

Linear Regression

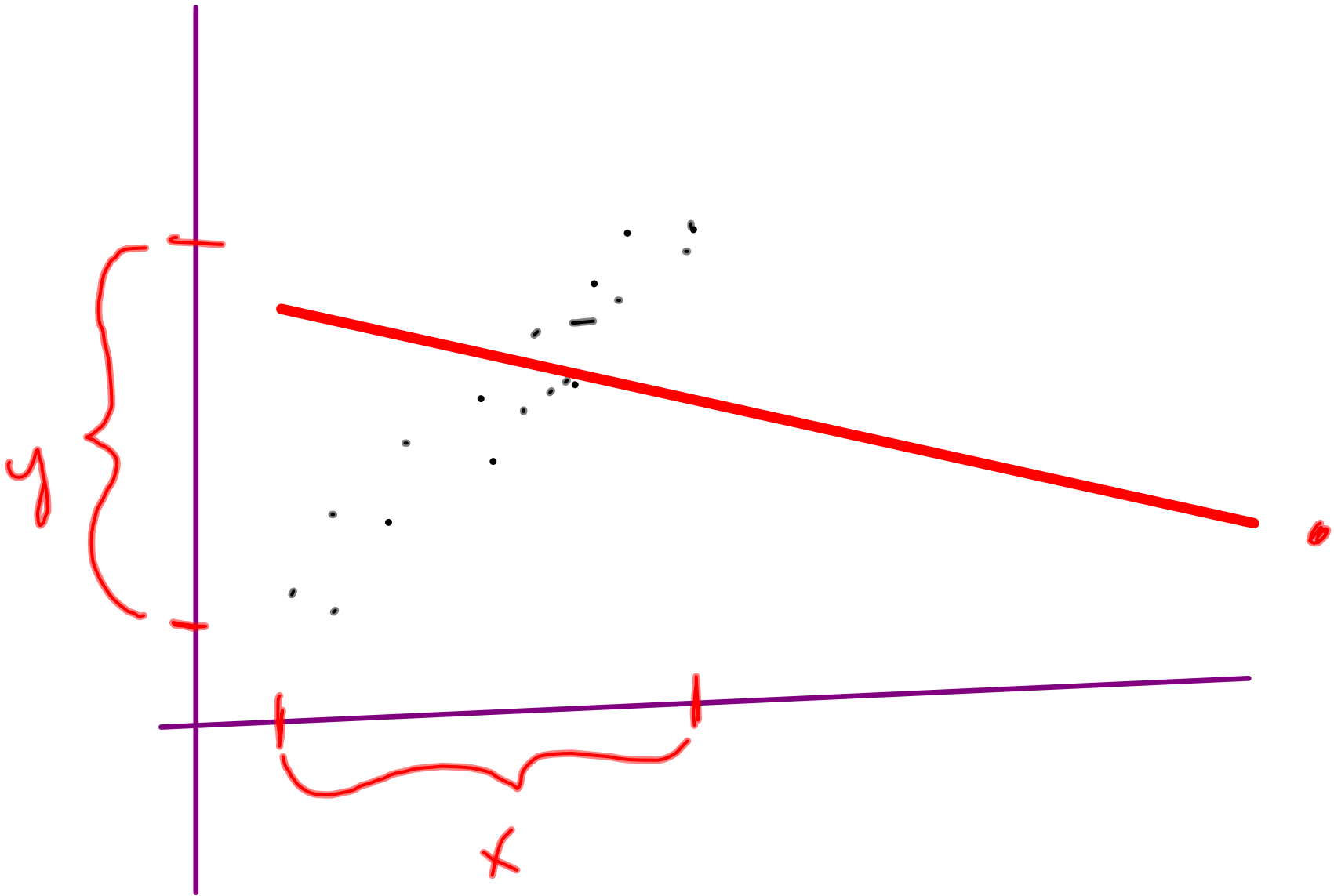
Scatterplot

Direction/Association (r^{\pm})

Form.

Strength (r)

Outlier/Influentials



Linear Regression $\hat{y} = a + bx$

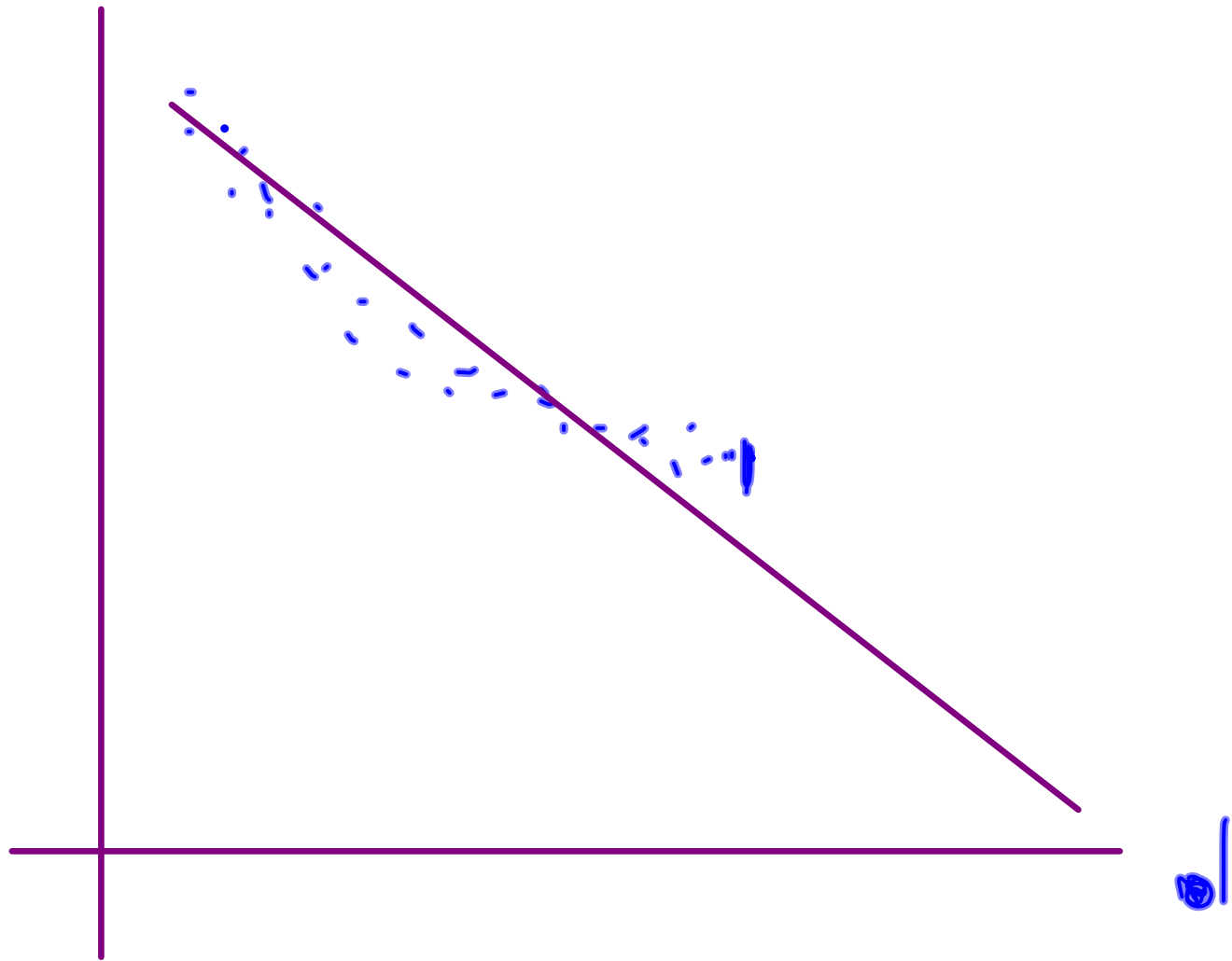
slope $b \rightarrow$ computer: coefficient of variable

intercept $a \rightarrow$ computer: coefficient of constant

prediction plug in x to find \hat{y}

extrapolation - making a prediction outside the range of data

The point (\bar{x}, \bar{y}) is on $\hat{y} = a + bx$



$$\text{Residual} = \text{Actual} - \text{Predicted}$$

(observed)

Residual Plot. Pattern is bad

Non-linear Regression

Linear

X	Y
1	3
2	7
3	11
4	15
5	20

↑
add same

Exp

x	y
1	3
2	6
3	12
4	24
5	49

↑
mult same

Exponential
($x, \log y$)

$$\log \hat{y} = a + bx$$

$$\hat{y} = 10^a \cdot (10^b)^x$$

Power
($\log x, \log y$)

$$\log \hat{y} = a + b \log x$$

$$\hat{y} = 10^a \cdot x^b$$

Make predictions for $x=5$
 $a=2$ $b=3$

$$\log \hat{y} = 2 + 3 \cdot 5$$

$$\log \hat{y} = 17$$

$$\hat{y} = 10^{17}$$


$$\log \hat{y} = 2 + 3 \log 5$$

$$\hat{y} = 10^{\square}$$

The Process

Scatterplot

→ Choose Regression
Residual Plot, r , r^2



$$\underline{r}$$

$$-1 \leq r \leq 1$$

independent of units, x, y

has no units

$$\underline{r^2}$$

Coefficient of Determination

% of variation in y that
can be explained by the
regression line on x