7.3.3 SPSS regression output.

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Male=0 Female=1, PIQ, WEIGHT, HEIGHT, VIQ, FSIQ[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: MRI_CNT

a) Variation explained by the model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.808[a]	.652	.585	46759.0

a. Predictors: (Constant), Male=0 Female=1, PIQ, WEIGHT, HEIGHT, VIQ, FSIQ

The variation explained here is 65.2%.

b) Testing whether the x-variables jointly are significant.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.27E+	6	2.117E	9.68	.000[a]
	Residual	6.78E+	31	2186409		
	Total	1.95E+	37			

a. Predictors: (Constant), Male=0 Female=1, PIQ, WEIGHT, HEIGHT, VIQ, FSIQ

b. Dependent Variable: MRI_CNT

If the regression is not significant, then y does not depend on the x's. The hypotheses may be written as:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0 \text{ (y does not depend on x's) model: } y_i = \beta_0 + \varepsilon_i$$

$$H_1: \textit{at least one of the } \beta_i \neq 0 \qquad \text{model: } y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

Test Statistic

Reject H_0 if observed $F > F_{k,n-k,\alpha}$. Conclude that y does depend on x.

Otherwise accept the null hypothesis and conclude that y does not depend on the x's.

The F-value is 9.684 with 6 and 31 degrees of freedom and p = 0.000, a result that is highly significant indicating that the x-variables jointly are significant.

c) Testing whether the x-variables individually are significant.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	20681	23516		.879	.386
	FSIQ	-9389.3	4651.6	-3.08	-2.01	.052
	VIQ	5388.7	2761.4	1.70	1.95	.060
	PIQ	6287.5	2526.2	1.95	2.48	.018
	WEIGHT	87.01	485.5	.028	.179	.859
	HEIGHT	6883.3	3207.9	.379	2.14	.040
	Male=0 Female=1	-42368	24529.	-.295	-1.72	.094

a. Dependent Variable: MRI_CNT

Interpretation of the coefficients:
The ***slopes*** (B) represent the amount by which y changes for every unit change in one of the x's while the rest of the x's remain constant. If PIQ increases by one unit while all the other variables remain constant the y variable (MRI_cnt) will increase by 6287.507.
The ***intercept*** represents the value of y when all the x's are zero.

Hypothesis test about β_i:

$H_0 : \beta_i = 0$ No linear relationship between x_i and y given the rest of the x - variables

$H_1 : \beta_i \neq 0$ There is a linear relationship between x_i and y given the rest of the x - variables

Test Statistic: $$T = \frac{\hat{\beta}_i}{std.Error(\hat{\beta}_i)}$$

Reject H_0 if $T > t_{n-k,\alpha/2}$ or $T < -t_{n-k,\alpha/2}$

Otherwise Accept the null hypothesis

$H_0 : \beta_4 = 0$ No relationship between PIQ and MRI_cnt given FSIQ,VIQ,HEIGHT, WEIGHT,SEX

$H_1 : \beta_4 \neq 0$ There is a linear relationship given the rest

Observed T = 2.489 and p = 0.018

Hence evidence to reject the null hypothesis and conclude that there is a linear relationship between PIQ and MRO_cnt given the rest of the variables.

Also from the table the HEIGHT seems significant given the rest of the variables.

Practical 7.1: Regression analysis

1. Go to SPSS and type the following data (in two columns) from the simple linear regression example.

age	1	1.5	2	2.5	3	3.5	4	4.5	5	6
words	3	22	272	446	896	1222	1540	1870	2072	2562

Use ANALYSE/REGRESSION/LINEAR to declare your y and x variable and fit the model. You can use the PLOTS option to get the residual plots.

i) Tick Histogram and Normal probability plot

ii) Declare ZRESID as Y and ZPRED as X.

2. The accompanying data is of 10 makes of car. The response variable is the fuel consumption (gallons per 100 miles), to be explained in terms of the weight of the car (measured in 1000 pounds). The data are taken from Hogg, R.V. and Ledolter, J. (1987) *Engineering Statistics.* New York: Macmillan. page 272.

Weight	**Fuel**
3.40	5.50
3.80	5.90
4.10	6.50
2.20	3.30
2.60	3.60
2.90	4.60
2.00	2.90
2.70	3.60
1.90	3.10
3.40	4.90

The data are on the shared drive (K):\Sctms\som\ma2013\data\fuel.sav.

i) Fit a simple linear regression line and show that the constant is not needed. Use the Options tab to fit a regression line through the origin.

ii) Check Cook's distance for any influential points. Repeat the regression line through the origin with the most influential point weighted out. Show that the slope is not greatly different.

iii) Refit the regression line with all 10 cars.

iv) Find a 95% confidence interval for the gradient of the regression line which passes through the origin.