# EXERCISES IN STATISTICS

## Series A, No. 9

1. The average length of a finger bone of 10 fossil skeletons of the proconsul hominid is 3.73cm, and the standard deviation is 0.34cm. Find 80% and 90% confidence intervals for the mean length of the bone in the species.

**Answer.** We are told that $\bar{x}_{10} = 3.73$ and that $s_{10} = \sqrt{\sum(x_i - \bar{x})^2/n} = 0.34$. On the assumption that $x_i \sim N(\mu, \sigma^2); i = 1, \ldots, n$ is a normally distributed random sample, there are

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{and} \qquad \frac{\bar{x} - \mu}{\sigma\sqrt{n}} \sim N(0, 1).$$

Since $\sigma^2$ is unknown, we must use

$$\hat{\sigma}^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = s^2 \frac{n}{n-1} \quad \text{whence} \quad \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n-1}} = \frac{0.34}{3}.$$

From the result that

$$\frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} = \left\{ \frac{(\bar{x} - \mu)}{\sigma\sqrt{n}} \Bigg/ \sqrt{\frac{\sum(x_i - \bar{x})^2}{\sigma^2(n-1)}} \right\} \sim \frac{N(0, 1)}{\sqrt{\frac{\chi^2(n-1)}{(n-1)}}} = t(n-1),$$

we may derive a probability statement of the form

$$P\left(\bar{x} - b\frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{x} + b\frac{\hat{\sigma}}{\sqrt{n}}\right) = Q.$$

The $t(9)$ tables indicate that

$$P\{t(9) > b\} = 0.95 \Longrightarrow b = 1.833 \quad \text{and} \quad P\{t(9) > b\} = 0.90 \Longrightarrow b = 1.383.$$

Therefore, the 80% confidence interval is $3.73 \pm (1.833 \times 0.34/3) = [3.522, 3.938]$ and the 90% confidence interval is $3.73 \pm (1.383 \times 0.34/3) = [3.573, 3.887]$

2. Mr. Smith has been threatened with the loss of his job if he persists in arriving late at the office. Prior to this threat, his average arrival time over 10 days was 10–46am. with a standard deviation of 16 minutes. For five working days since the threat, his arrival time has been 10–01am. with a standard deviation of 12 mins. Construct a 90% confidence interval for the extent to which Mr. Smith has improved his arrival time.

**Answer.** We assume that the arrival times before the threat are independent random variables $x_1, \ldots, x_n$ with $x_i \sim N(\mu_x, \sigma_x^2)$ and that the arrival times after the threat are likewise a random sample $y_1, \ldots, y_m$ with $y_j \sim N(\mu_y, \sigma_y^2)$. We assume that that $\sigma_x^2 = \sigma_y^2$ and use the result that

$$\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2).$$

where $\hat{\sigma}^2$ is the pooled estimator of the variance:

$$\hat{\sigma}^2 = \frac{\sum_i^n (x_i - \bar{x})^2 + \sum_j^m (y_j - \bar{y})}{n + m - 2} = \frac{\hat{\sigma}_x^2(n-1) + \sigma_y^2(m-1)}{n + m - 2}$$

$$= \frac{s_x^2 n + s_y^2 m}{n + m - 2} = \frac{(16^2 \times 10) + (12^2 \times 5)}{10 + 5 - 2} = \frac{2560 + 720}{13},$$

which gives $\hat{\sigma}^2 \simeq 252.3$ and $\hat{\sigma} \simeq 15.9$ The following probability statement can be made:

$$P\left\{(\bar{x} - \bar{y}) - \beta\hat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_x - \mu_y < (\bar{x} - \bar{y}) + \beta\hat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{m}}\right\} = Q$$

where $Q \in (0, 1)$ is a chosen probability value which determines $\beta$. If $Q = 0.9$, then the corresponding value from the $t(13)$ tables is $\beta = 1.771$. The 90% confidence interval for $\mu_x - \mu_y$ is given by

$$45 \pm 1.771 \times 15.9 \times \left(\sqrt{\frac{3}{10}} = 0.5477\right) = 45 \pm 15.42.$$

3. A factory that manufactures shafts of 5cm diameter has installed new lathes. Hitherto, the variance of the diameter of the shafts has been $0.49\text{mm}^2$. A sample of 20 shafts, produced by the new machines, has a variance of $0.25\text{mm}^2$ measured about the theoretical mean of 5cm. Find **(a)** a 95% confidence interval for the new variance, and **(b)** a 95% confidence interval for the ratio of the old and the new variances.

**Answer: (a)** The following statements are equivalent:

$$\left(a < \sum_i \frac{(x_i - \mu)^2}{\sigma^2} < b\right) \iff \left(\frac{1}{a} > \frac{\sigma^2}{\sum(x_i - \mu)^2} > \frac{1}{b}\right)$$
$$\iff \left(\frac{\sum(x_i - \mu)^2}{b} < \sigma^2 < \frac{\sum(x_i - \mu)^2}{a}\right),$$

whence the proabability statement on which the confidence interval is based is of the form

$$P\left(\frac{\sum(x_i - \mu)^2}{b} < \sigma^2 < \frac{\sum(x_i - \mu)^2}{a}\right) = Q.$$

We assume that the normal distribution provides an adequate approximation. Given that $Q = 0.95$ and $n = 20$, we find from the table of the $\chi^2(20)$ distribution that

$$P\{\chi^2(20) < a\} = 0.025 \Longrightarrow a = 9.59 \quad \text{and} \quad P\{\chi^2(20) < b\} = 0.975 \Longrightarrow b = 34.2.$$

Given that $\sum(x_i - \mu)^2/n = 0.25\text{mm}^2$, it follows that

$$\left(\frac{\sum(x_i - \mu)^2}{b} < \sigma^2 < \frac{\sum(x_i - \mu)^2}{a}\right) \Longrightarrow \left(\frac{0.25 \times 20}{34.2} < \sigma^2 < \frac{0.25 \times 20}{9.59}\right)$$
$$\Longrightarrow (0.145 < \sigma^2 < 0.521).$$

**Answer: (b)** To find the interval for the ratio of the new and the old variance, we divide this inequality throughout by the old variance, which is assumd to be known with certainty.

4. Two independent random samples of sizes $n = 16$ and $m = 10$, taken from independent normal distributions $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$ yield, respectively, $\bar{x} = 3.6$, $s_x^2 = 4.14$ and $\bar{y} = 13.6$, $s_y^2 = 7.6$. Find the 90% confidence interval for $\sigma_x^2/\sigma_y^2$ when $\mu_x$ and $\mu_y$ are unknown.

**Answer:** Let the two samples be denoted $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ with $E(x_i) = \mu_x$, $V(x_i) = \sigma_x^2$ for all $i$, and $E(y_j) = \mu_y$, $V(y_j) = \sigma_y^2$ for all $j$. The estimated variances are

$$\hat{\sigma}_x^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} \quad \text{and} \quad \hat{\sigma}_y^2 = \sum_{j=1}^{m} \frac{(y_j - \bar{y})^2}{m-1}.$$

Also, there are

$$\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma_x^2} \sim \chi^2(n-1), \quad \sum_{j=1}^{m} \frac{(y_j - \bar{y})^2}{\sigma_y^2} \sim \chi^2(m-1), \quad \text{and}$$

$$\left\{ \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{(m-1)\sigma_y^2} \Big/ \sum_{j=1}^{n} \frac{(x_j - \bar{x})^2}{(n-1)\sigma_x^2} \right\} = \left\{ \frac{\hat{\sigma}_y^2 \sigma_x^2}{\hat{\sigma}_x^2 \sigma_y^2} \right\} \sim F(m-1, n-1).$$

From tables, we can find numbers $\alpha$, $\beta$ such that

$$P(\alpha < F(m-1, n-1) < \beta) = Q, \quad \text{or, equivalently,}$$

$$P\left( \alpha \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} < \frac{\sigma_x^2}{\sigma_y^2} < \beta \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} \right) = Q,$$

where $Q$ is a preassigned probability value.

In this case, we have

$$n = 16, \quad m = 10, \quad \hat{\sigma}_x^2 = 4.14 \times \left\{ \frac{16}{15} \right\}, \quad \hat{\sigma}_y^2 = 7.6 \times \left\{ \frac{10}{9} \right\}.$$

Also

$$\frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} = \frac{\sum (x_i - \bar{x})(m-1)}{\sum (y_j - \bar{y})(n-1)} = \frac{s_x^2 n(m-1)}{s_y^2 m(n-1)} = \frac{4.14 \times 26 \times 9}{7.6 \times 10 \times 15} = 0.5230$$

The $F(9, 15)$ tables show that $P(F \le \beta) = 0.95$ implies $\beta = 2.59$. Also we must find a value $\alpha$ such that $P(F \ge \alpha) = 0.95$. But $P(F \ge \alpha) = P(F^{-1} \le \alpha^{-1})$; and the $F(15, 9)$ tables show $P(F^{-1} \le \alpha^{-1}) = 0.95$ is satisfied by $\alpha^{-1} = 3.01$. Therefore $\alpha = 1/3.01 = 0.332$ and the confidence interval is given by

$$0.332 \times 0.523 = 0.1736 < \frac{\sigma_y^2}{\sigma_x^2} < 1.355 = 2.59 \times 0.523.$$