**EXERCISE: CALCULATING STATISTICS**

STATISTICS 1403, PROBABILITY AND STATISTICS FOR THE LIFE SCIENCES

**Abstract.** Solve these problems, and mark your answers on the form provided.

**Calibrating a Test**

Two tests for e. coli contamination are compared to see if the readings from the new HGMF test can be calibrated from the current HEC test. The test data (HEC(x), HGMF(y)) has the following sufficient statistics:

\[
\begin{align*}
\sum_{i} x_i &= 17.46 & \sum_{i} y_i &= 21.42 \\
\sum_{i} x_i^2 &= 34.98 & \sum_{i} y_i^2 &= 42.54 \\
\sum_{i} x_i y_i &= 36.58 \\
n &= 18
\end{align*}
\]

1. **Mean and Variance of the Response**

What are the mean and variance of the HGMF(y) responses?

(a) 17.46, 18.04   (b) 0.97, 1.00   (c) 1.19, 1.06
(d) 0.97, 1.06   (e) 1.19, 1.00

2. **How Strong is the Relationship?**

What is the correlation between HEC(x) and HGMF(y) scores?

(a) -0.90   (b) 1.06   (c) 0.876
(d) 0.90   (e) 0.95
3. Fit a Straight Line

What is the equation that predicts HGMF, given the HEC score?
(a) $HGMF = 0.340 + 0.876 \ HEC$
(b) $HGMF = 0.340 + 0.812 \ HEC$
(c) $HGMF = 0.876 + 0.340 \ HEC$
(d) $HGMF = 13.78 + 3.20 \ HEC$
(e) $HGMF = 13.78 + 0.20 \ HEC$

4. Model Error?

OLS works by minimizing the sum of squared errors (SSE). So what is the minimized SSE for this model?
(a) 3.21  (b) 1.01  (c) 13.8
(d) 0.20  (e) 0.812

5. Make a Prediction

For an HEC value of $x_0 = 1.75$, what is the predicted value for the corresponding value of HGMF?
(a) 0.95  (b) 1.87  (c) 1.47
(d) 1.19  (e) 0.97

Useful Formulas

Mean and Variance.

$\bar{y} = \frac{\sum y_i}{n}$ \hspace{1cm} $s_y^2 = \frac{1}{n-1} \left[ \sum y_i^2 - \frac{(\sum y_i)^2}{n} \right]$

Correlation.

$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left[ \sum x_i^2 - n \bar{x}^2 \right] \left[ \sum y_i^2 - n \bar{y}^2 \right]}}$

Estimating Simple Regression Coefficients.

$\hat{\beta}_1 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$ \hspace{1cm} $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ \hspace{1cm} $\hat{y}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$

$SSE = \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i$