Statistics 12: Guide for Teachers

Introduction

Previous exposure to statistical ideas expected is minimal: Foundations/Pre-calc 10: data, graphs, "situations", experimental probability

Main themes for Statistics 12:

1. Appreciation of the role of statistics in research, decision making, and policy.
2. Types of research and the nature of evidence – what makes a good study?
3. Exploring, describing, modeling, and explaining variation.
4. Intuitive inferential concepts.
5. Communication of statistical ideas.

All course materials, including “instructor only” resources, are available in the online repository StatSpace at

https://statspace.elearning.ubc.ca

Teachers are invited to rate the resources and provide comments and suggestions.

Weekly guide

Each class is centred around an inclass activity. Students should be encouraged to work in small groups on these activities, with the teacher circulating around the class discussing problems and monitoring progress. In most cases online resources that are freely available support the activity.

There follows a week-by-week guide to the course.

Week 1: Role of data in research, decision making, science, commerce ...

Interesting questions tend to require data to address them.

History: data collection and appropriate studies took a long time to evolve.

Historical perspective on development of statistical ideas in research.

E.g., role of data, clinical trials (Lind study on scurvy, 1753), randomization (post 1880)

Critique studies – identifying the motivation for a study and the data obtained.

Describing the role of statistical thinking in the scientific method.

Notions of population and sample. Introduce first applet.

Specifying the population of interest in a research study. (target populations)

Activity 1: What is statistics? For what type of problem could statistics be useful?

Activity 2: Create a research question. What sort of data would be needed? How would you collect the data? Problems?

Activity 3: Possible studies – what is the population of interest? What information would be obtained and how? What problems would be encountered?

Learning outcomes: By the end of this week learners should be able to

1. Appreciate that Statistics helps us explore variation.
2. Communicate at least one reason why Statistics is necessary to explore research questions that require data.
3. Define the concept of target population in the context of a research study.
4. Identify the target population for a research study given a brief description of the study.
5. Identify situations where it may be practical to take a census from the target population in order to address the research question of interest.
6. Explain why for a particular study it would be impractical to take a census of the target population.
7. Describe, in general terms, taking a sample from a population.

Week 2: Types of studies 1: Understanding observational studies

Observational studies: issues with lurking variables.

Why appropriate in some cases? Examples where only observational study feasible.

Critical thinking about observational studies. Surveys - questionnaire issues (e.g., Gallop poll on religious conviction.).

Census@School – big international survey on school students.

Limitations to inference from observational studies.

Mostly taught via examples.

Activity 1a: Smoking and mortality
Activity 1b: Exploring Variation in Sample Mean
Activity 2: PED use in class
Activity 3: Depression and Dementia
Activity: Mid-life activity and dementia
Activity: Dietary fibre and mortality
(One of latter could be set as homework assignment.)

There are two group projects during the course, one observational and one experimental. Students should be assigned into groups of three or four. Best to use different group assignments for each project.

Project 1 set: Formulate a research question suitable to be addressed by an observational study; design an observational study fit for purpose, conduct the study (gather suitable data), describe data and how they address research question. Write report and present.

Learning outcomes: By the end of this week learners should be able to

1. Recognise when a study is observational.
2. Describe common features of observational studies.
3. Identify the likely target population for a given observational study.
4. Describe why, for certain research questions, it is only possible to conduct an observational study.
5. Identify a possible lurking (or confounding) variable in a given observational study.
6. Communicate to a general audience the key aspects and findings of a research study that is observational.
7. Identify likely flaws and sources of error in a given observational study.

Week 3: Types of studies 2 Designed experiments

Contrast with observational studies – intervention. Treatments, randomisation.

Gold standard: random sample from population(s), random allocation to treatments.

Why difficult? Practical considerations. All via examples. E.g. smoking.

Activity 1: Lind study on scurvy – the first clinical trial?
Activity 2: Diet type and weight.
Activity 3: Experiment on Smoking?
Activity 4: Therapeutic touch
Activity: Processed foods
Activity: EHS
Additional activities on: Effect of prayer, jargon and learning, flipped classes, dogs detecting cancer.

Learning outcomes: By the end of this week learners should be able to

1. Recognise when a study is an experiment.
2. Describe why, for certain research questions, it is only possible to conduct an experiment.
3. Identify the likely target population for a given experimental study.
4. Identify the experimental units in an experiment.
5. Explain the importance of randomization in experiments.
6. Identify the treatments in an experiment.
7. Explain the roles of blinding and double-blinding in experiments.
8. Critique an experiment by identifying an important strength and possible flaw in its design.

Project 2 set: Formulate a research question suitable to be addressed by an experiment; design an experiment fit for purpose, conduct the study (gather suitable data), describe data and how they address research question. Write report and present.

Week 4: Exploring and describing variation graphically (most ideas met in grade 10 Foundations/Pre-calc)

Examples of studies - ask students to create suitable graphics, summaries, for univariate data (though taking steps to think about bivariate, multivariate extensions)

Key tools: Bar charts, histograms, dotplots/strip charts, stemplots, boxplots. Compare/contrast.

Interpretation of data distributions via graphics. Limited mainly to univariate data here.

Recognising advantages and disadvantages of different approaches.

Specify which method is better for a given data set.

Software: students are free to use what they wish, hand-drawn fine for inclass activities. R notes are provided, however, for those interested.
Activity 1: Stemplots
Activity 2: Histograms
Activity 3: More on Histograms
Activity 4: Boxplots
Activity 5: Bar charts

Learning outcomes: By the end of this week learners should be able to

1. Distinguish between categorical and quantitative variables.
2. Create histograms, stemplots, dotplots, and boxplots for quantitative data.
3. Interpret histograms, stemplots, dotplots, and boxplots for quantitative data.
4. Use appropriate vocabulary when describing data distributions, including terms such as unimodal, bimodal, multi-modal, symmetric, (left and right) skewed, tails, outliers, and uniform.
5. Construct bar charts for presenting univariate categorical data.
6. Interpret bar and pie charts for presenting categorical data.
7. Decide which method is preferable for displaying a univariate data set.
8. Recall that good graphical displays include a title and clear labeling of variables, providing any units necessary.
9. Recognize and critique both good and bad examples of displays of univariate data.

Week 5: Exploring data distributions via summary statistics

Measures of centre, spread (range, variance, sd, IQR), include five-number summary.
Software: students are free to use what they wish, including MS Excel, for inclass activities. R notes provided, however, for those interested.
Activity 1: Measures of location: mean, median, and mode
Activity 2: Five-number summary
Activity 3: Measures of spread
Activity 4: More on measures of spread
Activity 5: Review of Summary Statistics

Learning outcomes: By the end of this week learners should be able to

1. Compute summary statistics such as the median, mean, mode, range, variance, and standard deviation for univariate data.
2. Interpret summary statistics such as the median, mean, mode, range, variance, and standard deviation for univariate data.
3. Compare summary statistics such as the median, mean, mode, range, variance, and standard deviation across more than one data set.
4. Compute the five-number summary to describe a set of data.
5. Interpret what the five-number summary indicates about a set of data.
6. Identify outliers in a data set.
7. Describe how outliers influence summary statistics.
8. Appreciate the role of units in the comparison of data sets and their summary statistics.
9. Apply standardisation to facilitate the comparisons of data sets.
10. Recognize how rescaling data, such as by changing the units, affects summary statistics.
11. Apply Chebychev's Inequality to estimate the proportion of a distribution falling within two (or three) standard deviations of the mean.
12. Apply Chebychev's Inequality to find ranges that must contain at least 75% and 89% of a distribution.

Week 6: Exploring association between variables

Categorical variables: Contingency tables – clustered/stacked bar charts

Quantitative variables: Scatterplots, Correlation - no formula, uses, does not always imply causation.

Activity 1: Bivariate categorical
Activity 2: Association across categorical variables
Activity 3: Intro Bivariate Quantitative Data
Activity 4: Introducing correlation
Activity 5: Review and extensions

Learning outcomes: By the end of this week learners should be able to

1. Create and interpret contingency tables, including their marginal totals.
2. Investigate association between two categorical variables via a contingency table and graphical methods.
3. Identify occurrences of Simpson's paradox.
4. Use graphical displays and summary statistics to explore association between a categorical variable and a quantitative variable.
5. Decide which method is preferable for displaying a bivariate data set.
6. Identify a possible relationship in bivariate quantitative data from a scatterplot.
7. Recall the properties of a sample correlation, characterising situations when the values 0 and ±1 are attained.
8. Interpret a sample correlation.
9. Recognize the limitations of correlation as a summary of bivariate data.
10. In the context of an example, explain why correlation need not necessarily imply causation.
11. Identify which variable, if either, is most naturally the response variable in an analysis of bivariate data.

Week 7: Models for variation 1

Binomial distribution - not an overly mathematical introduction.

What does the distribution model?

When is it appropriate?

What are the assumptions?

Dependence of shape of distribution on parameters (n and p)

Simulation and software: see activity notes for online resources

Based on examples. Use of software/applets

Activity 1: Introducing Binomial Variation
Activity 2: Introducing the Binomial Distribution
Activity 3: Understanding the Binomial Distribution
Activity 4: Exploring the Binomial Distribution
Activity 5: Applying the Binomial Distribution

Learning outcomes: By the end of this week learners should be able to

1. Describe, in general terms, the key features of phenomena for which a Binomial distribution could be a reasonable model.
2. Identify situations where a Binomial distribution would be a reasonable model.
3. Where a Binomial distribution may be reasonable for describing variation, identity the number of trials, the range of values possible, and the shape of the distribution.
4. Display graphically data resulting from trials that could reasonably be modelled by a Binomial distribution.
5. Use data assumed to be from a Binomial model to estimate the probability of "success" for the model.
6. Describe how the shape of a Binomial distribution depends on the two parameters.
7. Use software (online or otherwise) to compute any probability of interest for a given Binomial distribution.
8. For a situation where a Binomial model may be appropriate, identify parameter choices that could be of interest in the context of the research question(s).
9. For studies where a Binomial model may be appropriate, use software to explore a relevant research question.

Week 8: Models for variation 2

Normal (Gaussian) distribution - properties (not very mathematical!)

When useful?

Role of parameters (mean mu and sd. sigma)

Proportions 1, 2, 3, s.d.'s from mean. (68-95-99 rule)

Applications of Normal model. Software/applets - "discover" sampling distribution of mean

Software, applets
Activity 1: Introducing Normal Variation
Activity 2: Introducing the Normal Distribution
Activity 3: Understanding the Normal Distribution
Activity 4: Exploring the Normal Distribution
Activity 5: Applying the Normal Distribution

Learning outcomes: By the end of this week learners should be able to

1. Recall the shape of the Normal distribution.
2. Approximate proportions falling one, two, and three standard deviations either side of the mean for a Normal distribution.
3. Standardise an arbitrary Normal variable to a standard Normal.
4. Use software or otherwise to find the proportion (or probability) over a set of values for a Normal distribution.

5. Use software or otherwise to find a percentile (or quantile) for a Normal distribution.
6. Draw a suitable picture to illustrate a Normal percentile.
7. Identify situations when the Normal model would and would not be sensible.
8. Describe in simple terms the roles of the parameters in a Normal model.
9. Apply a graphical method to assess the appropriateness of a Normal model for a data set.
10. Describe the sampling distribution of the mean of a sample from a Normal distribution.

Week 9: Review week

Explore some real studies, milestones for projects: proposals submitted, feedback.

Discuss communication of statistical findings.
Test 2

Week 10: Models for variation 3: linear models for bivariate quantitative data

Review concepts from week 6: bivariate data, correlation, scatterplots …

Linear modelling: purpose, estimation, link to correlation, applications.
Least squares fitting via software.

Activity 1: Introducing linear models
Activity 2: Introducing least squares model fitting
Activity 3: Fitting Linear models
Activity 4: Applying Linear Models 1
Activity 5: Applying Linear Models 2

Learning outcomes: By the end of this week learners should be able to

1. Define the concept of least squares estimation in linear models.
2. Identify which variable is most naturally the response variable in a regression analysis.
3. Guess the correlation associated to a scatterplot to within 0.3 of the true value.
4. Fit a linear model to a bivariate data set via software.
5. Interpret the parameter estimates for a fitted linear model.
6. Identify the units of the parameter estimates for a fitted linear model.
7. Interpret a fitted linear model in the context of the data.
8. Use a linear model to find the expected value of the response for a given value of the explanatory variable.
9. Recall that the centroid of the bivariate data, as defined by the coordinate given by the sample means, lies on the least squares line.
10. Compute and interpret residuals in the context of a linear model.

11. Compute and interpret the R^2 statistic in a linear model example.
12. Identify data points that have high influence in the fitting of a linear model.

Week 11: Variation of sample statistics: means and proportions

Explore again via simulation how sample mean varies – CLT!

Explore how sample proportions vary as a special case of the sample mean.

Activity 1: Revisiting Variation in The Sample Mean
Activity 2: The Central Limit Theorem
Activity 3: CLT applications
Activity 4: Variation in sample proportions 1
Activity 5: Variation in sample proportions 2

Learning outcomes: By the end of this week learners should be able to

1. Distinguish between a population and a sample, and between population parameters and sample statistics.
2. Identify the population, sample, parameters, and statistics in a given scenario.
3. Describe the sampling distribution of the sample mean of a sample from a Normal distribution.
4. Describe properties of the sampling distribution of the sample mean in general situations, using the Central Limit Theorem where appropriate.
5. For the sample mean and proportion, identify features that will influence the sampling distribution of the statistic, such as sample size and population parameters.
6. Recognise conditions under which use of the Central Limit Theorem may not be justified.
7. Apply the Central Limit Theorem to problems involving averages from arbitrary distributions.

Week 12: Estimation

Intuition on interval estimation via simulations/applets: proportions, means.
Activity 1: Introducing interval estimation
Activity 2: Confidence interval for a mean
Activity 3: Confidence interval for a proportion
Activity 4: Applying the CLT for confidence intervals

Learning outcomes: By the end of this week learners should be able to

1. Compute a confidence interval for a proportion or mean given relevant sample statistics.

2. Explain how the Central Limit Theorem helps to find confidence intervals for a mean or proportion.
3. Explore the behavior of confidence intervals for a mean or proportion via simulation.
4. Interpret a confidence interval at a given confidence level for a mean and a proportion.
5. Identify features that determine the width of a confidence interval for a mean or a proportion.

Test 3

Week 13: Exploring research questions 1

Inference for proportion via simulation (randomisation/permutation) methods

Activity 1: Dolphin communication
Activity 2: Kissing direction
Activity 3: Brexit poll
Activity 4: Therapeutic Touch
Activity 5: Dogs and Colorectal Cancer

Learning outcomes: By the end of this week learners should be able to

1. Recognise when a research study relates to inference for a proportion (or probability).
2. Where appropriate, suggest a sensible default value for a proportion (or probability), departure from which would suggest an effect of interest.
3. Explain how software and physical randomization devices (such as dice, coins, and cards) can be used to conduct simulation-based inference for a proportion (or probability).
4. Use simulation to explore how often a sample proportion will be at least as extreme as the value observed in the data if the default value were the actual value of the proportion (or probability).
5. Compare a confidence interval for a proportion (or probability) applying the Central Limit Theorem with simulation-based inference for that proportion (or probability).
6. Communicate the conclusions following inference for a proportion (or probability) in the context of the study and the research question(s) of interest.

Week 14: Exploring research questions 2

Inference for a mean via simulation – CLT!

Includes paired studies.

Activity 1: Song Excerpt Length
Activity 2: Sleep time
Activity 3: Head Lice Eggs
Activity 4: BP differences by arm
Activity 5: Phthalate excretion

Learning outcomes: By the end of this week learners should be able to

1. Identify when a convenience sample has been taken in a study.
2. In the context of a study, explain whether or not convenience sampling may have biased the results.
3. Recognise when use of the Central Limit Theorem may be justified for inference about a population mean.
4. Use the Central Limit Theorem to find a 95% confidence interval for a mean of interest in a research study.
5. In the context of a study, interpret a 95% confidence interval for a mean of interest in the research study.
6. Identify which value of a mean is the most sensible default value in the context of a research study (that is, define the appropriate “null hypothesis”).
7. Identify which alternative hypothesis about a mean may be of most relevance in the context of a research study.
8. Explore a hypothesis about a population mean in the context of a study, determining whether the data appear more consistent with that hypothesis or an alternative hypothesis of interest.
9. Use an applet, or otherwise, to quantify how surprising the data from a study exploring a population mean appear assuming a default value for the mean.
10. Identify when a study uses a matched pairs design.
11. Explain the likely benefits of using a matched pairs design.
12. Communicate conclusions from inference for a population mean in context.

Week 15: Exploring research questions 3

Two-sample questions - via randomisation/permutation (call them "relabeling/reshuffling algorithms")

Two-way tables - via applets

Activity 1: Penguin Tagging
Activity 2: Yawn seeding
Activity 3: Sleep deprivation
Activity 4: Are Experienced Teachers Better?

Learning outcomes: By the end of this week learners should be able to

1. Identify when a study relates to a difference in population proportions.
2. In a context where inference is about the difference in two proportions, identify which value of the difference is the most sensible default value in the context of a research study (that is, define the appropriate “null hypothesis”).
3. In a context where inference is about the difference in two means, identify which value of the difference is the most sensible default value in the context of a research study (that is, define the appropriate “null hypothesis”).
4. In a context where inference is about the difference in two proportions, identify which alternative hypothesis about the difference may be of most relevance in the context of a research study.
5. In a context where inference is about the difference in two means, identify which alternative hypothesis about the difference may be of most relevance in the context of a research study.
6. Explain how physical randomization devices (such as cards and paper) can be used to conduct simulation-based inference for a difference in two proportions (or probabilities) or a difference in two means.
7. Explain how software can be used to conduct simulation-based inference for a difference in two proportions (or probabilities) or a difference in two means.
8. Explore a hypothesis about the difference in two proportions in the context of a study, determining whether the data appear more consistent with that hypothesis or an alternative hypothesis of interest.
9. Explore a hypothesis about the difference in means in the context of a study, determining whether the data appear more consistent with that hypothesis or an alternative hypothesis of interest.
10. In the context of a study, interpret a 95% confidence interval for a difference in two proportions of interest in the research study.
11. In the context of a study, interpret a 95% confidence interval for a difference in two means of interest in the research study.
12. Use an applet, or otherwise, to quantify how surprising the data from a study exploring a difference in two proportions appears assuming a default value for the difference.
13. Use an applet, or otherwise, to quantify how surprising the data from a study exploring a difference in two means appears assuming a default value for the difference.
14. Communicate conclusions from inference for a difference in two proportions in context.
15. Communicate conclusions from inference for a difference in two means in context.

Weeks 16/17: Exploring research questions 4

Questions about bivariate quantitative data: correlation, regression.

Activity 1: Foot length and height
Activity 2: Whale feeding
Activity 3: Name game

Learning outcomes: By the end of this week learners should be able to

1. Identify when a study relates to inference on the relationship between two quantitative variables.
2. In a context where inference is about the relationship between two quantitative variables, identify which value of a parameter is the most sensible default value in the context of a research study (that is, define the appropriate "null hypothesis").
3. In a context where inference is about the relationship between two quantitative variables, identify which alternative hypothesis about a parameter may be of most relevance in the context of the study.
4. Explain how physical randomization devices (such as using paper) can be used to conduct simulation-based inference for inference about the relationship between two quantitative variables.
5. Explain how software can be used to conduct simulation-based inference about the relationship between two quantitative variables.
6. Explore a hypothesis about the relationship between two quantitative variables in the context of a study, determining whether the data appear more consistent with that hypothesis or an alternative hypothesis of interest.
7. Use an applet, or otherwise, to quantify how surprising the data from a study exploring the relationship between two quantitative variables appears assuming a default value for the parameter of interest.
8. Describe whether a simulation-based exploration of the relationship between two quantitative variables depends on the choice of the alternative hypothesis.
9. Describe whether a simulation-based exploration of the relationship between two quantitative variables depends on the choice of the parameter of interest.
10. Communicate conclusions from inference for a study exploring the relationship between two quantitative variables in context.

Test 4

Weeks 17, 18: Project submissions/peer reviews/ presentations/Review sessions/Final Test

Resources

All resources listed are freely available.

On-line applets and visualization tools in StatSpace at
https://statspace.elearning.ubc.ca/

All course activities are available in StatSpace, along with solutions and additional notes for those registering as instructors.

Nathan Tintle *et al.* materials are at
http://www.isi-stats.com/isi/applets.html
includes assessment papers/presentations as well as sample syllabi; look under the “instructor resources” tab. E.g., "One proportion" applet: shows repeats of coin tosses, animating coins if desired. Also links to publications on simulation-based inference.
E.g., Descriptive statistics: sample data can be entered, displays and summary statistics created.
All referred to in the textbook:
Tintle, N., Chance, C., Cobb, G., Rossman, A., Roy, S., Swanson, T., VanderStoep, J. (2021): Introduction To Statistical Investigations. (2nd edt.). Wiley.

See also StatKey resources (Lock*5)
http://www.lock5stat.com/StatKey/index.html

SOCR applets:
https://rcompute.nursing.umich.edu/SOCR_Probability_Distribution_Calculator_and_Modeler/
(also https://myumi.ch/bvkpx. These great tools for exploring probability distributions have moved around online over the past decade; the above links are correct as of 2025).

University of Sussex visualisation applets: https://and.netlify.app/viz/
(nice, particularly for exposure to R code)

CAUSEWeb: www.causeweb.org
(Some good resources but many broken links.)

WeBWorK(iR): www.webwork.maa.org
(Online homework system, lots of statistics questions.)

LOCUS sample questions: https:/locus.statisticseducation.org/
(Concept-based test questions for schools.)

Online open access books suggested at reading and viewing materials in activities:
https://www.ck12.org/c/statistics/

Open Intro Stats: Modern approach, good for SBI
https://www.openintro.org/book/os/

OpenStax Statistics: More traditional, no SBI:
https://openstax.org/details/books/introductory-statistics

There is online forum for statistics educators in Canada on the Slack channel “SSC - Statistics Education Section”. Discussions are mainly on teaching at undergraduate level although contributions from school teachers would be welcome.

See also
https://iase-web.org/icots/11/proceedings/pdfs/ICOTS11_434_DUNHAM.pdf?1669865564
for some reflections and findings on teaching Statistics 12.

BD