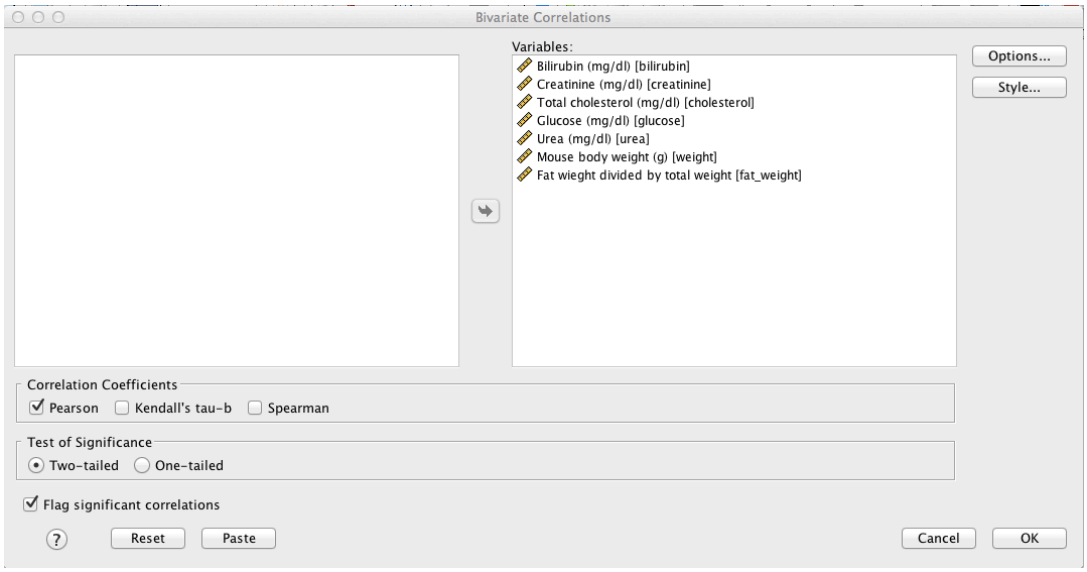
SPSS Tutorial 3: Linear regression

1. The dataset for this tutorial is called **impc_data.sav** (a *.sav file is an SPSS dataset, if you want to use other software, you can use **impc_data.txt** instead, though some information will be lost (variable descriptions, variable level definitions, etc.))

2. Start by looking at the Pearson correlation coefficient between all pairs of variables. Go to Analyze->Correlate->Bivariate. Move all the available variables into the box on the right side



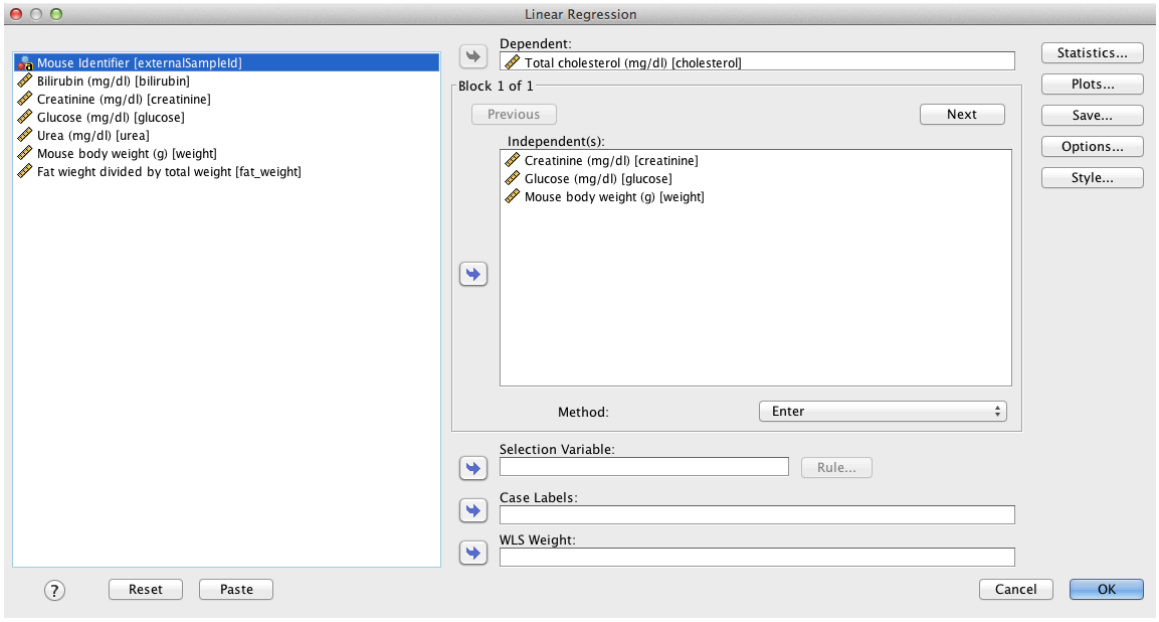
Press "OK" and look at the resulting correlation matrix.

Correlations

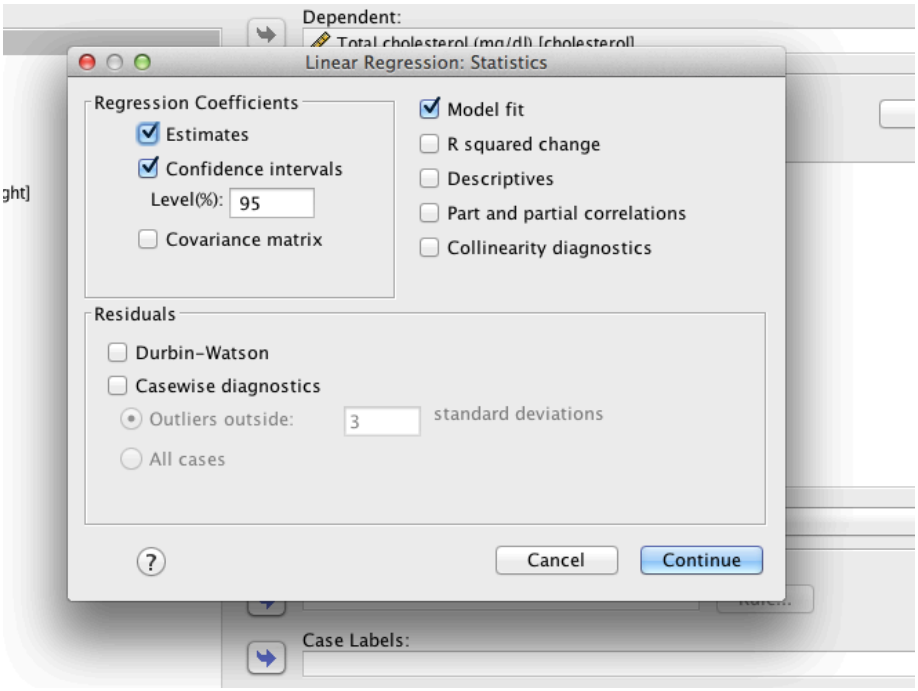
		Bilirubin (mg/dl)	Creatinine (mg/dl)	Total cholesterol (mg/dl)	Glucose (mg/dl)	Urea (mg/dl)	Mouse body weight (g)	Fat wieght divided by total weight
Bilirubin (mg/dl)	Pearson Correlation	1	.145**	-.087**	.147**	-.224**	-.177**	-.045
	Sig. (2-tailed)		.000	.001	.000	.000	.000	.085
	N	1471	1471	1471	1471	1471	1471	1471
Creatinine (mg/dl)	Pearson Correlation	.145**	1	.234**	.472**	.098**	.036	.163**
	Sig. (2-tailed)	.000		.000	.000	.000	.173	.000
	N	1471	1471	1471	1471	1471	1471	1471
Total cholesterol (mg/dl)	Pearson Correlation	-.087**	.234**	1	.370**	.502**	.532**	.546**
	Sig. (2-tailed)	.001	.000		.000	.000	.000	.000
	N	1471	1471	1471	1471	1471	1471	1471
Glucose (mg/dl)	Pearson Correlation	.147**	.472**	.370**	1	.091**	.086**	.227**
	Sig. (2-tailed)	.000	.000	.000		.000	.001	.000
	N	1471	1471	1471	1471	1471	1471	1471
Urea (mg/dl)	Pearson Correlation	-.224**	.098**	.502**	.091**	1	.354**	.109**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
	N	1471	1471	1471	1471	1471	1471	1471
Mouse body weight (g)	Pearson Correlation	-.177**	.036	.532**	.086**	.354**	1	.575**
	Sig. (2-tailed)	.000	.173	.000	.001	.000		.000
	N	1471	1471	1471	1471	1471	1471	1471
Fat wieght divided by total weight	Pearson Correlation	-.045	.163**	.546**	.227**	.109**	.575**	1
	Sig. (2-tailed)	.085	.000	.000	.000	.000	.000	
	N	1471	1471	1471	1471	1471	1471	1471

**. Correlation is significant at the 0.01 level (2-tailed).

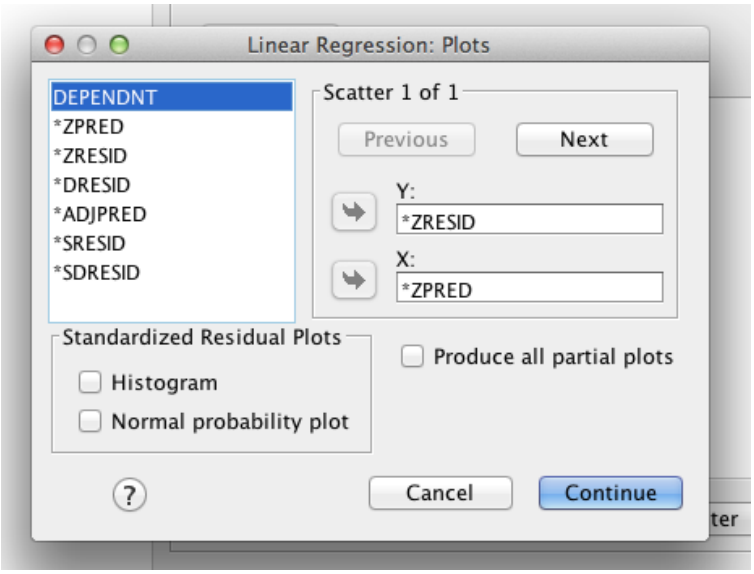
3. Run a multiple linear regression with total cholesterol as the outcome variable and creatinine, glucose, and body weight as the predictor variables.



Then click on the "Statistics" button and make sure "Estimates" and "Confidence Intervals" are checked off.



Next, click on "Plots" and put "ZRESID" as the y-axis variable, and "ZPRED" as the x-axis variable:



Press "Continue" and "OK" to run the regression function.

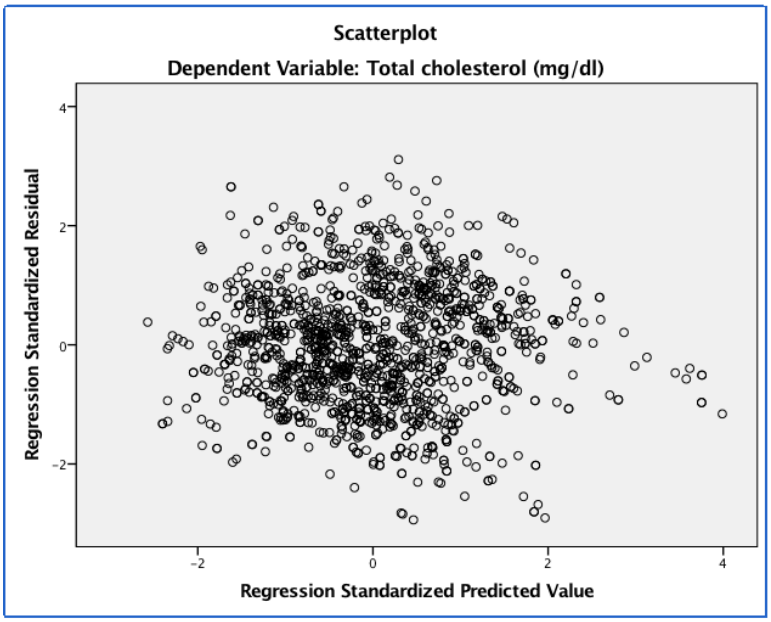
4. Check the table containing the estimates from the regression function:

Model		Unstandardized Coefficients		Standardized Coefficients	t	p-values	95.0% Confidence Interval for B	
		B	Std. Error	Beta		Sig.	Lower Bound	Upper Bound
1	(Constant)	-25.400	5.312		-4.782	.000	-35.820	-14.980
	Creatinine (mg/dl)	67.912	19.710	.079	3.446	.001	29.249	106.574
	Glucose (mg/dl)	.139	.011	.289	12.494	.000	.117	.161
	Mouse body weight (g)	3.438	.139	.505	24.738	.000	3.165	3.711

regression estimates

a. Dependent Variable: Total cholesterol (mg/dl)

The residual plot should also appear:



5. Do another multiple regression model with bilirubin as the outcome and glucose, urea, and mouse body weight as predictors. Which of the predictors are significant? What are the relationships between the predictors and the outcome? Does the residual plot look good?