

## Exercise 3 Solutions

1. (a) Let  $X_j$  be the random variables that generate the increase. The sample size is small (only 8). So we have to assume the  $X_j$  are from a normal population. With this assumption, a 95% two-sided confidence interval for the increase is given by

$$\left[ \bar{X} - t_{\alpha/2}(n-1)S/\sqrt{n}, \quad \bar{X} + t_{\alpha/2}(n-1)S/\sqrt{n} \right],$$

where  $\bar{X}$  and  $S$  are the sample mean and the sample variance respectively.

We have

$$\bar{x} = \frac{11}{8}, \quad t_{0.025}(7) = 2.356$$

$$s^2 = \frac{1}{n-1} \left[ \sum_{j=1}^n x_j^2 - n(\bar{x})^2 \right] = \frac{1}{7}(49 - 121/8) = 4.84.$$

Hence a 95% two-sided confidence interval for the increase is  $[-0.465, 3.215]$ .

- (b) The hypothesis at the level  $\alpha = 0.05$  to be tested is

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0,$$

where  $\mu$  denotes the average increase. Note that 0 lies in the 95% two-sided confidence interval given in (b). According to the duality between confidence intervals and tests, we should accept the claim that there is no evidence to indicate the increase between pulse rates at the level  $\alpha = 0.05$ .

2. Let  $X$  denote the number having the disease in the sample,  $X \sim B(n, p)$ .  
Test  $H_0 : p = 0.1$  against the alternative  $H_1 : p > 0.1$ . at significance level 0.05.

For large  $n$ ,  $X$  is approximately  $N(0.1n, 0.09n)$ . The approximate rejection region is

$$X > 0.1n + 1.645\sqrt{0.09n}.$$

If  $p = 0.13$  then  $X$  is approximately  $N(0.13n, 0.1131n)$ . We want

$$0.95 \leq P(X > 0.1n + 1.645\sqrt{0.09n} | p = 0.13)$$

Thus

$$-1.645 \geq (0.1n + 1.645\sqrt{0.09n} - 0.13n) / \sqrt{0.1131n}$$

$$n \geq (34.8906)^2 = 1217.36$$

A sample of at least 1218 is needed.

3.  $X_1, \dots, X_n \sim B(1, p)$ . We want to test  $H_0 : p = p_0$  against  $H_1 : p = p_1 (> p_0)$ . The rejection region according to the Neyman-Pearson Lemma is

$$\frac{\prod_{i=1}^n p_0^{x_i} (1 - p_0)^{1-x_i}}{\prod_{i=1}^n p_1^{x_i} (1 - p_1)^{1-x_i}} < c$$

for some constant  $c$ . This is equivalent to

$$x \log \left( \frac{p_0(1 - p_1)}{p_1(1 - p_0)} \right) < k,$$

where  $x = \sum_{i=1}^n x_i$ . Since  $\log \left( \frac{p_0(1-p_1)}{p_1(1-p_0)} \right) < 0$  the rejection region is

$$x > k_1,$$

for some constant  $k_1$ . The (approximate) rejection region, using the normal approximation is

$$X > np_0 + z_{1-\alpha} \sqrt{np_0(1 - p_0)}.$$

The rejection region does not depend on the value of  $p_1$  and so the test is uniformly most powerful against  $H_1 : p > p_0$ .

4. Let  $X$  be the breaking point of the string in standard technique and  $Y$  be the breaking point of the string in new technique. For the two-sample  $t$ -test and we have to assume

$$X \sim N(\mu_X, \sigma_X^2) \quad \text{and} \quad Y \sim N(\mu_Y, \sigma_Y^2).$$

Note that  $\mu_X$  and  $\mu_Y$  are the mean breaking points of the string in the standard technique and the new technique respectively.

- (a) Test  $H_0 : \sigma_X^2 = \sigma_Y^2$  against the general alternative.

We have

- i.  $\bar{x} = 138, \quad \bar{y} = 143;$
- ii.  $s_x^2 = 133.5, \quad s_y^2 = 68.5;$
- iii.  $s^2 = (4 \times s_x^2 + 4 \times s_y^2)/8 = 101,$

The test statistic is  $f = 133.5/68.5 = 1.949$ . The  $p$ -value is

$$p = 2 * P(F_{4,4} \geq 1.949) = 0.5340.$$

The data are consistent with the two populations having common variance. This test requires the normal assumption for the shape of the populations.

- (b) The hypothesis to be tested is

$$H_0 : \mu_X = \mu_Y \quad \text{vs} \quad H_1 : \mu_X < \mu_Y.$$

$$t_{m,n} = (\bar{x} - \bar{y}) / \left( s \sqrt{1/5 + 1/5} \right) = -0.786646.$$

The  $p$ -value is given by

$$p = \mathbf{P}(T_8 \leq -0.786646) = 0.22708.$$

**Conclusion:** The data are consistent with the null hypothesis, that is, there is no evidence to support the contention of the research department.