Fall 2000 Course 513-697: Applied Linear Models

Assignment 7. Hand in by Friday October 13.

---

1 NWNW4 Problems 4.4 part a, 4.11, 4.12, 4.13 Exercises 4.21, 4.24, 4.25

2 ***[Optional]*** Effect of Measurement Errors in X on estimate of slope (and intercept) #

Use as "X" the 1986, and "Y" the 1999 ages of the n=40 epidemiology students referred to in the Notes for Chapter 4.

Distort each X by randomly adding or subtracting 5 years to/from it: i.e. create X* = X +/- 5. Use a different "seed" from that in the SAS program shown in the Notes/www.

- Fit a regression $E[\,Y \mid X] = \beta_0^* + \beta_1^*\ X^*$; what are your estimates of β_0^* and β_1^* ?
- Keeping the same Y's and Xs, repeat the allocation of errors a large number of (say 1000) times; what is the behavior of the estimates <u>on average</u>, rather than in your 1st dataset?
- What happens if the range of X is bigger/smaller? if the size of the errors in X* is bigger/smaller ?
- Does the attenuation formula given in the Notes give accurate predictions as to the bias?

3 Effect of Measurement Errors in X and/or Y variable on estimate of slope (and intercept) #

Imagine you are trying to find out, from imperfect observations of F and C, what the two coefficients β_0 and β_1 are in the temperature relation $F = \beta_0 + \beta_1 C$.

For each of the following situations a-e, and using the true values of $\beta_0 = 32$ and $\beta_1 = 1.8$, simulate 1000 datasets and investigate the behaviour of 1000 estimates, b_0 and b_1, of β_0 and β_1. In each simulation, use samples of size 4, with temperatures of C = 14, 16,18 and 20.

a C measured perfectly, F measured with independent Gaussian(0, σ=1) errors. Check for evidence of bias in b_1 and check whether the empirical variance of b_1 lines up with the theoretical formula for var(b_1).

b C measured perfectly, F measured with independent Gaussian(0, σ=3) errors Again, check for evidence of bias and compare the empirical variance of b_1 with the theoretical formulae for var(b_1).

c F measured perfectly, C measured with independent Gaussian(0, σ=1) errors [Berkson type]
[imagine you are using an imperfect instrument that "tells" you when the temperature is 14C, 16C, etc and that the instrument doesn't make the same error each time]

d F measured perfectly, C measured with independent Gaussian(0, σ=3) errors [Classical type]
imagine <u>someone else chose situations</u> when C was exactly 14, 16, etc, but didn't tell you what C was, and instead asked you to independently record C using an imperfect instrument] and to use <u>your</u> recordings of C in your estimation of the line

In c-e, do your findings line up with the predictions in the Notes? If the patterns are difficult to see, you might change the number of simulations, the sizes of the errors, the range of C or the sample size.

\# The skeleton of SAS programs for the simulations in 2 and 3 can be found in:
`www.epi.mcgill.ca/hanley/c697/`

4 Question # 4, "Using Multiple Linear Regression: with more information, can one better predict Old Faithful?" from Homework due June 12, 200 in Course 678