ST505: Fall 2012 Homework 8 solution.

Part i: Problem 8.15

a) If $X_2 =$ type (=1 for small and 0 for large) then the model in (8.33) implies

$E(Y|X_1) = (\beta_0 + \beta_2) + \beta_1 X_1$ for small copiers and

$E(Y|X_1) = \beta_0 + \beta_1 X_1$ for large copiers.

So, there is a common slope β_1 for each type, with intercept β_0 for large copiers and intercept $\beta_0 + \beta_2$ for small copiers. The model here corresponds to two parallel lines for E(Y), one for small and one for large.

The parameter β_2 is the difference between the intercepts and so represents the copier effect (which with no interaction is the same at each X_1).

b) Here is the SAS output. You obviously get the same from R.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.92247	3.09969	-0.30	0.7675
number	1	15.04614	0.49000	30.71	<.0001
type	1	0.75872	2.77986	0.27	0.7862

Variable	DF	95% Confidence Limits	
Intercept	1	-7.17789	5.33294
number	1	14.05728	16.03500
type	1	-4.85125	6.36870

c) Note that .7582 is the estimate of the difference in intercepts for the two groups with a 95% CI of $[-4.85125, 6.3687]$.

d) If we had an overall random sample then you could compare copier types directly by just doing a two sample comparison of means. Assuming equal variances this is the standard two sample t-test but can also be viewed as the F-test in a one-way analysis of variance with two groups. If you have other variables that are related to the response (here X_1) then including it can help improve precision in estimating the difference. (In fact if you compare service without accounting for number you get an estimated diffference of 6.1954 with a confidence interval of [-20.5454, 32.9361].)

If it is not a random smaple and X_1 influences service then it should be included in the model.

e) There is a sign of a trend up indicating a potential interaction. Here an interaction means that there is different slope for each group. This is explored in the rest of the problem.

Part ii: Using Z_1 and Z_2 form products with X_1 $xz1$ and $xz2$. Then running a model with z1,z2 and these two products will give estimated intercepts and slopes directly.

The coefficient for Z_1 is β_{S0} so $\hat{\beta}_{S0} = -5.32808$ and the coefficient for $Z_1 * X_1$ is β_{S1} so $\hat{\beta}_{S1} = 16.1168$.

Similarly $\hat{\beta}_{L0} = 2.81311$ and $\hat{\beta}_{L1} = 14.33941$.

These would agree with running separate regressions for each type.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
z1	1	-5.32808	4.22346	-1.26	0.2142
z2	1	2.81311	3.64685	0.77	0.4449
xz1	1	16.11680	0.75641	21.31	<.0001
xz2	1	14.33941	0.61455	23.33	<.0001

Part iii: If type = 1 (small copier): $E(Y) = \beta_0 + \beta_2 + (\beta_1 + \beta_3)X_1,$.

If type = 0 (large): $E(Y) = \beta_0 + \beta_1 X_1$.

So, $\beta_0 = \beta_{L0}, \quad \beta_1 = \beta_{L1}, \quad \beta_2 = \beta_{S0} - \beta_{S1}, \quad \beta_3 = \beta_{S1} - \beta_{L1}$, since $\beta_0 + \beta_2 = \beta_{S0}$ and $\beta_1 + \beta_3 = \beta_{S1}$.

This shows that the the coefficient for $Z_1 = type$ is the difference in intercepts and the coefficient for $type * number = Z_1 * X_1$ is the different in slopes

The fit is below. Notice that the estimated differences agree with what you would get from the previous part (as they should)

The confidence interval for $\beta_{S0} - \beta_{L0}$ is the CI for β_2 which is [-19.41037, 3.12797]
The confidence interval for $\beta_{S1} - \beta_{L1}$ is the confidence interval for β_3 which is [-0.19084, 3.74561].

The t-test for type (with t=-1.46 and p-value = .1522) is testing $H_0 : \beta_{A0} = \beta_{B0}$.

The t-test for type*number=prod (with t= 1.822 and p-value = .0755)) is testing $H_0 : \beta_{A1} = \beta_{B1}$.

Some indication of unequal slopes. The test for equal slopes is more borderline; you would reject for $\alpha = .10$ but not for $\alpha = .05$. Probably best to keep different slopes for the two types in trying to predict service time.

The test for equal intercepts is non-significant. But, once again remember the issue of power and that we never prove the null. The test by itself is non-informative. Look at the CI also, which has 0 in it, but is also mostly on the negative side with a wide range. To proceed under the assumption that there is no type effect would be dangerous. Best to get more data.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	77222	25741	334.57	<.0001
Error	41	3154.43514	76.93744		

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	2.81311	3.64685	0.77	0.4449	-4.55184	10.17807
number	1	14.33941	0.61455	23.33	<.0001	13.09830	15.58052
type	1	-8.14120	5.58007	-1.46	0.1522	-19.41037	3.12797
prod	1	1.77739	0.97459	1.82	0.0755	-0.19084	3.74561

Consistent Covariance of Estimates

Variable	Intercept	number	type	prod
Intercept	13.195464537	-2.436778265	-13.19546454	2.4367782648
number	-2.436778265	0.5444660697	2.4367782648	-0.54446607
type	-13.19546454	2.4367782648	25.357564511	-4.042676856
prod	2.4367782648	-0.54446607	-4.042676856	0.8480982513

Part iv: Testing simultaneously for equal intercepts and slopes Using the test in SAS or the anova approach in R, leads to the F-test below. The chi-square test, which we get automatically from SAS but will need to customize in R does the test allowing unequal variances.

Test equal Results for Dependent Variable service

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	130.97094	1.70	0.1949
Denominator	41	76.93744		

Test equal Results using ACOV Estimates

DF	Chi-Square	Pr > ChiSq
2	3.74	0.1539

You can also use the full-reduced model approach with $SSE(H_0) = 3416.37702$ under the null model. Under the full model $SSE = 3154.43514$ and $MSE = 76.93744$ with 41 d.o.f. So $F = (3416.378 - 3154.435)/(2 * 76.93744)$ with 2 and 41 degrees of freedom.