Introduction to Some Ability and Knowledge Test Analyses

I. Basic Item Analyses for Ability Tests
II. Regression Analysis
III. Logistic Regression Analysis

I. Basic Item Analyses for Ability Tests

Item difficulty concept

Better performers overall should be more likely to get the question correct

So, the term *item difficulty* refers an easier item

$$P = \frac{\textit{number of correct responses}}{\textit{total number of responses}}$$

Example: 12 out of the 15 student in the class get the first question correct on the quiz

$$P = \frac{12}{15} = .80$$

I. Basic Item Analyses for Ability Tests

Item discrimination concept

- Those who score higher on a test are more likely to get an item correct
- Simple form of this is the item-total correlation
  - Respondents with higher overall score are more likely to get a particular item correct

I. Basic Item Analyses for Ability Tests

Item Discrimination Index

$$D = p_{high} - p_{low}$$

p_{high} is the proportion of high scorers on the test getting the item correct, and p_{low} is the proportion of low scorers on the test getting the item correct

High D values indicate bigger difference between higher and lower scorers, so the item does a better job of discriminating who knows the material more (but no conventional cutoff).

I. Basic Item Analyses for Ability Tests

If the item does a better job of distinguishing those who have higher underlying ability in one group than another, the item is not equally valuable in the two groups

That is, higher D values in one group indicate item does a better job of discriminating who knows the material more for one group than another group

I. Regression Analysis

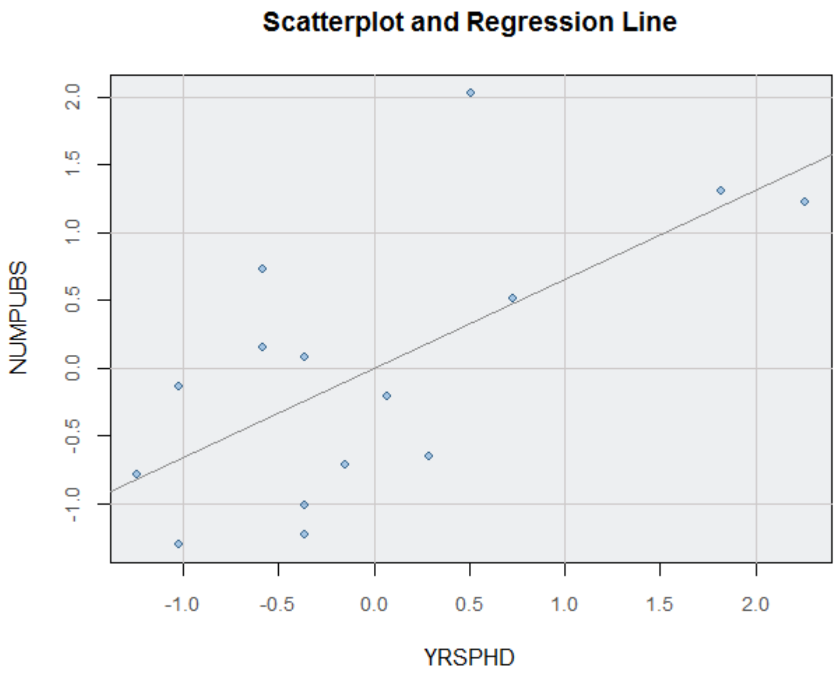
The statistical model for prediction is known as *regression analysis*, whose term comes from regression toward the mean (Galton, 1886)

Regression analysis is based on the equation of a line, which is used to summarize the relationship between two variables

I. Regression Analysis

We will use regression analysis for predicting item responses and for examining test and item bias

I. Regression Analysis



Above is a scatterplot for the number of publications for professors in a psychology department predicted by the number of years since receiving their PhDs

I. Regression Analysis

The equation for a line, $mx + b$, is used to summarize the trend of the points

In the equation for a line, m is the slope, which you should remember as "rise over run"—the amount that Y increases as X is incremented by one point

b is the intercept, or the point on the y-axis where the line intersects when X is equal to 0

I. Regression Analysis

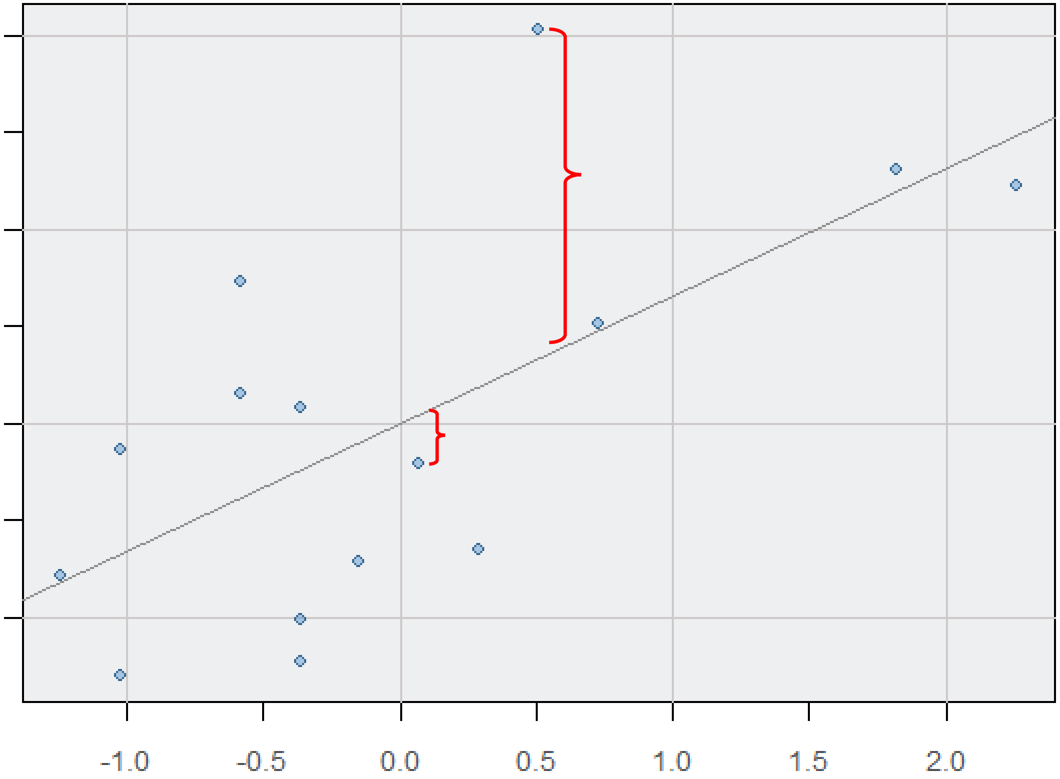
The line is comprised of the values the dependent variable (GPA in college) predicted by X (SAT score)—predicted or expected values called

In your text, the equation for the regression line is given as

$$\hat{Y} = a + b(X)$$

I. Regression Analysis

But as we can see in the scatterplot, the actual points are not always perfectly predicted



I. Regression Analysis

The statistical equation then adds a term that accounts for the error of prediction—the distance of a point from the line, known as the residual

$$Y = a + b(X) + r$$

The actual or observed score, Y, is equal to the equation for the line plus some error, r (sometimes this is *e*, instead)

Simple Logistic Regression

In most ability test analyses, we want to predict whether the item is correct using the persons underlying ability

The better the prediction, the stronger the relationships between ability and the probability of a correct response, $P(Y=1)$

The problem is that predicting the probability of a binary variable is that the line not usually a straight

Simple Logistic Regression

In predicting the probability the line is not straight

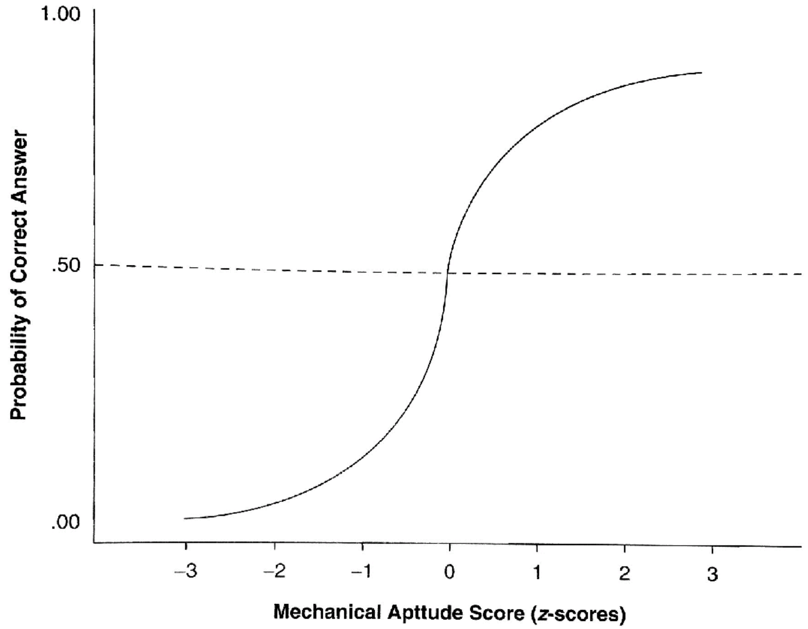


Figure 11.1 Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An Introduction*. Sage.

II. Logistic Regression

The predicted line has to be bent using an exponential transformation, which we call *logistic regression*

$$\ln\left(\frac{p}{1-p}\right) = a + bx$$

The predicted values, $\hat{Y}$, are replaced by the natural logarithm transformation, $\ln$, of the probability that $Y = 1$, represented by p

II. Logistic Regression

To get back to the predicted values, we need to transform the other side of the equation using the opposite of the natural log, the exponential transformation. This produces the probability that Y =1 at some value of x

$$P(Y=1)=\frac{e^{a+bx}}{1+e^{a+bx}}$$

e is Euler's constant, which is approximately 2.71828182845904 (or rounding a bit more, 2.72)

II. Logistic Regression

The exponential transformation, sometimes exp, is simply the constant *e* is raised to a power

For example, $e^3 \approx (2.72)^3 = 20.09$

We can return to the exponent value 3 by taking the natural log of the result, $\ln(20.09) = 3$