

## **Multiple Linear Regression**

### **How do both sex and ethnicity influence total GCSE scores?**

In social science research, it is rare that we would only want to include one explanatory variable in our regression analysis. It is more likely that we would want to investigate the effect that two or more factors have on an outcome, such as total GCSE score. This might be because variables may measure the same thing or have similar relationships – we want to know what the relationship is *controlling* for other variables. Multiple linear regression allows us to obtain predicted values for specific variables under certain conditions, such as GCSE scores between sexes, while controlling for the influence of other factors, such as ethnicity.

We are now going to add additional explanatory variables to our regression model and learn how to make predictions using a multiple linear regression model.

### **Multiple regression model:**

Using the same procedure outlined on the previous pages for a simple model, you can fit a linear regression model with **s1gcseptsnew** as the dependent variable and both **s1gender1** and the **dummy variables for ethnic group** as explanatory variables.

To fit a multiple linear regression, select **Analyze, Regression,** and then **Linear.**

In the dialogue box that appears, move **s1gcseptsnew** to the **Dependent(s)** box and **s1gender1, MIXED, ASIAN, BLACK,** and **OTHER** in the **Independent(s)** box. (Remember we are still using **WHITE** as a baseline, so you do not need to include this dummy variable in your multiple linear regression model.) You should get the following output:

Coefficients <sup>a</sup>						
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			
1	(Constant)	381.242	1.657		230.139	.000
	Black Ethnic Group	-49.069	6.639	-.063	-7.391	.000
	Asian Ethnic Group	7.646	3.783	.017	2.021	.043
	Mixed Ethnic Group	3.623	7.180	.004	.505	.614
	Other Ethnic Group	29.328	12.741	.020	2.302	.021
	S1 Gender	23.995	2.148	.095	11.169	.000

a. Dependent Variable: ks4 pts score on new basis not capped

For this model, our regression equation is:

$$\mathbf{s1gcseptsnew = 381.242 + (s1eth2 \text{ constant } \times \text{ethnicity}) + (s1gender \text{ constant } \times \text{sex})}$$

After our multiple linear regression, our predicted values are:

$$\mathbf{s1gcseptsnew = 381.242 + (-49.069 \times 1) + (23.995 \times 0) = 332.173 \text{ (Black Male)}}$$

$$\mathbf{s1gcseptsnew = 381.242 + (-49.069 \times 1) + (23.995 \times 1) = 356.168 \text{ (Black Female)}}$$

$$\begin{aligned}
s1gcseptsnew &= 381.242 + (7.646 \times 1) + (23.995 \times 0) = 388.888 \text{ (Asian Male)} \\
s1gcseptsnew &= 381.242 + (7.646 \times 1) + (23.995 \times 1) = 412.883 \text{ (Asian Female)} \\
s1gcseptsnew &= 381.242 + (3.623 \times 1) + (23.995 \times 0) = 384.865 \text{ (Mixed Male)} \\
s1gcseptsnew &= 381.242 + (3.623 \times 1) + (23.995 \times 1) = 408.86 \text{ (Mixed Female)} \\
s1gcseptsnew &= 381.242 + (29.328 \times 1) + (23.995 \times 0) = 410.570 \text{ (Other Male)} \\
s1gcseptsnew &= 381.242 + (29.328 \times 1) + (23.995 \times 1) = 434.565 \text{ (Other Female)} \\
s1gcseptsnew &= 381.242 + (23.995 \times 0) = 340.693 \text{ (White Male)} \\
s1gcseptsnew &= 381.242 + (23.995 \times 1) = 340.693 \text{ (White Female)}
\end{aligned}$$

*Taking into consideration the trends we saw in the gender and ethnicity simple linear regression models, how closely do the results of our multiple linear regression follow the established patterns? Do women still have higher total GCSE scores? Are differences with respect to ethnicity still seen?*

*Run another multiple linear regression, including **s1truan** in the model along with **s1gender1** and the ethnicity dummy variables. You'll need to create dummy variables for the categories in **s1truan**, and then select one of them to be the baseline category, remembering to leave that baseline category out of the multiple linear regression model. Do the predicted scores change at all when you control for the influence of **s1truan**? Are the trends we saw previously still illustrated in this model?*

### **Summary**

*Here, we've used multiple linear regression to determine the statistical significance of GCSE scores while controlling for sex and ethnic background. We've learned that there are still statistically significant relationships between GCSE score and ethnicity, and between GCSE score and sex. Finally, we've used the ethnicity and sex coefficients presented to us in the multiple linear regression to predict GCSE scores for people falling into the various ethnicity and sex categories.*

**\*\*\*Note:** as we are making changes to a dataset we'll continue using, please make sure to save your changes before you close down SPSS. This will save you having to repeat sections you've already completed!