STAT 200: Course Aims and Objectives

Attitudinal aims

In addition to specific learning outcomes, the course aims to shape the attitudes of learners regarding the field of Statistics. Specifically, the course aims to

1. Motivate in students an intrinsic interest in statistical thinking.

2. Instill the belief that Statistics is important for scientific research.

3. Provide a foundation and motivation for exposure to statistical ideas subsequent to the course.

Learning outcomes

Each numbered item states a learning aim for the course, and the items that follow indicate the learning outcomes (or objectives) through which that aim could be deemed to have been satisfied.

1. Demonstrate the ability to apply fundamental concepts in exploratory data analysis.

   (a) Distinguish between different types of data.

   (b) Interpret examples of methods for summarising data sets, including common graphical tools (such as boxplots, histograms and stemplots) and summary statistics (such as mean, median, mode, variance and IQR).

   (c) Assess which methods for summarising a data set are most appropriate to highlight interesting features of the data.

   (d) Identify the features that describe a data distribution.

   (e) Use an appropriate software tool for data summary and exploratory data analysis.

2. Design studies for obtaining data whilst avoiding common design flaws that incur bias, inefficiency and confounding.

(a) Recognise observational studies and experiments.

(b) Identify features common in experiments, including the experimental units, treatments, factors, control groups, randomisation and blocking.

(c) Recognise some common types of sampling design, such as simple random sampling, stratified sampling and multistage designs.

(d) Identify possible sources of bias and confounding in experiments and surveys.

(e) Identify possible hidden variables in a study.

3. Demonstrate an understanding of the basic concepts of probability and random variables.

(a) Recall rudimentary mathematical properties of probability.

(b) Describe the sample space for certain random experiments.

(c) Explain probability in terms of long–term relative frequencies in repetitions of experiments.

(d) Recall what are meant by the terms *independent*, *disjoint* and *complementary* events.

(e) Apply the definition of independence to attempt to determine whether an assumption of independence is justifiable in a given situation.

(f) Find probabilities of single events, complementary events and the unions and intersections of collections of events.

(g) Describe the main properties of probability distributions and random variables.

(h) Identify the random variable(s) of interest in a given scenario.

(i) Contrast discrete and continuous random variables.

(j) Find probabilities for distributions over finite discrete sets and for those continuous distributions for which probabilities can be found without the use of calculus.

(k) Calculate the mean and variance of a discrete random variable.

(l) Apply general properties of the expectation and variance operators.

(m) Interpret the mean of a random variable in terms of the Law of Large Numbers.

(n) Recall the key properties of the Normal distribution, including the preservation of Normality under linear transformation.

(o) Find the following for a Normal distribution: (i) the probability over a set of values, (ii) a percentile, (iii) the mean or variance given the other and either a percentile or a probability over a set and (iv) the mean and variance given probabilities over two sets.

(p) Recognise cases where the Binomial distribution could be an appropriate model.

(q) Recall characteristics of the Binomial distribution, such as the mean and variance.

(r) Compute probabilities for a Binomial distribution.

(s) Approximate Binomial probabilities using a Normal distribution, incorporating a continuity correction as appropriate.

4. Understand the concept of the sampling distribution of a statistic, and in particular describe the behaviour of the sample mean.

(a) Distinguish between a population and a sample, and between parameters and statistics.

(b) Recall the sampling distribution of the mean of a sample from a Normal distribution.

(c) Describe properties of the sampling distribution of the sample mean in general situations, using the Central Limit Theorem where appropriate.

(d) For statistics other than the sample mean, identify features that will influence the sampling distribution of the statistic.

(e) Apply the Central Limit Theorem to problems involving sums and averages of variables from arbitrary distributions.

5. Understand the foundations for classical inference involving confidence intervals and hypothesis testing.

(a) Describe the property of unbiasedness.

(b) Interpret a confidence interval and confidence level.

(c) Identify features that determine the width of a confidence interval.

(d) Identify the components of a classical hypothesis test, including the parameter of interest, the null and alternative hypotheses and the test statistic.

(e) Compute, or approximate, the P-value of a test statistic.

(f) Explain the two types of errors possible and define the associated probabilities.

6. Apply inferential methods relating to the means of Normal distributions.

(a) Recall the definition of a t statistic in terms of statistics of a sample from a Normal distribution.

(b) State and apply the definitions of the t, F and χ^2 distributions in terms of the standard Normal.

(c) Conduct inference for the mean of a Normal distribution where the underlying variance is either known or unknown, including the construction of confidence intervals and one and two–sided hypothesis tests.

(d) Recognise when inference should be based on matched pairs.

(e) Perform inference when data are in matched pairs.

(f) Identify where a two–sample inference is required.

(g) Conduct inference about the difference in the means of two Normal distributions, including cases where the underlying variances are known or unknown.

(h) Be aware of robustness issues in the context of one and two–sample t tests.

7. Apply and interpret basic summary and modelling techniques for bivariate data and use inferential methods in the context of simple linear models with Normally distributed errors.

(a) Identify a possible relationship in bivariate data from a scatterplot.

(b) Recall the properties of a sample correlation.

(c) Interpret a sample correlation.

(d) Recognise the limitations of correlation as a summary of bivariate data.

(e) Define the concept of least squares estimation in linear regression.

(f) Fit a linear model to a bivariate data set via software.

(g) Interpret the parameter estimates in a fitted linear model.

(h) Identify which variable is most naturally the response variable in a regression analysis.

(i) In the context of an example, explain why correlation need not necessarily imply causation.

(j) Evaluate the fit of a linear model by consideration of the residuals.

(k) Conduct inference for the slope and intercept parameters, including construction of confidence intervals and hypothesis tests.

(l) Calculate interval estimates for the mean and predicted responses at a given value of the explanatory variable.

(m) Estimate the underlying variance about the line.

(n) Explain the output from a software package for a regression analysis.

8. Demonstrate an appreciation of one–way analysis of variance (ANOVA).

(a) Identify situations where one–way ANOVA is and is not appropriate.

(b) State the modelling assumptions underlying ANOVA.

(c) State the null and alternative hypotheses for the ANOVA test.

(d) Explain the partitioning of the total sum of squares into the "within" and "between" group components.

(e) Identify the degrees of freedom associated with each sum of squares.

(f) Interpret an ANOVA table.

(g) Perform the F test in ANOVA, evaluating or approximating the P-value of the test statistic.

(h) Use the data to estimate the underlying within–group variance.

(i) Explain the output from a software package for an ANOVA study.

9. Interpret and analyse data that may be displayed in a two–way table.

   (a) Recognise when a two–way table is an appropriate way to summarise a data set.

   (b) Compute and interpret marginal and conditional distributions from two–way tables.

   (c) Describe the possible relationships between two categorical variables in a two–way table.

   (d) Perform the Chi-squared test for association for a two–way table, computing the test statistic if necessary.