1: Data was measured on 50 individuals for arm size ($x$) and brain size ($y$), with the following results:

$$
\bar{x} = 10, \quad \bar{y} = 50, \quad \sum_{i=1}^{50} (x_i - \bar{x})^2 = 100, \quad \sum_{i=1}^{50} (y_i - \bar{y})^2 = 400, \quad \sum_{i=1}^{50} (x_i - \bar{x})(y_i - \bar{y}) = 175
$$

(a) Compute the variances of $x$ and $y$, their covariance, and their correlation.

\[
\text{Var}(x) = \frac{100}{49} = 2.04, \quad \text{Var}(y) = \frac{400}{49} = 8.16, \quad \text{Cov}(x, y) = \frac{175}{49} = 3.57
\]

\[
\text{Corr}(x, y) = \frac{3.57 \sqrt{2.04} \cdot \sqrt{8.16}}{0.88} = 0.88
\]

(b) What is the best linear regression of arm size ($x$) on brain size ($y$)?

\[
b_{x|y} = \frac{3.57}{8.16} = 0.44, \quad a = \bar{x} - b_{x|y} \bar{y} = 10 - 0.44 \cdot 50 = -11.88
\]

Hence, the regression is $(\text{Arm size}) = -11.88 + 0.44(\text{Brain size})$

(c) What is the best linear regression of brain size ($y$) on arm size ($x$)?

\[
b_{y|x} = \frac{3.57}{2.04} = 1.75, \quad a = \bar{y} - b_{y|x} \bar{x} = 50 - 1.75 \cdot 10 = 32.58
\]

Hence, the regression is $(\text{Brain size}) = 32.50 + 1.75(\text{Arm size})$

(d) What fraction of the total variance in brain size does the regression account for?

\[
\text{Fraction of the total variance explained by the regression is the squared correlation, or } 0.88^2 = 0.766
\]

(e) Assuming the appropriate normality assumptions, compute the 95% confidence intervals for $\sigma_x^2$ and $\sigma_y^2$. (Potentially helpful tables are enclosed).

Since $\sum_{i=1}^{n}(x_i - \bar{x})^2 \simeq \sigma_x^2 \chi_{n-1}^2$, it follows that

\[
\sum_{i=1}^{n}(x_i - \bar{x})^2/\sigma_x^2 \sim \chi_{n-1}^2
\]

Define $\chi^2_n(\alpha/2)$ as satisfying

\[
\Pr(\chi^2_n < \chi^2_n(\alpha/2)) = \alpha/2 \quad \text{so that} \quad \Pr(\chi^2_n > \chi^2_n(1-\alpha/2)) = \alpha/2
\]

Thus the upper cutoff for $\sigma_x^2$ in an $(1 - \alpha)$ confidence interval of $\sigma_x^2$ satisfies

\[
\Pr \left( \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma_x^2} \leq \chi^2_n(\alpha/2) \right) \quad \text{or} \quad \sigma_x^2 \leq \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_n(\alpha/2)}
\]

The lower cutoff follows similarly, giving the $(1 - \alpha)$ confidence interval of $\sigma_x^2$ as

\[
\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_{n-1}(1-\alpha/2)} \leq \sigma_x^2 \leq \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_{n-1}(\alpha/2)}
\]

From tables $\chi^2_{19}(0.025) = 31.56$ and $\chi^2_{19}(0.975) = 70.22$, giving the 95% confidence interval for $\sigma_x^2$ as $1.42 \leq \sigma_x^2 \leq 3.17$ (note this confidence interval is not symmetric around the sample variance). Likewise, the interval for $\sigma_y^2$ is $5.68 \leq \sigma_y^2 \leq 12.68$. 

1
2: Use the properties of covariances to show that $E[(x - \mu_x)^2] = E[x^2] - \mu_x^2$.

$E[(x - \mu_x)^2] = E[x^2 - 2\mu_xx + \mu_x^2] = E[x^2] - 2\mu_xE[x] + \mu_x^2 = E[x^2] - 2\mu_x^2 + \mu_x^2 = E[x^2] - \mu_x^2$

3: What is the covariance between a particular data point $(x_i)$ and the sample mean $\bar{x}$?

$$\sigma \left( x_i, \frac{1}{n} \sum_{j=1}^{n} x_j \right) = \frac{1}{n} \sum_{j=1}^{n} \sigma(x_i, x_j) = \frac{\sigma^2(x_i)}{n} + \frac{1}{n} \sum_{j \neq i}^{n} \sigma(x_i, x_j)$$