Fall 2000 Course 513-697: Applied Linear Models

Assignment 4 Answers and Comments

2.17 **Alpha** (α) is the pre-set level. Here the P-value must have been smaller than α. If α had been preset to 0.01, then the P-value of 0.033 would have exceeded this α level.

2.18 A **t-statistic** is more versatile than its square **F**=t^2, because the sign of the difference is not lost in the squaring, and one can test **1-sided** hypotheses.

2.19 **Issue of 1- and 2-sided hypotheses and 1- and 2-tailed p-values**
A good idea to use say "sides" for describing H_a but maybe a different word. e.g., "tails" for the P-value. Otherwise it can get confusing, as when we use 1 (the upper) tail of the F distribution to get the P-value in connection with a 2-sided alternative $\beta \neq 0$. I am not sure that all other authors agree with calling the F-test a "1-sided" test. You might want to check around in various textbooks to see what they say!

Easiest way to see why F Ratio is a 1-sided test is to note that

(i) under H_0: [$\beta_1 = 0$]

distr'ns of MSR and MSE ***both*** center on σ^2

mean of ratio distr'n is approx 1 (Central F distribution has a longer right tail .. skewed)

(ii) under H_0: [$\beta_1 \neq 0$]

distr'ns of MSR and MSE center on $\sigma^2 + \beta^2 \times$ (+ve fn. of x's) and σ^2 , respectively

mean of ratio distr'n is ***further to right of 1*** (Non-Central F distribution)

2.20 In the situation where SD(Y) = SD(X), then r = b. Otherwise, b is a function of r and the SD of the observed Y's. So, r (and thus r^2) is not a point estimator of any single parameter.

2.22 First 10 observations gave r=0; could r from full 30 observations be non-zero? Yes, r could go either way. The fact that r=0 doesn't even mean that the 10 points all fall on a line: it could be that the relationship is u-shaped! Likewise, the 1st 10 could give a non-zero r, but untimately the r from the 30 could be zero - it would depend on the X order in which one made the Y observations if say one had a u-shaped relationship.

2.23 a Anova Table

Source	SS	df	Mean Square	F	P
Regression	6.43	1	6.43	34	0.000016
Residual	3.41	18	0.19		
Total	9.84	19			

b What is estimated by $MS_{resgression}$ and $MS_{residual}$? Some of you said that both are estimates of some variance parameter. Yes, $MS_{residual}$ is an estimate of σ^2. But $MS_{resgression}$ is an estimate of σ^2 **only if** $\beta_1 \neq 0$. Otherwise MS is an estimate of a sum of 2 parameters, $\sigma^2 + \beta_1^2\,\Sigma(x - \text{xbar})^2$.

Get used to **translating**

$E[MSE] = \sigma^2$ and

$E[MSR] = \sigma^2 + \beta_1^2 \Sigma(X - xbar)^2$

as

"MSE is an unbiased estimator of" σ^2 ;

"MSR is an unbiased estimator of" $\sigma^2 + \beta_1^2 \Sigma(X - xbar)^2$

and don't mix up statistics and parameters!!

c Test of β_1. I prefer to say "reject H_0" or "do not reject H_0" than to "accept" H_a or "accept" H_0.

f Question on r or r^2, which has the clearer operational interpretation? I am between two minds here. One one level, I prefer r because I like the link between r and b, and I don't like thinking about variance, which is in squared units. Imagine measuring the variability in fertility. The units for average fertility are number of children per woman. The units for the "variance" of fertility are "number of square children per square woman", while the standard deviation is in the same everyday units as the mean. On another level, r^2 has a cleaner meaning as a simple proportion... but what if I had to explain it to my in-laws...?

2.40 Non-uniformity of the error variance is a second-order issue here; one can still speak of $\beta_1 \neq 0$ as being evidence for *(at least)* a linear relation, and vice versa.

2.42 $E[MSR] = \sigma^2 + \beta_1^2 \Sigma(X - xbar)^2$. Thus, having a larger $\Sigma(X - xbar)^2$, either through spreading out the X's or having more of them (i.e., larger n) will --- if $\beta_1 \neq 0$ --- make for a larger MSR <u>on average</u> Remember that there is still random variation at play and that in any one realization, the observed MSE and MSR can be opposite from what is expected on the average. A larger $\Sigma(X - xbar)^2$ **increases the probability of a large MSR**, and so the probability of a statistically significant F value. So, it increases the power of the test of whether β_1 is zero.

Another way to look at it is to assess impact on $var[b_1] = \sigma^2 / \Sigma(X - xbar)^2$. Again, having a larger $\Sigma(X - xbar)^2$ decreases $var[b_1]$ and $SE[b_1]$.

If issue is precision with respect to the estimate of E[Y | X=8], rather than with respect the estimate of β_1 , then one has two design choices: study X's near or at 8, or X's that are more spread out.

The term $(x - xbar)^2 / \Sigma(X - xbar)^2$ in the variance of the estimate of E[Y | X=8] vanishes if we take all the measurements at x=xbar, and we are left with the usual σ^2 / n -- the same as for var[ybar].

If we take x's that are spread out from 8, we get the same var for our estimate, ***BUT AT A PRICE***: we have to assume the linear model is correct in the range of X's studied.

This was the same issue in the study of impairment at alcohol=legal limit. If we are interested in what happens at x=legal limit, why not make all our y observations on a person there, at x=0.08? If -- for logistic reasons -- we cannot, and have to "straddle" the limit, then in effect we are using a model to interpolate to the legal limit. If the model is good, then observations away from x=0.08 are almost as good as those at x=0.08.

This raises the question: assuming the model is correct, do we <u>lose</u> <u>anything</u> by straddling x=0.08 rather than making all our measurements there? I can think of one loss: it has to do with the price (df) of estimating a mean (all x's=0.08) versus estimating a line from which we interpolate to x=0.08. Any other inefficiency?

42a When testing H_0: $\beta_1 = 5$ vs. H_a: $\beta_1 \neq 5$ by means of a general linear test, the reduced model is

$$E[Y \mid X] = \beta_0 + 5X, \qquad \text{or} \qquad E[Y - 5X \mid X] = \beta_0 ;$$

so we use 1 df to fit β_0 -- using the average of the n observed (Y - 5X)'s -- leaving the n-1 independent residuals to estimate the variance of the ε's

The full model requires that we estimate both β_0 and β_1 , thereby using 2 degrees of freedom, and leaving n-2 for variance estimation.

42b When testing H_0: $\beta_0 = 2$ AND $\beta_1 = 5$ versus the alternative $\beta_0 \neq 2$ OR $\beta_1 = 5$ (or both!), the reduced model is

$$E[Y \mid X] = 2 + 5X, \qquad \text{or} \qquad E[Y - \{2 + 5X\} \mid X] = 0 ;$$

so we don't require any df to fit the systematicpart of the model, leaving us with n independent residuals [from 0] to estimate the variance of the ε's

Again, the full model requires that we estimate both β_0 and β_1 , thereby using 2 degrees of freedom, and leaving n-2 for variance estimation.