Statistics for Business and Economics

8th Edition

Chapter 14

Analysis of Categorical Data

Chapter Goals

After completing this chapter, you should be able to:

- Use the chi-square goodness-of-fit test to determine whether data fits specified probabilities
- Perform tests for the Poisson and Normal distributions
- Set up a contingency analysis table and perform a chi-square test of association
- Use the sign test for paired or matched samples
- Recognize when and how to use the Wilcoxon signed rank test for paired or matched samples

Chapter Goals

(continued)

After completing this chapter, you should be able to:

- Use a sign test for a single population median
- Apply a normal approximation for the Wilcoxon signed rank test
- Know when and how to perform a Mann-Whitney U-test
- Explain Spearman rank correlation and perform a test for association
- Use the Runs Test to test for randomness in a time series

Introduction

- Nonparametric Statistics
  - Fewer restrictive assumptions about data levels and underlying probability distributions
    - Population distributions may be skewed
    - The level of data measurement may only be ordinal or nominal

14.1

Goodness-of-Fit Tests: Specified Probabilities

- Does sample data conform to a hypothesized distribution?
  - Examples:
    - Do sample results conform to specified expected probabilities?
    - Are technical support calls equal across all days of the week? (i.e., do calls follow a uniform distribution?)
    - Do measurements from a production process follow a normal distribution?

Chi-Square Goodness-of-Fit Test

(continued)

- Are technical support calls equal across all days of the week? (i.e., do calls follow a uniform distribution?)
  - Sample data for 10 days per day of week:

	Sum of calls for this day:
Monday	**290**
Tuesday	**250**
Wednesday	**238**
Thursday	**257**
Friday	**265**
Saturday	**230**
Sunday	**192**
	Σ = 1722

Logic of Goodness-of-Fit Test

- If calls **are** uniformly distributed, the 1722 calls would be expected to be equally divided across the 7 days:

$$\frac{1722}{7} = 246 \quad \text{expected calls per day if uniform}$$

- Chi-Square Goodness-of-Fit Test: test to see if the sample results are consistent with the expected results

Observed vs. Expected Frequencies

	Observed O_i	Expected E_i
Monday	290	246
Tuesday	250	246
Wednesday	238	246
Thursday	257	246
Friday	265	246
Saturday	230	246
Sunday	192	246
TOTAL	1722	1722

Chi-Square Test Statistic

H_0: The distribution of calls is uniform over days of the week

H_1: The distribution of calls is not uniform

- The test statistic is

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \qquad (\text{where d.f.} = K - 1)$$

where:

K = number of categories
O_i = observed frequency for category i
E_i = expected frequency for category i

The Rejection Region

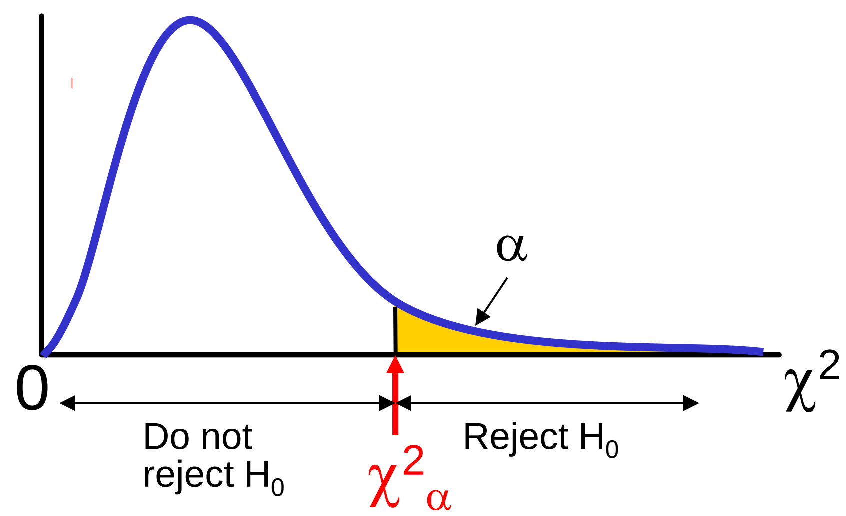
H_0: The distribution of calls is uniform over days of the week

H_1: The distribution of calls is not uniform

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

- Reject H_0 if $\chi^2 > \chi^2_\alpha$

(with k – 1 degrees of freedom)



Observed vs. Expected Frequencies

	Observed O_i	Expected E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
Monday	290	246	44	1936	7.870
Tuesday	250	246	4	16	0.065
Wednesday	238	246	-8	64	0.260
Thursday	257	246	11	121	0.492
Friday	265	246	19	361	1.467
Saturday	230	246	-16	256	1.041
Sunday	192	246	-54	2916	11.854
TOTAL	1722	1722			$\chi^2 = 23.049$

Chi-Square Test Statistic

H_0: The distribution of calls is uniform over days of the week

H_1: The distribution of calls is not uniform

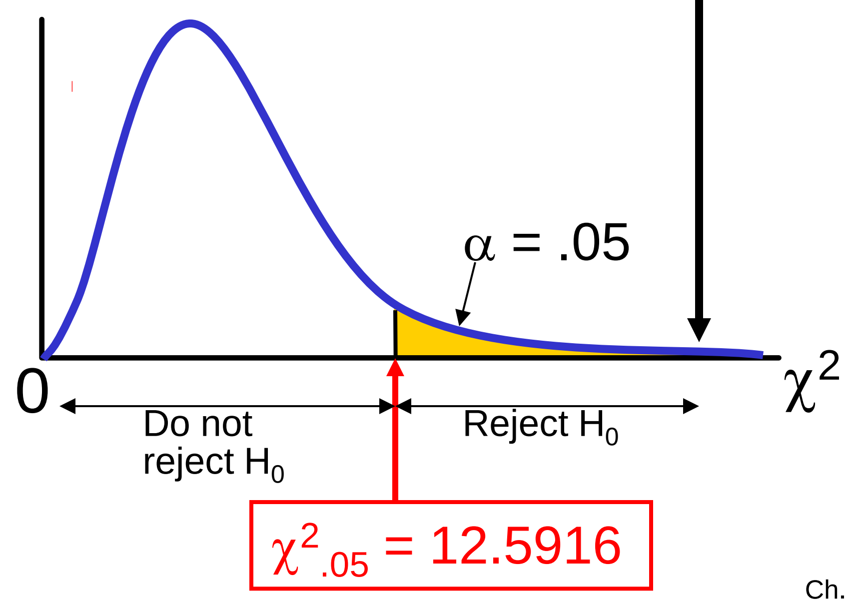
$$\chi^2 = \frac{(290-246)^2}{246} + \frac{(250-246)^2}{246} + \ldots + \frac{(192-246)^2}{246} = \boxed{23.049}$$

K – 1 = 6 (7 days of the week) so use 6 degrees of freedom:

$\chi^2_{.05} = 12.5916$

Conclusion:

$\chi^2 = 23.05 > \chi^2_\alpha = 12.5916$ so **reject H_0** and conclude that the distribution is not uniform



Goodness-of-Fit Tests: Population Parameters Unknown

14.2

Idea:

- Test whether data follow a specified distribution (such as binomial, Poisson, or normal) . . .
- . . . without assuming the parameters of the distribution are known
- Use sample data to estimate the unknown population parameters

Goodness-of-Fit Tests: Population Parameters Unknown

(continued)

- Suppose that a null hypothesis specifies category probabilities that depend on the estimation (from the data) of m unknown population parameters

- The appropriate goodness-of-fit test is the same as in the previously section . . .

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

- . . . except that the number of degrees of freedom for the chi-square random variable is

$$\boxed{\text{Degrees of Freedom} = (K - m - 1)}$$

- Where K is the number of categories

Test of Normality

- The assumption that data follow a normal distribution is common in statistics
- Evidence of normality was assessed in prior chapters

  (for example, with normal probability plots in Chapter 5)
- Here, a chi-square test is developed

Test of Normality

(continued)

- Two population parameters can be estimated using sample data:

$$\text{Skewness} = \frac{\sum_{i-1}^{n}(x_i - \bar{x})^3}{ns^3}$$

$$\text{Kurtosis} = \frac{\sum_{i-1}^{n}(x_i - \bar{x})^4}{ns^4}$$

- For a normal distribution,

Skewness = 0

Kurtosis = 3

Jarque-Bera Test for Normality

- Consider the null hypothesis that the population distribution is normal
- The Jarque-Bera Test for Normality is based on the closeness the sample skewness to 0 and the sample kurtosis to 3
- The test statistic is

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

- as the number of sample observations becomes very large, this statistic has a chi-square distribution with 2 degrees of freedom
- The null hypothesis is rejected for large values of the test statistic

Jarque-Bera Test for Normality

(continued)

- The chi-square approximation is close only for very large sample sizes
- The test statistic is compared to significance points from text Table 14.9

Sample size n	10% point	5% point	**Sample size n**	10% point	5% point
20	2.13	3.26	**200**	3.48	4.43
30	2.49	3.71	**250**	3.54	4.61
40	2.70	3.99	**300**	3.68	4.60
50	2.90	4.26	**400**	3.76	4.74
75	3.09	4.27	**500**	3.91	4.82
100	3.14	4.29	**800**	4.32	5.46
125	3.31	4.34	∞	4.61	5.99
150	3.43	4.39			

Example: Jarque-Bera Test for Normality

- The average daily temperature has been recorded for 200 randomly selected days, with sample skewness 0.232 and kurtosis 3.319
- Test the null hypothesis that the true distribution is normal

$$JB = n\left[\frac{(\text{Skewness})^2}{6} + \frac{(\text{Kurtosis} - 3)^2}{24}\right] = 200\left[\frac{(0.232)^2}{6} + \frac{(3.319 - 3)^2}{24}\right] = 2.642$$

- From Table 14.9 the 10% critical value for n = 200 is 3.48, so there is not sufficient evidence to reject that the population is normal

14.3

Contingency Tables

Contingency Tables

- Used to classify sample observations according to a pair of attributes
- Also called a cross-classification or cross-tabulation table
- Assume r categories for attribute A and c categories for attribute B
  - Then there are (r x c) possible cross-classifications

r x c Contingency Table

	Attribute B				
Attribute A	**1**	**2**	**. . .**	**c**	**Totals**
1	O_{11}	O_{12}	…	O_{1c}	R_1
2	O_{21}	O_{22}	…	O_{2c}	R_2
.	.	.	…	.	.
.	.	.	…	.	.
.	.	.	…	.	.
r	O_{r1}	O_{r2}	…	O_{rc}	R_r
Totals	C_1	C_2	…	C_c	n

Test for Association

- Consider n observations tabulated in an r x c contingency table
- Denote by O_{ij} the number of observations in the cell that is in the i^{th} row and the j^{th} column
- The null hypothesis is

H_0 : No association exists
between the two attributes in the population

- The appropriate test is a chi-square test with
(r-1)(c-1) degrees of freedom

Test for Association

(continued)

- Let R_i and C_j be the row and column totals
- The expected number of observations in cell row i and column j, given that H_0 is true, is

$$\boxed{E_{ij} = \frac{R_i C_j}{n}}$$

- A test of association at a significance level α is based on the chi-square distribution and the following decision rule

$$\boxed{\text{Reject } H_0 \text{ if } \quad \chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} > \chi^2_{(r-1)(c-1),\alpha}}$$

Contingency Table Example

Left-Handed vs. Gender

- Dominant Hand: Left vs. Right
- Gender: Male vs. Female

H_0: There is no association between hand preference and gender

H_1: Hand preference is not independent of gender

Contingency Table Example

(continued)

Sample results organized in a contingency table:

sample size = n = 300:

120 Females, 12 were left handed

180 Males, 24 were left handed

Gender	Hand Preference		
	Left	Right	
Female	12	108	120
Male	24	156	180
	36	264	300

Logic of the Test

H_0: There is no association between
hand preference and gender

H_1: Hand preference is not independent of gender

- If H_0 is true, then the proportion of left-handed females should be the same as the proportion of left-handed males
- The two proportions above should be the same as the proportion of left-handed people overall

Finding Expected Frequencies

120 Females, 12 were left handed

180 Males, 24 were left handed

Overall:

P(Left Handed) = 36/300 = .12

If no association, then

P(Left Handed | Female) = P(Left Handed | Male) = .12

So we would expect 12% of the 120 females and 12% of the 180 males to be left handed…

i.e., we would expect (120)(.12) = 14.4 females to be left handed
(180)(.12) = 21.6 males to be left handed

Expected Cell Frequencies

(continued)

- Expected cell frequencies:

$$E_{ij} = \frac{R_i C_j}{n} = \frac{(i^{th}\ \text{Row total})(j^{th}\ \text{Column total})}{\text{Total sample size}}$$

Example:

$$E_{11} = \frac{(120)(36)}{300} = 14.4$$

Observed vs. Expected Frequencies

Observed frequencies vs. expected frequencies:

Gender	Hand Preference		
	Left	Right	
Female	Observed = 12 Expected = 14.4	Observed = 108 Expected = 105.6	120
Male	Observed = 24 Expected = 21.6	Observed = 156 Expected = 158.4	180
	36	264	300

The Chi-Square Test Statistic

The Chi-square test statistic is:

$$\boxed{\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}} \quad \text{with} \quad \text{d.f.} = (r-1)(c-1)$$

- where:
  O_{ij} = observed frequency in cell (*i*, *j*)
  E_{ij} = expected frequency in cell (*i*, *j*)
  r = number of rows
  c = number of columns

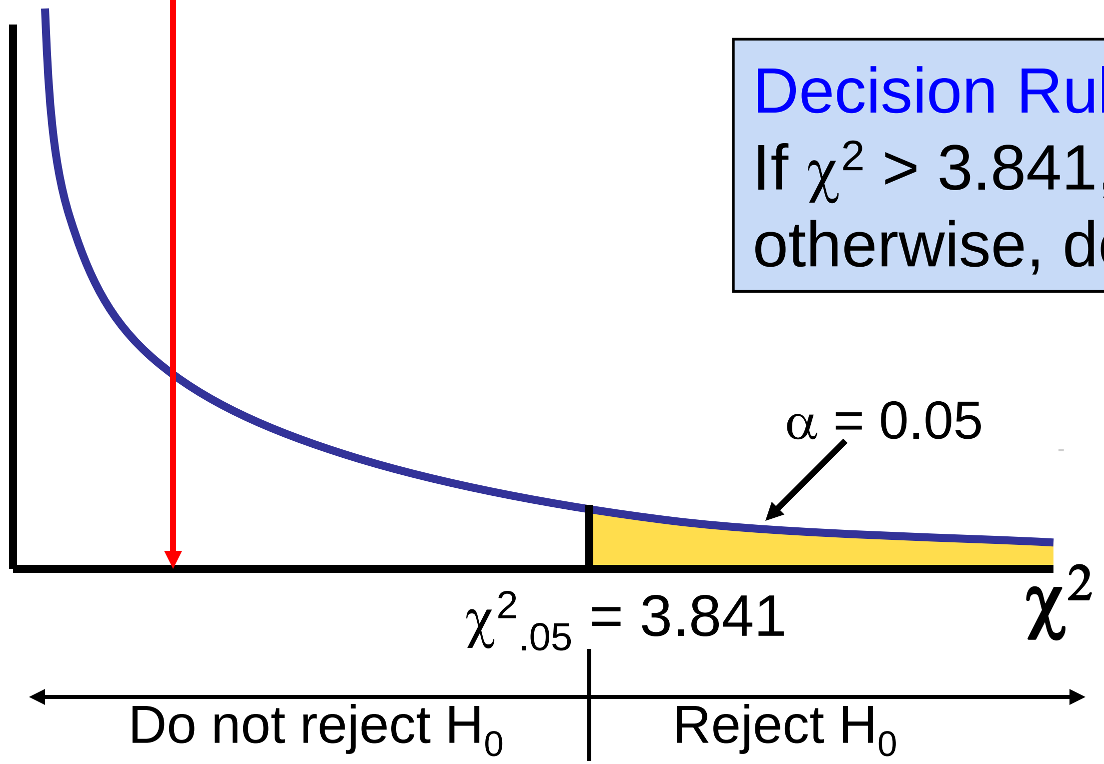
Observed vs. Expected Frequencies

Gender	Hand Preference: Left	Hand Preference: Right	
Female	Observed = 12 Expected = 14.4	Observed = 108 Expected = 105.6	120
Male	Observed = 24 Expected = 21.6	Observed = 156 Expected = 158.4	180
	36	264	300

$$\chi^2 = \frac{(12-14.4)^2}{14.4} + \frac{(108-105.6)^2}{105.6} + \frac{(24-21.6)^2}{21.6} + \frac{(156-158.4)^2}{158.4} = 0.7576$$

Contingency Analysis

$$\chi^2 = \boxed{0.7576} \quad \text{with} \quad \text{d.f.} = (r-1)(c-1) = (1)(1) = 1$$



Decision Rule:

If $\chi^2 > 3.841$, reject H_0, otherwise, do not reject H_0

Here, $\chi^2 = 0.7576 < 3.841$, so we **do not reject H_0** and conclude that gender and hand preference are not associated

14.4

Nonparametric Tests for Paired or Matched Samples

- A sign test for paired or matched samples:
  - Calculate the differences of the paired observations
  - Discard the differences equal to 0, leaving *n* observations
  - Record the sign of the difference as + or –
- For a symmetric distribution, the signs are random and + and – are equally likely

Sign Test

(continued)

- Define + to be a "success" and let P = the true proportion of +'s in the population
- The sign test is used for the hypothesis test

$$\boxed{H_0 : P = 0.5}$$

- The test-statistic S for the sign test is

$$\boxed{\text{S = the number of pairs with a positive difference}}$$

- S has a binomial distribution with P = 0.5 and
  n = the number of nonzero differences

Determining the p-value

- The p-value for a Sign Test is found using the binomial distribution with n = number of nonzero differences, S = number of positive differences, and P = 0.5

- For an upper-tail test, H_1: P > 0.5, $\boxed{\text{p-value} = P(x \geq S)}$

- For a lower-tail test, H_1: P < 0.5, $\boxed{\text{p-value} = P(x \leq S)}$

- For a two-tail test, H_1: $P \neq 0.5$, $\boxed{2P(x \geq S)}$

Sign Test Example

- Ten consumers in a focus group have rated the attractiveness of two package designs for a new product

Consumer	Rating		Difference	Sign of Difference
	Package 1	Package 2	Rating 1 – 2	
1	5	8	-3	**–**
2	4	8	-4	**–**
3	4	4	0	**0**
4	6	5	+1	**+**
5	3	9	-6	**–**
6	5	9	-4	**–**
7	7	6	-1	**–**
8	5	9	-4	**–**
9	6	3	+3	**+**
10	7	9	-2	**–**

Sign Test Example

(continued)

- Test the hypothesis that there is no overall package preference using α = 0.10

$$H_0 : P = 0.5$$

The proportion of consumers who prefer package 1 is the same as the proportion preferring package 2

$$H_1 : P < 0.5$$

A majority prefer package 2

- The test-statistic S for the sign test is

S = the number of pairs with a positive difference
= 2

- S has a binomial distribution with P = 0.5 and n = 9 (there was one zero difference)

Sign Test Example

(continued)

- The p-value for this sign test is found using the binomial distribution with $n = 9$, $S = 2$, and $P = 0.5$:

- For a lower-tail test,

$$\text{p-value} = P(x \leq 2 | n=9, P=0.5)$$
$$= 0.090$$

Since $0.090 < \alpha = 0.10$ we reject the null hypothesis and conclude that consumers prefer package 2

Wilcoxon Signed Rank Test for Paired or Matched Samples

- Uses matched pairs of random observations
- Still based on ranks
- Incorporates information about the magnitude of the differences
- Tests the hypothesis that the distribution of differences is centered at zero
- The population of paired differences is assumed to be symmetric

Wilcoxon Signed Rank Test for Paired or Matched Samples

(continued)

Conducting the test:

- Discard pairs for which the difference is 0
- Rank the remaining n absolute differences in ascending order (ties are assigned the average of their ranks)
- Find the sums of the positive ranks and the negative ranks
- The smaller of these sums is the Wilcoxon Signed Rank Statistic T:

$$\boxed{T = \min(T_+, T_-)}$$

Where T_+ = the sum of the positive ranks

T_- = the sum of the negative ranks

n = the number of nonzero differences

- The null hypothesis is rejected if T is less than or equal to the value in Appendix Table 10

Signed Rank Test Example

Consumer	Rating		Difference		
	Package 1	Package 2	Diff (rank)	Rank (+)	Rank (–)
1	5	8	-3 (4.5 tie)		**4.5**
2	4	8	-4 (7 tie)		**7**
3	4	4	0 (-)		
4	6	5	+1 (1.5 tie)	**1.5**	
5	3	9	-6 (9)		**9**
6	5	9	-4 (7 tie)		**7**
7	7	6	-1 (1.5 tie)		**1.5**
8	5	9	-4 (7 tie)		**7**
9	6	3	+3 (4.5 tie)	**4.5**	
10	7	9	-2 (3)		**3**

- Ten consumers in a focus group have rated the attractiveness of two package designs for a new product

$T_+ = 6$ $T_- = 39$

Signed Rank Test Example

(continued)

Test the hypothesis that the distribution of paired differences is centered at zero, using $\alpha = 0.10$

Conducting the test:

- The smaller of T_+ and T_- is the Wilcoxon Signed Rank Statistic T:

$$\boxed{\mathbf{T = \min(T_+, T_-) = 6}}$$

- Use Appendix Table 10 with $n = 9$ to find the critical value:

$$\boxed{\text{The null hypothesis is rejected if } T \leq 4}$$

- Since $T = 6 > 4$, we do not reject the null hypothesis; we do not have sufficient evidence that rankings are higher for package 2

Normal Approximation to the Sign Test

- If the number n of nonzero sample observations is large, then the sign test is based on the normal approximation to the binomial with mean and standard deviation

$$\mu = np = 0.5n$$
$$\sigma = \sqrt{np(1-p)} = \sqrt{0.25n} = 0.5\sqrt{n}$$

- The test statistic is

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

- Where S* is the test-statistic corrected for continuity:
  - For a two-tail test, $S^* = S + 0.5$, if $S < \mu$ or $S^* = S - 0.5$, if $S > \mu$
  - For upper-tail test, $S^* = S - 0.5$
  - For lower-tail test, $S^* = S + 0.5$

Normal Approximation to the Wilcoxon Signed Rank Test

A normal approximation can be used when

- Paired samples are observed
- The sample size is large ($n > 20$)
- The hypothesis test is that the population distribution of differences is centered at zero

Wilcoxon Matched Pairs Test for Large Samples

- The mean and standard deviation for Wilcoxon T :

$$E(T) = \mu_T = \frac{n(n+1)}{4}$$

$$Var(T) = \sigma_T^2 = \frac{(n)(n+1)(2n+1)}{24}$$

where n is the number of paired values

Wilcoxon Matched Pairs Test for Large Samples

(continued)

- Normal approximation for the Wilcoxon T Statistic:

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

  - If the alternative hypothesis is one-sided, reject the null hypothesis if

$$\frac{T - \mu_T}{\sigma_T} < -z_\alpha$$

  - If the alternative hypothesis is two-sided, reject the null hypothesis if

$$\frac{T - \mu_T}{\sigma_T} < -z_{\alpha/2}$$

Sign Test for Single Population Median

- The sign test can be used to test that a single population median is equal to a specified value
  - For small samples, use the binomial distribution
  - For large samples, use the normal approximation

14.5

Nonparametric Tests for Independent Random Samples

Used to compare two samples from two populations

Assumptions:

- The two samples are independent and random
- The value measured is a continuous variable
- The two distributions are identical except for a possible difference in the central location
- The sample size from each population is at least 10

Mann-Whitney U-Test

- Consider two samples
  - Pool the two samples (combine into a singe list) but keep track of which sample each value came from
  - rank the values in the combined list in ascending order
    - For ties, assign each the average rank of the tied values
  - sum the resulting rankings separately for each sample
- If the sum of rankings from one sample differs enough from the sum of rankings from the other sample, we conclude there is a difference in the population medians

Mann-Whitney U Statistic

- Consider n_1 observations from the first population and n_2 observations from the second
- Let R_1 denote the sum of the ranks of the observations from the first population
- The Mann-Whitney U statistic is

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

Mann-Whitney U Statistic

(continued)

- The null hypothesis is that the medians of the two population distributions are the same
- The Mann-Whitney U statistic has mean and variance

$$E(U) = \mu_U = \frac{n_1 n_2}{2}$$

$$Var(U) = \sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

- Then for large sample sizes (both at least 10), the distribution of the random variable

$$Z = \frac{U - \mu_U}{\sigma_U}$$

is approximated by the normal distribution

Decision Rules for Mann-Whitney Test

The decision rule for the null hypothesis that the two populations have the same medians:

- For a one-sided upper-tailed alternative hypothesis:

$$\text{Reject } H_0 \text{ if } \; Z = \frac{U - \mu_U}{\sigma_U} < -z_\alpha$$

- For a one-sided lower-tailed hypothesis:

$$\text{Reject } H_0 \text{ if } \; Z = \frac{U - \mu_U}{\sigma_U} > z_\alpha$$

- For a two-sided alternative hypothesis:

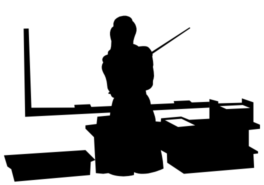
$$\text{Reject } H_0 \text{ if } Z = \frac{U - \mu_U}{\sigma_U} < -z_{\alpha/2} \quad \text{or} \quad \text{Reject } H_0 \text{ if } Z = \frac{U - \mu_U}{\sigma_U} > z_{\alpha/2}$$

Mann-Whitney U-Test Example

Claim: Median class size for Math is larger than the median class size for English

A random sample of 10 Math and 10 English classes is selected (samples do not have to be of equal size)

Rank the combined values and then determine rankings by original sample

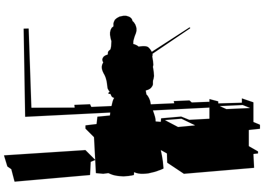


Mann-Whitney U-Test Example

(continued)

- Suppose the results are:

Class size (Math, M)	Class size (English, E)
23	30
45	47
34	18
78	34
34	44
66	61
62	54
95	28
81	40
99	96



Mann-Whitney U-Test Example

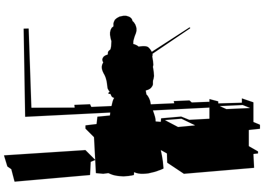
(continued)

Ranking for combined samples

tied

Size	Rank
18	1
23	2
28	3
30	4
34	6
34	6
34	6
40	8
44	9
45	10

Size	Rank
47	11
54	12
61	13
62	14
66	15
78	16
81	17
95	18
96	19
99	20

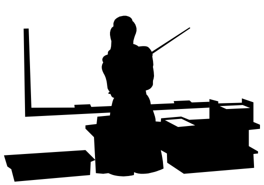


Mann-Whitney U-Test Example

(continued)

- Rank by original sample:

Class size (Math, M)	Rank	Class size (English, E)	Rank
23	2	**30**	4
45	10	**47**	11
34	6	**18**	1
78	16	**34**	6
34	6	**44**	9
66	15	**61**	13
62	14	**54**	12
95	18	**28**	3
81	17	**40**	8
99	20	**96**	19
	$\Sigma = 124$		$\Sigma = 86$



Mann-Whitney U-Test Example

(continued)

Claim: Median class size for Math is larger than the median class size for English

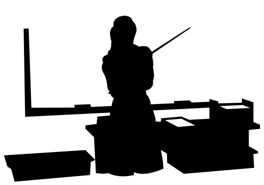
H_0: Median$_M$ ≤ Median$_E$

(Math median is not greater than English median)

H_A: Median$_M$ > Median$_E$

(Math median is larger)

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1 = (10)(10) + \frac{(10)(11)}{2} - 124 = 31$$



Mann-Whitney U-Test Example

(continued)

H_0: $\text{Median}_M \le \text{Median}_E$

H_A: $\text{Median}_M > \text{Median}_E$

$$Z = \frac{U - \mu_U}{\sigma_U} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = \frac{31 - \frac{(10)(10)}{2}}{\sqrt{\frac{(10)(10)(10 + 10 + 1)}{12}}} = -1.436$$

- The decision rule for this one-sided upper-tailed alternative hypothesis:

$$\text{Reject } H_0 \text{ if } \quad Z = \frac{U - \mu_U}{\sigma_U} < -z_\alpha$$

- For $\alpha = 0.05$, $-z_\alpha = -1.645$
- The calculated Z value is not in the rejection region, so we conclude that there is not sufficient evidence of difference in class size medians

Wilcoxon Rank Sum Test

- Similar to Mann-Whitney U test
- Results will be the same for both tests
- May be easier to apply

Wilcoxon Rank Sum Test

(continued)

- n_1 observations from the first population
- n_2 observations from the second population
- Pool the samples and rank the observations in ascending order
- Let T denote the sum of the ranks of the observations from the first population
  - (T in the Wilcoxon Rank Sum Test is the same as R_1 in the Mann-Whitney U Test)

Wilcoxon Rank Sum Test

(continued)

- The Wilcoxon Rank Sum Statistic, T, has mean

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

- And variance

$$Var(T) = \sigma_T^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

- Then, for large samples ($n_1 \geq 10$ and $n_2 \geq 10$) the distribution of the random variable

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

is approximated by the normal distribution

Wilcoxon Rank Sum Example

- We wish to test

H_0: $\text{Median}_1 \geq \text{Median}_2$

H_1: $\text{Median}_1 < \text{Median}_2$

- Use $\alpha = 0.05$
- Suppose two samples are obtained:
- $n_1 = 40$, $n_2 = 50$
- When rankings are completed, the sum of ranks for sample 1 is $\Sigma R_1 = 1475$ $= T$
- When rankings are completed, the sum of ranks for sample 2 is $\Sigma R_2 = 2620$

Wilcoxon Rank Sum Example

(continued)

- Using the normal approximation:

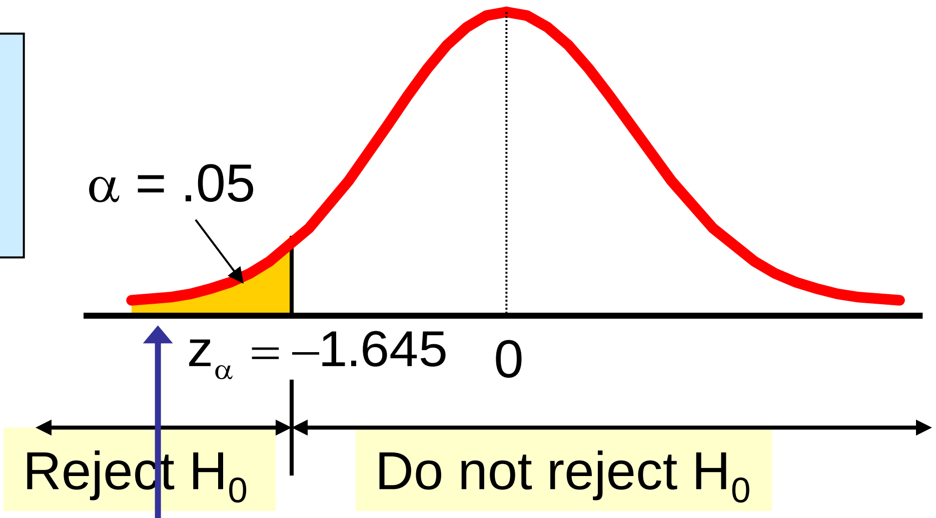
$$Z = \frac{T - \mu_T}{\sigma_T} = \frac{T - \frac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = \frac{1475 - \frac{(40)(40 + 50 + 1)}{2}}{\sqrt{\frac{(40)(50)(40 + 50 + 1)}{12}}} = -2.80$$

Wilcoxon Rank Sum Example

(continued)

H_0: Median$_1$ $\geq$ Median$_2$

H_1: Median$_1$ < Median$_2$



$$z = \frac{T - \mu_T}{\sigma_T} = \boxed{-2.80}$$

Since Z = -2.80 < -1.645, we reject H_0 and conclude that median 1 is less than median 2 at the 0.05 level of significance

14.6

Spearman Rank Correlation

- Consider a random sample $(x_1, y_1), \ldots, (x_n, y_n)$ of n pairs of observations
- Rank x_i and y_i each in ascending order
- Calculate the sample correlation of these ranks
- The resulting coefficient is called Spearman's Rank Correlation Coefficient
- If there are no tied ranks, an equivalent formula for computing this coefficient is

$$\boxed{r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}}$$

where the d_i are the differences of the ranked pairs

Spearman Rank Correlation

(continued)

- Consider the null hypothesis

H_0: no association in the population

- To test against the alternative of positive association, the decision rule is

$$\text{Reject } H_0 \text{ if } r_S > r_{S,\alpha}$$

- To test against the alternative of negative association, the decision rule is

$$\text{Reject } H_0 \text{ if } r_S < -r_{S,\alpha}$$

- To test against the two-sided alternative of some association, the decision rule is

$$\text{Reject } H_0 \text{ if } r_S < -r_{S,\alpha/2} \quad \text{or} \quad r_S > r_{S,\alpha/2}$$

14.7

A Nonparametric Test for Randomness

The Runs Test: Small Sample Size

- Consider a time series of $n \leq 20$ observations
- The runs test is used to determine whether a pattern in time-series data is random
- A run is a sequence of one or more occurrences above or below the median
- Denote observations above the median with “+” signs and observations below the median with “-” signs

The Runs Test for Randomness

(continued)

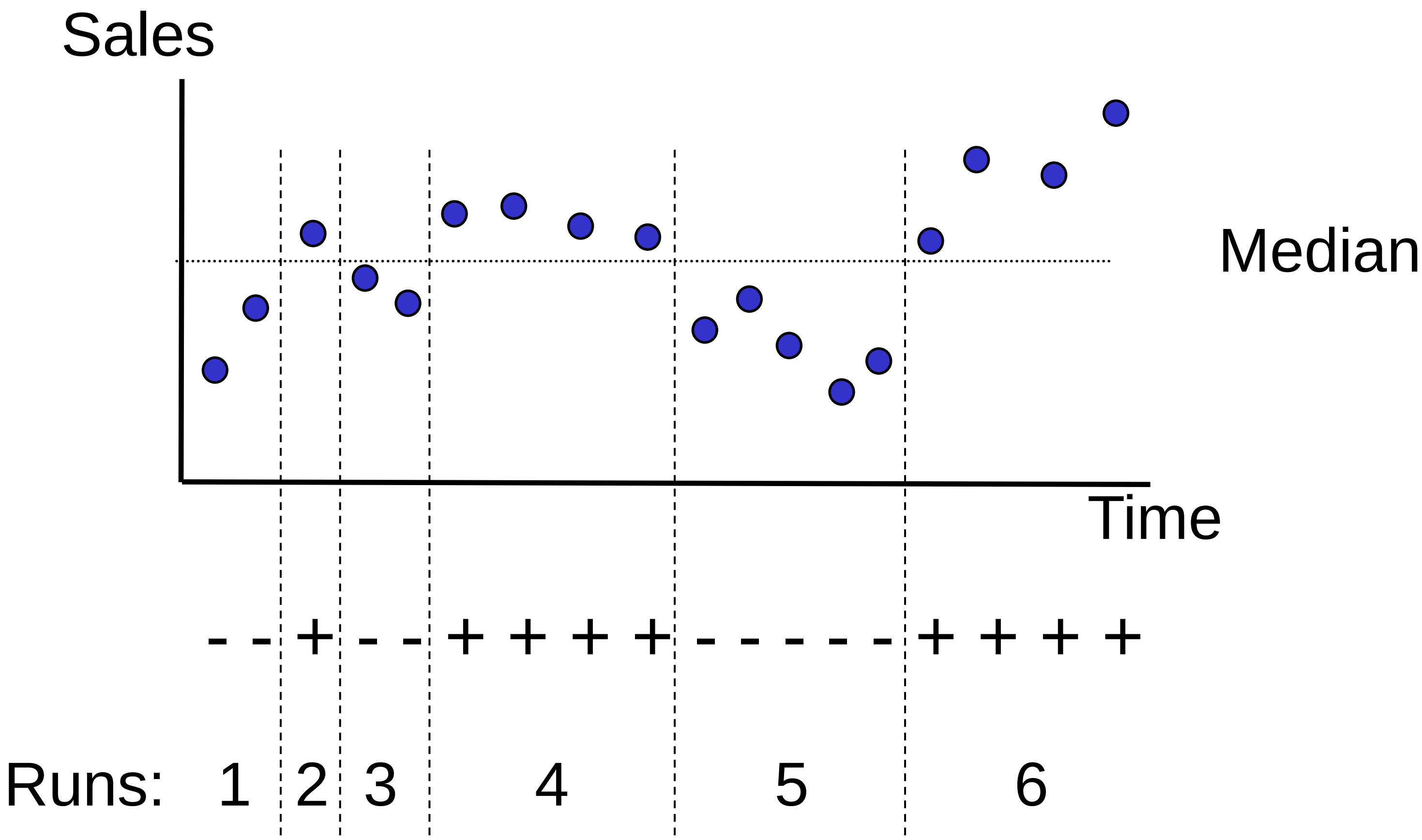
- Consider n time series observations
- Let R denote the number of runs in the sequence
- The null hypothesis is that the series is random
- Appendix Table 14 gives the smallest significance level for which the null hypothesis can be rejected (against the alternative of positive association between adjacent observations) as a function of R and n

The Runs Test for Randomness

(continued)

- If the alternative is a two-sided hypothesis on nonrandomness,
  - the significance level must be doubled if it is less than 0.5
  - if the significance level, α, read from the table is greater than 0.5, the appropriate significance level for the test against the two-sided alternative is $2(1 - \alpha)$

Counting Runs



n = 18 and there are R = 6 runs

Runs Test Example

n = 18 and there are R = 6 runs

- Use Appendix Table 14
  - n = 18 and R = 6
  - the null hypothesis can be rejected (against the alternative of positive association between adjacent observations) at the 0.044 level of significance
  - Therefore we reject that this time series is random using $\alpha = 0.05$

Runs Test: Large Sample Sizes

- Given n > 20 observations
- Let R be the number of sequences above or below the median

Consider the null hypothesis H_0: The series is random

- If the alternative hypothesis is positive association between adjacent observations, the decision rule is:

Reject H_0 if $$Z = \frac{R - \frac{n}{2} - 1}{\sqrt{\frac{n^2 - 2n}{4(n-1)}}} < -z_\alpha$$

Runs Test: Large Samples

(continued)

Consider the null hypothesis H_0: The series is random

- If the alternative is a two-sided hypothesis of nonrandomness, the decision rule is:

Reject H_0 if $Z = \dfrac{R - \frac{n}{2} - 1}{\sqrt{\dfrac{n^2 - 2n}{4(n-1)}}} < -z_{\alpha/2}$ or $Z = \dfrac{R - \frac{n}{2} - 1}{\sqrt{\dfrac{n^2 - 2n}{4(n-1)}}} > z_{\alpha/2}$

Example: Large Sample Runs Test

- A filling process over- or under-fills packages, compared to the median

OOO U OO U O UU OO UU OOOO UU O UUUU OOO UUU OOOO UU OO UUU O U OO UUUUU OOO U O UU OOO U OOOO UUU O UU OOO U OO UU O U OO UUU O UU OOOO UUU OOO

n = 100   (53 overfilled, 47 underfilled)

R = 45 runs

Example: Large Sample Runs Test

(continued)

- A filling process over- or under-fills packages, compared to the median
- n = 100 , R = 45

$$Z = \frac{R - \frac{n}{2} - 1}{\sqrt{\frac{n^2 - 2n}{4(n-1)}}} = \frac{45 - \frac{100}{2} - 1}{\sqrt{\frac{100^2 - 2(100)}{4(100-1)}}} = \frac{-6}{4.975} = -1.206$$

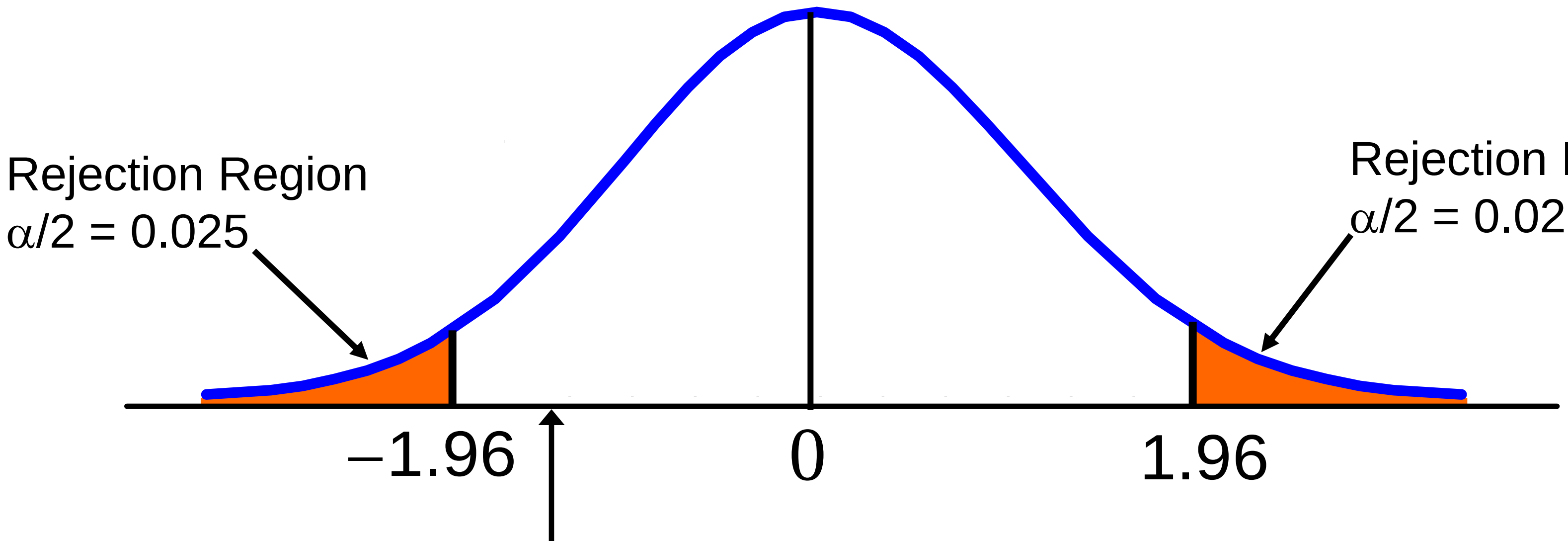
Example: Large Sample Runs Test

(continued)

H_0: Fill amounts are random

H_1: Fill amounts are not random

Test using $\alpha = 0.05$



Since $z = -1.206$ is not less than $-z_{.025} = -1.96$, we do not reject H_0

Chapter Summary

- Used the chi-square goodness-of-fit test to determine whether sample data match specified probabilities
- Conducted goodness-of-fit tests when a population parameter was unknown
- Tested for normality using the Jarque-Bera test
- Used contingency tables to perform a chi-square test for association
  - Compared observed cell frequencies to expected cell frequencies

Chapter Summary

(continued)

- Used the sign test for paired or matched samples, and the normal approximation for the sign test
- Developed and applied the Wilcoxon signed rank test, and the large sample normal approximation
- Developed and applied the Mann-Whitney U-test for two population medians
- Used the Wilcoxon rank-sum test
- Examined Spearman rank correlation for tests of association
- Used the Runs Test to test for randomness