Continuous Random Variables

General Information

- A function $f\colon \mathbb{R} \to \mathbb{R}$ is a *probability mass function* (pdf) of a continuous random variable X iff f is nonnegative and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.
- For any probability mass function f, we have $\mathrm{P}(a \le X \le b) = \int_a^b f(x)\,dx$. Whether the inequality is strict or nonstrict does not affect the above identity.
- A *mode* of X is any value m such that $f(m)$ is maximum.
- A *cumulative distribution function* (cdf) $F\colon \mathbb{R} \to [0,1]$ of a random variable X is defined by

$$F(x) \coloneqq P(X \le x) = \int_{-\infty}^{x} f(x)\,dx.$$

- When writing out the cdf as a piecewise function, we explicitly write out the range of values for each case. We reserve the use of "otherwise" for pdf's.
- Any cdf is continuous and nondecreasing.
- Let X be a continuous random variable with cdf F. To find the pdf g of any $y(X)$, we first find its cdf, then differentiate. We achieve this by reverse engineering $y(X) \le y$ to find an inequality that relates X with y. E.g. $e^X \le y$ iff $X \le \ln(y)$.
- A *median* of X is any value m such that $\mathrm{P}(X \le m) = F(m) = 1/2$.
- Mean/Expectation:

$$\mu = \mathrm{E}(X) \coloneqq \int_{-\infty}^{\infty} x f(x)\,dx \qquad \text{and} \qquad \mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx.$$

- Important property:

$$\mathrm{E}(ag(X) \pm bh(x)) = a\,\mathrm{E}(g(X)) \pm \mathrm{E}(h(X)).$$

- Variance:

$$\mathrm{Var}(X) \coloneqq \mathrm{E}(X^2) - [\mathrm{E}(X)]^2.$$

- Important property:

$$\mathrm{Var}(aX \pm b) = a^2\,\mathrm{Var}(X).$$

Special Continuous Random Variables

Definition 2.1

A continuous random variable X has a *normal distribution* with mean μ and standard deviation σ, denoted by $X \sim \mathrm{N}(\mu, \sigma^2)$, iff its pdf f is such that

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

General Information

- A normal distribution is symmetrical about the line $x = \mu$. That is

$$\mathrm{P}(X \leq \mu - \delta) = \mathrm{P}(X \geq \mu + \delta)$$

  for each $\delta > 0$. Note that the mean, median, and mode coincide with μ.
- Properties of the normal distribution. Let X and Y be independent, such that $X \sim \mathrm{N}(\mu, \sigma^2)$ and $Y \sim \mathrm{N}(m, s^2)$. Then, for any $n \in \mathbb{N}$ and $x, y \in \mathbb{R}$,
  - $nX \sim \mathrm{N}(n\mu, n^2\sigma^2)$,
  - $X_1 + X_2 + \cdots + X_n \sim \mathrm{N}(n\mu, n\sigma^2)$,
  - $aX \pm bY \sim \mathrm{N}(a\mu \pm bm, a^2\sigma^2 + b^2s^2)$.
- At times, the question may be phrased in a misleading manner. Try using some inference to figure out the intended interpretation.

Example 2.1

"The mass of the padding is 30% of the mass of a randomly selected light bulb of mass L. Find the probability that a light bulb with padding has mass c."
Then for any light bulb of mass L_1, the mass of the padding is $0.3L_2$ (and *not* $0.3L_1$). i.e. we are to find $\mathrm{P}(L_1 + 0.3L_2)$.

- A variable $Z \sim \mathrm{N}(0, 1)$ is said to follow the *standard* normal distribution.

  Note: Z is reserved for this purpose.
- Let $X \in \mathrm{N}(\mu, \sigma^2)$. Then, $\frac{X-\mu}{\sigma}$ follows the standard normal distribution.
- What `Tail` do we select for `invNorm`?

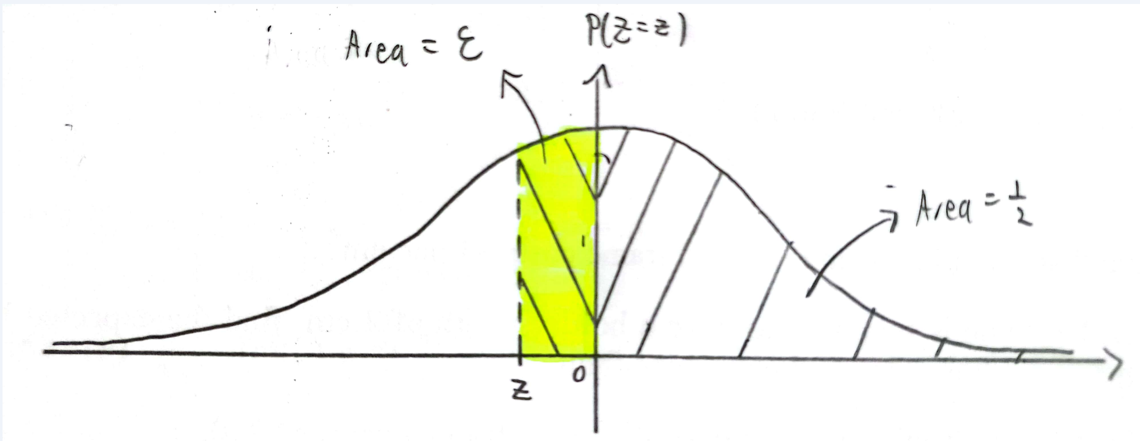
$\mathrm{P}(X < x) = p$	`LEFT`
$\mathrm{P}(-x < X < x) = p$	`CENTER`
$\mathrm{P}(X > x) = p$	`RIGHT`

- When using `invNorm` on an inequality, what should the sign be? For simplicity, we write $\mathscr{L}(p) = \texttt{invNorm}(p, 0, 1, \texttt{RIGHT})$, and $\mathscr{R}(p) = \texttt{invNorm}(p, 0, 1, \texttt{LEFT})$. Then,

$\mathrm{P}(Z > z) \geq p$	$z \leq \mathscr{L}(p)$
$\mathrm{P}(Z > z) \leq p$	$z \geq \mathscr{L}(p)$
$\mathrm{P}(Z < z) \geq p$	$z \geq \mathscr{R}(p)$
$\mathrm{P}(Z < z) \leq p$	$z \leq \mathscr{R}(p)$

Example 2.2

Suppose we want to find the least integer value of m for which $\mathrm{P}(Z > 1 - m) \geq 1/2$.
Then, using `invNorm (RIGHT)`, we infer that $z \leq 0$, *not* $z \geq 0$. An illustration:



Definition 2.2

continuous random variable X has a *uniform distribution* over the interval (a, b), which is denoted by $X \sim \mathrm{U}(a, b)$, iff its pdf f is such that

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2.3

continuous random variable Y has an (negative) exponential distribution, which we denote with $Y \sim \mathrm{Exp}(\lambda)$, iff its pdf g is such that

$$g(Y) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note

Let $Y \sim \mathrm{Exp}(\lambda)$, then

$$\mathrm{P}(Y > z + y \,|\, Y > y) = \mathrm{P}(Y > z).$$

- Expectation and variance:

Distribution	Expectation	Variance
$X \sim \mathrm{U}(a, b)$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
$Y \sim \mathrm{Exp}(\lambda)$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$

  Note: We need to remember the expectation and variance for the uniform distribution, as it is not provided in the MF26 formula sheet (unlike all other distributions).

- *Warning*: The G.C. tends to incorrectly process an integral if its upper and lower bounds contain $\pm$E99.

Sampling and Estimation

Definition 3.1

A sample is a finite subset of the population.

Definition 3.2

A random sample is a sample selected such that each member of the population has an equal probability of being selected into the sample.

Note

State, in context, what it means for the sample to be random.

> It means that every [a member of the population] has an equal probability of being selected into the sample.

Note

Explain why the sample would actually not be random.

> [Contextual reason], so not all the [members of the population] have an equal probability of being selected into the sample.

Definition 3.3

Any statistic T derived from a random sample and used to estimated an unknown population parameter θ is known as an *estimator*. It is an *unbiased* estimator iff $\mathrm{E}(T) = \theta$. If T is unbiased we commonly write $\hat{\theta}$ for T.

General Information

- Either write $\hat{\mu} = \bar{x} = \ldots$ or write out "Unbiased estimate of the population mean μ, $\bar{x} = \ldots$" Same holds for other population parameters θ.
- Estimators you should know:
-
Parameter	Estimator	Unbiased?	Formula
Population Mean μ	Sample Mean $\overline{X}$	✓	$\dfrac{X_1 + X_2 + \cdots + X_n}{n}$
Population Variance σ^2	Sample Variance σ_n^2	×	$\dfrac{\sum (X_i - \overline{X})^2}{n}$ $\dfrac{\sum X_i^2}{n} - \overline{X}^2$
	S^2	✓	$\dfrac{n}{n-1}\sigma_n^2$ $\dfrac{\sum (X_i - \overline{X})^2}{n-1}$ $\dfrac{1}{n-1}\left[\sum X_i^2 - \dfrac{(\sum X_i)^2}{n}\right]$
Population Proportion p	Sample Proportion P_s	✓	$\dfrac{X}{n}$

- Let X be a random variable following *any distribution*, and suppose we have a random sample $X_1, X_2, \ldots, X_n$ of size $n \geq 50$. Then by CLT (Central Limit Theorem), since $n \geq 50$ is large,

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{and} \qquad X_1 + X_2 + \cdots + X_n \sim \mathrm{N}(n\mu, \nu\sigma^2)$$

approximately.

- Assumptions when using CLT:
  - The sample is random.
  - Each X_i is independent and identically distributed.
- Suppose $X \sim \mathrm{N}(\mu, \sigma^2)$ is known and we pick a *particular* sample. Then,

Distribution	Is An Approximation?
$\overline{X} \sim \mathrm{N}(\mu, \sigma^2)$	No
$\overline{X} \sim \mathrm{N}(\bar{x}, \sigma^2)$	Yes
$\overline{X} \sim \mathrm{N}(\mu, s^2)$	Yes
$\overline{X} \sim \mathrm{N}(\bar{x}, s^2)$	Yes

  So, if we obtain any of the latter three in solving a question, we must write "$X \sim \mathrm{N}(__, __)$approximately" (even though we knew X *exactly* follows a normal distribution!)

- Pooled estimators. First assume we have two populations, from which we select a random sample of size n_1 and n_2. We let $\overline{X}_1$ and S_1^2 denote the sample mean and unbiased estimator for variance, respectively, for the first sample. Similarly define $\overline{X}_2$ and S_2^2, for the second sample.

Parameter	Unbiased Pooled Estimator
Mean	$\hat{\mu} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$
Variance	$S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$

The following definition is found in Hogg-McKean-Craig. Similar definitions are also found in Wackerly-Mendenhall-Schaefer and Nitis Mukhopadhyay.

Definition 3.4

Let $X_1, X_2, \ldots, X_n$ be a sample on a random variable X, where X has pdf $f(x;\theta)$, $\theta \in \Omega$. Let $0 < \alpha < 1$ be specified. Let $L = L(X_1, X_2, \ldots, X_n)$ and $U = U((X_1, X_2, \ldots, X_n))$ be two statistics. We say that the interval (L, U) is a $(1-\alpha)100\%$ *confidence interval* for θ iff

$$1 - \alpha = P_\theta[\theta \in (L, U)].$$

That is, the probability that the interval contains θ is $1-\alpha$, which is called the *confidence coefficient* or *confidence level* of the interval.

- We cannot write "a $1-\alpha$ (e.g. 0.95) confidence interval". The $1-\alpha$ must always be expressed as a *percentage.*
- Let $\hat{\theta}$ be a statistic that is normally distributed with mean θ and standard error $\sigma_{\hat{\theta}}$. We see that

$$\frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} = Z \sim \mathrm{N}(0, 1).$$

  Rewriting $\mathrm{P}(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$ gives

$$\mathrm{P}(\hat{\theta} - z_{1-\alpha/2}\sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{1-\alpha/2}\sigma_{\hat{\theta}}) = 1 - \alpha.$$

Hence, a $(1-\alpha)100\%$ confidence interval for θ is

$$(\hat{\theta} - z_{1-\alpha/2}\sigma_{\hat{\theta}}\ ,\ \hat{\theta} + z_{1-\alpha/2}\sigma_{\hat{\theta}}).$$

(Wackerly-Mendenhall-Schaefer)

- Let $0 < \alpha < 1$ and $X_1, X_2, \ldots, X_n$ be a sample on a random variable X with mean μ, where n is large. Then, an approximate $(1-\alpha)100\%$ confidence interval for μ is

  $$\left(\bar{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}\ ,\ \bar{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}\right).$$

  When the variance σ^2 is known, we can replace s with σ. If the distribution of X is known to be normal, in addition to σ^2 being known exactly, then the confidence interval is exact; it is not just an approximation.

  (Hogg-McKean-Craig)

- Let X be a Bernoulli random variable with probability of success p, where X is 1 or 0 if the outcome is success or failure, respectively. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the distribution of X, where n is large. Let $\hat{p} = \overline{X}$ be the sample proportion of successes. Then, an approximate $(1-\alpha)100\%$ confidence interval for p is given by

  $$\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

  (Letting $Y = X_1 + X_2 + \cdots + X_n \sim \mathrm{B}(n,p)$ gives $\hat{p} = Y/n$, which is the presentation used in the school's notes.)

  (Hogg-McKean-Craig)

Note

Standard phrasing for the interpretation of a $(1-\alpha)100\%$ confidence interval (a, b).

The probability that the interval (a, b) contains the true value of the [population mean/proportion in context] is $1-\alpha$.

Note

Standard phrasing for what is a $(1-\alpha)100\%$ confidence interval for θ?

It is an interval which has probability $1-\alpha$ of containing the true value of θ.

Note

Standard phrasing for whether mean/proportion in context has likely increased/decreased, when given suitable confidence intervals.

1. There is no conclusive result.

   Since the old and new $(1-\alpha)\%$ confidence intervals overlap, we are unable to conclude whether the [mean/proportion in context] has decreased or not. Hence, it is inconclusive from these figures as to whether the [context (e.g. an awareness campaign)] has been effective.

2. It has likely increased/decreased.

   The old $(1-\alpha)\%$ confidence interval is to the left/right of the new $(1-\alpha)\%$ confidence interval, such that they do not overlap. So, can conclude that the [mean/proportion in context] likely increased/decreased. Hence, these figures suggests that the [context (e.g. an awareness campaign)] has been effective.

Note

Advantage and disadvantage of a $(1-\beta)\%$ confidence interval compared to a $(1-\alpha)\%$ confidence interval, where $\beta < \alpha$.

Advantage:	A $(1-\beta)\%$ CI is more likely to contain the true mean.
Disadvantage:	A $(1-\beta)\%$ CI is less precise (or wider).

Note. Clearly state which is the advantage and disadvantage, as illustrated above.

G.C. Skills

Calculating statistics (i.e. $\bar{x}$, s, etc) by G.C. given data for a sample.

1. Keying in the data: `stat` $\implies$ `1:Edit` $\implies$ Key in the data into one of the lists $\texttt{L}_i$.
2. Calculating the statistic: `stat` $\implies$ `CALC` $\implies$ `1-Var Stats` ($\texttt{List:L}_i$) $\implies$ `Calculate`.
3. Getting the statistic for further calculations: `vars` $\implies$ `5:Statistics` $\implies$ Select the desired statistic.

G.C. Skills

Calculating the symmetric confidence interval by G.C.

Mean:	`stat` $\implies$ `TESTS` $\implies$ `7:ZInterval...`
Proportion:	`stat` $\implies$ `TESTS` $\implies$ `A:1-PropZInt...`

Statistics: Hypothesis Testing

Definition 4.1

The *null hypothesis* H_0 and *alternative hypothesis* H_1 are the hypotheses that we hope to reject and accept, respectively.

General Information

- Without going into details, a *critical region* C is just a set that defines the decision rule / test

$$\text{Reject } H_0 \text{ (Accept } H_1) \quad \text{if } (X_1, X_2, \ldots, X_n) \in C,$$

for any random sample $X_1, X_2, \ldots, X_n$ from the distribution of a random variable X.

Definition 4.2

The *significance level* $100\alpha\%$ of a test is the probability of rejecting H_0 when it is in fact true. i.e. $\alpha = \mathrm{P}(H_0 \text{ is rejected} \,|\, H_0 \text{ is true})$.

Note

Explain what the p-value means in context.

The p-value is the least value of the significance level to conclude that [H_1 in context].

Definition 4.3

The *p-value* is the lowest level of significance for which the null hypothesis will be rejected. In other words, for the null hypotheses

$$\text{(a) } \mu < \mu_0, \quad \text{(b) } \mu \neq \mu_0, \quad \text{(c) } \mu > \mu_0,$$

we have

$$\text{(a) } p\text{-value} = \mathrm{P}(Z \leq z_{\text{calc}}), \quad \text{(b) } p\text{-value} = \mathrm{P}(|Z| \geq |z_{\text{calc}}|), \quad \text{(c) } p\text{-value} = \mathrm{P}(Z \geq z_{\text{calc}}).$$

- A large sample hypothesis test for the mean.

  1. Test $H_0 : \mu = \mu_0$
     against $H_1 :$ (a) $\mu < \mu_0$, (b) $\mu \neq \mu_0$, or (c) $\mu > \mu_0$,
     at the $100\alpha\%$ significance level.
  2. Under H_0, we have $\overline{X} \sim \mathrm{N}(\mu_0, \hat{\sigma}^2/n)$ approximately. Or, if σ^2 is known exactly, then by CLT $\overline{X} \sim \mathrm{N}(\mu_0, \sigma^2/n)$ approximately.
  3. Test statistic:
$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1).$$

4. Find $z_{1-\alpha}$ or $z_{1-\alpha/2}$, which satisfies
   (a) $\mathrm{P}(Z < z_{1-\alpha}) = \alpha$,
   (b) $\mathrm{P}(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = \alpha$, or
   (c) $\mathrm{P}(Z > z_{1-\alpha})$.
5. Find the test statistic value
$$z_{\text{calc}} = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}}.$$
6. Reject H_0 iff
   (a) $z_{\text{calc}} < z_{1-\alpha}$,
   (b) $|z_{\text{calc}}| > z_{1-\alpha/2}$, or
   (c) $z_{\text{calc}} > z_{1-\alpha}$.
7. Write a conclusion.

4. Find the p-value using GC.
5. Reject H_0 iff p-value is less than α.

G.C. Skills

Calculating the p-value of a sample.

`stat` $\implies$ `TESTS` $\implies$ `1:Z-Test.`

- If we have a null hypothesis, such as
$$H_0 \colon \mu \leq \mu_0 \quad \text{or} \quad H_0 \colon \mu \geq \mu_0,$$
we can just use $H_0 \colon \mu = \mu_0$ instead.

Note

Standard phrasing for rejecting H_0.

> Since (a) $z_{\text{calc}} < z_{1-\alpha}$, (b) $|z_{\text{calc}}| > z_{1-\alpha/2}$, (c) $z_{\text{calc}} > z_{1-\alpha}$, or p-value $< \alpha$, we reject H_0. There is sufficient evidence at the significance level $100\alpha\%$ that [H_1 in context].

For *not* rejecting H_0, simply change to the appropriate inequality (such that z_{calc} is outside the critical region) and write "insufficient" instead of "sufficient".

Note

Explain, in context, the meaning of 'at the $\alpha\&$ level of significance'.

> The probability that [H_1 in context], when actually [H_0 in context], is $\alpha\%$.

Note

Explain why there is no need to assume that the distribution of X is normal/know anything about the population distribution of X.

> As the sample size n is large, by the Central Limit Theorem, the sample mean of [random variable X in context] will approximately follow a normal distribution.

Note. Spell "Central Limit Theorem" and "the sample mean" out *in full*. Do not use CLT or $\overline{X}$ for this question.

Correlation and Linear Regression

Note

A good scatter diagram should follow the guidelines below.

- The relative position of each point on the scatter diagram should be clearly shown.
- The range of values for the set of data should be clearly shown by marking out the extreme x and y values on the corresponding axis.
- The axes should be labeled clearly with the variables.

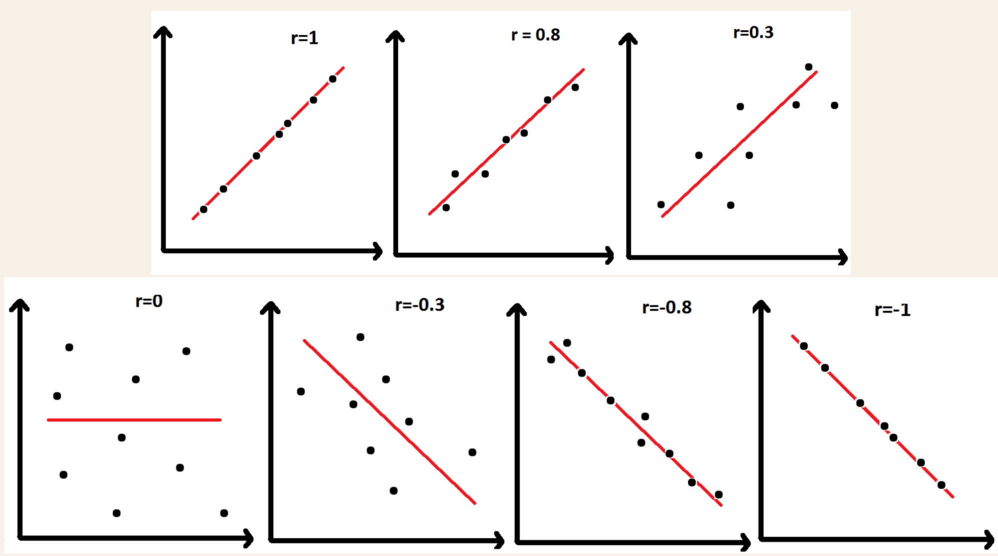
General Information

- The Product Moment Correlation Coefficient is a measure of the linear correlation between two variables. It is defined by

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}\right]\left[\sum y^2 - \dfrac{\left(\sum y\right)^2}{n}\right]}},$$

  which takes on a value from 0 to 1.

- When $r = 0$, there is no linear relationship. But, a nonlinear relationship may be present. Additionally, the regression lines are perpendicular.
- The closer the value of r is to 1 (or -1), the stronger the positive (or negative) linear correlation. Furthermore, the regression lines coincide.



- The regression line of y on x minimises the sum of squares deviation (error) in the y-direction. (i.e. we are assuming x is the independent variable whose values are known exactly.) It is given by

$$y = \bar{y} + b(x-\bar{x}), \qquad \text{where} \qquad b = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} = \frac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}.$$

- The point $(\bar{x}, \bar{y})$ always lies on both the regression lines of y on x, and x on y.
- Say we are given the value of one variable, and asked to approximate the the value of the other variable. Then, we should always use the line of the *dependent* variable on the *independent*.
- Estimations should not be taken for data outside the range of the sample provided, even if the value of r is close to 1.