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.

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 can use the result that

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

Assuming that $\sigma_x^2 = \sigma_y^2$, we use a pooled estimate of the variance:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_i^n (x_i - \bar{x})^2 + \sum_j^m (y_j - \bar{y})}{n + m + 2} = \frac{\hat{\sigma}_x^2(n-1) + \sigma_y^2(m-1)}{n + m + 2} \\ &= \frac{s_x^2 n + s_y^2 m}{n + m - 2} = \frac{(16^2 \times 10) + (12^2 \times 5)}{10 + 5 - 2} = \frac{2560 + 720}{13},\end{aligned}$$

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 β. 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 0.49mm^2. A sample of 20 shafts, produced by the new machines, has a variance of 0.25mm^2 measured about the theoretical mean of 5cm. Find a 95% confidence interval for the new variance, and a 95% confidence interval for the ratio of the old and the new variances.

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 estimates of the variances are

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

Also, we have

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

$$\left\{ \sum_{i=1}^{n} \frac{(x_i - \mu_x)^2}{n\sigma_x^2} \Big/ \sum_{j=1}^{m} \frac{(y_j - \mu_y)^2}{m\sigma_y^2} \right\} \sim F(n, m).$$

From tables, we can find numbers α, β such that

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

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

where Q is a preassigned probability value.

In this case, we have

$$n = 20, \quad m = 10, \quad \hat{\sigma}_x^2 = 0.64, \quad \hat{\sigma}_y^2 = 0.25, \quad Q = 0.90.$$

The $F(20, 10)$ tables show that $P(F \leq \beta) = 0.95$ implies $\beta = 2.77$. Also we must find a value α such that $P(F \geq \alpha) = 0.95$. But $P(F \geq \alpha) = P(F^{-1} \leq \alpha^{-1})$; and the $F(10, 20)$ tables show $P(F^{-1} \leq \alpha^{-1}) = 0.95$ is satisfied by $\alpha^{-1} = 2.35$. Therefore $\alpha = 1/2.35$ and the confidence interval is given by

$$\frac{1}{2.35} \times \frac{0.25}{0.64} < \frac{\sigma_y^2}{\sigma_x^2} < 2.77 \times \frac{0.25}{0.64}.$$

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 σ_x^2/σ_y^2 when μ_x and μ_y are unknown.

   Find the 90% confidence interval for σ_x^2/σ_y^2 on the assumption that $\mu_x = 4$ and $\sum(x_i - \mu_x)^2/n = 5.3$ and that $\mu_y = 12$ and $\sum(y_i - \mu_y)^2/m = 7.5$.

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 estimates of the variances are

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

Also, we have

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

$$\left\{ \sum_{i=1}^{n} \frac{(x_i - \mu_x)^2}{n\sigma_x^2} \Big/ \sum_{j=1}^{m} \frac{(y_j - \mu_y)^2}{m\sigma_y^2} \right\} \sim F(n, m).$$

From tables, we can find numbers α, β such that

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

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

where Q is a preassigned probability value.

In this case, we have

$$n = 20, \quad m = 10, \quad \hat{\sigma}_x^2 = 0.64, \quad \hat{\sigma}_y^2 = 0.25, \quad Q = 0.90.$$

The $F(20, 10)$ tables show that $P(F \leq \beta) = 0.95$ implies $\beta = 2.77$. Also we must find a value α such that $P(F \geq \alpha) = 0.95$. But $P(F \geq \alpha) = P(F^{-1} \leq \alpha^{-1})$; and the $F(10, 20)$ tables show $P(F^{-1} \leq \alpha^{-1}) = 0.95$ is satisfied by $\alpha^{-1} = 2.35$. Therefore $\alpha = 1/2.35$ and the confidence interval is given by

$$\frac{1}{2.35} \times \frac{0.25}{0.64} < \frac{\sigma_y^2}{\sigma_x^2} < 2.77 \times \frac{0.25}{0.64}.$$