Chapter 11 Exercise Set A

1 Tall and Chubby; Short and skinny.

2a False, the dot is less than 180.

2b True, he is above average with regard to his height.

3

Actual y	Predicted y	Error	(Error)^2
57	64	7	49
63	62	1	1
43	40	-3	9
51	52	1	1
49	45	-4	16

signs don't matter!

15.2 = ave

$\sqrt{15.2} = 3.90$

4 a → .2
b → 1
c → 5 (scaling)

5 A few thousand dollars; the prediction is likely to be ±2 RMS errors away from the prediction.

6 7; the guesses will be closer to predicted on average.

7a ± 8 points; 1 RMS error

7b ± 16 points; 2 RMS errors.

8a If you ignore husband's income than the RMS error is the SD so 20,000.

8b The Solid one, it is the regression line and all predictions are on it by definition.

Chapter 11 Exercise Set B

1 First year score is y so $\sqrt{1-(.6)^2} \times 10$.

2a Guess the average first year score! (65)

2b 10, because we only have one variable it is the SD.

2c Use the regression method now with x being LSAT score and y being the first-year score.

3 Person B makes the smaller RMS error. Person A's RMS error is the SD, one variable. Person B's RMS error will be $\sqrt{1-(.6)^2} \times .5 = .4$

Person A's error = .5
Person B's error = .4

So it reduces the error by 1/5

Chapter 11 Exercise Set C

1 i - something is wrong. The histogram should be centered about 0.
ii - (b) the spread is almost to ±50. The only option that makes sense is b because c is too skinny.
iii - (c) the spread is short, ±10 or so, this corresponds to an RMS of 5.

2 i - (a) although the plot isn't symmetrical, it still has the correct error of about ±2,000 2 RMS errors.
ii - (a) Same logic as above, the spread is at ±2,000, this one looks better.
iii - something is wrong, the residuals look like a scatter-plot that is football shaped not a residual plot.

3 (a) - look at the scatter plot!
Range 6 - 1 so $6-1 = 5/4 = 1.25 \approx 1.0$.
So I guess the SD of y is 1.0.

(b) look at the residual plot!
Range is about 1.5 to -1.5 so
$-1.5 - 1.5 = -3/4 = .75 \approx .60$
so I guess the SD of residuals is .60.

(c) It will look like the residual plot so .60.

Chapter 11 Exercise Set D

1a $\sqrt{1-.8^2} \times 15 = \sqrt{1-.64} \times 15 = \sqrt{.36} \times 15 = .6 \times 15 =$ about 9 True

1b True, the scatter diagram is homoscedastic so the error for any x is going to be 9.

1c False, the scatter diagram is heteroscedastic so the RMS error should not be used.

2a Son's height is y.

$\sqrt{1-.5^2} \times 2.7 = \sqrt{1-.25} \times 2.7 = \sqrt{.75} \times 2.7 = 2.34$"

2b ① $\frac{72-68}{2.7} = 1.48$

② $1.48 \times .5 = .74$

③ $y = .74(2.7) + 69 = 71$"

2c 2.34 or so, an RMS error.

2d ① $\frac{66-68}{2.7} = -.74$

② $-.74 \times .5 = -.37$

③ $y = -.37(2.7) + 69 = 68$

It is likely to be off by 2.34 again.

3a y is income.

$\sqrt{1-.37^2} \times 20,000 = \sqrt{1-.1369} \times 20,000 = \sqrt{.8631} \times 20,000 = 18580.64$

3b ① $\frac{16-13}{3.4} = .88$

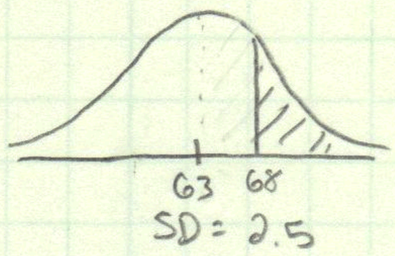
② $.88 \times .37 = .33$

③ $y = .33(20,000) + 18,000 = 24,529.41$

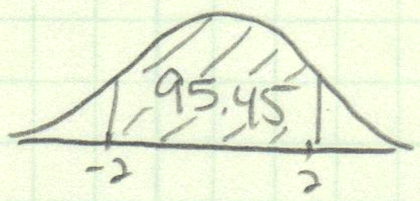
3c The scatter plot is heteroscedastic so we can't use the RMS error to predict the error. More info is needed.

Chapter 11 Exercise Set E

1a One Variable! Normal Curve!



$\frac{68-63}{2.5} = 2$ $\quad 2 \approx 95.45$



TAIL! $\frac{100-95.45}{2} = 2.275\%$

1b Vertical Strip!

New Average: y is women's height. use regression.

① $\frac{72-68}{2.7} = 1.48$

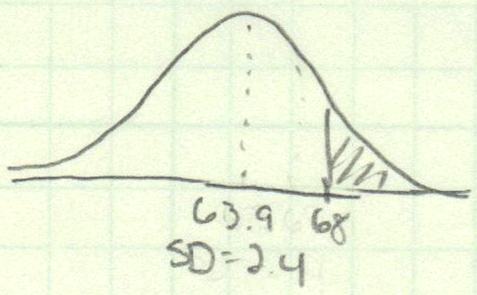
② $1.48 \times .25 = .37$

③ $y = .37(2.5) + 63 \approx 63.9"$

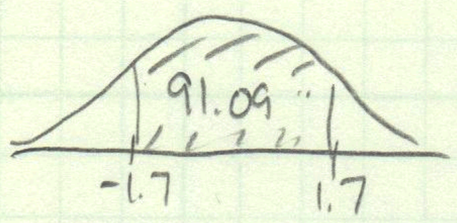
New SD: Use RMS error!

$\sqrt{1-.25^2} \times 2.5 = \sqrt{1-.0625} \times 2.5 = \sqrt{.9375} \times 2.5 = 2.4$

Normal curve now!

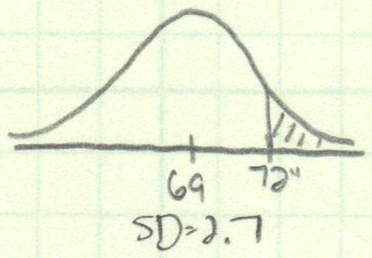


$\frac{68-63.9}{2.4} = 1.70$ $\quad 1.70 \approx 91.09\%$

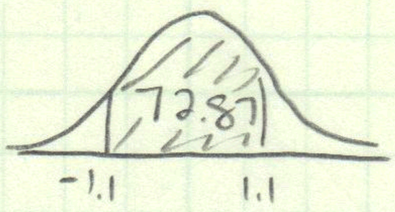


TAIL! $\frac{100-91.09}{2} = 4.455\%$

12a One Variable! Normal Curve



$\frac{72-69}{2.7} = 1.11$ $\quad 1.10 \approx 72.87\%$



TAIL $\quad \frac{100-72.87}{2} = 13.565\%$

12b Vertical Strip!

New average: y is son's height use regression

① $\frac{72-68}{2.7} = 1.48$

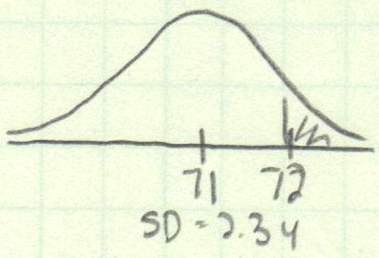
② $1.48 \times .5 = .74$

③ $.74(2.7) + 69 = 71"$

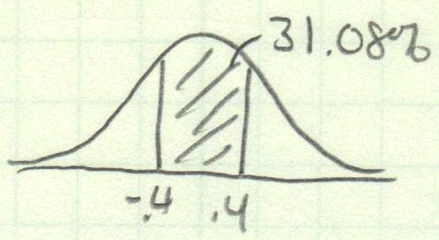
New SD: Use RMS error.

$\sqrt{1-.5^2} \times 2.7 = \sqrt{1-.25} \times 2.7 = \sqrt{.75} \times 2.7 = 2.34"$

Normal Curve Now!

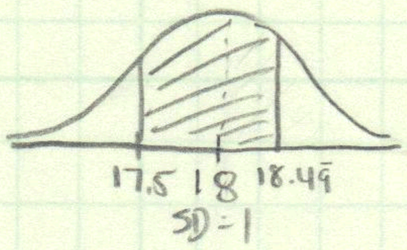


$\frac{72-71}{2.34} = .42$ $\quad .40 \approx 31.08\%$



TAIL! $\quad \frac{100-31.08}{2} = 34.46\%$

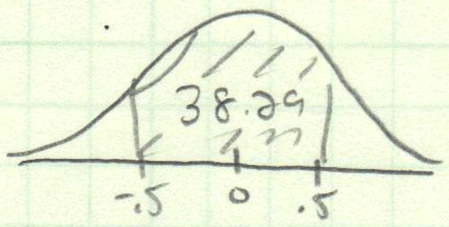
3a One variable! Normal Curve "Tricky" - Nearest inch!



$\frac{18.5-18}{1} = .5$

$\frac{17.5-18}{1} = -.5$

$.5 \approx 38.29$



The pictures match so <u>38.29%</u>

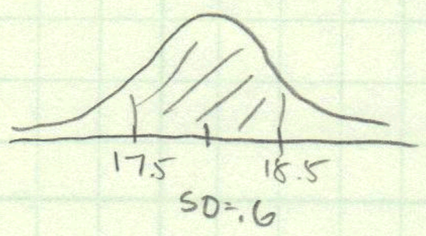
3b Vertical Strip! "Tricky" - Nearest inch!

Regression to find New average BUT 68 is one of the points of average so the other has to be 18". = New Average.

New SD: RMS error. y is forearm length!

$\sqrt{1-.8^2} \times 1 = \sqrt{1-.64} \times 1 = \sqrt{.36} \times 1 = .6$

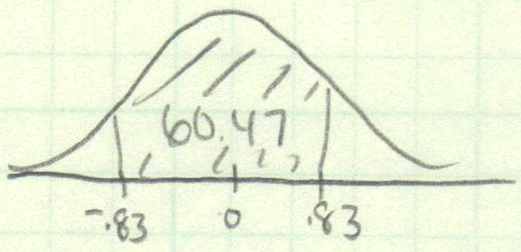
Normal Curve Now!



$\frac{18.5-18}{.6} = .833$

$\frac{17.5-18}{.6} = -.833$

$.85 \approx 60.47\%$



The pictures match so <u>60.47%</u>

Chapter 11 Review Exercises

1 (v)

2 Yes! The RMS error is way too big it should be a give or take number but 3.12 covers almost the entire spread of 0-4.0. I think the scatter plot is heteroscedastic.

3a y is height at 18.

$\sqrt{1-.80^2} \times 2.5 = \sqrt{.36} \times 2.5 = .6 \times 2.5 = 1.5"$

3b y is height at 6

$\sqrt{1-.8^2} \times 1.7 = .6 \times 1.7 = 1.02"$

4a by more than one RMS error! $100 - 68 \approx 32\% \approx 1/3$.

RMS error $= \sqrt{1-.6^2} \times 15 = \sqrt{1-.36} \times 15 = \sqrt{.64} \times 15 = .8 \times 15 = \underline{12}$.

4b y is final score.

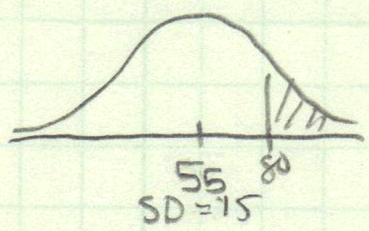
① $\frac{80-50}{25} = 1.2$

② $1.2 \times .60 = .72$

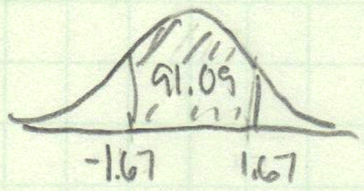
③ $.72(15) + 55 = 65.8$.

4c 12 points, the RMS error.

5a Normal Curve, one variable!



$\frac{80-55}{15} = 1.67$ $\quad 1.70 \approx 91.09$



TAIL! $\frac{100 - 91.09}{2} = 4.455\%$

5b Vertical Strip.

New average: y is final, use regression.

① $\frac{80-50}{25} = 1.2$

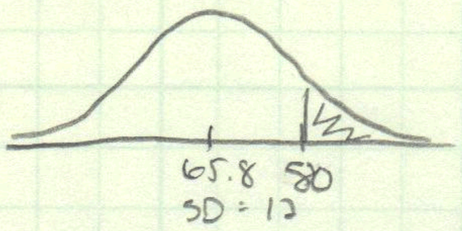
② $1.2 \times .60 = .72$

③ $y = .72(15) + 55 = 65.8$

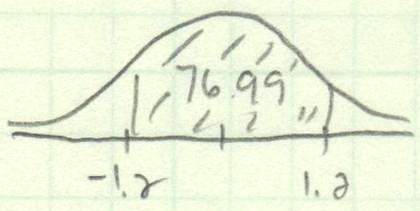
New SD!

RMS error $= \sqrt{1-.6^2} \times 15 = 12$.

Normal Curve Now!



$\frac{80\ 65.8}{12} = 1.18$ $\quad 1.2 \approx 76.99\%$



TAIL $\quad \frac{100-76.99}{2} = 11.505\%$.

6 No. Association is not causation, confounding factors exist.

7 Lii, the regression effect will pull scores toward average for the second test.

8 a) Replication will make the test more accurate, it could be chance error someone does well or not.

b) Lifting weights. You may be getting stronger or it could just be chance error, repeat with similar weight a few days after to see if you can still do it.

9 No! This is an example of the regression fallacy. Blaming someones decline or incline on anything other than the regression fallacy.

10 No! Look at the differences between 2005 actual and 2006 predicted; IF regression were used the 2006 predicted should look like a line, but they don't!

11 Look at where the clustering happened. 13 and 17 years. Most people stop after 12 and 16 years so this sample must have counted an extra year, i.e. kindergarten.

12 His blood pressure should decline as his educational level goes up because of the negative correlation coefficient. While his educational level is quite high, his blood pressure is almost ave, "way" high for his educational level. Doing the regression method shows this.

① $\frac{20-13}{3} = 2.33$

② $2.33 \times -.1 = -.233$

③ $-.233(11) + 119 = 116$

His 118 is higher than it should be.