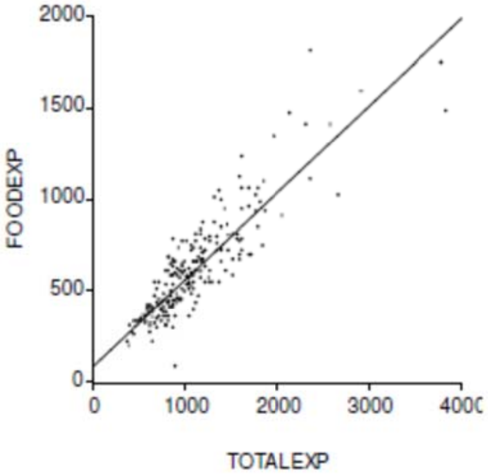
**HW 2 EE325 Answers**

1. Question 2.2, 2.5, 2.15 and 2.16

   2.2 The distinction between the sample regression function and the population regression function is important, for the former is an estimator of the latter; in most situations we have a sample of observations from a given population and we try to learn something about the population from the given sample.

   2.5 A model that is linear in the parameters; it may or may not be linear in the variables.

   2.15
   (a) The scattergram and the regression line look as follows:

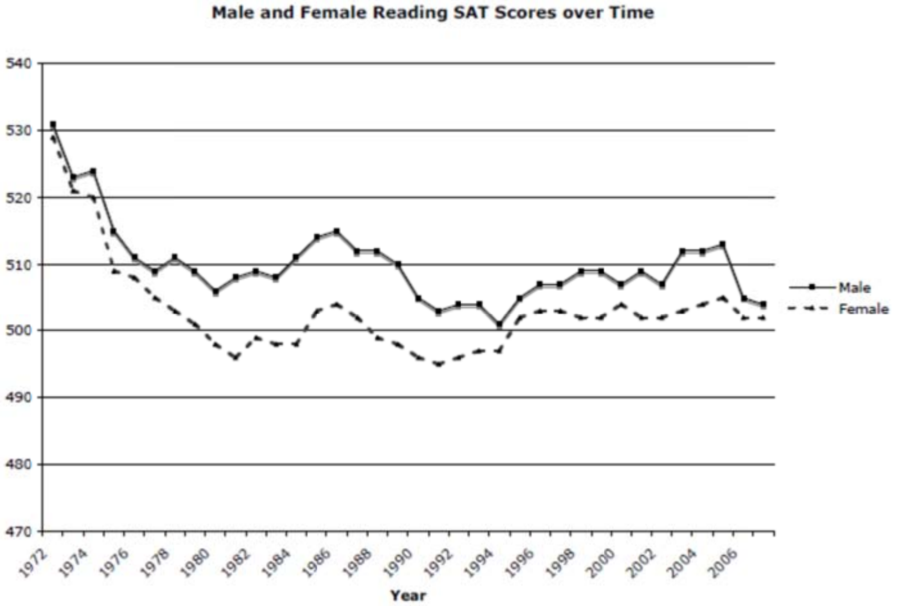



   (b) As total expenditure increases, on the average, expenditure on food also increases. But there is greater variability between the two after the total expenditure exceeds the level of Rs. 2000.

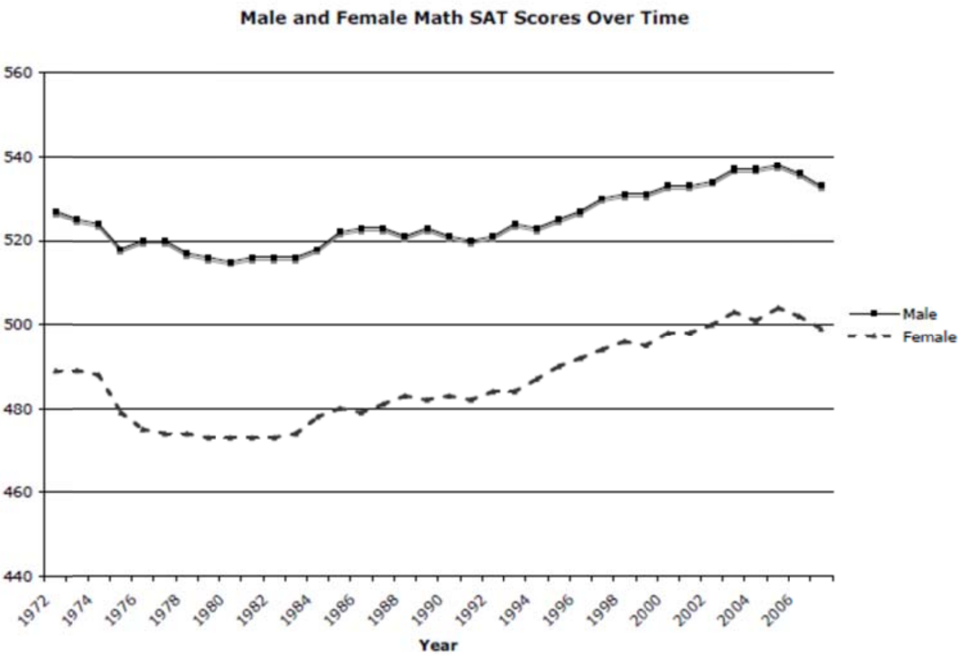
   (c) We would not expect the expenditure on food to increase linearly (i.e., in a straight line fashion) for ever. Once basic needs are satisfied, people will spend relatively less on food as their income increases. That is, at higher levels of income consumers will have more discretionary income. There is some evidence of this from the scattergram shown in (a): At the income level beyond Rs. 2000, expenditure on food shows much more variability.

2.16

(a) The scatter plot for male and female verbal scores is as follows:



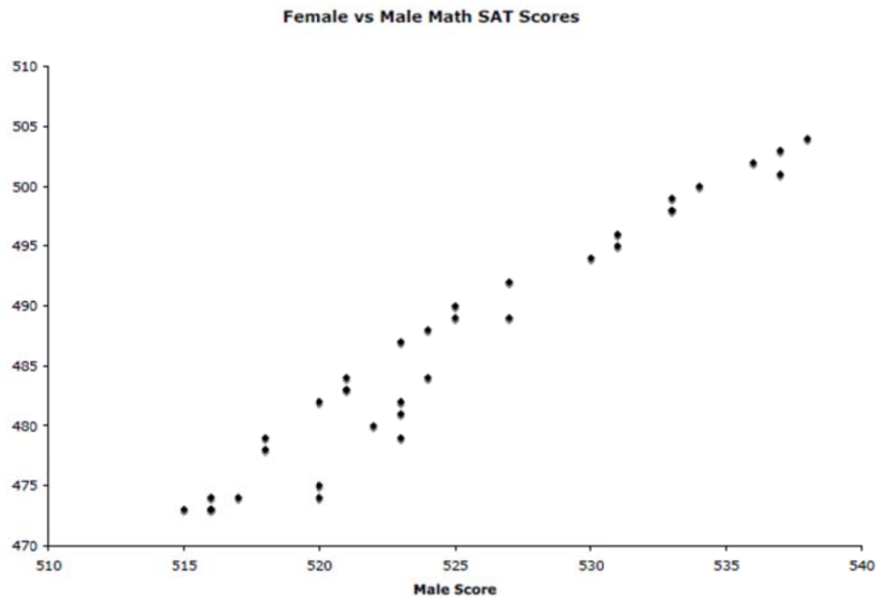

And the corresponding plot for male and female math score is as follows:




(b) Over the years, the male and female reading scores show a slight downward trend, although there seems to be a leveling in the mid-1990's. The math scores, however, show a slight increasing trend, especially starting in the early 1990's. In both graphs it seems the male scores are generally higher than the female scores, of course with year-to-year variation.

(c) We can develop a simple regression model regressing the math score on the verbal score for both sexes.

(d)




As the graph shows, the two genders seem to move together, although the male scores are always higher than the female scores.

2. Show the derivation

If X and Y are two **independent random variables**, then

$$var(X + Y) = var(X) + var(Y)$$

3. If $X_1, X_2, X_3$ is a random sample from a population with mean $u$ and variance $\sigma^2$ .

$X_1, X_2, X_3$ are not independent $Cov(X_1, X_2) = Cov(X_1, X_3) = Cov(X_2, X_3) = \frac{1}{4}\sigma^2$

$\overline{X}$ is estimator used to estimate mean value. $\overline{X} = \frac{1}{3}(X_1 + X_2 + X_3)$

Find $E(\overline{X})$ and $\text{var}(\overline{X})$

$$\overline{X} = \frac{1}{3}(X_1 + X_2 + X_3)$$

$$E(\overline{X}) = E\left(\frac{1}{3}\sum_{i=1}^{3} X_i\right)$$

$$= \frac{1}{3}3E(X_i) = \mu$$

$$
\begin{aligned}
Var(\bar{X}) &= \frac{1}{9}Var(X_1 + X_2 + X_3) \\
&= \frac{1}{9}(Var(X_1) + Var(X_2) + Var(X_3) + 2Cov(X_1, X_2) + 2Cov(X_1, X_3) + 2Cov(X_2, X_3)) \\
&= \frac{1}{9}(\sigma^2 + \sigma^2 + \sigma^2 + 2(\frac{1}{4}\sigma^2) + 2(\frac{1}{4}\sigma^2) + 2(\frac{1}{4}\sigma^2)) \\
&= \frac{1}{9}(3\sigma^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2) \\
&= \frac{1}{9}(3\sigma^2 + \frac{3}{2}\sigma^2) = \frac{\sigma^2}{2}
\end{aligned}
$$

4. Given $X_1, X_2, X_3, X_4, X_5$ are independent identically distributed random variables from population with mean $u$ and variance $\sigma^2$. $\bar{X}$ is estimator used to estimate mean value.
   $\bar{X} = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$

   a. Find $E(\bar{X})$ and $\text{var}(\bar{X})$ in term of $\mu$ and $\sigma$

$$
\begin{aligned}
E(\bar{X}) &= E\left(\frac{1}{5}\sum_{i=1}^{5} X_i\right) \\
&= \frac{1}{3}5E(X_i) = \mu
\end{aligned}
$$

$$
\begin{aligned}
Var(\bar{X}) &= \frac{1}{25}Var(X_1 + X_2 + X_3 + X_4 + X_5) \\
&= \frac{1}{25}(Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4) + Var(X_5)) \\
&= \frac{1}{25}(\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2) \\
&= \frac{1}{5}\sigma^2
\end{aligned}
$$

b. Given $\tilde{X} = \frac{1}{8}X_1 + \frac{1}{4}X_2 + \frac{1}{8}X_3 + \frac{1}{4}X_4 + \frac{1}{4}X_5$ is another estimator of $\mu$. Show that $\tilde{X}$ is an unbiased estimator of $u$

$$\tilde{X} = \frac{1}{8}X_1 + \frac{1}{4}X_2 + \frac{1}{8}X_3 + \frac{1}{4}X_4 + \frac{1}{4}X_5$$
$$= \frac{1}{8}E(X_1) + \frac{1}{4}E(X_2) + \frac{1}{8}E(X_3) + \frac{1}{4}E(X_4) + \frac{1}{4}E(X_5)$$
$$= \frac{1}{8}\mu + \frac{1}{4}\mu + \frac{1}{8}\mu + \frac{1}{4}\mu + \frac{1}{4}\mu = \mu$$

$\tilde{X}$ is unbiased estimator of $\mu$

c. Between $\bar{X}$ and $\tilde{X}$, which one is the better estimator for $\mu$? Why?

$$Var(\tilde{X}) = Var(\frac{1}{8}X_1 + \frac{1}{4}X_2 + \frac{1}{8}X_3 + \frac{1}{4}X_4 + \frac{1}{4}X_5)$$
$$= Var(\frac{X_1 + 2X_2 + X_3 + 2X_4 + 2X_5}{8})$$
$$= \frac{14\sigma^2}{64}$$
$$\therefore Var(\bar{X}) < Var(\tilde{X})$$

$\bar{X}$ is better estimator for $\mu$