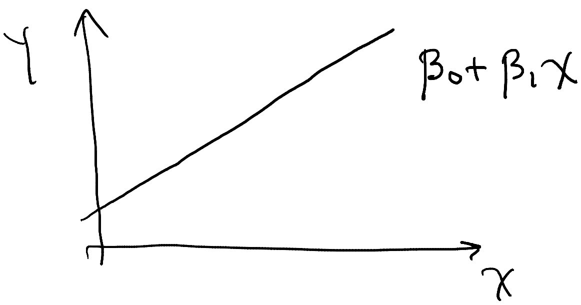
04/19/2023

Chapter 10 - Correlation and Linear Regression (continued)

Linear Regression

Suppose a straight-line adequately summarizes the relationship between X and Y.

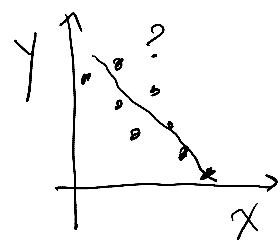


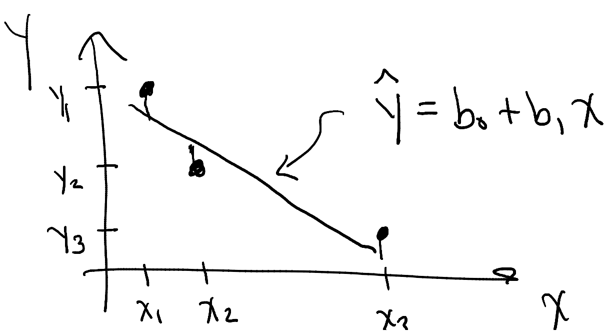
β_1, β_0 are unknown since we just have the sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.

Estimate β_1 and β_0 based on the sample.

Terminology

1. Y is called the dependent or <u>response</u> variable.
2. X is called the independent, <u>explanatory</u>, or predictor variable.
3. A method used to fit a line to bivariate data is called the <u>least squares</u> method.





$\hat{y}$ = predicted or fitted value of y, which depends on x

b_0 = value of $\hat{y}$ when $x=0$ (vertical intercept) = estimate of β_0

b_1 = change in $\hat{y}$ when x changes by 1 unit (slope) = estimate of β_1

One can show that

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Recall $\hat{y} = b_0 + b_1 x$,

which is the <u>least-square line</u> or the <u>fitted regression line</u>. In most cases, use software to find b_1, b_0.

Ex: (continued) 04/24/2023

x = Family income (\$ 1000)

y = House size (100 ft^2)

$n=7$

$$b_1 = 0.345\ldots = \frac{6310 - \frac{(264)(168)}{7}}{10974 - \frac{(264)^2}{7}}$$

$$b_0 = 9.555\ldots$$
$$= \bar{y} - b_1 \bar{x}$$

$$= \frac{158}{7} - (0.345\cdots)\frac{264}{7}$$

In summary,

$$\hat{y} = 9.555 + 0.345X$$

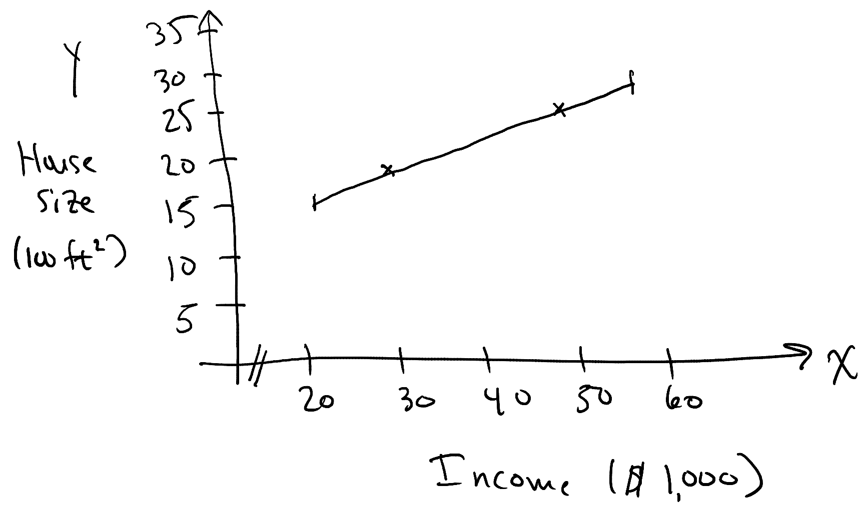
Q: What is the predicted house size for a family whose annual income is \$30,000?

$X = 30$

$$\hat{y} = 9.555 + 0.345(30)$$
$$= 19.905$$

or about $1990.5\ ft^2$

Q: How does one overlay the fitted regression line on a scatterplot?



$$\hat{y} = 9.555 + 0.345X$$

$X = 50$, $\hat{y} = 26.805$

Remark: Do not use regression line outside of the range of X-values in the sample since X and Y relationship outside this range may not be linear. i.e., do not extrapolate.

<u>Ex</u>: (continued)

What does $b_0 = 9.555\ldots$ mean? How to interpret?

What does $b_1 = 0.345\ldots$ mean? How to interpret?

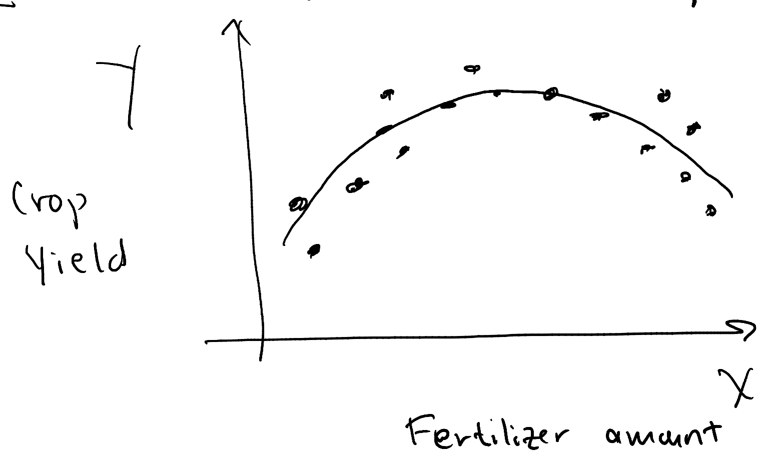
When $x=0$, $\hat{y} = b_0 + b_1 \underset{0}{\underset{\shortparallel}{x}} = b_0 = 9.555$

For this data, the predicted house size is 955.5 ft^2 for an annual income of \$0, but $x=0$ is outside range of x-values in sample.

$b_1 = 0.345$ (100 ft^2)

When x increases by 1 unit (or \$1000), the predicted house size increases by 0.345 (or 34.5 ft^2).

<u>Ex</u>: Relationship between x and y may be more complicated.



$\hat{y} = b_0 + b_1 x + b_2 x^2$