

15.37 The runs test is used to test the null hypothesis of randomness. Either a large or a small number of runs indicates non randomness, and a two-tailed test is used. The data are shown below:

W, W, W, W, B, W, W, W, B, B, W, B, B

There are $n_1 = 5$ blacks hired and $n_2 = 8$ whites hired. The number of runs observed is $R = 6$. From Table 10, the p -value is $2P(R \leq 6) = 2(.347) = .694$. We do not reject the null hypothesis of randomness. The data do not suggest a nonrandom racial selection in the hiring of the union's members.

15.39a. In this exercise it is necessary to calculate $P(R \leq 11)$, where $n_1 = 11$ and $n_2 = 23$. Since it is known that the quantity

$$z = \frac{R - E(R)}{\sigma_R}$$

is approximately normally distributed for large n_1 and n_2 (say, $n_1 \geq 10, n_2 \geq 10$), we may use the normal approximation to calculate $P(R \leq 11)$. The first step is to

determine the z value corresponding to an R value of 11. Note that

$$E(R) = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(11)(23)}{11+23} + 1 = 15.88$$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} = \frac{2(11)(23)[2(11)(23) - 11 - 23]}{(11+23)^2(11+23-1)} = \frac{238,832}{38,148}$$

$$= 6.2607$$

$$\sigma_R = \sqrt{6.2607} = 2.50$$

Hence the corresponding z value will be

$$z = \frac{R - E(R)}{\sigma_R} = \frac{11 - 15.88}{2.50} = -1.95$$

Thus,

$$P(R \leq 11) \approx P(Z \leq -1.95) = P(Z \geq 1.95) = .0256$$

b. The hypothesis to be tested is

H_0 : randomness of occurrence vs. H_a : non randomness of occurrence

The test statistic is R , the number of runs observed. The reader may verify that the observed value of R is $R = 11$, with $n_1 = 11$ and $n_2 = 23$. Using a large-sample approximation, the standardized test statistic is

$$z = \frac{R - E(R)}{\sigma_R} = -1.95 \quad (\text{calculated above})$$

Since an unusually small or unusually large number of runs would imply a non randomness of defectives, a two-tailed test is employed and the rejection region is $z < -1.96$ or $z > 1.96$. Since the test statistic, $z = -1.95$, does not fall in the rejection region, the null hypothesis is not rejected. Hence, there is not sufficient evidence of a non randomness of defectives.

15.40a. The hypothesis to be tested is

H_0 : process is stable (random fluctuation) vs. H_a : process is unstable (nonrandom fluctuation)

and the test statistic is R , the number of runs observed. The mean of the 16 measurements is calculated to be

$$\bar{y} = \frac{\sum y_i}{n} = \frac{1082.7}{16} = 67.67$$

The measurements are classified as A if they lie above the mean and as B if they fall below. The sequence of runs generated by using this procedure is

A A A A A B B B B B B A B A B A

Notice that the observed value of the test statistic is $R = 7$, with $n_1 = 8$, $n_2 = 8$. Nonrandom fluctuation will be implied by a small number of runs. (That is, the process would be higher than the mean for a period of time and then would drop below the mean for a longer period of time. This would imply a lack of stability in the process.) Consulting Table 10, with $n_1 = 8$ and $n_2 = 8$, the p -value is $P(R \leq 7) = .214$. Hence the null hypothesis is not rejected.

b. If the time period is divided into two equal parts, the data appear as shown in the table at the right. The hypothesis to be tested (using the t test described in Section 10.6) is	Group I	Group II
	68.2	65.3
	71.6	64.2
	69.3	67.6
	71.6	68.6
$H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$	70.4	66.8
	65.0	68.9
Then	63.6	66.8
	64.7	70.1

$$\bar{y}_1 = \frac{\sum y_{1j}}{n_1} = \frac{544.4}{8} = 68.05 \qquad \bar{y}_2 = \frac{\sum y_{2j}}{n_2} = \frac{538.30}{8} = 67.29$$

$$s^2 = \frac{\sum_j y_{1j}^2 - \frac{(\sum_j y_{1j})^2}{n_1} + \sum_j y_{2j}^2 - \frac{(\sum_j y_{2j})^2}{n_2}}{n_1 + n_2 - 2} = \frac{37,119.06 - \frac{(544.4)^2}{8} + 36,247.15 - \frac{(538.30)^2}{8}}{14} = \frac{72.64 + 26.29}{14} = 7.066$$

The test statistic is

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{68.05 - 67.29}{\sqrt{7.066 \left(\frac{1}{8} + \frac{1}{8} \right)}} = \frac{.76}{1.328} = .57$$

Using a two-tailed rejection region (with $\alpha = .10$), the critical value of t is $t_{.05,14} = 1.761$, and the rejection region is $|t| > 1.761$. The null hypothesis cannot be rejected.

15.41 Let A represent an observation from population A , and let B represent an observation from population B . Now referring to Exercise 15.18, the observations are arranged according to rank, and the population from which they were drawn is noted. Using the ranks obtained in Exercise 15.18 to arrange the observations, the sequence of runs is as follows:

A B A B A B B B A B B A A B A B A A

Notice that the 9th and 10th and the 13th and 14th letters in the sequence represent the two pairs of tied observations. If the tied observations were reversed in the sequence of runs, we would still obtain $R = 13$. Hence the order of the tied observations is irrelevant.

Consider the alternative situation that asserts that the two distributions are not identical. If the alternative is true, we would expect a small number of runs because most of the measurements for population A will fall below those for population B (or vice versa). Hence small values for R will tend to contradict the null hypothesis. A one-tailed test of hypothesis is employed with a lower-tailed rejection region. Table 10 is then used to find the p -value. For $n_1 = n_2 = 9$ and $R = 13$, the p -value is $P(R < 13) = .956$. The null hypothesis is not rejected. This is the same conclusion that was reached when the Mann-Whitney U test was employed.

15.42 Refer to Exercise 15.19. The test statistic is R , the number of runs observed. If the alternative is true, and the experimental batteries have greater mean life, then most of the observations from plant B will be smaller than those for plant A . Consequently, the expected number of runs will be small.

To use a large-sample test, the values for R , $E(R)$, and $V(R)$ must be determined. Then when n_1 and n_2 are large, the quantity $Z = \frac{R - E(R)}{\sigma_R}$ will be approximately normally distributed with mean 0 and variance 1. Since we are interested in a one-tailed test of hypothesis (that is, a small number of runs will tend to contradict the null hypothesis), the rejection region will be $z \leq -1.645$.

$$E(R) = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(15)(16)}{30} + 1 = 16$$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} = \frac{2(15)(16)[2(15)(16) - 15 - 16]}{(30)^2(29)} = \frac{210}{29} = 7.24137$$

$$\sigma_R = \sqrt{7.24137} = 2.69$$

The reader may verify that $R = 15$, and the standardized test statistic is

$$z = \frac{R - E(R)}{\sigma_R} = \frac{15 - 16}{2.69} = -.37$$

Hence, the null hypothesis is not rejected, and there is insufficient evidence to indicate a difference between the distributions of the two types of batteries.

11.1 Calculate the following:

$$\begin{aligned} \sum x_i &= 0 & \sum y_i &= 7.5 & \sum x_i y_i &= -6 \\ \sum x_i^2 &= 10 & \sum y_i^2 &= 15.25 & n &= 5 \\ S_{xy} &= \sum x_i y_i - \frac{1}{n} (\sum x_i) (\sum y_i) = -6 \\ S_{xx} &= \sum x_i^2 - \frac{1}{n} (\sum x_i)^2 = 10 \end{aligned}$$

Then

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = -\frac{6}{10} = -.6$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{7.5}{5} - 0 = 1.5.$$

The least squares straight line is $\hat{y} = 1.5 - .6x$

The observed points and the fitted lines are shown in Figure 11.1.

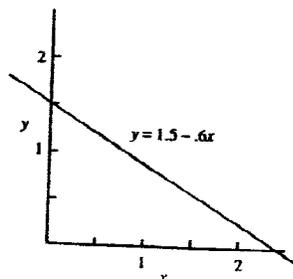


Figure 11.1

11.3 Calculate

$$\begin{aligned} \sum x_i &= 36 & \sum y_i &= 346.9 & \sum x_i y_i &= 1764.4 \\ \sum x_i^2 &= 204 & \sum y_i^2 &= 16,045.29 & n &= 8 \\ S_{xy} &= 203.35 & S_{xx} &= 42 \end{aligned}$$

Then

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{54,243}{54,714} = 4.84167$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 43.3625 - 21.7875 = 21.575.$$

The least squares straight line is $\hat{y} = 21.575 + 4.842x$. The positive slope suggests an increase in sales over time. In particular we expect sales to increase by 4,842 dollars per year.

11.5 a. The data are plotted in Figure 11.2.

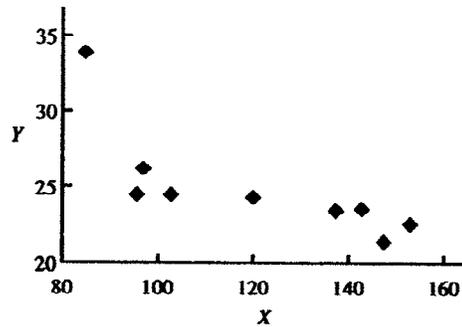


Figure 11.2

b. Calculate

$$\begin{aligned} \sum x_i &= 1076 & \sum y_i &= 216 & \sum x_i y_i &= 25,431 \\ \sum x_i^2 &= 133,336 & \sum y_i^2 &= 5228 & n &= 9 \\ S_{xy} &= -393.0 & S_{xx} &= 4694.22 & & \end{aligned}$$

Then

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = -.0837$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 24 + 10.0092 = 34.0092$$

c. The least squares line is graphed in Figure 11.3.

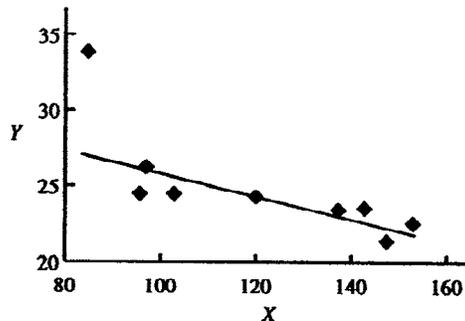


Figure 11.3

d. The estimate of y when $x = 125$ is $\hat{y} = 34.0092 - .0837(125) = 23.5467$.

11.6 We need to minimize $SSE = \sum_{i=1}^n [y_i - \hat{\beta}_1 x_i]^2$. Consider

$$\begin{aligned} \frac{d SSE}{d \hat{\beta}_1} &= - \sum_{i=1}^n 2 [y_i - \hat{\beta}_1 x_i] x_i = 0 \\ &= -2 \left[\sum_{i=1}^n (x_i y_i - \hat{\beta}_1 x_i^2) \right] = 0 \\ \Rightarrow \sum_{i=1}^n x_i y_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= 0 \end{aligned}$$

Implying

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$