Biostats VHM 801 Course
Fall 2006, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to Final Exam, December 2006

Question 3

Subquestion a)

Although it is not described how the data were collected, the data suggest that the ages of the sparrows were somehow selected to represent the range between 3 and 17 days. The shown age distribution would be very implausible if the sparrows were randomly selected from a population. Therefore, age is a (partly) controlled variable and should be considered as explanatory, whereas the wing length is a response variable. The random variation is in the wing lengths, not in the ages. As an an additional justification for *not* taking age as the response, it can be said that it biologically is close to nonsense to state that the age would depend on the wing length; the direction of the causation is the other way round. Note that after it has been established that age is not the response, we cannot take age as the dependent (Y) variable in a linear regression. The natural model is therefore a linear regression with age (x) and wing length (Y) related by the equation:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, 13,$$

where the errors ε_i are assumed independent and $\sim \mathrm{N}(0, \sigma)$. The Minitab listing gives parameter estimates and standard errors for the regression parameters. Using $t^* = t_{.975}(11) = 2.201$ we have the following estimates and 95% confidence intervals:

$$\begin{aligned}
\hat{\beta}_0 &= 0.7283, \quad 95\%\ \text{CI: } 0.7283 \pm 2.201 \cdot 0.1444 = (0.41, 1.05), \\
\hat{\beta}_0 &= 0.2679, \quad 95\%\ \text{CI: } 0.2679 \pm 2.201 \cdot 0.0132 = (0.239, 0.297), \\
\hat{\sigma} &= 0.2132.
\end{aligned}$$

The listing shows an R^2-value of 97.4% which indicates a strong relation between the two variables (alternatively, the correlation is 0.987). A prediction of the wing length from a bird's age will be quite precise.

Subquestion b)

According to the model, a bird of 14 days of age would have a predicted wing length of 4.48 cm, with a 95% prediction interval ranging from 3.98 cm to 4.98 cm. The reported value of Jones of 5.0 cm for the wing length therefore falls just outside the prediction interval. Strictly speaking this means that there is evidence at the 5% level that Jones' bird does not agree with the 13 birds on which the equation is based. As the observed value is just outside the interval and the representativity of a line fitted based on 13 birds can be questioned, it seems preferable to be cautious and state that Jones' bird seems unusually large for a bird of that age but there is no strong evidence against it from the equation fitted by the 13 birds. It could also be useful to examine the fitted line plot to see if there is any indication of a lack of fit for older birds. A linear growth equation is not going to work "forever" (at some point, growth will fade off). If the curve actually seemed to have a downward curvature for the oldest birds, the evidence against Jones' bird might be stronger than indicated above.

Subquestion c)

The fitted regression equation:

$$\text{wing length} = 0.7283 + 0.2679 \cdot \text{age},$$

can be used to estimate age (x) from a measured wing length ($Y = 4.5$), by inserting the value of Y and solving for x:

$$4.5 = 0.7283 + 0.2679x \;\Rightarrow\; x = (4.5 - 0.7283)/0.2679 = 14.08.$$

That is, a sparrow with a wing length of 4.5 cm is estimated to be $14.08 \approx 14$ days old. To obtain a confidence interval for this estimate one would try to work out a standard error (SE) of the estimate $\hat{x} = (Y - \hat{\beta}_0)/\hat{\beta}_1$. With an estimated SE, one could then use the general formula for confidence intervals: estimate $\pm\ t^* \times$ SE, with the $t^* = 2.201$ already used. However, the SE is not easy to compute, and it may be argued that a symmetric confidence interval is not the best solution. An alternative approach, with somewhat cumbersome formulae, can be found under the heading "Inverse Prediction" in more comprehensive statistics textbooks, e.g. *Biostatistical Analysis* by J.E. Zar.