Fall 2000 Course 513-697: Applied Linear Models

Assignment 3 hand in on Monday September 18

---

1 NWNW4 **Problems 2.1, 2.2, 2.4*, 2.6**(parts a-d)* **2.9, 2.10, 2.11, 2.12, 2.13***

* `data in www.epi.mcgill.ca/hanley/c697/ ; to save time, program & output given below.`

2 Analysis of Rates of Fatal Crashes on rural interstate highways in New Mexico in the 5 years 1982-1986 (55 mph limit) and in 1987 (65 mph limit). Data from Oct. 27 article in JAMA by Gallaher et al. 1989;262:2243-2245.

	55 mph					65 mph
YEAR	1982	1983	1984	1985	1986	1987
Rate per 10^8 vehicle miles	2.8	2.0	2.1	1.7	1.9	2.9
N OF YEARS	5					1
MEAN(Rate)	2.100					2.9
VARIANCE(Rate)	0.175					0.0

The authors argued that it was inappropriate to compare the 1987 rate with the average of the 1982-1986 rates, since rates seem to have been falling over the 5 years. The authors first fitted a regression line to the rates for 5 years before the change, then predicted within what range the rate would be for 1987 if the downward trend continued. The following is output from Systat.

DEP VAR: Rate N:5 MULTIPLE R:0.794 MULTIPLE R^2: 0.630

STANDARD ERROR OF ESTIMATE: **0.294**

(This "STANDARD ERROR OF ESTIMATE" *is a misnomer; It is the square root of the average squared residual and might be called the "average residual")*

VARIABLE	COEFF.	STD ERROR	T	P(2 TAIL)
CONSTANT	**418.740**	**184.345**	2.272	0.108
YEAR	**-0.210**	**0.093**	-2.260	0.109

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
REGRESSION	0.441	1	0.441	5.108	0.109
RESIDUAL	0.259	3	0.086		

a **Interpret the fitted "constant" of 418.740. Why does it have such a large standard error? Rewrite the fitted model using a more appropriate "beginning of time" (don't worry about being Y2K compliant! you could even use the Microsoft definition of the "beginning of time").**

b **Interpret the -0.210 and its standard error 0.093 [for parts a and b use your parents in law as your intended readership]**

c **Scientists often interpret an absolute value of "b / SE(b)" of 2.0 or more as "P<0.05(2-sided)". Here b/SE(b) is -2.26, but P(2 tail) is 0.109!! Explain.**

d **Use equations 2.4 and 2.4a (p46) to quickly hand-calculate the b_1. What weights do the 5 different rates receive in the calculation? Why are these weights appropriate?**

e **Obtain the 5 fitted values and thus verify by hand that the 0.294 is in fact the square root of the "average" squared residual.**

3 Analysis of Rates of Fatal Crashes: **fill in the ______ 's and .... 's:**

- "fitted" (predicted) rate for 1987 = _______ - _______ × _______ = 1.47

  (slightly different from authors' because of rounding)

- range of variation for 1987 rate:

$$1.47 \pm t_{\_\_\_\_,95} \times \_\_\_\_ \times \sqrt{1 + \frac{1}{....} + \frac{[1987 - 1984]^2}{\Sigma_{year}[year - 1984]^2})} =$$

$$1.47 \pm \_\_\_\_ \times \_\_\_\_ \times \sqrt{1 + \frac{1}{....} + \frac{....}{....}} =$$

$$1.47 \pm \_\_\_\_ = 0.14 \text{ to } 2.80.$$

The observed value of 2.9 is just outside the 95% range of random variation predicted for 1987. In fact, using the SD of 1.45 [the 0.4205 obtained by multiplying the 0.294 by the radical, the 2.9 is t = (2.9 - 1.47)/ 0.4205 = 3.40 SD's above expected, and since the estimated SD is based on only 3 df, this deviate is somewhere between the 97.5% and the 99%ile. It is not clear whether the p-value in the article is 1- or 2-sided, or indeed whether the authors calculated it in the same way as here.

4 Blood Alcohol and Eye Movements:

www.epi.mcgill.ca/hanley/c678/ datasets: alcohol and smooth pursuit

Questions are at end of documentation file

SAS Program and output for NKNW Problem 2.4

```
DATA prob119;
INPUT gpa entrance;
LINES;
    3.1    5.5
    2.3    4.8
    3.0    4.7
    1.9    3.9
    2.5    4.5
    3.7    6.2
    3.4    6.0
    2.6    5.2
    2.8    4.7
    1.6    4.3
    2.0    4.9
    2.9    5.4
    2.3    5.0
    3.2    6.3
    1.8    4.6
    1.4    4.3
    2.0    5.0
    3.8    5.9
    2.2    4.1
    1.5    4.7
;
PROC MEANS;

PROC REG; MODEL gpa = entrance;
RUN;
```

```
 Variable   N          Mean       Std Dev       Minimum       Maximum
   -------------------------------------------------------------------
   GPA       20     2.5000000     0.7196490     1.4000000     3.8000000
   ENTRANCE  20     5.0000000     0.6928203     3.9000000     6.3000000
   -------------------------------------------------------------------

Dependent Variable: GPA
                           Analysis of Variance

                           Sum of         Mean
  Source          DF      Squares       Square      F Value       Prob>F

  Model            1      6.43373      6.43373       33.998       0.0001
  Error           18      3.40627      0.18924
  C Total         19      9.84000

      Root MSE       0.43501     R-square       0.6538
      Dep Mean       2.50000     Adj R-sq       0.6346
      C.V.          17.40057

                           Parameter Estimates

                    Parameter      Standard    T for H0:
   Variable  DF      Estimate         Error   Parameter=0    Prob > |T|

   INTERCEP   1     -1.699561    0.72677682        -2.338        0.0311
   ENTRANCE   1      0.839912    0.14404759         5.831        0.0001
```

NKNW Problem 2.6

```
DATA prob121;
INPUT broken transfer;
LINES;
   16.0    1.0
    9.0    0.0
   17.0    2.0
   12.0    0.0
   22.0    3.0
   13.0    1.0
    8.0    0.0
   15.0    1.0
   19.0    2.0
   11.0    0.0
;
proc means;

proc reg;
 model broken = transfer;

run;
```

```
 Variable   N          Mean       Std Dev       Minimum       Maximum
   -------------------------------------------------------------------
   BROKEN    10    14.2000000     4.4422217     8.0000000    22.0000000
   TRANSFER  10     1.0000000     1.0540926             0     3.0000000
   -------------------------------------------------------------------

Dependent Variable: BROKEN

                           Analysis of Variance

                           Sum of         Mean
  Source          DF      Squares       Square      F Value       Prob>F

  Model            1    160.00000    160.00000       72.727       0.0001
  Error            8     17.60000      2.20000
  C Total          9    177.60000

      Root MSE       1.48324     R-square       0.9009
      Dep Mean      14.20000     Adj R-sq       0.8885
      C.V.          10.44535

                            Parameter Estimates

                   Parameter      Standard    T for H0:
   Variable  DF     Estimate         Error   Parameter=0    Prob > |T|

   INTERCEP   1    10.200000    0.66332496        15.377        0.0001
   TRANSFER   1     4.000000    0.46904158         8.528        0.0001
```

1. Least Squares and the Combination of Observations

Adrien Marie Legendre (1752–1833)

THE METHOD of least squares was the dominant theme — the leitmotif — of nineteenth-century mathematical statistics. In several respects it was to statistics what the calculus had been to mathematics a century earlier. "Proofs" of the method gave direction to the development of statistical theory, handbooks explaining its use guided the application of the higher methods, and disputes on the priority of its discovery signaled the intellectual community's recognition of the method's value. Like the calculus of mathematics, this "calculus of observations" did not spring into existence without antecedents, and the exploration of its subtleties and potential took over a century. Throughout much of this time statistical methods were commonly referred to as "the combination of observations." This phrase captures a key ingredient of the method of least

from Stephen Stigler's book *History of Statistics*