Chapter 4 Contingency Tables

Section 4.1 The 2×2 contingency table

	Class 1	Class 2	
Population 1	O_{11}	O_{12}	n_1
Population 2	O_{21}	O_{22}	n_2
	C_1	C_2	N

2×2 contingency table

Question: Does the treatment significantly alter the proportion of objects in each of the two categories?

- The Chi-squared test for differences in probabilities

Assumptions

1. Each sample is a random sample.
2. The two samples are independent.
3. Each observation may be categorized either into class 1 or class 2.

	Class 1	Class 2	
Population 1	O_{11}	O_{12}	n_1
Population 2	O_{21}	O_{22}	n_2
	C_1	C_2	N

p_1 = probability that an observation from population 1 will be in class 1

p_2 = probability that an observation from population 2 will be in class 1

Statistical test:

Setting 1	Setting 2	Setting 3
$H_0 : p_1 \le p_2$	$H_0 : p_1 \ge p_2$	$H_0 : p_1 = p_2$
$H_a : p_1 > p_2$	$H_a : p_1 < p_2$	$H_a : p_1 \ne p_2$

	Class 1	Class 2	
Population 1	O_{11}	O_{12}	n_1
Population 2	O_{21}	O_{22}	n_2
	C_1	C_2	N

Test statistic: $T_1 = \dfrac{\dfrac{O_{11}}{n_1} - \dfrac{O_{21}}{n_2}}{\sqrt{\dfrac{C_1 C_2}{N^2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Another expression of T_1: $T_1 = \dfrac{\sqrt{N}\,(O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1 n_2 C_1 C_2}}$

Setting 1. $H_0 : p_1 \le p_2$ $H_a : p_1 > p_2$

Reject H_0 if $T_{1(\text{obs})} > z_{1-\alpha}$.

p-value $= P\left(Z \ge T_{1(\text{obs})}\right)$

Setting 2. $H_0 : p_1 \ge p_2$ $H_a : p_1 < p_2$

Reject H_0 if $T_{1(\text{obs})} < z_{\alpha}$.

p-value $= P\left(Z \le T_{1(\text{obs})}\right)$

Setting 3. $H_0 : p_1 = p_2$ $H_a : p_1 \ne p_2$

Reject H_0 if $T_{1(\text{obs})} > z_{1-\alpha/2}$ or $T_{1(\text{obs})} < z_{\alpha/2}$.

p-value $= 2 \min\left\{P\left(Z \ge T_{1(\text{obs})}\right), P\left(Z \le T_{1(\text{obs})}\right)\right\}$

EXAMPLE 1

Two carloads of manufactured items are sampled randomly to determine if the proportion of defective items is different for the two carloads. From the first carload 13 of the 86 items were defective. From the second carload 17 of the 74 items were considered defective.

	Defective	Nondefective	
Carload 1			
Carload 2			

p_1 = probability that an item from carload 1 is defective

p_2 = probability that an item from carload 2 is defective

EXAMPLE 2

At the U.S. Naval Academy a new lighting system was installed throughout the midshipmen's living quarters. It was claimed that the new lighting system resulted in poor eyesight due to a continual strain on the eyes of the midshipmen. Random samples are taken before and after the installation of the new lights. The results are shown in the table.

	Good Vision	Poor Vision
Old Lights	O_{11}=714	O_{12}=111
New Lights	O_{21}=662	O_{22}=154

Let p_1 be the probability that a randomly selected graduating midshipman had good vision under the old lighting system.

Let p_2 be the probability that a randomly selected graduating midshipman had good vision under the new lighting system.

- The χ^2 distribution can be used for the two-sided test

  $H_0 : p_1 = p_2$ vs. $H_a : p_1 \neq p_2$

	Class 1	Class 2	
Population 1	O_{11}	O_{12}	n_1
Population 2	O_{21}	O_{22}	n_2
	C_1	C_2	N

Test statistic: $T_1' = (T_1)^2 = \dfrac{N(O_{11}O_{22} - O_{12}O_{21})^2}{n_1 n_2 C_1 C_2}$

Reject H_0 if $T_1'{}_{(\text{obs})} > \chi^2_{1,1-\alpha}$.

- Fisher's Exact Test

	Column 1	Column 2	
Row 1	x	$r - x$	r
Row 2	$c - x$	$N - r - c + x$	$N - r$
	c	$N - c$	N

Assumptions:

1. Each observation is classified into exactly one cell.

2. The row and column totals are fixed, not random.

p_1 = probability of an item in row 1 being classified into column 1

p_2 = probability of an item in row 2 being classified into column 1

	Column 1	Column 2	
Row 1	x	$r-x$	r
Row 2	$c-x$	$N-r-c+x$	$N-r$
	c	$N-c$	N

Statistical test:

Setting 1	Setting 2	Setting 3
$H_0: p_1 \le p_2$	$H_0: p_1 \ge p_2$	$H_0: p_1 = p_2$
$H_a: p_1 > p_2$	$H_a: p_1 < p_2$	$H_a: p_1 \ne p_2$

Test statistic: T_2 = number of items in cell (row 1, column 1)

When $p_1 = p_2$, the distribution of T_2 is a hypergeometric distribution.

$$P(T_2 = x) = \frac{\binom{r}{x}\binom{N-r}{c-x}}{\binom{N}{c}} \quad (x = 0, 1, 2, \cdots, \min(r, c))$$

	Column 1	Column 2	
Row 1	x	$r-x$	r
Row 2	$c-x$	$N-r-c+x$	$N-r$
	c	$N-c$	N

Setting 1. $H_0: p_1 \le p_2$ $H_a: p_1 > p_2$

$$p\text{-value} = P\left(T_2 \ge T_{2(\text{obs})}\right)$$

Setting 2. $H_0: p_1 \ge p_2$ $H_a: p_1 < p_2$

$$p\text{-value} = P\left(T_2 \le T_{2(\text{obs})}\right)$$

Setting 3. $H_0: p_1 = p_2$ $H_a: p_1 \ne p_2$

$$p\text{-value} = 2\min\left\{P\left(T_2 \ge T_{2(\text{obs})}\right), P\left(T_2 \le T_{2(\text{obs})}\right)\right\}$$

EXAMPLE 3

Fourteen newly hired business majors, 10 males and 4 females, all equally qualified, are being assigned by the bank president to their new jobs. Ten of the new jobs are as tellers, and four are as account representatives. The null hypothesis is that males and females have equal chances at getting the more desirable account representative jobs. The one-sided alternative of interest is that females are more likely than males to get the account representative jobs. Only one female is assigned a teller position. Can the null hypothesis be rejected?

	Representative	Teller
Male		
Female		

Section 4.2 The $r \times c$ contingency table

- The chi-squared test for differences in probabilities

	Class 1	Class 2	$\cdots$	Class c	
Population 1	O_{11}	O_{12}	$\cdots$	O_{1c}	n_1
Population 2	O_{21}	O_{22}	$\cdots$	O_{2c}	n_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Population r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	n_r
	C_1	C_2	$\cdots$	C_c	N

Question: Does the treatment significantly alter the proportion of objects in each of the c categories?

Assumptions

1. Each sample is a random sample.
2. All the samples are independent.
3. Each observation is categorized into exactly one category or class.

	Class 1	Class 2	$\cdots$	Class c	
Population 1	O_{11}	O_{12}	$\cdots$	O_{1c}	n_1
Population 2	O_{21}	O_{22}	$\cdots$	O_{2c}	n_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Population r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	n_r
	C_1	C_2	$\cdots$	C_c	N

p_{ij} = probability that an observation from population i will be in column j

$(i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c)$

Statistical test:

$H_0 : p_{1j} = p_{2j} = \cdots = p_{rj} \quad (j = 1, 2, \cdots, c)$

(All of the probabilities in the same column are the same.)

$H_a : p_{ij} \neq p_{kj}$ for at least one pair of i and k and for some j

(At least two probabilities in the same column are different.)

	Class 1	Class 2	$\cdots$	Class c	
Population 1	O_{11}	O_{12}	$\cdots$	O_{1c}	n_1
Population 2	O_{21}	O_{22}	$\cdots$	O_{2c}	n_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Population r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	n_r
	C_1	C_2	$\cdots$	C_c	N

Test Statistic: $T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$ $\left(E_{ij}=\frac{n_i C_j}{N}\right)$

O_{ij} = observed number of observations in cell (i, j)

E_{ij} = expected number of observations in cell (i, j)

Another expression of T: $T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}} - N$ $\left(E_{ij}=\frac{n_i C_j}{N}\right)$

	Class 1	Class 2	$\cdots$	Class c	
Population 1	O_{11}	O_{12}	$\cdots$	O_{1c}	n_1
Population 2	O_{21}	O_{22}	$\cdots$	O_{2c}	n_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Population r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	n_r
	C_1	C_2	$\cdots$	C_c	N

Reject H_0 if $T_{(\text{obs})} > \chi^2_{(r-1)(c-1),1-\alpha}$

p-value $= P\left(T \geq T_{(\text{obs})}\right)$ $\left(T \sim \chi^2_{(r-1)(c-1)}\right)$

Minimum requirement for the χ^2 approximation to be valid:

1. All E_{ij}'s are greater than 0.5.

2. At least half of the E_{ij} values are greater than 1.

If the minimum requirement is not met, some rows or columns should be combined. The other option is to omit rows or columns with very few points.

EXAMPLE 1

A sample of students randomly selected from private high schools and a sample of students randomly selected from public high schools were given standardized achievement tests with the following results.

	0-275	276-350	351-425	426-500
Private School	6	14	17	9
Public School	30	32	17	3

Is there any significant difference between the test scores of private and public high school students?

- The chi-squared test for independence

	Column 1	Column 2	$\cdots$	Column c	
Row 1	O_{11}	O_{12}	$\cdots$	O_{1c}	R_1
Row 2	O_{21}	O_{22}	$\cdots$	O_{2c}	R_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Row r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	R_r
	C_1	C_2	$\cdots$	C_c	N

Assumptions:

1. The sample of N observations is a random sample. (Each observation has the same probability as every other observation of being classified in row i and column j, independently of the other observations.)
2. Each observation is classified into exactly one row according to one criterion and into one column according to another criterion.

	Column 1	Column 2	$\cdots$	Column c	
Row 1	O_{11}	O_{12}	$\cdots$	O_{1c}	R_1
Row 2	O_{21}	O_{22}	$\cdots$	O_{2c}	R_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Row r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	R_r
	C_1	C_2	$\cdots$	C_c	N

p_{ij} = probability that an observation will be in row i column j
$(i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c)$

$p_{i.}$ = probability that an observation will be in row i $(i = 1, 2, \cdots, r)$

$p_{.j}$ = probability that an observation will be in column j $(j = 1, 2, \cdots, c)$

	Column 1	Column 2	$\cdots$	Column c	
Row 1	O_{11}	O_{12}	$\cdots$	O_{1c}	R_1
Row 2	O_{21}	O_{22}	$\cdots$	O_{2c}	R_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Row r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	R_r
	C_1	C_2	$\cdots$	C_c	N

Statistical test:

$H_0: p_{ij} = p_{i.}\, p_{.j}$ for all i and j

(The event "an observation is in row i" is independent of the event "an observation is in column j" for all i and j.)

$H_a: p_{ij} \neq p_{i.}\, p_{.j}$ for at least one pair of i and j

(The event "an observation is in row i" is not independent of the event "an observation is in column j" for all i and j.)

	Column 1	Column 2	$\cdots$	Column c	
Row 1	O_{11}	O_{12}	$\cdots$	O_{1c}	R_1
Row 2	O_{21}	O_{22}	$\cdots$	O_{2c}	R_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Row r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	R_r
	C_1	C_2	$\cdots$	C_c	N

Test Statistic: $T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} \quad \left(E_{ij}=\frac{R_iC_j}{N}\right)$

Another expression of T: $T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}^2}{E_{ij}} - N \quad \left(E_{ij}=\frac{R_iC_j}{N}\right)$

Reject H_0 if $T_{(\text{obs})} > \chi^2_{(r-1)(c-1),1-\alpha}$

p-value $= P\left(T \ge T_{(\text{obs})}\right) \quad \left(T \sim \chi^2_{(r-1)(c-1)}\right)$

Minimum requirement for the χ^2 approximation to be valid:

1. All E_{ij}'s are greater than 0.5.
2. At least half of the E_{ij} values are greater than 1.

If the minimum requirement is not met, some rows or columns should be combined. The other option is to omit rows or columns with very few points.

EXAMPLE 2

A random sample of students at a certain university was classified according to the college in which they were enrolled and also according to whether they graduated from a high school in the state or out of the state. The results were put into a 2×4 contingency table.

	Engineering	Arts and Sciences	Home Economics	Other
In State	16	14	13	13
Out of State	14	6	10	8

Is there any correlation between the students' college choices and their status (in state or out of state)?

EXAMPLE (Flu Vaccine)

A survey was conducted to evaluate the effectiveness of a new flu vaccine that had been administrated in a community. The following table shows the outcomes of 1000 residents in the community.

	No Vaccine	One Shot	Two Shots	Total
Flu	24	9	13	46
No Flu	289	100	565	954
Total	313	109	578	1000

Do the data provide sufficient evidence to indicate that the two classifications (vaccine category and flu occurrence category) are dependent? Use α=0.01.

- The chi-squared test with fixed marginal totals

	Column 1	Column 2	$\cdots$	Column c	
Row 1	O_{11}	O_{12}	$\cdots$	O_{1c}	R_1
Row 2	O_{21}	O_{22}	$\cdots$	O_{2c}	R_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Row r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	R_r
	C_1	C_2	$\cdots$	C_c	N

Assumptions:

1. Each observation is classified into exactly one cell.
2. The row totals and column totals are fixed, not random.
3. Each observation has the same probability as every other observation of being classified into cell (i, j).

Statistical test:

$H_0 : p_{ij} = p_{i.}\, p_{.j}$ for all i and j

(The event "an observation is in row i" is independent of the event "an observation is in column j" for all i and j.)

$H_a : p_{ij} \neq p_{i.}\, p_{.j}$ for at least one pair of i and j

(The event "an observation is in row i" is not independent of the event "an observation is in column j" for all i and j.)

Test Statistic: $T = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$ $\left(E_{ij} = \frac{R_i C_j}{N}\right)$

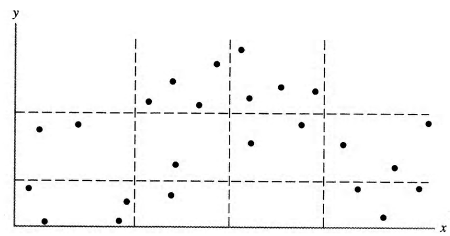
Another expression of T: $T = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{O_{ij}^2}{E_{ij}} - N$ $\left(E_{ij} = \frac{R_i C_j}{N}\right)$

Reject H_0 if $T_{(\text{obs})} > \chi^2_{(r-1)(c-1),\,1-\alpha}$

$p\text{-value} = P\left(T \geq T_{(\text{obs})}\right)$ $\left(T \sim \chi^2_{(r-1)(c-1)}\right)$

EXAMPLE 3

The chi-squared test with fixed marginal totals is use to determine whether or not two random variables X and Y are independent. The scatter plot of 24 (x, y) pairs is as follows.



The points are classified into 3×4 cells using horizontal lines and vertical lines. The result can be summarized by a 3×4 contingency table.

	Column 1	Column 2	Column 3	Column 4
Row 1	0	4	4	0
Row 2	2	1	2	3
Row 3	4	1	0	3

H_0 : There is no correlation between X and Y.

H_a : There exists correlation between X and Y.

Section 4.3 The median test

The median test is a special application of the chi-squared test with fixed marginal totals.

The purpose of the median test is to determine whether or not several population distributions possess the same median.

- The Median Test

  Suppose c population medians are to be compared. A random sample is drawn from each of the population distributions.

  H_0 : All the populations have the same median.

  H_a : At least two of population medians are different.

  M = grand median in all samples

  a = total number of observations above the grand median in all samples

  b = total number of observations less than or equal to the grand median

  N = total number of observations $(N = a + b)$

	Sample 1	Sample 2	$\cdots$	Sample c	
> Median	O_{11}	O_{12}	$\cdots$	O_{1c}	a
≤ Median	O_{21}	O_{22}	$\cdots$	O_{2c}	b
	n_1	n_2	$\cdots$	n_c	N

Test Statistic: $T = \dfrac{N^2}{ab}\sum_{j=1}^{c}\dfrac{\left(O_{1j} - \dfrac{a\,n_j}{N}\right)^2}{n_j}$

Another expression of T: $T = \dfrac{N^2}{ab}\sum_{j=1}^{c}\dfrac{O_{1j}^2}{n_i} - \dfrac{Na}{b}$

If $a = b$, then $T = \sum_{i=1}^{c}\dfrac{\left(O_{1j} - O_{2j}\right)^2}{n_i}$

Reject H_0 if $T_{(\text{obs})} > \chi^2_{c-1,1-\alpha}$

p-value $= P\left(T \geq T_{(\text{obs})}\right)$ $\left(T \sim \chi^2_{c-1}\right)$

EXAMPLE I

Four different methods of growing corn were randomly assigned to a large number of different plots of land and the yield per acre was computed for each plot.

	Method		
1	2	3	4
83	91	101	78
91	90	100	82
94	81	91	81
89	83	93	77
89	84	96	79
96	83	95	81
91	88	94	80
92	91		81
90	89		
	84		

The grand median of all samples is 89.

	Sample 1	Sample 2	Sample 2	Sample 4	
> 89	6	3	7	0	16
≤ 89	3	7	0	8	18

Section 4.4 The measure of dependency

- Use the χ^2 test statistic as a measure of dependency.

EXAMPLE 1 (This is an example in Section 4.2.)

A sample of students randomly selected from private high schools and a sample of students randomly selected from public high schools were given standardized achievement tests with the following results.

	0-275	276-350	351-425	426-500	
Private School	6	14	17	9	46
Public School	30	32	17	3	82

Is there any correlation between the test score and the type of school (private or public)?

- Cramer's Contingency Coefficient

	Column 1	Column 2	$\cdots$	Column c	
Row 1	O_{11}	O_{12}	$\cdots$	O_{1c}	R_1
Row 2	O_{21}	O_{22}	$\cdots$	O_{2c}	R_2
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Row r	O_{r1}	O_{r2}	$\cdots$	O_{rc}	R_r
	C_1	C_2	$\cdots$	C_c	N

$$R_1 = \frac{T}{N(q-1)} \qquad q = \min\{r, c\}$$

It can be shown that $R_1 \le 1$.

If the value of R_1 is close to 1, it is an indication that the row classification and the column classification are not independent.

Cramer's coefficient = $\sqrt{R_1}$

EXAMPLE 1 (Continued)

	0-275	276-350	351-425	426-500	
Private School	6	14	17	9	46
Public School	30	32	17	3	82
	36	46	34	12	128

- Pearson's Contingency Coefficient

$$R_2 = \sqrt{\frac{T}{N+T}}$$

- Pearson's Mean-Square Contingency Coefficient

$$R_3 = \frac{T}{N}$$

- Tschuprow's Coefficient

$$R_4 = \sqrt{\frac{T}{N\sqrt{(r-1)(c-1)}}}$$

Section 4.5 The chi squared goodness-of-fit test

Class 1	Class 2	$\cdots$	Class c	Total
O_1	O_2	$\cdots$	O_c	N

O_i = observed number (frequency) of points in class i $(i = 1, 2, \cdots, c)$

p_i = probability that a point will be in class i $(i = 1, 2, \cdots, c)$

E_i = expected number (frequency) of points in class i $(i = 1, 2, \cdots, c)$

$H_0 : p_1 = p_1^*,\ p_2 = p_2^*, \cdots, p_c = p_c^*$

$H_a : p_i \neq p_i^*$ for at least one i

Test statistic: $T = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}$

Another expression: $T = \sum_{i=1}^{c} \frac{O_i^2}{E_i} - N$

Class 1	Class 2	$\cdots$	Class c	Total
O_1	O_2	$\cdots$	O_c	N

Reject H_0 if $T_{(\text{obs})} > \chi^2_{c-1,1-\alpha}$.

p-value $= P\left(T \geq T_{(\text{obs})}\right) \quad \left(T \sim \chi^2_{c-1}\right)$

Minimum requirement for the χ^2 approximation to be valid:

1. All E_i's are at least 1.

2. No more than 20% of E_i's should be smaller than 5.

If the minimum requirement is not met, some cells should be combined. The other option is to omit cells with very few points.

EXAMPLE 1

A certain computer program is supposed to furnish random digits. If the program is accomplishing its purpose, the computer prints out digits (2, 3, 7, 4, etc.) that seem to be observations on independent and identically distributed random variables, where each digit $0, 1, 2, \cdots, 8, 9$ is equally likely (probability 0.1) to be obtained.

H_0 : The numbers appear to be random digits. $(p_0 = p_1 = \cdots = p_9 = 0.1)$

H_a : Some digits are more likely than others.

1578748416	4705188926	6936349612
4653843213	0282868892	3928057043
5101259393	9837006785	3011679938
7122863085	6528271107	2956427027
2671728075	9759178719	9373309535
8363265100	2546793732	2212122529
9453087720	3976759377	9593511031
5605373242	1819898287	3872181027
3494768396	9296177240	8620774591
4659773922	9246724287	8326143939

	0	1	2	3	4	5	6	7	8	9	Total
Observed	22	28	41	35	19	25	25	40	30	35	300

EXAMPLE 2

Efron and Morris (1975) presented data on the first 18 major league baseball players to have 45 times at bat in 1970. The players' names and the number of hits they got in their 45 times at bat are given as follows.

Clemente	18	Kessinger	13	Scott	10
F. Robinson	17	L. Alvarado	12	Petrocelli	10
F. Howard	16	Santo	11	E. Rodriguez	10
Johnstone	15	Swoboda	11	Campaneris	9
Berry	14	Unser	10	Munson	8
Spencer	14	Williams	10	Alvis	7

Test the null hypothesis that the data follow a binomial distribution with $n = 45$.

EXAMPLE 3

Fifty two-digit numbers were drawn at random from a telephone book, and the chi-squared test for goodness of fit is used to see if they could have been observations on a normally distributed random variable. The numbers, after being arranged in order from the smallest to the largest, are as follows.

23	23	24	27	29	31	32	33	33	35
36	37	40	42	43	43	44	45	48	48
54	54	56	57	57	58	58	58	58	59
61	*61*	62	63	64	65	66	68	68	70
73	73	74	75	*77*	81	87	89	93	97

H_0 : These numbers are from a normal distribution.

H_a : These numbers are not from a normal distribution.