

# Five Prototypes of the General Linear Model

Five prototypes are presented with the hope that they will be information collectors for the statistics contained in the remainder of the manual.

The five Prototypes are:

[Counting and Measuring](#)

[Bigger Numbers get even Bigger when Multiplied](#)

[Sums of Products of Similar Numbers Across Sets Get Bigger Results](#)

[Using a proportion to Compare Things](#)

[How Well does the Model Fit the Data](#)

## Prototype 1

### Counting and Measuring

Wundt asked people to make judgments about "psychophysical phenomenon" -- about weights for example, he would say, "Does this weigh more than this?" and point at two weights. He was the first one to try to measure things of the mind. Thurstone measured attitude and achievement. In these examples there is some error in judgment on the part of the participant. Some people are better at making judgments about weights than others. The same is true for the "strength" of an attitude, emotion or achievement. Psychological measurements (in fact all measurements) contain error and consequently our assessments and the mathematical models (statistics) must make provisions for such error. Psychology is not at the level of measurement of other sciences. For example, other sciences have "scopes"; telescopes, the microscopes, stethoscopes, and the sphygmomanometers. The measurement of personality and intellectual attributes has been harder to come by--we have no scopes.

As a result of lack of precision in measurement the statistics that we use must consider this "error of measurement." Later in this chapter you will see that this is variously called "error variance", "residual" and "measurement error. This problem of measuring the mind is seen by some as an impossibility of overcome. Emmanuel Kant said it. Popper restated it with fervor.

The first prototype is that our assessment tools will contain error of measurement and our analytical methods must estimate the degree of error.

[Wundt](#)

[Thurstone](#)

## Prototype 2

### Big Numbers get even Bigger Results when Multiplied

The set of numbers in Box A shows that when you square numbers (multiply each number by itself) that the results get proportionately larger with larger numbers.

Box A	
Number	Number squared
1	1
2	4
3	9
4	16
5	25

Each number of the set (1 through 5) is squared resulting in the set 1, 4, 9, 16, and 25. Notice the difference between the square of 1 and 2 (their squares are 1 and 4) is 3. Whereas the difference between the squares of 4 and 5 (their squares are 16 and 25) is 9. The important characteristic is the difference between the original numbers were the same (1) while the difference between their squares are 3 and 9 respectively. The rate of change is proportionately larger for larger numbers. That is, they get bigger quicker.

This is the second prototype is that the results of squaring large numbers be disproportionately larger than squaring small numbers.

One more example might be helpful to solidify this second prototype. Add 1 to 5 and you get 6; multiply 6 times 6 and the result is 36; the difference between 25 (5 X 5) and 36 (6 X 6) is 11. So once again the "squared numbers get bigger, faster." It will happen all the way to infinity.

## Prototype 3

### Sums of Products of Similar Numbers Across Sets Get Bigger Results

In Box B there are two sets of numbers (each a set of 1 through 5). These can also be thought of as five pairs of numbers. The 1 from the first and the 1 from the second set is the first pair, the 2 from the first set and the 2 from the second set make up the second pair, and so on. The third set is the product (the number in set # 1 times the number in set # 2) of the pairs from the first two sets. Each member of the pair is multiplied to obtain the product. Since each member of the pair is identical calculating the product is the same as squaring a single set of numbers.

Box B			
	Set # 1	Set # 2	Product
	1	1	1
	2	2	4
	3	3	9
	4	4	16
	5	5	25
Sum	15	15	55

In Box C the second set of numbers is reversed so that the number at the opposite ends of the range are multiplied. The first pair is 1 and 5, and the last pair is 5 and 1. The sum of the products results in the smallest possible number. In Box B where the pairs were the most similar the result was the largest possible number and in Box C the pairs were the most different the smallest number resulted. Multiplying pairs and summing the results tells us something about the arrangement. You get the smallest result when you multiply the smallest and the largest with their opposite.

Box C			
	Set # 1	Set # 2	Product
	1	5	5
	2	4	8
	3	3	9
	4	2	8
	5	1	5
Sum	15	15	35

In Box D the second set of numbers has been changed around a little bit so that the summed products of pairs is somewhere in between the largest possibility and the smallest possibility. That is, since the two smallest are together the next two biggest are together. That indicates that small things are going together with small things and large things with large things. So we say that's a relationship between those two sets of numbers then.

When the similar pairs of numbers are drawn from a population the resulting summed product (sum of cross products) will be larger than when the pairs are dissimilar. Small numbers multiplied by small numbers and large numbers multiplied by large numbers and summed will produce the largest result when compared to any other combination. On the other hand if small numbers are multiplied by large numbers and middle range numbers multiplied by middle range numbers and summed the smallest number will result. When the extremes are dissimilar and the mid-range similar the smallest product will result.

Box D			
	Set # 1	Set # 2	Product
	1	5	5
	2	2	4
	3	3	9
	4	4	16
	5	1	5
Sum	15	15	39

### Extensions of prototypes #2 and #3

Intuitive understanding of how prototypes #2 and #3 indicate whether or not two sets of numbers are similar (related). The principle underlying the combination of these two prototypes is that when two sets of numbers are paired together so that the numbers of each set are paired with their most similar size in the other set the resulting sum products will be larger than if they are not paired with their most similar size. The purpose is to get an intuitive grasp of this principle. Only two sets of two numbers each are used.

This first example shows the most similar pairs together (the 2 goes with the other 2 and the 5 goes with the other 5).

$$2 \times 2 = 4$$

$$5 \times 5 = 25$$

This next example shows the dissimilar pairs of the 2 of the first set goes with the 5 of second set and the 5 of the first set goes with the 2 of the second set.

$$2 \times 5 = 10$$

$$5 \times 2 = 10$$

—

20

Notice in the first set that in the 5 X 5 pair that there are 5 rows of 5-one more row of 5s than there are in all pairs of the second set where there are only 4 rows of 5s combined. That is, when bigger numbers are multiplied together they produce even bigger results. When the larger numbers are paired with larger numbers then the resulting products will be larger than if they were not paired together. This occurs even when the two sets of numbers are on different scales (the numbers in each set do not need to be the same). Two more sets of two numbers show the principle.

This example shows the smallest pair of the set together (2 and 6) and the largest pair of the set together (4 and 8).

$$2 \times 6 = 12$$

$$4 \times 8 = 32$$

—

44

The next example shows the smallest number of the first set paired with the largest number of the second set paired together (2 and 8) and the largest number of the first set paired with the smallest number of the second set (4 and 6).

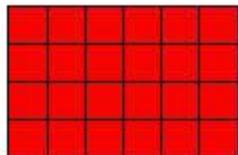
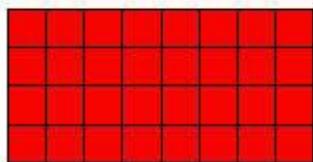
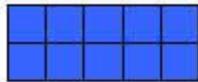
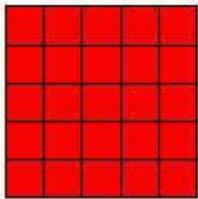
$$2 \times 8 = 16$$

$$4 \times 6 = 24$$

—

40

This example is not as dramatic as the first in showing the bigger result of big numbers but it still exists. Notice in the first example in the larger pair (4 and 8) there are 4 rows of 8s compared to the second example where there are 4 rows of 6s (a loss of 8). At the same time when the 2 is multiplied by the 8 there is only a gain of 4 over the 2 times the 6 in the first set. The principle holds that when pairs numbers of similar size are multiplied of two sets the sum of the products will be larger than any other possible pairing.



Prototype 4

Using a Proportion to Compare Things

One more prototype is needed before a relationship can actually be assessed. We know how big (or how much, or how far) something is by comparing it to something familiar. For example, if we hear that someone weighs 250 pounds we think that's pretty big. We know that because the average weight of a person is about 160 pounds. But how much bigger is 250 than the average person. We divide 160 into 250 and find that it is 1.5625 and think the 250 person is about 1 and half times bigger. We might have done it the other way around and divided 250 in 160 and found that it was .64 and found that the average person is about 6/10ths or 64% the size of the large person (we get the 64% by multiplying 100 times .64).

In prototype # 4 we are going to compare prototype # 2 with prototype # 3 by the use of a proportion or ratio. Are the squares (squaring each number and adding them up) bigger than the products (multiplying the number in one set times the number in the other set) of the two sets. The degree to which the products are as large as the squares is the degree to which the two sets are related (this concept is key to understanding the general linear model). If we compute a ratio between those two results (sum of products and sum of squares), it in fact will indicate the relationship between those two sets of numbers.

Most statistics are concerned with a relationship between two or more sets of numbers. Consequently, the concept of a relationship between two or more sets of numbers is central to the concept of statistics. The prototypes that have been presented are all that is necessary for conceptual understanding but some added calculation are needed for a correlation, t-test or regression are known. Before the relationship between two sets of numbers can be determined both sets need to have a range and "anchor" point. The average or mean of the set is used for that anchor. The steps that were carried out in the previous sets will be performed on set below using the differences from the mean. The first set of numbers will be identified as X and the second set identified as Y.

Set A	Set B	Set C	Set D	Set E	Set F	Set G
	$\bar{X}$ minus the mean		$\bar{Y}$ minus the mean	little $x$ squared	little $y$ squared	little $x$ times little $y$
$\bar{X}$	$X - \bar{X}$ (little $x$ )	$\bar{Y}$	$Y - \bar{Y}$ (little $y$ )	$x^2$	$y^2$	$x \times y$
1	-2	1	-2	4	4	4
2	-1	2	-1	1	1	1
3	0	3	0	0	0	0
4	1	4	1	1	1	1
5	2	5	2	4	4	4
$\bar{X}$	$\bar{0}$	$\bar{Y}$	$\bar{0}$	$\bar{10}$	$\bar{10}$	$\bar{10}$
$\bar{X} = 3$		$\bar{Y} = 3$				

Set A and set C are the same sets we have been working with Set B is  $X$  minus the mean ( $X - 3$ ) or  $x$  (little  $x$ ) and Set D is  $Y$  minus the mean of  $Y$  ( $Y - 3$ ) or  $y$  (little  $y$ ). Set E and Set F are the squares of little  $x$  and little  $y$  respectively. Set G is the product of the little  $y$  times little  $x$ .

It should be noted that "larger numbers multiplied by themselves getting larger faster" applies to "absolute values" (disregarding the signs) in this case. That can be seen where  $-2$  times  $-2$  is equal to  $4$ , whereas  $-1$  times  $-1$  is  $1$ . Remember squaring a set of numbers and adding them together will result in the largest possible result for that set of numbers. That is seen in little  $x$  squared and little  $y$  squared. Consequently, multiplying  $x$  times  $y$  and adding those together will indicate something about the relationship between the two sets. That can be done by comparing the result of (the sum of little  $x$  squared), (the sum of little  $y$  squared), and the (sum of little  $x$  times little  $y$ -- or sum of the cross products).

The formal method of making that comparison is called the Pearson Correlation Coefficient. It is accomplished by the forth prototype -- the ratio. In this case the two squared sets need to be averaged since there are two of them and only one of the cross products. If all problems were as simple as this one we could merely add  $10$  and  $10$  together and divide by  $2$  giving the result of  $10$ . However, these numbers will usually be different and simple arithmetic would not take into account "large numbers produce larger number" we must multiply the sum of  $x^2$  time the sum of  $y^2$  and then take the square root of that. In this case the result is still  $10$ . The final step is to divide this result into the sum of little  $xy$  ( $x$  times  $y$ ) that is divide (producing a ratio)  $10$  by  $10$  the result is  $1.00$  indicating a perfect correlation. The formula that we have just worked out is:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

That's about as complicated as it gets logically. Compute some sort of "sum of squares" and divide that by some other set of "sum of squares". That will indicate some sort of relationship between the two or more sets.

The divisor may not be quit intuitive yet. Why do we multiply them and take the square root? The top number (numerator)  $\sum xy$  is made of only one set of products --  $x$  times  $y$  while the divisor (denominator) is made up of two sets of squares ( $x$  times  $x$ , and  $y$  times  $y$ ). Multiplying them combines them into the proper scale but they then need to be reduced back to a comparable size and the  $\sum xy$  value. Taking the square root provides the proper reduction.

Changing the numbers in the previous problem will show the characteristics in another way. In the numbers below the  $2$  and the  $3$  have been changed in the  $Y$  set of numbers. That should change the relationship between the two sets.

Set A	Set B	Set C	Set D	Set E	Set F	Set G
	X minus the mean		Y minus the mean	little x squared	little y squared	little x times little y
$X$	$X - \bar{X}$  (little x)	$Y$	$Y - \bar{Y}$  (little y)	$x^2$	$y^2$	$x \times y$
1	-2	1	-2	4	4	4
2	-1	3	0	1	0	0
3	0	2	-1	0	1	0
4	1	4	1	1	1	1
5	2	5	2	4	4	4
$\bar{15}$	$\bar{0}$	$\bar{15}$	$\bar{0}$	$\bar{10}$	$\bar{10}$	$\bar{9}$
$\bar{X} = 3$		$\bar{Y} = 3$				

Notice that the only changes made in the sums was the sum of xy. It has changed to 9 rather than 10. That will result in a lower correlation.

$$r = \frac{9}{\sqrt{10 * 10}} = .9$$

Another example is needed to get to a real world example. In this example the scale of the Y variable is changed while the correlation remains the same. A constant of 6 has been added to each of the numbers of the Y variable.

Set A	Set B	Set C	Set D	Set E	Set F	Set G
	X minus the mean		Y minus the mean	little x squared	little y squared	little x times little y
X	$X - \bar{X}$ (little x)	Y	$Y - \bar{Y}$ (little y)	$x^2$	$y^2$	$x \times y$
1	-2	7	-2	4	4	4
2	-1	9	0	1	0	0
3	0	8	-1	0	1	0
4	1	10	1	1	1	1
5	2	11	2	4	4	4
$\bar{15}$	$\bar{0}$	$\bar{45}$	$\bar{0}$	$\bar{10}$	$\bar{10}$	$\bar{9}$
$\bar{X} = 3$		$\bar{Y} = 9$				

Notice that this change does not make a difference in the correlation itself. The only change is in the sum of Y and consequently, the mean of Y.

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = \frac{9}{\sqrt{10 \cdot 10}} = .9$$

Even though the change does not make a difference in the correlation it will make a difference in prototype # 5.

The following set of numbers shows a negative relationship. Consequently, it also demonstrates what happens when dissimilar pairs of numbers are multiplied.

Set A	Set B	Set C	Set D	Set E	Set F	Set G
	$X$ minus the mean		$Y$ minus the mean	little $x$ squared	little $y$ squared	little $x$ times little $y$
$X$	$X - \bar{X}$  (little $x$ )	$Y$	$Y - \bar{Y}$  (little $y$ )	$x^2$	$y^2$	$x \times y$
1	-2	5	2	4	4	-4
2	-1	4	1	1	1	-1
3	0	3	0	0	0	0
4	1	2	-1	1	1	-1
5	2	1	-2	4	4	-4
$\bar{15}$	$\bar{0}$	$\bar{15}$	$\bar{0}$	$\bar{10}$	$\bar{10}$	$\bar{-10}$
$\bar{X} = 3$		$\bar{Y} = 3$				

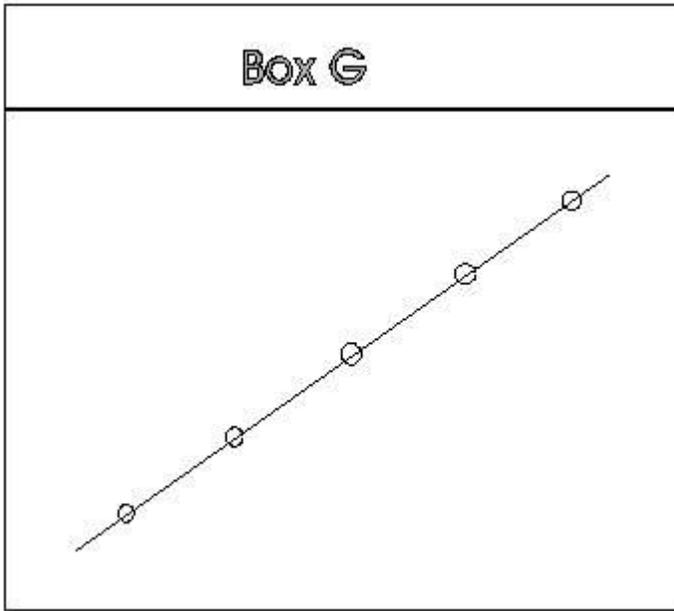
Notice how all of the absolute results all remain the same as the above example of the perfect correlation. However, the signs changes in the sum of  $xy$ . Consequently, you can see that it will now be a perfect negative correlation.

## Prototype 5

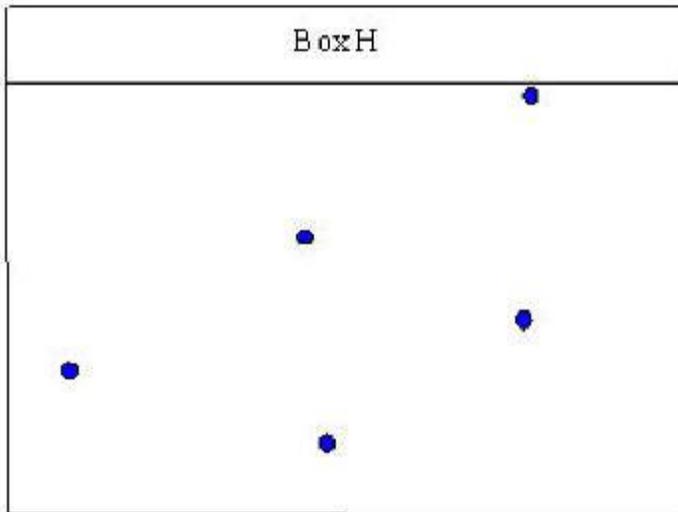
### How well does the Model Fit the data?

The basic idea of this concept is to make a prediction about the data (or anything in fact that can be turned into data). You will see later how model or fit can be applied to this concept. It is the prediction compared to the actual obtained scores. The mean can be used as a prediction. For example, you might be asked to guess how much Fred weighs. If that is all the information you have your best guess would be the average weight of men. On the other hand if you also knew how tall Fred was then your guess could be much improved. Such improvement is the focus of this section. The prototype will be the regression line. It is the basis of the general linear model.

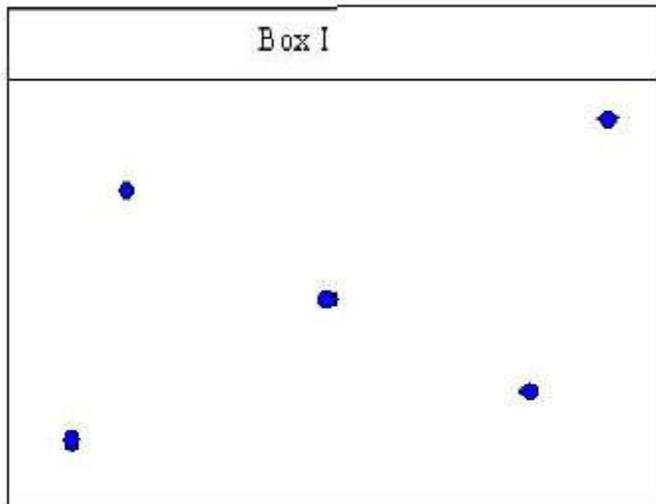
To make this prediction we need a straight line that passes closest to all of the points. In Box G it is easy to find a line that would pass closest to all of the points. In fact the line can pass through all the points.



In Box H it is not as clear where to draw a line that would pass through all of the points.



Box I is similar in that one does not quite know where to draw a line that will be the closest to all of the points in the box.



One way to make the assessment would be to measure the distance from each point and add up those distances and then draw a new line a make the measurements again and repeat the procedure until one found the line that would result in the shortest measures. There is a mathematical way to find the solution called the method of least squares. The points of pairs of numbers can be plotted by having one set of measures plotted vertically (y axis) and one set of numbers plotted horizontally (x axis). Two numbers are needed to identify where the line should be drawn: (1) the slope of the line and (2) where to begin the line.

The slope of the line (for predicting y when x is known) is determined as:

$$slope = b = \frac{\sum xy}{\sum x^2}$$

The slope of the line for predicting x when y is known the formula is:

$$slope = b = \frac{\sum xy}{\sum y^2}$$

The convention in statistics is that x variables are predictors and y variables are the criterion or predicted variables, we will use that convention.

The second characteristic that is needed is where to start or the intercept of y when x is 0. Or what is the value of y when x is 0. It is the mean of y minus the slope times the mean of x. The formula is:

$$\text{constant} = \alpha = \bar{Y} - (b\bar{X})$$

Using the results of these two formulae we can now plot the regression line. In order to keep use connected to the task of learning to use the computer and SPSS the graph is generated from the SPSS package. The following set of data will be used in this example (you have seen it before).

Set A	Set B	Set C	Set D	Set E	Set F	Set G
	X minus the mean		Y minus the mean	little x squared	little y squared	little x times little y
X	$X - \bar{X}$  (little x)	Y	$Y - \bar{Y}$  (little y)	$x^2$	$y^2$	$x \times y$
1	-2	1	-2	4	4	4
2	-1	2	-1	1	1	1
3	0	3	0	0	0	0
4	1	5	2	1	4	2
5	2	4	1	4	1	2
$\bar{15}$	$\bar{0}$	$\bar{15}$	$\bar{0}$	$\bar{10}$	$\bar{10}$	$\bar{9}$
$\bar{X} = 3$		$\bar{Y} = 3$				

The computation for the slope is:

$$\text{slope} = b = \frac{\sum xy}{\sum x^2} = \frac{9}{10} = .9$$

The computation for the intercept (constant) is:

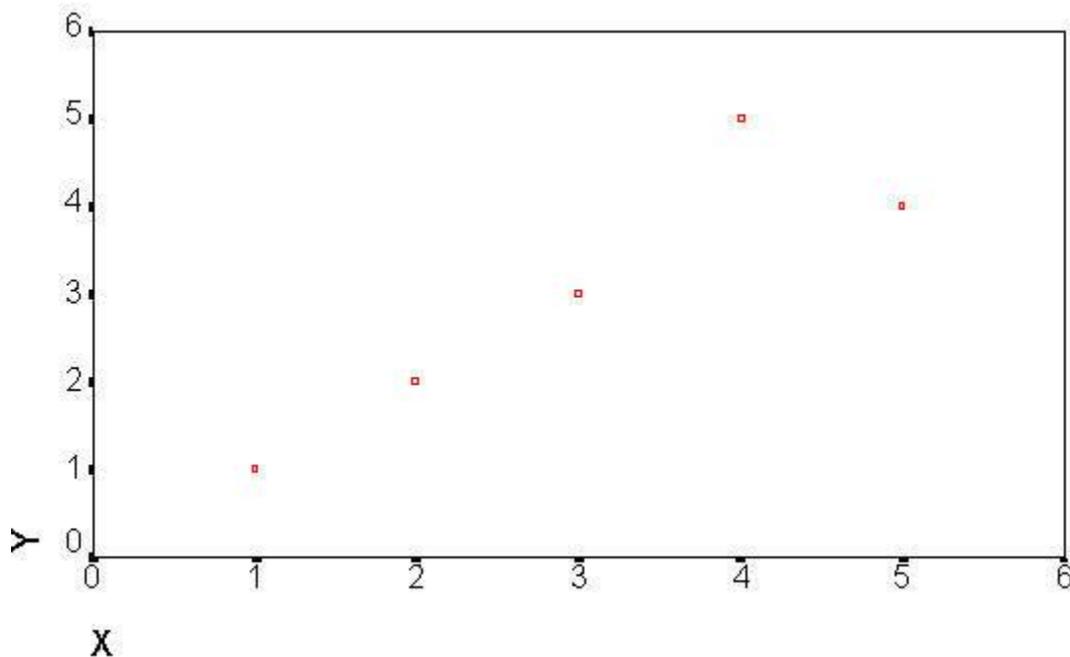
$$\text{constant} = a = \bar{Y} - (b\bar{X}) = 3 - (.9 \times 3) = .3$$

This regression can now be plotted as a regression that is the line that comes closest to the points of the scatterplot. The SPSS program will plot everything but the regression as seen in the following Figure. The following syntax file will produce a plot that will include everything but the regression --that has been drawn in for our purposes.

Plots of the data might be helpful in representing Prototype # 5. You can get those in a crude from the SPSS program (not that SPSS is crude). The following is a syntax file that will generate the plot needed:

```
File Name = crsgra6.sps  
get file='rddalcrssam1.sav'.  
GRAPH  
/SCATTERPLOT(BIVAR)=x WITH y  
/MISSING=LISTWISE.
```

The following is the produced.



The next plot is the same plot that contains further explanation of the data points.

The next plot has been further modified to show the regression line as computed above. The regression line was drawn by starting at .3 on the Y axis when X was equal to 0 and incrementing .9 on the Y axis for each increment of 1 on the X axis. The formula use to generate the regression line was:

$$Y' = Y \text{ primed} = a + (b \text{ times } X).$$

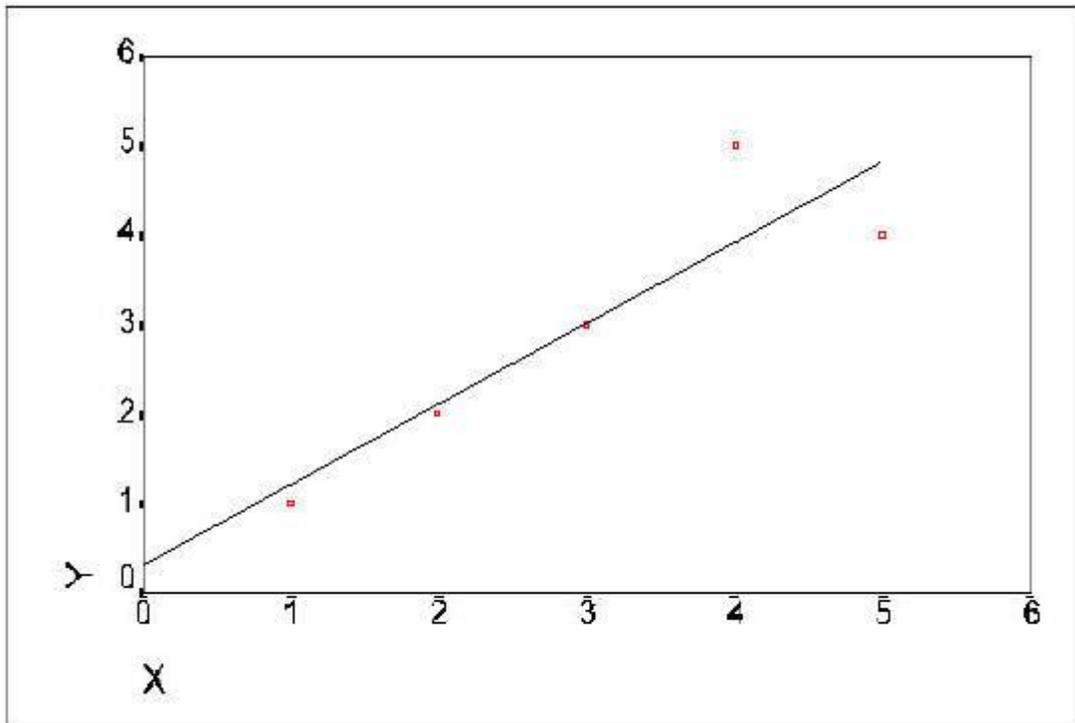
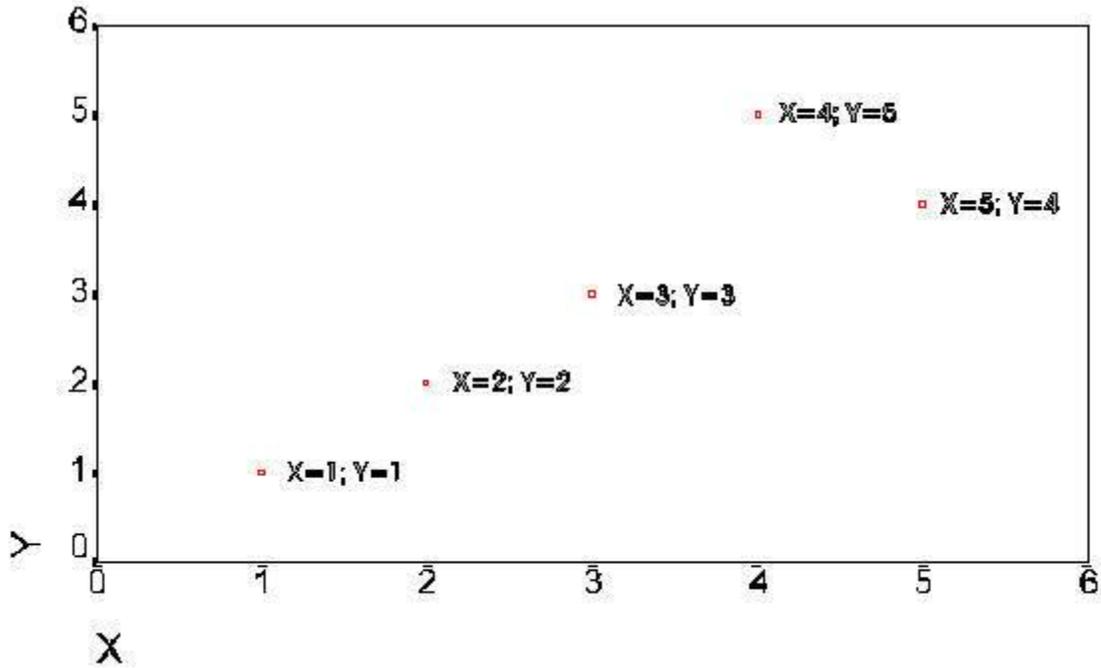


Figure # 3. Scatterplot with regression line.

The model is obtain in the following manner: (1) find a straight which passes closest to all of the points of the variables when they are plotted on the x and y axis. (2) Use this line to predict y scores from the x scores. (3) The difference between the predicted score and the actual score is the error. (4) Square each error score and sum the squares. (5) Compare the sum of squares error

to the total sum of squares. The comparison will result in relationship of the variables or the fit. There are no new computations here -- it has all been done in the above example. Only the concept is added. The correlation itself indicates the fit. This is another way to conceptualize the relationship. It becomes useful in the conceptualization of complex multivariate statistics.

This sum of the differences (lines drawn from the regression line to the observed values) is the error in prediction: the degree to which the model does not fit the data. The error variance is actually the sums of the squares of the length of these lines.

The regression line is the line that will come closest to all of the observed values. If the lines drawn from the regression line to the observed values were added together it would be the smallest of the values for another possible line that could be drawn through the observed values. This graph represents Prototype # 5. The regression line is the prediction (or model) and the lines from the regression line to the actual data points is the error in prediction. This represents the fit of the model to the data.

The regression line can be generated in SPSS in the following manner:

- Click on Graphs
- Click on Scatter
- Click on Simple
- Click on Define
- Select X variable
- Click on the Delta Button to move the variable into the X-axis box
- Select Y variable

- Click on the Delta Button to move the variable into the X-axis box
- Click OK
- Double Click on the chart itself
- Click Chart
- Click Options
- Click Total
- Click OK

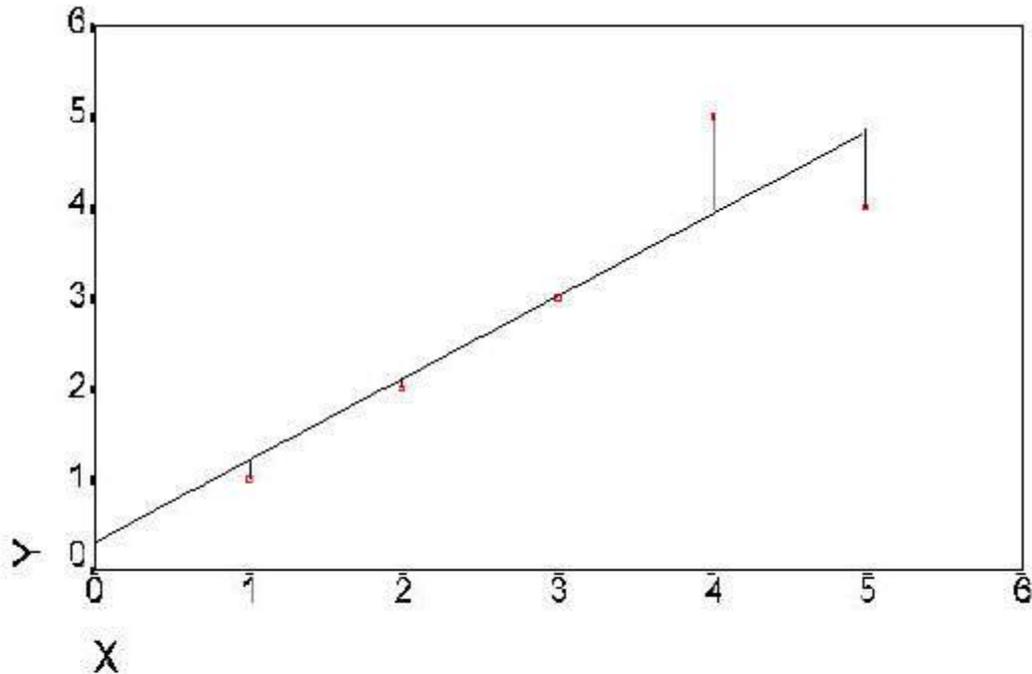


Figure # 4. Actual scores and their differences from the regression line.

The program regression presents the above concepts visually.

## Using the Five Prototypes

That completes the 5 prototypes needed to understand most statistics, now we can add operations to them. Three different "sums of squares" (Prototypes #2 and #3) need to be understood and compared (using Prototype # 4). Particularly, "sums of squares total" (SST), "sums of squares between" (or sometimes called sums of squares regression) - (SSB or SSR), and "sums of squares error" (SSE). SSE was presented in the last Figure. Further it will be useful to then present three sums of squares by three different method (1) numerically, (2) geometrically, (3) as formulae, and finally (4) and Venn diagrams. You should recognize that these are four ways of presenting the same thing.

The three sums of squares (SST, SSB, and SSE) are the basis of the "general linear model." Creative distribution of the "sums of squares regression" among the variable can be used to assess many different hypotheses or models. In each case (numerically, geometrically, formulae, and Venn diagrammatically) the above example will include SST, SSB (SSR), and SSE. At the same time I will "show my work" so that information needed for each calculation needed will also be given.

1	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
2	$X$	$x$	$x^2$	$Y$	$y$	$y^2$	$xy$	$Y'$	$E$	$E^2$		
3			C2*C2			C5*C5	C2*C5		C4-C8			
4		$X - \bar{X}$	$x * x$		$Y - \bar{Y}$	$y * y$	$x * y$	$a + bX$	$Y - Y'$	$(Y - Y')^2$	$(Y' - \bar{Y})$	$(Y' - \bar{Y})^2$
5	actual score	deviation score	deviation squared	actual score	deviation score	deviation squared	cross product	predicted score	error actual minus predicted	error squared		
6	observed score	little x	little x squared	observed score	little y	little y squared		Y primed	residual score			
7	raw score	X minus the mean	little x times little x	raw score	Y minus the mean	little y times little y	little x times little y	Y predicted from X	error in prediction		Y primed minus the mean of Y	
8	1	-2	4	1	-2	4	4	12	-.2	.04	-1.8	3.24
9	2	-1	1	2	-1	1	1	21	-.1	.01	-.9	.81
10	3	0	0	3	0	0	0	30	0	.00	.0	0
11	4	1	1	5	2	4	2	39	1.1	1.21	.9	.81
12	5	2	4	4	1	1	2	48	-.8	.64	1.8	3.24
13	15	0	10	15	0	10	9	15.0	0.0	1.90	0	8.10
14	3			3								
15	mean of X		Sum of little x Squared	mean of Y		sum of little y squared	sum of cross products			error sum of squares		between sum of squares
16						SST				SSE		SSB

Table 2-3. Rows 1 through 7 are either mathematical notation or verbal description of mathematical calculations of the numbers in the column. Rows 8 through 12 are associated numbers involved the calculation. Row 13 is the sum of the numbers in the column while row 14 is the mean for the column. Row 15 is the usual verbal description of the sum in the column and row 16 is an abbreviation of that description.

### B. Geometrically.

The geometric presentation of the model was started with Figure # 1 in the discussion of the prototypes but it was not completed (although the prototypes were completed). The "error sum of squares" was presented in Figure # 4; the "total sum of squares", and "between sum of squares" are presented in the next two figures.

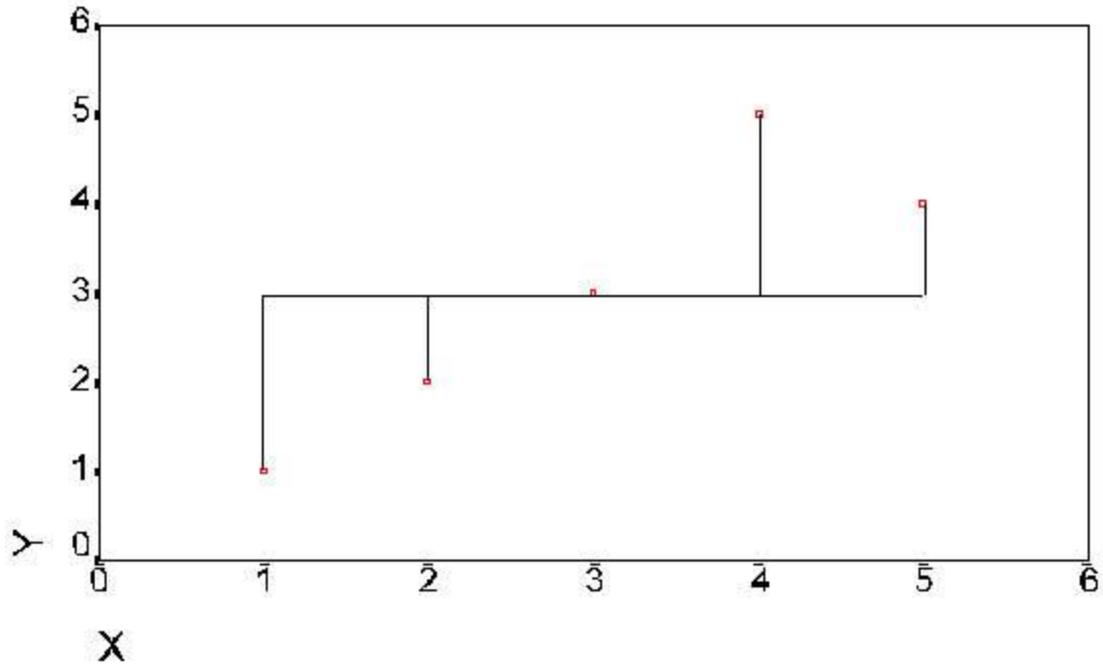


Figure # 5. Distances from the mean -- total sum of squares (same as little y squared).

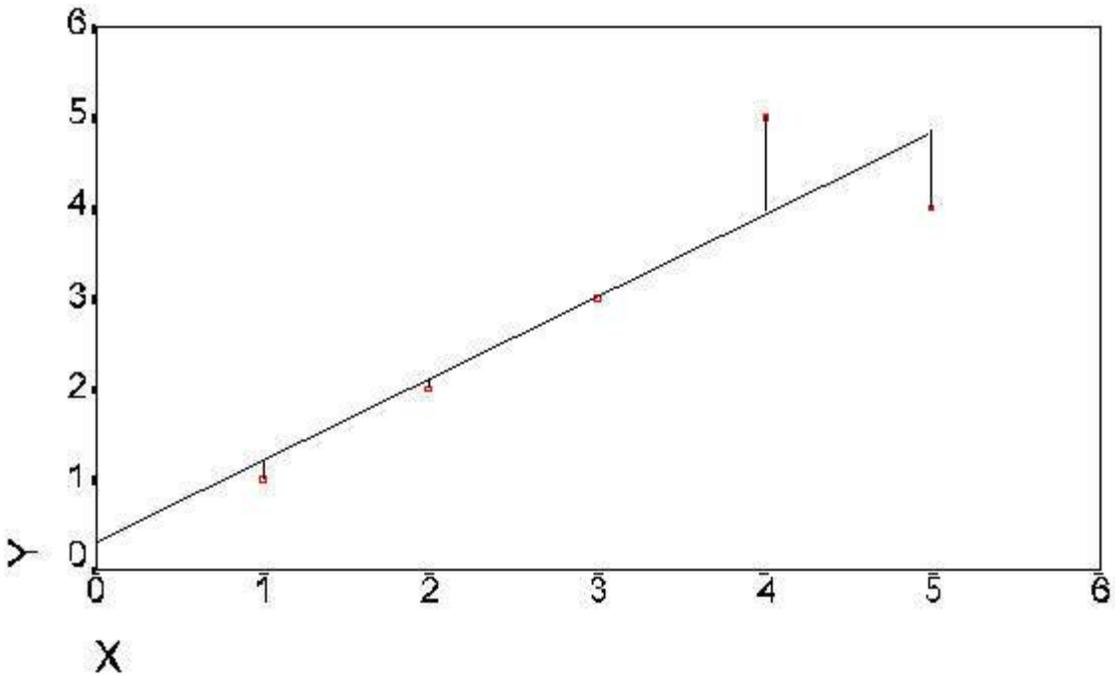


Figure 6. Distances of difference between data points and regression line.

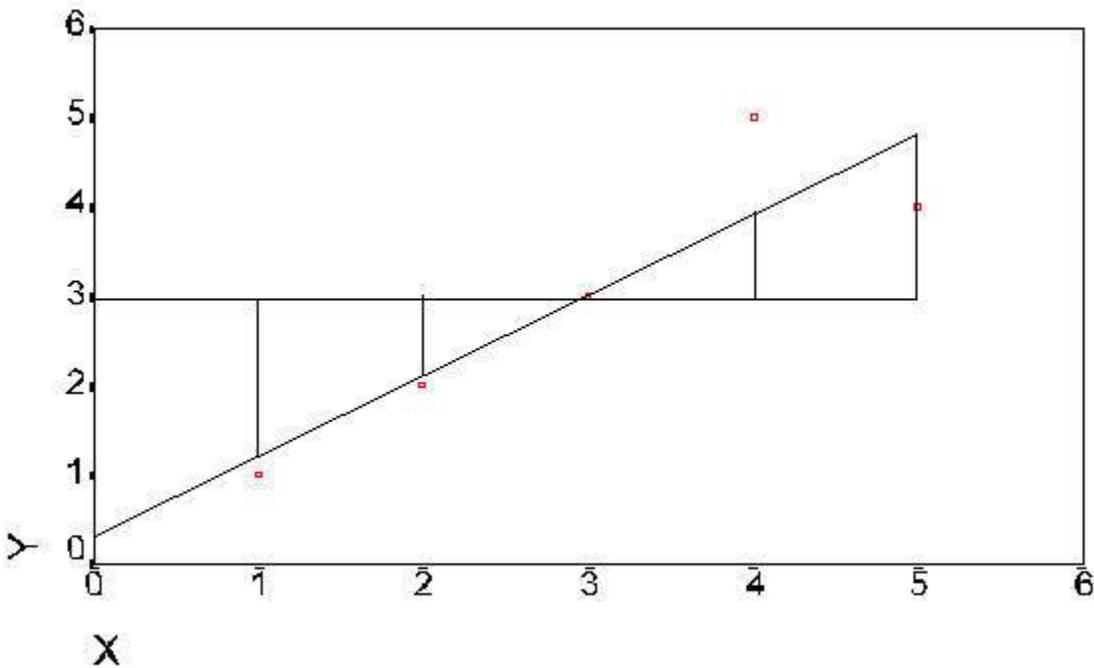


Figure 7. Distances between regression line and mean of Y

This section now gives the formulae and their names for a lot that is statistical. Think of it as learning a new vocabulary (not a set of formulas). Its a way of talking. You may use either the name or the formula. It will get you a long way. Only the standard deviation will be new from you have already covered.

These formulae will cover the essence of all of the statistics covered in this manual -- that is they will work of the intuitive genotype if not the actual statistic. The general linear model can be understood using this set. The statistics it will help you to understand are correlation, anova, (t-test), regression, multiple regression, manova, factor analysis, discriminant function, canonical analysis, and structural equation modeling.

We will next follow through with the above example so that you have a concrete reference to come back to. The are few numbers so that you can work in through easily.

$X$	$x$	$x^2$	$Y$	$y$	$y^2$	$xy$	$Y'$	$E$	$E^2$		
	$X - \bar{X}$	$x * x$		$Y - \bar{Y}$	$y * y$	$x * y$	$a + bX$	$Y - Y'$	$(Y - Y')$	$(Y' - \bar{Y})$	$(Y' - \bar{Y})$
actual score	deviation score	deviation squared	actual score	deviation score	deviation squared	cross product	predicted score	error actual minus predicted	error squared		
observed score	little x	little x squared	observed score	little y	little y squared		Yprimed	residual score			
	X minus the mean	little x times little x		Y minus the mean	little y times little y	little x times little y	Y predicted from X	error in prediction		Y primed minus the mean of Y	
1	-2	4	1	-2	4	4	1.2	-.2	.04	-1.8	3.24
2	-1	1	2	-1	1	1	2.1	-.1	.01	-.9	.81
3	0	0	3	0	0	0	3.0	0	.00	0	0
4	1	1	5	2	4	2	3.9	1.1	1.21	9	81
5	2	4	4	1	1	2	4.8	-.8	.64	1.8	3.24
15	0	10	15	0	10	9	15	0	19	0	8.1
3			3								

All values of the formulas above are represented in this example. There are five observations of X (Raw Score X); therefore N = 5. Incidentally, there are also five observations of Y. The values of X are 1, 2, 3, 4, and 5. The sum of X is 15. Fifteen divided by 5 is 3 (sum of X divided by N) resulting in the mean of X -- ditto for Y.

#	Description	Label	Symbol	Notation	Full Notation or Formula	Individual Case Values
1	Observation of the variable specified by "i" case number and the "j" group or variable	observation		$X$	$X_{ij}$	1, 2, 3, 4, 5
2	Number of observations in each (jth) group (set)	number of cases		$N$	$N_j$	5
3	Sum of X for each (jth) group or set	Sum of X, Total, Sum		$\Sigma X$	$\Sigma X_{ij}$	15
4	Mean for each (jth) group or set	Average, Mean	$\bar{X}$	$\frac{\Sigma X}{N}$	$\frac{\Sigma X_{ij}}{N_j}$	3
5	Deviation Score (X minus the mean)	Little x, deviation score	$x$	$X - \bar{X}$	$x_{ij}$	-2, -1, 0, 1, 2
9	Sum of little x			$\Sigma x$	$\Sigma x_{ij}$	0
10	Sum of little x squared	Sum of Squares	SS	$\Sigma x^2$	$\Sigma x_{ij}^2$	4+1+0+1+4=10
11	Population variance of each (jth) group or set	Population variance		$\frac{\Sigma x^2}{N}$	$\frac{\Sigma x_{ij}^2}{N_j}$	$\frac{10}{5} = 2$
12	Sample variance of each (jth) group or set	Sample variance		$\frac{\Sigma x^2}{N+1}$	$\frac{\Sigma x_{ij}^2}{N_j+1}$	$\frac{10}{5-1} = 2.5$
13	Population standard deviation for each (jth) group or set	Population standard deviation	$\sigma$	$\sqrt{\frac{\Sigma x^2}{N}}$	$\sqrt{\frac{\Sigma x_{ij}^2}{N_j}}$	$\sqrt{\frac{10}{5}} = 1.414$
14	Sample standard deviation for each (jth) group or set	Sample standard deviation	SD	$\sqrt{\frac{\Sigma x^2}{N-1}}$	$\sqrt{\frac{\Sigma x_{ij}^2}{N_j-1}}$	$\sqrt{\frac{10}{5-1}} = 1.58$
15	z-score for each (ijth) observation	z-score	$z$	$\frac{X_{ij} - \bar{X}_j}{SD_j}$		

16	Cross products for two groups (sets) of scores	Cross Products		$\Sigma xy$		
17	Covariance	Covariance		$\frac{\Sigma xy}{N}$		$\frac{9}{5} = 1.8$
18	Slope of the regression line	slope, b, beta	b	$\frac{\Sigma xy}{\Sigma x^2}$		$\frac{9}{10} = .9$
19	The value of Y at point where X is zero (0)	Constant, a, intercept	a	$\bar{Y} - b\bar{X}$	$\bar{Y}_j - b_j \bar{X}_i$	$3 - (.9 * 3) = 3$
20	Y predicted from X	Y hat, Y primed, Y'	Y'	$a + bX$		case 1= 3+(.9*1)=1.2 2= 3+(.9*2)=2.1 3= 3+(.9*3)=3.0 4= 3+(.9*4)=3.9 5= 3+(.9*5)=4.8
21	Regression sum of squares	sum of squares, between sum of squares	SSR, SSB	$\frac{\Sigma xy * \Sigma xy}{\Sigma x^2}$		$\frac{9 * 9}{10} = 8.1$
22	Residual sum of squares	Error, within sum of squares	SSE, SSW	SST - SSR SST - SSB	$\Sigma (Y_i - Y')^2$	$.04 + .01 + 0 + .21 + .64 = 1.9$
23	Total sum of squares	Total sum of squares	SST	$\Sigma y^2$		$4 + 1 + 0 + 4 + 1 = 10$
24	Pearson correlation coefficient	correlation	r	$\frac{\sqrt{\frac{SSB}{SST}}}{\frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}}$		$\frac{\sqrt{\frac{8.1}{10}}}{\frac{9}{\sqrt{10 * 10}}} = .90$

## IMPORTANT

For purposes of intuitive understanding the following concepts are important:  $\sum x^2$ ,  $\sum y^2$ ,  $\sum xy$ ,  $\sum Y'$ ,  $\sum (Y - Y')^2$ , b (beta), a (constant) and  $Y'$  (Y primed). These are all you need to understand the majority of statistics in this book. Get those and you got it. It should be remembered that they may be understood in terms of the five prototypes.

$$\frac{\sum x^2}{N-1}$$

$$\frac{\sum x_y^2}{N_j-1}$$