JOHN ABBOTT CEGEP/COLLEGE

Probability and Statistics

SCIENCE PROGRAM
COURSE OUTLINE
WINTER 2025

General Information.

Discipline: Mathematics *Course code*: 201-SN1-RE
Ponderation: 2-1-2 *Credits*: $1\frac{2}{3}$
Prerequisite: Secondary IV mathematics, Science option or Technical and Scientific option (SN4 or TS4)
Objective: Solve problems related to the natural sciences using statistical methods and probability concepts (0M01)

Your teacher will give you his/her schedule and availability.
Students are strongly advised to seek help promptly from their teacher if they encounter difficulties in the course.

Introduction. Probability and Statistics is one of the four required mathematics courses in the Science program. It is usually taken in the first semester and it introduces the student to statistical tools for producing, describing, and drawing conclusions from scientific data.

The main purpose of the course is the attainment of objective 0M01 ("Solve problems related to the natural sciences using statistical methods and probability concepts"). To achieve this, the course presents methods of basic statistics in a way that emphasizes working with data and mastering statistical reasoning. Using real examples from both the natural and health sciences, the course discusses not only the traditional topics of probability and inference, but also the production of data (experiments and observational studies) and the analysis of data (data visualization).

This course is a prerequisite for 360-PRO-NS, the Science Project course. It contributes to Exit Profile goals 3, 4, 5, and 7 of the Science program. (Goals 1–6 are prescribed by the Ministry, whereas goal 7 is specific to John Abbott; it states that students should "demonstrate a nuanced understanding of human impacts on the natural world and planetary life support systems in the context of global climate and environmental disruptions.") Also, the course can contribute to the Environmental Studies certificate; for more information, talk to your teacher or contact the certificate coordinator.

Teaching Methods. This is a 45-hour course with two 90-minute classes per week. Classes are primarily lectures, but other teaching methods may be used (for example, labs, problem-solving sessions, class discussions, and assigned reading for independent study). Regular homework requiring an average of two hours per week should be expected. Students are responsible for all homework problems assigned by the teacher.

Software. Students will be introduced to R and its companion interface RStudio, a freely available language and environment for statistical computing and graphics. Those without a personal computer or laptop will have access to college computers (for example, in the Math Study Area) where this software is installed.

Reference. There is no required textbook for this course. A set of exercises will be provided by your teacher. A good reference for the course material is

The Practice of Statistics in the Life Sciences, Fourth Edition, by Brigitte Baldi and David S. Moore (Macmillan Learning, 2022).

Note that this book is not available for purchase at the bookstore.

Course Costs. A scientific, non-graphing, non-programmable calculator is necessary. A suitable model (Sharp EL-531) is available at the bookstore for around $24.

Evaluation Plan. The Final Evaluation in this course consists of the Project (which involves R) and the Final Exam (two hours during the final exam period), which covers all elements of the competency.

A student's Final Grade will be calculated like this:

Final exam	30%
Project	10%
2 written, in-class tests	40%
Minor assessments	20%

The minor assessments in the evaluation plan, which consist of labs, quizzes, or assignments, are detailed in an appendix that is distributed to students along with this course outline. The Final Exam is set by the 201-SN1-RE course committee (which consists of all instructors currently teaching the course), and is marked by each individual instructor. Every effort is made to ensure equivalence between the various sections of the course.

Students must be available until the end of the final examination period to write exams.

Other Resources.

Math Website.
http://departments.johnabbott.qc.ca/departments/mathematics

Math Study Area. Located in H-200A and H-200B; the common area is usually open from 8:30 to 17:30 on weekdays as a quiet study space. Computers and printers are available for math-related assignments. It is also possible to borrow course materials when the attendant is present.

Math Help Centre. Located near H-211; teachers are on duty from 8:30 until 15:30 to give math help on a drop-in basis.

Academic Success Centre. The Academic Success Centre, located in H-139, offers study skills workshops and individual tutoring.

College Policies.

Policy No. 7 - IPESA, Institutional Policy on the Evaluation of Student Achievement: https://www.johnabbott.qc.ca/wp-content/uploads/2021/05/Policy-No.-7-IPESA-FINAL.pdf.

Religious Holidays (Article 3.2.13 and 4.1.6). Students who wish to miss classes in order to observe religious holidays must inform their teacher of their intent in writing within the first two weeks of the semester.

Student Rights and Responsibilities: (Article 3.2.18). It is the responsibility of students to keep all assessed material returned to them and/or all digital work submitted to the teacher in the event of a grade review. (The deadline for a Grade Review is 4 weeks after the start of the next regular semester.)

Student Rights and Responsibilities: (Article 3.3.6). Students have the right to receive graded evaluations, for regular day division courses, within two weeks after the due date or exam/test date, except in extenuating circumstances. A maximum of three (3) weeks may apply in certain circumstances (ex. major essays) if approved by the department and stated on the course outline. For evaluations at the end of the semester/course, the results must be given to the student by the grade submission deadline (see current Academic Calendar). For intensive courses (i.e.: intersession, abridged courses) and AEC courses, timely feedback must be adjusted accordingly.

Academic Procedure: Academic Integrity, Cheating and Plagiarism (Article 9.1 and 9.2). Cheating and plagiarism are unacceptable at John Abbott College. They represent infractions against academic integrity. Students are expected to conduct themselves accordingly and must be responsible for all of their actions.

College definition of Cheating: Cheating means any dishonest or deceptive practice relative to examinations, tests, quizzes, lab assignments, research papers or other forms of evaluation tasks. Cheating includes, but is not restricted to, making use of or being in possession of unauthorized material or devices and/or obtaining or providing unauthorized assistance in writing examinations, papers or any other evaluation task and submitting the same work in more than one course without the teacher's permission. It is incumbent upon the department through the teacher to ensure students are forewarned about unauthorized material, devices or practices that are not permitted.

College definition of Plagiarism: Plagiarism is a form of cheating. It includes copying or paraphrasing (expressing the ideas of someone else in one's own words), of another person's work or the use of another person's work or ideas without acknowledgement of its source. Plagiarism can be from any source including books, magazines, electronic or photographic media or another student's paper or work.

Course Content. The topics listed below follow the presentation in the reference text *The Practice of Statistics in the Life Sciences*, Fourth Edition, by Brigitte Baldi and David S. Moore (Macmillan Learning, 2022). Your teacher may supplement this list during the semester. Regular work done as the course progresses should make it easier for you to master the course.

1. *Producing and Looking at Data*
   - 1.1 Picturing distributions with graphs
   - 1.2 Describing distributions with numbers
   - 1.3 Examining relationships with scatterplots
   - 1.4 Correlation and regression
   - 1.5 Samples, observational studies, and experiments
2. *Probability*
   - 2.1 Probability rules
   - 2.2 Independence and conditional probabilities
   - 2.3 Normal distributions
   - 2.4 Binomial distributions
   - 2.5 Sampling distributions
3. *Statistical Inference*
   - 3.1 Estimation and hypothesis tests
   - 3.2 Inference about a population mean
   - 3.3 Inference about a population proportion
   - 3.4 Inference for two-way tables

For more details, see the list of Intermediate Learning Objectives below.

OBJECTIVE	STANDARD
Statement of the Competency Solve problems related to the natural sciences using statistical methods and probability concepts (0M01).	**Performance Criteria for the Competency as a Whole** • Correct use of mathematical terminology and syntax • Appropriate use of necessary computer and statistical tools • Algebraic manipulation in accordance with established rules
Elements of the Competency	**Performance Criteria**
1. Use descriptive statistics to analyze data.	• Appropriate use of statistical vocabulary • Appropriate representation of a frequency distribution in the form of a table or graph • Appropriate calculation of measures of central tendency, variability and position • Accurate interpretation of tables, graphs and measurements
2. Use probability concepts in aleatory situations.	• Correct use of counting techniques (permutations and combinations) • Exact calculation of probabilities (independence and conditional probabilities) • Accurate probability distribution of discrete and continuous variables • Correct resolution of problems involving binomial and normal distributions
3. Use statistical inference to describe a population.	• Accurate recognition of the conditions for applying the Central Limit Theorem • Accurate interpretation of the margin of error • Appropriate use of confidence intervals to estimate a population mean (when $n \geqslant 30$ and when $n < 30$) and a population proportion (when $n \geqslant 30$) • Correct use of hypothesis tests for a population mean (when $n \geqslant 30$ and when $n < 30$) and for a population proportion (when $n \geqslant 30$) • Correct resolution of problems calling for the use of statistical inference methods
4. Determine the nature and strength of the relationship between two variables.	• Accurate determination of the equation of the regression line • Accurate interpretation of the correlation coefficient and coefficient of determination • Correct resolution of problems involving the concept of a regression line • Accurate determination of the dependency relationship between two qualitative variables • Correct resolution of problems involving the chi-squared test of independence

Intermediate Learning Objectives. It is understood that each item in the following list is preceded by: "The student is expected to . . . ".

1.1 Picturing distributions with graphs

1.1.1. Identify a variable as categorical or quantitative.
1.1.2. Make a pie chart or bar graph of the distribution of a categorical variable and interpret the results.
1.1.3. Make a time plot of a quantitative variable recorded over time. Recognize trends and cycles in time plots.
1.1.4. Make a histogram or dotplot of the distribution of a quantitative variable.
1.1.5. Describe the shape, centre, and spread of a quantitative distribution from a histogram or dotplot. Recognize outliers and give plausible explanations for them.

1.2 Describing distributions with numbers

1.2.1. Decide which measures of centre and spread are more appropriate: the mean and standard deviation or the five-number summary.
1.2.2. Find the five-number summary and draw a boxplot. Assess centre, spread, symmetry, and skewness from a boxplot.
1.2.3. Find the mean, variance, and standard deviation for a set of observations.
1.2.4. Understand that the median is more resistant than the mean. Recognize how the skewness of a distribution affects where the mean is located relative to the median.
1.2.5. Know the basic properties of the standard deviation.

1.3 Examining relationships with scatterplots

1.3.1. Identify the explanatory and response variables in situations where one variable explains or influences another.
1.3.2. Make a scatterplot to display the relationship between two quantitative variables measured on the same subjects.
1.3.3. Describe the form, direction, and strength of the overall pattern in a scatterplot. Recognize outliers and give plausible explanations for them.

1.4 Correlation and regression

1.4.1. Judge whether it is appropriate to use correlation to describe the relationship between two quantitative variables. Find the correlation.
1.4.2. Know the basic properties of correlation.
1.4.3. Find the slope and intercept of the least-squares regression line. Explain what they mean.
1.4.4. Use the regression line to predict y for a given x. Recognize the dangers of extrapolation.
1.4.5. Interpret the square of the correlation.
1.4.6. Understand that correlation and regression can both be strongly influenced by a few extreme observations.
1.4.7. Recognize that even a strong correlation does not mean that there is a cause-and-effect relationship between x and y.
1.4.8. Give plausible explanations for the observed association between two variables: direct cause and effect, the influence of lurking variables, or both.

1.5 Samples, observational studies, and experiments

1.5.1. Identify the population in a sampling situation.
1.5.2. Identify potential confounding variables and understand the limitations of the conclusions that can be drawn from observational studies.
1.5.3. Recognize whether a study is an observational study or an experiment.
1.5.4. Recognize bias due to confounding of explanatory variables with lurking variables.
1.5.5. Identify the factors (explanatory variables), treatments, response variables, and subjects in an experiment.
1.5.6. Outline the design of a completely randomized experiment.
1.5.7. Recognize the placebo effect. Recognize when double-blinding should be used.
1.5.8. Explain why randomized comparative experiments can give good evidence for cause-and-effect relationships.

2.1 Probability rules

2.1.1. Recognize probabilities as long-run proportions (under the frequentist interpretation) and be aware of other interpretations.
2.1.2. Identify the sample space of a random event.
2.1.3. Calculate probabilities using basic probability rules (addition, complement).
2.1.4. Find probabilities in a discrete probability model by adding the probabilities of individual outcomes. Interpret probabilities in a continuous probability model as areas under a density curve.
2.1.5. State the definition of a random variable. Identify a random variable as discrete or continuous.
2.1.6. Calculate the mean and standard deviation of a discrete random variable.

2.2 Independence and conditional probabilities

2.2.1. Understand the idea of independence. Judge when it is reasonable to assume independence as part of a probability model.
2.2.2. Use the multiplication rule to calculate joint and conditional probabilities.
2.2.3. State and apply Bayes's Rule. Use the formula or a tree diagram to find a conditional probability based on other, known probabilities.

2.3 Normal distributions

2.3.1. Find and interpret the standardized value (z-score) of an observation.
2.3.2. Compute probabilities for a normal random variable.
2.3.3. Find the point that splits a given normal distribution into two complementary areas.

2.4 Binomial distributions

2.4.1. Recognize the binomial setting: a fixed number of independent success/failure trials each having the same probability of success.
2.4.2. Understand how to use counting techniques to derive the binomial probability formula.
2.4.3. Find probabilities of events in a binomial setting.
2.4.4. Find the mean and standard deviation of a binomial random variable.
2.4.5. Recognize when you can use the normal approximation to a binomial distribution.
2.4.6. Calculate probabilities using the normal approximation.

2.5 Sampling distributions

2.5.1. Identify the parameters and statistics in a statistical study.
2.5.2. Interpret the sampling distribution of a statistic as the probability distribution of its possible values.
2.5.3. Recognize when a problem involves the mean $\overline{x}$ of a sample.
2.5.4. Identify the mean and standard deviation of the sampling distribution of $\overline{x}$ when the mean and standard deviation of the population are known.
2.5.5. Understand that $\overline{x}$ has approximately a normal distribution when the sample is large (central limit theorem). Use this distribution to calculate probabilities that concern $\overline{x}$.
2.5.6. Recognize when a problem involves the proportion $\widehat{p}$ of successes in a sample.
2.5.7. Identify the mean and standard deviation of the sampling distribution of $\widehat{p}$ when the population proportion is known.

2.5.8. Understand that $\widehat{p}$ has approximately a normal distribution when the sample is large. Use this distribution to calculate probabilities that concern $\widehat{p}$.

3.1 Estimation and hypothesis tests

3.1.1. State in non-technical language what is meant by the confidence level of a confidence interval.
3.1.2. Obtain a confidence interval for the mean of a normal population with known standard deviation.
3.1.3. Understand how the margin of error of a confidence interval changes with the sample size and level of confidence.
3.1.4. State the null and alternative hypotheses in a testing situation where the parameter in question is a population mean.
3.1.5. State in non-technical language the meaning of the P-value.
3.1.6. Obtain the one-sample z test statistic and the P-value for both one-sided and two-sided tests about the mean of a normal population with known standard deviation.
3.1.7. Use the P-value to assess the strength of the evidence against the null hypothesis.
3.1.8. Recognize that significance testing does not measure the size or importance of an effect. Explain why a small effect can be significant in a large sample and why a large effect can fail to be significant in a small sample.
3.1.9. Identify the two types of error that can be made in hypothesis testing.
3.1.10. Identify practices such as P-hacking, data snooping, and publication bias that undermine the reliability of the scientific record.

3.2 Inference about a population mean

3.2.1. Check that the t procedures are appropriate in a given setting.
3.2.2. Use the one-sample t procedure to obtain a confidence interval for the mean of a population. Conclude in context.
3.2.3. Carry out a one-sample t test for the hypothesis that a population mean has a specified value against a one-sided or two-sided alternative. Obtain the P-value and conclude in context.

3.3 Inference about a population proportion

3.3.1. Check that the large-sample z procedures can safely be used in a given setting.
3.3.2. Use the large-sample z procedure to obtain a confidence interval for a population proportion. Conclude in context.
3.3.3. Carry out a one-sample z test for the hypothesis that a population proportion has a specified value against a one-sided or two-sided alternative. Obtain the P-value and conclude in context.

3.4 Inference for two-way tables

3.4.1. Organize data on two categorical variables as a two-way table of counts.
3.4.2. Use the expected counts in a two-way table to check whether the chi-squared test can safely be used.
3.4.3. Carry out a chi-squared test for two-way tables. Obtain the P-value and conclude in context.

The following intermediate learning objectives will be explored, at least in part, using examples related to Exit Profile goal 7.2 (*Identify human factors that complicate effective human action from being taken [regarding the climate emergency]*): 1.3.3, 1.4.4, 1.4.6, 1.5.2, 1.5.4, 3.1.10.