

## Multiple Regression

PSY 5102: Advanced Statistics for  
Psychological and Behavioral Research 2

### Goals

- Understand when to use multiple regression
- Understand the multiple regression equation and what the regression coefficients represent
- Understand different methods of regression
  - Simultaneous
  - Hierarchical
  - Stepwise
- Understand how to conduct a multiple regression using SPSS
- Understand how to interpret multiple regression
- Understand the assumptions of multiple regression and how to test them

### What is Multiple Regression?

- Linear Regression is a model used to predict the value of one variable from another
- Multiple Regression is a natural extension of this model:
  - We use it to predict values of an outcome from **several** predictors
  - It is a hypothetical model of the relationship between several variables
  - The outcome variable should be continuous and measured at the interval or ratio scale

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

## Multiple Regression as an Equation

- ◎ With multiple regression the relationship is described using a variation of the equation of a straight line
- ◎ Simple Linear Regression:  

$$Y = A + BX + e$$
- ◎ Multiple Regression:  

$$Y = A + B_1X_1 + B_2X_2 + \dots + B_nX_n + e$$

---

---

---

---

---

---

---

## Basic Terminology

- “A” is the Y-intercept
  - The Y-intercept is the value of the Y variable when all Xs = 0
  - This is the point at which the regression line (simple linear regression) or regression plane (multiple regression) crosses the Y-axis (vertical)
- “B<sub>1</sub>” is the regression coefficient for variable 1
- “B<sub>2</sub>” is the regression coefficient for variable 2
- “B<sub>n</sub>” is the regression coefficient for n<sup>th</sup> variable

---

---

---

---

---

---

---

## Basic Terminology

- An important idea in multiple regression is **statistical control** which refers to mathematical operations that remove the effects of other variables from the relationship between a predictor variable and an outcome variable
  - Terms that are used to refer to this statistical control include
    - Partialling
    - Controlling for
    - Residualizing
    - Holding constant
  - The resulting association between a predictor variable and the outcome variable (after controlling for the other predictors) is referred to as **unique association**

---

---

---

---

---

---

---

## Regression: An Example

- A record company boss was interested in predicting record sales from advertising
- Data
  - 200 different album releases
- Outcome variable:
  - Sales (CDs and Downloads) in the week after release
- Predictor variables
  - The amount (in £s) spent promoting the record before release (from our linear regression lecture)
  - Number of plays on the radio (new variable)

---



---



---



---



---



---

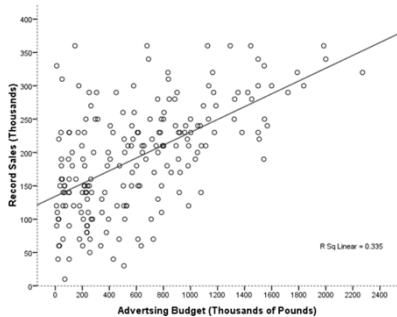


---



---

## The Model with One Predictor




---



---



---



---



---



---

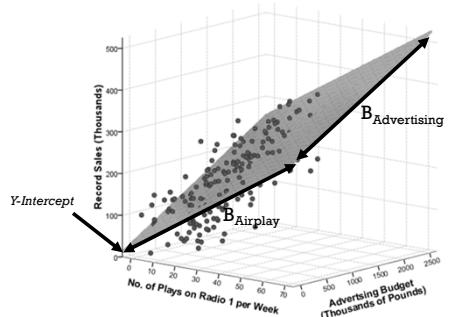


---



---

## The Model with Two Predictors




---



---



---



---



---



---



---



---

## Methods of Regression

### ◎ Simultaneous:

- All predictors are entered simultaneously

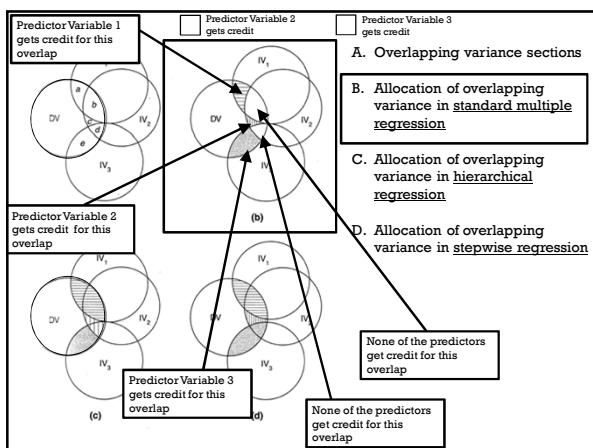
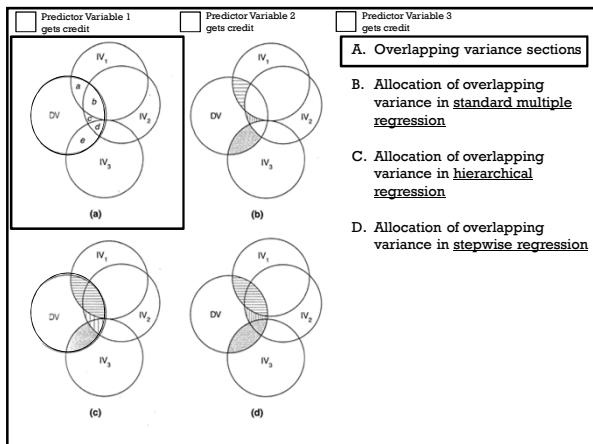
#### • Hierarchical (also referred to as Sequential):

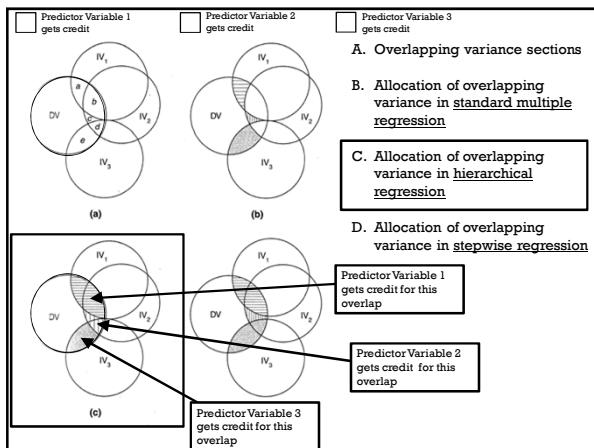
- Researcher decides the order in which variables are entered into the model

### • Stepwise:

- Predictors are selected using their semi-partial correlation with the outcome

- These methods differ in how they allocate shared (or overlapping) variance among the predictors in the model






---



---



---



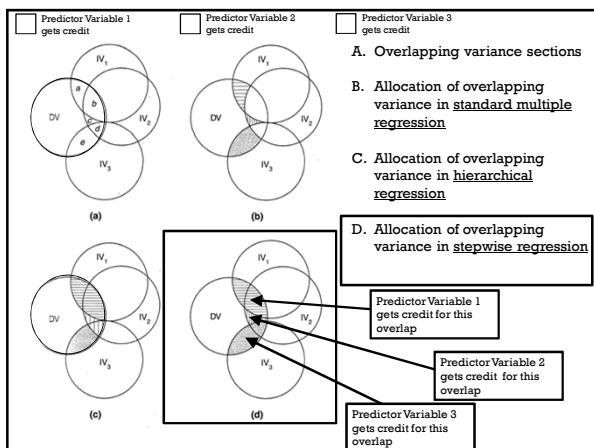
---



---



---




---



---



---



---



---



---

## Simultaneous Regression

• All variables are entered into the model on the same step

- This is the simplest type of multiple regression
- Each predictor variable is assessed as if it had been entered into the regression model after all other predictors
- Each predictor is evaluated in terms of what it adds to the prediction of the outcome variable that is different from the predictability afforded by all the other predictors (i.e., how much unique variance in the outcome variable does it explain?)

---



---



---



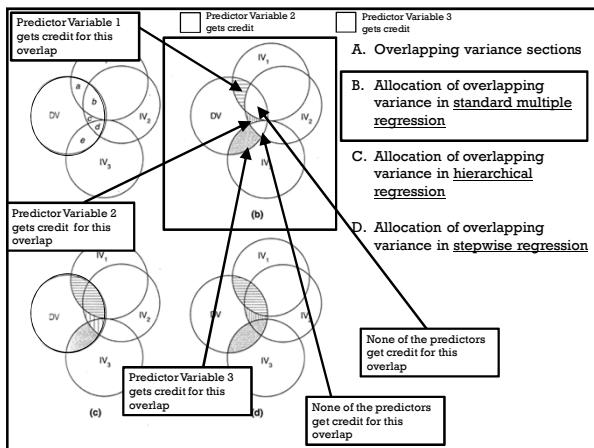
---



---



---




---

---

---

---

---

---

---

---

---

## Hierarchical Regression

- The researcher decides the order in which the predictors are entered into the regression model
- For example, the researcher may enter known predictors (based on past research) into the regression model first with new predictors being entered on subsequent steps (or blocks) of the model

---

---

---

---

---

---

---

---

---

## Hierarchical Regression

- It is the “best” method of regression:
  - Based on theory testing
  - It allows you to see the unique predictive influence of a new variable on the outcome because known predictors are held constant in the model
  - The results that you obtain depend on the variables that you enter into the model so it is important to have good theoretical reasons for including a particular variable
- Weakness of hierarchical multiple regression:
  - It is heavily reliant on the researcher knowing what he or she is doing!

---

---

---

---

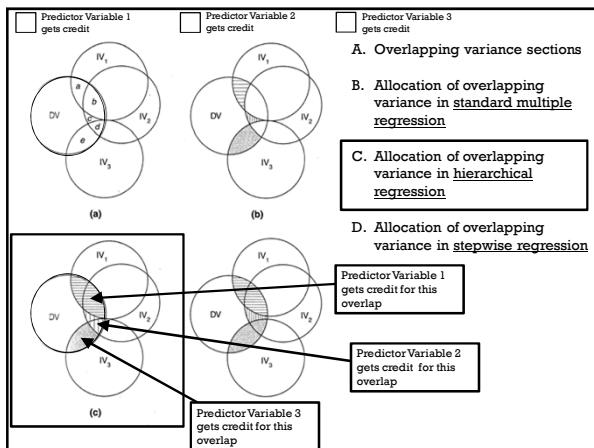
---

---

---

---

---




---



---



---



---



---



---



---



---



---

## Stepwise Regression

- Variables are entered into the model based on mathematical criteria
  - This is a controversial approach and it is not one that I would recommend you use in most cases
- Computer selects variables in steps
  - There are actually three types of "stepwise" regression
    - Forward selection:** Model starts out empty and predictor variables are added one at a time provided they meet statistical criteria for entry (e.g., statistical significance). Variables remain in the model once they are added
    - Backward deletion:** Model starts out with all predictors and variables are deleted one at a time based on a statistical criteria (e.g., statistical significance)
    - Stepwise:** Compromise between forward selection and backward deletion. Model starts out empty and predictor variables are added one at a time based on a statistical criteria...but they can also be deleted at a later step if they fail to continue to display adequate predictive ability

---



---



---



---



---



---



---



---



---

## Stepwise Regression: Step 1

- Step 1
  - SPSS looks for the predictor that can explain the most variance in the outcome variable

---



---



---



---



---



---



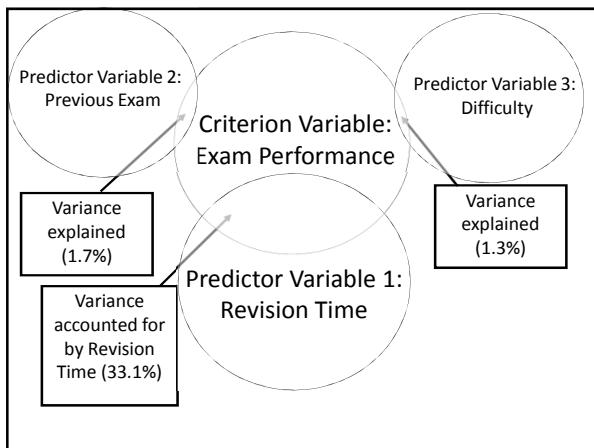
---



---



---




---

---

---

---

---

---

---

Correlations				
		Exam Mark (%)	First Year Exam Grade	Difficulty of Question
Exam Mark (%)	Pearson Correlation	1.000	.130**	-.113*
	Sig. (2-tailed)		.005	.014
	N	474	474	474
First Year Exam Grade	Pearson Correlation	.130**	1.000	-.011
	Sig. (2-tailed)	.005		.047
	N	474	474	474
Difficulty of Question	Pearson Correlation	-.113*	-.012	1.000
	Sig. (2-tailed)	.014		.000
	N	474	474	474
Revision Time	Pearson Correlation	.575**	.047	-.236**
	Sig. (2-tailed)	.000		1.000
	N	474	474	474

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

\* Correlation is significant at the 0.05 level (2-tailed).

---

---

---

---

---

---

---

## Stepwise Regression: Step 2

**Step 2:**

- Having selected the 1<sup>st</sup> predictor, a second one is chosen from the remaining predictors
- The *semi-partial correlation* is used as the criterion for selection
  - The variable with the strongest semi-partial correlation is selected on Step 2
  - The same process is repeated until the remaining predictor variables fail to reach the statistical criteria (e.g., statistical significance)

---

---

---

---

---

---

---

## Semi-Partial Correlation

◎ Partial correlation:

- Measures the relationship between two variables, controlling for the effect that a third variable has on them both

◎ A semi-partial correlation:

- Measures the relationship between two variables controlling for the effect that a third variable has on only one of the others
- In stepwise regression, the semi-partial correlation controls for the overlap between the previous predictor variable(s) and the new predictor variable...without removing the variability in the outcome variable that is accounted for by the previous predictor variables

---



---



---



---



---



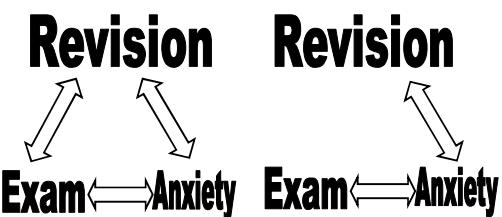
---



---



---



Partial Correlation

Semi-Partial Correlation

---



---



---



---



---



---



---



---

## Semi-Partial Correlation in Regression

◎ The semi-partial correlation

- Measures the relationship between a predictor and the outcome, controlling for the relationship between that predictor and any others already in the model
- It measures the **unique contribution** of a predictor to explaining the variance of the outcome

---



---



---



---



---



---



---



---

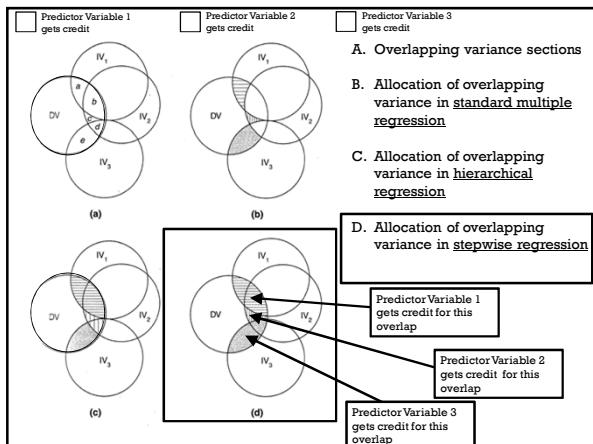
## Problems with Stepwise Methods

⦿ Rely on a mathematical criterion

- Variable selection may depend upon only slight differences in the semi-partial correlation
- These slight numerical differences can lead to major theoretical differences
- These models are often “distorted” by random sampling variation and there is a danger of over-fitting (including too many predictor variables that do not explain meaningful variability in the outcome) or under-fitting (leaving out important predictor variables)

⦿ Should be used only for exploratory purposes

⦿ If stepwise methods are used, then you should cross-validate your results using a second sample



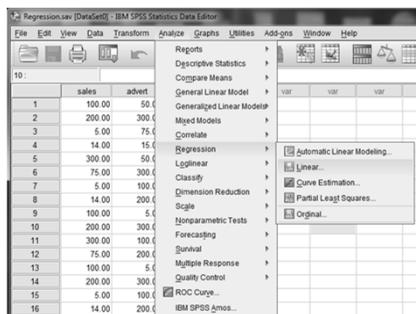
## Sample Size for Multiple Regression

- ⦿ There is no exact number required for multiple regression
- ⦿ You need enough participants to generate a “stable” correlation matrix
- ⦿ One suggestion is to use the following formula to estimate the number of participants you need for multiple regression
  - $50 + 8m$  (where  $m$  is the number of predictors in your model)
  - If there is a lot of “noise” in the data you may need more than that
  - If little noise you can get by with less
  - If you are interested in generalizing your regression model, then you may need at least twice that number of participants

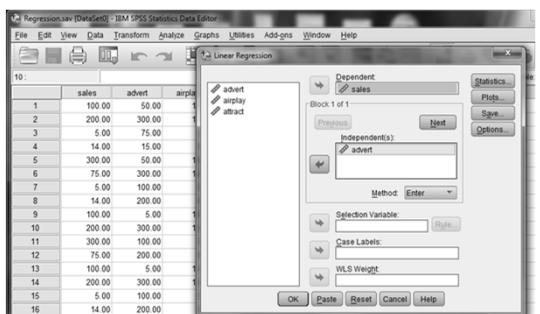
## Multiple Regression Using SPSS

Regression.sav [DataSet0] - IBM SPSS Statistics Data Editor									
	File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Add-on
5 : attract									
		sales	advert	airplay	attract				
1		100.00	50.00	16.00	8.00				
2		200.00	300.00	14.00	9.00				
3		5.00	75.00	2.00	2.00				
4		14.00	15.00	8.00	4.00				

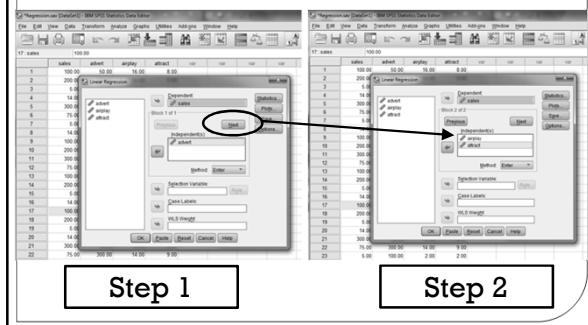
## Multiple Regression Using SPSS



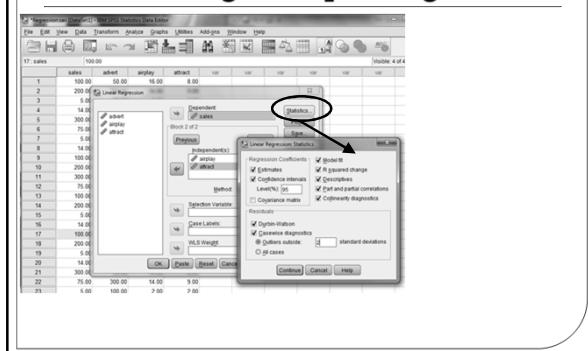
## Multiple Regression Using SPSS



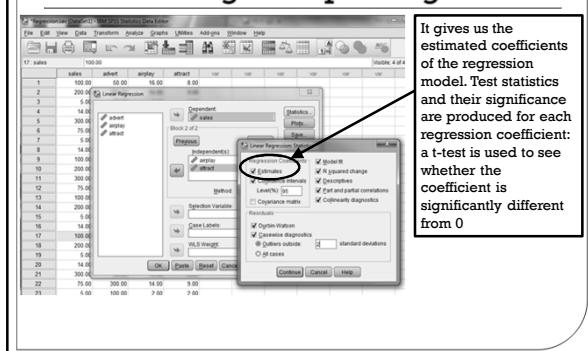
## Conducting Multiple Regression



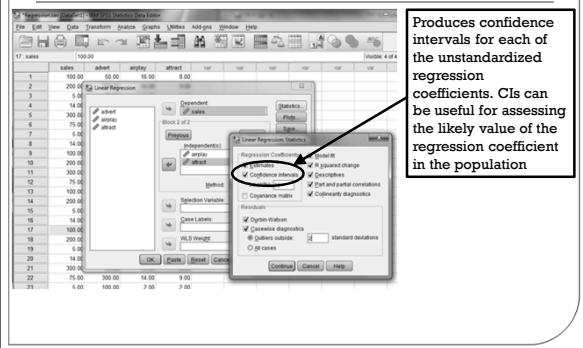
## Conducting Multiple Regression



## Conducting Multiple Regression

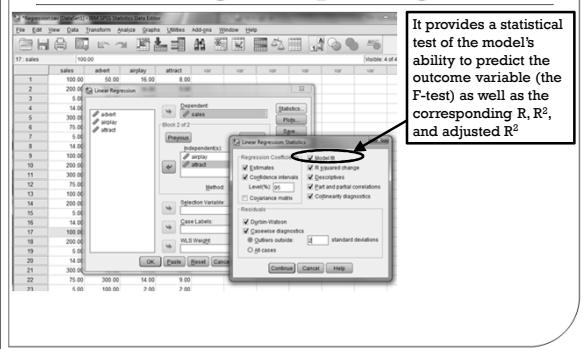


## Conducting Multiple Regression



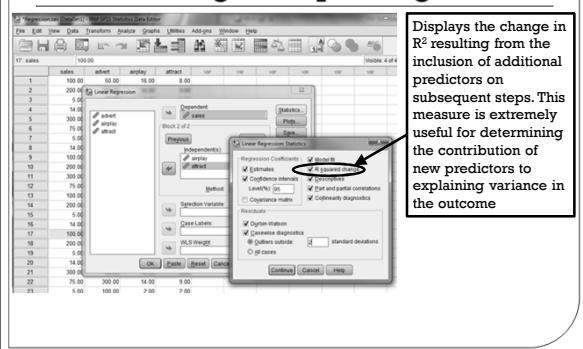
Produces confidence intervals for each of the unstandardized regression coefficients. CIs can be useful for assessing the likely value of the regression coefficient in the population

## Conducting Multiple Regression



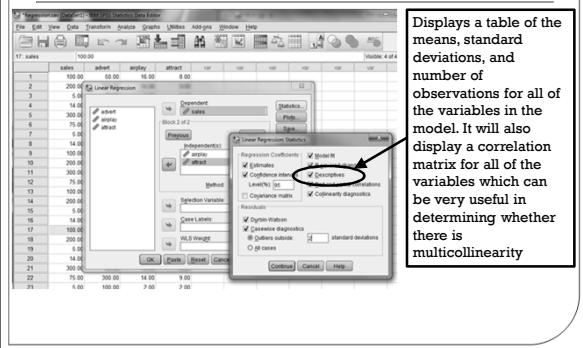
It provides a statistical test of the model's ability to predict the outcome variable (the F-test) as well as the corresponding R,  $R^2$ , and adjusted  $R^2$ .

## Conducting Multiple Regression



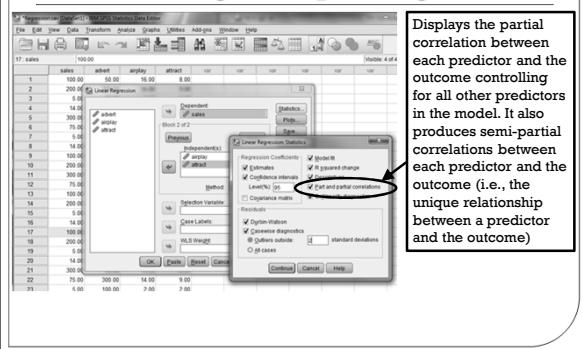
Displays the change in  $R^2$  resulting from the inclusion of additional predictors on subsequent steps. This measure is extremely useful for determining the contribution of new predictors to explaining variance in the outcome.

## Conducting Multiple Regression



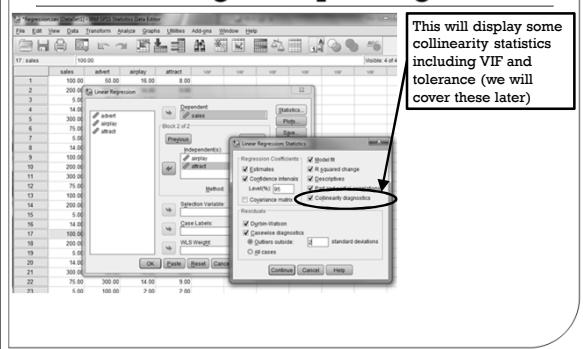
Displays a table of the means, standard deviations, and number of observations for all of the variables in the model. It will also display a correlation matrix for all of the variables which can be very useful in determining whether there is multicollinearity.

## Conducting Multiple Regression



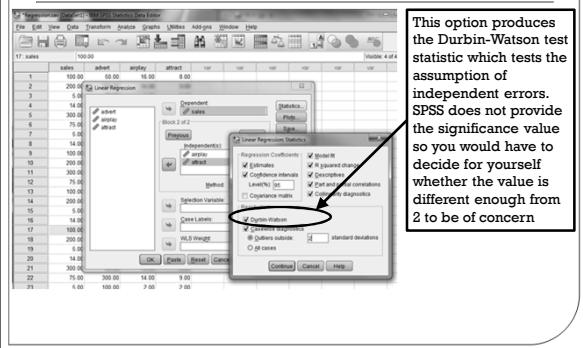
Displays the partial correlation between each predictor and the outcome controlling for all other predictors in the model. It also produces semi-partial correlations between each predictor and the outcome (i.e., the unique relationship between a predictor and the outcome)

## Conducting Multiple Regression



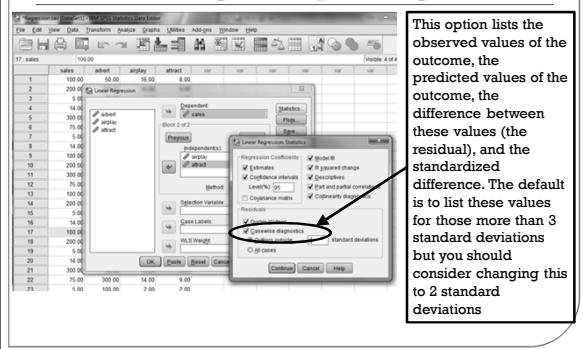
This will display some collinearity statistics including VIF and tolerance (we will cover these later)

## Conducting Multiple Regression



This option produces the Durbin-Watson test statistic which tests the assumption of independent errors. SPSS does not provide the significance value so you would have to decide for yourself whether the value is different enough from 2 to be of concern.

## Conducting Multiple Regression



This option lists the observed values of the outcome, the predicted values of the outcome, the difference between these values (the residual), and the standardized difference. The default is to list these values for those more than 3 standard deviations but you should consider changing this to 2 standard deviations

R and R<sup>2</sup>

### • $R$ (multiple correlation)

- The correlation between the observed values of the outcome, and the values predicted by the model

### • $R^2$ (multiple correlation squared)

- The proportion of variance in the outcome that is accounted for by the predictor variables in the model

### • Adjusted $R^2$

- An estimate of  $R^2$  in the population rather than this particular sample (it captures *shrinkage*)
  - It has been criticized because it tells us nothing about how well the regression model would perform using a different set of data

## Output: Model Summary

Model Summary<sup>2</sup>

Model	R	R Square	Adjusted R Square	Std. Err. of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.875 <sup>b</sup>	.335	.331	65.9914	.335	99.587	1	188	.000	
2	.875 <sup>c</sup>	.335	.331	77.0773	.330	96.447	2	186	.000	1.900

a. Predictors: (Constant), Advertising Budget (thousands of pounds)

b. Predictors: (Constant), Advertising Budget (thousands of pounds); Attractiveness of Band; No. of plays on Radio 1 per week

c. Dependent Variable: Record Sales (thousands)

This is the R for Model 1 which reflects the degree of association between our first regression model (advertising budget only) and the criterion variable

# Output: Model Summary

Model Summary<sup>9</sup>

Model	R	R Square	Adjusted R Square	Std. Err. of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.574	.325	.331	65.9914	.335	99.587	1	198	.000	
2	.818 <sup>b</sup>	.665	.660	47.0573	.330	98.447	2	198	.000	1.950

- a. Predictors: (Constant), Advertising Budget (thousands of pounds)
- b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week
- c. Dependent Variable: Record Sales (thousands)

This is  $R^2$  for Model 1 which represents the amount of variance in the criterion variable that is explained by the model ( $SS_{\text{Model}}$ ) relative to how much variance there was to explain in the first place ( $SS_{\text{Total}}$ ). To express this value as a percentage, you should multiply this value by 100 (which will tell you the percentage of variation in the criterion variable that can be explained by the model).

# Output: Model Summary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Err. of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.578 <sup>a</sup>	.335	.331	65.9914	.335	99.587	1	198	.000	
2	.816 <sup>b</sup>	.665	.660	47.0873	.330	98.447	3	198	.000	1.960

a. Predictors: (Constant), Advertising Budget (thousands of pounds)

b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week

c. Dependent Variable: Record Sales (thousands)

This is the Adjusted  $R^2$  for the first model which corrects  $R^2$  for the number of predictor variables included in the model. It will always be less than or equal to  $R^2$ . It is a more conservative estimate of model fit because it penalizes researchers for including predictor variables that are not strongly associated with the criterion variable.

### Output: Model Summary

Model Summary<sup>f</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.578 <sup>a</sup>	.335	.331	65.9914	.335	99.587	1	198	.000
2	.814 <sup>b</sup>	.665	.660	47.9373	.330	98.447	2	198	.000

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week  
c. Dependent Variable: Record Sales (thousands)

This is the standard error of the estimate for the first model which is a measure of the accuracy of predictions. The larger the standard error of the estimate, the more error in our regression model

### Output: Model Summary

Model Summary<sup>f</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.578 <sup>a</sup>	.335	.331	65.9914	.335	99.587	1	198	.000
2	.814 <sup>b</sup>	.665	.660	47.9373	.330	98.447	2	198	.000

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week  
c. Dependent Variable: Record Sales (thousands)

The change statistics are not very useful for the first step because it is comparing this model (i.e., one predictor) with an empty model (i.e., no predictors)...which is going to be the same as the R<sup>2</sup>

### Output: Model Summary

Model Summary<sup>f</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.578 <sup>a</sup>	.335	.331	65.9914	.335	99.587	1	198	.000
2	.814 <sup>b</sup>	.665	.660	47.9373	.330	98.447	2	198	.000

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week  
c. Dependent Variable: Record Sales (thousands)

This is the R for Model 2 which reflects the degree of association between this regression model (advertising budget, attractiveness of band, and airplay on radio) and the criterion variable

## Output: Model Summary

**Model Summary<sup>a</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.578 <sup>b</sup>	.335	.331	65.9914	.335	99.587	1	198	.000	1.950
2	.811 <sup>b</sup>	.665	.665	47.9373	.330	98.447	2	198	.000	1.950

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week  
c. Dependent Variable: Record Sales (thousands)

This is  $R^2$  for Model 2 which represents the amount of variance in the criterion variable that is explained by the model ( $SS_M$ ) relative to how much variance there was to explain in the first place ( $SS_B$ ). To express this value as a percentage, you should multiply this value by 100 (which will tell you the percentage of variation in the criterion variable that can be explained by the model).

## Output: Model Summary

**Model Summary<sup>a</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.578 <sup>b</sup>	.335	.331	65.9914	.335	99.587	1	198	.000	1.950
2	.811 <sup>b</sup>	.665	.665	47.9373	.330	98.447	2	198	.000	1.950

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week  
c. Dependent Variable: Record Sales (thousands)

This is the Adjusted  $R^2$  for Model 2 which corrects  $R^2$  for the number of predictor variables included in the model. It will always be less than or equal to  $R^2$ . It is a more conservative estimate of model fit because it penalizes researchers for including predictor variables that are not strongly associated with the criterion variable.

## Output: Model Summary

**Model Summary<sup>a</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.578 <sup>b</sup>	.335	.331	65.9914	.335	99.587	1	198	.000	1.950
2	.811 <sup>b</sup>	.665	.665	47.9373	.330	98.447	2	198	.000	1.950

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week  
c. Dependent Variable: Record Sales (thousands)

This is the standard error of the estimate for Model 2 which is a measure of the accuracy of predictions. The larger the standard error of the estimate, the more error in our regression model.

### Output: Model Summary

**Model Summary<sup>a</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.578 <sup>b</sup>	.305	.231	65.9914	.244	99.067	1	199	.000 <sup>c</sup>
2	.811 <sup>b</sup>	.666	.600	47.034	.330	98.447	2	198	.000 <sup>c</sup>

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week  
c. Dependent Variable: Record Sales (thousands)

The change statistics tell us whether the addition of the other predictors in Model 2 significantly improved the model fit. A significant "F Change" value means that there has been a significant improvement in model fit (i.e., more variance in the outcome variable has been explained by Model 2 than Model 1).

### Output: ANOVA

**ANOVA<sup>a</sup>**

Model	Sum of Squares		df	Mean Square	F	Sig.
	Regression	Residual				
1	433687.833	862264.167	198	4354.870	99.587	.000 <sup>b</sup>
Total	1295952.0		199			
2	861377.418	434574.582	196	2217.217	129.498	.000 <sup>b</sup>
Total	1295952.0		199			

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of Plays on Radio 1 per Week  
c. Dependent Variable: Record Sales (thousands)

The ANOVA for Model 1 tells us whether the model, overall, results in a significantly good degree of prediction of the outcome variable.

### Output: ANOVA

**ANOVA<sup>a</sup>**

Model	Sum of Squares		df	Mean Square	F	Sig.
	Regression	Residual				
1	433687.833	862264.167	198	4354.870	99.587	.000 <sup>b</sup>
Total	1295952.0		199			
2	861377.418	434574.582	196	2217.217	129.498	.000 <sup>b</sup>
Total	1295952.0		199			

a. Predictors: (Constant), Advertising Budget (thousands of pounds)  
b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of Plays on Radio 1 per Week  
c. Dependent Variable: Record Sales (thousands)

The ANOVA for Model 2 tells us whether the model, overall, results in a significantly good degree of prediction of the outcome variable. However, the ANOVA does not tell us about the individual contribution of variables in the model.

## Analysis of Variance: ANOVA

### • The F-test

- Examines whether the variance explained by the model ( $SS_M$ ) is significantly greater than the error within the model ( $SS_E$ )
  - It tells us whether using the regression model is significantly better at predicting values of the outcome than simply using their means

## Output: Regression Coefficients

Model	Coefficients <sup>a</sup>							
	Unstandardized Coefficients		Standardized Coefficients		t	Sig.		
	Beta	Std. Error	Beta	Std. Error				
1	(Constant)	134.140	7.537		17.799	.000	119.278	149.002
	Advertising Budget (Thousands of pounds)	.005	.010	.578	9.979	.000	.077	.115
2	(Constant)	26.813	17.350		-1.534	.127	-60.830	7.604
	Advertising Budget (Thousands of pounds)	.085	.007	.511	12.261	.000	.071	.099
	No. of plays on Radio 1 per week	3.367	.278	.512	12.123	.000	2.820	3.915
	Attractiveness of Band	11.086	2.438	.192	4.548	.000	6.279	15.894

This is the Y-intercept for Model 1 (i.e., the place where the regression line intercepts the Y-axis). This means that when £0 in advertising is spent, the model predicts that 134,140 records will be sold (unit of measurement is "thousands of records")

## Output: Regression Coefficients

Model	Coefficients*			95% Confidence Interval for B		
	Unstandardized Coefficients		Standardized Coefficients			
	B	Std. Error	Beta	Lower Bound	Upper Bound	
1	(Constant)	134.140	7.537		119.278	149.002
	Advertising Budget (Thousands of pounds)	.096	.010	.578	.979	.000
2	(Constant)	-26.613	17.350		-1.534	.127
	Advertising Budget (Thousands of pounds)	.085	.007	.511	12.261	.000
	No. of plays on Radio 1 per week	3.367	.278	.512	12.123	.000
	Atntractiveness of Band	11.098	2.438	.192	4.548	.000
	a. Dependent Variable: Recert Sales (Thousands)					

This t-test compares the value of the Y-intercept with 0. If it is significant, then it means that the value of the Y-intercept (134.14 in this example) is significantly different from 0

## Output: Regression Coefficients

Coefficients*						
Model	Unstandardized Coefficients			Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.	95% Confidence Interval for B
1	(Constant) 134.140	7.537		17.799	.000	119.278 149.002
	Advertising Budget (thousands of pounds) .096	.010	.578	9.979	.000	.077 .115
2	(Constant) -26.613	17.350		-1.534	.127	-60.930 7.604
	Advertising Budget (thousands of pounds) .085	.007	.511	12.261	.000	.071 .099
	No. of plays on Radio 1 per week 3.367	.278	.512	12.123	.000	2.820 3.915
	Attractiveness of Band 11.086	2.438	.192	4.548	.000	6.279 15.894

a. Dependent Variable: Record Sales (thousands)

This is the unstandardized slope or gradient coefficient. It is the change in the outcome associated with a one unit change in the predictor variable. This means that for every 1 point the advertising budget is increased (unit of measurement is thousands of pounds) the model predicts an increase of 96 records being sold (unit of measurement for record sales was thousands of records)

## Output: Regression Coefficients

Coefficients*						
Model	Unstandardized Coefficients			Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.	95% Confidence Interval for B
1	(Constant) 134.140	7.537		17.799	.000	119.278 149.002
	Advertising Budget (thousands of pounds) .096	.010	.578	9.979	.000	.077 .115
2	(Constant) -26.613	17.350		-1.534	.127	-60.930 7.604
	Advertising Budget (thousands of pounds) .085	.007	.511	12.261	.000	.071 .099
	No. of plays on Radio 1 per week 3.367	.278	.512	12.123	.000	2.820 3.915
	Attractiveness of Band 11.086	2.438	.192	4.548	.000	6.279 15.894

a. Dependent Variable: Record Sales (thousands)

This is the standardized regression coefficient ( $\beta$ ). It represents the strength of the association between the predictor and the criterion variable. If there is more than one predictor, then  $\beta$  may exceed +/-1

## Output: Regression Coefficients

Coefficients*						
Model	Unstandardized Coefficients			Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.	95% Confidence Interval for B
1	(Constant) 134.140	7.537		17.799	.000	119.278 149.002
	Advertising Budget (thousands of pounds) .096	.010	.578	9.979	.000	.077 .115
2	(Constant) -26.613	17.350		-1.534	.127	-60.930 7.604
	Advertising Budget (thousands of pounds) .085	.007	.511	12.261	.000	.071 .099
	No. of plays on Radio 1 per week 3.367	.278	.512	12.123	.000	2.820 3.915
	Attractiveness of Band 11.086	2.438	.192	4.548	.000	6.279 15.894

a. Dependent Variable: Record Sales (thousands)

This t-test compares the magnitude of the standardized regression coefficient ( $\beta$ ) with 0. If it is significant, then it means that the value of  $\beta$  (0.578 in this example) is significantly different from 0 (i.e., the predictor variable is significantly associated with the criterion variable)

## Output: Regression Coefficients

Coefficients <sup>a</sup>								
Model	B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B		
							Lower Bound	Upper Bound
1	(Constant)	134.140	7.537	17.799	.000	119.278	149.002	
	Advertising Budget (Thousands of pounds)	.096	.010	.578	9.979	.000	.077	.115
2	(Constant)	-26.613	17.350	-1.534	.127	-60.930	7.604	
	Advertising Budget (Thousands of pounds)	.085	.007	.511	12.261	.000	.071	.099
	No. of plays on Radio 1 per week	3.367	.278	.512	12.123	.000	2.820	3.915
	Attractiveness of Band	11.686	2.438	.192	4.548	.000	6.279	15.894

a. Dependent Variable: Record Sales (thousands)

This is the Y-intercept for Model 2 (i.e., the place where the regression plane intercepts the Y-axis). This means that when £0 in advertising is spent, the record receives 0 airplay, and the attractiveness of the band is 0, then the model predicts that -26,613 records will be sold (unit of measurement is "thousands of records")

## Output: Regression Coefficients

Coefficients <sup>a</sup>								
Model	B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B		
							Lower Bound	Upper Bound
1	(Constant)	134.140	7.537	17.799	.000	119.278	149.002	
	Advertising Budget (Thousands of pounds)	.096	.010	.578	9.979	.000	.077	.115
2	(Constant)	-26.613	17.350	-1.534	.127	-60.930	7.604	
	Advertising Budget (Thousands of pounds)	.085	.007	.511	12.261	.000	.071	.099
	No. of plays on Radio 1 per week	3.367	.278	.512	12.123	.000	2.820	3.915
	Attractiveness of Band	11.686	2.438	.192	4.548	.000	6.279	15.894

a. Dependent Variable: Record Sales (thousands)

This t-test compares the value of the Y-intercept for Model 2 with 0. If it is significant, then it means that the value of the Y-intercept (-26,613 in this example) is significantly different from 0

## Output: Regression Coefficients

Coefficients <sup>a</sup>								
Model	B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B		
							Lower Bound	Upper Bound
1	(Constant)	134.140	7.537	17.799	.000	119.278	149.002	
	Advertising Budget (Thousands of pounds)	.096	.010	.578	9.979	.000	.077	.115
2	(Constant)	-26.613	17.350	-1.534	.127	-60.930	7.604	
	Advertising Budget (Thousands of pounds)	.085	.007	.511	12.261	.000	.071	.099
	No. of plays on Radio 1 per week	3.367	.278	.512	12.123	.000	2.820	3.915
	Attractiveness of Band	11.686	2.438	.192	4.548	.000	6.279	15.894

a. Dependent Variable: Record Sales (thousands)

This is the regression information for advertising budget when airplay and attractiveness are also included in the model

## Output: Regression Coefficients

Model	Coefficients*								
	Unstandardized Coefficients		Standardized Coefficients		t	Sig.			
	B	Std. Error	Beta						
1	(Constant) 134.140	.096	.537	17.799	.000	119.278	149.002		
	Advertising Budget (thousands of pounds)	.010	.578	9.979	.000	.077	.115		
2	(Constant) -26.613	.085	.17.350	.007		-1.534	.127	-60.930	7.604
	Advertising Budget (thousands of pounds)	.007	.511	12.261	.000	.071	.099		
	No. of plays on Radio 1 per week	.3367	.278	.512	12.123	.000	2.820	3.915	
	Attractiveness of Band	11.086	2.438	.192	4.548	.000	6.279	15.894	

a. Dependent Variable: Record Sales (thousands)

This is the unstandardized slope or gradient coefficient for airplay when advertising budget and attractiveness are also included in the model. It is the change in the outcome associated with a one unit change in the predictor variable. This means that for every additional airplay the model predicts an increase of 3,367 records being sold (unit of measurement for record sales was thousands of records)

## Output: Regression Coefficients

Model	Coefficients*								
	Unstandardized Coefficients		Standardized Coefficients		t	Sig.			
	B	Std. Error	Beta						
1	(Constant) 134.140	.096	.537	17.799	.000	119.278	149.002		
	Advertising Budget (thousands of pounds)	.010	.578	9.979	.000	.077	.115		
2	(Constant) -26.613	.085	.17.350	.007		-1.534	.127	-60.930	7.604
	Advertising Budget (thousands of pounds)	.007	.511	12.261	.000	.071	.099		
	No. of plays on Radio 1 per week	.3367	.278	.512	12.123	.000	2.820	3.915	
	Attractiveness of Band	11.086	2.438	.192	4.548	.000	6.279	15.894	

a. Dependent Variable: Record Sales (thousands)

This is the unstandardized slope or gradient coefficient for attractiveness of the band when advertising budget and airplay are also included in the model. It is the change in the outcome associated with a one unit change in the predictor variable. This means that for every additional point on attractiveness the model predicts an increase of 11,086 records being sold (unit of measurement for record sales was thousands of records)

## Output: Regression Coefficients

Model	Coefficients*								
	Unstandardized Coefficients		Standardized Coefficients		t	Sig.			
	B	Std. Error	Beta						
1	(Constant) 134.140	.096	.537	17.799	.000	119.278	149.002		
	Advertising Budget (thousands of pounds)	.010	.578	9.979	.000	.077	.115		
2	(Constant) -26.613	.085	.17.350	.007		-1.534	.127	-60.930	7.604
	Advertising Budget (thousands of pounds)	.007	.511	12.261	.000	.071	.099		
	No. of plays on Radio 1 per week	.3367	.278	.512	12.123	.000	2.820	3.915	
	Attractiveness of Band	11.086	2.438	.192	4.548	.000	6.279	15.894	

a. Dependent Variable: Record Sales (thousands)

This is the standardized regression coefficient ( $\beta$ ). It represents the strength of the association between airplay and record sales. It is standardized so it can be compared with other  $\beta$ s

## Output: Regression Coefficients

Coefficients*						
Model	Unstandardized Coefficients			Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.	95% Confidence Interval for B
1	(Constant) 134.140	7.537		17.799	.000	119.278 149.002
	Advertising Budget (thousands of pounds) .096	.010	.578	9.979	.000	.077 .115
2	(Constant) -26.613	17.350		-1.534	.127	-60.930 7.604
	Advertising Budget (thousands of pounds) .085	.007	.511	12.261	.000	.071 .099
	No. of plays on Radio 1 per week 3.367	.278	.512	12.123	.000	2.820 3.915
	Attractiveness of Band 11.086	2.438	.192	4.548	.000	6.279 15.894

a. Dependent Variable: Record Sales (thousands)

This is the standardized regression coefficient ( $\beta$ ). It represents the strength of the association between the attractiveness of the band and record sales. It is standardized so it can be compared with other  $\beta$ s

## Output: Regression Coefficients

Coefficients*						
Model	Unstandardized Coefficients			Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.	95% Confidence Interval for B
1	(Constant) 134.140	7.537		17.799	.000	119.278 149.002
	Advertising Budget (thousands of pounds) .096	.010	.578	9.979	.000	.077 .115
2	(Constant) -26.613	17.350		-1.534	.127	-60.930 7.604
	Advertising Budget (thousands of pounds) .085	.007	.511	12.261	.000	.071 .099
	No. of plays on Radio 1 per week 3.367	.278	.512	12.123	.000	2.820 3.915
	Attractiveness of Band 11.086	2.438	.192	4.548	.000	6.279 15.894

a. Dependent Variable: Record Sales (thousands)

This t-test compares the magnitude of the standardized regression coefficient ( $\beta$ ) with 0. If it is significant, then it means that the value of  $\beta$  (0.512 in this example) is significantly different from 0 (i.e., airplay is significantly associated with record sales)

## Output: Regression Coefficients

Coefficients*						
Model	Unstandardized Coefficients			Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.	95% Confidence Interval for B
1	(Constant) 134.140	7.537		17.799	.000	119.278 149.002
	Advertising Budget (thousands of pounds) .096	.010	.578	9.979	.000	.077 .115
2	(Constant) -26.613	17.350		-1.534	.127	-60.930 7.604
	Advertising Budget (thousands of pounds) .085	.007	.511	12.261	.000	.071 .099
	No. of plays on Radio 1 per week 3.367	.278	.512	12.123	.000	2.820 3.915
	Attractiveness of Band 11.086	2.438	.192	4.548	.000	6.279 15.894

a. Dependent Variable: Record Sales (thousands)

This t-test compares the magnitude of the standardized regression coefficient ( $\beta$ ) with 0. If it is significant, then it means that the value of  $\beta$  (0.192 in this example) is significantly different from 0 (i.e., attractiveness of the band is significantly associated with record sales)

## How to Interpret Regression Coefficients

- ⦿ Unstandardized regression coefficients ( $B$ s):
  - the change in the outcome variable associated with a single-unit change in the predictor variable
- ⦿ Standardized regression coefficients ( $\beta$ s):
  - tell us the same thing as the unstandardized coefficient but it is expressed in terms of standard deviations (i.e., the number of standard deviations you would expect the outcome variable to change in accordance with a one standard deviation change in the predictor variable)

---



---



---



---



---



---



---



---



---



---



---



---

## How well does the model fit the data?

- ⦿ There are two ways to assess the accuracy of the model in the sample
  - Residual Statistics
    - Standardized Residuals
    - In an average sample, 95% of standardized residuals should lie between  $\pm 2$  and 99% of standardized residuals should lie between  $\pm 2.5$
    - Outliers: Any case for which the absolute value of the standardized residual is 3 or more, is likely to be an outlier
  - Influential cases
    - Mahalanobis distance: Measures the distance of each case from the mean of the predictor variable (absolute values greater than 15 may be problematic but see Barnett & Lewis, 1978)
    - Cook's distance: Measures the influence of a single case on the model as a whole (absolute values greater than 1 may be cause for concern according to Weisberg, 1982)

---



---



---



---



---



---



---



---



---



---



---



---

## Can you predict future scores?

- ⦿ Can the regression be generalized to other data?
- ⦿ Randomly separate a data set into two halves
  - Estimate regression equation with one half
  - Apply it to the other half and see if it fits
- ⦿ Collect a second sample
  - Estimate regression equation with first sample
  - Apply it to the second sample and see if it predicts

---



---



---



---



---



---



---



---



---



---



---



---

## Generalization

- ⦿ When we run a regression, we hope to be able to generalize the sample model to the entire population
- ⦿ To do this, several assumptions must be met
- ⦿ Violating these assumptions prevents us from being able to adequately generalize our results beyond the present sample

---



---



---



---



---



---



---



---



---

## Straightforward Assumptions

- ⦿ Variable Type:
  - Outcome must be continuous
  - Predictors can be continuous or dichotomous
- ⦿ Non-Zero Variance:
  - Predictors must not have zero variance
- ⦿ Linearity:
  - The relationship we model is, in reality, linear
  - We will talk about curvilinear associations later
- ⦿ Independence:
  - All values of the outcome should come from a different person
- ⦿ Normality:
  - The outcome variable must be normally distributed (no skewness or outliers)
  - Normally distributed predictors can make interpretation easier but this is not required

---



---



---



---



---



---



---



---



---

## The More Tricky Assumptions

- ⦿ No multicollinearity
  - Predictors must not be highly correlated because this can cause problems estimating the regression coefficients
- ⦿ No multivariate outliers
  - Leverage – an observation with an extreme value on a predictor variable is said to have high leverage
  - Influence – individual observations that have an undue impact on the coefficients
  - Mahalanobis distance and Cook's distance
- ⦿ Homoscedasticity (homogeneity of variance)
  - For each value of the predictors the variance of the error term should be constant
- ⦿ Independent errors
  - For any pair of observations, the error terms should be uncorrelated
- ⦿ Normally-distributed errors
- ⦿ Model specification
  - The model should include relevant variables and exclude irrelevant variables
  - The model may only include linear terms when the relationships between the predictors and the outcome variable are non-linear

---



---



---



---



---



---



---



---



---

## Multicollinearity

- Multicollinearity exists if predictors are highly correlated
    - The best prediction occurs when the predictors are moderately independent of each other, but each is highly correlated with the outcome variable
  - Problems stemming from multicollinearity
    - Increases standard errors of regression coefficients (creates untrustworthy coefficients)
    - It limits the size of R because the variables may be accounting for the same variance in the outcome (i.e., they are not accounting for unique variance)
    - Difficult to distinguish between the importance of predictors because of their overlap (they appear to be interchangeable with each other)
  - This assumption can be checked with collinearity diagnostics
    - Look at correlation matrix of predictors for values above .80
    - Variance Inflation Factor (VIF): Indicates whether the variable has a strong linear relationship with the other predictors
    - Tolerance =  $1 - R^2$

Coefficients <sup>a</sup>						
Model	Correlations			Collinearity Statistics		
	Zero-order	Partial	Part	Tolerance	VIF	
1	Advertising Budget (thousands of pounds)	.578	.578	.578	1.000	1.000
2	Advertising Budget (thousands of pounds)	.578	.659	.503	.986	1.015
	No. of plays on Radio 1 per week	.599	.655	.501	.959	1.043
	Attractiveness of Band	.326	.309	.188	.963	1.038

a. Dependent Variable: Record Sales (thousands)

Tolerance should be more than 0.2 or it may be indicative of problems (Menard, 1995)

Coefficients <sup>a</sup>						
Model		Correlations		Collinearity Statistics		
		Zero-order	Partial	Part	Tolerance	VIF
1	Advertising Budget (thousands of pounds)	.578	.578	.578	1.000	1.000
2	Advertising Budget (thousands of pounds)	.578	.659	.507	.986	1.015
	No. of plays on Radio 1 per week	.599	.655	.501	.949	1.043
	Attractiveness of Band	.326	.309	.188	.985	1.038

a. Dependent Variable: Record Sales (thousands)

VIF should be less than 10 or it may be indicative of problems  
(Myers, 1990)

## Checking Assumptions about Errors

- ◎ Homoscedacity/Independence of Errors:
  - Plot ZRESID against ZPRED
  
- ◎ Normality of Errors:
  - Normal probability plot

---

---

---

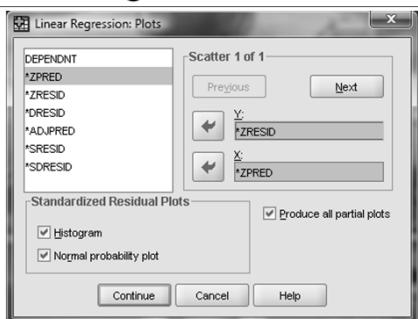
---

---

---

---

## Regression Plots




---

---

---

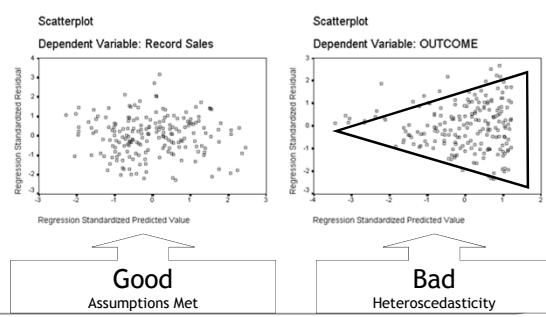
---

---

---

---

## Homoscedasticity: ZRESID vs. ZPRED




---

---

---

---

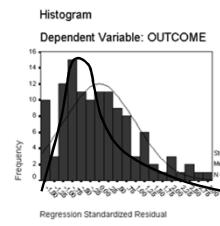
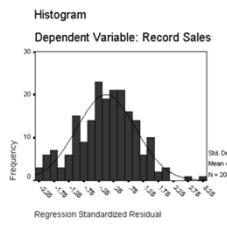
---

---

---

## Normality of Errors: Histograms

Residuals should be normally distributed

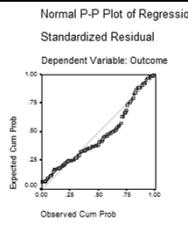
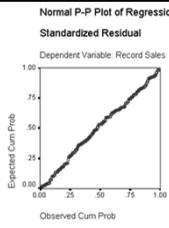


Good

Bad

## Normality of Errors: Normal Probability Plot

The straight line represents a normal distribution so deviations from that tell us that the residuals for our data are not normally distributed



Good

Bad