

Handout 09-Inference for Population

JIN, ICK HOON

2009. 07. 06.

1. What makes the least squares line the best fit?

The least squares line is the best fit line because it minimizes the distance between the line and every point in the y direction.

2. Interpret the intercept, $b_0 = 4$, and the slope, $b_1 = 1$.

The intercept is just the y point for $x = 0$. It is often not a valid data point, but it 'locates' the line on the y -axis. So, $b_0 = 4$ says $y = 4$ if $x = 0$. The slope tells you how much the y value changes (up or down) for each unit increase (up) of x . So, $b_1 = 1$ says y goes up 1 if x goes up 1.

5. Describe the difference between an influential point and a general outlier.

An influential point usually lies outside the x range of the data. It is influential if removing it changes the regression equation. An outlier is a point that lies far off the line.

6. What does adding (or removing) an influential point do to the regression line? And the correlation?

If an influential point is removed, the slope, and therefore the intercept, noticeably change. That is, the relationship between x and y is changed. The correlation may get stronger or weaker, depending on the location of the influential point.

7. What does adding (or removing) an outlier do to the regression line?
The correlation?

Adding or removing an outlier will have little effect on the regression line. Adding an outlier to a dataset will weaken (decrease it if $r > 0$, increase it if $r < 0$) the correlation. Removing it will strengthen the correlation.

8. Why is it that even strong correlation does not imply causation?

It is always possible that a strong correlation is actually caused by some lurking variable. This variable may not be measured or may not even be known. A good researcher eliminates all other possibilities before claiming causation.

9. Describe the following in reference to a Two-Way Table and how you would calculate it.
- (a) **Joint Probability:** A joint probability is the probability of being in both a particular row and a particular column. It is the cell count divided by the grand total of the table.
 - (b) **Marginal Probability:** A marginal probability is part of the distribution of either the row variable ignoring the column variable or the column variable ignoring the row variable. It is the row (or column) total divided by the grand total.
 - (c) **Conditional Probability:** A conditional probability is the probability of being in a particular cell GIVEN that you are in a particular row (or column). It is the cell count divided by the GIVEN row (or column) total.