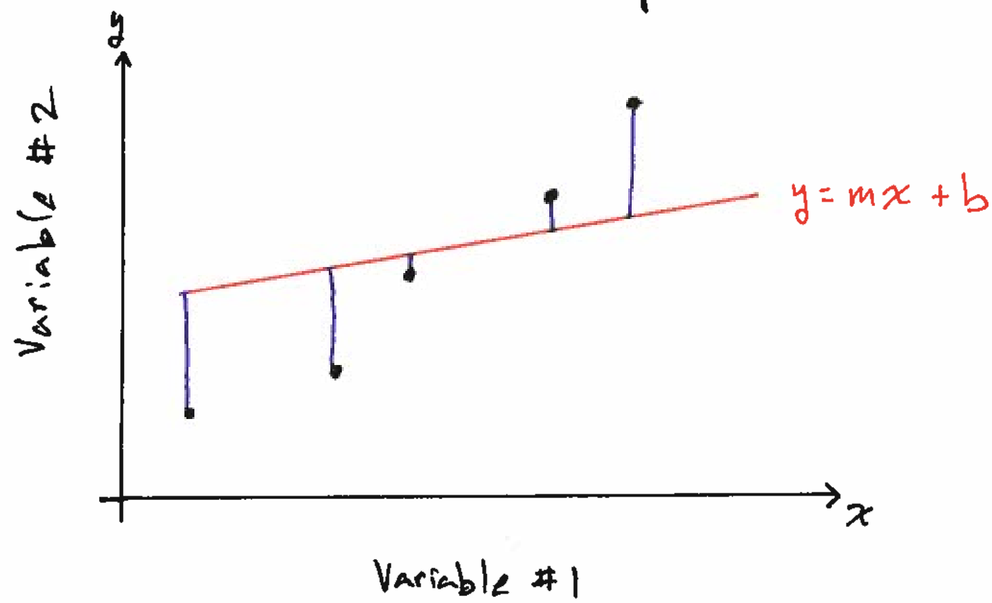
§11.2 Fitting the Model: The Least Squares Approach

Suppose we have data of the form $\{(x_k, y_k)\}_{k=1}^{n}$. That is data $\{(x_1,y_1),(x_2,y_2),(x_3,y_3),\dots,(x_n,y_n)\}$. Also suppose that we have reason to believe that there is a straight line relationship between the x_k's and the y_k's. That is the data points fall roughly along a line. How do we find the "best" line that represents this relationship?



Suppose we have a line given by $y = mx + b$. We can calculate the following quantities:

$$q_1 = \sum (mx_k + b - y_k)$$ sum of purple distances

and $$q_2 = \sum (mx_k + b - y_k)^2$$ sum of squares of distances

We will require that m and b are chosen so that $q_1 = 0$ and q_2 is "as small as possible"

See page 592 for the formulas and example 11.2 in the book for a process to find the equation that gives this <u>Linear Least Squares Regression Line</u>.

We will use our TI calculator.

<u>Example</u> Find the "best fit" line for the following data: $\{(1,2.1),(2,2.8),(3,3.7),(4,4.2),(5,4.9)\}$

See the calculator:

$$y = 1.44 + 0.7x$$

See the Scatter Plot on the calculater.

<u>Example</u> Find the Linear Least Squares line for the data on quiz/final exam scores from section 2.8.

Quiz%	97	98	94	92	98	89	96	91	75	70	87	97	80
FE score	180	172	169	164	158	158	149	146	138	130	129	127	124

Quiz %	79	95	86	87	72	77	84	81	73	85	88	59
FE score	123	119	118	118	116	114	112	105	105	97	78	76

See the calculater:

$$y = 1.72x - 17.6$$

We will talk about the r and r^2 values in the next section.

See scatter plot on the calculater.