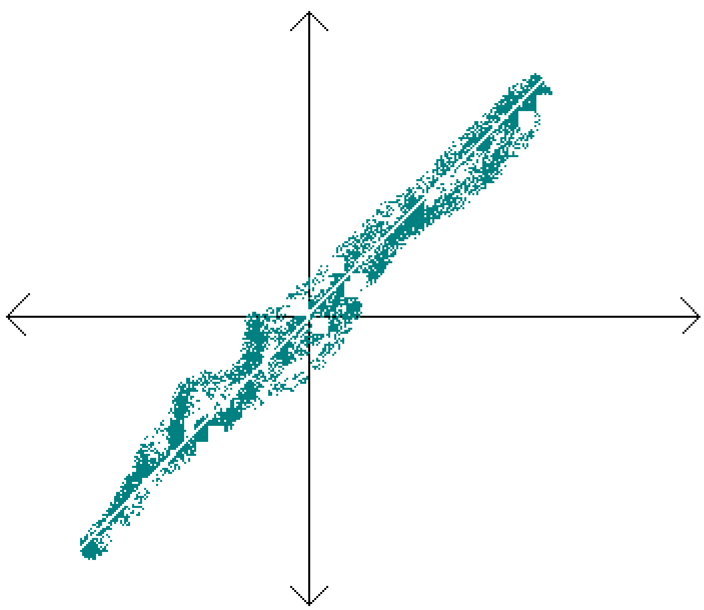
MATH 140
DR. MCLOUGHLIN'S CLASS
REGRESSION FORMULAE
HANDOUT VI DRAFT 1

Let *U* = S be a well defined universe (the sample space) for our work S will always be able to be a subset of $\mathbb{R}$ (the reals, of course).

The simple ordinary least squares (OLS) regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Let D = {(X_1, Y_1), (X_2, Y_2), (X_3, Y_3) . . . , (X_n ,Y_n) } be a finite paired-data set.



What we do is create a line using OLS (beyond the scope of the course – Math 260 Linear Algebra) to minimize the distance between the 'line of best fit' and the actual data to create:

$$\hat{Y}_i = b_0 + b_1 X_i \quad \text{for } i \in \mathbb{N}_n$$

Recall the arithmetic mean of the X - sample is the value $\overline{X}$ where $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j = \frac{X_1 + X_2 + \ldots + X_n}{n}$

Y - sample is the value $\overline{Y}$ where $\overline{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j = \frac{Y_1 + Y_2 + \ldots + Y_n}{n}$

the variance, S_X^2, is defined as $S_X^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}$ and S_Y^2, is defined as $S_Y^2 = \frac{\sum_{k=1}^{n}(Y_k - \overline{Y})^2}{n-1}$

The standard deviations is the square root of the variance.(for each).

Recall from previous handouts:

$\overline{X}$ is $\hat{\mu}_X$ and $\overline{Y}$ is $\hat{\mu}_Y$

$S_X = \sqrt{\dfrac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}}$, S_X is $\hat{\sigma}_X$, etc.

$$r_{XY} = \frac{\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y})}{\sqrt{\sum_{k=1}^{n}(X_k - \overline{X})^2 \sum_{k=1}^{n}(Y_k - \overline{Y})^2}} = \frac{\sum_{k=1}^{n}(X_k - \overline{X})(Y_k - \overline{Y})}{n\sqrt{(S_x)(S_y)}} = \frac{\sum_{k=1}^{n}(z_{X_k})(z_{Y_k})}{n}$$

r_{XY} is $\hat{\rho}_{XY}$

$r_{XY} \in [-1,1]$

We compute

$$b_1 = \frac{\left(\sum_{k=1}^{n}(X_k)(Y_k)\right) - \left(\dfrac{\sum_{k=1}^{n}(X_k)\sum_{k=1}^{n}(Y_k)}{n}\right)}{\left(\sum_{k=1}^{n}(X_k^2)\right) - \dfrac{\left(\sum_{k=1}^{n}(X_k)\right)^2}{n}}$$ indeed $b_1 = \dfrac{S_Y}{S_X} \cdot r_{XY}$

and $b_0 = \overline{Y} - b_1\overline{X}$

Computational examples:

-