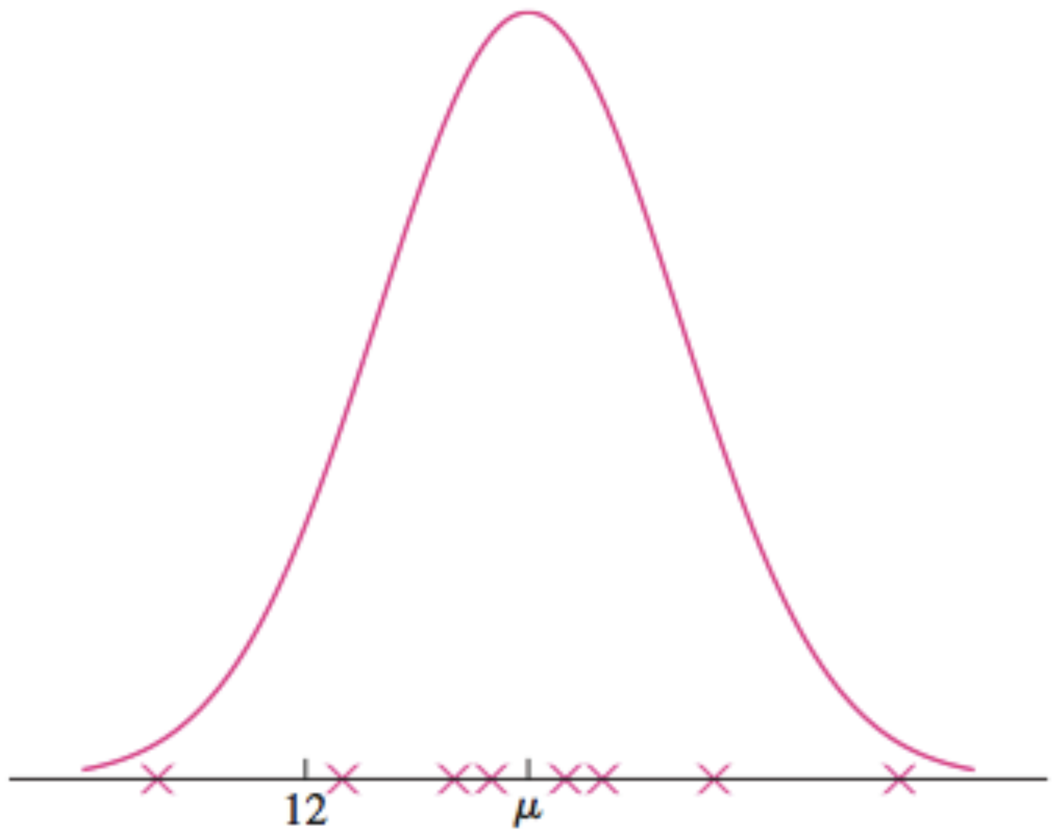
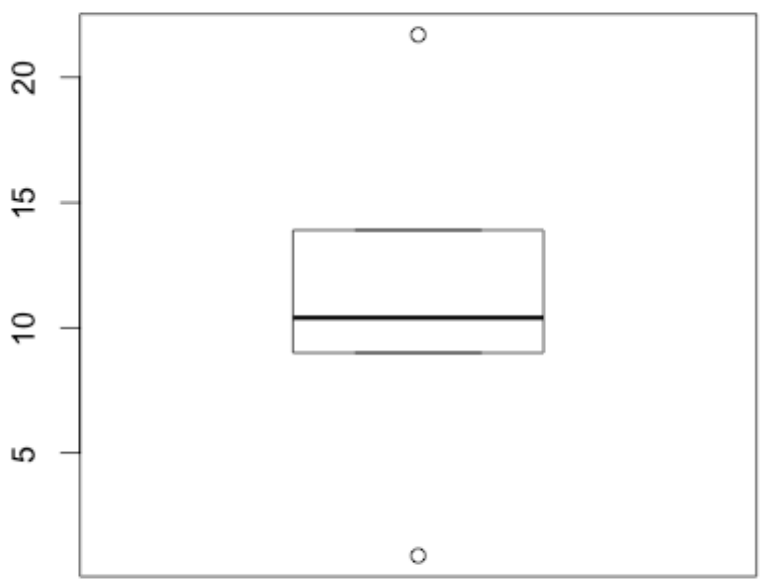


FIGURE 6.17 The true median is greater than 12. Sample observations are more likely to be above 12 than below 12. Furthermore, the observations above 12 will tend to have larger differences from 12 than the observations below 12. Therefore S_+ is likely to be large.

How to Compute: Example



- The nickel content, in parts per thousand by weight, is measured for 6 welds: 9.3, 0.9, 9.0, 21.7, 11.5, 13.9
- Let μ represent the mean nickel content for this type of weld
- We wish to test $H_0: \mu \geq 12$ vs $H_1: \mu < 12$
- Student's t test is not appropriate (two outliers)
- Population can be assumed to be continuous
- Also looks reasonably symmetric

Assumptions met; can use Wilcoxon signed-rank test

How to Compute the Test Statistic

- Under H_0, mean = median = μ_0
- First take the difference of each observation from μ_0
- Next assign each observation a **rank:**
  - The difference closest to zero (ignoring sign) is rank 1
  - Next closest to zero (ignoring sign) is rank 2
  - … and so on
- Next add the appropriate sign back to the rank value
- Finally, sum up the positive ranks (S_+) and the absolute values of the negative ranks (S_-)
- Compare either S_+ or S_- to the null distribution in Table A.5 to obtain a tail probability (we'll use S_+)

How to Compute: Example

- In this example, $\mu_0 = 12$ so we compute the ranks as follows:

x	$x - 12$	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

- This leads to $S_+ = 2 + 5 = 7$ and $S_- = 1 + 3 + 4 = 8$
- Note that $S_+ + S_-$ will always equal $1 + 2 + \ldots + n = n(n+1)/2$
- Looking at Table A.5, we find that the probability of observing a S_+ of 4 or less is 0.1094
  - The probability of observing a S_+ of 7 or less must be even greater than this -> do not reject H_0

Notes

- To find the p-value for a two sided Wilcoxon signed-rank test, multiply the tail area by two

- How to handle ties in rank:
  - Assign the average value of the ranks they would receive if they were not tied

    4, 5, 6, 6, 7 -> ranks 1, 2, 3.5, 3.5 5

- How to handle differences of zero
  - Neither above nor below μ_0, so remove them entirely from the sample (decrease n appropriately)

Summary: Wilcoxon Signed-Rank Test

Let $X_1,\dots,X_n$ be a small ($n < 30$) sample from a **continuous, symmetric** population with mean μ

1. Set up the null H_0 and alternative H_1 hypotheses (see table below)
2. State the level of significance α you will use
3. Calculate the test statistic S_+ = sum of the positive ranks of the differences of each observation from μ_0
4. Assume H_0 is true and look up the P-value using Table A.5

H_0	H_1	P-value
$\mu \le \mu_0$	$\mu > \mu_0$	Area to the right of S_+
$\mu \ge \mu_0$	$\mu < \mu_0$	Area to the left of S_+
$\mu = \mu_0$	$\mu \ne \mu_0$	2*(tail area of S_+)

5. Make a conclusion based on the P-value

Example 6.16

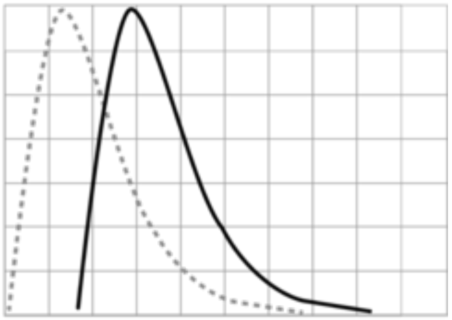
Using the sample of 6 welds from before, we want to test $H_0: \mu = 16$ versus $H_1: \mu \neq 16$

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

$S_+ = 3$ -> From Table A.5 the left-tail area (probability of observing S_+ less than or equal to 3 under H_0) = 0.0781, so the pvalue is 2*0.0781 = 0.1562

Wilcoxon Rank-Sum Test (Mann-Whitney Test)

- Analogous to the two-sample t-test in Sec. 6.7
- Use when sampled data cannot be assumed to originate from normal distributions
- Two assumptions are necessary:
  1. Populations are continuous
  2. Probability density functions are identical to each other in shape and spread



Wilcoxon Rank-Sum Test: The Idea

If there is no difference in the population means, then the sum of the overall ranks of the values in one sample should be approximately equal to the other

- If the sum of the ranks of the values in both samples are approximately equal, do not reject the null hypothesis of equal means

- If the sum of the ranks of the values in one sample is very different than the other, reject the null hypothesis of equal means

How to Compute

- For two random samples $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ where $m \leq n$, rank each of the m + n values from smallest to largest

- Formulate the hypotheses H_0 and H_1

- Compute the test statistic W = sum of the ranks corresponding to the X values

- Consult Table A.6 for the null distribution of W to obtain a p-value for the tail area corresponding to H_1

How to Compute – Example 6.19

- Resistances (mΩ) are measured for 5 wires of one type and 6 wires of another type:

  X: 36, 28, 29, 20, 38

  Y: 34, 41, 35, 47, 49, 46
- Use the Wilcoxon rank-sum test to test H_0: $\mu_X \geq \mu_Y$ versus H_1: $\mu_X < \mu_Y$
- Order the 11 values and assign ranks:

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*			

- W = 1+2+3+6+7=19
- From Table A.6, find left-tail (direction of H_1) area of 0.0260 -> Reject H_0

Summary: Wilcoxon Rank-Sum Test

Let $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ where $m \leq n$ be two small samples from **continuous** populations with means μ_X and μ_Y. Assume distributions have the **same shape and spread**.

1. Set up the null H_0 and alternative H_1 hypotheses (see table below)
2. State the level of significance α you will use
3. Calculate the test statistic W = sum of the ranks of the X values
4. Assume H_0 is true and look up the P-value using Table A.6

H_0	H_1	P-value
$\mu_X \leq \mu_Y$	$\mu_X > \mu_Y$	Area to the right of W
$\mu_X \geq \mu_Y$	$\mu_X < \mu_Y$	Area to the left of W
$\mu_X = \mu_Y$	$\mu_X \neq \mu_Y$	2*(tail area of W)

5. Make a conclusion based on the P-value

Notes

- Again deal with ties by assigning the average rank that would be assigned if the values were not tied
- Distribution-free methods are not **assumption-free**
  - Still had to assume continuous populations as well as symmetry (for the Wilcoxon signed-rank) or equal spread and shape (for the Wilcoxon rank-sum)
- Can apply the Wilcoxon signed-rank test to the differences in a paired sample

Example – Exercise 6.9.7

- A new postsurgical treatment is being compared with a standard treatment
- 7 subjects receive the new treatment while 7 others receive the standard (control) treatment
- The recovery times (in days) are:

Treatment (X):	12, 13, 15, 19, 20, 21, 27
Control (Y):	18, 23, 24, 30, 32, 35, 40

- We suspect that the recovery times do not follow a normal distribution, but are continuous and boxplots show that they have similar same shape and spread
- **Can we conclude that the mean recovery time differs between treatment and control?**

Next

- Note that we are not covering the large-sample approximations of the Wilcoxon tests

- The Chi-Square Tests for homogeneity and independence