University of York

Department of Health Sciences

Applied Biostatistics

Suggested answers to Exercise: Variability

Question 1

(a) *The mean dose was 2.89 with standard deviation 2.12 mGy. Why did the authors say 'The standard deviation suggests that very few X-ray units irradiate the patient above 7 mGy (less than 5% based on this work)'?* There would be only about 5% of observations outside mean ± 2 standard deviations, i.e. between $2.89 - 2\times2.12 = -1.35$ and $2.89 + 2\times2.12 = 7.13$. The dose cannot be negative, so we expect about 5% of doses to be above 7.13, so in fact we might expect just over 5% to lie above 7 mGy.

Question 2

(a) *What does `SD' stand for and what does it mean?* `SD' stands for standard deviation. This is a measure of the variability or spread of the data. Approximately 95% of the observations will lie between the mean plus or minus 2 standard deviations.

(b) *In Table 1 there are two obvious misprints. What are they?* The standard deviations for height cannot be correct. These would imply that some heights could be below the mean minus 2 SD = $1.63 - 2 \times 0.6 = 0.43$ m, or above the mean plus 2 SD = $1.63 + 2 \times 0.6 = 2.83$ m, clearly impossible. In fact the standard deviations were 0.06m and 0.08 m.

(c) *In Table 2, there are two obvious misprints in the units in which variables are measured. What are they?* The units for peak speed should be (deg/sec) or (deg sec^{-1}), not (deg/sec^{-1}). I think the units for PVAS should be cm, not mm. It is implausible that almost all of the subjects will put their marks in the first 10mm of the 100mm line.

(d) *In Table 2, which variables must have a skew distribution?* PVAS and HAQ must have skew distributions, because the standard deviation is greater than half the mean, which would imply negative values if the distributions were symmetrical.