Multiple Linear Regression

How do sex and ethnicity influence police confidence 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 confidence in the police. 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 levels of police confidence 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 above for a simple model, you can fit a linear regression model with **policeconf1** as the dependent variable and both **sex** 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 **policeconf1** to the **Dependent(s)** box and **sex**, **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[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.789	.031		440.338	.000
	Respondent Mixed Ethnic Group	1.073	.250	.021	4.285	.000
	Respondent Asian Ethnic Group	-.860	.108	-.039	-7.974	.000
	Respondent Black Ethnic Group	.533	.145	.018	3.681	.000
	Respondent Other Ethnic Group	-.728	.188	-.019	-3.869	.000
	Adult number 1 (respondent): Sex	-.444	.042	-.051	-10.600	.000

a. Dependent Variable: I have confidence in the police

Now that we have run a multiple linear regression on the combined effect of sex and ethnicity on GCSE scores, have the predicted values changed?

Remember that in our simple sex linear regression, the predicted value of police confidence score was 13.325 in females and 13.761 in males. In our simple ethnicity linear regression, the predicted value of police confidence score was 14.617 for Mixed respondents, 12.711 for Asian respondents, 14.067 for Black respondents, 13.550 for White respondents, and 12.81 for respondents of all other ethnicities.

After our multiple linear regression, our values are:

policeconf1 = 13.789 + (1.073 x 1) + (-0.444 x 1) = 14.418 (Mixed Female)
policeconf1 = 13.789 + (1.073 x 1) + (-0.444 x 0) = 14.862 (Mixed Male)
policeconf1 = 13.789 + (-0.860 x 1) + (-0.444 x 1) = 12.485 (Asian Female)
policeconf1 = 13.789 + (-0.860 x 1) + (-0.444 x 0) = 12.929 (Asian Male)
policeconf1 = 13.789 + (.533 x 1) + (-0.444 x 1) = 13.878 (Black Female)
policeconf1 = 13.789 + (.533 x 1) + (-0.444 x 0) = 14.322 (Black Male)
policeconf1 = 13.789 + (-0.728 x 1) + (-0.444 x 1) = 12.617 (Other Female)
policeconf1 = 13.789 + (-0.728 x 1) + (-0.444 x 0) = 13.061 (Other Male)
policeconf1 = 13.789 + (-0.444 x 1) = 13.345 (White Female)
policeconf1 = 13.789 + (-0.444 x 0) = 13.789 (White Male)

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 lower mean police confidence scores? Are differences with respect to ethnicity still seen?

*Run another multiple linear regression, including **wattack** in the model along with **sex1** and the ethnicity dummy variables. You'll need to create dummy variables for the categories in **wattack**, 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 **wattack**? 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 police confidence scores while controlling for sex and ethnic background. We've learned that there is still a statistically significant relationship between police confidence score and ethnicity, and between police score and sex. Finally, we've used the ethnicity and sex coefficients presented to us in the multiple linear regression to predict police confidence scores for people falling into the various ethnicity and sex categories.

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