Econ 120B, Spring 2003 --- Ramu Ramanathan --- Answers to Homework #2

a)

Inspecting the signs in Model 1 – are they what you have expected?

	VARIABLE		COEFFICIENT SIGNS		Expected Signs
2)	home	0.405549	(+)		(+) .
3)	inst	-0.526420	(-)		(-)
4)	svc	2.03873	(+) Wrong sign	(-)
5)	tv	0.756508	(+)		(+)
6)	age	1.19351	(+)		(+)
7)	air	-5.11114	(-)		(-)
8)	y	0.00165517	(+)		(+)

[If an alternative explanation provided by you is sensible, you will get full credit.]

Home (+) : we would expect a higher number of subscribers of each system if we have more homes passed by each system, holding all else constant.

Inst (-) : we would expect a higher installation fee to have negative partial effect on the number of subscribers of each system, holding all else constant.

Svc (-): we would expect higher monthly service charge to have a negative partial effect on the number of subscribers of each system, holding all else constant – but the coefficient has the wrong sign.

TV (+): we would expect the more television signals carried by each cable system, the more subscribers of each system, holding all else constant.

Age (+): we would expect that the longer the system has been installed, the greater the market share it has earned, this should have a positive partial effect on its number of subscribers, holding all else constant.

Air (-): everything else equal, we would expect that the more people can free ride the less subscribers would be for each system.

Y (+): If we believe that the cable television service is a normal good, we would expect that the higher the per capita income for each television market with cable, the more subscribers would be for each system.

b)   Testing individual coefficients for significance at 5% level in Model 1 (note we need the normality of errors assumption to carry out the following hypothesis testing)

We perform seven two-tailed t-tests for each i, i=(2,3 …8), at 5% level:

$H_0 : \beta_i = 0$
$H_1 : \beta_i \neq 0$

Each t-statistic, $t_c = \beta_i$ hat $/ S_{\beta i \text{ hat}}$, has the t-distribution with 32 degrees of freedom under the null hypothesis. The column T STAT in the Gretl output for Model 1 contains all seven t-statistics of interest. (Check for yourself using the values in the column COEFFICIENT divide by the values in the column STDERROR.)

Choose the significance level to be 5%, compute the critical value $t_{32}*$ (0.025)

Look up in the t-table corresponding to 32 d.f. and find $t_{32}*$ (0.025) such that the area to the right of it is one half the level of significance 5%.

$t_{32}*$ (0.025) is between 2.021 and 2.042. If any t-statistic is greater than 2.042 in absolute value, we would reject the corresponding null hypothesis that the coefficient is zero at 5% level.

We tabled the results:I   *[Using p-values instead is also acceptable.]*

	VARIABLE	COEFFICIENT	T STAT	Reject/Do not reject H_0 at 5%
2)	home	β_2 hat = 0.405549	11.586	Reject .
3)	inst	β_3 hat = -0.526420	-1.106(-)	Fail to reject
4)	svc	β_4 hat = 2.03873	0.959	Fail to reject
5)	tv	β_5 hat = 0.756508	1.100(+)	Fail to reject
6)	age	β_6 hat = 1.19351	2.374	Reject
7)	air	β_7 hat = -5.11114	-3.366	Reject
8)	y	β_8 hat = 0.00165517	0.477	Fail to reject

Y, SVC, INST, and TV are initial candidates for omission from the model because there is no evidence that they have significant partial effects on SUB. Note that the p-values for β_3 hat, β_4 hat, β_5 hat, and β_8 hat are all beyond 0.25. You should be alert that a Type I error of 25% and beyond is unacceptably high.

c)

First, we omitted variable Y, it is the variable with the least significant regression coefficient, i.e., it has the highest p-value in the model estimates excluding the constant. On average, β_8 hat is expected to be closest to zero, so any bias caused by omitting the variable Y is likely to be small.

In Model 2, after dropping the variable Y, all of the model selection criteria have improved, i.e. decreased. Also the omission of Y has improved the precision of the remaining coefficients. In Model 1 and Model 2, you should compare the values in the column STDERROR. We see that some of the coefficients are more significant (by looking at their p-values), e.g. the constant, AGE, SVC and AIR.

The variable with the highest p-value in model 2 is now TV. We drop TV and obtain model 3. In model 3, we drop SVC (it has the highest p-value in model 3). In model 4, we drop INST. Note that based on a process known as data-based model simplification, we omitted Y, TV, SVC, and INST one at a time. In model 5, all the remaining regression coefficients are significant at 10% or lower level.

d)

Model 5 is the best model: *[Using R bar square to judge a model is inadequate. You need to use all 8 Model selection statistics.]*

Out of the eight model selection statistics, five of the statistics have the lowest value. And all the coefficients in model 5, excluding the constant, are extremely significant, even at 1% level. Also, the coefficient for HOME, AGE and AIR are not very different among Model 5, Model 4 and Model 3, thus the bias in omitting SVC and INST might not be serious.

e)

Test the best model (Model 5) for overall significance at 1% level

H_0: $\beta_2 = \beta_6 = \beta_7 = 0$ *[Note: in the multiple regression model it is not*
H_1:   At least one of them is not zero   *correct to say rho xy = 0 as the null hypothesis]*

Construct the test statistic:

$$F_c = \frac{R^2 / (k-1)}{(1-R^{2)}/(n-k)} = \frac{0.872436 / (4-1)}{(1-0.872436)/(40-4)} = 82.08$$

F_c has the F distribution with d.f. 3 and 36 under the null hypothesis.

Construct the critical value:

$F *_{(3,36)} (0.01)$ is between 4.31 and 4.51.

Since the F-statistic is far greater than 4.51, we reject the corresponding null hypothesis that $\beta_2 = \beta_6 = \beta_7 = 0$.

f)

Testing individual coefficients for significance at 1% level in Model 5

We perform three two-tailed t-tests at 1% level:

$H_0 : \beta_i = 0$ *[You are test beta not beta hat]*
$H_1 : \beta_i \neq 0$ (i = 2, 6, 7)

Each t-statistic, $t_c = \beta_i$ hat $/ S_{\beta i \, hat}$, has the t-distribution with 36 degrees of freedom under the null hypothesis. The column T STAT in the Gretl output for Model 5 contains all three t-statistics of interest.

Choose the significance level to be 1%, compute the critical value $t_{36}* (0.005)$

$t_{36}*$ (0.005) is between 2.704 and 2.750. If any t-statistic is greater than 2.750 in absolute value, we would reject the corresponding null hypothesis that the coefficient is zero at 1% level.

We tabled the results:

	VARIABLE	COEFFICIENT	T STAT	Reject/Do not reject H_0 at 1%
2)	home	β_2 hat = 0.411502	13.290	Reject (significantly different from 0)
6)	age	β_6 hat = 1.13974	2.784	Reject (significantly different from 0)
7)	air	β_7 hat = -3.46183	-3.336	Reject (significantly different from 0)

g)

Model 1 is the unrestricted model
Model 5 is the restricted model

H_0: $\beta_3 = \beta_4 = \beta_5 = \beta_8 = 0$ (i.e. 4 restrictions)
H_1: At least one of them is not zero

The test for this joint hypothesis is known as the Wald test, that is, the null hypothesis for this Wald F-test is that the coefficients of all the omitted variables are zero.

h)

$$F_c = \frac{(SSR_R - SSR_U) / (k - m)}{SSR_U / (n - k)} = \frac{(5595.61 - 4923.91) / (8 - 4)}{4923.91 / (40 - 8)} = 1.091328$$

Under the null hypothesis, this test statistic F_c has the F distribution of d.f. 4 and 32.

The critical value:

$F*_{(4,32)}$ (0.1) is between 2.09 and 2.14.

The null hypothesis cannot be rejected at 10% level because F_c is less than the F* value. We can conclude that the coefficients for Y, TV, SVC and INST are jointly insignificant at 10% level.

i)   Use model 6 to generate the forecast of sub when home=58, age=9 and air=6. Use that and the corresponding std. error to construct a 95% C.I. for the forecast.

From model 6 output, read off the coefficient estimate for the constant, we have b_1 *hat* equals 26.2231. This is our SUB_f *hat*. The standard error for b_1 *hat* is the standard error for SUB_f *hat*, $S_{subf\,hat}$.
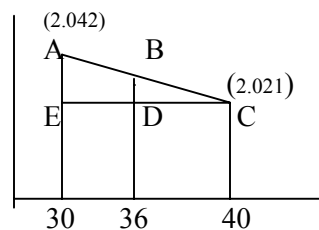
.
95% C.I. for the forecast:

$P\ (SUB_f\,hat - t*S_{subf\,hat} \leq SUB_f \leq SUB_f\,hat + t*S_{subf\,hat}) = 0.95$

$SUB_f \quad [SUB_f\,hat - t*S_{subf\,hat}\ ,\ SUB_f\,hat + t*S_{subf\,hat}]$

$t_{36}*\ (0.025)$ is between 2.021 and 2.042, but for t*, we need to interpolate the actual value needed.

We will use the properties of congruent triangles:

We know that $\triangle$ CBD and $\triangle$ CAE are congruent and hence BD/AE = DC/EC.

BD/0.021 = 4/10 which gives BD = 0.0084

So $t_{36}*\ (0.025) = 2.021 + 0.0084 = 2.0294$

So the 95% C.I. for the forecast is:

[*26.2231 – 2.0294*2.02345, 26.2231 + 2.0294*2.02345*]

i.e. [22.1167, 30.3295]