UQ, Semester 1, 2019,
Companion to STAT2201 Exam
Formulae and Tables

Probability and Monte Carlo

- An experiment that can result in different outcomes, even though it is repeated in the same manner every time, is called a **random experiment**.
- The set of all possible outcomes of a random experiment is called the **sample space** of the experiment, and is denoted as Ω.
  - A sample space is **discrete** if it consists of a finite or countably infinite set of outcomes.
  - A sample space is **continuous** if it contains an interval (either finite or infinite) of real numbers, vectors or similar objects.

- An **event** is a subset of the sample space of a random experiment.
  - The **union** of two events is the event that consists of all outcomes that are contained in either of the two events or both. We denote the union as $E_1 \cup E_2$.
  - The **intersection** of two events is the event that consists of all outcomes that are contained in both of the two events. We denote the intersection as $E_1 \cap E_2$.
  - The **complement** of an event in the sample space is the set of outcomes in the sample space that are not in the event. We denote the complement of the event E as $\overline{E}$. The notation E^c is also used. Note that $E \cup \overline{E} = \Omega$.
- Two events, denoted E_1 and E_2 are **mutually exclusive** if: $E_1 \cap E_2 = \varnothing$ where $\varnothing$ is called the **empty set** or **null event**.
- A collection of events, $E_1, E_2, \ldots, E_k$ is said to be **mutually exclusive** if for all pairs,

$$E_i \cap E_j = \varnothing.$$

- The definition of the complement of an event implies that: $(E^c)^c = E$.
- The distributive law for set operations implies that

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \qquad \text{and} \qquad (A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

- DeMorgan's laws imply that

$$(A \cup B)^c = A^c \cap B^c \qquad \text{and} \qquad (A \cap B)^c = A^c \cup B^c.$$

- Union and intersection are commutative operations: $A \cap B = B \cap A$ and $A \cup B = B \cup A$.

- **Probability** is used to quantify the likelihood, or chance, that an outcome of a random experiment will occur.
- Whenever a sample space consists of a finite number N of possible outcomes, each **equally likely**, the probability of each outcome is $1/N$.

- For a discrete sample space, the **probability of an event** E, denoted as $P(E)$, equals the sum of the probabilities of the outcomes in E.
- If Ω is the sample space and E is any event in a random experiment,
  (1) $P(\Omega) = 1$.
  (2) $0 \leq P(E) \leq 1$.
  (3) For two events E_1 and E_2 with $E_1 \cap E_2 = \varnothing$ (disjoint),
  $$P(E_1 \cup E_2) = P(E_1) + P(E_2).$$
  (4) $P(E^c) = 1 - P(E)$.
  (5) $P(\emptyset) = 0$.
- The probability of event A or event B occurring is,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
- If A and B are mutually exclusive events,
$$P(A \cup B) = P(A) + P(B).$$
- For a collection of **mutually exclusive events**,
$$P(E_1 \cup E_2 \cup \cdots \cup E_k) = P(E_1) + P(E_2) + \cdots + P(E_k).$$

- The probability of an event B under the knowledge that the outcome will be in event A is denoted $P(B \mid A)$ and is called the **conditional probability** of B given A.
- The **conditional probability** of an event B given an event A, denoted as $P(B \mid A)$, is
$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \quad \text{for} \quad P(A) > 0.$$
- The **multiplication rule** for probabilities is: $P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B)$.
- For an event B and a collection of mutual exclusive events, $E_1, E_2, \ldots, E_k$ where their union is Ω. The **law of total probability** yields,
$$\begin{aligned} P(B) &= P(B \cap E_1) + P(B \cap E_2) + \cdots + P(B \cap E_k) \\ &= P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k). \end{aligned}$$

- Two events A and B are **independent** if any one of the following equivalent statements is true:
  (1) $P(A \mid B) = P(A)$.
  (2) $P(B \mid A) = P(B)$.
  (3) $P(A \cap B) = P(A)P(B)$.

  Observe that **independent** events and **mutually exclusive** events, are completely different concepts. Don't confuse these concepts.
- For **multiple events** $E_1, E_2, \ldots, E_n$ are independent if and only if for any subset of these events
$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = P(E_{i_1})\ P(E_{i_2})\ \ldots\ P(E_{i_k}).$$

Distributions

- A **random variable** X is a numerical (integer, real, complex, vector, etc.) summary of the outcome of the random experiment. The **range** or **support** of the random variable is the set of possible values that it may take. Random variables are usually denoted by capital letters.
- A **discrete random variable** is an integer/real-valued random variable with a finite (or countably infinite) range.
- A **continuous random variable** is a real-valued random variable with an interval (either finite or infinite) of real numbers for its range.
- The **probability distribution** of a random variable X is a description of the probabilities associated with the possible values of X. There are several common alternative ways to describe the probability distribution, with some differences between discrete and continuous random variables.
- While not the most popular in practice, a unified way to describe the distribution of any scalar valued random variable X (real or integer) is the **cumulative distribution function**,

$$F(x) = P(X \le x).$$

- It holds that

  (1) $0 \le F(x) \le 1$.

  (2) $\lim_{x \to -\infty} F(x) = 0$.

  (3) $\lim_{x \to \infty} F(x) = 1$.

  (4) If $x \le y$, then $F(x) \le F(y)$. That is, $F(\cdot)$ is non-decreasing.

- Distributions are often summarised by numbers such as the **mean**, μ, **variance**, σ^2, or **moments**. These numbers, in general do not identify the distribution, but hint at the general location, spread and shape.
- The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$ and is particularly useful when working with the Normal distribution.

<u>**Discrete Random Variable**</u>

- Given a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, the **probability mass function** of X is,

$$p(x) = P(X = x).$$

  Note: Often in literature, the notation used is $f(x)$ (as a pdf of a continuous random variable).

- A probability mass function, $p(x)$ satisfies:

  (1) $p(x_i) \ge 0$.

  (2) $\sum_{i=1}^{n} p(x_i) = 1$.

- The **cumulative distribution function** of a discrete random variable X, denoted as $F(x)$, is

$$F(x) = \sum_{x_i \le x} p(x_i).$$

- $P(X = x_i)$ can be determined from the *jump* at the value of x. More specifically

$$p(x_i) = P(X = x_i) = F(x_i) - \lim_{x \uparrow x_i} F(x_i).$$

- The **mean** or **expected value** of a discrete random variable X, is

$$\mu = E(X) = \sum_x x\, p(x).$$

- The **expected value** of $h(X)$ for some function $h(\cdot)$ is:

$$E\Big[h(X)\Big] = \sum_x h(x)\, p(x).$$

- The k'th **moment** of X is,

$$E(X^k) = \sum_x x^k\, p(x).$$

- The **variance** of X, is

$$\sigma^2 = V(X) = E\big((X-\mu)^2\big) = E\big(X^2\big) - \mu^2 = \sum_x (x-\mu)^2\, p(x) = \sum_x x^2\, p(x) - \mu^2.$$

- A random variable X has a **discrete uniform distribution** if each of the n values in its range, $x_1, x_2, \ldots, x_n$, has equal probability. I.e.

$$p(x_i) = 1/n.$$

- **Binomial Distribution $X \sim Bin(n, p)$:**
The Binomial distribution is related to the n independent and identical Bernoulli trials as follows:

  (1) There are n Bernoulli trials.
  (1) The trials are independent.
  (2) Each trial results in only two possible outcomes, labelled as "success" and "failure".
  (3) The probability of a success in each trial denoted as p is the same for all trials.

  The random variable X that equals the number of trials that result in a success is a **binomial random variable** with parameters $0 \le p \le 1$ and $n = 1, 2, \ldots$. The probability mass function of X is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

  with **mean** and **variance** given by

$$E(X) = n\,p \qquad \text{and} \qquad V(X) = n\,p\,(1-p).$$

- **Geometric distribution $X \sim G(p)$:**
The random variable X that equals the number of trials until the first success of a Bernoulli experiment with success probability $0 \le p \le 1$, has a geometric distribution. The probability mass function of X is

$$p(x) = (1-p)^{x-1} p, \qquad x \ge 1.$$

  with **mean** and **variance** given by

$$E(X) = \frac{1}{p} \qquad \text{and} \qquad V(X) = \frac{1-p}{p^2}.$$

- **Useful Formula:**

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}, \qquad \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

Continuous Random Variable

- Given a continuous random variable X, the **probability density function** (pdf) is a function, $f(x)$ such that,
  (1) $f(x) \geq 0$.
  (2) $f(x) = 0$ for x not in the range.
  (3) $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
  (4) For small Δx, $f(x)\,\Delta x \approx P(X \in [x, x + \Delta\, x))$.
  (5) $P(a \leq X \leq b) = \int_a^b f(x)dx =$ area under $f(x)$ from a to b.
- Given the pdf, $f(x)$ we can get the cdf as follows:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du \qquad \text{for} \qquad -\infty < x < \infty.$$

- Given the cdf of a continuous random variable, $F(x)$ we can get the pdf:

$$f(x) = \frac{d}{dx}\, F(x).$$

- The **mean** or **expected value** of a continuous random variable X, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)dx.$$

- The **expected value** of $h(X)$ for some function $h(\cdot)$ is:

$$E\Big[h(X)\Big] = \int_{-\infty}^{\infty} h(x)f(x)\, dx.$$

- The k'th **moment** of X is,

$$E(X^k) = \int_{-\infty}^{\infty} x^k\, f(x)\, dx.$$

- The **variance** of X, is

$$\sigma^2 = V(X) = E\big((X-\mu)^2\big) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2.$$

- **Uniform Distribution $X \sim U(a, b)$:**
  A continuous random variable X on the domain $[a, b]$, (a, b), $[a, b)$, $(a, b]$ has the probability density function

$$f(x) = \frac{1}{b-a}, \qquad a \leq x \leq b,$$

  with **mean** and **variance** given by

$$\mu = E(X) = \frac{a+b}{2} \qquad \text{and} \qquad \sigma^2 = V(X) = \frac{(b-a)^2}{12}.$$

- **Normal Distribution** $X \sim N(\mu, \sigma^2)$**:**
  A random variable X with probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty,$$

  is a **normal random variable** with parameters μ where $-\infty < \mu < \infty$, and $\sigma > 0$. The mean and variance are

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2.$$

- A normal random variable with a mean and variance of:

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

  is called a **standard normal random variable** and is denoted as Z. The cumulative distribution function of a standard normal random variable is denoted as

$$\Phi(z) = F_Z(z) = P(Z \leq z),$$

  and is tabulated.

- It is very common to compute $P(a < X < b)$ for $X \sim N(\mu, \sigma^2)$. This is the typical way:

$$\begin{aligned} P(a < X < b) &= P(a - \mu < X - \mu < b - \mu) \\ &= P\Big(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\Big) \\ &= P\Big(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\Big) \\ &= \Phi\Big(\frac{b-\mu}{\sigma}\Big) - \Phi\Big(\frac{a-\mu}{\sigma}\Big). \end{aligned}$$

  We get:

$$F_X(b) - F_X(a) = F_Z\Big(\frac{b-\mu}{\sigma}\Big) - F_Z\Big(\frac{a-\mu}{\sigma}\Big).$$

- **Exponential Distribution** $X \sim Exp(\lambda)$**:**
  $(\lambda > 0)$ The probability density function of X is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \leq x < \infty$$

  with cdf

$$\overline{F}(x) = 1 - F(x) = P(X > x) = e^{-\lambda x}.$$

  The **mean** and **variance** are:

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2}.$$

  The exponential distribution is the only continuous distribution with range $[0, \infty)$ exhibiting the **lack of memory property**. For an exponential random variable X,

$$P(X > t + s \,|\, X > t) = P(X > s).$$

- **Useful Formula:** $a, b, \delta \in \mathbb{R}$

$$\int_a^b x e^{\delta x} \, \mathrm{d}x = \frac{1}{\delta}\left[e^{\delta b} - e^{\delta a}\right]$$

Joint Probability Distributions

- A joint probability distribution of two random variables is also referred to as a **bivariate probability distribution**.

- A **joint probability mass function** for discrete random variables X and Y, denoted as $p_{XY}(x, y)$, satisfies the following properties:

  (1) $p_{XY}(x, y) \geq 0$ for all x, y.

  (2) $p_{XY}(x, y) = 0$ for (x, y) not in the range.

  (3) $\sum\sum p_{XY}(x, y) = 1$, where the summation is over all (x, y) in the range.

  (4) $p_{XY}(x, y) = P(X = x,\ Y = y)$.

- A **joint probability density function** for continuous random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies the following properties:

  (1) $f_{XY}(x, y) \geq 0$ for all x, y.

  (2) $f_{XY}(x, y) = 0$ for (x, y) not in the range.

  (3) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\ dx\ dy = 1$.

  (4) For small $\Delta x,\ \Delta y$: $\quad f_{XY}(x, y)\,\Delta x\,\Delta y \approx P\Big((X, Y) \in [x, x + \Delta\, x) \times [y, y + \Delta\, y)\Big)$.

  (5) For any region R of two-dimensional space,

  $$P\Big((X, Y) \in R\Big) = \iint_R f_{XY}(x, y)\ dx\ dy.$$

- A **joint probability density function** can also be defined for $n > 2$ random variables (as can be a **joint probability mass function**). The following needs to hold:

  (1) $f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) \geq 0$.

  (2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) dx_1\ dx_2 \dots dx_n = 1$.

- Most of the concepts in this section, carry over from bivariate to general multivariate distributions ($n > 2$).

- The **marginal distributions** of X and Y as well as the **conditional distribution** of X given a specific value $Y = y$ and vice versa can be obtained from the joint distribution.

- If the random variables X and Y are independent, then $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ and similarly in the discrete case.

- The **expected value of a function of two random variables** is:

$$E\Big[h(X, Y)\Big] = \iint h(x, y) f_{XY}(x, y)\ dx\ dy \qquad \text{for } X, Y \text{ continuous.}$$

- The **covariance** is a common measure of the relationship between two random variables (say X and Y). It is denoted as $\mathrm{cov}(X, Y)$ or σ_{XY}, and is given by:

$$\sigma_{XY} = E\Big[(X - \mu_X)(Y - \mu_Y)\Big] = E(XY) - \mu_X \mu_Y.$$

- The covariance of a random variable with itself is its variance.
- The **correlation** between the random variables X and Y, denoted as ρ_{XY}, is

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

- For any two random variables X and Y, $-1 \leq \rho_{XY} \leq 1$.
- If X and Y are independent random variables, $\sigma_{XY} = 0$ and $\rho_{XY} = 0$. The opposite case does not always hold: In general $\rho_{XY} = 0$ does not imply independence. But for jointly Normal random variables it does. In any case, if $\rho_{XY} = 0$ then the random variables are called uncorrelated.
- When considering several random variables, it is common to consider the (symmetric) **Covariance Matrix**, Σ with $\Sigma_{i,j} = \text{cov}(X_i, X_j)$.

- **Independence:**
Two discrete random variables X, Y are independent if

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

for all $x \in \Omega_X$ and $y \in \Omega_Y$, where Ω_X is the range of X and Ω_Y is the range of Y.
Two continuous random variables X, Y are independent if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

where $f_X(x) = \int_{y \in \Omega_Y} f_{X,Y}(x,y)\,\mathrm{d}y$ is the marginal probability density function of X and $f_Y(y) = \int_{x \in \Omega_X} f_{X,Y}(x,y)\,\mathrm{d}x$ is the marginal probability density function of Y.

- The **probability density function** of a **bivariate normal distribution** is

$$f_{XY}(x,y;\sigma_X,\sigma_Y,\mu_X,\mu_Y,\rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$
$$\times \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right] \right\}$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$,
with parameters $\sigma_X > 0$, $\sigma_Y > 0$, $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, and $-1 < \rho < 1$.

- Given random variables $X_1, X_2, \ldots, X_n$ and constants $c_1, c_2, \ldots, c_n$, the (scalar) **linear combination** (with possible affine term b),

$$Y = c_1X_1 + c_2X_2 + \cdots + c_nX_n + b$$

is often a random variable of interest.

- The mean of the linear combination is the linear combination of the means,

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \cdots + c_nE(X_n) + b$$

This holds even if the random variables are not independent.

- The variance of the linear combination is as follows:

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_n^2V(X_n) + 2\sum_{i<j}\sum c_ic_j\text{cov}(X_i,X_j).$$

- If $X_1, X_2, \ldots, X_n$ are **independent** (or even if they are just uncorrelated).

$$V(Y) = c_1^2 V(X_1) + c_2^2 V(X_2) + \cdots + c_n^2 V(X_n).$$

- In case the random variables $X_1, \ldots, X_n$ were jointly Normal then, $Y \sim \text{Normal}\Big(E(Y), V(Y)\Big)$. That is, **linear combinations of Normal random variables remain Normally distributed**.

- A collection of random variables, $X_1, \ldots, X_n$ is said to be **i.i.d.**, or **independent and identically distributed** if they are mutually independent and identically distributed. This means that the (n - dimensional) joint probability density is a product of the individual densities.
- In the context of statistics, a **random sample** is often modelled as an i.i.d. vector of random variables. $X_1, \ldots, X_n$.
- An important linear combination associated with a random sample is the **sample mean**:

$$\overline{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n.$$

- If X_i has mean μ and variance σ^2 then sample mean (of an i.i.d. sample) has,

$$E(\overline{X}) = \mu, \qquad V(\overline{X}) = \frac{\sigma^2}{n}.$$

Descriptive Statistics

- **Descriptive statistics** deals with summarizing **data** using numbers, qualitative summaries, tables and graphs.
- Here are some types of **data configurations**:
  1. Single sample: $x_1, x_2, \ldots, x_n$.
  2. Single sample over time (time series): $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$ with $t_1 < t_2 < \ldots < t_n$.
  3. Two samples: $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$.
  4. Generalizations from two samples to k samples (each of potentially different sample size, $n_1, \ldots, n_k$).
  5. Observations in tuples: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
  6. Generalizations from tuples to vector observations (each vector of length ℓ),

$$(x_1^1, \ldots, x_1^\ell), \ldots, (x_n^1, \ldots, x_n^\ell).$$

- Individual **variables** may be **categorical** or **numerical**. Categorical variables (taking values in one of several categories) may be **ordinal** meaning that they can be sorted (e.g. "low", "moderate", "high"), or not (e.g. "cat", "dog", "fish").

- A **statistic** is a quantity computed from a sample (assume here a single sample $x_1, \ldots, x_n$). Here are very common and useful statistics:
  1. The **sample mean**: $\bar{x} = \dfrac{x_1 + \cdots + x_n}{n} = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$.

2. The **sample variance**: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2}{n-1}$.
3. The **sample standard deviation**: $s = \sqrt{s^2}$.
4. The **sample correlation coefficient** r_{xy} is an estimate for the correlation coefficient, ρ, presented in the previous unit:

$$r_{xy} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2 \sum_{i=1}^{n}(x_i - \overline{x})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}n}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2\right)\left(\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2\right)}}.$$

5. **Order statistics** work as follows: Sort the sample to obtain the sequence of sorted observations, denoted $x_{(1)}, \ldots, x_{(n)}$ where, $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$. Some common order statistics:
   (a) The **minimum** $\min(x_1, \ldots, x_n) = x_{(1)}$.
   (b) The **maximum** $\max(x_1, \ldots, x_n) = x_{(n)}$.
   (c) The **median**

$$\text{median} = \begin{cases} x_{(\frac{n+1}{2})} & \text{if } n \text{ is odd,} \\ \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}\right) & \text{if } n \text{ is even.} \end{cases}$$

   Note that the median is the 50'th percentile and the 2nd quartile (see below).
   (d) The q th **quantile** ($q \in [0,1]$) or alternatively the $p = 100q$ **percentile** (measured in percents instead of a decimal), is the observation such that p percent of the observations are less than it and $(1-p)$ percent of the observations are greater than it. In cases (as is typical) that there is not such a precise observation, it is a linear interpolation between two neighbouring observations (as is done for the median when n is even). In terms of order statistics, the q th quantile is approximately (not taking linear interpolations into account) $x_{([q*n])}$. Here $[z]$ denotes the nearest integer in $\{1, \ldots, n\}$ to z.
   (e) The first **quartile**, denoted $Q1$ is the 25th percentile. The second quartile ($Q2$) is the median. The **third quartile**, denoted $Q3$ is the 75th percentile. Thus half of the observations lie between $Q1$ and $Q3$. In other words, the quartiles break the sample into 4 quarters. The difference $Q3 - Q1$ is the **interquartile range**.
   (f) The **sample range** is $x_{(n)} - x_{(1)}$.

- In a **cumulative frequency plot** the height of each bar is the total number of observations that are less than or equal to the upper limit of the bin.

- The **Empirical Cumulative Distribution Function** (ECDF) is,

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{x_i \le x\}.$$

  Here $\mathbf{1}\{\cdot\}$ is the **indicator function**. The ECDF is a function of the data, defined for all x.

- Given a **candidate distribution** with cdf $F(x)$, a **probability plot** is a plot of the ECDF (or sometimes just it's jump points) with the y-axis stretched by the inverse of the cdf $F^{-1}(\cdot)$. The monotonic transformation of the y-axis is such that if the data comes from the candidate $F(x)$, the points would appear to lie on a straight line. Names of variations of probability plots are the **P-P plot** and **Q-Q plot** (these plots are similar to the probability plot). A very common probability plot is the **Normal probability plot** where the candidate distribution is taken to be Normal($\overline{x}, s^2$).

- The Normal probability plot can be useful in identifying distributions that are symmetric but that have tails that are "heavier" or "lighter" than the Normal.

Statistical Inference Ideas

- **Statistical Inference** is the process of forming judgements about the **parameters of a population**, typically on the basis of **random sampling**.
- The random variables $X_1, X_2, \ldots, X_n$ are an (i.i.d.) **random sample** of size n if
  (a) the X_i's are independent random variables and
  (b) every X_i has the same probability distribution.
- A **statistic** is any function of the observations in a random sample, and the probability distribution of a statistic is called the **sampling distribution**.
- Any function of the observation, or any **statistic**, is also a random variable. We call the probability distribution of a statistic a **sampling distribution**. A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.
- The most common statistic we consider is the **sample mean**, $\overline{X}$, with a given value denoted by $\overline{x}$. As an estimator, the sample mean is an estimator of the population mean, μ.

- **Central Limit Theorem** (for sample means):
  If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population with mean μ and finite variance σ^2 and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

  as $n \to \infty$, is the standard normal distribution.
- This implies that $\overline{X}$ is approximately normally distributed with mean μ and standard deviation $\sigma/\sqrt{n}$.
- The **standard error** of $\overline{X}$ is given by $\sigma/\sqrt{n}$. In most practical situations σ is not known but rather estimated in this case, the **estimated standard error**, (denoted in typical computer output as "SE"), is $s/\sqrt{n}$ where s is the point estimator,

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2}{n-1}}.$$

- **Central Limit Theorem** (for sums):
  Manipulate the central limit theorem (for sample means and use $\sum_{i=1}^{n} X_i = n\overline{X}$. This yields,

$$Z = \frac{\sum_{i=1}^{n} X_i - n\,\mu}{\sqrt{n\sigma^2}},$$

  which follows a standard normal distribution as $n \to \infty$.

- This implies that $\sum_{i=1}^{n} X_i$ is approximately normally distributed with mean $n\,\mu$ and variance $n\,\sigma^2$.

- Knowing the sampling distribution (or the approximate sampling distribution) of a statistic is the key for the two main tools of statistical inference that we study:

  (a) **Confidence intervals** – a method for yielding error bounds on **point estimates**.

  (b) **Hypothesis testing** – a methodology for making conclusions about population parameters.

- The formulas for most of the statistical procedures use **quantiles of the sampling distribution**. When the distribution is $N(0,1)$ (standard normal), the α quantile is denoted z_α and satisfies:

$$\alpha = \int_{-\infty}^{z_\alpha} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx.$$

  A common value to use for α is 0.05 and in procedures the expressions $z_{1-\alpha}$ or $z_{1-\alpha/2}$ appear. Note that in this case $z_{1-\alpha/2} = 1.96 \approx 2$.

- A **confidence interval** estimate for μ is an interval of the form $l \le \mu \le u$, where the end-points l and u are computed from the sample data. Because different samples will produce different values of l and u, these end points are values of random variables L and U, respectively. Suppose that

$$P\big(L \le \mu \le U\big) = 1 - \alpha.$$

  The resulting **confidence interval** for μ is

$$l \le \mu \le u.$$

  The end-points or bounds l and u are called the **lower-** and **upper-confidence limits** (bounds), respectively, and $1-\alpha$ is called the **confidence level**.

- If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1-\alpha)\%$ **confidence interval** on μ is given by

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

- Note that it is roughly of the form, $\bar{x} - 2\ \mathrm{SE}(\bar{x}) \le \ \mu \ \le \bar{x} + 2\ \mathrm{SE}(\bar{x})$.

- Confidence interval formulas give insight into the **required sample size**: If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error $|\bar{x} - \mu|$ will not exceed a specified amount Δ when the sample size is not smaller than

$$n = \left(\frac{z_{1-\alpha/2}\ \sigma}{\Delta}\right)^2.$$

- A **statistical hypothesis** is a statement about the parameters of one or more populations. The **null hypothesis**, denoted H_0 is the claim that is initially assumed to be true based on previous knowledge. The **alternative hypothesis**, denoted H_1 is a claim that contradicts the null hypothesis.

- For some arbitrary value μ_0, a **two-sided alternative hypothesis** would be expressed as follows:
$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0,$$
whereas a **one-sided alternative hypothesis** would be expressed as:
$$H_0 : \mu = \mu_0 \qquad H_1 : \mu < \mu_0 \qquad \text{or} \qquad H_0 : \mu = \mu_0 \qquad H_1 : \mu > \mu_0.$$
- The standard scientific research use of hypothesis is to "hope to reject" H_0 so as to have statistical evidence for the validity of H_1.
- An hypothesis test is based on a **decision rule** that is a function of the **test statistic**. For example: Reject H_0 if the test statistic is below a specified threshold, otherwise don't reject.
- Rejecting the null hypothesis H_0 when it is true is defined as a **type I error**. Failing to reject the null hypothesis H_0 when it is false is defined as a **type II error**.

	H_0 **Is True**	H_0 **Is False**
Fail to reject H_0:	No error	Type II error
Reject H_0:	Type I error	No error

$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true}).$

$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \mid H_0 \text{ is false}).$

- The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

- A typical example of a **simple hypothesis test** has $H_0 : \mu = \mu_0$ vs. $H_1 : \mu = \mu_1$, where μ_0 and μ_1 are some specified values for the population mean. This test isn't typically practical but is useful for understanding the concepts at hand.
- Assuming that $\mu_0 < \mu_1$ and setting a threshold, τ, reject H_0 if the $\overline{x} > \tau$, otherwise don't reject.
- Explicit calculation of the relationships of τ, α, β, n, σ, μ_0 and μ_1 is possible in this case.

- In most hypothesis tests used in practice (and in this course), a specified level of type I error, α is predetermined (e.g. $\alpha = 0.05$) and the type II error is not directly specified.
- The probability of making a type II error β increases (power decreases) rapidly as the true value of μ approaches the hypothesized value.
- The probability of making a type II error also depends on the sample size n - increasing the sample size results in a decrease in the probability of a type II error.
- The population (or natural) variability (e.g. described by σ) also affects the power.

- The P**-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data. That is, the P-value is based on the data. It is computed by considering the location of the test statistic under the sampling distribution based on H_0. It can also be viewed as the probability of observing a set of data which is as consistent or more consistent with the alternative hypothesis than the observed data, when the null hypothesis is true.
- It is customary to consider the test statistic (and the data) significant when the null hypothesis H_0 is rejected; therefore, we may think of the P-value as the smallest α at which the data are significant. In other words, the P-value is the **observed significance level**.

- Clearly, the P-value provides a measure of the credibility of the null hypothesis. Computing the exact P-value for a statistical test is not always doable by hand.
- It is typical to report the P-value in studies where H_0 was rejected (and new scientific claims were made). Typical ("convincing") values can be of the order 0.001.

- **A General Procedure for Hypothesis Tests is**

  (1) **Parameter of interest:** From the problem context, identify the parameter of interest.

  (2) **Null hypothesis, H_0:** State the null hypothesis, H_0.

  (3) **Alternative hypothesis, H_1:** Specify an appropriate alternative hypothesis, H_1.

  (4) **Test statistic:** Determine an appropriate test statistic.

  (5) **Reject H_0 if:** State the rejection criteria for the null hypothesis.

  (6) **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute the value.

  (7) **Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

Single Sample Inference

- The setup is a sample $x_1, \ldots, x_n$ (collected values) modelled by an i.i.d. sequence of random variables, $X_1, \ldots, X_n$.
- The parameter at question in this unit is the population mean, $\mu = E[X_i]$. A point estimate is $\overline{x}$ (described by the random variable $\overline{X}$).
- We devise hypothesis tests and confidence intervals for μ, distinguishing between the (unrealistic but simpler) case where the population variance, σ^2, is known, and the more realistic case where it is not known and estimated by the sample variance, s^2.
- For very small samples, the results we present are valid only if the population is normally distributed. But for non-small samples (e.g. $n > 20$, although there isn't a clear rule), the central limit theorem provides a good approximation and the results are approximately correct.

- **Testing Hypotheses on the Mean, Variance Known (Z-Tests)**

  Model: $X_i \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with μ unknown but σ^2 known.

  Null hypothesis: $H_0 : \mu = \mu_0$.

  Test statistic: $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}, \qquad Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$

Alternative Hypotheses	P-value	Rejection Criterion for Fixed-Level Tests
$H_1 : \mu \neq \mu_0$	$P = 2\big[1 - \Phi\big(\lvert z\rvert\big)\big]$	$z > z_{1-\alpha/2}$ or $z < z_{\alpha/2}$
$H_1 : \mu > \mu_0$	$P = 1 - \Phi\big(z\big)$	$z > z_{1-\alpha}$
$H_1 : \mu < \mu_0$	$P = \Phi\big(z\big)$	$z < z_{\alpha}$

- Note: For $H_1 : \mu \neq \mu_0$, a procedure identical to the preceding fixed significance level test is:

$$\text{Reject } H_0 : \mu = \mu_0 \quad \text{if either} \quad \bar{x} < a \text{ or } \bar{x} > b$$

where

$$a = \mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad b = \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Compare these results with the confidence interval formula (presented in previous unit):

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

- In this case, if H_0 is not true and H_1 holds with a specific value of $\mu = \mu_1$, then it is possible to compute the probability of type II error, β.

- In the (very realistic) case where σ^2 is not known, but rather estimated by S^2, we would like to replace the test statistic, Z, above with,

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}},$$

but in general, T no longer follows a Normal distribution.

- Under $H_0 : \mu = \mu_0$, and for moderate or large samples (e.g. $n > 100$) this statistic is approximately Normally distributed just like above. In this case, the procedures above work well.

- But for smaller samples, the distribution of T is no longer Normally distributed. Nevertheless, it follows a well known and very famous distribution of classical statistics: **The Student-*t* Distribution**.

- The probability density function of a Student-t Distribution with a parameter v, referred to as **degrees of freedom**, is,

$$f(x\,;\,v) = \frac{\Gamma\big[(v+1)/2\big]}{\sqrt{\pi v}\Gamma(v/2)} \cdot \frac{1}{\Big[\big(x^2/v\big)+1\Big]^{(v+1)/2}} \qquad -\infty < x < \infty,$$

where $\Gamma(\cdot)$ is the Gamma-function. It is a symmetric distribution about 0 and as $v \to \infty$ it approaches a standard Normal distribution.

- The following mathematical result makes the t-distribution useful: Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from a Normal distribution with mean μ and variance σ^2. The random variable, T has a t-distribution with $n-1$ degrees of freedom.

- Now, knowing the distribution of T (and noticing it depends on the sample size, n), allows us to construct hypothesis tests and confidence intervals when σ^2 is not known, analogous to the (Z-tests and confidence intervals) presented above.

- If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, a $100(1-\alpha)\%$ **confidence interval** on μ is given by

$$\bar{x} - t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\alpha/2,n-1}\frac{s}{\sqrt{n}},$$

where $t_{1-\alpha/2,n-1}$ is the $1-\alpha/2$ quantile of the t distribution with $n-1$ degrees of freedom.

- A related concept is a $100(1-\alpha)\%$ **prediction interval** (PI) on a single future observation from a normal distribution is given by

$$\bar{x} - t_{1-\alpha/2,n-1} s\sqrt{1+\frac{1}{n}} \le X_{n+1} \le \bar{x} + t_{1-\alpha/2,n-1} s\sqrt{1+\frac{1}{n}}.$$

This is the range where we expect the $n+1$ observation to be, after observing n observations and computing $\bar{x}$ and s.

- **Testing Hypotheses on the Mean, Variance Unknown (T-Tests)**

Model: $X_i \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with both μ and σ^2 unknown.

Null hypothesis: $H_0 : \mu = \mu_0$.

Test statistic: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}, \qquad T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}.$

Alternative Hypotheses	**P-value**	**Rejection Criterion for Fixed-Level Tests**
$H_1 : \mu \neq \mu_0$	$P = 2\big[1 - F_{n-1}\big(\lvert t\rvert\big)\big]$	$t > t_{1-\alpha/2,n-1}$ or $t < t_{\alpha/2,n-1}$
$H_1 : \mu > \mu_0$	$P = 1 - F_{n-1}\big(t\big)$	$t > t_{1-\alpha,n-1}$
$H_1 : \mu < \mu_0$	$P = F_{n-1}\big(t\big)$	$t < t_{\alpha,n-1}$

Note that here, $F_{n-1}(\cdot)$ denotes the cdf of the t-distribution with $n-1$ degrees of freedom. As opposed to $\Phi(\cdot)$, it is not tabulated in standard tables and like $\Phi(\cdot)$ it cannot be explicitly evaluated. So to calculate P-values, we use software.

Linear Regression

- The collection of statistical tools that are used to model and explore relationships between variables that are related in a nondeterministic manner is called **regression analysis**. Of key importance is the conditional expectation,

$$E(Y \mid x) = \mu_{Y\mid x} = \beta_0 + \beta_1 x \qquad \text{with} \qquad Y = \beta_0 + \beta_1 x + \epsilon,$$

where x is not random and ϵ is a Normal random variable with $E(\epsilon) = 0$ and $V(\epsilon) = \sigma^2$.

- **Simple Linear Regression** is the case where both x and y are scalars, in which case the data is,

$$(x_1, y_1), \ldots, (x_n, y_n).$$

Then given estimates of β_0 and β_1 denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$ we have

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \qquad i = 1, 2, \ldots, n,$$

where e_i, are the **residuals** and we can also define the **predicted observation**,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

Ideally it would hold that $y_i = \hat{y}_i$ ($e_i = 0$) and thus **total mean squared error**

$$L := SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2,$$

would be zero. But in practice, unless $\sigma^2 = 0$ (and all points lie on the same line), we have that $L > 0$.

- The standard (classic) way of determining the statistics $(\hat{\beta}_0, \hat{\beta}_1)$ is by minimisation of L. The solution, called the **least squares estimators** must satisfy

$$\frac{\partial L}{\partial \beta_0}\Big|_{\hat{\beta}_0\hat{\beta}_1} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial L}{\partial \beta_1}\Big|_{\hat{\beta}_0\hat{\beta}_1} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0$$

Simplifying these two equations yields

$$n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0\sum_{i=1}^{n} x_i + \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i$$

These are called the **least squares normal equations**. The solution to the normal equations results in the **least squares estimators** $\hat{\beta}_0$ and $\hat{\beta}_1$. Using the sample means, $\overline{x}$ and $\overline{y}$ the estimators are,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}, \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}.$$

- The following quantities are also of common use:

$$S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x}) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x^i\right)\left(\sum_{i=1}^{n} y^i\right)}{n}$$

Hence,

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

Further,

$$SS_T = \sum_{i=1}^{n}(y_i - \overline{y})^2, \qquad SS_R = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2, \qquad SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

- The **Analysis of Variance Identity** is

$$\sum_{i=1}^{n}\Big(y_i-\bar{y}\Big)^2=\sum_{i=1}^{n}\Big(\hat{y}_i-\bar{y}\Big)^2+\sum_{i=1}^{n}\Big(y_i-\hat{y}_i\Big)^2$$

or,

$$SS_T = SS_R + SS_E.$$

Also, $SS_R = \hat{\beta}_1 S_{xy} \quad \longrightarrow \quad SS_E = SS_T - \hat{\beta}_1 S_{xy}.$

- An **Estimator of the Variance**, σ^2 is

$$\hat{\sigma}^2 := MS_E = \frac{SS_E}{n-2}$$

- A widely used measure for a regression model is the following ratio of sum of squares, which is often used to judge the adequacy of a regression model:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}.$$

$$E\Big(\hat{\beta}_0\Big) = \beta_0, \qquad V\Big(\hat{\beta}_0\Big) = \sigma^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{XX}}\right]$$

$$E\Big(\hat{\beta}_1\Big) = \beta_1, \qquad V\Big(\hat{\beta}_1\Big) = \frac{\sigma^2}{S_{XX}}.$$

- In simple linear regression, the **estimated standard error of the slope** and the **estimated standard error of the intercept** are

$$se\Big(\hat{\beta}_1\Big) = \sqrt{\frac{\hat{\sigma}^2}{S_{XX}}} \qquad \text{and} \qquad se\Big(\hat{\beta}_0\Big) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{XX}}\right]}$$

- The **Test Statistic for the Slope** is

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2/S_{XX}}}$$

$$H_0 : \beta_1 = \beta_{1,0} \qquad\qquad H_1 : \beta_1 \neq \beta_{1,0}$$

Under H_0 the test statistic T follows a **t - distribution** with "$n-2$ degree of freedom".

- An alternative is to use the F statistic as is common in **ANOVA** (Analysis of Variance) – not covered fully in the course.

$$F = \frac{SS_R/1}{SS_E/(n-2)} = \frac{MS_R}{MS_E}.$$

Under H_0 the test statistic F follows an **F - distribution** with "1 degree of freedom in the numerator and $n-2$ degrees of freedom in the denominator".

Analysis of Variance Table for Testing Significance of Regression

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	MS_R/MS_E
Error	$SS_E = SS_T - \hat{\beta}_1 S_{xy}$	$n-2$	MS_E	
Total	SS_T	$n-1$		

- There are also confidence intervals for β_0 and β_1 as well as prediction intervals for observations. We don't cover these formulas.

- To check the regression model assumptions we plot the residuals e_i and check for (i) Normality. (ii) Constant variance. (iii) Independence.

Logistic Regression:

- Take the response variable, Y_i as a Bernoulli random variable. In this case notice that $E(Y) = P(Y = 1)$.
- The **logit response function** has the form

$$E\Big(Y\Big) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp\Big(\beta_0 + \beta_1 x\Big)}.$$

- Fitting a logistic regression model to data yields estimates of β_0 and β_1.
- The following formula is called the **odds**

$$\frac{E\Big(Y\Big)}{1 - E\Big(Y\Big)} = \exp\Big(\beta_0 + \beta_1 x\Big).$$

Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

This table was generated using the "CDF" command in Minitab.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9983	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

t-Distribution Quantiles

ν	$Q(.9)$	$Q(.95)$	$Q(.975)$	$Q(.99)$	$Q(.995)$	$Q(.999)$	$Q(.9995)$
1	3.078	6.314	12.706	31.821	63.657	318.317	636.607
2	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.849
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

This table was generated using the "INVCDF" command in Minitab.