MAT 2377
Probability and Statistics for Engineers

Final Review

Iraj Yadegari (uOttawa)

Fall 2021

Note that:

- There is no make-up exam for MAT2377.
- The final exam will cover all topics of the course.
- The focus of the final exam is not on particular chapters.
- This is a short review of the course. If a topic is missing in this review, it does not mean that it is not covered in the exam.

Summary Chapter 1

- Probability: $0 \leq P(A) \leq 1$; $P(S) = 1$; $P(\varnothing) = 0$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- $P(A) = P(A \cap B) + P(A \cap B^c)$
- Mutually exclusive: $A \cap B = \varnothing$; $P(A \cap B) = 0$
- INDEPENDENCE: $P\left(A \cap B\right) = P(A) \times P(B)$
- Conditional Probability: $P(B|A) = \frac{P(A \cap B)}{P(A)}$; $P(A) \neq 0$.
- Bayes: $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$ such that $P(B) = P(B|A_1)P(A_1) + ... + P(B|A_k)P(A_k)$.
- $P(A \cap B) = P(A) \times P(B|A) = P(B) \times P(A|B) = P(B \cap A)$;
- $P(A|B) \neq P(B|A)$

Summary of Chapter 2

If X is a discrete random variable with p.m.f. $f(x)$ and c.d.f. $F(x)$, then

- $0 < f(x) \leq 1$ for all $x \in X(\mathcal{S})$;
- $\Sigma_{s\in\mathcal{S}} f(X(s)) = \Sigma_{x\in X(\mathcal{S})} f(x) = 1$;
- for any event $A \subseteq \mathcal{S}$, $P(X \in A) = \Sigma_{x\in A} f(x)$;
- for any $a, b \in \mathbb{R}$,

$$P(a < X) = 1 - P(X \leq a) = 1 - F(a)$$
$$P(X < b) = P(X \leq b) - P(X = b) = F(b) - f(b)$$

- for any $a, b \in \mathbb{R}$,

$$P(a \leq X) = 1 - P(X < a) = 1 - \Big(P(X \leq a) - P(X = a)\Big)$$
$$= 1 - F(a) + f(a)$$

We can use these results to compute the probability of a **discrete** r.v. X falling in various intervals:

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$
$$P(a \leq X \leq b) = P(a < X \leq b) + P(X = a) = F(b) - F(a) + f(a)$$
$$P(a < X < b) = P(a < X \leq b) - P(X = b) = F(b) - F(a) - f(b)$$
$$P(a \leq X < b) = P(a \leq X \leq b) - P(X = b) = F(b) - F(a) + f(a) - f(b)$$

Expectation and variance of a Discrete R.V.

The **expectation** of a discrete random variable X is defined as

$$\mathrm{E}[X] = \sum_x x \cdot P(X = x) = \sum_x x f(x)\,,$$

where the sum extends over all values of x taken by X.

$$\mathrm{E}[X^2] = \sum_x x^2 P(X = x) = \sum_x x^2 f(x)$$

.

The variance of a discrete random variable X is the **expected squared difference from the mean**:

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mu_X)^2] = \sum_x (x - \mu_X)^2 P(X = x).$$

This is also sometimes written as

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}^2[X]$$

General Properties of Expectation and Variance

For all $a \in \mathbb{R}$:

- $\mathrm{E}[aX] = a\mathrm{E}[X]$;
- $\mathrm{E}[X + a] = \mathrm{E}[X] + a$;
- $\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$; therefore $SD(aX) = |a|SD(X)$.
- $\mathrm{Var}[X + a] = \mathrm{Var}[X]$.
- For any random variables X and Y: $\quad \mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

For INDEPENDENT variables X and Y, we have

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

X	**Description**	$P(X = x)$	**Domain**	$\mathrm{E}[X]$	$\mathrm{Var}[X]$
Uniform (Discrete)	Equally likely outcomes	$\frac{1}{b-a+1}$	$a, \ldots, b$	$\frac{a+b}{2}$	$\frac{(b-a+2)(b-a)}{12}$
Binomial	Number of successes in n independent trials	$\binom{n}{x} p^x (1-p)^{n-x}$	$0, \ldots, n$	np	$np(1-p)$
Poisson	Number of arrivals in a fixed period of time	$\frac{\lambda^x \exp(-\lambda)}{x!}$	$0, 1, \ldots$	λ	λ
Geometric	Number of trials until 1^{st} success	$(1-p)^{x-1} p$	$1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	Number trials until k^{th} successes	$\binom{x-1}{k-1}(1-p)^{x-k} p^k$	$k, k+1, \ldots$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$

Summary of Chapter 3: Continuous Random Variables

- Discrete data are data with a **finite** or **countably infinite** number of possible outcomes.
- Continuous data are data which come from a continuous interval of possible outcomes. It means that continuous data are with **uncountably infinitely many outcomes**.
- In the discrete case, the probability mass function $f_X(x) = P(X = x)$ was the main object of interest. In the continuous case, the analogous role is played by the **probability density function** (**p.d.f**.), still denoted by $f_X(x)$, but $f_X(x) \neq P(X = x)$.

The **(cumulative) distribution function** (c.d.f.) of any such random variable X is still defined by

$$F_X(x) = P(X \leq x)\,,$$

viewed as a function of a real variable x; but $P(X \leq x)$ is not simply computed by adding a few terms of the form $P(X = x_i)$. Note that

$$\lim_{x\to-\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x\to+\infty} F_X(x) = 1.$$

We can describe the **distribution** of the random variable X *via* the following relationship between $f_X(x)$ and $F_X(x)$:

$$f_X(x) = \frac{d}{dx}F_X(x).$$

Probability Density Functions (p.d.f.)

The **probability density function** (p.d.f.) of a continuous random variable X is an **integrable** function $f_X : X(\mathcal{S}) \to \mathbb{R}$ such that

- $f_X(x) > 0$ for all $x \in X(\mathcal{S})$ and $\lim_{x\to\pm\infty} f_X(x) = 0$;
- $\int_{\mathcal{S}} f_X(x)\, dx = 1$;

- for any event $A = (a,b) = \{X | a < X < b\}$,

$$P(A) = P((a,b)) = \int_a^b f_X(x)\,dx;$$

- for any x,

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt;$$

- for any x,

$$P(X > x) = 1 - P(X \leq x) = 1 - F_X(x) = \int_x^{\infty} f_X(t)\,dt;$$

- for any $a, b \in \mathbb{R}$,

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$$
$$= F_X(b) - F_X(a) = \int_a^b f(x)\,dx.$$

- for any $a \in \mathbb{R}$,

$$P\left(X = a\right) = \lim_{\Delta \to 0} P\left(a \leq X \leq a + \Delta\right) = \lim_{\Delta \to 0} \int_a^{a+\Delta} f_X(x)\,dx = 0.$$

Expectation and variance of Continuous RVs

For a continuous random variable X with p.d.f. $f_X(x)$, the **expectation** of X is defined as

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

In a similar way to the discrete case, the **mean** of X is defined to be $\mathrm{E}[X]$, and the **variance** and **standard deviation** of X are, as before,

$$\mathrm{Var}[X] \overset{\text{def}}{=} \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \overset{\text{comp. formula}}{=} \mathrm{E}[X^2] - \Big(\mathrm{E}[X]\Big)^2,$$

$$\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]}\,.$$

Standard Normal Distribution

An **very** important example of continuous distributions is that of the special probability distribution function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,.$$

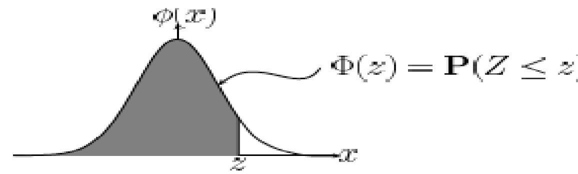
The corresponding cumulative distribution function is denoted by

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(t)\,dt\,.$$

A random variable Z with this c.d.f. is said to have a **standard normal distribution**, and we write $Z \sim \mathcal{N}(0,1)$.

Standard Normal Table

Table 1. Normal Distribution Function
Lower tail of the standard normal distribution is tabulated



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.0
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7
0.60	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7
0.70	0.7580	0.7612	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8
1.20	0.8849	0.8869	0.8888	0.8906	0.8925	0.8943	0.8962	0.8980	0.8997	0.9

Normal Approximation with Continuity Correction

Let $X \sim \mathcal{B}(n, p)$. Recall that $\mathrm{E}[X] = np$ and $\mathrm{Var}[X] = np(1-p)$.

If n is large, we may approximate X by a normal random variable in the following way:

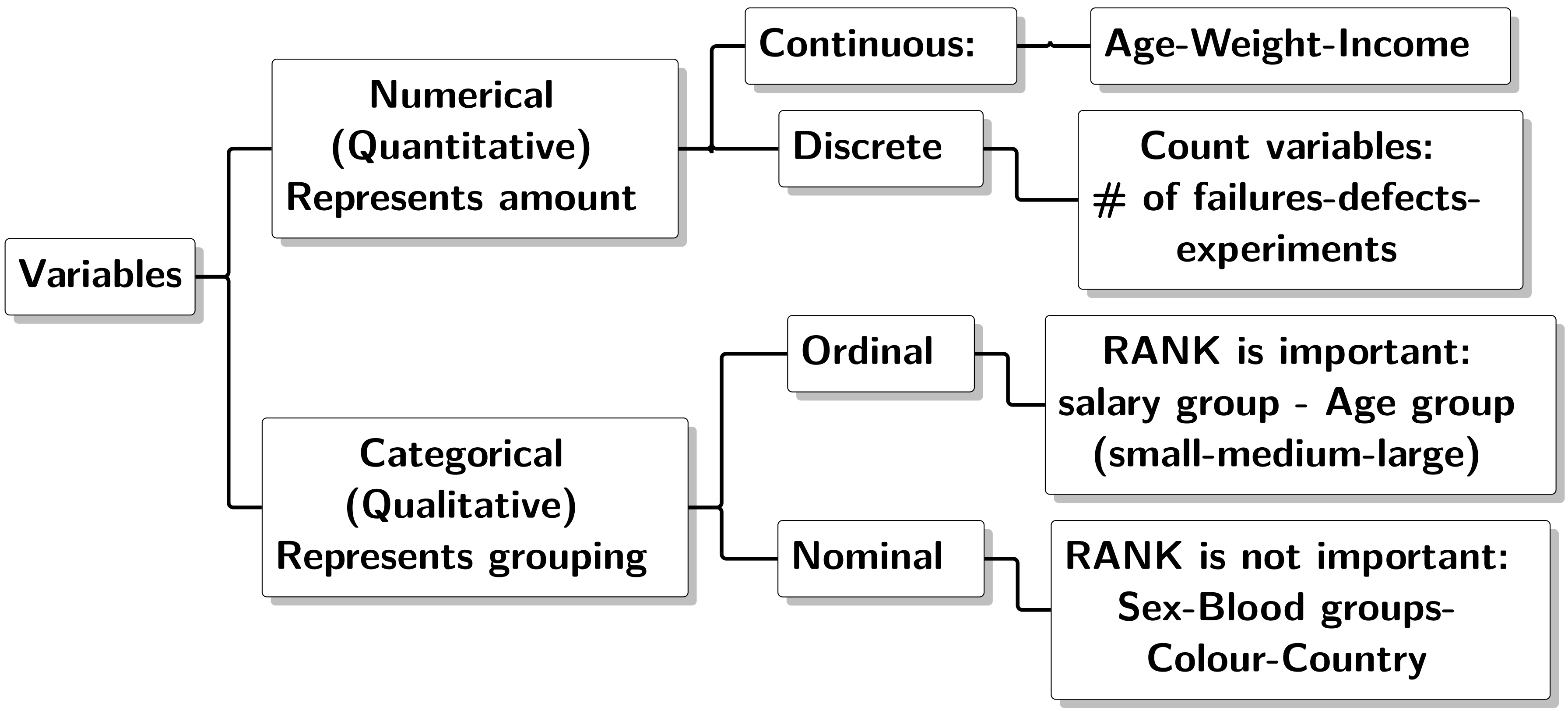
$$P(X \leq x) = P(X < x + 0.5) = P\left(Z < \frac{x - np + 0.5}{\sqrt{np(1-p)}}\right)$$

and

$$P(X \geq x) = P(X > x - 0.5) = P\left(Z > \frac{x - np - 0.5}{\sqrt{np(1-p)}}\right).$$

X	**Example**	$f(x)$	**Domain**	$\mathrm{E}[X]$	$\mathrm{Var}[X]$
Uniform	Select a point at random from $[a,b]$	$\frac{1}{b-a}$	$a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	Meas. errors; children heights; breaking strengths, etc.	$\frac{\exp(-(x-\mu)^2/2\sigma^2)}{\sigma\sqrt{2\pi}}$	$-\infty < x < \infty$	μ	σ^2
Exponential	Waiting time to first arrival in a Poisson process with rate λ	$\lambda e^{-\lambda x}$	$0 \leq x < \infty$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	Waiting time to rth arrival in a Poisson process with rate λ	$\frac{x^{r-1}}{(r-1)!}\lambda^r e^{-\lambda x}$	$0 \leq x < \infty$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$

Summary of Chapter 4-1: descriptive statistics



Statistical Summaries

A variable can be described with two type of measures: **centrality**, **spread**.

- **Centrality** measures: **median**, **mean**, (mode, less frequent).

- **Spread** (variation or dispersion) measures: **variance, standard deviation** (sd), **inter-quartile range** (IQR), range (less frequent), (**skew** and **kurtosis** are also used sometimes).

The median, range and the quartiles are easily calculated from an **ordered** list of the data.

(Sample) Median

The **median** med$(x_1, \ldots, x_n)$ of a sample of size n is a numerical value which splits the ordered data into 2 equal subsets: half the observations are below the median, **and** half above it.

- If n is **odd**, then the **position** of the median is $(n+1)/2$, that is to say, the median observation is the $\frac{n+1}{2}^{\text{th}}$ ordered observation.

- If n is **even**, then the median is the average of the $\frac{n}{2}^{\text{th}}$ and the $(\frac{n}{2}+1)^{\text{th}}$ ordered observations.

The procedure is simple: Order the data, and follow the even/odd rules.

(Sample) Mean

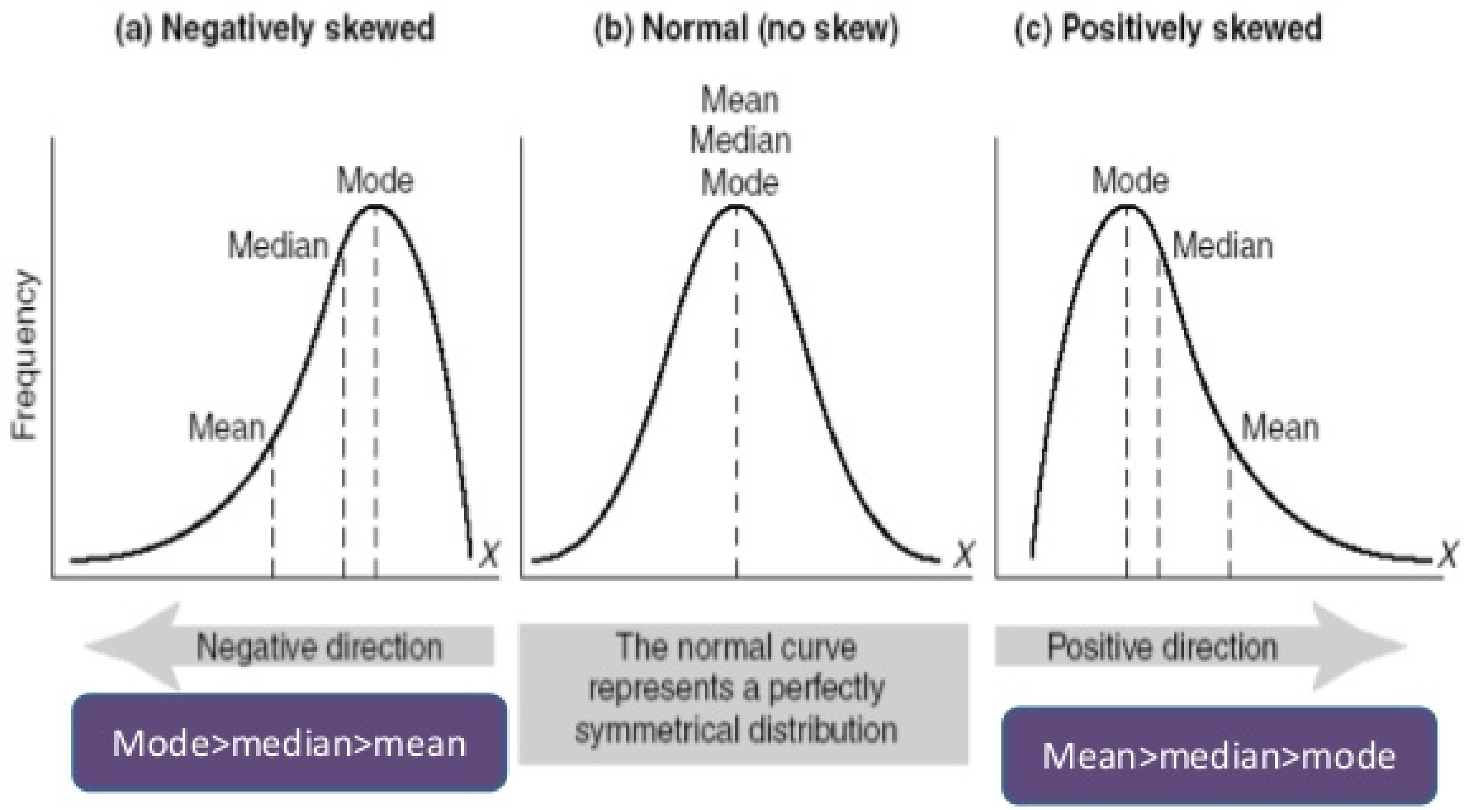
The **mean** of a sample is simply the arithmetic average of its observations. For observations $x_1, x_2, \ldots, x_n$, the sample mean is

$$\text{AM}(x_1, \ldots, x_n) = \overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$$

Other means exist, such as the **harmonic** mean and the **geometric** mean:

$$\text{HM}(x_1, \ldots, x_n) = \frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}} \quad \text{and} \quad \text{GM}(x_1, \ldots, x_n) = \sqrt[n]{x_1 \cdots x_n}.$$

The median is **robust** against extreme values, but mean is affected by extremes.



Measures of Spread

A) sample standard deviation

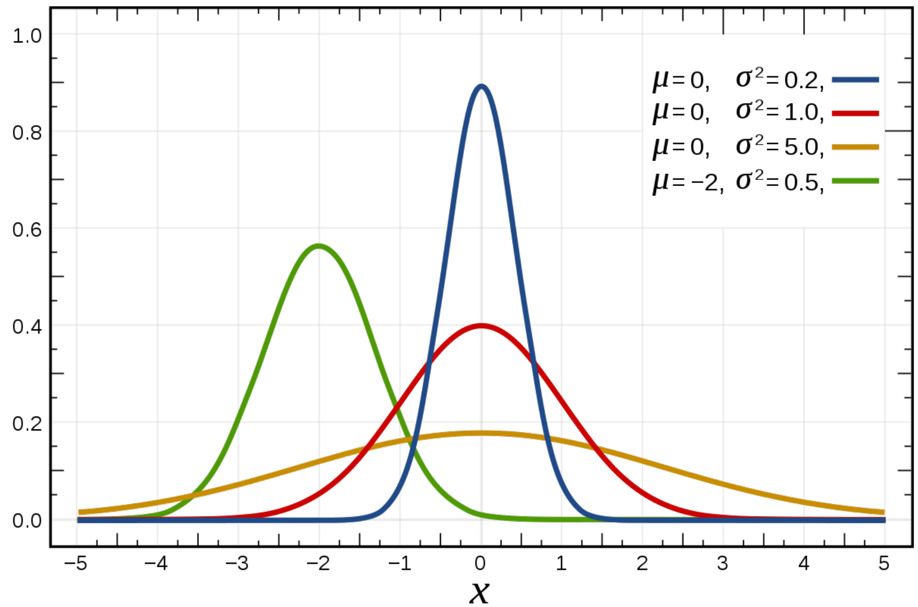
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right).$$

B) The **sample range** is

$$\text{range}(x_1, \dots, x_n) = \max\{x_i\} - \min\{x_i\} = y_n - y_1,$$

where $y_1 \leq \ldots \leq y_n$ is the ranked data.

C) The **inter-quartile range** is IQR $= Q_3 - Q_1$.


1.0
0.8
0.6
0.4
0.2
0.0
−5
−4
−3
−2
−1
0
1
2
3
4
5
x
μ=0, σ²=0.2,
μ=0, σ²=1.0,
μ=0, σ²=5.0,
μ=−2, σ²=0.5,

Quartiles

Another way to provide information about the spread of the data is with the help of **quartiles**.

The **lower quartile** $Q_1(x_1, \ldots, x_n)$ of a sample of size n, or Q_1, is a numerical value which splits the ordered data into 2 unequal subsets: 25% of the observations are below Q_1, **and** 75% of the observations are above Q_1.

Similarly, the **upper quartile** Q_3 splits the ordered data into 75% of the observations below Q_3, **and** 25% of the observations above Q_3.

The median can be interpreted as the **middle quartile**, Q_2: 50% of the observations are below Q_2, **and** 50% of the observations are above Q_2.

How to calculate?

Sort the sample observations $\{x_1, x_2, \ldots, x_n\}$ in an **increasing order** as

$$y_1 \leq y_2 \leq \ldots \leq y_n.$$

The smallest y_1 has **rank** 1 and the largest y_n has **rank** n.

- The lower quartile Q_1 is computed as the average of ordered observations with ranks $\lfloor \frac{n}{4} \rfloor$ and $\lfloor \frac{n}{4} \rfloor + 1$.
- Similarly, Q_3 is computed as the average of ordered observations with ranks $\lceil \frac{3n}{4} \rceil$ and $\lceil \frac{3n}{4} \rceil + 1$.
- The median can be interpreted as the **middle quartile**, Q_2.
- Operators $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are defined such that $\lfloor 2.31 \rfloor = 2$ and $\lceil 2.31 \rceil = 3$

Outliers

An outlier is an observation that lies outside the overall pattern in a distribution.

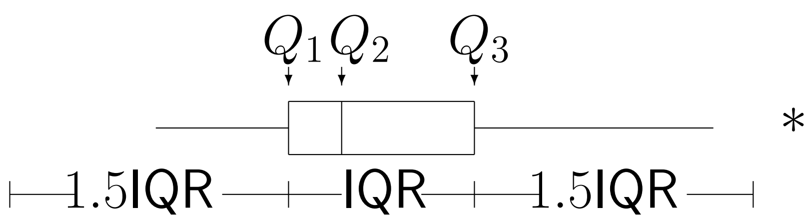
Let x be an observation in the sample. It is a **suspected outlier** if

$$x < Q_1 - 1.5\,\mathsf{IQR} \quad \text{or} \quad x > Q_3 + 1.5\,\mathsf{IQR},$$

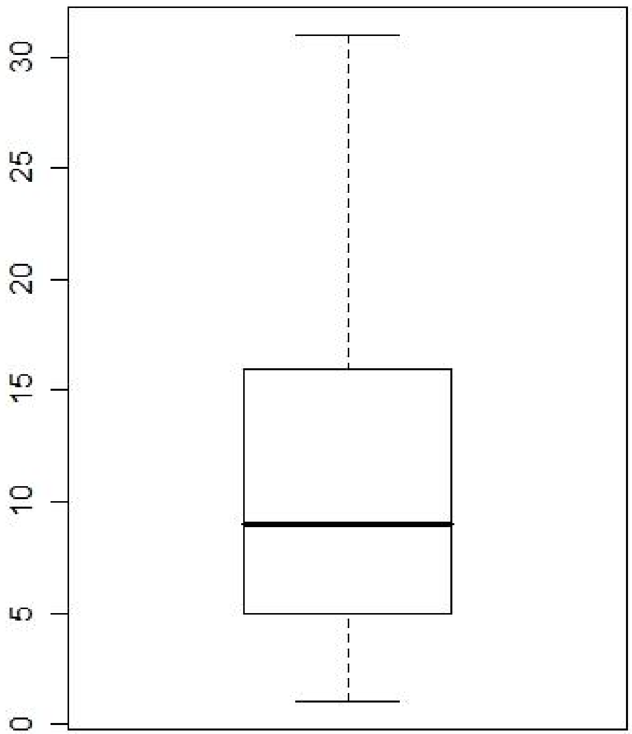
where IQR $= Q_3 - Q_1$ it the **inter-quartile range** $Q_3 - Q_1$.

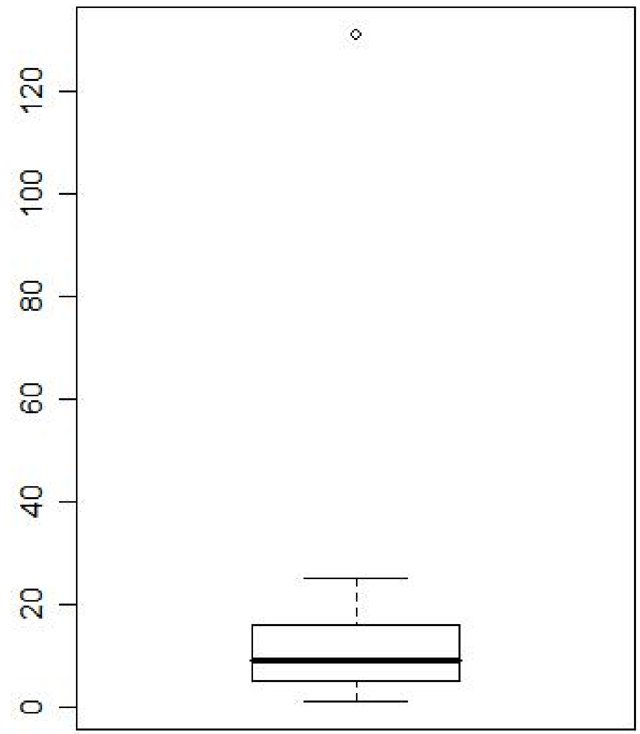
Box plot

The **boxplot** is a quick and easy way to present a graphical summary of a univariate distribution.



- The main part is a box, with endpoints at the lower and upper quartiles, and with a "belt" at the median.
- A line is extending from Q_1 to the smallest value less than $1.5\,$IQR to the left of Q_1.
- A line is extending from Q_3 to the largest value less than $1.5\,$IQR to the right of Q_3.
- Suspected outliers are represented by *.


0
5
10
15
20
25
30

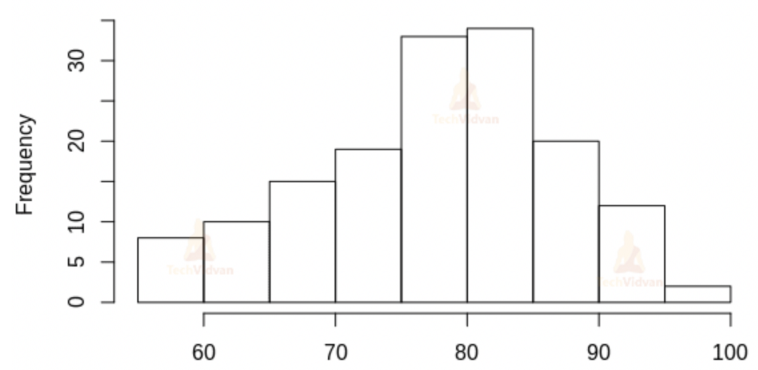

0
20
40
60
80
100
120

Histogram

Histograms also provide an indication of the distribution of the sample.

Histograms should contain the following information:

- the range of the histogram is $r = \max\{x_i\} - \min\{x_i\}$;
- the number of bins should approach $k = \sqrt{n}$, where n is the sample size;
- the bin width should approach r/k,
- and the frequency of observations in each bin should be added to the chart.

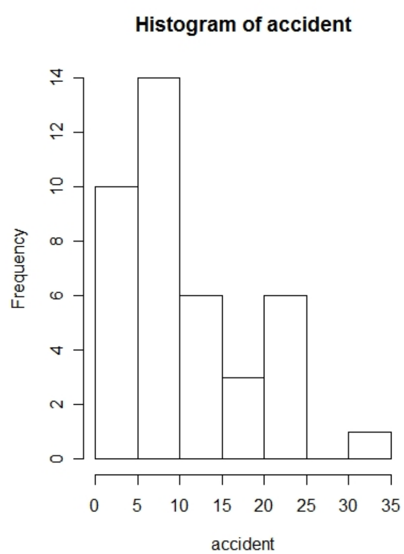


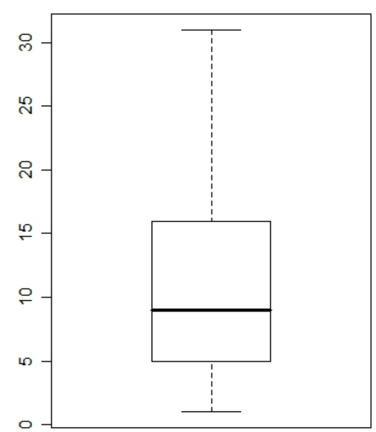
Skewness

Boxplots give an easy graphical means of getting an impression of the shape of the data set. The shape is used to suggest a mathematical model for the situation of interest.

The data set is **right skewed** if the boxplot is stretched to the right.

Similar observations can be inferred from the histogram.





If the data distribution is symmetric then the (population) median and mean are equal and the first and third (population) quartiles are equidistant from the median.

If data is stretched to the right or left, then distribution of data is Asymmetric (skewed).

If $Q_3 - Q_2 > Q_2 - Q_1$ then the data distribution is **skewed to the right**.

If $Q_3 - Q_2 < Q_2 - Q_1$ then the data distribution is **skewed to left**.

Summary of Chapter 4-2: Sampling distributions

The **sample mean** is a typical statistic of interest:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i\,.$$

If $X_1, \ldots, X_n$ are iid with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2$ for all $i = 1, \ldots, n$, then

$$\mathrm{E}\left[\overline{X}\right] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}(n\mu) = \mu$$

$$\mathrm{Var}\left[\overline{X}\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \left[\frac{1}{n}\right]^2 \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\left(n\sigma^2\right) = \frac{\sigma^2}{n}\,.$$

Sum of Independent Normal RVs

If $\{X_1, \ldots, X_n\}$ is a random sample from a population **with mean** μ **and variance** σ^2, then

- $\mathrm{E}\left[\Sigma_{i=1}^n X_i\right] = n\mu$ and $\mathrm{Var}\left[\Sigma_{i=1}^n X_i\right] = n\sigma^2$;
- $\mathrm{E}\left[\overline{X}\right] = \mu$ and $\mathrm{Var}\left[\overline{X}\right] = \sigma^2/n$;
- furthermore, if the population distribution is **normal**, then $\Sigma_{i=1}^n X_i$ and $\overline{X}$ are also normal, i.e.

$$\sum_{i=1}^n X_i \sim \mathcal{N}\left(n\mu, n\sigma^2\right) \quad \text{and} \quad \overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Central Limit Theorem

Theorem: If $\overline{X}$ is the mean of a random sample of size n taken from an **unknown** population with mean μ and finite variance σ^2, then $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$, has the standard normal distribution $\mathcal{N}(0,1)$ as $n \to \infty$.

More precisely, the result is a **limiting** result. If we view the standardized

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

as functions of n, **regardless of whether the original X_i's are normal or not,** for each z we have

$$\lim_{n\to\infty} P\left(Z_n \leq z\right) = \Phi(z) \quad \text{and} \quad P\left(Z_n \leq z\right) \approx \Phi(z) \text{ if } n \text{ is large enough.}$$

Sampling Distribution – Difference Between 2 Means

Theorem: Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ_1 and variance σ_1^2, and $Y_1, \ldots, Y_m$ be another random sample, independent of X, from a population with mean μ_2 and variance σ_2^2. If $\overline{X}$ and $\overline{Y}$ are the respective sample means, then

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

has standard normal distribution as $n, m \to \infty$. This is also a **limiting** result.

Sample Mean with Unknown Population Variance

Theorem: let $X_1, \ldots, X_n$ be independent normal random variables with mean μ and standard deviation σ . Let $\overline{X}$ and S^2 be the sample mean and sample variance, respectively. Then the random variable

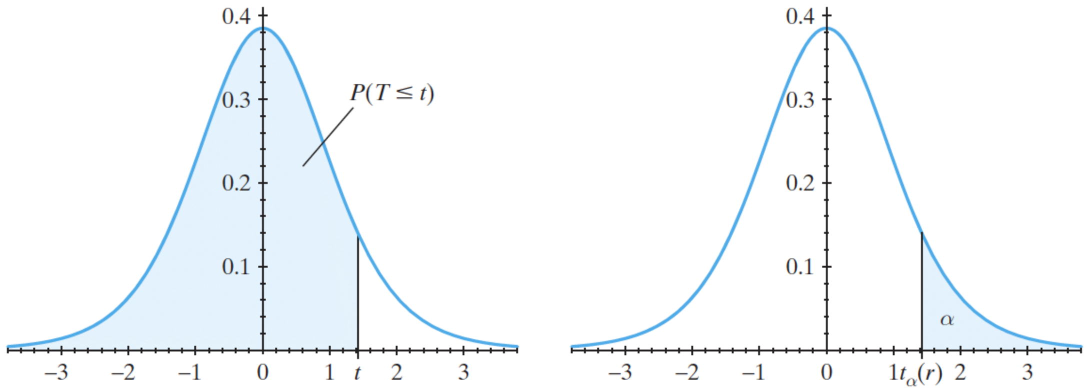
$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

follows a **Student t-distribution with $\nu = n - 1$ degrees of freedom**.

t-**Table:** let $t_\alpha(\nu)$ represent the critical t-value above which we find an area equal to α , i.e. $P(T > t_\alpha(\nu)) = \alpha$, where $T \sim t(\nu)$.

For all ν, the Student t-distribution is a symmetric distribution around zero, so we have $t_{1-\alpha(\nu)} = -t_\alpha$.

Table VI The t Distribution



$$P(T \le t) = \int_{-\infty}^{t} \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\,\Gamma(r/2)(1+w^2/r)^{(r+1)/2}}\,dw$$

$$P(T \le -t) = 1 - P(T \le t)$$

	$P(T \le t)$						
	0.60	0.75	0.90	0.95	0.975	0.99	0.995
r	$t_{0.40}(r)$	$t_{0.25}(r)$	$t_{0.10}(r)$	$t_{0.05}(r)$	$t_{0.025}(r)$	$t_{0.01}(r)$	$t_{0.005}(r)$
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169

Summary of Chapter 5: Confidence Intervals

Sample: $\{X_1, \ldots, X_n\}$. **Objective:** predict μ with confidence level α.

- If population is **normal** with **known** variance σ^2, the **exact** $100(1-\alpha)\%$ C.I. is
$$\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$
- If population is **non-normal** with **known** variance σ^2 and n is '**big**', the **approximate** $100(1-\alpha)\%$ C.I. is
$$\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

- If population is **normal** with **unknown** variance, the **exact** $100(1-\alpha)\%$ C.I. is
$$\overline{X} \pm t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}.$$

- If population has **unknown** variance and n is '**big**', the **approximate** $100(1-\alpha)\%$ C.I. is
$$\overline{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}.$$

- If population has **unknown** variance and n is '**small**', you are S.O.O.L.

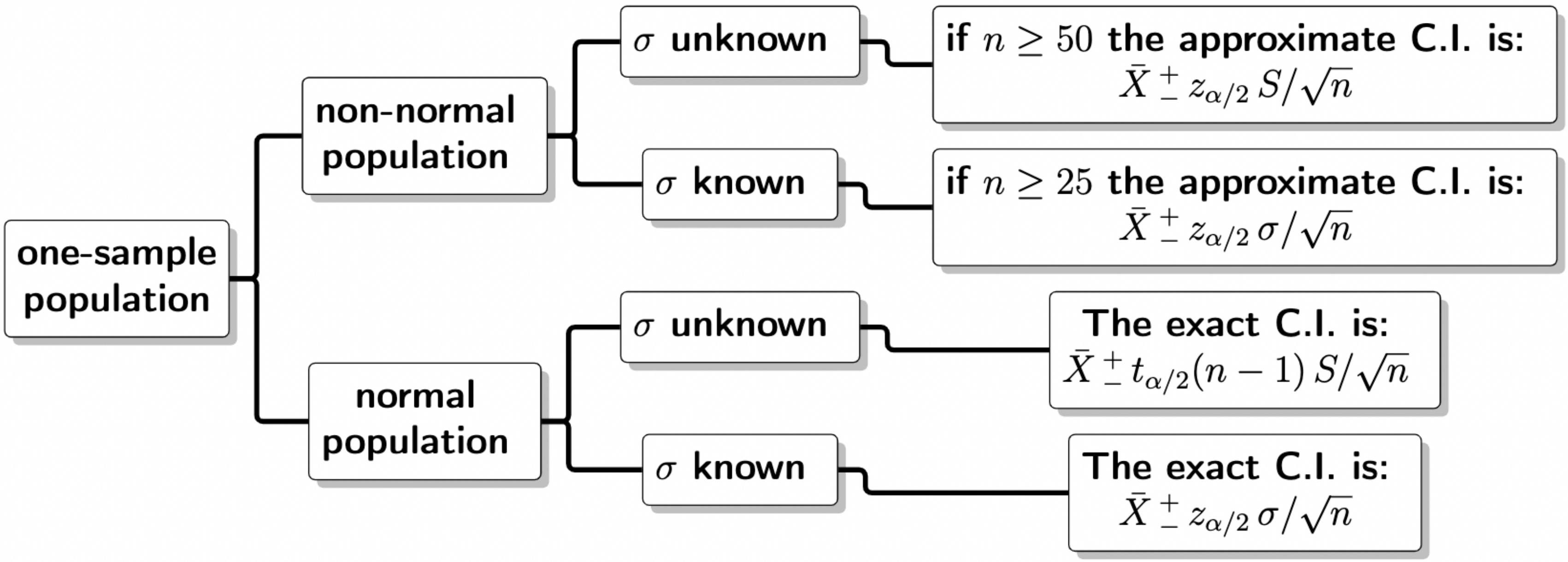


Figure 12: Confidence interval for the mean of a population

C.I. for a Proportion

If $X \sim \mathcal{B}(n,p)$ (number of successes in n trials), then the point estimator for p is $\hat{P} = \frac{X}{n}$.

Recall that $\mathrm{E}[X] = np$ and $\mathrm{Var}[X] = np(1-p)$. We can standardize any random variable:

$$Z = \frac{X-\mu}{\sigma} = \frac{n\hat{P}-np}{\sqrt{np(1-p)}} = \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

is approximately $\mathcal{N}(0,1)$.

$$\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}.$$

Summary of Chapter 6: Hypothesis testing

Two types of errors can be committed when testing H_0 against H_1.

	Decision: reject H_0	**Decision:** fail to reject H_0
Reality: H_0 is True	Type I Error	No Error
Reality: H_0 is False	No Error	Type II Error

- If we reject H_0 when H_0 is true $\Rightarrow$ we have committed a **type I error**.
- If we fail to reject H_0 when H_0 is false $\Rightarrow$ **type II error**.

Probability of Committing Errors and Power

$$\boxed{\alpha = P\Big(\text{reject } H_0 \mid H_0 \text{ is true}\Big).}$$

$$\boxed{\beta = P\Big(\text{fail to reject } H_0 \mid H_0 \text{ is false}\Big).}$$

$$\boxed{\text{Power} = \mathrm{P}\Big(\text{reject } \mathrm{H}_0 \mid H_0 \text{ is false}\Big) = 1 - \beta.}$$

Conventional values of α, β, and Power are 0.05, 0.2, and 0.8, respectively.

One-sample testing

Procedure: to test for $H_0 : \mu = \mu_0$, where μ_0 is a constant.

Step 1: set $H_0 : \mu = \mu_0$

Step 2: select an alternative hypothesis H_1 (what we are trying to show using the data). Depending on context, we choose one of these alternatives:

- $H_1 : \mu < \mu_0$ (one-sided test)
- $H_1 : \mu > \mu_0$ (one-sided test)
- $H_1 : \mu \neq \mu_0$ (two-sided test)

Step 3: choose $\alpha = P(\text{type I error})$: typically $\alpha = 0.01$ or 0.05.

Step 4: for the observed sample $\{x_1, \ldots, x_n\}$, compute the observed value of the test statistics $z_0 = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$.

Step 5: determine the **critical region** as follows:

Alternative Hypothesis	Critical Region (Rejection Area)
$H_1 : \mu > \mu_0$	$z_0 > z_\alpha$
$H_1 : \mu < \mu_0$	$z_0 < -z_\alpha$
$H_1 : \mu \neq \mu_0$	$\lvert z_0 \rvert > z_{\alpha/2}$

where z_α is the critical value satisfying $P(Z > z_\alpha) = \alpha$, and $Z \sim \mathcal{N}(0, 1)$:

α	z_α	$z_{\alpha/2}$
0.05	1.645	1.960
0.01	2.327	2.576

Step 6: compute the associated $p-$**value** as follows:

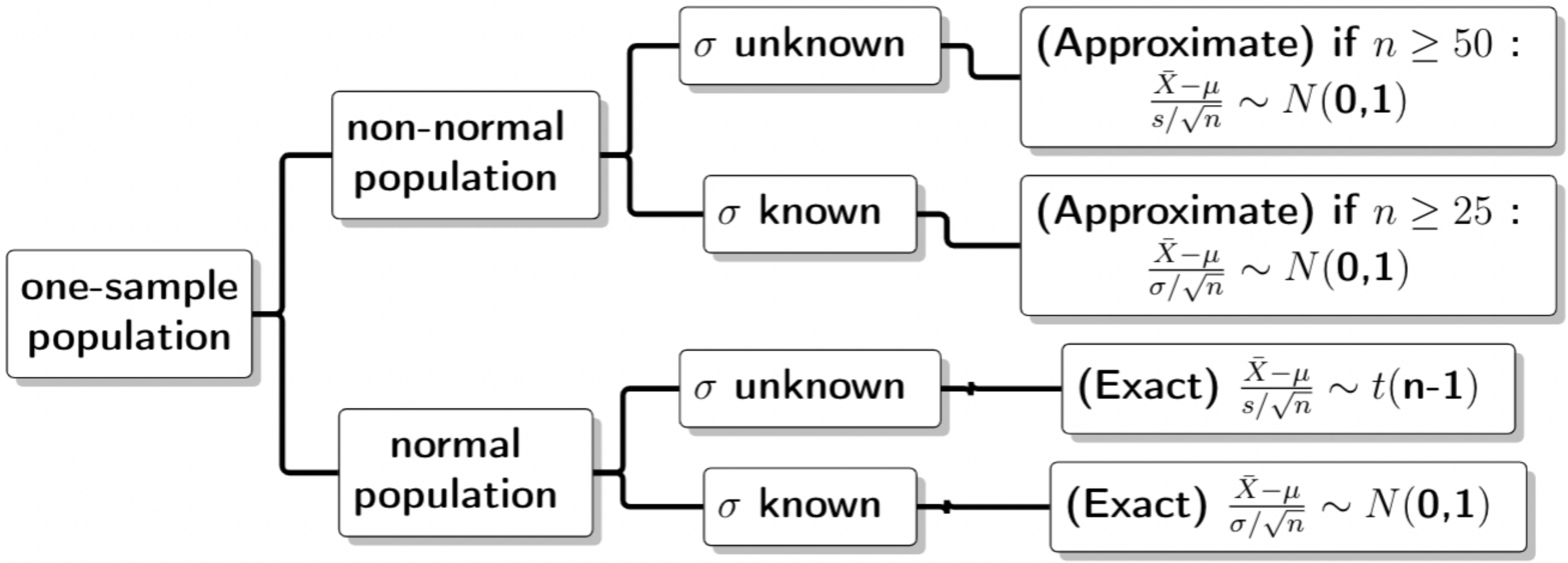
Alternative Hypothesis	$p-$**Value**
$H_1 : \mu > \mu_0$	$P(Z > z_0)$
$H_1 : \mu < \mu_0$	$P(Z < z_0)$
$H_1 : \mu \neq \mu_0$	$2 \cdot \min\{P(Z > z_0), P(Z < z_0)\} = 2\,P(Z > \lvert z_0 \rvert)$

where $Z \sim \mathcal{N}(0, 1)$.

Decision-Rule based on p-value: if the $p-$value $\leq \alpha$, then we **reject** H_0 **in favour of** H_1. If the $p-$value $> \alpha$, we **fail to reject** H_0**.**

Decision-Rule based on critical region: if the z_0 is in the critical region, then we **reject** H_0 **in favour of** H_1. If z_0 is not in the critical region, we **fail to reject** H_0**.**

NOTE: both decision-rules are equivalent.


one-sample population
non-normal population
normal population
σ unknown
σ known
σ unknown
σ known
(Approximate) if $n \geq 50$: $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim N(\mathbf{0,1})$
(Approximate) if $n \geq 25$: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(\mathbf{0,1})$
(Exact) $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(\mathbf{n\text{-}1})$
(Exact) $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(\mathbf{0,1})$

Two-sample test (independent populations)

Let $X_{1,1},\ldots,X_{1,n}$ be a random sample from a normal population with unknown mean μ_1 and variance σ_1^2; let $Y_{2,1},\ldots,Y_{2,m}$ be a random sample from a normal population with unknown mean μ_2 and variance σ_2^2, with both populations **independent** of one another. We want to test

$$\boxed{\mathrm{H}_0 : \mu_1 = \mu_2 \ \text{ against } \ H_1 : \mu_1 \neq \mu_2.}$$

Let $\overline{X} = \frac{1}{n}\Sigma_{i=1}^{n} X_i$, $\overline{Y} = \frac{1}{m}\Sigma_{i=1}^{m} Y_i$. The observed values are again denoted by lower case letters: $\overline{x}$, $\overline{y}$.

Case 1: σ_1^2 and σ_2^2 are Known

Alternative Hypothesis	Critical Region
$H_1 : \mu_1 > \mu_2$	$z_0 > z_\alpha$
$H_1 : \mu_1 < \mu_2$	$z_0 < -z_\alpha$
$H_1 : \mu_1 \neq \mu_2$	$\lvert z_0 \rvert > z_{\alpha/2}$

where $z_0 = \frac{\overline{x}-\overline{y}}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}$, z_α satisfies $P(Z > z_\alpha) = \alpha$, and $Z \sim \mathcal{N}(0,1)$.

Alternative Hypothesis	$p-$Value
$H_1 : \mu_1 > \mu_2$	$P(Z > z_0)$
$H_1 : \mu_1 < \mu_2$	$P(Z < z_0)$
$H_1 : \mu_1 \neq \mu_2$	$2 \cdot \min\{P(Z > z_0), P(Z < z_0)\} = 2\,P(Z > \lvert z_0 \rvert)$

- **Two Independent Populations**
  - **non-normal populations**
    - **Variances Unknown**
      - **(Approximate) if** $n \geq 50$ **and** $m \geq 50$**:**
        $$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{1/n+1/m}} \sim N(\mathbf{0,1})$$
        S_p^2 **is** $\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}$
    - **Variances Known**
      - **(Approximate) if** $n \geq 25$ **and** $m \geq 25$ **:**
        $$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}} \sim N(\mathbf{0,1})$$
  - **normal populations**
    - **Variances Unknown Unequal**
      - $$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{S_1^2/n+S_2^2/m}} \sim t(df)$$
        df **is estimated using Welch Approximation**
    - **Variances Unknown Equal**
      - $$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{S_p\sqrt{1/n+1/m}} \sim t(\mathbf{n+m\text{-}2})$$
        S_p^2 **is** $\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}$
    - **Variances Known**
      - $$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}} \sim N(\mathbf{0,1})$$

Two-Sample Test (Paired)

Let $X_{1,1}, \ldots, X_{1,n}$ be a random sample from a normal population with unknown mean μ_1 and unknown variance σ^2; let $X_{2,1}, \ldots, X_{2,n}$ be a random sample from a normal population with unknown mean μ_2 and unknown variance σ^2, with both populations **not independent** of one another (i.e., it's possible that the 2 samples come from the same population, or are measurements on the same units). We want to test

$$\boxed{H_0 : \mu_1 = \mu_2 \text{ against } H_1 : \mu_1 \neq \mu_2.}$$

In order to do so, we compute the differences $D_i = X_{1,i} - X_{2,i}$ and consider the $t-$test (as we do not know the variance). The test statistic is

$$T_0 = \frac{\overline{D}}{S_D/\sqrt{n}} \sim t(n-1),$$

where

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \,,$$

and

$$S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2 \,.$$

Example: $n = 10$ engineers' knowledge of basic statistical concepts was measured on a scale from $0-100$ before and after a short course in statistical quality control. The result are as follows:

Engineer	1	2	3	4	5	6	7	8	9	10
Before $X_{1,i}$	43	82	77	39	51	66	55	61	79	43
After $X_{2,i}$	51	84	74	48	53	61	59	75	82	48

Let μ_1 and μ_2 be the mean score before and after the course, respectively, with normally distributed scores. Test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$.

Solution: The differences $D_i = X_{1,i} - X_{2,i}$ are:

Engineer	1	2	3	4	5	6	7	8	9	10
Before X_{1i}	43	82	77	39	51	66	55	61	79	43
After X_{2i}	51	84	74	48	53	61	59	75	82	48
Difference D_i	-8	-2	3	-9	-2	5	-4	-14	-3	-5

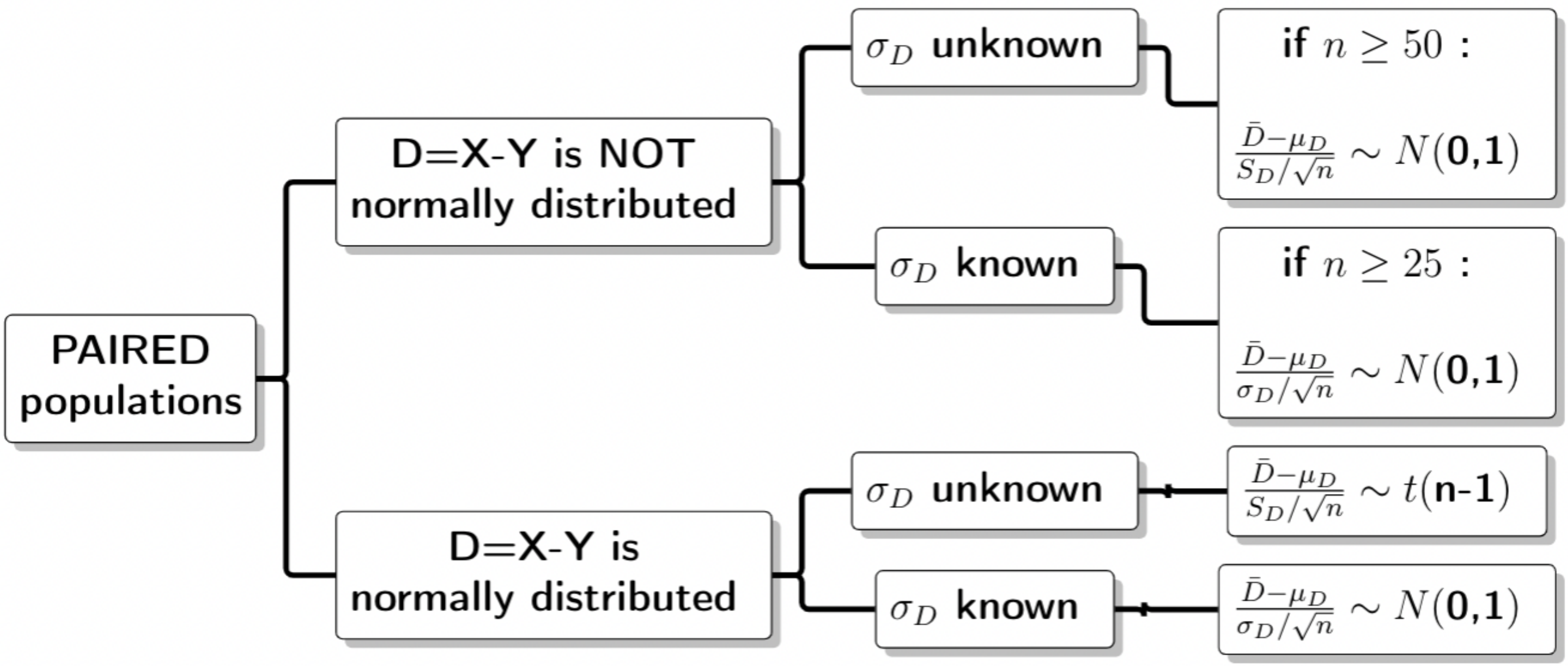
The observed sample mean is $\overline{d} = -3.9$, and the observed sample variance is $s_D^2 = 31.21$. The test statistic is

$$T_0 = \frac{\overline{D} - 0}{S_D/\sqrt{n}} \sim t(n-1), \text{ with observed value } t_0 = \frac{-3.9}{\sqrt{31.21/10}} \approx -2.21.$$

We compute

$$P(\overline{D} \leq -3.9) = P(T(9) \leq -2.21) = P(T(9) > 2.21).$$

But $t_{0.05}(9) = 1.833 < t_0 = 2.21 < t_{0.01}(9) = 2.821$, so we reject H_0 when $\alpha = 0.05$, but we do not reject H_0 when $\alpha = 0.01$.


PAIRED populations
D=X-Y is NOT normally distributed
σ_D unknown
if $n \geq 50$:
$\frac{\bar{D}-\mu_D}{S_D/\sqrt{n}} \sim N(\mathbf{0,1})$
σ_D known
if $n \geq 25$:
$\frac{\bar{D}-\mu_D}{\sigma_D/\sqrt{n}} \sim N(\mathbf{0,1})$
D=X-Y is normally distributed
σ_D unknown
$\frac{\bar{D}-\mu_D}{S_D/\sqrt{n}} \sim t(\mathbf{n-1})$
σ_D known
$\frac{\bar{D}-\mu_D}{\sigma_D/\sqrt{n}} \sim N(\mathbf{0,1})$

Summary of Chapter 7: Correlation & Linear regression

Sample Coefficient of Correlation

For paired data (x_i, y_i), $i = 1, \ldots, n$, the **sample correlation coefficient** of x and y is

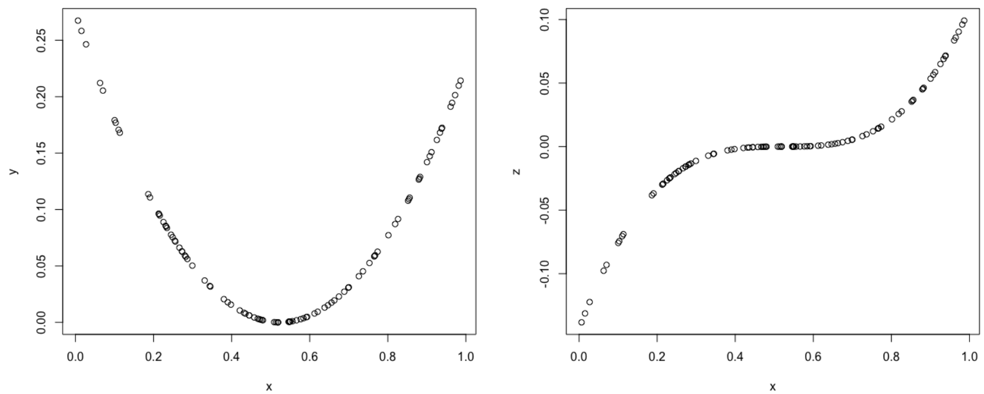
$$\boxed{\mathsf{r}_{XY} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}}.}$$

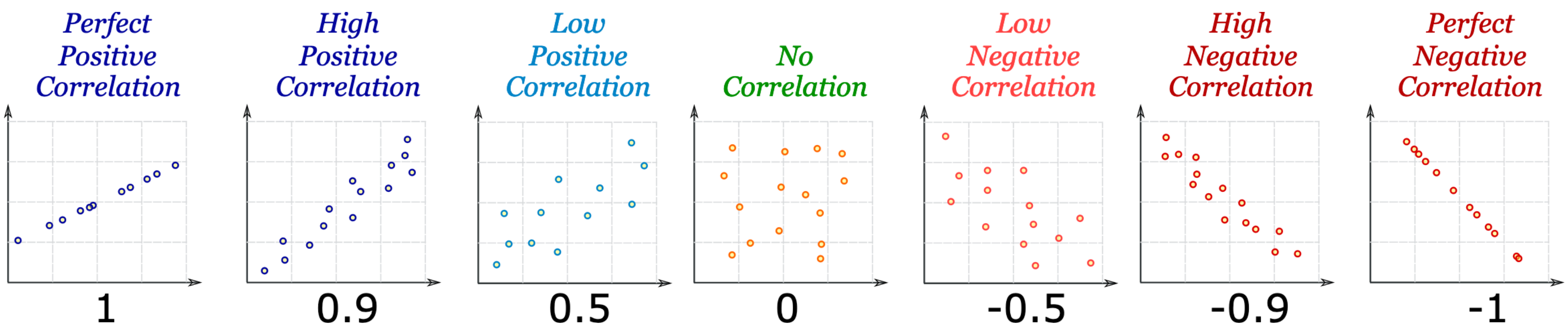
The coefficient r_{XY} is defined only if $S_{xx} \neq 0$ and $S_{yy} \neq 0$, i.e. neither x_i nor y_i are constant.

The variables x and y are **uncorrelated** if $r_{XY} = 0$ (or very small, in practice), and **correlated** if $\rho_{XY} \neq 0$ (or $|r_{XY}|$ is "large", in practice).

- r_{XY} is unaffected by changes of scale or origin. Adding constants to x does not change $x - \overline{x}$ and multiplying x and y by constants changes both the numerator and denominator equally;
- r_{XY} is symmetric in x and y (i.e. $r_{XY} = r_{YX}$)
- $-1 \leq r_{XY} \leq 1$;
- if $r_{XY} = \pm 1$, then the observations (x_i, y_i) all lie on a straight line with a positive (negative) slope;
- the sign of r_{XY} reflects the trend of the points;
- a high correlation coefficient value $|r_{XY}|$ does not necessarily imply a **causal relationship** between the two variables;

- note that x and y can have a very strong **non-linear** relationship without r_{XY} reflecting it (-0.12 on the left, 0.93 on the right).





Simple Linear Regression

Regression analysis can be used to describe the relationship between a **predictor variable** (or regressor) X and a **response variable** Y. Assume that they are related through the model

$$\boxed{\mathsf{Y}{=}\beta_0 + \beta_1 X + \varepsilon,}$$

where ε is a **random error** and β_0, β_1 are the **regression coefficients**.

It is assumed that $\mathrm{E}[\varepsilon] = 0$, and that the error's variance $\sigma_\varepsilon^2 = \sigma^2$ is constant.

Suppose that we have observations $(x_i, y_i),\ i = 1, \ldots, n$ so that

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n.$$

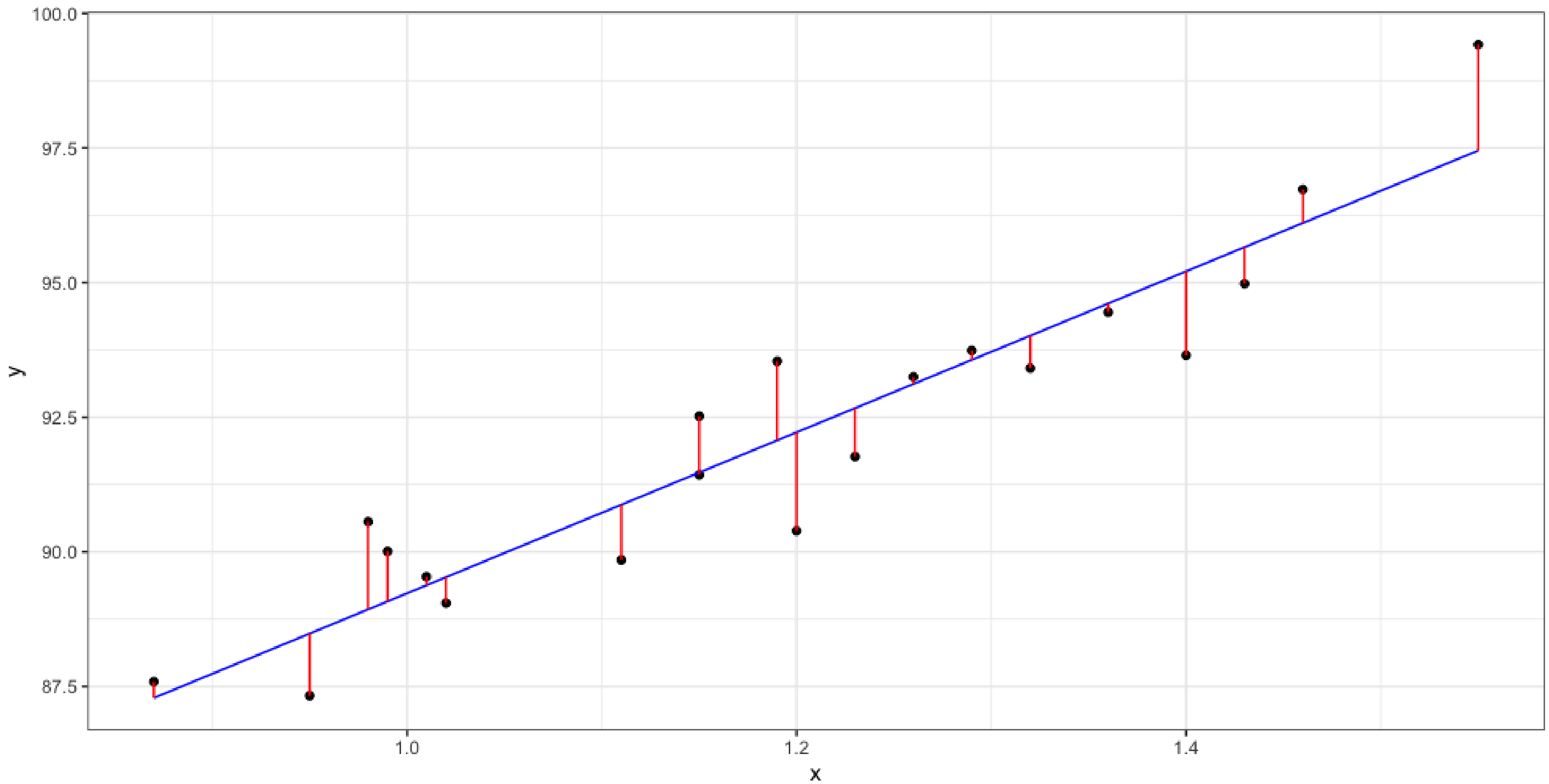
The aim is to find **estimators** b_0, b_1 of the unknown parameters β_0, β_1, in order to obtain the **estimated (fitted) regression line**

$$\hat{y}_i = b_0 + b_1 x_i$$

The **residual** or error in predicting y_i using $\hat{y}_i$ is thus

$$e_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i, \quad i = 1, \ldots, n.$$

How do we find the estimators? How do we determine if the fitted line is a good model for the data?



residuals: $e_i = y_i - \hat{y}_i$

Consider the **Sum of Squared Errors** (SSE):

$$\boxed{\text{SSE} = \Sigma_{i=1}^{n} e_i^2 = \Sigma_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2.}$$

(It can be shown that $\text{SSE}/\sigma^2 \sim \chi^2(n-2)$, but that's outside the scope of this course). **The optimal values of b_0 and b_1 are those that minimize the SSE**. As such, solving

$$0 = \frac{\text{dSSE}}{\text{d}b_0} = -2\Sigma(y_i - b_0 - b_1 x_i) = -2n(\overline{y} - b_0 - b_1\overline{x})$$

$$0 = \frac{\text{dSSE}}{\text{d}b_1} = -2\Sigma(y_i - b_0 - b_1 x_i)x_i = -2\left(\Sigma\, x_i y_i - n b_0 \overline{x} - b_1 \Sigma\, x_i^2\right)$$

yields the **least squares estimators** b_0, b_1 or β_0, β_1, respectively.

$$S_{xy} = \Sigma(x_i - \overline{x})(y - \overline{y}) = \Sigma\, x_i y_i - n\overline{x}\overline{y}$$
$$S_{xx} = \Sigma(x_i - \overline{x})^2 = \Sigma\, x_i^2 - n\overline{x}^2,$$

$$\boxed{\mathsf{b}_1 = \frac{S_{xy}}{S_{xx}} \quad , \qquad b_0 = \overline{y} - b_1\overline{x}.}$$

Estimating σ^2

For the regression error, the **unbiased estimator** of σ^2 is in fact

$$\boxed{\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} = \text{MSE} = \frac{\text{SSE}}{n-2} = \frac{S_{yy} - b_1 S_{xy}}{n-2},}$$

where the SSE has $n-2$ **degrees of freedom**, because 2 parameters had to be estimated in order to obtain $\hat{y}_i$: b_0 and b_1.

Properties of the Least Square Estimators

$$\mathrm{E}[b_0] = \beta_0, \qquad \sigma_{b_0}^2 = \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}} \right] = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n S_{xx}},$$

$$\mathrm{E}[b_1] = \beta_1, \qquad \sigma_{b_1}^2 = \sigma^2 / S_{xx}.$$

We say that b_0 and b_1 are **unbiased estimators** of β_0 and β_1, respectively. The **estimated standard errors** (replacing σ^2 by $\mathrm{MSE} = \hat{\sigma}^2$ in the expressions for $\sigma_{b_1}^2$ and $\sigma_{b_0}^2$ above) are

$$\mathrm{se}(b_0) = \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}} \right]} \quad \text{and} \quad \mathrm{se}(b_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}.$$

Hypothesis testing for the Intercept β_0

$$\boxed{\mathrm{H}_0 : \beta_0 = \beta_{0,0} \text{ against } H_1 : \beta_0 \neq \beta_{0,0}.}$$

$$\boxed{\mathrm{Z}_0 = \frac{b_0 - \beta_{0,0}}{\sqrt{\sigma^2 \frac{\sum x_i^2}{n S_{xx}}}} \sim \mathcal{N}(0,1).}$$

But σ^2 is not known, so the test statistic with $\hat{\sigma} = \mathrm{MSE}$

$$\boxed{\mathrm{T}_0 = \frac{b_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \frac{\sum x_i^2}{n S_{xx}}}} \sim t(n-2)}$$

follows a Student $t-$distribution with $n-2$ degrees of freedom.

Alternative Hypothesis	**Critical/Rejection Region**
$H_1 : \beta_0 > \beta_{0,0}$	$t_0 > t_\alpha(n-2)$
$H_1 : \beta_0 < \beta_{0,0}$	$t_0 < -t_\alpha(n-2)$
$H_1 : \beta_0 \neq \beta_{0,0}$	$\lvert t_0 \rvert > t_{\alpha/2}(n-2)$

where t_0 is the observed value of T_0 and $t_\alpha(n-2)$ is the $t-$value satisfying $P(T > t_\alpha(n-2)) = \alpha$, and $T \sim t(n-2)$.

Reject H_0 if t_0 in the critical region.

Hypothesis testing for the Slope β_1

$$\boxed{\mathsf{H}_0 : \beta_1 = \beta_{1,0} \ \text{ against } \ H_1 : \beta_1 \neq \beta_{1,0}.}$$

$$\boxed{\mathsf{Z}_0 = \frac{b_1 - \beta_{1,0}}{\sqrt{\sigma^2 / S_{xx}}} \sim \mathcal{N}(0, 1).}$$

But σ^2 is not known, so the test statistic with $\hat{\sigma}^2 = \mathrm{MSE}$

$$\boxed{\mathsf{T}_0 = \frac{b_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t(n-2)}$$

follows a Student $t-$distribution with $n-2$ degrees of freedom.

Alternative Hypothesis	Critical/Rejection Region
$H_1 : \beta_1 > \beta_{1,0}$	$t_0 > t_\alpha(n-2)$
$H_1 : \beta_1 < \beta_{1,0}$	$t_0 < -t_\alpha(n-2)$
$H_1 : \beta_1 \neq \beta_{1,0}$	$\lvert t_0 \rvert > t_{\alpha/2}(n-2)$

where t_0 is the observed value of T_0 and $t_\alpha(n-2)$ is the $t-$value satisfying $P(T > t_\alpha(n-2)) = \alpha$, and $T \sim t(n-2)$.

Reject H_0 if t_0 in the critical region.

Significance of Regression

Given a regression line, we may want to test whether it is **significant**. The test for **significance of the regression** is

$$\boxed{\mathrm{H}_0 : \beta_1 = 0 \text{ against } H_1 : \beta_1 \neq 0.}$$

If we reject H_0 in favour of H_1, then the evidence suggests that there is a linear relationship between X and Y.

C.I. for the Intercept β_0 and the Slope β_1

$$\boxed{\beta_0: \qquad b_0 \pm t_{\alpha/2}(n-2)\mathrm{se}(b_0) = b_0 \pm t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2 \frac{\sum x_i^2}{nS_{xx}}}}$$

$$\boxed{\beta_1: \qquad b_1 \pm t_{\alpha/2}(n-2)\mathrm{se}(b_1) = b_1 \pm t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}$$

Confidence Intervals for the Mean Response

The predicted value can be read directly from the regression line:

$$\hat{\mu}_{Y|x_0} = b_0 + b_1 x_0.$$

$\mathrm{E}[\hat{\mu}_{Y|x_0}] = \mu_{Y|x_0}$ and $\mathrm{Var}[\hat{\mu}_{Y|x_0}] = \sigma^2 \left[\frac{1}{n} + \frac{(x_0-\overline{x})^2}{S_{xx}}\right]$.

With the usual $t_{\alpha/2}(n-2)$, the $100(1-\alpha)\%$ C.I. for the **mean response** $\mu_{Y|x_0}$ (or for the line of regression) is

$$\hat{\mu}_{Y|x_0} \pm t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{(x_0-\overline{x})^2}{S_{xx}}\right]}.$$

where $\hat{\mu}_{Y|x_0} = b_0 + b_1 x_0$.

Predicting New Observations

If x_0 is the value of interest for the regressor (predictor), then the estimated value of the response variable Y is

$$\hat{y} = \hat{Y}_0 = b_0 + b_1 x_0.$$

a $100(1-\alpha)\%$ **prediction interval** for Y_0:

$$(b_0 + b_1 x_0) \pm t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2\left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]},$$

where $t_{\alpha/2}$ is the critical value of Student's $t-$distribution with $n-2$ degrees of freedom at α.

Variance Decomposition

- (x_i, y_i), $i = 1, \ldots, n$
- $\hat{y}_i$, $i = 1, \ldots, n$

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

- $SST = \Sigma_{i=1}^{n}(y_i - \bar{y})^2$ is total sum of squares=Total variation
- $SSE = \Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2$ is sum of squared errors=Unexplained variation
- $SSR = \Sigma_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ is Regression Sum of Squares= Explained variation

$$\boxed{\text{SST=SSE+SSR}}$$

ANOVA

The test for **significance of regression**,

$$\boxed{\text{H}_0 : \beta_1 = 0 \text{ against } H_1 : \beta_1 \neq 0,}$$

can be restated in term of the **analysis-of-variance** (ANOVA), given by the following table:

Source of Variation	Sum of Square	df	Mean Square	F^*	$p-$Value
Regression	SSR	1	MSR	$\frac{\text{MSR}}{\text{MSE}}$	$P(F > F^*)$
Error	SSE	$n-2$	MSE		
Total	SST	$n-1$			

In this table, the $F-$statistic $F^* \sim F(1, n-2)$, and

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \quad \text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2, \quad \text{SST} = \sum_{i=1}^{n}(y_i - \overline{y})^2,$$

$$\text{MSR} = \frac{\text{SSR}}{1}, \quad \text{MSE} = \frac{\text{SSE}}{n-2}, \quad \text{and} \quad F^* = \frac{\text{MSR}}{\text{MSE}} = \frac{\text{SSR}/1}{\text{SSE}/n-2}$$

The **rejection region** for the null hypothesis $H_0 : \beta_1 = 0$ is still given by

$$\left|\frac{b_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2/S_{xx}}}\right| > t_{\alpha/2}(n-2),$$

but it can also be written as $F^* > f_\alpha(1, n-2)$, where $f_\alpha(1, n-2)$ is the critical $F-$value of the F-distribution with $\nu_1 = 1$ and $\nu_2 = n-2$ df.

Coefficient of Determination

For observations (x_i, y_i), $i = 1, \ldots, n$, we define the **coefficient of determination** as

$$\boxed{\mathsf{R}^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}.}$$

The coefficient of determination is the proportion of the variability in the response that is explained by the fitted model. Note that R^2 always lies between 0 and 1; when $R^2 \approx 1$, the fit is considered to be very good.