# Multiple Regression

PSYC 300B - Lecture 6
Dr. J. Nicol

Diagnostic Statistics

- Diagnostics are tools that enable you to see how good or bad your regression equation fits the sample data

- They are a way of assessing the regression model, they are not a way of justifying the removal of data points to effect some desirable change in the regression parameters (Belsey et al., 1980)

Outliers affect the estimates of the regression coefficients
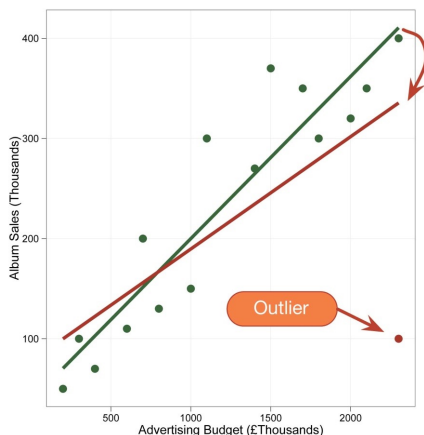
Assumptions of Regression

- *Linearity*
  - The outcome variable should be linearly related to the predictor variable(s)
- *Independent errors*
  - Residuals of cases should not be highly correlated
  - Tested with the Durbin-Watson test — the test statistic varies between 0–4, where 2 is ideal and less than 1 or greater than 3 are generally bad

Assumptions of Regression

- *Homoscedasticity/Homogeniety of variance*
  - The residuals at each level of the predictor(s) should have the same variance
- *Normally-distributed residuals*
  - Assumed that the residuals in the model are random, normally distributed with a mean of 0
  - Save standardized residuals and submit them to a Shapiro Wilk's test of normality

Assumptions of Regression

- *Multicollinearity*
  - Predictor variables should not correlate too highly
  - The correlation matrix is useful for getting an idea of the relationships between predictors and the outcome, and for a look for evidence of multicollinearity
  - If there is no multicollinearity in the data then there should be no substantial correlations (i.e., $r > 0.90$) between predictors

Assumptions of Regression

- The variance inflation factor (VIF) and tolerance are statistics that indicate whether a predictor has a strong linear relationship with the other predictor(s)

- No hard and fast rules, but as guidelines, if the VIF is greater than 10 there is cause for concern (Bowerman & O'Connell, 1990) and if tolerance is below 0.2 it indicates a potential problem (Menard, 1995)

Scatterplots show that the predictors have linear relationships with the outcome variable (and there are no obvious outliers)
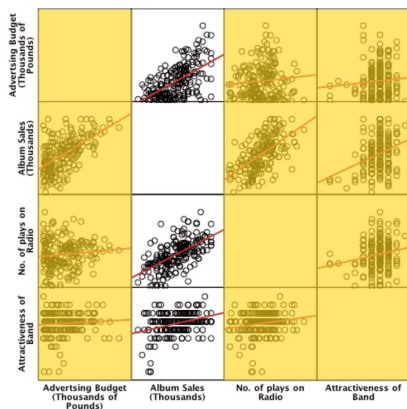

Plot of standardized predicted values (ZPred) against standardized residuals (ZResid) shows random values and indicates that the assumptions of homoscedasticity and linearity have not been violated
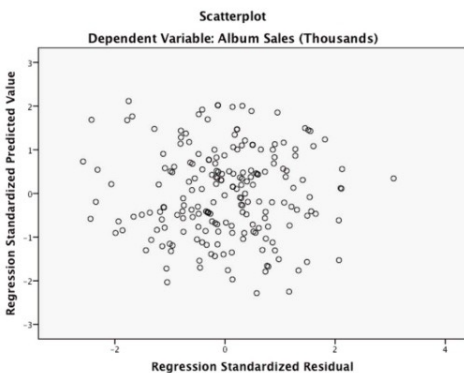
Plotting standardized residuals against predicted values is useful for testing the assumptions of independent errors, homoscedasticity, and linearity
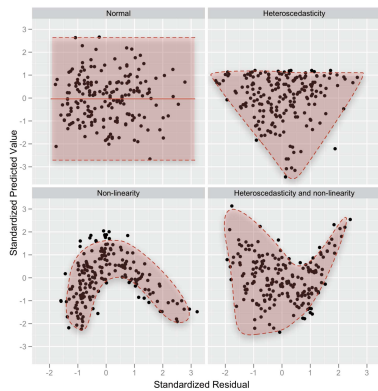
The histogram of the standardized residuals and normal probability plot indicates that the assumption of normally distributed residuals has not been violated
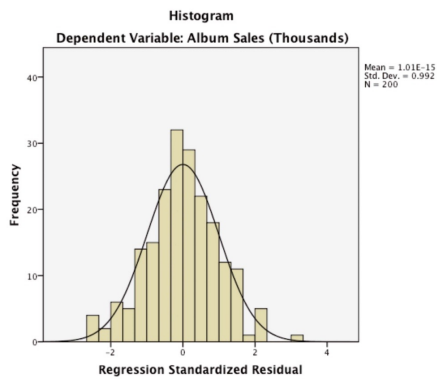
The correlations matrix does not indicate that there is a problem with multicollinearity in the predictor variables

Correlations

		Album Sales (Thousands)	Advertsing Budget (Thousands of Pounds)	No. of plays on Radio	Attractivenes s of Band
Pearson Correlation	Album Sales (Thousands)	1.000	.578	.599	.326
	Advertsing Budget (Thousands of Pounds)	.578	1.000	.102	.081
	No. of plays on Radio	.599	.102	1.000	.182
	Attractiveness of Band	.326	.081	.182	1.000
Sig. (1-tailed)	Album Sales (Thousands)	.	.000	.000	.000
	Advertsing Budget (Thousands of Pounds)	.000	.	.076	.128
	No. of plays on Radio	.000	.076	.	.005
	Attractiveness of Band	.000	.128	.005	.
N	Album Sales (Thousands)	200	200	200	200
	Advertsing Budget (Thousands of Pounds)	200	200	200	200
	No. of plays on Radio	200	200	200	200
	Attractiveness of Band	200	200	200	200

Durbin-Watson statistic indicates that the assumption of independent errors has not been violated

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
					Change Statistics					
1	.578[a]	.335	.331	65.991	.335	99.587	1	198	.000	
2	.815[b]	.665	.660	47.087	.330	96.447	2	196	.000	1.950

a. Predictors: (Constant), Advertsing Budget (Thousands of Pounds)
b. Predictors: (Constant), Advertsing Budget (Thousands of Pounds), Attractiveness of Band, No. of plays on Radio
c. Dependent Variable: Album Sales (Thousands)

Collinearity statistics further confirm that multicollinearity is not a problem

Coefficients[a]

Model		Zero-order	Partial	Part	Tolerance	VIF
		Correlations			Collinearity Statistics	
1	Advertsing Budget (Thousands of Pounds)	.578	.578	.578	1.000	1.000
2	Advertsing Budget (Thousands of Pounds)	.578	.659	.507	.986	1.015
	No. of plays on Radio	.599	.655	.501	.959	1.043
	Attractiveness of Band	.326	.309	.188	.963	1.038

a. Dependent Variable: Album Sales (Thousands)