Erick Castillo.
Horsep Akopyan
Ryan Agustin
Austin Venezia.

MATH 440B
SCHILLING

LAB: REGRESSION PRACTICE

1. A sample of 98 students from Math 140 was surveyed with regard to their age and weight. Here are the summary statistics:

Recall: $\hat{y}_i = \bar{y} + r\frac{s_y}{s_x}(x_i - \bar{x})$

		Age (yrs.)	Weight (lbs.)
Females:	Mean:	20.9	126.0
	S.D.:	3.8	18.1
	Corr.:	$r = 0.07$	

		Age (yrs.)	Weight (lbs.)
Males:	Mean:	22.9	163.0
	S.D.:	5.8	21.6
	Corr.:	$r = 0.00$	

(a) For each sex, give the equations for predicting weight from age using simple linear regression:

Females: $\hat{y} = 119.03 + 0.333x_i$ Males: $\hat{y} = 163$ (solved on separate piece of paper)

(b) For each sex, predict the weight of:

	F	M
an 18 year old student:	125.02	163
an 28 year old student:	128.35	163
an 38 year old student:	131.68	163.

(c) Interpret the slope of the least squares line to explain in simple terms how the line predicts that weight will vary with age for each sex. Be specific:

Females: A 1 yr increase in age will increase the weight of a student by 0.333 lbs.

Males: No increase in age (yrs) will impact the weight of this group of students.

(d) What is the predictive meaning of the intercepts? Do they have any relevance here? Explain.

The intercepts can be interpreted as a person w/ 0 yrs. of age will weigh 119.03 lbs (females) or 163 lbs (males). This holds no relevance, and is mostly just the best intercept to fit the LS line given the data.

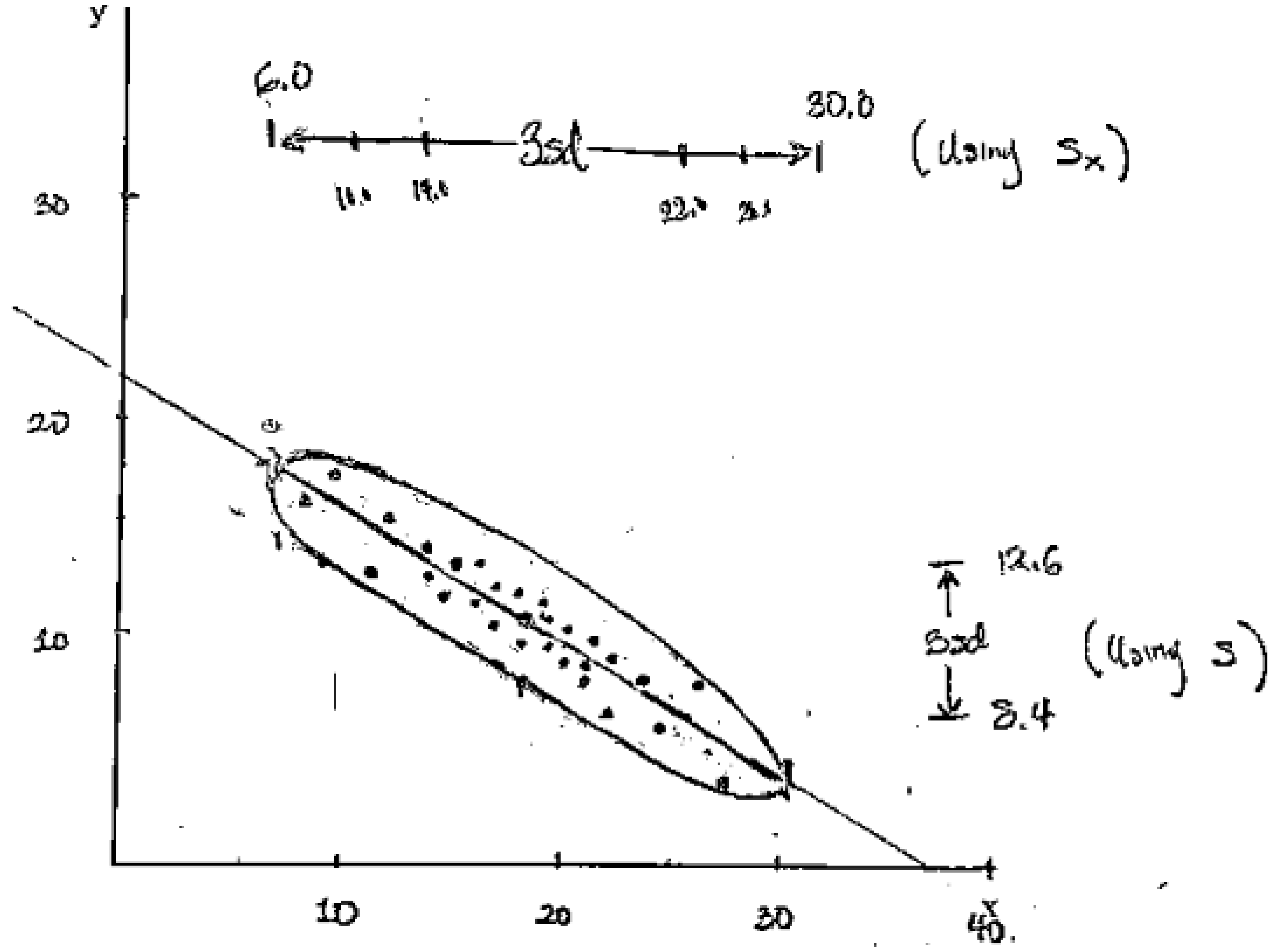
(e) What fraction of the variation in weight is accounted for by the age variable, according to the data?

Females: $(0.07)^2 = 0.0049$

Males: $(0.00)^2 = 0$.

$$r \approx \frac{(D/d)^2 - 1}{(D/d)^2 + 1} \quad ; \quad \hat{y}_i = \bar{y} + r\frac{s_y}{s_x}(x_i - \bar{x})$$

2. Thirty observations (x_i, y_i) were taken on two variables believed to have come from a bivariate normal distribution. The summary statistics are: $\bar{x} = 18.0$, $s_x = 4.0$, $\bar{y} = 10.5$, $s_y = 2.5$, $r = -0.96$. Unfortunately, the original data are lost. Carefully reconstruct the probable appearance of the scatterplot on the axes below, and draw the least squares line through it. Use the formula presented in class to determine the eccentricity of the scatter cloud.



$$\hat{y} = 10.5 + (-0.96)\left(\frac{2.5}{4.0}\right)(x_i - 18.0)$$
$$\hat{y} = 10.5 - 0.6(x_i - 18.0)$$
$$\hat{y} = 10.5 - 0.6x_i + 10.8$$
$$\hat{y} = 21.3 - 0.6x_i$$

$$s^2 = (1 - r^2)s_y^2$$
$$s^2 = (1 - (-0.96)^2)(2.5)^2$$
$$s^2 = 0.49$$
$$s = 0.7$$

99% of the data is within 3 sd's.

$$r \approx \frac{(D/d)^2 - 1}{(D/d)^2 + 1}$$
$$-0.96\left(D/d\right)^2 - 0.96 \approx \left(D/d\right)^2 - 1$$
$$1.96\left(D/d\right)^2 \approx 0.04$$
$$\frac{D^2}{d^2} \approx \frac{1}{49}$$

Note the plot is 7x wider on the "decline" (d) than on the "incline" (D).