

Ch 12.3 and Ch 12.1 Linear regression

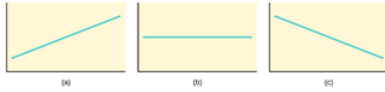
Ch 12.1 Linear Equations

A linear equation has the form of $y = a + bx$ where the graph is in the form of a line.

y is the dependent variable or explanatory variable.

x is the independent variable or predictor or response variable.

The goal is to use x to predict y.



when $b > 0$, the line has a positive slope, graph (a).

when $b < 0$, the line has a negative slope, graph (c).

when $b = 0$, the line is a horizontal line, graph (b).

The value of a is called y-intercept, it is the value of y when $x = 0$.

Ex1. The cost of ordering x items includes a fix shipping cost of \$4.99 and \$3.20 per item. Write the cost y and number of item x. Interpret the slope and y intercept of the equation.

Ans: equation is $y = 4.99 + 3.2x$

slope is 3.2 that is a cost of \$3.2 per item.

y-intercept is 4.99 which is the cost when 0 item are ordered. 4.99 is not meaningful in real life.

Ch 12.3 The regression equation

Match pairs sample can be used to find the equation of the “best fit line” also known as “linear regression line” or “least-squares line”.

The line of best fit is used to predict y given a known value of x. (note: the prediction is a point estimate.)

Terms: Given a matched pair data (x, y)

x – explanatory variable, independent variable

y – response variable, predictor variable, dependent variable.

Line of best fit is $\hat{y} = b_0 + b_1x$

where \hat{y} is the predicted value of y. (y is the observed value in the data.)

where b_0 is the y-intercept (predicted y value when $x = 0$)

b_1 is the slope (rate of change of y per change of x)

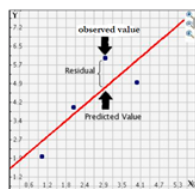
Coefficient of determination (r^2)

r^2 shows the proportion of variation of y that can be predicted by change of x. It tells how good linear prediction is.

How to determine the line of best fit?

The criterion to determin the line that is better than all others is based on the vertical distances between the original data points and the regression line. The distance is also known as residuals.

Residual (ϵ)= observed y – predicted y is $y - \hat{y}$.



The best fit line is the line that satisfy the “least-squares propriety” if the sum of squares of the residuals (SSE) is the smallest sum possible. (Calculus are used to build this.)

This also results in (\bar{x}, \bar{y}) always on the line.

Find equation of line of best fit:

Method 1: (use Statdisk)

- Enter match data to two columns of Statdisk, use Analysis/Correlation and Regression/

Enter significance, select data columns, evaluate

output: b_0 and b_1 for equation is $\hat{y} = b_0 + b_1x$

x is dependent variable, \hat{y} = predicted value of y .

Method 2: use formula $b_1 = r \frac{s_y}{s_x}$, $b_0 = \bar{y} - b_1\bar{x}$

Note: slope has the same sign as r .

use x to predict or estimate y .

the line of best fit is different if x and y switch.

Ex1. Given following matched pair data:

children	A	B	C	D	E	F
shoe size	8	10	8.5	9	11	8
scores	65	92	71	82	95	80

a) Find the best fit-line. Interpret slope.

Enter shoe size and scores to Statdisk. Click Analysis/Correlation and Regression/

Enter significance, select data columns, evaluate.

output: $b_0 = 3.861$, $b_1 = 8.474$ (round to 3 dec. places), $r^2 = 0.767$ or 76.7%, $\hat{y} = 3.861 + 8.474x$ is the linear regression line.

The score will increase 8.574 points for every increase in shoe size of the child.

b) Find correlation of determination. Interpret in the context of this problem.

Ans:

$r^2 = 76.7\%$ means 76.7% of variation in math scores can be predicted by the variation in shoe size.

Ch 12.3 and Ch 12.1 Linear regression is shared under a [not declared](#) license and was authored, remixed, and/or curated by LibreTexts.