STATISTICS FOR MANAGEMENT (IDS 570)

HOMEWORK 2
DUE DATE: THURSDAY, SEPTEMBER 18 AT 12 PM

Problem 1 (Questions 2.32 and 2.33)**.** How strong is the relationship between the score on the first test and the score on the final exam in an elementary statistics course? Here are data for eight students from such a course:

First-test score	153	144	162	149	127	158	158	153
Final-exam score	145	140	145	170	145	175	170	160

(a) Do you think that one of these variables should be an explanatory variable and the other a response variable? Give reasons for your answer.
(b) Find the correlation.
(c) Give some possible reasons why this relationship is so weak.

Here are the data for the second test and the final exam for the same students

Second-test score	158	162	144	162	136	158	175	153
Final-exam score	145	140	145	170	145	175	170	160

(d) Explain why you should use the second-test score as the explanatory variable.
(e) Find the correlation.
(f) Why do you think the relationship between the second-test score and the final-exam score is stronger than the relationship between the first-test score and the final-exam score?

Problem 2. Bergans of Norway has been making outdoor gear since 1908. The following data show the temperature rating (F$^\circ$) and the price (\$) for 11 models of sleeping bags produced by Bergans.

Model	**Temperature Rating (F$^\circ$)**	**Price(\$)**
Ranger 3-Seasons	12	319
Ranger Spring	24	289
Ranger Winter	3	389
Rondane 3-Seasons	13	239
Rondaner Summer	38	149
Rondane Winter	4	289
Senja Ice	5	359
Senja Snow	15	259
Senja Zero	25	229
Super Light	45	129
Tight & Light	25	199

(a) Develop a scatter diagram for these data with temperature rating (F$^\circ$) as the explanatory variable.
(b) What does the scatter diagram developed in part (a) indicate about the relationship between temperature rating (F$^\circ$) and price?
(c) Use the least square method to develop the estimated regression equation.
(d) Predict the price for a sleeping bag with a temperature of 20 F$^\circ$.

Problem 3. Elliptical trainers are becoming one of the more popular exercise machines. Their smooth and steady low-impact motion makes them a preferred choice for individuals with knee and ankle problems. But selecting the right trainer can be difficult process. Price and quality are two important factors in any purchase decision. Are higher prices generally associated with higher quality elliptical trainers? *Consumer Reports* conducted extensive test to develop an overall rating based on ease of use, ergonomics, construction, and exercise range. The following data show the price and rating for eight elliptical trainers tested.

Brand and Model	**Price ($)**	Rating
Precor 5.31	3700	87
Keys Fitness CG2	2500	84
Octane Fitness Q37e	2800	82
LifeFitness X1 Basic	1900	74
NordicTrack AudioStrider	1000	73
Schwinn 430	800	69
Vision Fitness X6100	1700	68
ProFrom XP 520 Razer	600	55

With x = price ($) and y = rating, the estimated regression equation is $\hat{y} = 58.158 + 0.008449x$. For these data, SSE = Sum of Squares due to Error $= \sum_i (y_i - \hat{y}_i)^2$.

(a) Compute the coefficient of determination r^2.
(b) Did the estimated regression equation provide a good fit? Explain.
(c) What is the value of the sample correlation coefficient? Does it reflect a strong or weak relationship between price and rating?

Problem 4. Consider the least squares regression line $\hat{y} = b_1 x + b_0$.

(a) Show that the slope of the regression line is
$$b_1 = r\frac{s_y}{s_x}$$
(b) Show that the intercept of the regression line is $b_0 = \bar{y} - b_1\bar{x}$, and that the regression line passes through the point of averages $(\bar{x}, \bar{y})$.
(c) Show that the equation of the regression line can be written as
$$\frac{\hat{y} - \bar{y}}{s_y} = r\left(\frac{x - \bar{x}}{s_x}\right).$$
In other words, the estimate of y in y-standard units equals r times the given x in x-standard units.
(d) **The "regression effect"**. Consider a student who scored 1.5 SDs above average on the Stat 135 midterm. Will the regression estimate of the students final exam score be 1.5 SDs above average, more than 1.5 SDs above average, or fewer than 1.5 SDs above average? Now consider a student who scored 1.5 SDs below average on the midterm. Will the regression estimate of the students final exam score be 1.5 SDs below average, more than 1.5 SDs below average, or fewer than 1.5 SDs below average? Justify your answers.

Problem 5. The fitted values. The estimated values $\hat{y}$ are often called the fitted values because they are obtained by fitting the regression model to the data. Use the fact that $\hat{y}$ is a linear function of x to show that

(a) $\bar{\hat{y}} = \bar{y}$. That is, the average of the fitted values equals the average of the original values.
(b) $s_{\hat{y}}^2 = r^2 s_y^2$. Check that this gives sensible answers when $r = 0$ and when $r = 1$.

Problem 6. The residuals. For each $i = 1, 2, 3, \ldots, n$, define $e_i = y_i - \hat{y}_i$ to be the *ith* residual, that is, the error in the regression estimate of y_i.

(a) Show that $\bar{e} = 0$.
(b) Show that $s_e^2 = (1 - r^2)s_y^2$. This implies that the larger r gets, the less overall error there is in the regression. Check that the answer is sensible when $r = 0$ and when $r = 1$.
(c) Show that the residuals and x are uncorrelated. This explains why the "residual plot" should show no trend upwards or downwards.