EHW 8

1. sum of residuals – small as possible → 0

Patterns = BAD in residual plot ⇒ not a good linear model

5. look at defns.

6. Wednesdays notes
8. expl = temp.
12. good if |r| close to 1 and residual plot has no pattern

WH 8

5.4

5. Create a residual plot for the values from problem 17 in section 5.3. Look at the residual plot for this data and the value of the correlation to determine if the LSRL is a good model for the data. Is there an influential point? If so, identify it, and remove it from your data. Sketch the new LSRL for the data.

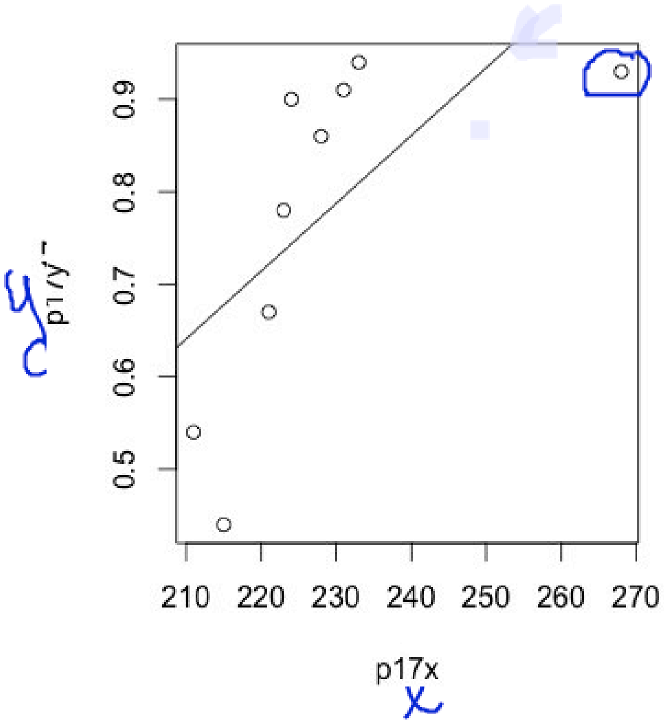
17. The following 9 observations compare the Quetelet index, x (a measure of body build) and dietary energy density, y.

x	221	228	223	211	231	215	224	233	268
y	.67	.86	.78	.54	.91	.44	.9	.94	.93

a. Make a scatter-plot of the data.
b. Compute the LSRL.
c. Provide an interpretation of the slope of this line in the context of these data.
d. Find the correlation coefficient for the relationship. Interpret this number.
e. Find the coefficient of determination for the relationship. Interpret this number.

Public Page

scatter plot




```
> lsrlp17=lm(p17y~p17x)
> lsrlp17

Call:
lm(formula = p17y ~ p17x)

Coefficients:
(Intercept)         p17x
   -0.89846      0.00733
```

$\hat{y} = -.89846 + .00733x$

on calc

LinReg

```
> cor(p17x,p17y)
[1] 0.6578901
> 0.6578901^2
[1] 0.4328194
>
```

r

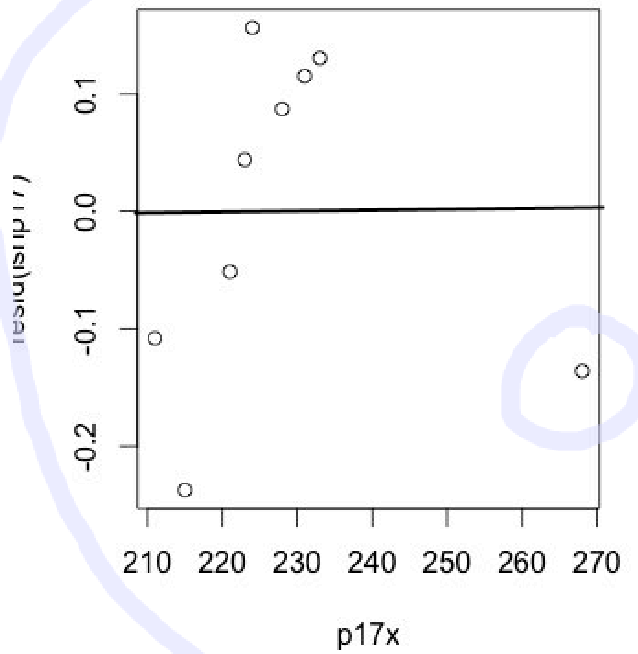
r^2

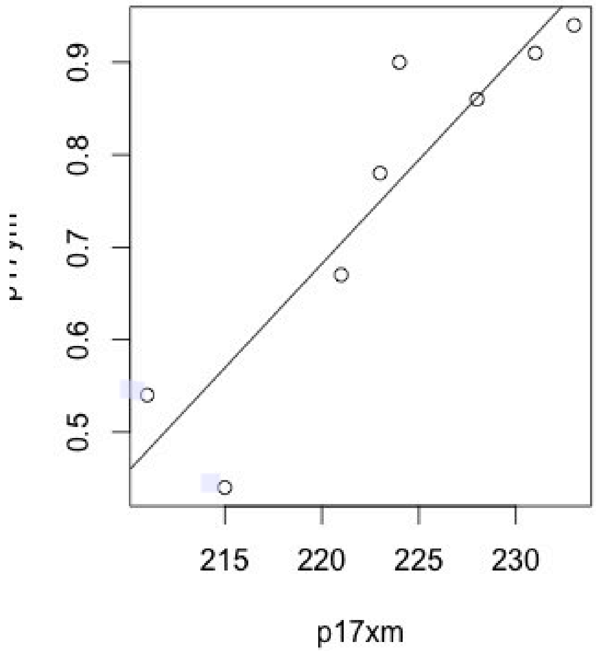
strong or weak?
or moderate

- linear pattern?
- positive or neg.?

```
> resid(lm(p17y~p17x))
          1           2           3           4
-0.05150453  0.08718448  0.04383518 -0.10820312
          5           6           7           8
 0.11519405 -0.23752368  0.15650504  0.13053377
          9
-0.13602119
> plot(p17x,resid(lsrlp17))
```

$L_3 = L_2 - Y_1(L_1)$

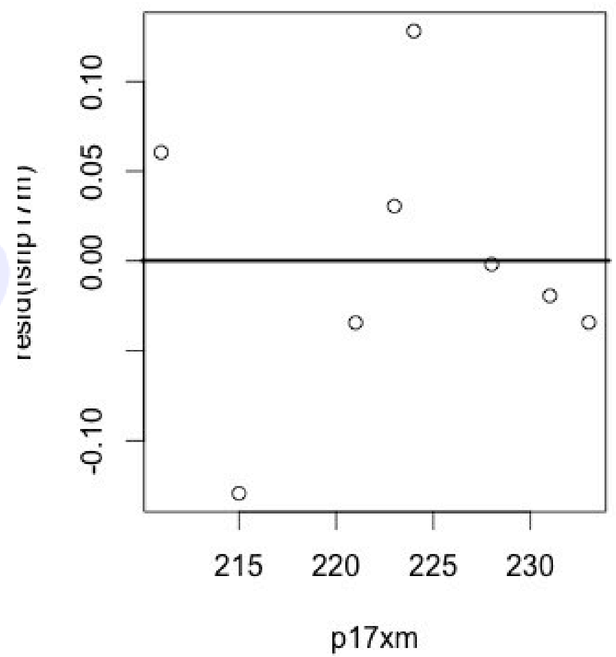





```
> plot(p17x,resid(lsrlp17))
> p17xm=c(221,228,223,211,231,215,224,233)
> p17ym=c(.67,.86,.78,.54,.91,.44,.9,.94)
> plot(p17xm,p17ym)
> lsrlp17m=lm(p17ym~p17xm)
> abline(lsrlp17m)
>
```

```
> cor(p17xm,p17ym)
[1] 0.9130748
> 0.9130748^2
[1] 0.8337056
```

Stronger



5.6 #4

4. The following two-way table describes the age and marital status of American women in 1991. The table entries are in thousands of women.

Age	Single	Married	Widowed	Divorced
18-24	9,008	3,352	8	257
25-39	6,658	21,769	248	3,224
40-64	1,975	24,462	2,570	4,755
65+	900	7,255	8,464	925

total

total 18541

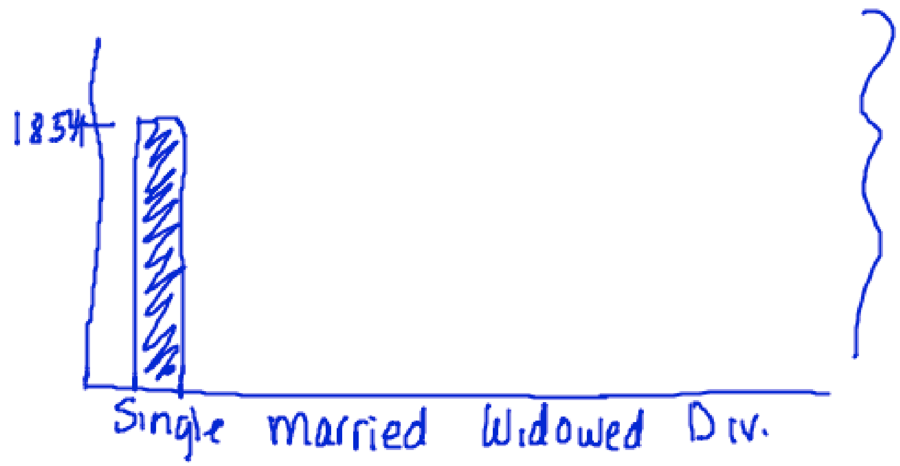
a. Construct the marginal distributions of this table in counts.

b. Draw a bar chart to display the marginal distribution of the marital status for all adult women (use percentages).

c. What percent of adult American women under the age of 25 have never married?

d. What percent of 25-39 year old American women are divorced?

e. Compare the conditional distributions of marital status for women aged 18 to 24 and women aged 40 to 64. Briefly describe the most important differences between the two groups of women, and back up your description with percentages.



$$\frac{\text{single \& under 25}}{\text{under 25}} = \frac{9008}{}$$

$$\frac{\text{div and 25-39}}{\text{25-39 yrs.}} = \frac{3224}{}$$

e. 18-24 cond. distr.

Single	Married	Wid.	Div.
$\frac{9008}{12625}$	$\frac{3352}{12625}$	$\frac{8}{12625}$	$\frac{257}{12625}$

5.5

4. The population in the city of Houston from 1900 to 2010 is given below:

	Year	Population
	Year	Population
0	1900	44,633
10	1910	78,800
20	1920	138,276
30	1930	292,352
	1940	384,514
	1950	596,163
	1960	938,219
	1970	1,233,505
	1980	1,595,138
	1990	1,631,766
	2000	1,953,631
110	2010	2,100,263

a. Give a scatter-plot and residual plot of the data.

b. Based on the graphs in part a, propose a model for the data. Show me evidence to support your conclusion. Go through all necessary steps to construct a model of the type you chose.

$$\sqrt{y} = a + bx$$

$$\boxed{\hat{y} = (a+bx)^2}$$