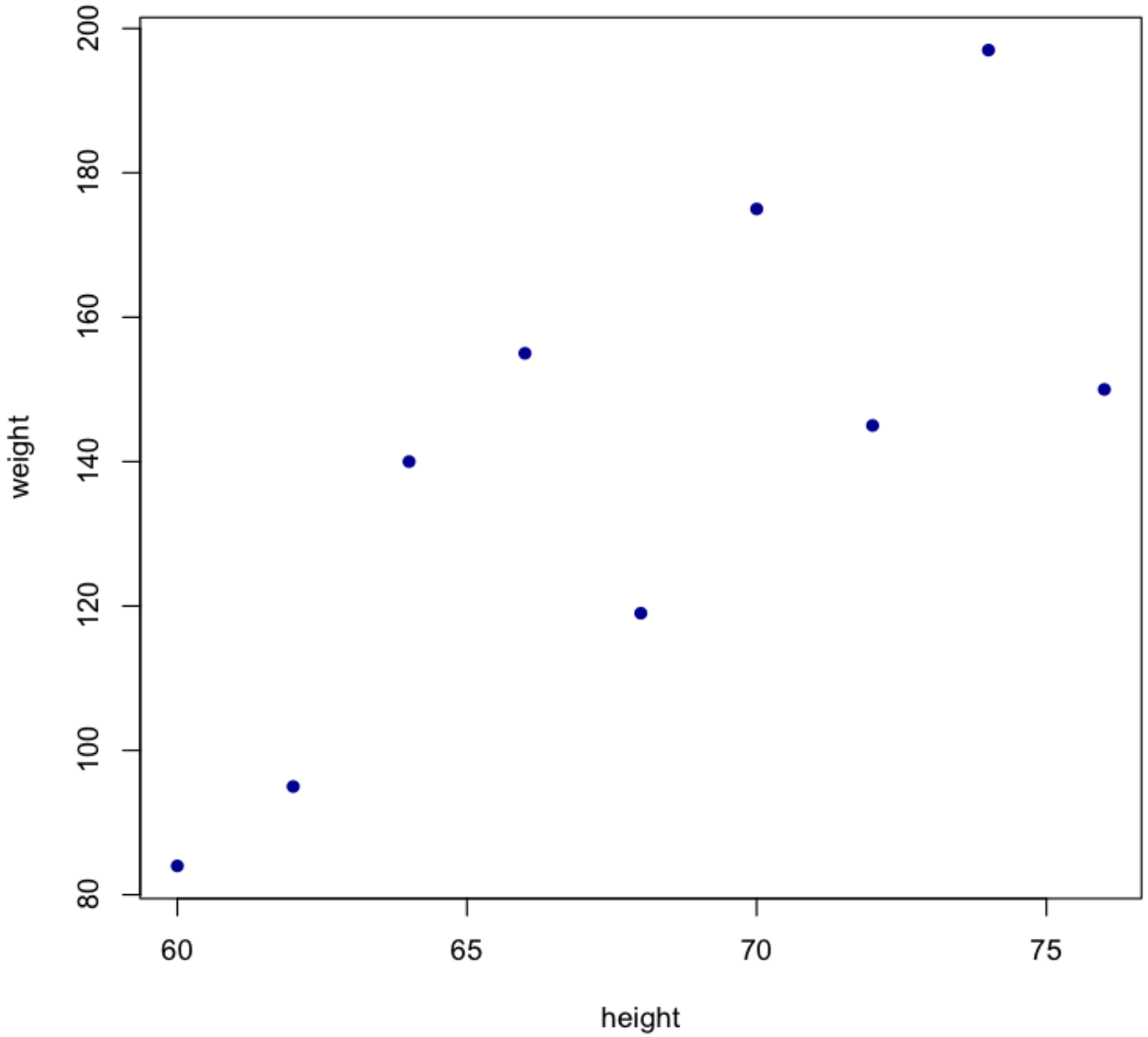
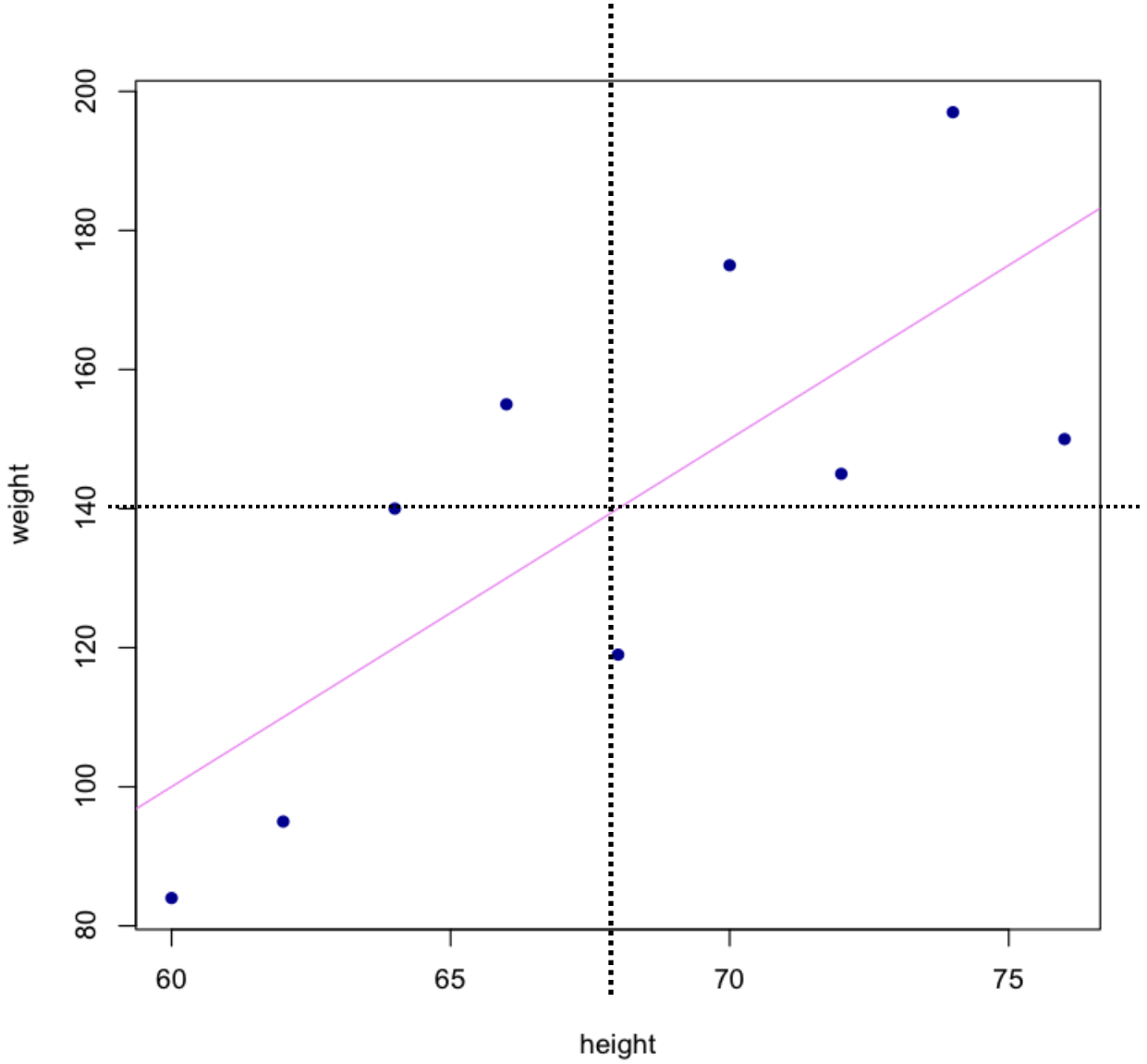
Example: Predicted weight for someone who is 65 inches tall?

Height	Weight
60	84
62	95
64	140
66	155
68	119
70	175
72	145
74	197
76	150



Example:

Height	Weight
60	84
62	95
64	140
66	155
68	119
70	175
72	145
74	197
76	150



Height Weight data:

$\sum X = 612$ $\sum Y = 1260$

$\sum X^2 = 41856$ $\sum Y^2 = 186826$

$\sum (XY) = 86880$

$n = 9$

$\overline{X} = 68$ $\overline{Y} = 140$

$b = 5$

$a = -200$

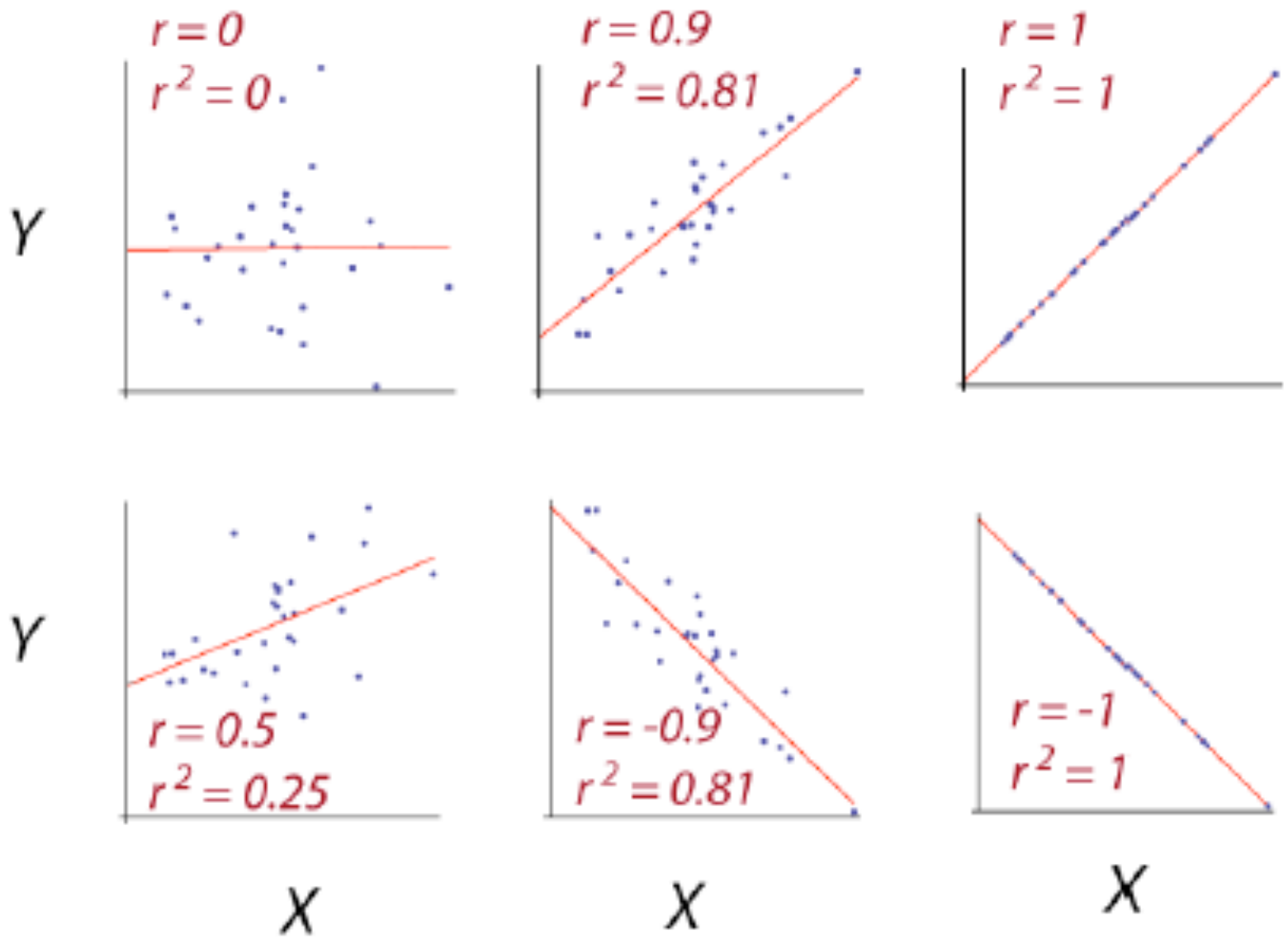
$$\hat{Y} = -200 + 5X$$

R^2 predicts the amount of variance in Y explained by the regression line

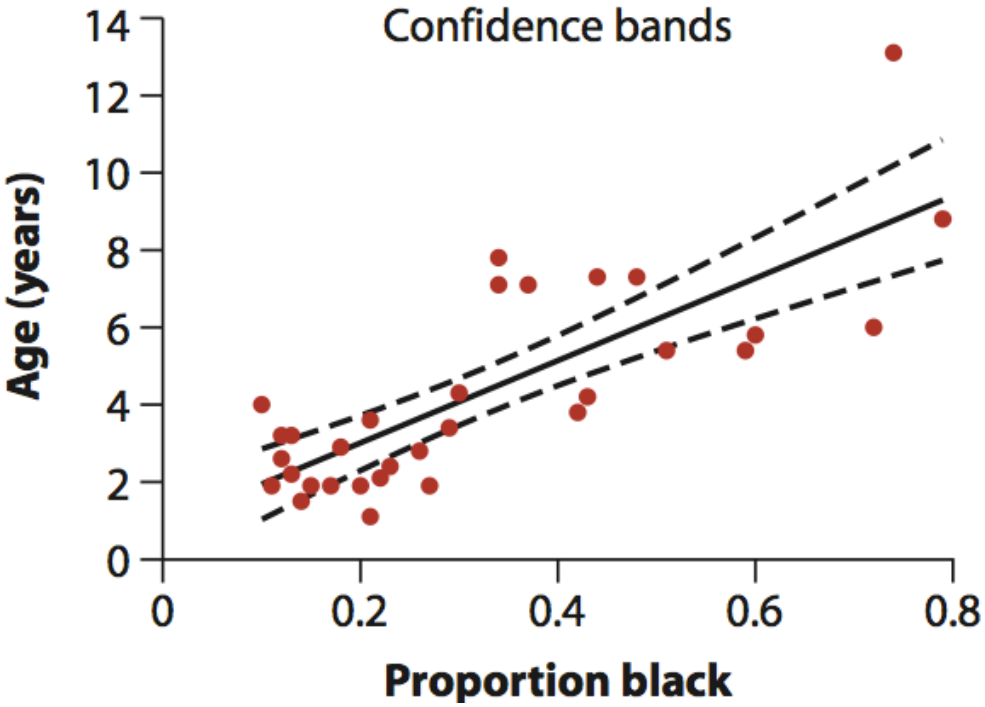
- We saw this in ANOVA where R^2 gave 'precision' of model (ie. Ability of model to explain variation)
- The coefficient of determination
- Sometimes written as r^2
- Square of the correlation coefficient, r

$$R^2 = \frac{SS_{regression}}{SS_{Total}}$$

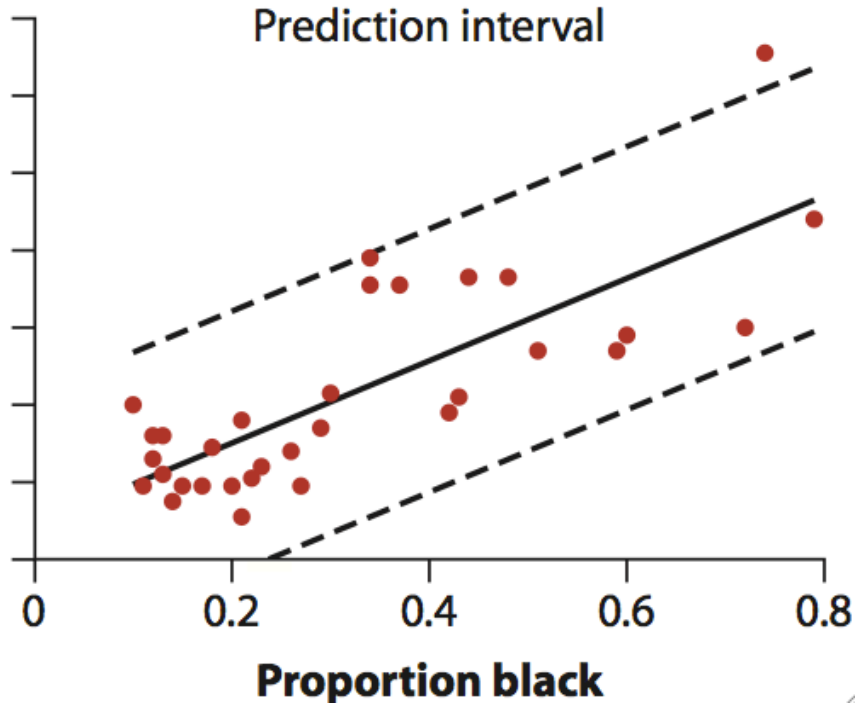
Regression Overview



Prediction confidence:



Predicted mean age



Predicted specific age

Prediction confidence:

The purpose of regression is to **predict**.
There are two types of prediction:

1. $\overline{Y}$ for a given X

2. Single Y for a given X

Both of the above will generate $\hat{Y}$ with the same value but the prediction of a single Y point will have a lower precision

Caution! Do not extrapolate beyond the range of the data

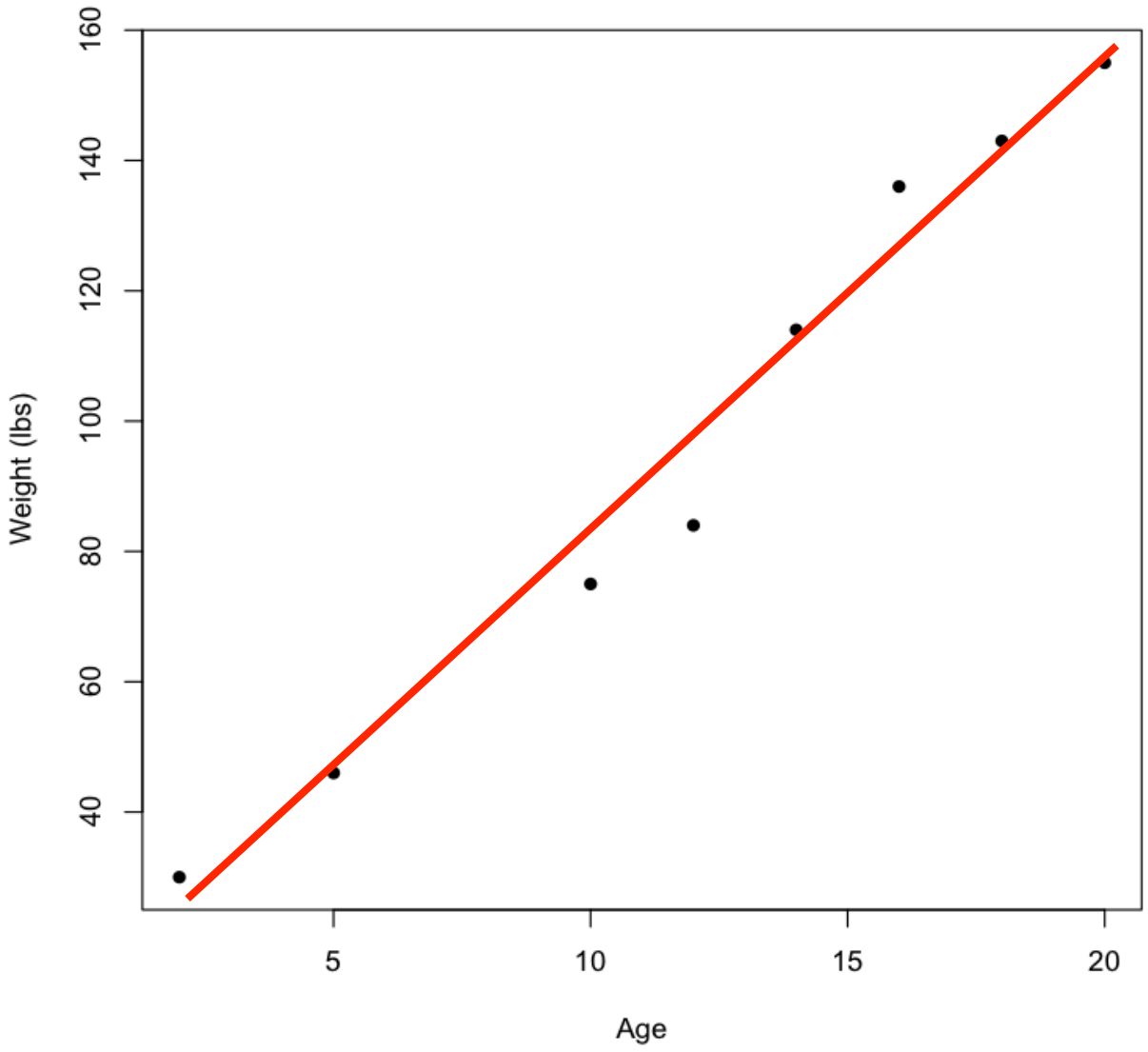
No reason to believe that the relationship between Y and X remains linear beyond the given range

Why we don't extrapolate: Measurements taken over the course of an individual's life:

Age	Weight (lbs)	Time to run one mile	Bench Press (lbs)
2	30		
5	46		
10	75		
12	84	5:40	
14	114	5:05	
16	136	4:40	160
18	143	4:35	180
20	155	4:30	

Regression Assumption Violations

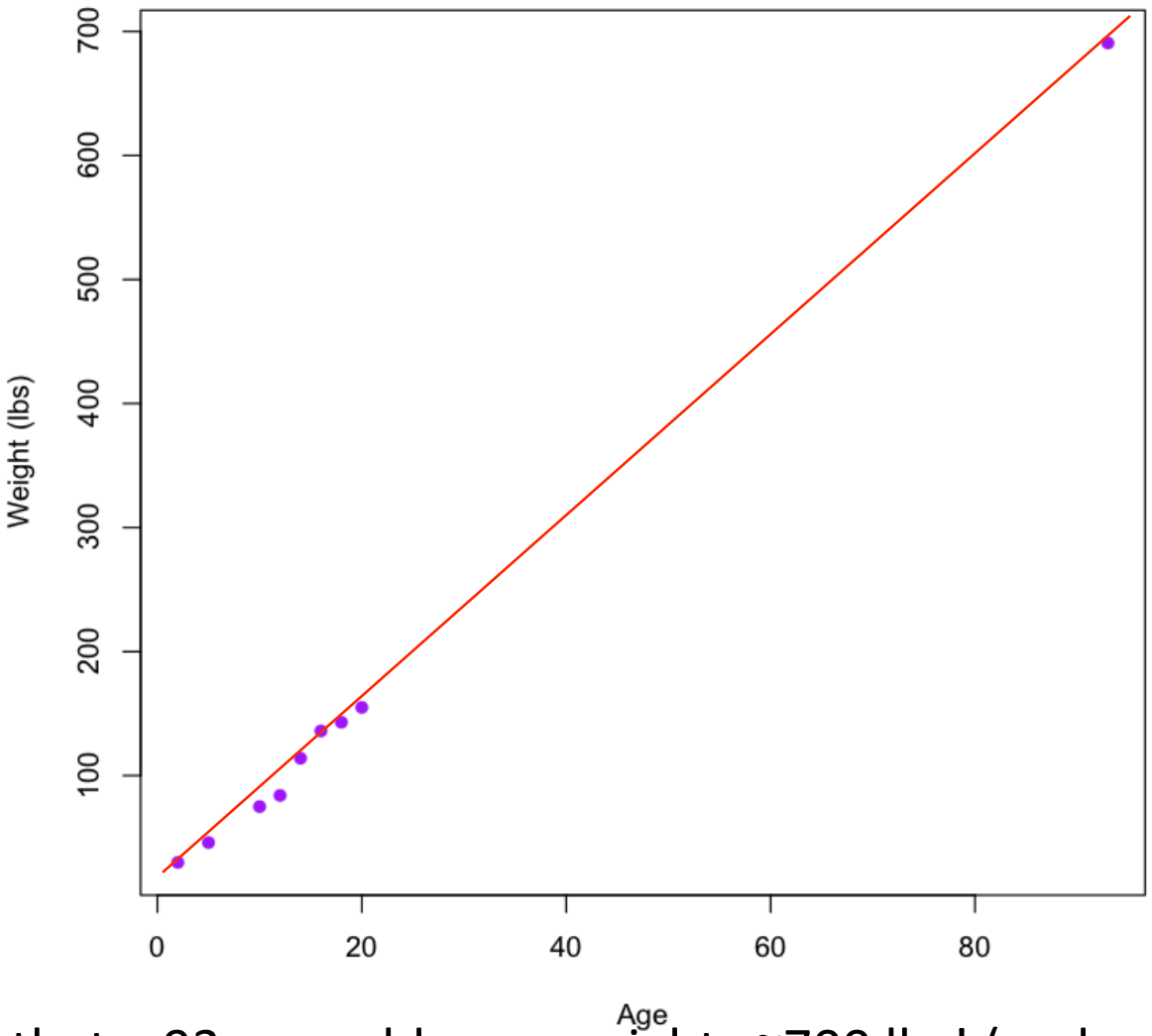
Example: Measurements taken over the course of an individual's life:



Y=9+7.33x

Regression Assumption Violations

Example: Measurements taken over the course of an individual's life:



Y=9+7.33x

This means that a 93 year old man weights ~700 lbs! (and can benchpress half a tonne and running a mile takes him -5 minutes)