

Multiple Regression

Dr. Kate Williams

Multiple Regression design

- Used to explore the relationship between one DV and **more than one** IVs (or predictor variables).
 - Standard
 - Hierarchical (Enters in Blocks)
 - Stepwise (to be avoided! Adds/deletes predictors based on strength)
- Ideal for investigating complex real-life research questions
 - Less used in lab-based experimental designs
 - Ideal for longitudinal prospective designs
- Often used to test how “good” a theory is
 - To what extent do the variable identified by a theory explain a given behaviour
 - Can add additional variables/new element to theory and test this

Why do Multiple Regression?

- Research questions:
 - Which of a **range** of variables are most strongly related to a specific variable of interest?
 - Do these variables do better than chance or simple mean scores at **predicting** scores on a variable of interest?
 - How might these variables **combine** to predict a variable?
- Multiple variables:
 - Could just carry out many simple correlations
 - Provide several Bivariate associations (2 variables correlated)
 - Still only a non-causal correlation produced
 - Will not indicate our overall success with all variables used

Why do Multiple Regression?

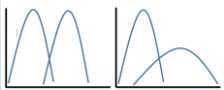
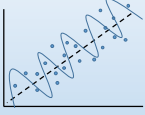
- Important extension of 'regression' over 'correlation':
 - **Combined** association effect of predictor variables
 - **Multiple** regression
 - Treating predictors (separate variables) as a combined variable in relation to the criterion (DV)
 - How 'associated' they are as a whole
 - How 'predictive' of the criterion they are as a whole
 - NB: Will also look at each in isolation

Assumptions

- Continuous data
 - Criterion should be measured on a scale at either interval or ratio level.
 - Predictors should be measured on ratio, interval or ordinal scale.
- Normal distribution of data
 - Also check for outliers
- Linear relationships between predictors and criterion
 - Non-linear relationships will require transformations



Assumptions

- Homogeneity of variance (Homoscedasticity)
 - Two side-by-side graphs. The left one shows two bell curves of different widths. The right one shows a scatter plot with a dashed regression line and data points that are spread evenly around the line.
 - A scatter plot with a dashed regression line. The data points are spread more widely around the line as the value of the predictor variable increases, indicating a non-constant variance.
- Collinearity between IVs (multicollinearity)
 - Correlations between the predictor variables
 - Ideally should be avoided as it creates problems when trying to infer relative contribution of individual predictors

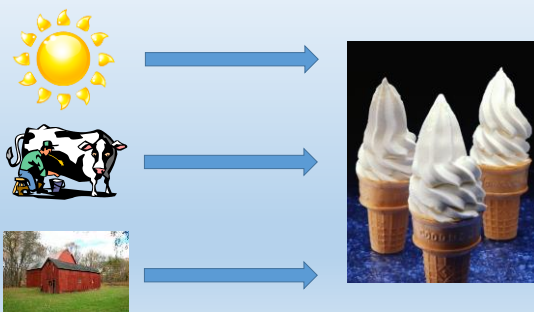
Assumptions

- Requires a large number of observations
 - The estimate of R depends on the number of predictor variables and number of participants
 - Depending on which book you read, the minimum no. of participants per predictor differs.
 - NB: would still need to consider the N for the power you require.
- How do you choose which variables to use as “predictors” in a regression design



- Regression can be used to analyse a more ‘causal’ relation between variables
- Regression analyses amount of change in the criterion based on change in the predictor variables
- Can be used to make **predictions** for the DV based on scores from a range of predictors

- Ice-cream sales example
 - What guides selection of predictors?



•Ice-cream sales example

- Was this “empirically driven good science?”
- Better approach if funded research project?
- Would you actually expect a “good fit” from this model?

Key Terminology

- Beta (standard Regression coefficients)
 - Is a measure of how strongly each predictor influences the criterion
 - Is measured in units of standard deviation (how wide data values are spread)
 - E.g., a beta value of 2.5 indicates a change of one standard deviation in the predictor will result in a change of 2.5 standard deviations in the criterion
- The higher the beta value, the greater the impact of the predictor on the criterion

Key Terminology

- When there is only one predictor, then beta is equivalent to the correlation coefficient between the predictor and the criterion.
 - This is the equivalence of the correlation between two variables
- When there is more than one predictor, you cannot compare the contribution of each predictor by comparing the correlation coefficients.
 - So, the **beta regression coefficient** is computed to assess the **strength of the relationship** between each predictor to the criterion

Multiple Regression

- Multiple Regression statistics of interest:
 - Multiple regression indicates how good a 'fit' all the predictors are in explaining the DV
 - Can assess the degree to which the DV can be predicted from the set of predictors
 - R = multiple correlation coefficient
 - A measure of the correlation between the observed value and predicted value of the criterion
 - (between 0 and 1)
 - The greater the R value, the less residual values

Multiple Regression

- Of more use is the R^2
 - The % of variance in the DV that 'model' explains
 - However, it tends to overestimate the success of the model
- **Adjusted R^2** is the best/most useful value
 - Takes into account the number of variables in the model adjusted for sample size
 - E.g., if we had an adjusted R square value of 0.75, we can say the model accounts for 75% of the variance of the criterion.
- Beta values represent amount of contribution (how good a fit) for each of the independent variables

Selection Methods

- There are different ways the relative contribution of each predictor can be assessed
- In the simultaneous method (which SPSS calls the enter method), the researcher specifies the set of predictors that make up the model
- The success of this model in predicting the criterion is then assessed.

Hierarchical Multiple Regression

- In contrast, hierarchical methods enter the variables into the model in a specific order.
 - The order should reflect theoretical considerations.
- Sometimes we want to see how much an additional variable or variables add to the predictive model after controlling for the effects of other variables.....
 - Age, SES
 - Social desirability responding
 - Baseline measures
- Allows researchers to:
 - explore the predictive power of specific variables over and above other variables that may share some the variance.
 - statistically control for the effects of specific variables
- Additional variables are entered in “stages”
