

Midterm 01 - Solutions

Linear Models – Fall 2015

2015-10-12

I.

1. (4 pts) After accounting for height, males are on average 4.6336 cm taller than women. A 95.44% confidence interval is given by $4.6336 \pm 2 \cdot 2.1478$.

3/4 if you did not correctly describe what the estimate actually meant in a non-mathematical context

2. (6 pts) The first 4 rows of out3's model matrix are:

$$\begin{pmatrix} 1 & 161 & 0 \\ 1 & 0 & 170 \\ 1 & 0 & 174 \\ 1 & 173 & 0 \end{pmatrix}$$

The first four rows of out4's model matrix are:

$$\begin{pmatrix} 1 & 161 & 0 & 0 \\ 1 & 170 & 1 & 170 \\ 1 & 174 & 1 & 174 \\ 1 & 173 & 0 & 0 \end{pmatrix}$$

4/6 if you instead wrote out $[X \cdot \text{diag}(\hat{\beta}_{OLS})]$, an odd thing to do but a fairly common error; 5/6 if you had it without an intercept

3. (4 pts) The estimate for height:genderF gives the increase in one unit of weight expected with an increase in one unit of height for females. The height:genderM is the same for the males. We do not have a formal hypothesis test showing whether these are equal, but each are within a standard error of one another so it seems very unlikely that a formal test would indicate that they are statistically significant (they aren't).

3/4 if you failed to see why the difference is not significant; 3/4 if you incorrectly thought that one of these was a baseline value

4. (4 pts) First of all, height:genderM has a different meaning in model out4 than in out3. In out4 it is the offset from the baseline, whereas in out3 it is the total effect for males. Additionally, these parameters may not be significant because while there is evidence that gender plays a role in the regression it is not clear if this is a different slope or intercept. When both are in the same model, there is not enough evidence to say that one or the other is the important factor.

full credit for either explanation

5. (4 pts) The F-test statistic is given by:

$$F = \frac{(SSR_R - SSR_U)/df_1}{(SSR_U)/df_2}$$
$$F = \frac{(8.259^2 \cdot 98 - 8.151^2 \cdot 96)/2}{(8.151^2 \cdot 96)/96}$$
$$F = 0.769$$

2/4 if you set up the F test but did not know how to find the sum of squared residuals; 4/4 if it was mostly correct with some sort of typo or minor error

II.

1. (8 pts) The ‘best’ answers are below, but I accepted any reasonable explanation of what assumptions were broken. As long as your conclusion of unbiased/biased matched the violations you stated (and they were reasonable) I also gave credit for that part as well.

(a) The spherical errors assumption is broken; specifically they appear to be heteroskedastic. Spherical errors are not needed for the OLS estimator to be unbiased, however, so the estimate will be unbiased. (In the non-linearity case it isn’t even clear what unbiased would mean given that there is no β .)

(b) The signal in the errors indicates that the true model is not linear. It would also be correct to say that the errors are endogenous. In either case the estimator should be unbiased.

(c) The outlying points indicate that the errors are non-normal (I simulated them from a Cauchy). It would also be correct to say that the spherical errors assumption is violated, with the assumption that the outlier is not a ‘heavy tail’ but just a point with a much higher variance. In either case, the estimator should be unbiased.

(d) The point all the way to the right has high leverage, but this is not a violation of the model assumptions. Therefore the estimator should be unbiased.

essentially one point per assumption and biased/unbiased; graded very leniently as technically many answers could be justified

2. (4 pts) Both of these matrices are projections into a linear subspace, P onto the column space of X and M onto the space perpendicular to the column space of X . Once we have projected a vector into a linear space, projecting again should not change the result; hence $P^2 = P$ and $M^2 = M$.

full credit for anything describing these as projections into spaces; 2/4 for a proof that these are true without any context

3. (4 pts) The matrix P is a symmetric idempotent matrix and therefore $u^t P u$ has a chi-squared distribution. The degrees of freedom is given by the rank, or trace, of the matrix P , which we know to be equal to p .

3/4 if you had everything except for the rank correct

4. (4 pts) Strict exogeneity says that the expected value of the errors are zero even when conditioned on all values of X ; in other words: $\mathbb{E}(\epsilon_i|X) = 0$. The weak form only requires conditioning on the data associated with an observation, so that $\mathbb{E}(\epsilon_i|x_i) = 0$. An example of a set-up that satisfies weak but not strong exogeneity was given in problem set 3:

$$Y_t = \beta Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim_{iid} \mathcal{N}(0, \sigma^2)$$

Clearly it satisfied weak exogeneity, almost by construction, but for example:

$$\begin{aligned} \mathbb{E}(\epsilon_1|X_2) &= \mathbb{E}(\epsilon_1|Y_1) \\ &= Y_1 \end{aligned}$$

Note: I'm providing this detailed explanation here to help explain what is going on; you did not need to do this for the exam.

3/4 if you had everything except for the rank correct; 2/4 if you confused zero means $\mathbb{E}\epsilon = 0$ with weak exogeneity

5. (6 pts) Prediction intervals provide a region in which a new observation, taken from an independent sample from the same generating model, will fall with some fixed probability. Confidence intervals instead give a region where the means of the new observations are expected to be; in other words, it bound $X\beta$ rather than $X\beta + \epsilon$.

5/6 if you explained the difference correctly but swapped the two definitions; 3/6 if you said one is a factor of s larger but failed to explain the meaning of this

6. (2 pts) The Gauss-Markov theorem says that the ordinary least squares estimator has the smallest variance amongst all linear, unbiased estimators. It is stronger than Cramer-Rao because it does not assume normality. Many people answer that Gauss-Markov is stronger because it also includes the fact that the estimator is unbiased and linear ...this is not really the spirit of the question but given that it does in fact answer the question I gave credit for that as well.

1/2 if you correctly explained the Gauss-Markov theorem but failed to relate that to the Cramer-Rao bound