Course timetable and suggested exercises:

Throughout the semester, there will be weekly updates. You should check this page regularly. Here is the link to the bank of questions.

(http://aix1.uottawa.ca/~glamothe/questions/introEN.html)

Suggested exercises are minimum practice you need at home.

Week 1 (Jan 9th and Jan 11th)

We have started to define a few concepts. Important terminology: random experiment, event, occurrence of an event, mutually exclusive events and exhaustive events. We defined operations on events: union, intersection, complement. We gave the DeMorgan Laws.

We talked about the counting techniques. A permutation is an arrangement of all or part of a set of objects (ordered samples without replacement). If we are interested in the numbers of ways of selecting r objects from m without regard to order, that is combination.

- If the order doesn't matter, it is a Combination.
- If the order does matter it is a Permutation.

If you say : "My fruit salad is a combination of apples, grapes and bananas" you don't care what order the fruits are in, they could also be "bananas, grapes and apples" or "grapes, apples and bananas", it’s the same fruit salad. (Combination)

"The combination to the safe is 472". Now we do care about the order. "724" won't work, nor will "247". It has to be exactly 4-7-2. (Permutation)

If the order doesn't matter, it is a Combination.

If the order does matter it is a Permutation.

The factorial function (symbol: !) just means to multiply a series of descending natural numbers. Examples:

$4! = 4 \times 3 \times 2 \times 1 = 24$

$7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5{,}040$

$1! = 1$

(The formulas for permutation and combination can be found in the notes)

We discussed different approaches to interpret probabilities. We defined probabilities through axioms on Thursday and construct some probability rules. We gave the definition of a probability by using three axioms: positivity, certainty and additivity. We constructed some probability rules, in particular addition rules.

You should now be able to do some of the suggested problems from the probability section of the bank of questions. Questions: 2, 3, 10, 11, 12, 13, 22, 28- 33 and 35.

Week 2 (Jan 16^{th} and Jan 18^{th})

This week we talked about the conditional probability, the total probability rule, Bayes Rule, and the notion of independent events.

A conditional probability is the probability of an event given that (by assumption) another event has occurred.[1] If the event of interest is A and the event B is known or assumed to have occurred, "the conditional probability of A given B", or "the probability of A under the condition B", is usually written as P(A|B).

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

The relationship between *P*(*A*|*B*) and *P*(*B*|*A*) is given by Bayes' theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

The events A and B are defined to be independent if the occurrence of A does not affect the probability of B, and vice versa.

$$\begin{aligned} &P(A \cap B) = P(A)P(B) \\ \Leftrightarrow\ &P(A|B) = P(A) \\ \Leftrightarrow\ &P(B|A) = P(B) \end{aligned}$$

You should now be able to do some of the suggested problems from the probability section of the bank of questions. Questions: 1,2,4,5,6,7,9,14,16,17,18,19, 20 and 21

Week 3 (Jan 23rd and Jan 25th)

This week we defined a random variable as a real valued function from the sample space to the real numbers. A random variable allows us to define events in terms of numbers. For a random variable X, we will specify probabilities in different ways:

- With a cumulative distribution function F, where F(x)=P(X<=x), where x is a real number. It is a cumulation of probabilities up to the value x.
- With a probability mass function f (if X is discrete), where f(x)=P(X=x) for x in the range of the random variable.

We defined the expectation of a h(X), where X is a discrete random variable. From this, we defined different descriptive parameters of the distribution of a random variable: The mean which is E[X], its variance

V[X]=E[(X-E[X])^2] and the standard deviation (V[X])^{1/2}.

The mean is a measure of central tendency or of location of the distribution and the variance is a measure (in square units) of the dispersion of the distribution or variability of the random variable. The standard deviation is also a measure of dispersion but in the original units of measurement.

We saw some an alternate formula for the computation of the variance and also results concerning a linear transformation of a random variable. For those that are interested here are proofs for these formulas: Expectation.pdf

A continuous random variable is a random variable where the data can take infinitely many values (the range is an interval). For example, a random variable measuring the time taken for something to be done is continuous For continuous random variables, the probability that X takes on any particular value x is 0. That is, finding P(X = x) for a continuous random variable X is not going to work. Instead, we'll need to find the probability that X falls in some interval (a, b), that is, we'll need to find P(a < X < b). We'll do that using a probability density function ("p.d.f.").

For any continuous random variable with probability density function f(x), we have that:

The probability density function ("p.d.f.") of a continuous random variable X with support S is an integrable function f(x) satisfying the following:

(1) f(x) is positive everywhere in the support S, that is, f(x) > 0, for all x in S

(2) The area under the curve f(x) in the support S is 1, that is:

$\int_S f(x)dx=1$

(3) If f(x) is the p.d.f. of x, then the probability that x belongs to A, where A is some interval, is given by the integral of f(x) over that interval, that is:

$P(X\in A)=\int_A f(x)dx$

An implication of the fact that $P(X = x) = 0$ for all x when X is continuous is that you can be careless about the endpoints of intervals when finding probabilities of continuous random variables. That is:

The cumulative distribution function ("c.d.f.") of a continuous random variable X is defined as:

$$F(x)=\int_{-\infty}^{x} f(t)dt \qquad \text{for } -\infty < x < \infty.$$

If the probability distribution of X admits a probability density function f(x), then the expected value and the variance can be computed as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x.$$

$$\mathrm{Var}(X) = \sigma^2 = \int (x-\mu)^2 f(x)\,dx = \int x^2 f(x)\,dx - \mu^2$$

Here are some suggested exercises that involve working with a discrete random variable bank of questions: 3, 8, 19, 22, 37, 38, 43, and 46.

Week 4 (Jan 30th and Feb 1st)

We have finished the discussion regarding Continuous Random Variables. Here are some suggested exercises that involve working with a Continuous Random Variables from random variables bank of questions: **4, 5, 24, 35 and 43.**

This week we have started Discrete Joint Probability. Joint probability distributions describe situations where by both outcomes represented by random variables occur. While we only X to represent the random variable, we now have X and Y as the pair of random variables. Joint probability distributions are defined in the form below:

$$f(x,y) = \mathrm{P}(X = x, Y = y)$$

The probability function, also known as the probability mass function for a joint probability distribution f(x,y) is defined such that:

$f(x,y) \geq 0$ for all (x,y)

Which means that the joint probability should always greater or equal to zero as dictated by the fundamental rule of probability:

$\sum_x \sum_y f(x,y) = 1$

Which means that the sum of all the joint probabilities should equal to one for a given sample space.

$f(x,y) = P(X = x, Y = y)$

The mass probability function f(x,y) can be calculated in a number of different ways depend on the relationship between the random variables X and Y.

As we saw in the section on probability concepts, these two variables can be either independent or dependent.

If X and Y are Independent:

$$f(x,y) = f(x) \times f(y)$$

For more information and details, please check the notes.

Here are some suggested exercises that involve working with a discrete joint distribution from random variables bank of questions: **53-55 and 58**

Today, we started to introduce some discrete probability models. We defined a Bernoulli trial. For independent Bernoulli trials with a constant probability of success, we defined the binomial distribution (for the number of successes among n trials

For questions in the bank of questions involving these distributions, try the following question from bank of questions: **1, 2, 9, 14, 20, 26, 32, 39, 44, and 52** (Random variables)

Week 5 (Feb 6th and Feb 8th)

This week we have finished the discussion of discrete probability models. For independent Bernoulli trials with a constant probability of success, we defined the binomial distribution (for the number of successes among n trials) and the geometric distribution (for the number of trials required to observe one success).We defined a negative binomial model. Suppose there is a sequence of independent Bernoulli trials, each trial having two potential outcomes called "success" and "failure". In each trial the probability of success is p and of failure is (1 − p). We are observing this sequence until a predefined number r of failures has occurred. Then the random number of successes we have seen, X, will have the negative binomial (or Pascal) distribution:

$$f(k;r,p) \equiv \Pr(X=k) = \binom{k+r-1}{k} p^k (1-p)^r \quad \text{for } k = 0,1,2,\dots$$

We also defined the Poisson distribution and mentioned how it is related to a Poisson process. We will finish the discussion concerning the Poisson process after the break: We will define three types of variables: (i) number of events over a fixed length interval, which has a Poisson distribution; (ii) length (which could be time) of an interval required to observe one event, which has an exponential distribution; (iii) length (which could be time) of an interval required to observe r events, which has an Erlang distribution. For questions in the bank of questions involving these distributions, try the following question from bank of questions: 7, 10, 12,13, 21,23, 29, 33,36,40, 41,45, 47, 50 (Random variables)

Topics for Test 1:

Test 1 will be held on Feb 15th at 8:30 in class. The test includes all the materials from the beginning up to and including Negative Binomial (Lecture 1- (most of the) Lecture 9)

Week 6 (Feb 13th and Feb 15th)

We finished the discussion regarding the Exponential and Erlang distributions.

The exponential distribution is the probability distribution that describes the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate. It is a particular case of the gamma distribution. It is the continuous analogue of the geometric distribution, and it has the key property of being memoryless. (Lack of memory)

The Erlang distribution with shape parameter k = 1 simplifies to the exponential distribution. It is a special case of the Gamma distribution. It is the distribution of a sum of k independent exponential variables with mean 1 / λ each.

For questions in the bank of questions involving these distributions, try the following question from bank of questions: 16, 23, 30 (Random variables)

Week 7 (Feb27th –March 1st)

We have introduced the normal distribution. We learned how to use Table III, which is a table for the c.d.f. of a standard normal (i.e. a normal with mean 0 and standard deviation 1). When working with a normal distribution with a particular mean and standard deviation, we must use the standardization theorem to use Table III. The standardization theorem says that if X is normal with a mean mu and standard deviation sigma, then Z=(X-mu)/sigma has a standard normal distribution.

For questions in the bank of questions involving the normal distribution, try the following questions from bank of questions : 17, 18, 25, 42, 48 and 51 (Random variables)

We have started Chapter 4 and discussed the estimation of the population mean, the population standard deviation and a population proportion using the observed values of a random sample. For the estimation, we use the sample mean, the sample standard deviation and the sample proportion, respectively.

Today, we discussed ways to describe the distribution of the sample.

- Measures of central tendency: median and mean
- Measures of dispersion: standard deviation, range and interquartile range (IQR)

We talk about the measure of centre (Sample Mean, Sample Median and Mode) and also measure of dispersion (Sample variance, Range and IQR).

We talk about how to locate median, Q1 (lower quartile) and Q3 (upper quartile) in the class. IQR=Q3-Q1.

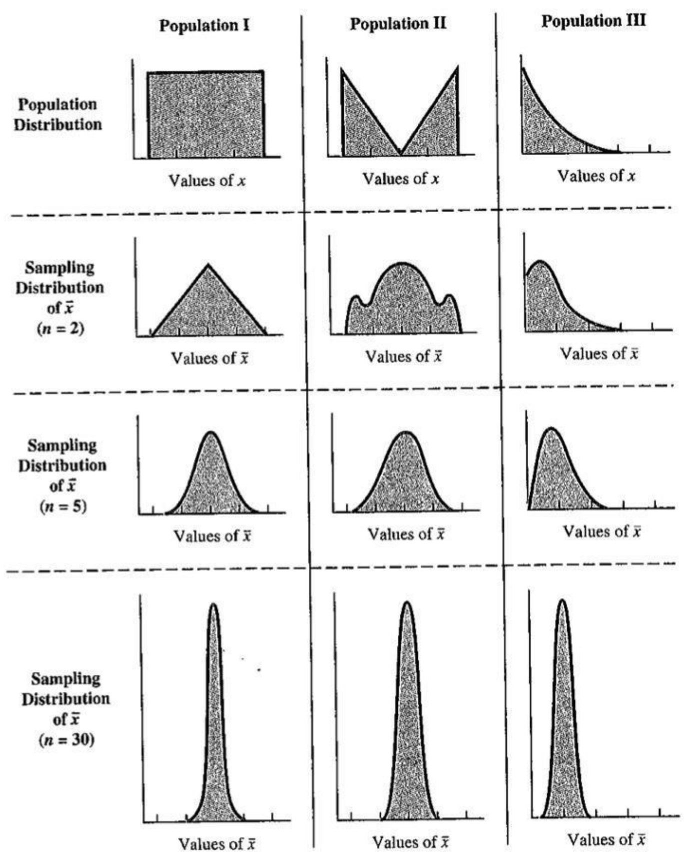
The IQR is the length of the box in your boxplot. An outlier is any value that lies more than one and a half times the length of the box from either end of the box. That is, if a data point is below Q1 – 1.5×IQR or above Q3 + 1.5×IQR, it is viewed as being too far from the central values to be reasonable.

For questions in the bank of questions involving descriptive statistics, try the following questions from bank of questions : 1, 7, 8, 9, 10, 11 (Descriptive Statistics)

Week 8 (March 6th –March 8th)

For the last two lectures, we have been discussing point estimation and sampling distributions for mean, variance and proportion. We have also discussed the estimation of the population mean, the population standard deviation and a population proportion using the observed values of a random sample. For questions in the bank of questions involving the sampling distribution of the sample mean and involving point estimation of a mean, of a variance or of a proportion, try the following questions from bank of questions : 2, 4, 5, 6, 7, 8, 9, 10, 16, 19, 20, 22, 23, 26, 27, 28, 30, 31 (Point Estimation and Sampling Distributions)

Central Limit Theorem: We see two results concerning the sampling distribution of a sample mean: (i) if the population is normally distributed, then the sample mean is also normally distributed, (ii) if the population is not necessarily distributed, however the sample size is large enough (n>=30), then we can approximate the sampling distribution for the sample mean with a normal distribution.



We have also started using soft R.

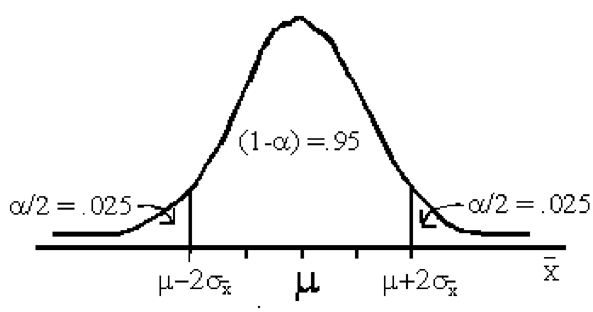
Week 9 (March 13th- 15th)

We started the discussion on the estimation of a mean using a confidence interval. There are two important things that we learned: (i) how to construct a confidence interval for the mean; (ii) finding the appropriate sample size n to be confident that the error in estimating the mean is no more than E.

Confidence Interval:

We discuss the interval estimation for the population mean and proportion. We saw three cases.

The 95% confidence interval for μ



1. Conditions: (i) Normal population or large sample (n>30); (ii) population standard deviation is known;

2. Conditions: normal population and population standard deviation is unknown (n<30);

3. Conditions: (i) large sample size (n at least 40) or Normal population (n at least 30) and population standard deviation is unknown.

For the first and third case, we construct a z-based confidence interval. For the second case, we construct a t-based confidence interval.

For questions in the bank of questions involving confidence intervals for estimating a population mean, try the following questions from bank of questions: 1-15 (Confidence intervals).

Z-scores (Z(alpha/2)) for confidence intervals:

Confidence Level	z
0.90	1.645
0.92	1.755
0.95	1.96
0.96	2.055
0.98	2.325
0.99	2.575

Topics for Test 1:

Test 1 will be held on March 22nd at 8:30 in class. The test includes all the materials after Test 1 (Starting from Poisson distribution) up to and including CI for proportions. ((part of) Lecture 9-Lecture 17)

Week 10 (March 20th- 22nd)

Hypothesis Testing

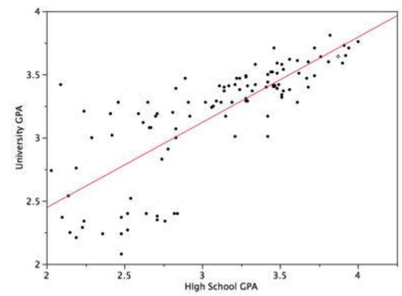
We introduce some basic notions from hypothesis testing. We learned how to formulate a null hypothesis and an alternative hypothesis to test hypotheses concerning a population parameter. We discussed errors of Type I (i.e. rejection H_0 when H_0 is true) and of Type II (i.e. failing to reject H_0 when H_1 is true). We defined the critical region method and a p-value as a measure the significance of our evidence against the null in favour of the alternative. We interpret a small p-value as strong evidence against the null hypothesis.

We use a significance level to determine which level we will consider the evidence as being significant. Decision Rule: if the p-value is less than the level of significance, then we have significance evidence against H0 in favor of H1.

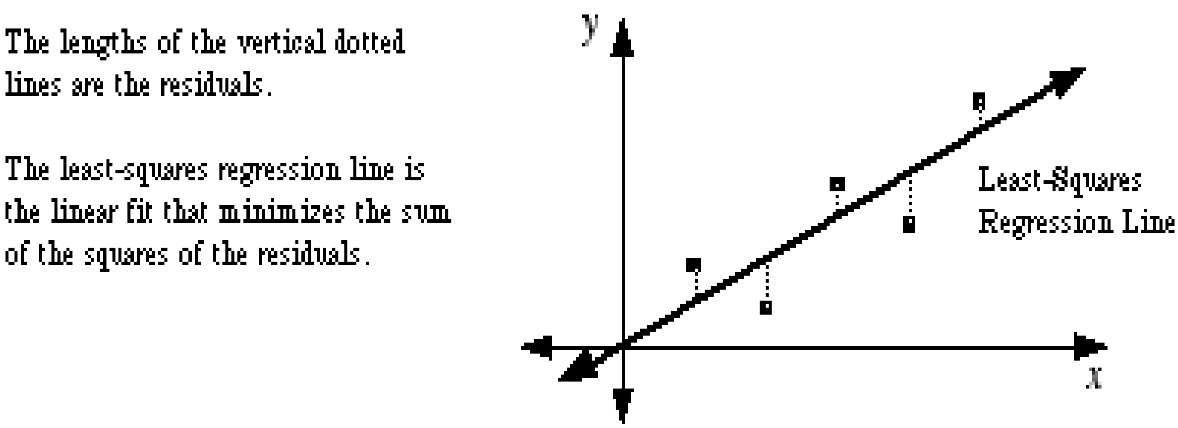
Please try the following questions from the bank of questions: 1-6, 9, 15 (Hypothesis testing for mean and proportion).

Week 11 and 12 (March 27th- April 5th)

We have finished the discussion regarding hypothesis testing and have started Chapter 7. Linear regression consists of finding the best-fitting straight line through the points. The best-fitting line is called a regression line.



The linear fit that matches the pattern of a set of paired data as closely as possible. Out of all possible linear fits, the least-squares regression line is the one that has the smallest possible value for the sum of the squares of the residuals.



We have learnt in the class to find the regression line and Pearson's correlation.

Please try the following questions from the bank of questions: 1 and 4.

(Regression)

We did review and the practice final. Good luck on the Final exam!