Fall 2000 Course 513-697: Applied Linear Models

Assignment 5. Hand in by Friday September 29.

---

1 NWNW4 Problems 3.1, 3.2, 3.9, 3.20 pp 144-149

Two years ago, I assigned 3.3; the tasks in it over-emphasize formal statistical testing for assumptions that may not be critical -- one would not do all of these tests in everyday practice; nevertheless, read through them to see what is involved, but don't do any calculations.

2 The following data show the caries experience of children in 21 communities according to the (natural) fluoride concentration of their public water supply. DMF denotes "Decayed, Missing or Filled".

Dental caries experience of 7257 children
12-14 years of age in 21 communities.

Community	DMF Teeth per 100 children	Fluoride Concentration in ppm
1	236	1.9
2	246	2.6
3	252	1.8
4	258	1.2
5	281	1.2
6	303	1.2
7	323	1.3
8	343	0.9
9	412	0.6
10	444	0.5
11	556	0.4
12	652	0.3
13	673	0.0
14	703	0.2
15	706	0.1
16	722	0.0
17	733	0.2
18	772	0.1
19	810	0.0
20	823	0.1
21	1037	0.1

a Suppose public health officials asked you to fit a straight line to these data. They wish to use the fitted line to choose the amount of fluoride to add to the water supply.

Is a straight line fit is a good summary of the pattern in the observed data?

b Suggest (and fit) other forms for the observed relationship.

If you wish to use computer rather than a calculator, the data are under "caries data" in `www.epi.mcgill.ca/hanley/c697/`

3 Is caffeine "cleared" faster from smokers than non-smokers?

Data and documentation are in `http://www.epi.mcgill.ca/hanley/c622/`

As a prelude to answering this question, we will reduce the relevant time-concentration data on each subject to an estimate of the caffeine "halflife" in the subject (ss) in question.

• For each of your 6 subjects, plot out their concentration curves. First try to estimate *by eye* as best you can when you think caffeine concentration has "peaked" [call this c_{max}] and then fit a curve to the concentrations from that timepoint on.

Pharmacologists tell us that caffeine is eliminated from c_{max} onwards according to a negative exponential "decay" curve i.e.,

$$c[t] = c_{max} \times \exp[-\beta t]$$

where t is the time since the peak was reached.

[*they tell us that alcohol is eliminated according to a linear decay curve.]*

From β you can get the halflife $t_{1/2}$ -- the time by which concentration has become half what it was, as

$$\text{halflife} = (-\ln 0.5) / \beta \ .$$

[This is also how one calculates median survival if survival times follow a negative exponential distribution]

• For each subject, treat both c_{max} and β as parameters to be estimated

The data are in the caffeine data (main) file, and the sas program to read them is in the first half of the sas file.

You can have SAS fit the regression for subject xx for time > t by using the "`where`" statement

e.g.

```
proc reg;
model  y = x;
where ( subject = xx and time > t);
```

• Graphically check the fit for each subject and comment on any worrysome fits/subjects/datapoints.