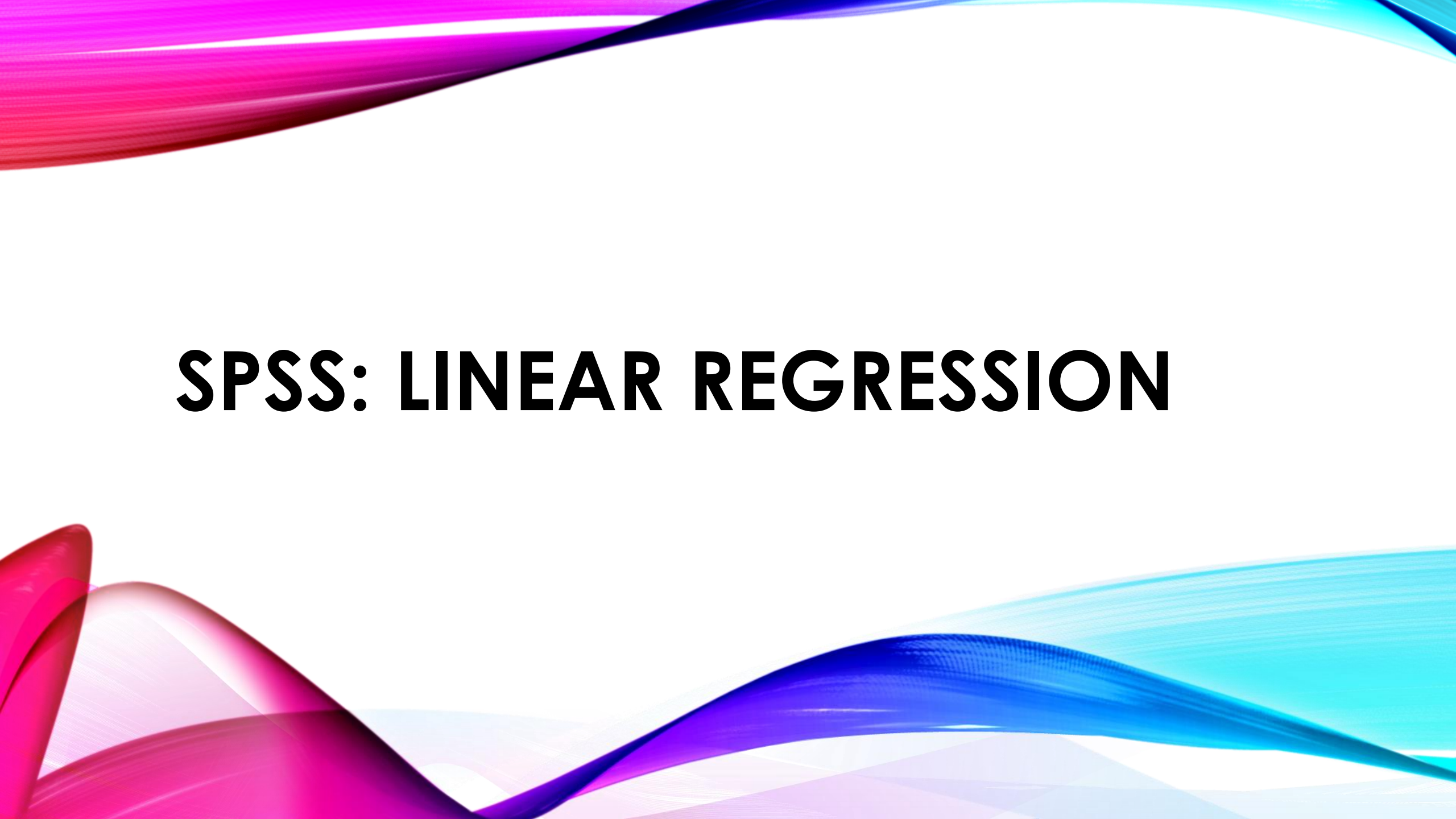




SPSS: LINEAR REGRESSION



LINEAR REGRESSION

- An extension of correlation when looking to **predict** scores on a DV
- Explores **prediction** rather than relationships
 - i.e., causality can be inferred in some cases (only if variables have been manipulated..)
- Linear regression also tells use **by how much** one variable will change if the other variable changes.
 - Because of this, we can theoretically predict an estimate of new 'scores' based on one data collection point (used often in applied research).

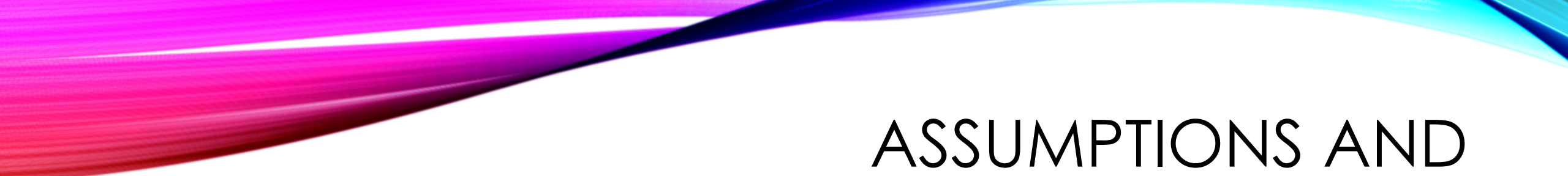


LINEAR REGRESSION

- ONE predictor variable (IV) and ONE criterion variable (DV)
 - In regression terminology, we refer to the predictor as x and the criterion as y .
- Less used in lab-based experimental designs and far more applied designs

ASSUMPTIONS AND REQUIREMENTS FOR DATA

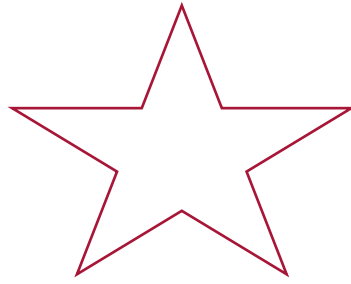
- Regression analyses are fussy! Makes many assumptions about data (and not all that forgiving if these are violated).
1. Continuous data is best (interval or ratio)
 2. Normal distribution of data (normality)
 - For any fixed value of x , y is normally distributed
 3. Linear relationship between IV and DV (linearity)
 4. Homoscedasticity
 - The variance of residual is the same for any value of x
 - Check using residuals scatterplot



ASSUMPTIONS AND REQUIREMENTS FOR DATA

5. Outliers

- Very sensitive to outliers so always check during cleaning and screening (but do NOT be tempted to just remove outliers!)
- Standard residuals should not be below -3.3 / above 3.3



STUDY EXAMPLE

- Predicting final year exam marks based on previous mock exam scores.
- We need:
 - One continuous DV = Final year exam mark
 - One continuous IV = Mock exam mark
- Note:
 - There are often multiple things you need to do to/with data to get them to the point of being able to be analysed!

RESULTS

- Main elements:
 - Regression model (ANOVA table)
 - **$F(1, 13) = 322.58, p < .001$**
 - R^2 (model summary table)
 - **0.96**
 - Prediction of IV (coefficients table – standardised co-efficients)
 - **$\beta = 0.98, p < .001$**

Here we can see we have a significant regression equation. The ANOVA results get reported as any other ANOVA.

We need to convert this to a %.
This suggests that 96% of the variation in final exam marks can be explained by mock exam marks

We have a significant predictor; mock exam marks significantly predict final exam marks.

We can use these values to state by how much SD in final marks increase (.98) when mock exam marks increase by 1 SD.

RESULTS

- Also of interest to us...
 - Confidence intervals:
 - The lower and upper CI's were 0.78 and 1.00

The confidence limits here are narrow, so we can state that it is 95% likely that the population slope is between these two values

- Plus....
 - Intercept (a) = 8.45
 - Slope (b) = 0.89

INTERPRETING RESULTS

A simple linear regression was carried out to determine whether finals marks on an exam could be predicted by marks scored on a mock exam. A significant regression equation was found ($F(1, 13) = 322.58, p < .001$). The R^2 value was 0.96, suggesting that 96% of the variation in the final exam marks can be explained by the mock exam marks. For every one mark increase in SD of the mock exam scores, final exam marks increased by .98 SD's. This can be interpreted further in that for every 15.13 points that mock exam scores increased, final exam scores increased by 13.45. The range between confidence limits were narrow, highlighting that it is 95% likely that the population slope is between 0.78 and 1.00.

PREDICTING NEW CASES

- We can use the information provided in your linear regression output to predict new cases using the regression equation:

$$Y = bX + a$$

Where:

Y = final exam mark

X = mock exam mark

a = 8.45

b = 0.89

To predict a new case, you just need a value for X.