

Group Questions

Data referenced throughout review

An Educational Testing Service (ETS) research scientist used multiple regression analysis to model y , the final grade point average (GPA) of business and management doctoral students (*Journal of Educational Statistics*, Spring 1993). Potential independent variables measured for each doctoral student in the study include:

- Quantitative Graduate Management Aptitude Test (GMAT) score
- Verbal GMAT score
- Undergraduate GPA
- First-year graduate GPA
- Student cohort (i.e., year in which the student entered doctoral program: 1988, 1990, 1992)

Model Building

- _____ means the smaller number of β s, the better. Simpler models are easier to understand and appreciate, and therefore have a "beauty" that their more complicated counterparts often lack.

Selecting Variables

What is the following describing?

- A regression in which a statistical software program begins by fitting all possible one-variable models to the data. The independent variable that produces the largest t value is declared the best one-variable predictor of y . A new variable is added until the given criteria for the t value can no longer be met.

Identifying Variables

Identify the following variables as quantitative or qualitative:

- Graduate Management Aptitude Test (GMAT) score
- Verbal GMAT score
- Undergraduate GPA
- First-year graduate GPA
- Student cohort (i.e., year in which the student entered doctoral program: 1988, 1990, 1992)

Purpose of Stepwise

- When and why would we use a variable selection technique such as stepwise regression?

Testing our Model

- If I have selected Graduate Management Aptitude Test (GMAT) score, Verbal GMAT score and Undergraduate GPA as my Independent variables for predicting graduate GPA, what would the first order model be?
- What are the null and alternative hypotheses regarding the global utility of the model?
- What about the hypothesis tests regarding the contribution of undergraduate GPA?
- How do I test these hypotheses?

Assumptions and Pitfalls

- What is multicollinearity and how can you detect it?

Residual Analysis

An observation that is larger than 2 or 3s is an _____.

Residual Analysis

- What is the difference between homoscedastic and heteroscedastic and which is preferable?

Residual Analysis

- How can we use residual plots to detect departures from the assumption of equal variances? (Include a sketch of what you are looking for and what indicates a violation of this assumption.)

Residual Analysis

- What are the assumptions about the random error term?

Assesses the Model

- What is the purpose of reviewing Standardized Residuals, Leverages, Cook's D, and/or DFFITS? Select one and tell how you might use it for this purpose.

Definitions

- Process of collecting sample data

Definitions

- In a(n) _____, subjects are assigned randomly to each of the cells formed by the treatments. Randomization is an effort to control for all unmeasured factors.

Definitions

- Originally developed for use in survival analysis, where the response y is typically measures as 0 or 1, depending on whether the experimental unit "survives."

Definitions

Individuals on which an experiment is conducted _____.

Explanations

- Describe a randomized block design and when it is advantageous to use it.

Design

- Few people want to eat discolored french fries. Potatoes are kept refrigerated before being cut for French fries to prevent spoiling and preserve flavor. But immediate processing of cold potatoes causes discoloring due to complex chemical reactions. The potatoes must therefore be brought to room temperature before processing. In this experiment, tasters will rate the color and flavor of french fries prepared from several groups of potatoes. The potatoes will be fresh picked or stored for a month at room temperature or stored for a month refrigerated. They will then be sliced and cooked either immediately or after an hour at room temperature.
- What are the factors and their levels?
- What are the treatments?
- What are the response variables?
- Describe the design of the experiment.
- If you want each taster to rate fries from all treatments, how could you use randomization in presenting fries to the tasters?