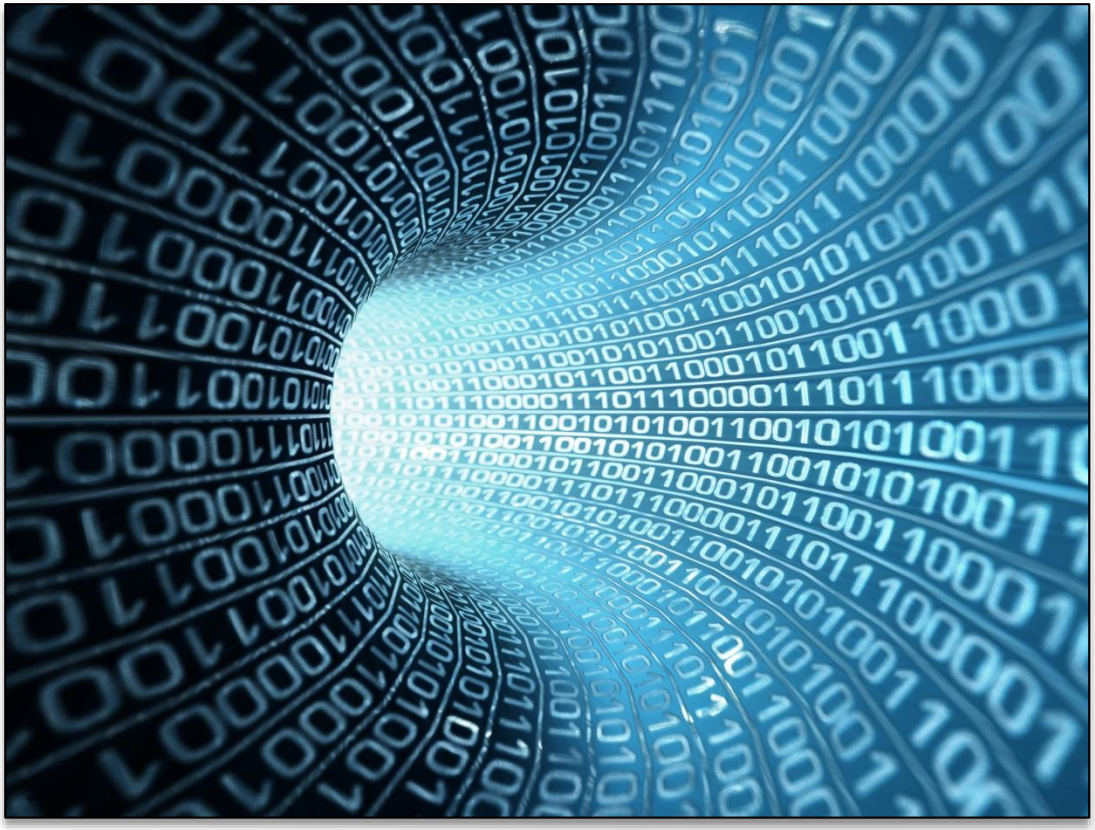
A Brief Introduction to Quantitative Analysis



Concepts

- Descriptive statistics
- Inferential statistics
    - Population
    - Sample
    - Statistical significance
    - P-value
- Correlation Coefficient
    - Pearson correlation
- T-test
- Linear regression
    - Coefficient
    - R Square

Descriptive and inferential statistics

Descriptive statistics

Goal: To summarize a group

Why do it?

- It is hard to think about characteristics of a large group as only the individuals' characteristics.
- e.g. Visualize 4,000 persons ages (without thinking an average and a distribution)

How? With simple maths, like averages, variance, etc.

Examples:

- Mean
- Median
- Standard deviation

Inferential statistics

Goal: To draw conclusions about a larger, unseen, population from a smaller sample of that population.

Why do it?

- We want to know about society, not just our sample.
- Censuses of whole populations are very expensive.

How?

- More complex mathematics.
- Asks what do we know about the true 'population parameter', given the information in the sample.
- Significance tests, which generally give the estimated probability that the population parameter is zero.

Examples:

- Any measure with a standard error
- Any measure with a significance test or p-value

Descriptive statistics

The most commonly used descriptive statistics are:

- Mean
- Standard deviation
- Minimum
- Maximum
- Number of valid cases (N)

Descriptive statistics

Mean – tells us the central tendency

Standard deviation – tells us the variation within the sample

Minimum and Maximum – tells us the range of values

Number of valid cases (N) – tells us sample size, and if for this variable there is a lot of missing data (such as people not answering that survey question).

Inferential statistics

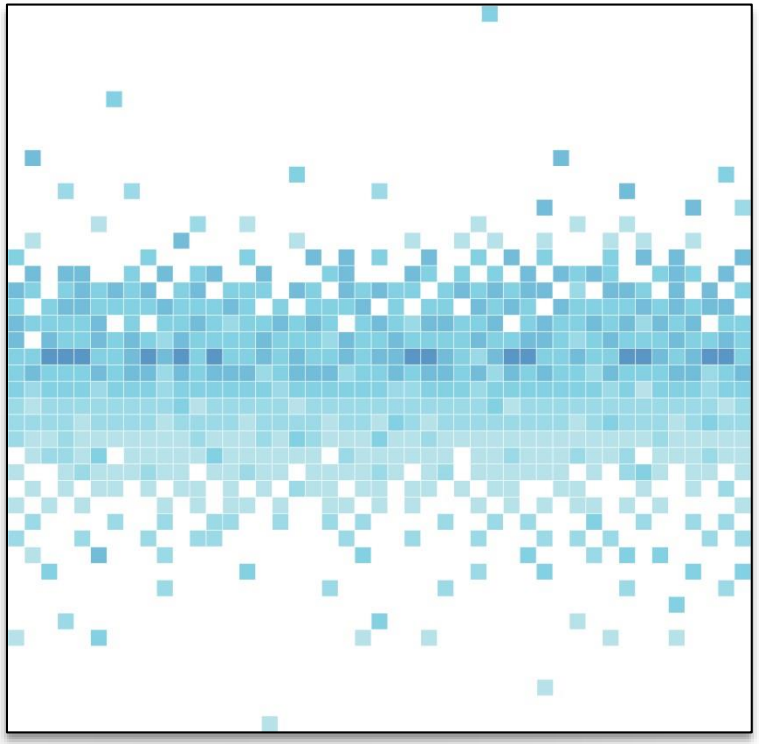
Important terms

Population: the group which you are drawing inferences about. E.g. Australian voters

Sample: the set of cases (individuals) you have to analyse, assumed to be drawn as a true random sample from the population. E.g. A sample of 1,000 voters surveyed by phone.

Population parameter: this is a true value in the population, which you are trying to estimate. E.g. the true mean age of Australian voters.

Sample statistic: this is a number you calculate on your sample, in an attempt to estimate the population parameter. E.g. the mean age of your sample of 1,000 voters.



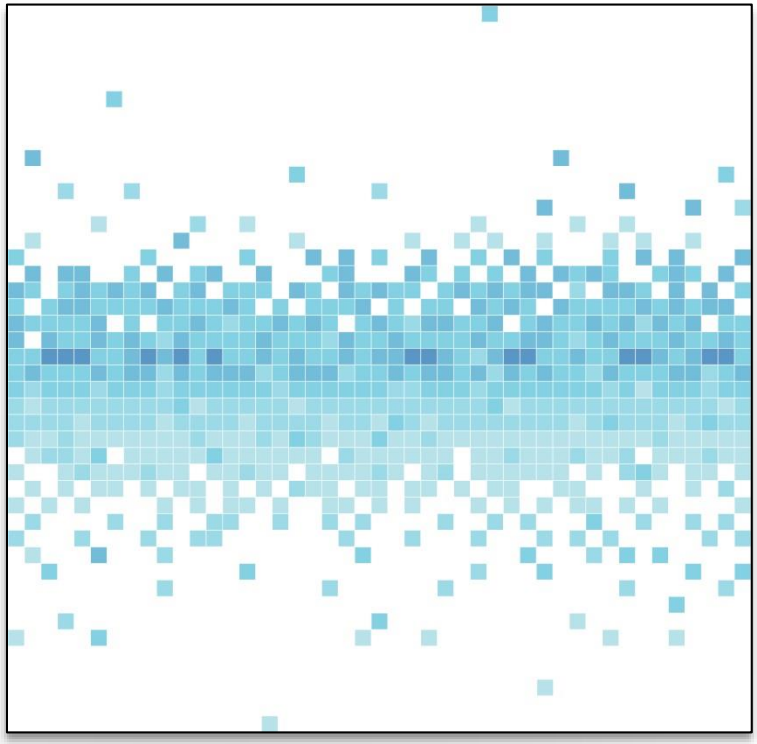
Inferential statistics

Important terms

Null hypothesis: This is – roughly speaking – the hypothesis that "nothing is happening". Most of the time the null hypothesis is that the population parameter is zero.

P-value/statistical significance: Roughly speaking - the chance that our population parameter is zero. This is expressed as a number between 0 and 1. The critical value is $p < 0.05$, which means a less that 5% chance that our population parameter is zero (or whatever our null hypothesis is).

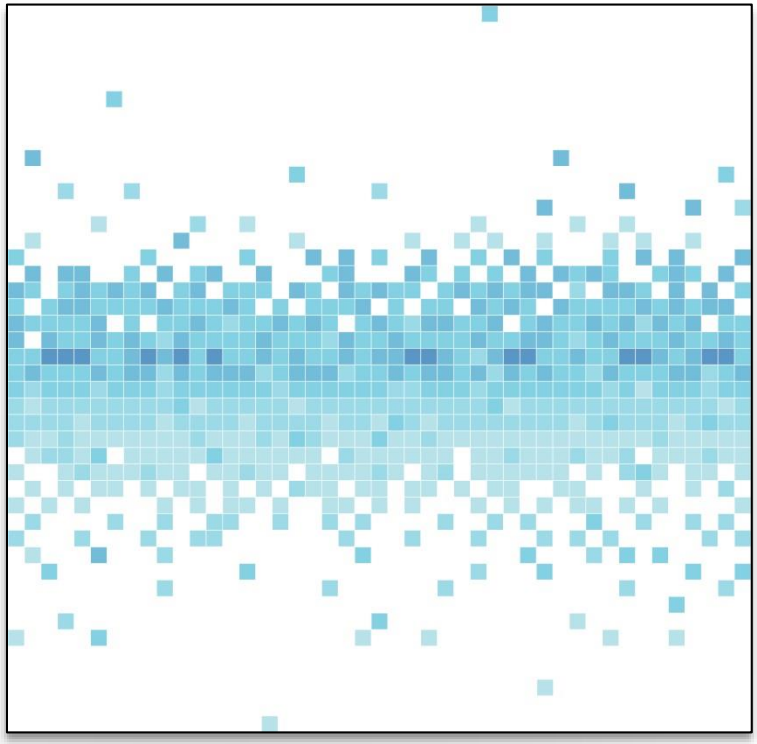
Confidence interval: This is the range within which we expect (normally with 95% confidence) the population parameter to lie.



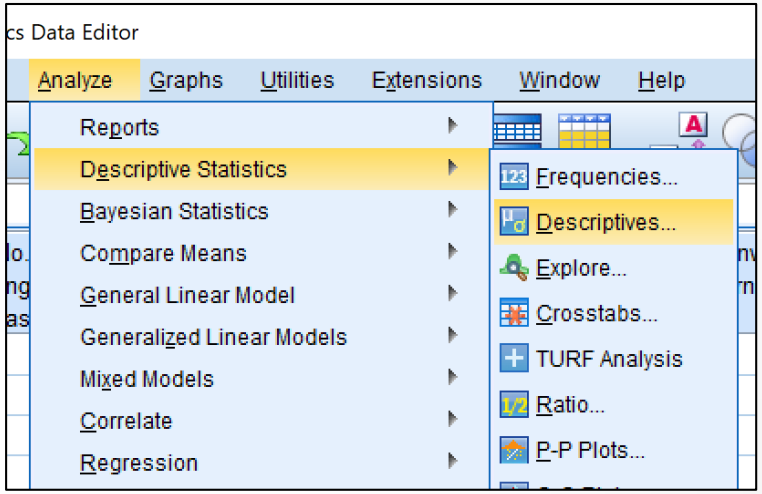
An example

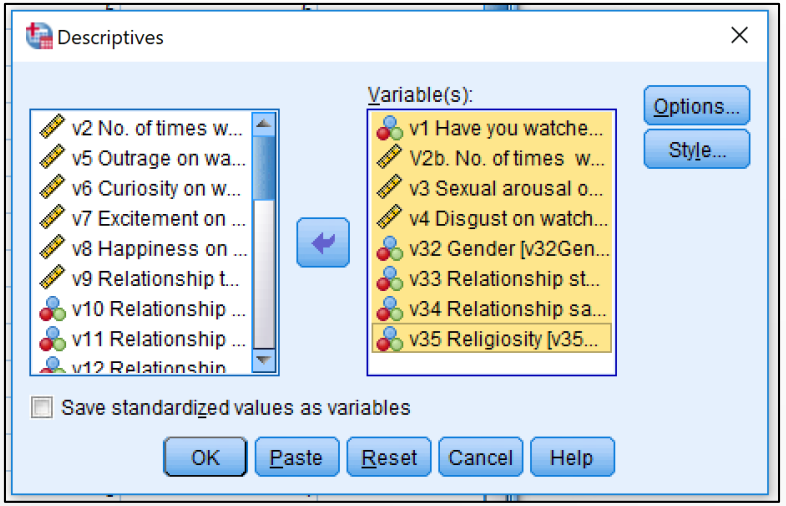
Dataset: Pornography consumption by students in Singapore.

- 595 respondents
- Approved for use for teaching by Ethics Committee (SMU IRB)
- Anonymized data.
- Available on methods101.com
  >SPSS Intro
  >Practice Datasets



An example


cs Data Editor
Analyze
Graphs
Utilities
Extensions
Window
Help
Reports
Descriptive Statistics
Bayesian Statistics
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Frequencies...
Descriptives...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...


Descriptives
Variable(s):
v2 No. of times w...
v5 Outrage on wa...
v6 Curiosity on w...
v7 Excitement on ...
v8 Happiness on ...
v9 Relationship t...
v10 Relationship ...
v11 Relationship ...
v1 Have you watche...
V2b. No. of times w...
v3 Sexual arousal o...
v4 Disgust on watch...
v32 Gender [v32Gen...
v33 Relationship st...
v34 Relationship sa...
v35 Religiosity [v35...
Options...
Style...
Save standardized values as variables
OK
Paste
Reset
Cancel
Help

An example

Descriptive statistics

- 86% watched porn
- 9.65 times in last month
- Gender (1=female): 40% female respondents
- 58% in relationship
- 6 out of 7 average relationships satisfaction
- N for relationships satisfaction = 252 (only those in relationships answer this question)

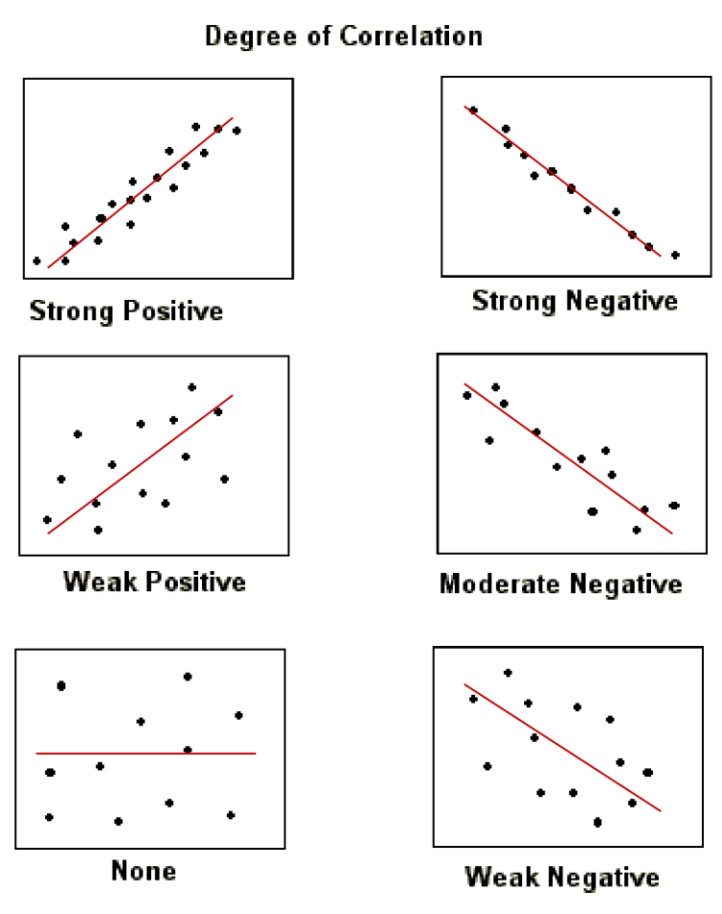
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
v1 Have you watched porn	596	0	1	.86	.348
V2b. No. of times watching porn missong code zero	596	0	267	9.65	18.294
v3 Sexual arousal on watching porn	596	1	7	5.44	1.644
v4 Disgust on watching porn	596	1	7	3.18	1.850
v32 Gender	596	0	1	.40	.491
v33 Relationship status	596	0	1	.58	.494
v34 Relationship satisfaction	252	1	7	6.00	1.066
v35 Religiosity	596	1	7	3.29	1.955
Valid N (listwise)	252				

Correlation

How to measure relationship between two variables?

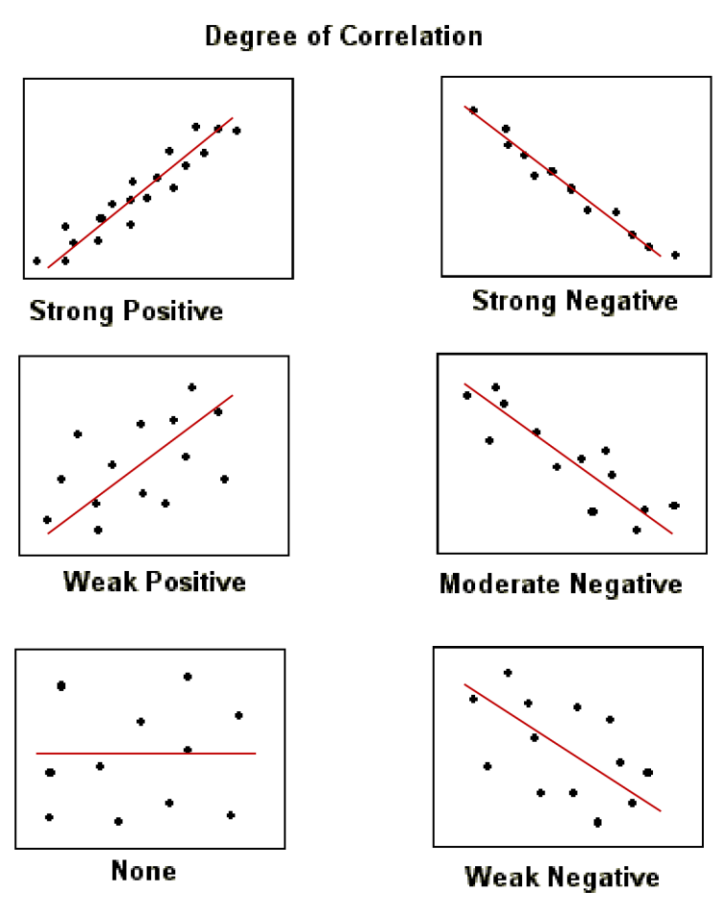
Correlation coefficients are one of the most important measures.



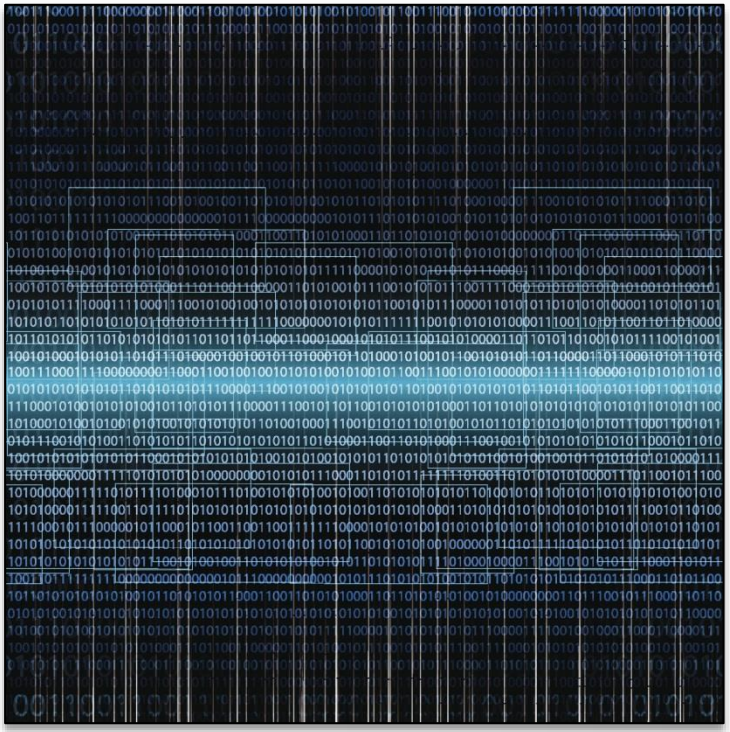
Correlation

A correlation coefficient is:

- a number between 1 and -1 that indicates the strength of the relationship between two variables,
- with 1 indicating that they completely covary in a positive direction,
- -1 indicating that they completely covary in an opposite direction, and
- 0 indicating that they are statistically independent.



Correlation



Interpreting effect size

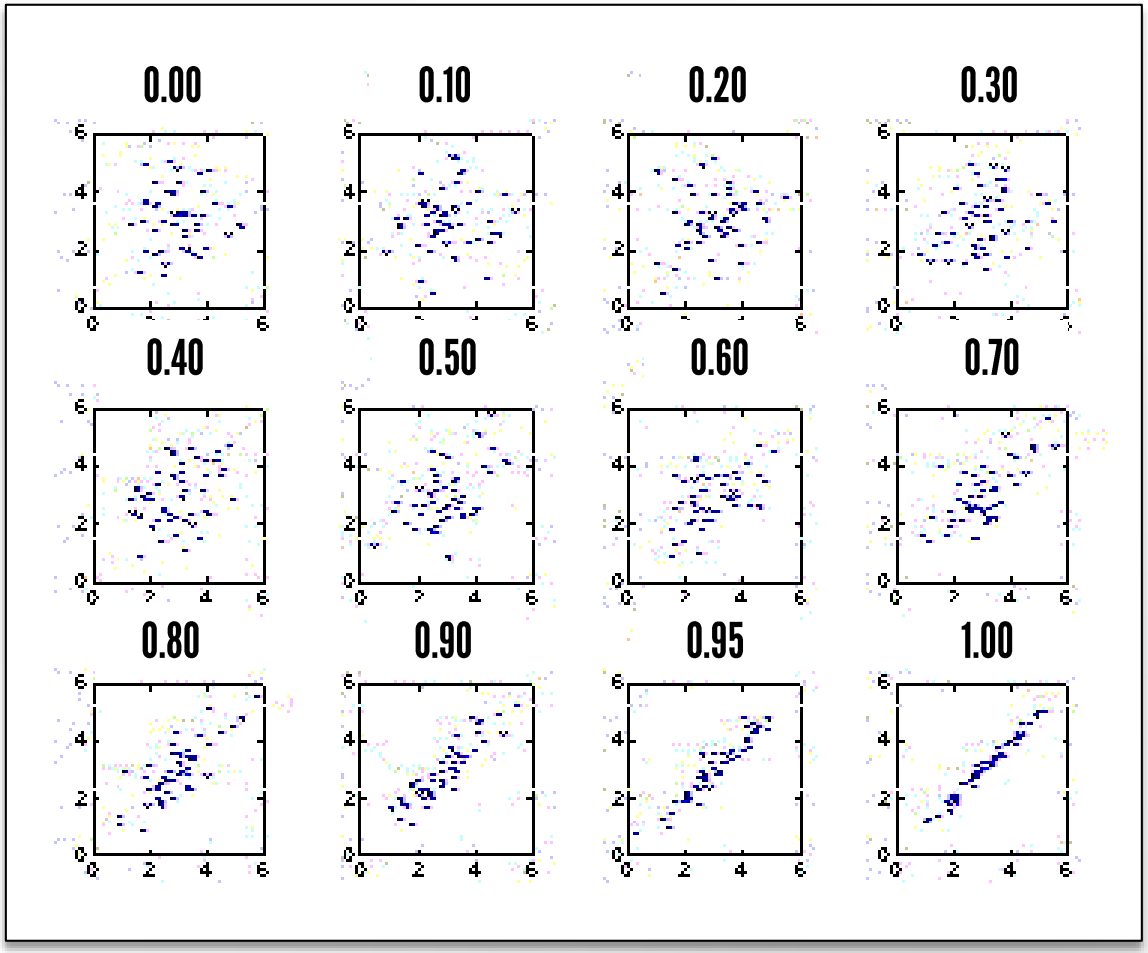
- ±0.1 - 0.2= small effect
- ±0.3 - 0.4 = medium effect
- ±.5+ = large effect

Coefficient of determination, r^2

- By squaring the value of *r* you get the proportion of variance in one variable shared by the other.

Pearson

http://blog.majesticseo.com/wp-content/uploads/2013/04/fig9.gif

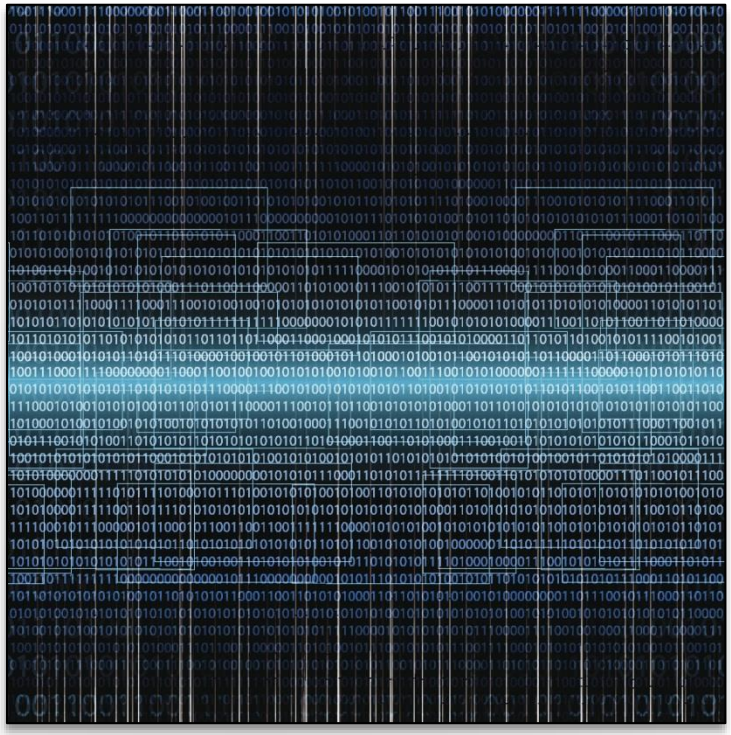


Correlation: Which one to choose?

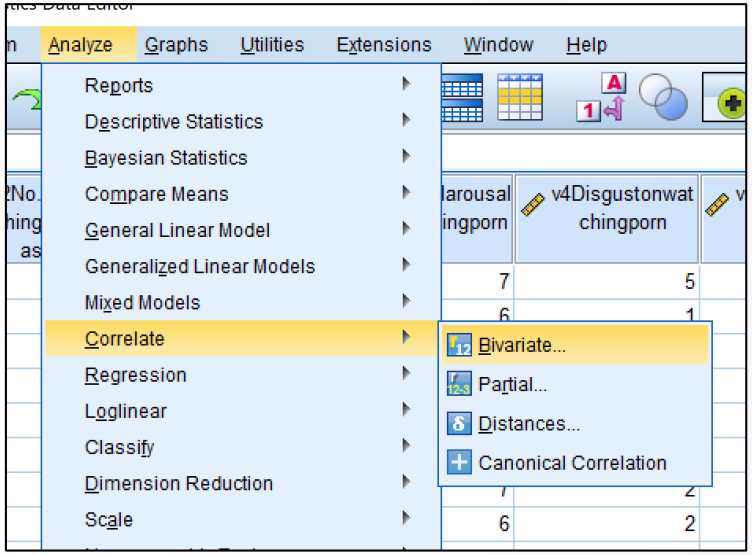
	Interval		Likert		Ordinal - many ranks		Ordinal - few ranks		Binary	
	Name of statistic	Calculated with...	Name of statistic	Calculated with...	Name of statistic	Calculated with...	Name of statistic	Calculated with...	Name of statistic	Calculated with...
Interval	Pearson	Pearson								
Likert	Pearson	Pearson	Pearson	Pearson						
Ordinal - many ranks	Spearman	Spearman	Spearman	Spearman	Spearman	Spearman				
Ordinal - few ranks	Kendall's	Kendall's	Kendall's	Kendall's	Kendall's	Kendall's	Kendall's	Kendall's		
Binary	point-bis	Pearson	point-bis	Pearson	Spearman	Spearman	Kendall's	Kendall's	Phi	Pearson

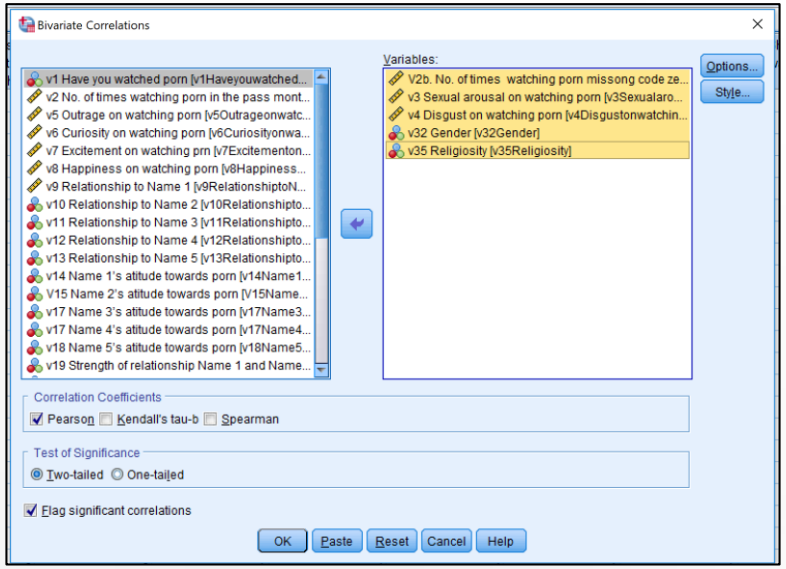
Correlation: Which one to choose?

- LESSON: Pearson is generally (i.e. most of the time) the correct correlation coefficient statistic to use.
- If you aren't sure what correlation coefficient to use then check your textbook or Google.



Example


Analyze
Graphs
Utilities
Extensions
Window
Help
Reports
Descriptive Statistics
Bayesian Statistics
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Dimension Reduction
Scale
Bivariate...
Partial...
Distances...
Canonical Correlation
v4Disgustonwatchingporn
7
5
6
1
6
2


Bivariate Correlations
Variables:
v1 Have you watched porn [v1Haveyouwatched...
v2 No. of times watching porn in the pass mont...
v5 Outrage on watching porn [v5Outrageonwatc...
v6 Curiosity on watching porn [v6Curiosityonwa...
v7 Excitement on watching prn [v7Excitementon...
v8 Happiness on watching porn [v8Happiness...
v9 Relationship to Name 1 [v9RelationshiptoN...
v10 Relationship to Name 2 [v10Relationshipto...
v11 Relationship to Name 3 [v11Relationshipto...
v12 Relationship to Name 4 [v12Relationshipto...
v13 Relationship to Name 5 [v13Relationshipto...
v14 Name 1's atitude towards porn [v14Name1...
V15 Name 2's atitude towards porn [V15Name...
v17 Name 3's atitude towards porn [v17Name3...
v17 Name 4's atitude towards porn [v17Name4...
v18 Name 5's atitude towards porn [v18Name5...
v19 Strength of relationship Name 1 and Name...
V2b. No. of times watching porn missong code ze...
v3 Sexual arousal on watching porn [v3Sexualaro...
v4 Disgust on watching porn [v4Disgustonwatchin...
v32 Gender [v32Gender]
v35 Religiosity [v35Religiosity]
Options...
Style...
Correlation Coefficients
Pearson
Kendall's tau-b
Spearman
Test of Significance
Two-tailed
One-tailed
Flag significant correlations
OK
Paste
Reset
Cancel
Help

Example

Correlation coefficients

- Significant weak to moderate positive correlation between arousal and number of times watching porn (r=0.280, p<0.001)
- Significant weak to moderate negative correlation between disgust and being female, and number of times watching porn (r=-0.292 and -0.240, p<0.001)
- No significant relationship between religiosity and number times watching porn (r=0.044, p=0.289)

Correlations

		V2b. No. of times watching porn missong code zero	v3 Sexual arousal on watching porn	v4 Disgust on watching porn	v32 Gender	v35 Religiosity
V2b. No. of times watching porn missong code zero	Pearson Correlation	1	.280**	-.292**	-.240**	.044
	Sig. (2-tailed)		.000	.000	.000	.289
	N	596	596	596	596	596
v3 Sexual arousal on watching porn	Pearson Correlation	.280**	1	-.396**	-.424**	-.063
	Sig. (2-tailed)	.000		.000	.000	.127
	N	596	596	596	596	596
v4 Disgust on watching porn	Pearson Correlation	-.292**	-.396**	1	.374**	.266**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	596	596	596	596	596
v32 Gender	Pearson Correlation	-.240**	-.424**	.374**	1	.101*
	Sig. (2-tailed)	.000	.000	.000		.013
	N	596	596	596	596	596
v35 Religiosity	Pearson Correlation	.044	-.063	.266**	.101*	1
	Sig. (2-tailed)	.289	.127	.000	.013	
	N	596	596	596	596	596

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

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

T-tests comparison of means

- T-test comparision of means:
  - Is used to compare the mean value of a dependent variable in two groups,
  - Generally these two groups are an experimental group and a control group.

T-tests comparison of means

- T-test comparision of means:
  - Paired sample t-test is used when the experimental and control group are the same individuals (e.g. pre and post tests on the same subjects).
  - Independent sample t-tests are used when there are different individuals in the two groups. It has more 'degrees of freedom' – meaning 'more information'.

T-tests comparison of means

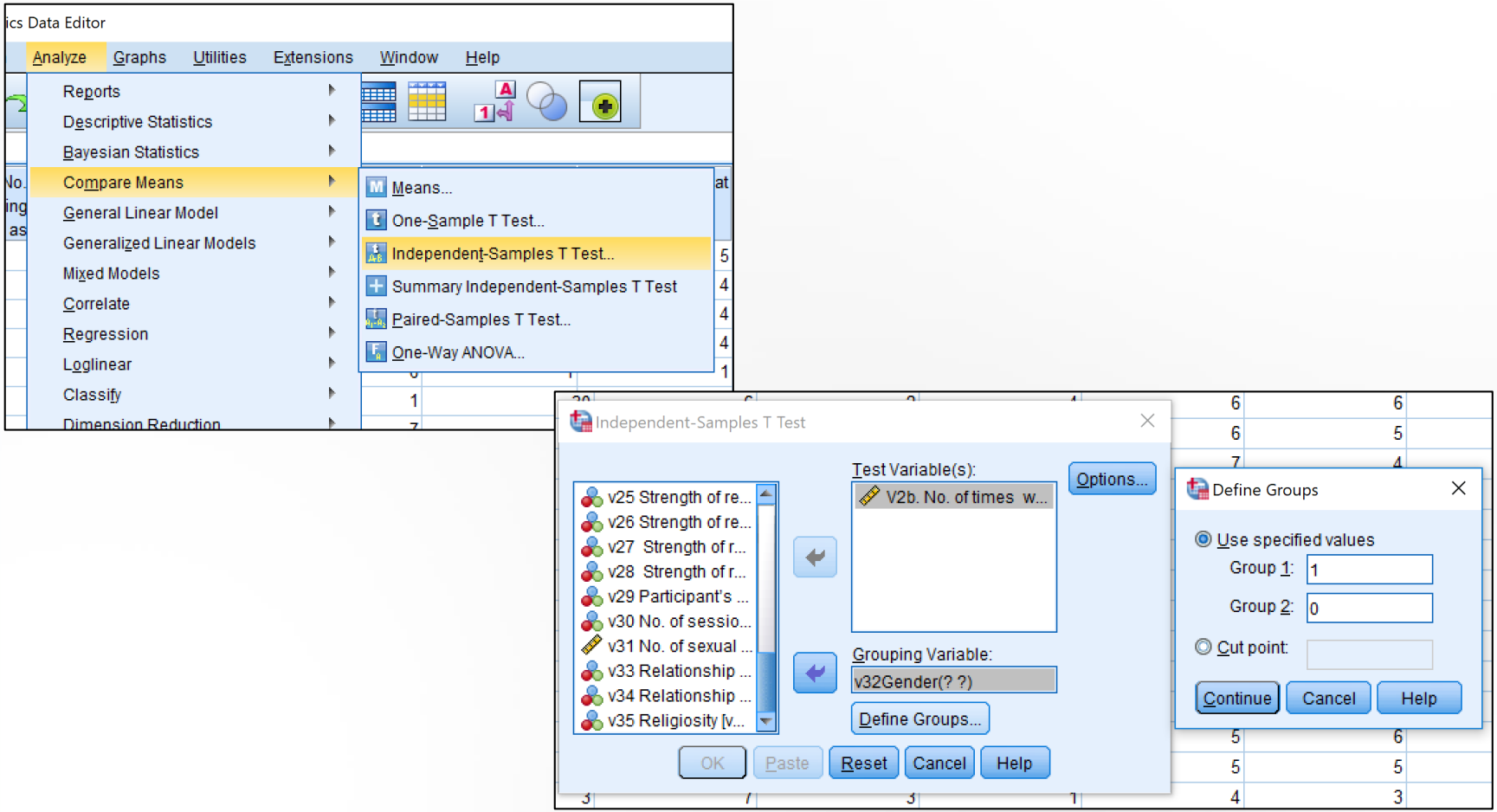
- T-test comparision of means:
  - For the independent sample t-test we test for equal variance of the dependent variable using Levane's Test. If this is significant, then equal variance is NOT assumed.

Rationale to the *t*-test

$$t = \frac{\text{observed difference between sample means} - \text{expected difference between population means (if null hypothesis is true)}}{\text{estimate of the standard error of the difference between two sample means}}$$

- Principle of the t-test is very similar to that of the z-score
- How? It is a score/standard error, which is then looked up on a statistical significance table.

Example


ics Data Editor
Analyze
Graphs
Utilities
Extensions
Window
Help
Reports
Descriptive Statistics
Bayesian Statistics
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Dimension Reduction
Means...
One-Sample T Test...
Independent-Samples T Test...
Summary Independent-Samples T Test
Paired-Samples T Test...
One-Way ANOVA...
Independent-Samples T Test
v25 Strength of re...
v26 Strength of re...
v27 Strength of r...
v28 Strength of r...
v29 Participant's ...
v30 No. of sessio...
v31 No. of sexual ...
v33 Relationship ...
v34 Relationship ...
v35 Religiosity [v...
Test Variable(s):
V2b. No. of times w...
Grouping Variable:
v32Gender(? ?)
Define Groups...
Options...
OK
Paste
Reset
Cancel
Help
Define Groups
Use specified values
Group 1:
1
Group 2:
0
Cut point:
Continue
Cancel
Help

Example

Comparison of means

- Females (1) = 4.29/month
- Males (0) = 13.23/month
- Levene's Test: $p<0.05$ so equal variance NOT assumed
- Read bottom row: $p<0.001$
- **Conclusion:** On average women watch porn 8.9 times less per month than men, and this difference is statistically significant ($p<0.001$)

T-Test

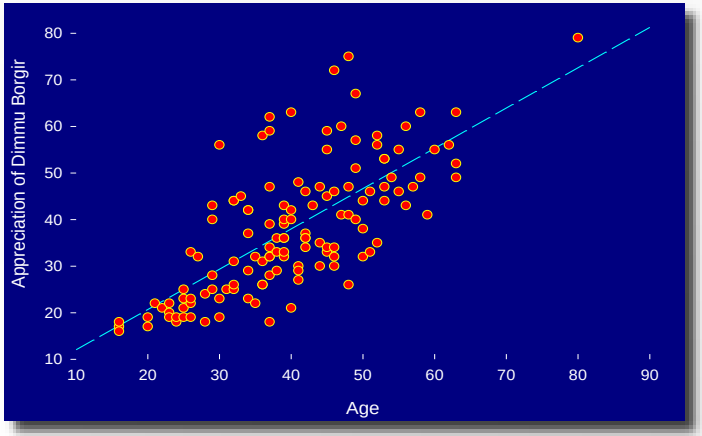
Group Statistics

	v32 Gender	N	Mean	Std. Deviation	Std. Error Mean
V2b. No. of times watching porn missong code zero	1	239	4.29	18.197	1.177
	0	357	13.23	17.487	.925

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
V2b. No. of times watching porn missong code zero	Equal variances assumed	19.329	.000	-6.020	594	.000	-8.944	1.486	-11.861	-6.026
	Equal variances not assumed			-5.973	496.392	.000	-8.944	1.497	-11.886	-6.002

Linear Regression



Summary

Regression:

- A model for predicting the value of one variable from the value of other variables.
- Linear Regression: continuous dependent variable
- Logistic Regression: binary dependent variable

Linear Regression

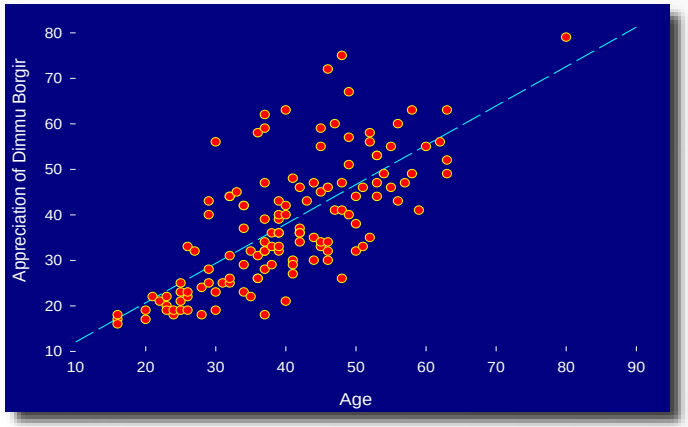
- Uses the equation of a straight line: $y = b_1x_1 + b_2x_2 + \ldots + b_0 + e$
- y = dependent variable
- x_1, x_2, etc = independent variable
- b_1 = slope of relationship between y and x_1
- b_0 = value of y when all $x = 0$

Summary

Basic assumptions

- Outcome is continuous (not binary)
- Predictors continuous or dichotomous (not categorical)
- Relationship between y and x is linear
- Cases in the sample are independent

What is Regression?



A way of predicting the value of one variable from another.

- It is a hypothetical model of the relationship between two variables.
- The model used is a linear one.
- Therefore, we describe the relationship using the equation of a straight line.

Describing a Straight Line

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

b_i

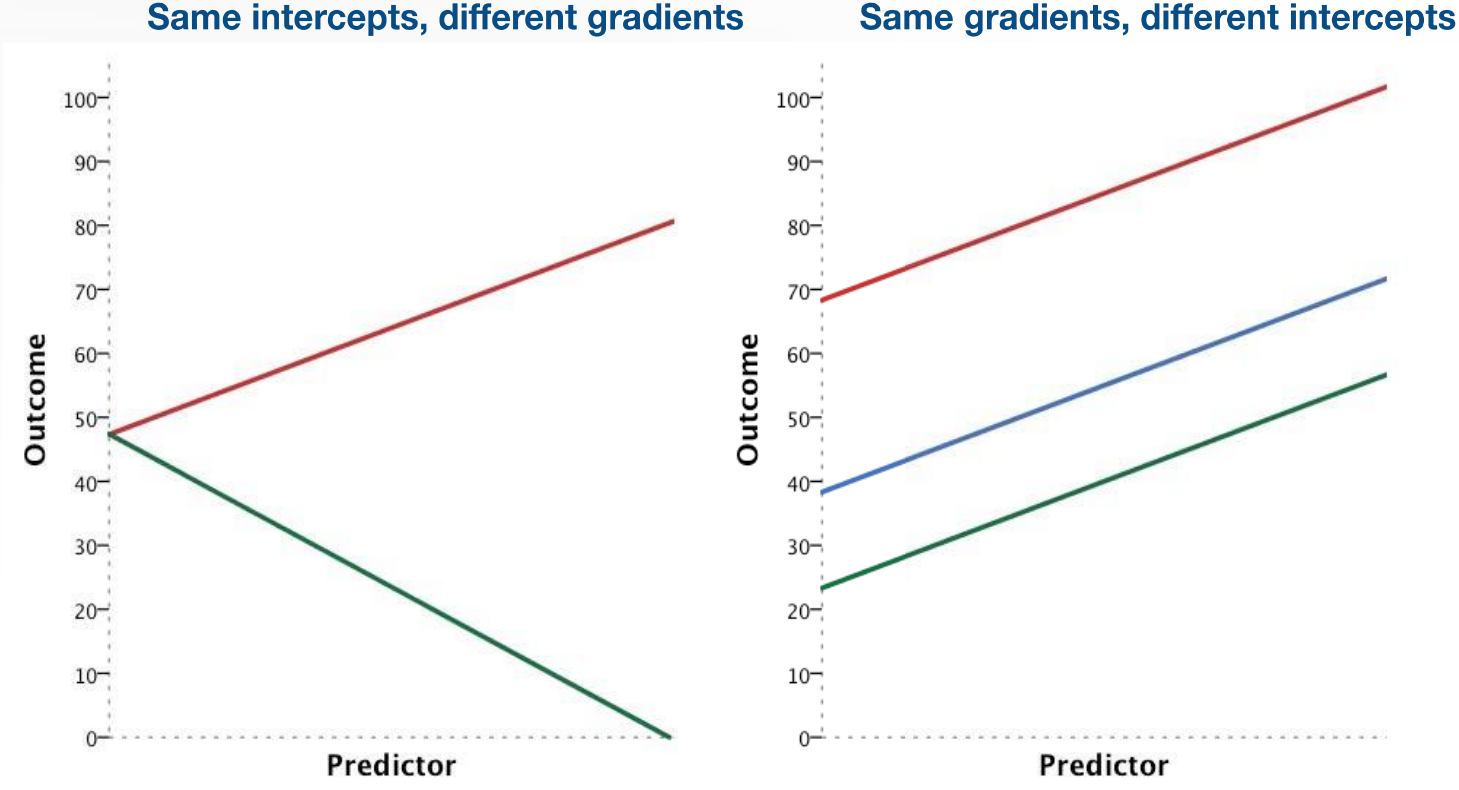
- Regression coefficient for the predictor
- Gradient (slope) of the regression line
- Direction/Strength of Relationship

b_0

- Intercept (value of Y when X = 0)
- Point at which the regression line crosses the Y-axis (ordinate)

Intercepts and Gradients

FIGURE 8.2
Lines that share the same intercept but have different gradients, and lines with the same gradients but different intercepts

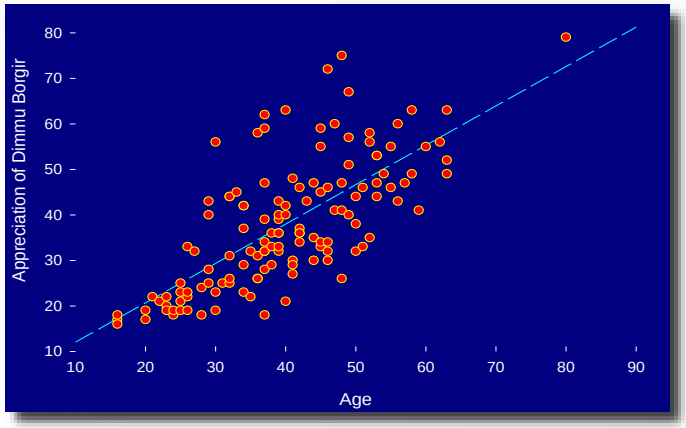


Testing the Model: R^2

R^2

- The proportion of variance accounted for by the regression model.
- The Pearson Correlation Coefficient Squared

Regression: An Example



A record company boss was interested in predicting album sales from advertising.

Data

- 200 different album releases

Outcome variable:

- Sales (CDs and Downloads) in the week after release

Predictor variable:

- The amount (in £s) spent promoting the album before release.

Step One: Graph the Data

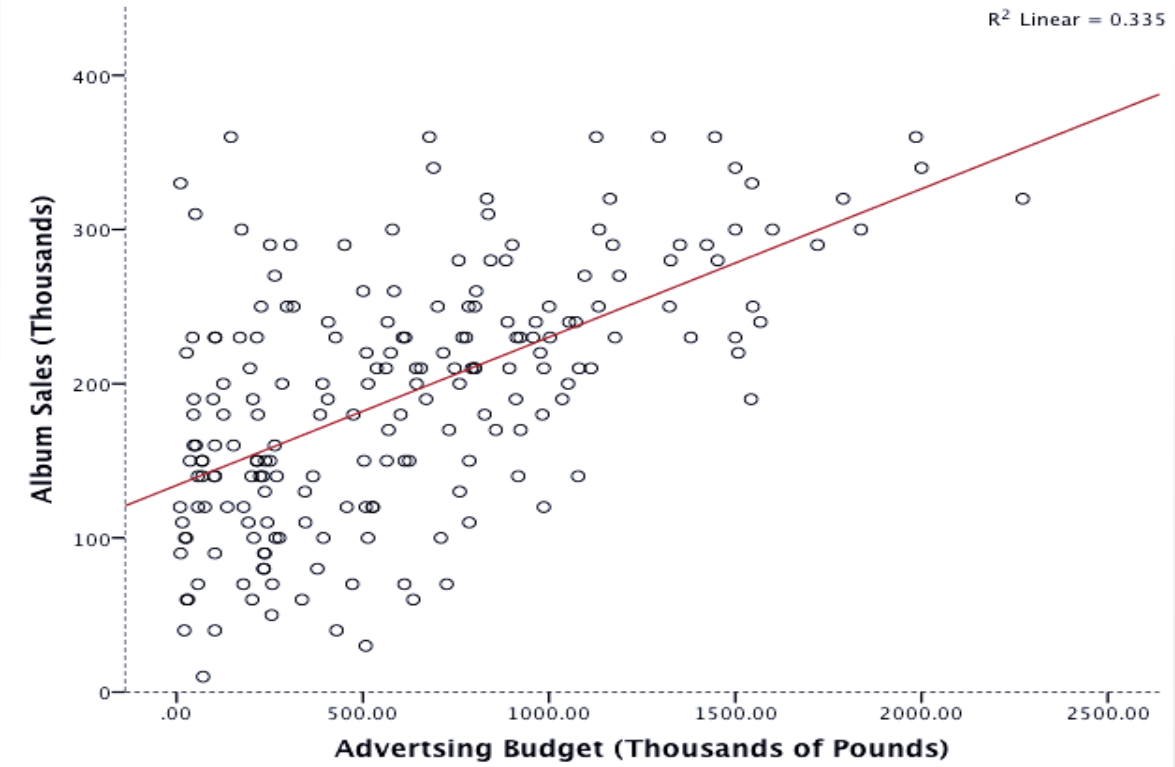


FIGURE 8.12 Scatterplot showing the relationship between album sales and the amount spent promoting the album

Regression Using SPSS

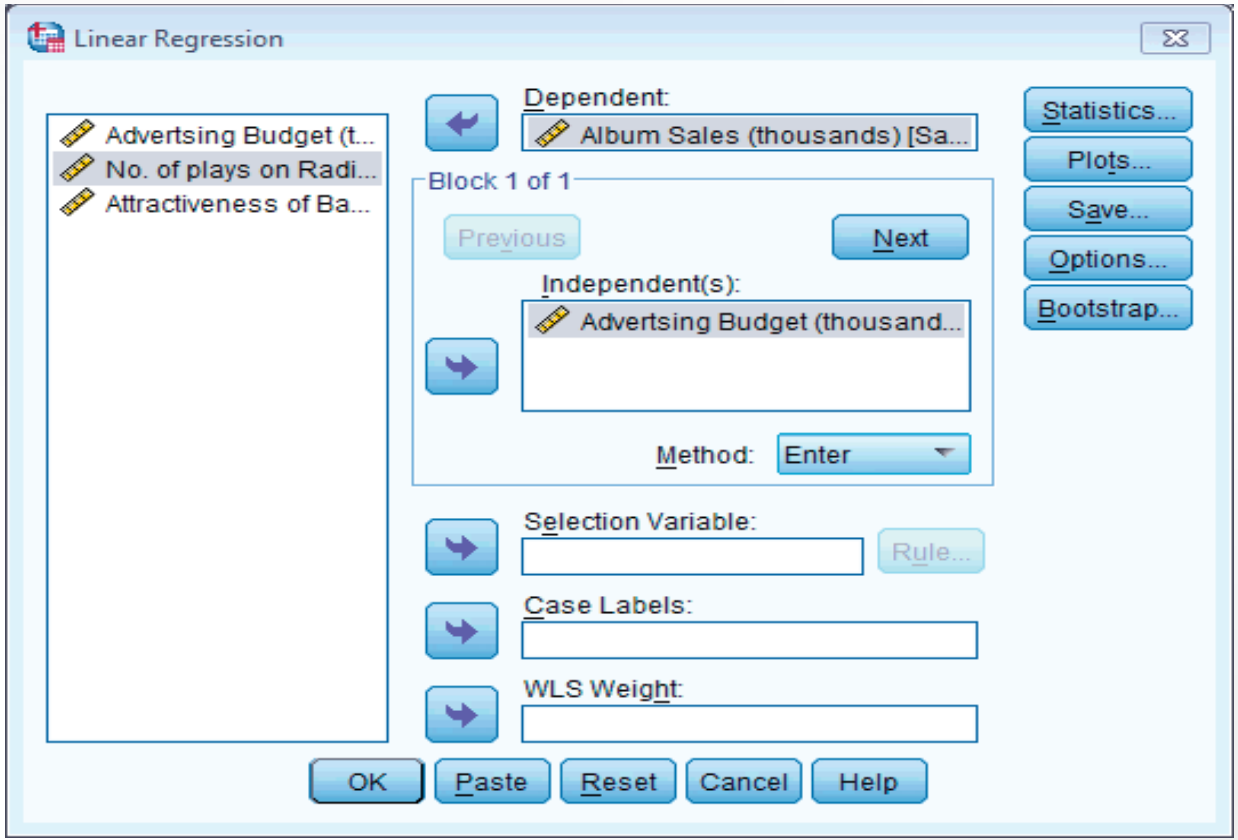


FIGURE 8.13
Main dialog box for regression

Output: Model Summary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.578[a]	.335	.331	65.9914

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

SPSS Output: Model Parameters

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	134.140	7.537		17.799	.000
	Advertsing Budget (thousands of pounds)	.096	.010	.578	9.979	.000

a. Dependent Variable: Album Sales (thousands)

$$\text{album sales}_i = b_0 + b_1 \text{advertising budget}_i$$

$$= 134.14 + \left(0.096 \times \text{ advertising budget}_i\right)$$

Using The Model

$$\begin{aligned}\text{album sales}_i &= 134.14 + (0.096 \times \text{advertising budget}_i) \\ &= 134.14 + (0.096 \times 100) \\ &= 143.74\end{aligned}$$

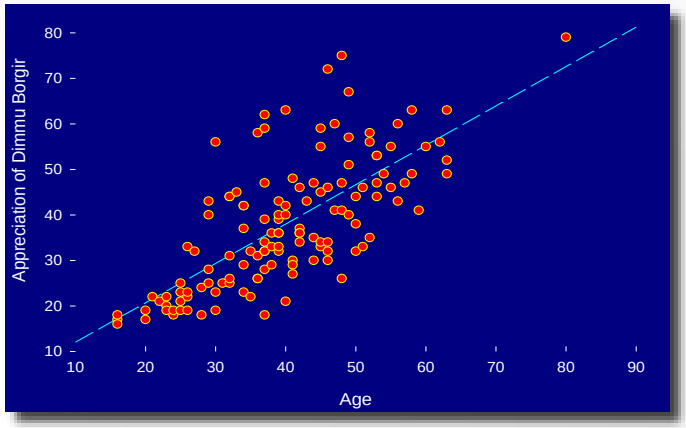
What is Multiple Regression?

Linear Regression is a model 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.

Regression: An Example



A record company boss was interested in predicting album sales from advertising.

Data

- 200 different album releases

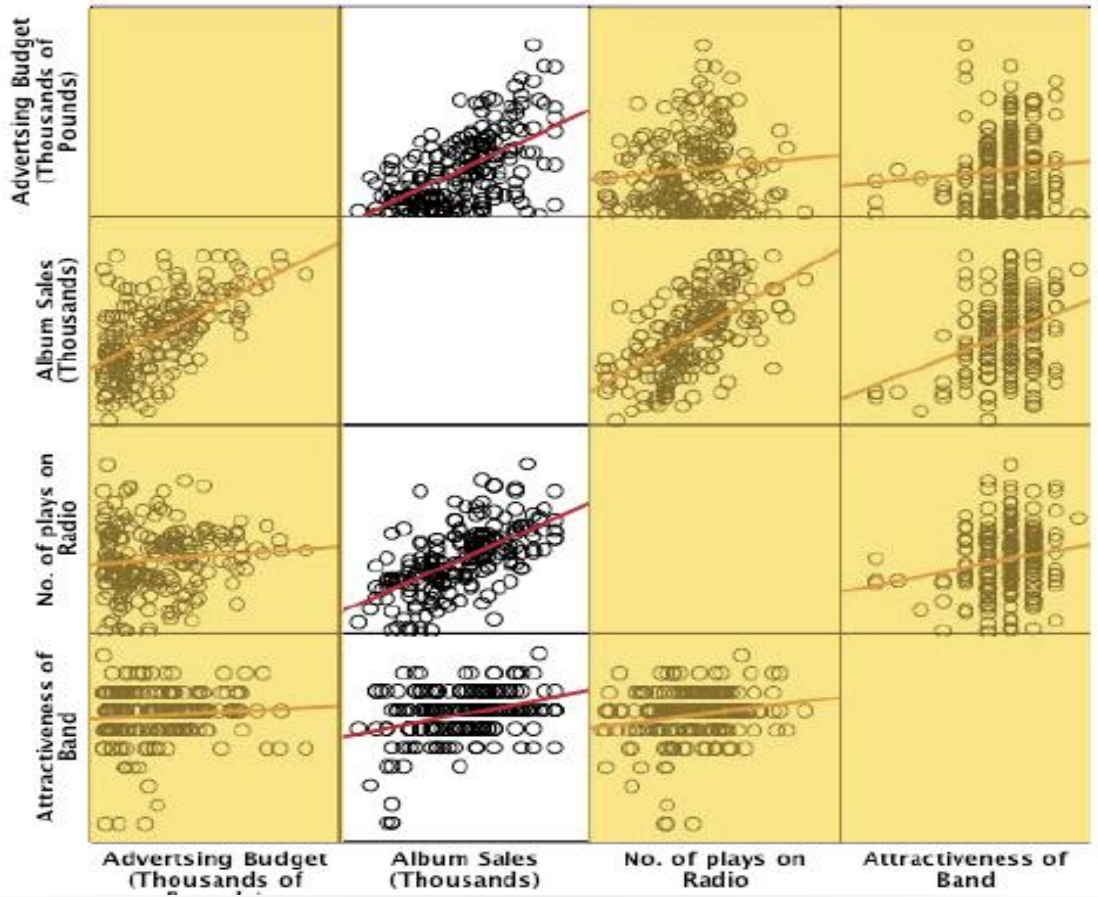
Outcome variable:

- Sales (CDs and Downloads) in the week after release

Predictor variables

- The amount (in £s) spent promoting the album before release (see last lecture)
- Number of plays on the radio (new variable)
- Attractiveness of band (new variable)

FIGURE 8.14
Matrix scatterplot of the relationships between advertising budget, airplay, and attractiveness of the band and album sales



Multiple Regression as an Equation

With multiple regression the relationship is described using a variation of the equation of a straight line.

$$y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n + \varepsilon_i$$

b_0

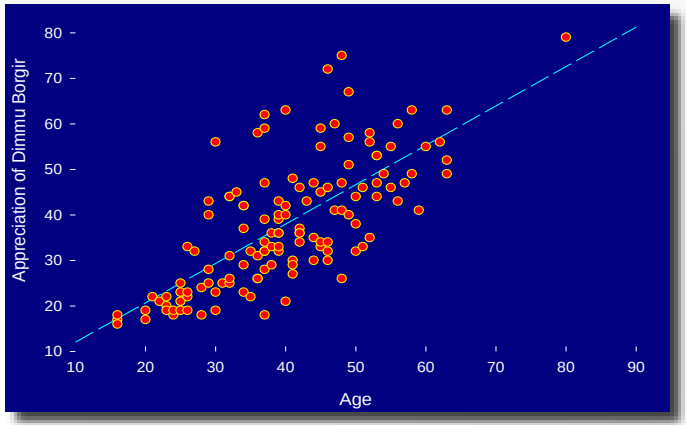
b_0 is the intercept.

The intercept is the value of the Y variable when all Xs = 0.

This is the point at which the regression plane crosses the Y-axis (vertical).

$$y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n + \varepsilon_i$$

Beta Values



b_1 is the regression coefficient for variable 1.

b_2 is the regression coefficient for variable 2.

b_n is the regression coefficient for n^{th} variable.

$$y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n + \varepsilon_i$$

The Model with Two Predictors

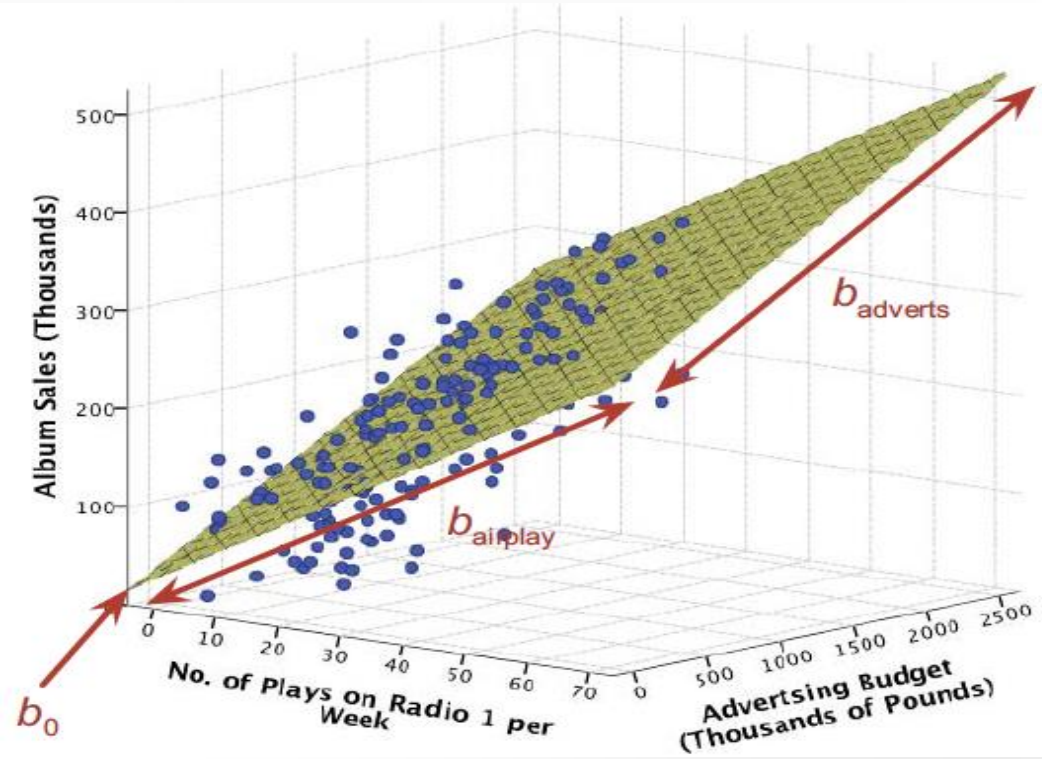


FIGURE 8.3 Scatterplot of the relationship between album sales, advertising budget and radio play

Methods of Regression

Hierarchical:

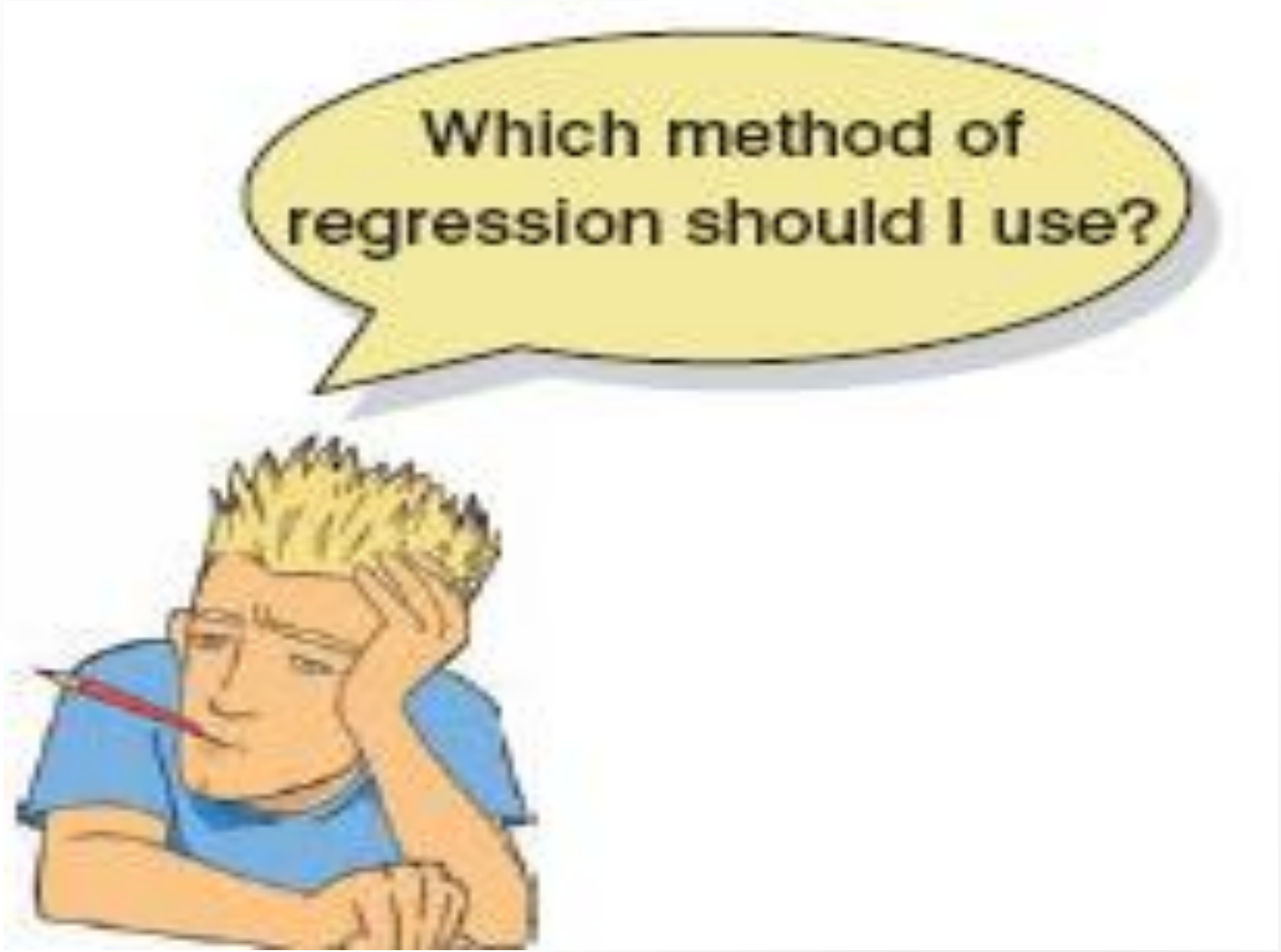
- Experimenter decides the order in which variables are entered into the model.

Forced Entry:

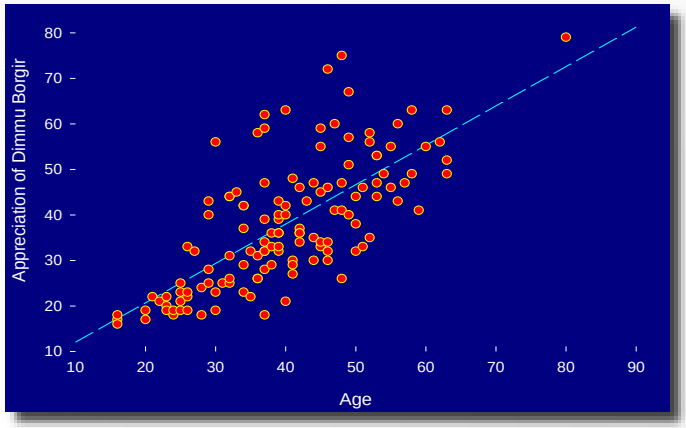
- All predictors are entered simultaneously.

Stepwise:

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


Which method of regression should I use?

Hierarchical Regression



Known predictors (based on past research) are entered into the regression model first.

New predictors are then entered in a separate step/block.

Experimenter makes the decisions.

Hierarchical Regression

It is the best method:

- Based on theory testing.
- You can see the unique predictive influence of a new variable on the outcome because known predictors are held constant in the model.

Bad Point:

- Relies on the experimenter knowing what they're doing!

Forced Entry Regression

All variables are entered into the model simultaneously.

The results obtained depend on the variables entered into the model.

- It is important, therefore, to have good theoretical reasons for including a particular variable.

Stepwise Regression

Variables are entered into the model based on mathematical criteria.

Computer selects variables in steps.

- Can do 'Forward selection', where computer adds variables one by one.
- Can do 'Backward selection', where computer adds all variables, then drops them out one by one until only significant variables are left.

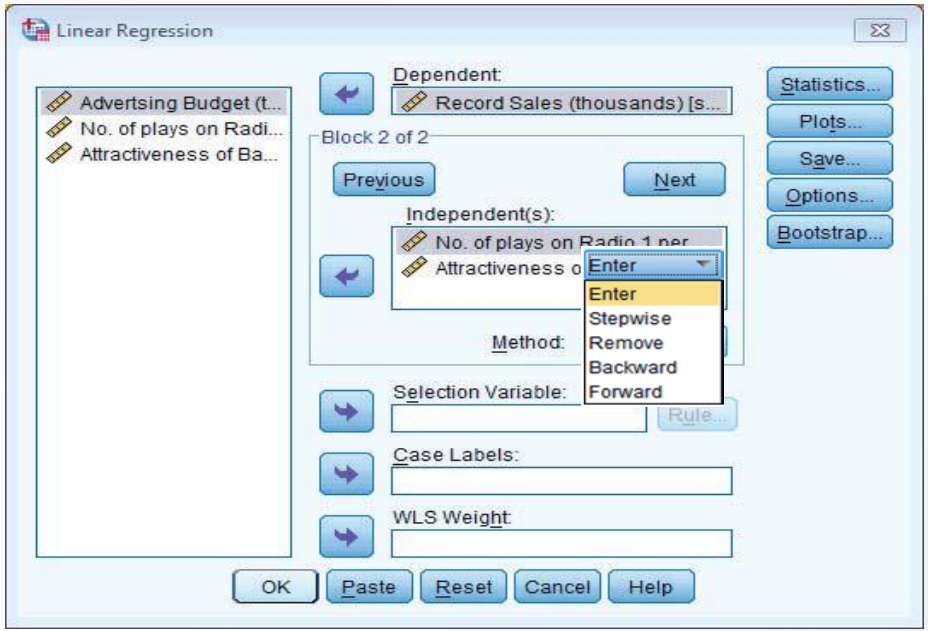
Problems with Stepwise Methods

Rely on a mathematical criterion.

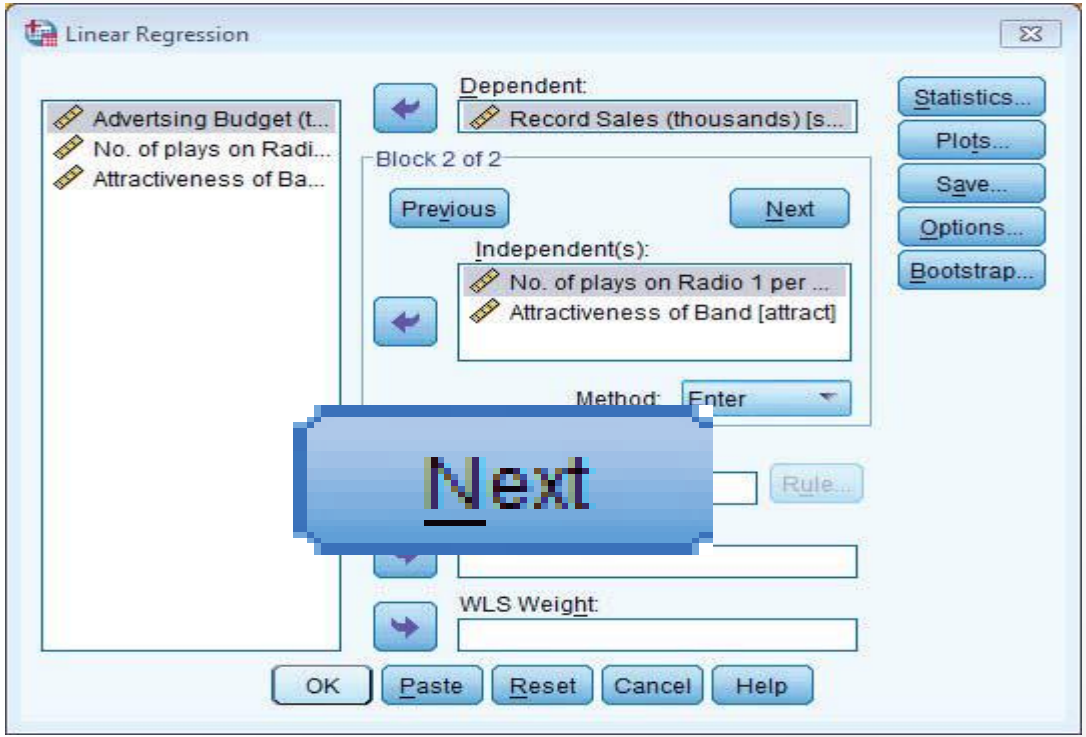
- Variable selection may depend upon only slight differences in significance of highly correlated variables.
- These slight numerical differences can lead to major theoretical differences.

Should be used only for exploration.

Doing Multiple Regression



Doing Multiple Regression



Output: betas

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	134.140	7.537		17.79	.000	119.278	149.002
	Advertsing Budget (Thousands of Pounds)	.096	.010	.578	9.97	.000	.077	.115
2	(Constant)	-26.613	17.350		-1.53	.127	-60.830	7.604
	Advertsing Budget (Thousands of Pounds)	.085	.007	.511	12.26	.000	.071	.099
	No. of plays on Radio	3.367	.278	.512	12.12	.000	2.820	3.915
	Attractiveness of Band	11.086	2.438	.192	4.54	.000	6.279	15.894

a. Dependent Variable: Album Sales (Thousands

How to Interpret Coefficients

Coefficients (b):

- the change in the outcome associated with a unit change in the predictor.

Standardised beta values:

- tell us the same but expressed as standard deviations of the normalized variables.

Beta Values

$b_1 = 0.087$.

- So, as advertising increases by £1, album sales increase by 0.087 units.

$b_2 = 3367$.

- So, each time (per week) a song is played on the radio its sales increase by 3367 units.

Constructing a Model

$$y = b_0 + b_1 X_1 + b_2 X_2$$

$$\text{Sales} = 41124 + 0.087\,\text{Adverts} + 3367\,\text{plays}$$

£1 Million Advert ; 15 plays

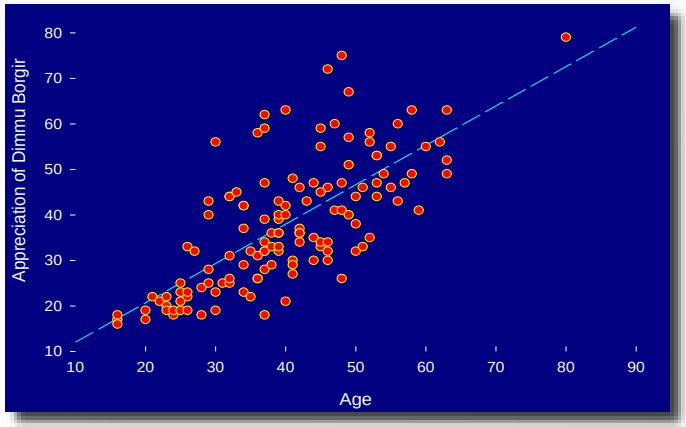
$$\begin{aligned}\text{Sales} &= 41124 + (0.087 \times 1{,}000{,}000) + (3367 \times 15) \\ &= 41124 + 87000 + 50505 \\ &= 178629\end{aligned}$$

Reporting the Model

Table 8.2: Linear regression model of predictors of album sales.

	b	SE	p
Attractiveness	11.09	2.22	.001
Plays on BBC Radio 1	3.37	0.32	.001
Advertising Budget	0.09	0.01	.001
Constant	134.14	7.95	.097
R^2	.34		

Generalization



When we run 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 stops us generalizing conclusions to our target population.

Straightforward Assumptions

Variable Type:

- Outcome must be continuous
- Predictors can be continuous or dichotomous.

Linearity:

- The relationship we model is, in reality, linear.

Independence:

- All values of the outcome should come from a different person.

The More Tricky Assumptions (non-examinable)

No Multicollinearity:

- Predictors must not be highly correlated.

Homoscedasticity:

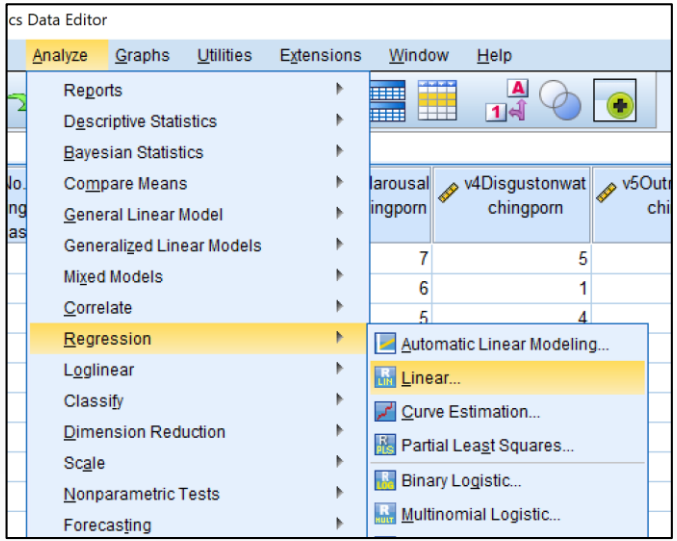
- For each value of the predictors the variance of the error term should be constant.

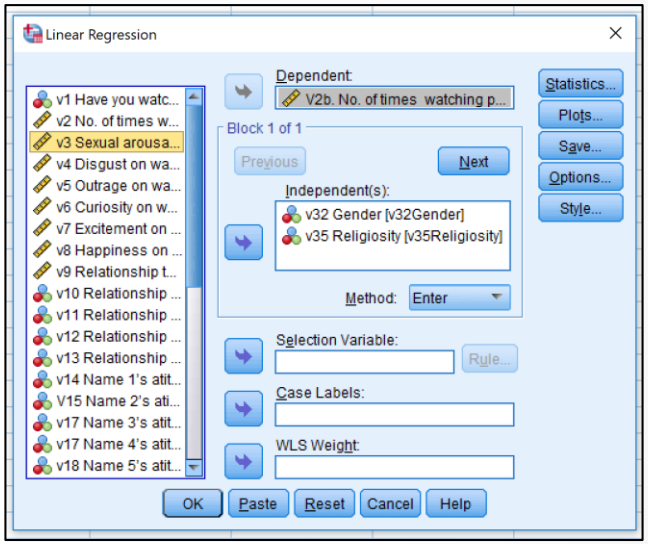
Independent Errors:

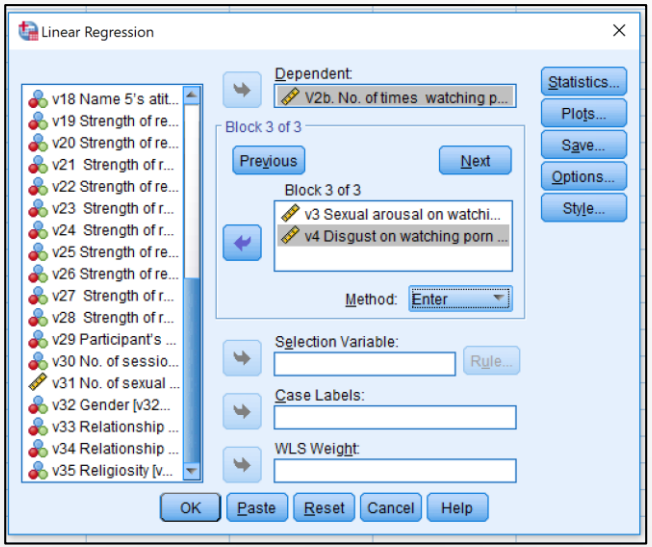
- For any pair of observations, the error terms should be uncorrelated.

Normally-distributed Errors

Example


cs Data Editor
Analyze
Graphs
Utilities
Extensions
Window
Help
Reports
Descriptive Statistics
Bayesian Statistics
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Automatic Linear Modeling...
Linear...
Curve Estimation...
Partial Least Squares...
Binary Logistic...
Multinomial Logistic...
v4Disgustonwatchingporn
v5Outchi
7
5
6
1
5
4


Linear Regression
v1 Have you watc...
v2 No. of times w...
v3 Sexual arousa...
v4 Disgust on wa...
v5 Outrage on wa...
v6 Curiosity on w...
v7 Excitement on ...
v8 Happiness on ...
v9 Relationship t...
v10 Relationship ...
v11 Relationship ...
v12 Relationship ...
v13 Relationship ...
v14 Name 1's atit...
V15 Name 2's ati...
v17 Name 3's atit...
v17 Name 4's atit...
v18 Name 5's atit...
Dependent:
V2b. No. of times watching p...
Block 1 of 1
Previous
Next
Independent(s):
v32 Gender [v32Gender]
v35 Religiosity [v35Religiosity]
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
Style...
OK
Paste
Reset
Cancel
Help


Linear Regression
v18 Name 5's atit...
v19 Strength of re...
v20 Strength of re...
v21 Strength of r...
v22 Strength of re...
v23 Strength of r...
v24 Strength of r...
v25 Strength of re...
v26 Strength of re...
v27 Strength of r...
v28 Strength of r...
v29 Participant's ...
v30 No. of sessio...
v31 No. of sexual ...
v32 Gender [v32...
v33 Relationship ...
v34 Relationship ...
v35 Religiosity [v...
Dependent:
V2b. No. of times watching p...
Block 3 of 3
Previous
Next
Block 3 of 3
v3 Sexual arousal on watchi...
v4 Disgust on watching porn ...
Method:
Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
Style...
OK
Paste
Reset
Cancel
Help

Example

Linear Regression

Model 1 shows that

- gender and religiosity explain about 6% of variance in frequency of watching porn.
- women watch on average 9.2 times less than men. In Model 1, religiosity is not significant

Model 2 shows that

- the addition of arousal and disgust increase the r-square to approximately 13%.
- Once we control for sexual arousal and disgust, the gender difference becomes much less, with women watching 3.8 less times than men.
- In model 2 religiosity is significant, with each unit (out of 7) of religiosity being associated with 1.2 more times watching porn in a month.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.249[a]	.062	.059	17.746
2	.373[b]	.139	.133	17.030

a. Predictors: (Constant), v35 Religiosity, v32 Gender

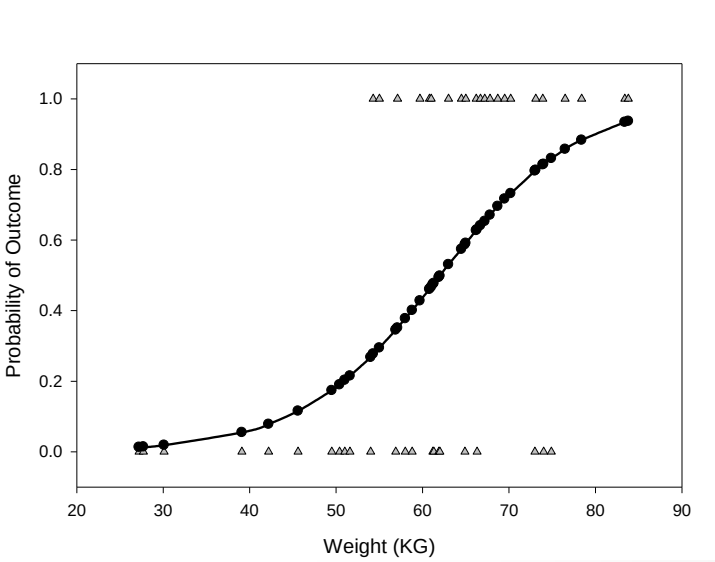
b. Predictors: (Constant), v35 Religiosity, v32 Gender, v3 Sexual arousal on watching porn, v4 Disgust on watching porn

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	11.229	1.499		7.490	.000
	v32 Gender	-9.203	1.491	-.247	-6.173	.000
	v35 Religiosity	.641	.374	.069	1.715	.087
2	(Constant)	5.064	3.612		1.402	.161
	v32 Gender	-3.818	1.622	-.102	-2.354	.019
	v35 Religiosity	1.157	.371	.124	3.118	.002
	v3 Sexual arousal on watching porn	1.725	.490	.155	3.523	.000
	v4 Disgust on watching porn	-2.229	.438	-.225	-5.085	.000

a. Dependent Variable: V2b. No. of times watching porn missong code zero

Logistic Regression



To predict an outcome variable that is categorical from one or more categorical or continuous predictor variables.

Used because having a categorical outcome variable violates the assumption of linearity in normal regression.

With One Predictor

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1X_1i)}}$$

Outcome

- We predict the *probability* of the outcome occurring

b_0 and b_0

- Can be thought of in much the same way as multiple regression
- Note the normal regression equation forms part of the logistic regression equation

With Several Predictor

$$P(Y) = \frac{1}{1+e^{-(b_0+b_1X_1i+b_2X_2i+\ldots+b_nX_{ni})}}$$

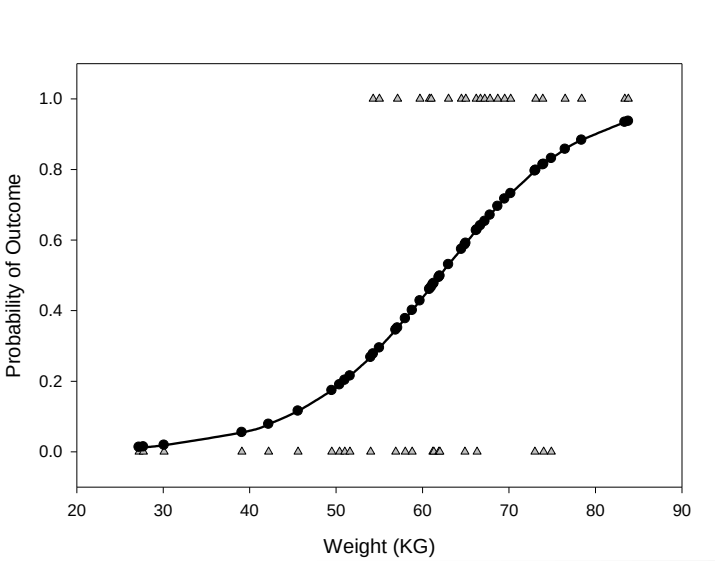
Outcome

- We still predict the *probability* of the outcome occurring

Differences

- Note the multiple regression equation forms part of the logistic regression equation
- This part of the equation expands to accommodate additional predictors

An Example



Predictors of a treatment intervention.

Participants

- 113 adults with a medical problem

Outcome:

- Cured (1) or not cured (0).

Predictors:

- Intervention: intervention or no treatment.

Output: Model Summary

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	144.158[a]	.084	.113

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Output: Model Summary

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Intervention(1)	1.229	.400	9.447	1	.002	3.417	1.561	7.480
	Constant	−.288	.270	1.135	1	.287	.750		

a. Variable(s) entered on step 1: Intervention.

The odds ratio: exp(*B*)

$$\text{Odds ratio} = \frac{\text{Odds after a unit change in the predictor}}{\text{Original odds}}$$

Indicates the change in odds resulting from a unit change in the predictor.

- OR > 1: Predictor ↑, Probability of outcome occurring ↑.
- OR < 1: Predictor ↑, Probability of outcome occurring ↓.

Variables in the Equation

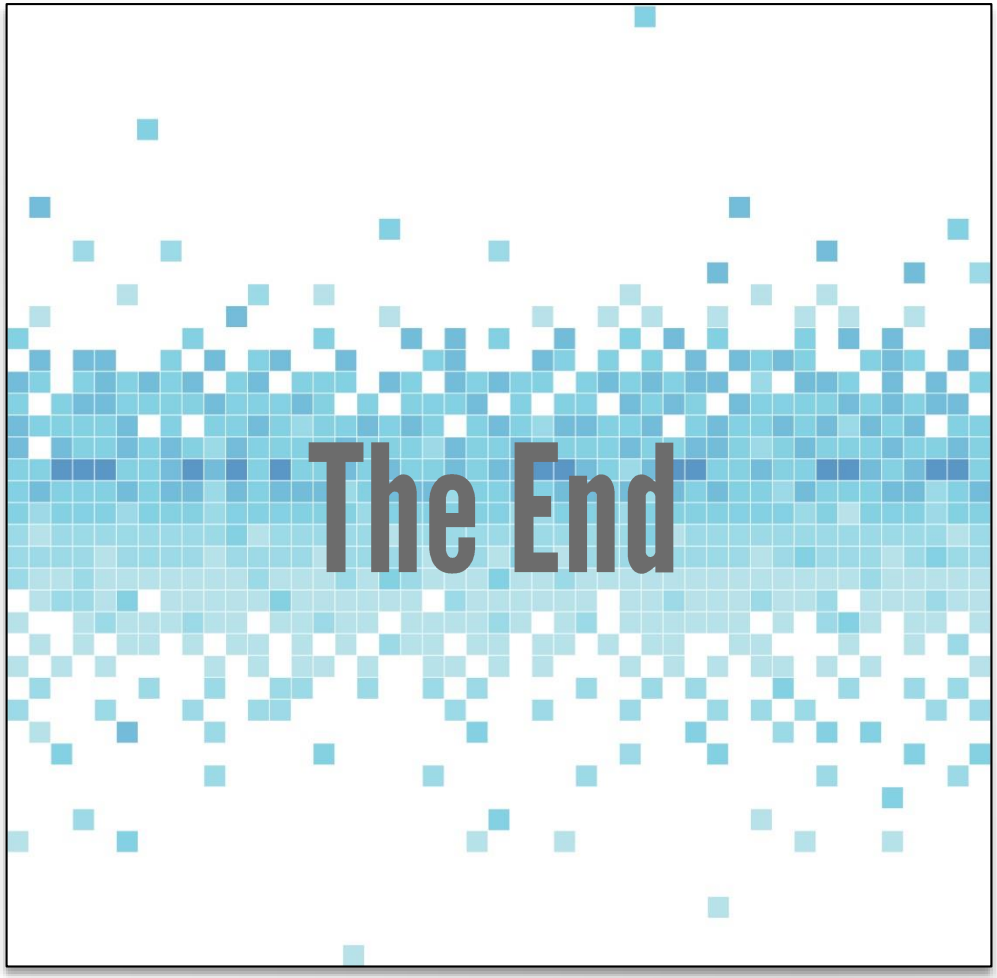
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Intervention(1)	1.229	.400	9.447	1	.002	3.417	1.561	7.480
	Constant	−.288	.270	1.135	1	.287	.750		

a. Variable(s) entered on step 1: Intervention.

Reporting the Analysis

Table 8.3: XXXXX.

	B	SE	p	Odds
Intervention	1.23	0.40	.002	3.42
Constant	-0.29	0.27	0.290	0.75
R^2 (Nagelkerke)	0.11			


The End