

1. 1.4. This illustrates the presence of the random error term  $\epsilon_i$ , which implies that different  $Y$  observations at the same  $X$  can (and usually will) differ.
2. 1.5 No,  $E\{Y_i\}$  is an expected value and so is constant. The expression the student gave has a random quantity,  $\epsilon_i$ , in it.
3. 1.7
  - (a) No. Knowing the expected value and variance of  $Y$  does not allow us to say anything about the exact probability of  $Y$  falling between 195 and 200.
  - (b) With a normal model, then  $Y \sim N(100 + 5(20), 25) = N(200, 25)$ ; that is has a normal distribution with mean 200 and variance 25. Now,  $P(195 < Y < 205) = P((195 - 200)/5 < Z < (205 - 200)/5) = P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = .8413 - (.1587) = .6826$ .
4. 1.8.  $E(Y)$  is still 104, but that doesn't mean that another  $Y$  value will have the same value (108) that the first  $Y$  value did.
5. 1.30 The expected value of  $Y$  is the same for any  $X$ ; that is there is not relationship between  $X$  and  $E(Y|X)$ . The regression plot will be a flat line at height  $\beta_0$ .
6. 1.31. This one is a bit subtle. Suppose we view the regression term  $\beta_0 + \beta_1 X$  to reflect the expected hardness at time  $X$  over a random sample of units. In problem 1.22 the error term will essentially incorporate "among unit effects" (due to the fact that all units do not have the same expected hardness at time  $X$ ) and "within unit effects" (due to the fact that the hardness of a specific unit at time  $X$  is still random due to various factors that might influence the hardening over time and also that there might be measurement error in measuring hardness). The unit effect only enters into the one error term associated with the observation for that unit. In problem 1.31 the data is from just one unit observed over time. With the regression line specific to that unit, then the error term will include just the "within unit effects" associated with that one unit and one might expect these to be correlated over time since whatever factors enter into the the within unit effects, such as environmental conditions, may persist over time.
7. # R example showing basics of reading and listing  
 # a data file with variable names in the first row  
 # (indicated by header=T)  
 # and plotting one variable versus the next.

```
breakdata<-read.table('f:/s505/breakage2.dat',header=T)
breakdata
attach(breakdata)
# attach lets us refer to the variable names as given
# by the header. Without this we would need to refer to
# the variable as, for example phdata$true
plot(transfers,number)           #plots measured versus true
lmout<-lm(number~transfers)      #fits simple linear regression
                                #model. Information save in phout
summary(lmout)

> breakdata
  number transfers
1      16         1
2       9         0
```

3	17	2
4	12	0
5	22	3
6	13	1
7	8	0
8	15	1
9	19	2
10	11	0

```
> summary(lmout)
Call:
lm(formula = number ~ transfers)
Residuals:
    Min       1Q   Median       3Q      Max
-2.2    -1.2     0.3     0.8     1.8
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.2000    0.6633   15.377 3.18e-07 ***
transfers     4.0000
Residual standard error: 1.483 on 8 degrees of freedom
Multiple R-squared:  0.9009,    Adjusted R-squared:  0.8885
F-statistic: 72.73 on 1 and 8 DF,  p-value: 2.749e-05
    0.4690    8.528 2.75e-05 ***
Signif. codes:  0 ***
```

SAS

```
option ls=80 nodate;
data a;
infile 'e:/s505/breakage.dat';      /* specifies file to read from */
input number transfers; /* names input variables */
proc print;
run;
proc reg;
model number=transfers;
plot number*transfers; /* this will produce scatterplot and
                        plot fitted line. Produces higher resolution
                        plot in graph window. */
run;
```

Obs	number	transfers
1	16	1
2	9	0
3	17	2
4	12	0
5	22	3
6	13	1
7	8	0
8	15	1
9	19	2
10	11	0

The REG Procedure  
Dependent Variable: number

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	160.00000	160.00000	72.73	<.0001
Error	8	17.60000	2.20000		
Corrected Total	9	177.60000			

The REG Procedure  
Model: MODEL1  
Dependent Variable: number

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	10.20000	0.66332	15.38	<.0001
transfers	1	4.00000	0.46904	8.53	<.0001

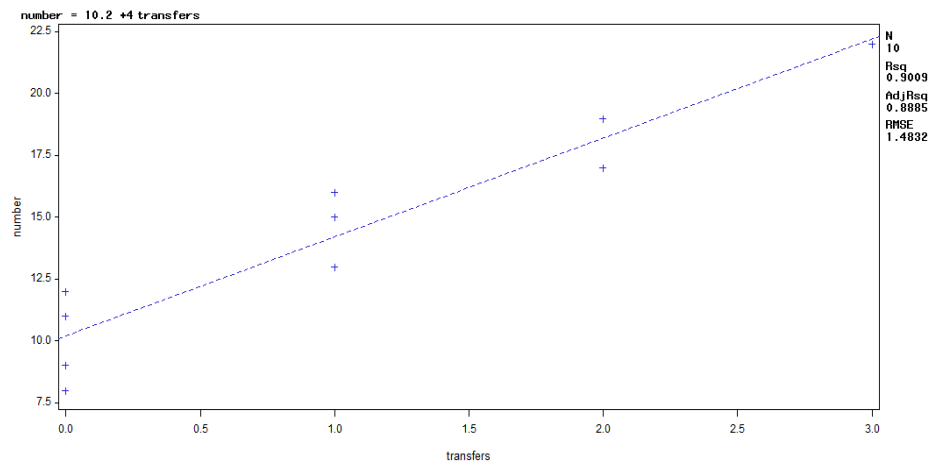


Figure 1: Plot of data and fit from SAS. R will be similar but without fitted line.