Linear Regression and the Coefficient of Determination

Example 1 (from previous lecture):
Is there a connection between the average weight of watermelons a vine produces and the root depth of the vine?
Suspection: vines with deeper roots have a better water supply, and thus larger average melons.
Large watermelon field, 15 vines are chosen at random. At the end of 8 weeks the watermelons are removed from each vine, weighed, and then the average weight (in pounds) is determined. Root depth of each vine is measured (in inches).
Plot a scatter diagram.

Root depth:	*26*	*14*	*18*	*10*	*26*	*21*	*7*	*26*	*13*	*19*	*17*	*13*	*16*	*28*	*23*
Mean weight:	*20*	*10*	*13*	*9*	*19*	*17*	*8*	*15*	*9*	*13*	*12*	*7*	*9*	*17*	*14*

1. *x – root depth, y – mean weight*
for each pland there is an ordered pair (x,y) : (root depth, mean weight)

Correlation coefficient: $r \approx 0.901$ – there is a strong positive linear correlation between the root length and the watermelon weight

Thus $r^2 \approx 0.81$, i.e. 81% of the behaviour (variation) of the *y*-variable can be explained by the corresponding behaviour (variation) of the *x*-variable if we use the equation of the least-squares line. The remaining 19% of the behaviour (variation) of the *y*-variable is due to random chance or to the possibility of lurking variables that influence *y*.

Root depth:	*26*	*14*	*18*	*10*	*26*	*21*	*7*	*26*	*13*	*19*	*17*	*13*	*16*	*28*	*23*
Mean weight:	*20*	*10*	*13*	*9*	*19*	*17*	*8*	*15*	*9*	*13*	*12*	*7*	*9*	*17*	*14*

X	Y	X^2	Y^2	XY
26	20	676	400	520
14	10	196	100	140
Skipped data values				
13	7	169	49	91
16	9	256	81	144
28	17	784	289	476
23	14	529	196	322
Σx=277	Σy=192	Σx^2=5695	Σy^2=2698	Σxy=3882

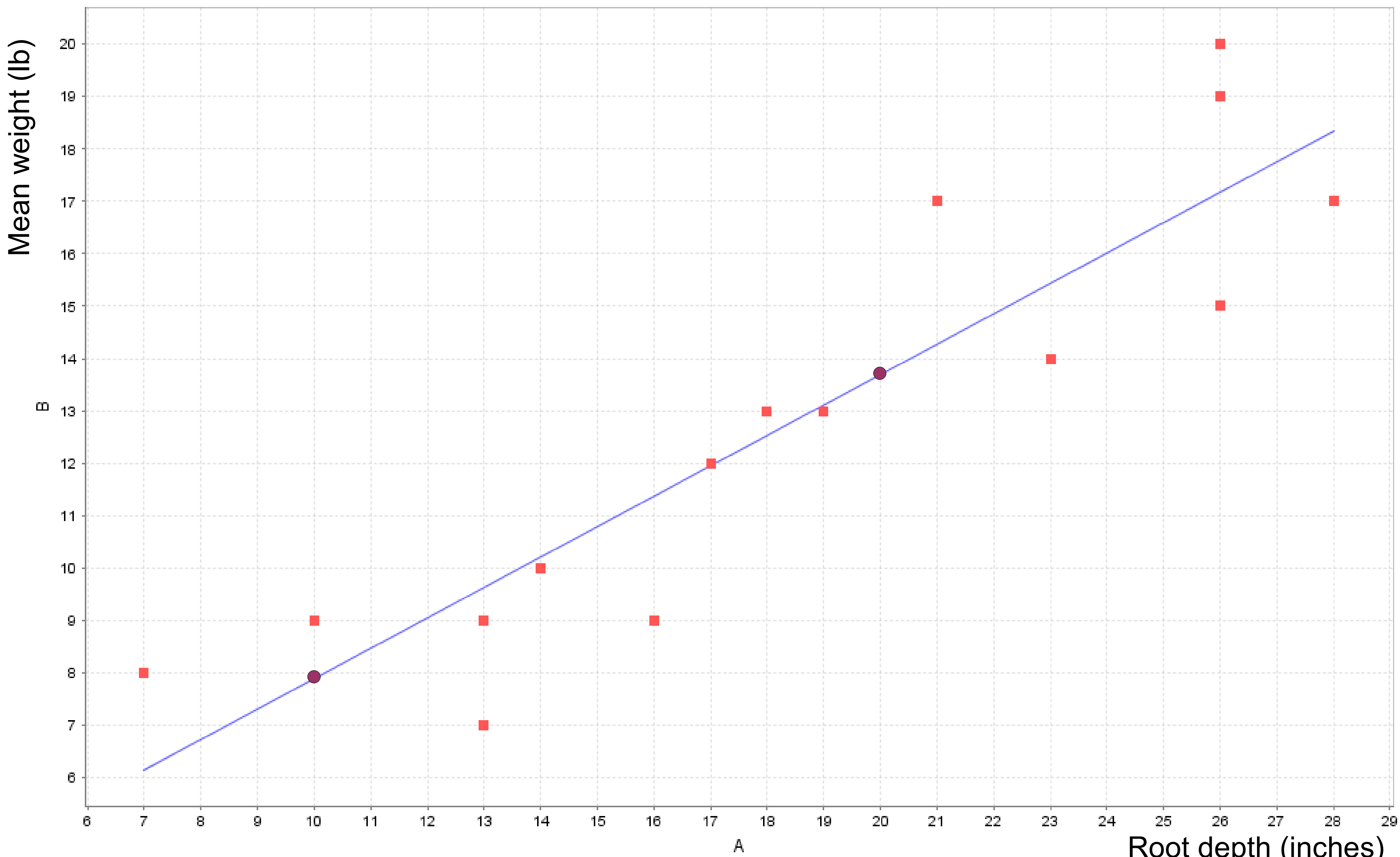
$$b=\frac{(n\sum xy-(\sum x)(\sum y))}{(n\sum x^2-(\sum x)^2)}=\frac{(15*3882-277*192)}{(15*5695-277^2)}=\frac{5046}{8696}\approx 0.58$$ ← slope

$$a=\bar{y}-b\bar{x}=\frac{\sum y}{n}-b\frac{\sum x}{n}=\frac{(\sum y-b\sum x)}{n}=\frac{(192-\frac{5046}{8696}277)}{15}\approx 2.08$$ ← y - intercept

Thus the equation of the least-squares line is
y = 2.08 + 0.58x

Let's draw the least-squares line now:

Root depth (x):	*26*	*14*	*18*	*10*	*26*	*21*	*7*	*26*	*13*	*19*	*17*	*13*	*16*	*28*	*23*
Mean weight (y):	*20*	*10*	*13*	*9*	*19*	*17*	*8*	*15*	*9*	*13*	*12*	*7*	*9*	*17*	*14*



$y = 2.08 + 0.58x$

two points: when x=10 , y = 2.08+0.58*10=7.88 *(10, 7.88)*

when x=20, y = 2.08+0.58*20=13.68 *(20,13.68)*

! Point $(\bar{x},\bar{y})$ or *(18.47,12.8)* is always on the least-squares line