Fall 2000 Course 513-697: Applied Linear Models

Assignment 2. Hand in at start of class on Wednesday September 13

---

1 NWNW4 Exercises 1.39, 1.40, 1.41 textbook pages 41-42 .

2 *(designed along the lines of exercise 1.42, but aimed to encourage you not to unquestioningly accept the assumptions just because the textbook makes them!)*

You are interested in estimating μ = the average weight of blank sheets of paper of a "standard" size and weight. You have just 2 "datapoints". One of these comes from a single sheet of paper weighed on its own, the other from two <u>other</u> sheets [without staples!] weighed together.

	x	y
datapoint	# sheets weighed	total weight of the x sheets[s]
1	1	145
2	2	302

a Using only the math tools learned in elementary school, and before going on to part b, make an estimate of μ. Explain the basis for this estimate. Can you give a rough idea of the "inexactness" of your estimate?

*Other students may arrive at a different estimate from you -- not because of arithmetic errors, but because they used a different method [a different estimat**<u>or</u>**]*

b Besides the difference in x, two possible explanations for the variations in y are (i) the sheets of paper are of uniform weight, but the measuring instrument, while corrected calibrated, cannot reproduce the same result from weighing to weighing, instead giving readings that fluctuate around the true value with a standard deviation of σ (ii) the instrument is perfect calibrated and produces perfectly(!!) reproducible measurements, but because of fluctuations during manufacture, there is slight random variation [standard deviation = σ, of the same magnitude as in (i)] in the weights of individual sheets.

Carefully write out separate statistical models for the observed y's under (i) and (ii). They may look somewhat like the models in exercises 1.41 and 1.42 but you need to pay particular attention to the variance of ε_1 and that of ε_2 !!

c [not necessarily obvious]

If (i) is correct, what is your estimate of μ.? If (ii) is correct, what is your estimate ? .

Why should (i) or (ii) influence the estimation? Explain your reasoning [no formal models or calculations required at this stage!].

d Which model for the error [(i) or (ii)] meets the specifications of model 1.1 on page 10?

e Do you think the variance part of the model suggested in exercise 1.42 is realistic? Explain your reasoning.

3 *(challenging but instructive as to degrees of freedom really are)*

Fill in the missing values in the 6 situations given in section 5-6 "SSE & the estimator of the (common) X-specific variation of E" of the "Chapter 5" notes for course 678 version given in 1999.

Course 678 web page->1999 Course Material -link -> "Chapter 5" notes

Fall 2000 Course 513-697: Applied Linear Models

Assignment 2. Hand in at start of class on Wednesday September 13

---

4 [**OPTIONAL**] Blood Alcohol and Eye Movements

`www.epi.mcgill.ca/hanley/c678/ -> Datasets: alcohol & smooth pursuit data`

Questions are at end of documentation file.

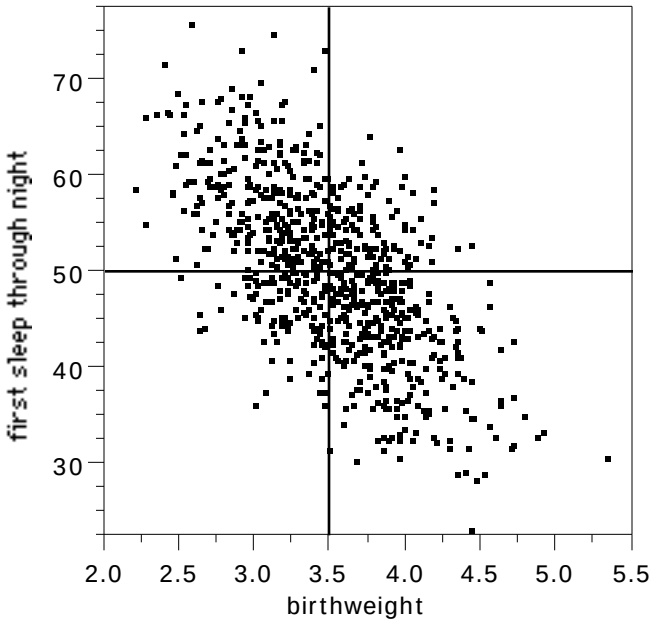
5 [**OPTIONAL]** Refer to the hurricane data used to discuss "Equation 5.7 -- CENTERED VERSION" in course 678 web page->1999 Course Material -link -> "Chapter 5" notes.

a Replicate the fitted regression coefficients, using uncentered "X's".

b Round to 1 decimal the fitted intercept and slope of the uncentered model and project "forward" from the rounded intercept to the 1980-1989 decade. Compare this fitted value with that obtained by the same amount of rounding applied to coefficients fitted to a centered model, and to the fit with that obtained with no rounding of the coefficients of the uncentered model. Comment.

c Although the text shows that the last decade is "(1990-1999)" the data (Y=2) are only for the 5-years 1990-1994. Describe some ways to incorporate this "complication' into the analysis.

6 [**OPTIONAL]** A study of babies examined the relationship between their weight at birth and the age at which they first slept through the night. The birth weights averaged 3.5 Kg with an SD of 0.5 Kg. The ages at which they first slept all night averaged 50 days with an SD of 10 days [JH's consultants believe that the mean and SD for this latter variable are much larger!]. The correlation between the two variables was –0.60. Here is a scatterplot of the raw data.



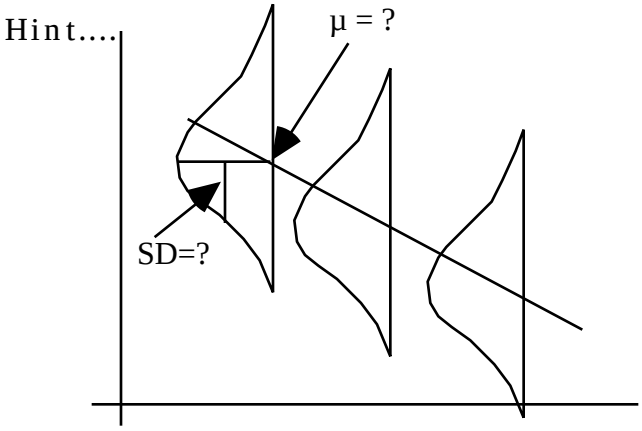
a Draw in the regression line. Hint: use the "centered" form: $\mu_{y|x} = \mu_y + \beta (x - \mu_x)$
[also, you may wish to use (general???) relationship 15.10b on page 638, or look at page 7 of "Notes on M&M Chapters 2 and 9" under 1999 material Chapter 5 of www.epi.mcgill.ca/hanley/c678/]

b If the birthweights were in grams rather than Kg, what would β be? What would the correlation be? Likewise, if the age was measured in weeks, what would change?

Assignment 2. Hand in at start of class on Wednesday September 13

---

c If we consider a baby that weighed 2.5 Kg at birth, what is the probability that it will sleep through the night before it is 10 weeks (70 days) old? before it is 10 weeks old? You don't need to DO the calculation, just indicate HOW to. What distributional assumptions do you have to make?



d If we consider all babies that weigh 2.5 Kg at birth, what is the probability that the average age at which they will first sleep through the night is less than 70 days? less than 75 days?