Homework Answers:

4. **Off to college.** Choice D. $\widehat{gpa\_U} = 0.22 + 0.72 gpa\_HS = 0.22 + 0.72(3.8) = \underline{2.956}$ The residual is $\underline{3.5}$ - $\underline{2.956}$ = 0.544.

6. **Cereals.** given

   $\widehat{Potassium} = 38 + 27 Fiber = 38 + 27(9) = 281$ mg. According to the model, we expect cereal with 9 grams of fiber to have 281 milligrams of potassium.

8. **More cereal.**

   A negative residual means that the potassium content is actually lower than the model predicts for a cereal with that much fiber.

10. **Another bowl.**

    The model predicts that cereals will have approximately 27 more milligrams of potassium for each additional gram of fiber.

12. **Human Development Index.**

    a) Fitting a linear model to the association between HDI and GDPPC would be misleading, since the relationship is not straight.

    b) If you fit a linear model to these data, the residuals plot will be curved downward.

14. **Residuals.**

    a) The scattered residuals plot indicates an appropriate linear model.

    b) The curved pattern in the residuals plot indicates that the linear model is not appropriate. The relationship is not linear.

    c) The fanned pattern indicates that the linear model is not appropriate. The model's predicting power decreases as the values of the explanatory variable increase.

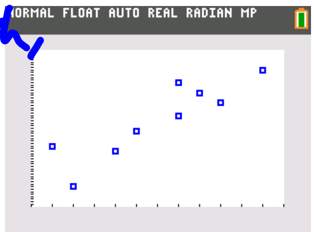
*Name*____________________________________ *Statistics Chapter 7: Regression and Back?*

In *MathWorld* you can use the equation *Perimeter* = *4* x *side length* to find the perimeter of a square using the length of its side, OR, you can use that same equation to work backwards to find the length of side of a square if you know its area. ***Let's see if in Statistics, regression models can also be used forward and backward.***

Enter the following measurements of 9 male subjects into calculator lists:

x waist (in)	32	36	38	33	39	40	42	35	38
y weight (lb)	175	181	200	159	196	192	205	173	187

1. Examine a scatterplot of the (*waist*, *weight*) data. Is it appropriate to fit a linear model to this data?
   x y
   yes (No outliers, straight enough, quant. variables)



2. Find the correlation and equation the best fit line for the (*waist*, *weight*) data:
   r = .886 $\widehat{weight}$ = 38.2 + 3.98 waist
   rel. strong (+)

3. Use your (*waist*, *weight*) model to predict the weight of a man with a 37 inch waist.

$$\widehat{weight} = 38.2 + 3.98(37)$$
$$= 185.3 \text{ lbs.}$$

NORMAL FLOAT AUTO REAL RADIAN MP
LinReg
y=a+bx
a=38.19379845
b=3.976744186
r^2=0.7843405488
r=0.8856300293

4. Use algebra to solve the equation from question 2 for the *waist* variable.

$\widehat{weight} = 38.2 + 3.98\ waist$

$-38.2 \qquad -38.2$

$\frac{weight - 38.2}{3.98} = \frac{3.98\ waist}{3.98}$

$waist = \frac{weight - 38.2}{3.98} = \frac{-38.2}{3.98} + \frac{1}{3.98} weight$

waist = $-9.605 + .251\ weight$

5. Use the equation from question 4 to find the waist of a man *predicted to weigh* 210 pounds.

$waist = -9.605 + .251(210) = 43.1''$

6. Now reverse the role of the variables and find the correlation and equation of the best fit line for the (*weight*, *waist*) data using regression: $r =$ .886 $\widehat{waist} =$ .446 + .197 weight

L2 L1

7. Use the equation from question 6 to predict the waist of a man who weighs 210 pounds.

$\widehat{waist} = .446 + .197(210) = 41.8''$

```
NORMAL FLOAT AUTO REAL RADIAN MP

LinReg
y=a+bx
a=0.446366782
b=0.1972318339
r²=0.7843405488
r=0.8856300293
```

8. Why are the equations in questions 4 and 6 (and consequently the waist measurements found for a 210 pound man) different? *Hint: The slope of each regression equation is* $\frac{rs_y}{s_x}$.

9. (C-level) What is the only (and almost never true) condition under which the (x, y) regression equation solved for x gives the same predictions as the (y,x) regression equation?

A Tale of Two Regressions

- You might be tempted to use a regression equation backwards: plug in a *y*-value and predict an *x*-value.
- * But that doesn't work. Our equation is not built to minimize predictions in the *x*-direction.
- * If you want to make predictions about the *y*-variable, you need to swap the variables and find a new regression equation.

(switch L_1, L_2 and recalculate)

Just Checking

Let's look at the student height/weight relationship one last time. Earlier you created the linear model $\widehat{Weight} = -149 + 4.4Height$ to use students' heights to estimate their weights.

Now suppose we know a student weighs 200 pounds and want to estimate his height. That earlier model won't work, because it uses *actual heights* to find *estimated weights*. Now we know the *actual weight* x and want to *estimate height*. y We need a new model.

18. In this new model, which is the explanatory variable wt. and which is the response variable? Ht.

19. Create the appropriate model. You may use the summary statistics, or enter the actual data into your calculator to find the equation of the regression line. Either way, be sure to think carefully about which variable is x and which is y.

20. How tall do you estimate the 200-pound student is?

y *Height* (inches)		x *Weight* (pounds)
66		147
62		124
71		189
64		141
75		172
70		144
64		112
71		165
69		129
68		177
68	**Mean**	150
4	**SD**	25
	$r = 0.7$	

(Check your answers on page 193.)

18. *Weight* is explanatory; *Height* is response.

19. $\widehat{Height} = 51.2 + 0.112Weight$

20. $51.2 + 0.112(200) = 73.6$ inches tall