

STAT220: WINTER 2006: MIDTERM REVIEW

1. Design of studies

Controlled experiments Study units, population, sample, variable, response, factors, treatment, treatment group, control group, placebo, blind and double-blind studies, randomized study

Observational Studies Controlling for variables, confounding variables, sources of bias
Distinguishing between Controlled Experiments and Observational Studies

Sample Surveys Sampling unit, population, parameter, sample, statistic
Sources of bias: nonresponse bias, selection bias, other sources

A large sample does not protect against selection bias

Sampling procedures simple random sample, stratified sample, quota sample, convenience sample
(describe, identify, strengths and weaknesses)

2. Displaying data

Types of variables qualitative vs. quantitative, discrete vs. continuous
Categorical variables: nominal, ordinal; ordered categorical, interval scale

Forms of histogram display: Histograms, bars of equal/different widths, back-to-back, stacked
Pie charts, Bar charts, Stem-and-leaf diagrams

3. Summary statistics, and the Normal curve

Measures of "center": Median, Average, Mean, Percentiles.

Measures of "spread": Standard Deviation (SD), Range, inter-quartile range,

How to define them, find their values, "guesstimate" them from histograms.

The normal curve: Standard units (SD units), z-scores

Normal approximation for data

Normal curve arithmetic: Finding z-scores, percentiles, percentages

4. Correlation and Association

Scatter diagrams — make, read, interpret: describe basic shapes, identify outliers, non-linearity

Identify independent and dependent variables in a problem and on scatter diagrams

Association — positive, negative, weak, strong

Correlation coefficient, r : r is a number between -1 and +1

Interpretation of r – measures the degree of clustering of the scatter diagram about the SD line

Sensitivity of the correlation coefficient to outliers and to non-linearity

5 summary statistics for data on 2 variables: 2 means, 2 SD's, and r

Association is not cause! – confounding factors, time trends, etc.

5. Regression

Draw, interpret, and describe the basic "picture" of what is going on in regression

Cloud of points, axes, independent and dependent variables, point of averages

SD line, regression line, regression estimate: SD line versus regression line, and their slopes

Computation of regression estimates for a dependent variable, given the value of the independent variable – how to figure out what is required in a given situation, and how to compute it.

Residuals, prediction errors: Residual = observed value - predicted value

Rms error = sd of residuals: a measure of the likely size of the residuals

Measures how well the line fits the cloud of points

Computing the rms error

Figuring the normal approximation in vertical strips

Recognizing when to use or not to use regression estimation