

Fall 2000 Course 513-697: Applied Linear Models  
Assignment 5 Answers and Comments

- 3.1 (1) Residuals are in original Y units whereas semi-studentized residuals are scaled so their SD is approx 1.
- (2)  $\epsilon$ 's are unobservable "true" deviations of Y from hypothetical or assumed unobservable mean  $E[Y | X] = \beta_0 + \beta_1 X$ , and thus  $E[\epsilon] = 0$ . The  $e$ 's are observed residuals measured about the observed line. Their mean is (by construction) zero.
- (3)  $\epsilon$  is the error term and  $e$  is the observed (constructed) residual.

3.2 See text.

3.9 The + + - - - + + pattern of residuals suggests a quadratic relation would be better than a simple linear one. I don't think of going to  $(X, X^2)$  as a "transformation" per se [since it adds terms to the model, rather than rescales terms] but some of you (and I think the textbook) see it that way.

3.19 [NOT asked this year .. quite tricky!] **Plot  $e$  vs  $\hat{Y}$  or  $e$  vs  $Y$  ???**  
 $e$  vs  $\hat{Y}$  won't show systematic tendencies if model fits;  $e$  vs  $Y$  will, even if fit is quite good.

$e$  vs  $Y$  will be a positive relation. I found it helpful to experiment with the Excel spreadsheet (EMS) to see what is happening before I worked out the following algebra.

**Theoretically**, to see this, without loss of generality (w.l.o.g.),

$$\text{take } \bar{X} = \bar{Y} = 0 \quad \Rightarrow b_0 = 0,$$

$$\text{and } SD(X) = SD(Y) = 1 \Rightarrow b_1 = \text{corr}[Y, X] = r \text{ and } SD[e] = \sqrt{1 - r^2} \quad SD(Y) = \sqrt{1 - r^2}$$

Then observed correlation of  $e$ 's and  $Y$ 's is

$$\begin{aligned} \text{corr}[e, Y] &= \text{ave}[e \times Y] / \{SD[e] \times SD[Y]\} \\ &= \text{ave}\{(Y - b_1 X) \times Y\} / SD[e] \\ &= \{\text{ave}[Y^2] - b_1 \text{ave}[X \times Y]\} / SD[e] = \{\text{ave}[Y^2] - b_1 r\} / SD[e] \\ &= \{\text{ave}[Y^2] - r^2\} / SD[e] \\ &= \{1 - r^2\} / SD[e] \\ &= \sqrt{1 - r^2} \end{aligned}$$

**Numerical Example ...** Y=GPA vs X=EntryScore

General Linear Models Procedure

Dependent Variable: GPA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.43372807	6.43372807	34.00	0.0001
Error	18	3.40627193	0.18923733		
Corrected Total	19	9.84000000			

R-Square	C.V.	Root MSE	GPA Mean
0.653834	17.40057	0.43501	2.50000

Pearson Correlation Coefficients

	GPA	RESIDUAL
GPA	1.00000	0.58836
RESIDUAL	<u>0.58836</u>	1.00000

**Check:**  $\sqrt{1 - r^2} = \sqrt{1 - 0.653834} = \sqrt{0.346166} = \underline{\underline{0.58836}}$

**Intuitively** ... Y has 2 parts (1) its (fitted) expectation,  $\hat{Y}$ , and (2) its (fitted) residual, e.  
 If we correlate e with Y, then Y still contains the e portion, i.e. Y and e have e in common.

**Extreme examples**; w.l.o.g., suppose  $\bar{Y} = 0$ .

(i) Suppose  $b_1 = 0$  (so  $r = 0$ ) so that  $Y = \hat{Y} + e = \bar{Y} + e = 0 + e = e$ ;

e and Y are perfectly positively correlated!

(ii) Suppose  $r=1$ ; then  $\hat{Y} = Y$ , and so  $Y = \hat{Y} + e = Y + 0$  i.e.,  $e = 0$  and so

Y and e are uncorrelated.

3.20 The (vertical) distribution of Y|X does not change just because we locate Y|X over some new  $X'$  = some transform of X. In contrast, transforming from Y|X to Y'|X where Y' is some function of Y, certainly will change the vertical distribution of these Y's. The only transformation of Y that leaves the distribution Gaussian is a linear transform.

3.21 Rather than reverse-engineer it, it is easier to start with each  $(Y_{ij} - \hat{Y}_{ij})$  and re-express it as

$$Y_{ij} - \hat{Y}_{ij} = (Y_{ij} - \bar{Y}_j) + (\bar{Y}_j - \hat{Y}_{ij}).$$

The square of this involves the squares of the two components and twice their crossproduct.

The task then reduces to showing that the sum (over i and j) of these crossproducts is zero.

The keys are that  $\hat{Y}_{ij}$  is the same for all i within j, and that within any j, the sum  $(Y_{ij} - \bar{Y}_j)$  is 0 by definition of a mean.

## 2 Dental Caries (Y) and Fluoride (X)

a Simple "Y vs X plot shows a curvilinear relationship, a "law of diminishing returns" with an asymptote (irreducible minimum) somewhere beyond 1.2 or 1.5 parts per million.

Residual (from linear fit) vs X (or predicted) shows the same thing:- a straight line underestimates Y at low and high X (positive residuals), and overestimates at middle X values (negative residuals). Some of you went to this plot first without ever plotting the Y vs X, and somehow thought it implied u-shaped relation -- implying increasing caries at very high F levels.

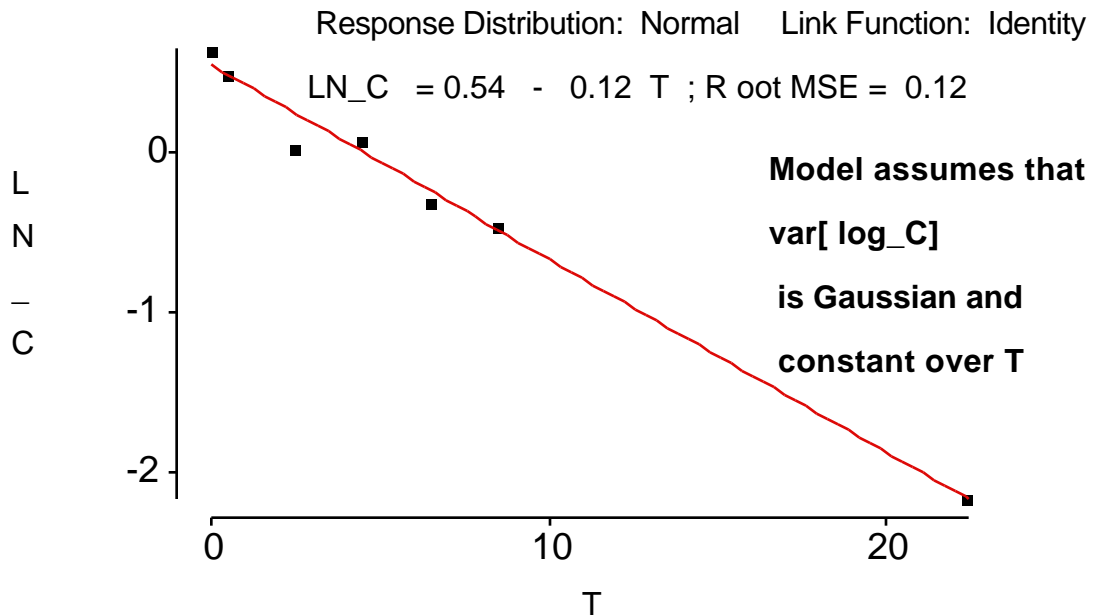
b Y vs 1/X or Y vs  $Y_0 \exp[-bX]$  suggest themselves. By the way, do you think of the latter as a transformation of the Y [to log] or of the X?

## 3 Caffeine clearance in smokers and non-smokers.

Most of you fitted  $\log\{\text{caffeine level}\}$  to time, using the model

$$E[\log\{\text{caffeine}\}] = \mu_0 + \mu_1 T; \quad (C = \text{Caffeine}; T = \text{time} - t_{\max})$$

$$\log[C] | T \sim \text{Gaussian}[\mu_0 + \mu_1 T, \sigma^2]$$



*In this model, the estimation procedure does not even know that the "Y" is already a log of something else, and it is the vertical variation in the logC that is assumed Gaussian and constant over T. i.e. the RMSE of 0.12 is in the logC scale.*

*You are essentially assuming that the C's are log-normal and that logC has same variance regardless of T.*

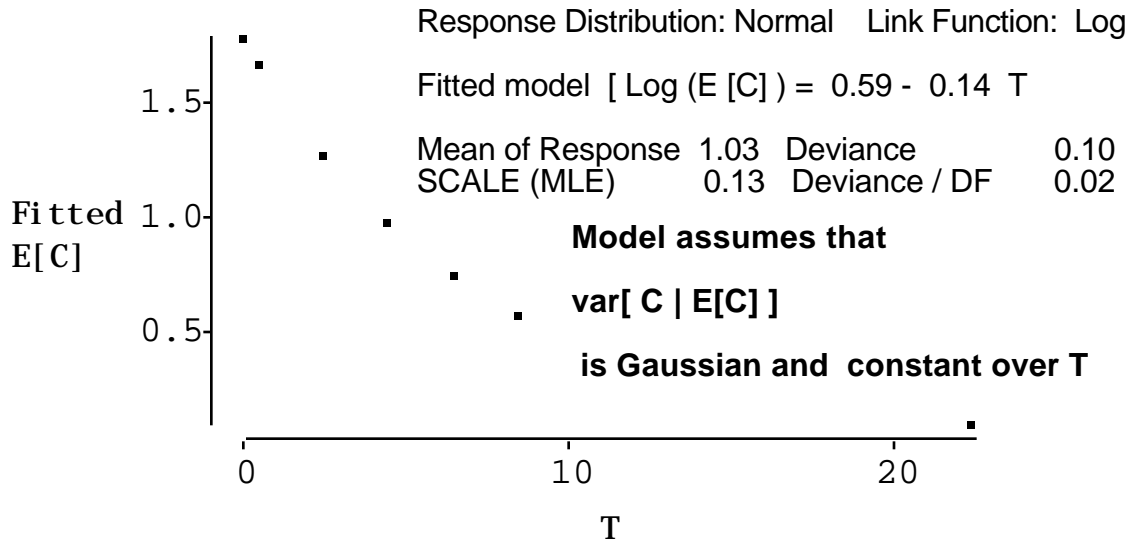
At least one student (last year) did something that at first sight might look the same, but is quite different conceptually

$$E[C] = \exp[\beta_0 + \beta_1 T];$$

i.e.  $\log[E[C]] = \beta_0 + \beta_1 T;$

$$C | \text{time} \sim \text{Gaussian}[\exp[\beta_0 + \beta_1 T], \sigma^2]$$

This latter model models the **log of the mean of C**, whereas the former models the **mean of the log of Y**. The latter is possible in this course by use of one of the generalized linear models.. ie log link and Gaussian variance.



If we took the  $\log[C]$ 's from linear model, equal variance assumptions in the former, and then "anti-logged" them, their means on the "regular" C scale would be curvilinear in T and their variances would be decreasing with T [constant relative errors applied to lower C levels at large T]

Half-life estimates obtained as  $\log[2]/b_1$  ; Thus, a  $b_1$  with high SE translates into a half-life estimate with higher SE. The highest SE's are from subjects where (i) spread of T was not as wide,(ii) fewer observations or (iii) more variation from fitted line (larger RMSE).