Chapter 3 Some Tests Based on the Binomial Distribution

Section 3.1 The binomial test and estimation of p

Suppose X has a binomial distribution with total number of trials n and success probability p for each trial. The purpose of this section is to conduct statistical test about p and to construct confidence interval of p.

Statistical test about p:

Setting 1 (S1)	Setting 2 (S2)	Setting 3 (S3)
$H_0: p \le p^*$	$H_0: p \ge p^*$	$H_0: p = p^*$
$H_a: p > p^*$	$H_a: p < p^*$	$H_a: p \ne p^*$

Test statistic: $T = X$

Decision rules (critical region method):

(S1) Reject H_0 if $x_{(\text{obs})} > y_{1-\alpha}$. $(Y \sim b(n, p^*))$

(S2) Reject H_0 if $x_{(\text{obs})} \le y_{\alpha}$. $(Y \sim b(n, p^*))$

(S3) Reject H_0 if $x_{(\text{obs})} > y_{1-\alpha/2}$ or $x_{(\text{obs})} \le y_{\alpha/2}$. $(Y \sim b(n, p^*))$

Example (Problem of the critical region method)

Suppose X has a binomial distribution with $n = 20$ and parameter p. It is desired to test $H_0: p \le 0.5$ vs. $H_a: p > 0.5$ using $\alpha = 0.05$.

Decision rules (p-value method): Reject H_0 when p-value $< \alpha$.

(S1) $p\text{-value} = P\left(Y \ge x_{(\text{obs})}\right) \quad \left(Y \sim b(n, p^*)\right)$

(S2) $p\text{-value} = P\left(Y \le x_{(\text{obs})}\right) \quad \left(Y \sim b(n, p^*)\right)$

(S3) $p\text{-value} = 2\min\left\{P\left(Y \le x_{(\text{obs})}\right), P\left(Y \ge x_{(\text{obs})}\right)\right\} \quad \left(Y \sim b(n, p^*)\right)$

EXAMPLE 1

It is estimated that at least half of the men who currently undergo an operation to remove prostate cancer suffer from a particular undesirable side effect. In an effort to reduce the likelihood of this side effect the FDA studied a new method of performing the operation. Out of 19 operations only 3 men suffered the unpleasant side effect. Is it safe to conclude the new method of operating is effective in reducing the side effect? $(\alpha = 0.05)$

Decision rules $(n > 20)$ (critical region method):

(S1) Reject H_0 if $x_{(\text{obs})} > np^* + z_{1-\alpha}\sqrt{np^*(1-p^*)}$.

(S2) Reject H_0 if $x_{(\text{obs})} \le np^* + z_{\alpha}\sqrt{np^*(1-p^*)}$.

(S3) Reject H_0 if $x_{(\text{obs})} > np^* + z_{1-\alpha/2}\sqrt{np^*(1-p^*)}$ or

$x_{(\text{obs})} \le np^* + z_{\alpha/2}\sqrt{np^*(1-p^*)}$ $\left(x_{(\text{obs})} \le np^* - z_{1-\alpha/2}\sqrt{np^*(1-p^*)}\right)$.

Decision rules $(n > 20)$ (p-value method): Reject H_0 when p-value $< \alpha$.

(S1) $p\text{-value} = P\left(Z \ge \dfrac{x_{(\text{obs})} - np^* - 0.5}{\sqrt{np^*(1-p^*)}}\right)$

(S2) $p\text{-value} = P\left(Z \le \dfrac{x_{(\text{obs})} - np^* + 0.5}{\sqrt{np^*(1-p^*)}}\right)$

(S3) $p\text{-value} = 2\min\left\{P\left(Z \le \dfrac{x_{(\text{obs})} - np^* + 0.5}{\sqrt{np^*(1-p^*)}}\right), P\left(Z \ge \dfrac{x_{(\text{obs})} - np^* - 0.5}{\sqrt{np^*(1-p^*)}}\right)\right\}$

EXAMPLE 2

Under simple Mendelian inheritance a cross between plants of two particular genotypes may be expected to produce progeny one-fourth of which are "dwarf" and three-fourths of which are "tall." In an experiment to determine if the assumption of simple Mendelian inheritance is reasonable in a certain situation, a cross results in progeny having 243 dwarf plants and 682 tall plants. Do the data provide sufficient evidence to negate the assumption of simple Mendelian inheritance? $(\alpha = 0.05)$

Confidence interval of p :

1. $(n \le 30)$

   Use Table A4 to find lower and upper confidence limits directly.

2. $(n > 30)$

$$\left(\frac{Y}{n} - z_{1-\alpha/2}\sqrt{\frac{Y(n-Y)}{n^3}}, \frac{Y}{n} + z_{1-\alpha/2}\sqrt{\frac{Y(n-Y)}{n^3}}\right)$$

$$\left(\text{Another expression: } \left(\frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{Y(n-Y)}{n^3}}, \frac{Y}{n} + z_{1-\alpha/2}\sqrt{\frac{Y(n-Y)}{n^3}}\right)\right)$$

EXAMPLE 3

In a certain state 20 high schools were selected at random to see if they met the standards of excellence proposed by a national committee on education. It was found that 7 schools did qualify and accordingly were designated "excellent." What is a 95% confidence interval for p, the proportion of all high schools in the state that would qualify for the designation "excellent"?

p = the probability of a school being designated "excellent"

a. Using Table A4.

b. Using normal approximation.

Section 3.2 The quantile test and estimation of x_p

Let $X_1, X_2, \cdots, X_n$ be a random sample from a population distribution, and let x_p be the pth quantile of the population distribution. $\left(P\left(X \le x_p\right) = p\right)$

The purpose of this section is to conduct statistical test about x_p and to construct confidence interval for x_p.

Statistical test about x_p:

Setting 1 (S1)	Setting 2 (S2)	Setting 3 (S3)
$H_0: x_p \le x^*$	$H_0: x_p \ge x^*$	$H_0: x_p = x^*$
$H_a: x_p > x^*$	$H_a: x_p < x^*$	$H_a: x_p \ne x^*$

T_1 = number of observations less than or equal to x^*

T_2 = number of observations less than x^*

Setting 1. $H_0 : x_p \le x^*$ $\left(H_0 : P(X \le x^*) \ge p\right)$

$H_a : x_p > x^*$ $\left(H_a : P(X \le x^*) < p\right)$

Test statistic: T_1

Reject H_0 if $T_{1\,(\text{obs})} \le y_\alpha$. $\left(Y \sim b(n, p)\right)$

Setting 2. $H_0 : x_p \ge x^*$ $\left(H_0 : P(X \ge x^*) \ge 1 - p\right)$

$H_a : x_p < x^*$ $\left(H_a : P(X \ge x^*) < 1 - p\right)$ $\left(H_a : P(X < x^*) > p\right)$

Test statistic: T_2

Reject H_0 if $T_{2\,(\text{obs})} > y_{1-\alpha}$. $\left(Y \sim b(n, p)\right)$

Setting 3. $H_0 : x_p = x^*$ $\left(H_0 : P(X \le x^*) \ge p \text{ and } P(X \ge x^*) \ge 1 - p\right)$

$H_a : x_p \ne x^*$ $\left(H_a : P(X > x^*) < p \text{ or } P(X \ge x^*) < 1 - p\right)$

Test statistics: T_1 and T_2

Reject H_0 if $T_{1\,(\text{obs})} \le y_{\alpha/2}$ or $T_{2\,(\text{obs})} > y_{1-\alpha/2}$. $\left(Y \sim b(n, p)\right)$

EXAMPLE 1

Entering college freshmen have taken a particular high school achievement examination for many years, and the upper quartile is well established at a score of 193. A particular high school sends 15 of its graduates to college, where they take the exam and get the following scores:

189, 233, 195, 160, 212, 176, 231, 185, 199, 213, 202, 193, 174, 166, 248.

It is assumed that these 15 students represent a random sample of all students from that high school who go on to college. One way of comparing college students from that high school with other college students is by testing the hypothesis that the above scores come from a population whose upper quartile is 193. Do the data provide sufficient evidence to conclude that the population upper quartile is not 193? ($\alpha = 0.05$)

p-value of the test:

(S1) $p-\text{value}=P\left(Y\le T_{1(\text{obs})}\right)\ \left(Y\sim b(n,p)\right)$

(S2) $p-\text{value}=P\left(Y\ge T_{2(\text{obs})}\right)\ \left(Y\sim b(n,p)\right)$

(S3) $p-\text{value}=2\min\left\{P\left(Y\le T_{1(\text{obs})}\right),P\left(Y\ge T_{2(\text{obs})}\right)\right\}\ \left(Y\sim b(n,p)\right)$

EXAMPLE 1 (Continue)

$$p-\text{value}=2\min\left\{P\left(Y\le T_{1(\text{obs})}\right),P\left(Y\ge T_{2(\text{obs})}\right)\right\}\quad \left(Y\sim b(15,0.75)\right)$$
$$=2\min\left\{P(Y\le 7),P(Y\ge 6)\right\}=2\times\min\left\{0.0173,0.9992\right\}=0.0346$$

Decision rules $(n>20)$ (critical region method):

(S1) Reject H_0 if $T_{1,\,(\text{obs})}\le np+z_{\alpha}\sqrt{np(1-p)}$.

(S2) Reject H_0 if $T_{2\,(\text{obs})}>np+z_{1-\alpha}\sqrt{np(1-p)}$.

(S3) Reject H_0 if $T_{1\,(\text{obs})}\le np+z_{\alpha/2}\sqrt{np(1-p)}$ or

$T_{2\,(\text{obs})}>np+z_{1-\alpha/2}\sqrt{np(1-p)}$.

Decision rules $(n>20)$ (p-value method): Reject H_0 when p-value $<\alpha$.

(S1) $p\text{-value}=P\left(Z\le\dfrac{T_{1\,(\text{obs})}-np+0.5}{\sqrt{np(1-p)}}\right)$

(S2) $p\text{-value}=P\left(Z\ge\dfrac{T_{2\,(\text{obs})}-np-0.5}{\sqrt{np(1-p)}}\right)$

(S3) $p\text{-value}=2\min\left\{P\left(Z\le\dfrac{T_{1\,(\text{obs})}-np+0.5}{\sqrt{np(1-p)}}\right),P\left(Z\ge\dfrac{T_{2\,(\text{obs})}-np-0.5}{\sqrt{np(1-p)}}\right)\right\}$

EXAMPLE 2

The time interval between eruptions of Old Faithful geyser is recorded 112 times to see whether the median interval is less than or equal to 60 minutes (null hypothesis) or whether the median interval is greater than 60 minutes (alternative hypothesis). Out of 112 recorded times, 8 are 60 minutes or less. Run a statistical test at level of significance 0.05.

Confidence interval for x_p:

1. $(n \le 20)$ Use the $b(n, p)$ distribution table (Table A3).

   To find the lower confidence limit, read down the column for p until reaching an entry approximately equal to $\alpha/2$. Let the corresponding value of y in the table be $r-1$. Then the lower confidence limit of a $1-\alpha$ confidence interval of x_p is $X_{(r)}$.

   To find the upper confidence limit, read down the column for p until reaching an entry approximately equal to $1-\alpha/2$. Let the corresponding value of y in the table be $s-1$. Then the upper confidence limit of a $1-\alpha$ confidence interval of x_p is $X_{(s)}$.

   $P\left(X_{(r)} < x_p < X_{(s)}\right) \ge 1-\alpha$

   If the population distribution is continuous, then $P\left(X_{(r)} < x_p < X_{(s)}\right) = 1-\alpha$.

2. $(n > 20)$

$$r^* = np + z_{\alpha/2}\sqrt{np(1-p)}$$

$$s^* = np + z_{1-\alpha/2}\sqrt{np(1-p)}$$

A $1-\alpha$ confidence interval of x_p is $\left(X_{(r^*)}, X_{(s^*)}\right)$.

Usually, r^* and s^* are not integers.

If r^* and s^* are not integers, use the largest integer which is less than r^* and the smallest integer which is greater than s^*.

EXAMPLE 3

Sixteen transistors are selected at random from a large batch of transistors and are tested. The number of hours until failure is recorded for each one. The ordered data are listed here.

46.5, 47.2, 49.1, 56.5, 56.8, 59.2,59.9, 63.2, 63.3, 63.4, 63.7, 64.1, 67.1
67.7, 73.3, 78.5

Find a confidence interval for the upper quartile, with a confidence coefficient close to 90%.

Section 3.4 The sign test

Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_{n'}, Y_{n'})$ be independent bivariate random variables.

A pair (X_i, Y_i) is classified as "+" if $X_i < Y_i$ $(i = 1, 2, \cdots, n')$.

A pair (X_i, Y_i) is classified as "−" if $X_i > Y_i$ $(i = 1, 2, \cdots, n')$.

A pair (X_i, Y_i) is classified as "0" or "tie" if $X_i = Y_i$ $(i = 1, 2, \cdots, n')$.

The pairs are assumed to be internally consistent.

If $P(+) > P(-)$ for one pair, then $P(+) > P(-)$ for all pairs.

If $P(+) < P(-)$ for one pair, then $P(+) < P(-)$ for all pairs.

If $P(+) = P(-)$ for one pair, then $P(+) = P(-)$ for all pairs.

Statistical tests:

Setting 1 (S1)	Setting 2 (S2)	Setting 3 (S3)
$H_0 : P(+) \le P(-)$	$H_0 : P(+) \ge P(-)$	$H_0 : P(+) = P(-)$
$H_a : P(+) > P(-)$	$H_a : P(+) < P(-)$	$H_a : P(+) \ne P(-)$

Disregard all tied pairs.

$T =$ number of "+" signs

Let n be the total number of "+" signs and "−" signs.

Setting 2. $H_0 : P(+) \geq P(-)$ $H_a : P(+) < P(-)$

Reject H_0 if $T_{(obs)} \leq y_\alpha$ $(Y \sim b(n, 0.5))$.

p-value $= P(Y \leq T_{(obs)})$ $(Y \sim b(n, 0.5))$

Setting 1. $H_0 : P(+) \leq P(-)$ $H_a : P(+) > P(-)$

Reject H_0 if $T_{(obs)} \geq n - y_\alpha$ $(Y \sim b(n, 0.5))$

Note: "$T_{(obs)} \geq n - y_\alpha$" is equivalent to "$n - T_{(obs)} \leq y_\alpha$".

p-value $= P(Y \geq T_{(obs)})$ $(Y \sim b(n, 0.5))$

Setting 3. $H_0 : P(+) = P(-)$ $H_a : P(+) \neq P(-)$

Reject H_0 if $T_{(obs)} \leq y_{\alpha/2}$ or $T_{(obs)} \geq n - y_{\alpha/2}$.

p-value $= 2 \min\{P(Y \geq T_{(obs)}), P(Y \leq T_{(obs)})\}$ $(Y \sim b(n, 0.5))$

EXAMPLE 1

An item A is manufactured using a certain process. Item B serves the same function as A but is manufactured using a new process. The manufacturer wishes to determine whether B is preferred to A by the consumer, so she selects a random sample consisting of 10 consumers, gives each of them one A and one B, and asks them to use the items for some period of time. At the end of the allotted period of time the consumers report their preferences to the manufacturer. Eight consumers preferred B to A, one preferred A to B, and one reported "no preference." Can the manufacturer conclude that the consumer population prefers B to A? Use $\alpha = 0.05$.

Large sample case: $(n > 20)$

(S1) Reject H_0 if $T_{(\text{obs})} \geq n - \left(np + z_\alpha \sqrt{np(1-p)}\right) = \frac{n}{2} - z_\alpha \frac{\sqrt{n}}{2}$.

$$p\text{-value} = P\left(Z \geq \frac{T_{(\text{obs})} - np - 0.5}{\sqrt{np(1-p)}}\right) = P\left(Z \geq \frac{2T_{(\text{obs})} - n - 1}{\sqrt{n}}\right)$$

(S2) Reject H_0 if $T_{(\text{obs})} \leq np + z_\alpha \sqrt{np(1-p)} = \frac{n}{2} + z_\alpha \frac{\sqrt{n}}{2}$.

$$p\text{-value} = P\left(Z \leq \frac{T_{(\text{obs})} - np + 0.5}{\sqrt{np(1-p)}}\right) = P\left(Z \leq \frac{2T_{(\text{obs})} - n + 1}{\sqrt{n}}\right)$$

(S3) Reject H_0 if $T_{(\text{obs})} \leq \frac{n}{2} + z_{\alpha/2} \frac{\sqrt{n}}{2}$ or $T_{(\text{obs})} \geq \frac{n}{2} - z_{\alpha/2} \frac{\sqrt{n}}{2}$.

$$p\text{-value} = 2\min\left\{P\left(Z \leq \frac{2T_{(\text{obs})} - n + 1}{\sqrt{n}}\right), P\left(Z \geq \frac{2T_{(\text{obs})} - n - 1}{\sqrt{n}}\right)\right\}$$

EXAMPLE 2

In what was perhaps the first published report of a nonparametric test, Arbuthnott (1710) examined the available London birth records of 82 years and for each year compared the number of males born with the number of females born, and found that the number of males born were all greater than the number of females born. For each year, denote the event "more males than females were born" by "+" and the opposite event by "−" (there were no ties). Test $H_0: P(+) = P(-)$ vs. $H_a: P(+) \neq P(-)$ using $\alpha = 0.05$. Find the p-value of the test.

EXAMPLE 3

Ten homing pigeons were taken to a point 25 kilometers west of their loft and released singly to see whether they dispersed at random in all directions (the null hypothesis) or whether they tended to proceed eastward toward their loft. Field glasses were used to observe the birds until they disappeared from view, at which time the angle of the vanishing point was noted. These 10 angles are:

20, 35, 350, 120, 85, 345, 80, 320, 280, and 85 degrees.

Is there sufficient evidence to indicate that the pigeons tend to fly homeward? Find the p-value of the test.

Use the sign test to compare two population medians:

Let "+" represent case "$Y > X$", and let "−" represent case "$Y < X$".

Setting 1. $H_0 : M_Y \le M_X$ vs. $H_a : M_Y > M_X$

This test is equivalent to $H_0 : P(+) \le P(-)$ vs. $H_a : P(+) > P(-)$.

Setting 2. $H_0 : M_Y \ge M_X$ vs. $H_a : M_Y < M_X$

This test is equivalent to $H_0 : P(+) \ge P(-)$ vs. $H_a : P(+) < P(-)$.

Setting 3. $H_0 : M_Y = M_X$ vs. $H_a : M_Y \ne M_X$

This test is equivalent to $H_0 : P(+) = P(-)$ vs. $H_a : P(+) \ne P(-)$.

Section 3.5 Some variation of the sign test

- McNemar test for significance of changes

  Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_{n'}, Y_{n'})$ be independent bivariate random variables.

  The possible values of (X_i, Y_i) are $(0,0), (0,1), (1,0), (1,1)$ $(i = 1, 2, \cdots, n')$.

  The pairs are assumed to be internally consistent.

  If $P(X_i = 0) > P(Y_i = 0)$ for one pair, then it is true for all pairs.

  If $P(X_i = 0) < P(Y_i = 0)$ for one pair, then it is true for all pairs.

  Similarly, if $P(X_i = 1) > P(Y_i = 1)$ for one pair, then the same is true for all pairs; if $P(X_i = 1) < P(Y_i = 1)$ for one pair, then the same is true for all pairs.

  Statistical test:

  $H_0 : P(X = 0) = P(Y = 0)$ vs. $H_a : P(X = 0) \neq P(Y = 0)$

  ($H_0 : P(X = 1) = P(Y = 1)$ vs. $H_a : P(X = 1) \neq P(Y = 1)$)

$P(X=0)=P(Y=0)$ is equivalent to $P(X=0, Y=1)=P(X=1, Y=0)$.

The test becomes

$H_0: P(X=0, Y=1)=P(X=1, Y=0)$

$H_a: P(X=0, Y=1)\neq P(X=1, Y=0)$.

	$Y=0$	$Y=1$
$X=0$	a pairs	b pairs
$X=1$	c pairs	d pairs

Small sample case $(n=b+c\leq 20)$

Test statistic: $T_2=b$

Reject H_0 if $T_{2\,(\text{obs})}\leq y_{\alpha/2}$ or $T_{2\,(\text{obs})}>n-y_{\alpha/2}$ $\left(Y\sim b(n,0.5)\right)$.

p-value $=2\min\{P(Y\leq b), P(Y\geq b)\}$ $\left(Y\sim b(n,0.5)\right)$

Large sample case $(n=b+c>20)$

Test statistic: $T_1=\dfrac{(b-c)^2}{b+c}$

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

p-value $=P\left(\chi^2_1\geq T_{1(\text{obs})}\right)$

EXAMPLE 1

Prior to a nationally televised debate between the two presidential candidates, a random sample of 100 persons stated their choice of candidates as follows. Eighty-four persons favored the Democratic candidate, and the remaining 16 favored the Republican. After the debate the same 100 people expressed their preference again. Of the persons who formerly favored the Democrat, exactly one-fourth of them changed their minds, and also one-fourth of the people formerly favoring the Republican switched to the Democratic side. The results are summarized in the following 2×2 contingency table.

		After		
		Democrat	Republican	
Before	Democrat			84
	Republican			16

- Cox and Stuart test for trend

  Let $X_1, X_2, \cdots, X_{n'}$ be a sequence of random variables arranged in a particular order, such as the order in which the random variables are observed. The purpose of the test is to check if a trend exists in the sequence.

Setting 1 (S1)

H_0 : There is no trend in the sequence.

H_a : The later random variables are likely to be greater than the earlier ones. (An upward trend exists in the sequence.)

Setting 2 (S2)

H_0 : There is no trend in the sequence.

H_a : The later random variables are likely to be smaller than the earlier ones. (A downward trend exists in the sequence.)

Setting 3 (S3)

H_0 : There is no trend in the sequence.

H_a : The later random variables are likely to be either greater than or smaller than the earlier ones.

(An upward trend or a downward trend exists in the sequence.)

Let $c = \begin{cases} n/2 & \text{if } n \text{ is even} \\ (n+1)/2 & \text{if } n \text{ is odd}. \end{cases}$

Group $X_1, X_2, \cdots, X_{n'}$ into pairs $(X_1, X_{1+c}), (X_2, X_{2+c}), \cdots, (X_{n'-c}, X_{n'})$.

Class (X_i, X_j) as "+" if $X_i < X_j$ and class (X_i, X_j) as "−" if $X_i > X_j$.

Eliminate tied pairs. Let n be the total number of untied pairs.

Test statistic: T = total number of "+" signs

Setting 1 (S1)

H_0 : There is no trend in the sequence.

H_a : An upward trend exists in the sequence.

Reject H_0 if $T_{(\text{obs})} > n - y_\alpha$ $\left(Y \sim b(n, 0.5)\right)$.

p-value $= P\left(Y \ge T_{(\text{obs})}\right)$ $\left(Y \sim b(n, 0.5)\right)$

Setting 2 (S2)

H_0 : There is no trend in the sequence.

H_a : A downward trend exists in the sequence.

Reject H_0 if $T_{(\text{obs})} \le y_\alpha$ $\left(Y \sim b(n, 0.5)\right)$.

p-value $= P\left(Y \le T_{(\text{obs})}\right)$ $\left(Y \sim b(n, 0.5)\right)$

Setting 3 (S3)

H_0 : There is no trend in the sequence.

H_a : An upward trend or a downward trend exists in the sequence.

Reject H_0 if $T_{(\text{obs})} > n - y_{\alpha/2}$ or $T_{(\text{obs})} \leq y_{\alpha/2}$ $\left(Y \sim b(n, 0.5)\right)$.

p-value $= 2 \min\left\{P\left(Y \leq T_{(\text{obs})}\right), P\left(Y \geq T_{(\text{obs})}\right)\right\}$ $\left(Y \sim b(n, 0.5)\right)$

When $n > 20$, normal approximation needs to be used.

EXAMPLE 2

Total annual precipitation is recorded yearly for 19 years. This record is examined to see if the amount of precipitation is tending to increase or decrease. The precipitation in inches was

45.25, 45.83, 41.77, 36.26, 45.37, 52.25, 35.37, 57.16, 35.37, 58.32, 41.05, 33.72, 45.73, 37.90, 41.72, 36.07, 49.83, 36.24, and 39.90.

Is there sufficient evidence to indicate a trend (either increasing or decreasing) in the sequence?

EXAMPLE 3

On a certain stream the average rate of water discharge is recorded each month (in cubic feet per second) for a period of 24 months.

H_0 : The rate of discharge is not decreasing.
H_a : The rate of discharge is decreasing.

The rate of discharge is known to follow a yearly cycle so that nothing is learned by pairing stream discharges for two different months. However, by pairing the same months in two successive years the existence of a trend can be investigated. The following data were collected.

Month	First Year	Second Year	Month	First Year	Second Year
Jan	14.6	14.2	Jul	92.8	88.1
Feb	12.2	10.5	Aug	74.4	80.0
Mar	104	123	Sep	75.4	75.6
Apr	220	190	Oct	51.7	48.8
May	110	138	Nov	29.3	27.1
Jun	86.0	98.1	Dec	16.0	15.7

What is the conclusion of the test?

EXAMPLE 4

Cochran (1937) compares the reactions of several patients with each of two drugs, to see if there is a positive correlation between the two reactions for each patient.

Patient	Drug 1	Drug 2	Patient	Drug 1	Drug 2
1	+0.7	+1.9	6	+3.4	+4.4
2	−1.6	+0.8	7	+3.7	+5.5
3	−0.2	+1.1	8	+0.8	+1.6
4	−1.2	+0.1	9	0.0	+4.6
5	−0.1	−0.1	10	+2.0	+3.4

Ordering the pairs according to the reaction from drug 1 gives

Patient	Drug 1	Drug 2	Patient	Drug 1	Drug 2
2	−1.6	+0.8	1	+0.7	+1.9
4	−1.2	+0.1	8	+0.8	+1.6
3	−0.2	+1.1	10	+2.0	+3.4
5	−0.1	−0.1	6	+3.4	+4.4
9	0.0	+4.6	7	+3.7	+5.5