- 6.6.2 Moments
- 6.6.3 Distributions
- 6.7 Bivariate Normal Distribution

c7. Confidence Interval

- 7.0 Pre-requisites
- 7.1 Introduction to $E(\bar{X})$ and $Var(\bar{X})$
- 7.2 Limit Theorems
- 7.4 Sampling Distribution
  - 7.4.1 Sampling Distribution of Sample Mean
  - 7.4.2 Sampling Distribution of Sample Variance

c8 Statistical Testing

- 8.1 Introduction
- 8.2 Asymmetric Case (One-sample Analysis) and Symmetric Case (Two-sample Analysis)
- 8.3 Test Error
- 8.4 Goodness of Fit Test
  - 8.4.1 Chi-squared Test (χ^2)
  - 8.4.2 Kolmogorov-Smirov (KS) Test

c9 Linear Regression

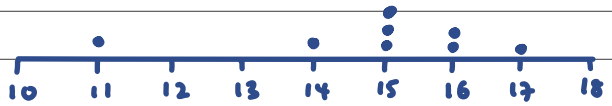
- 9.1 Introduction
- 9.2 Properties of Parameter Estimates
- 9.3 Confidence Interval (Population Parameters: a and b)
- 9.4 Confidence Interval ($E(Y|X=x_0)$ and $Y_{X=x_0}$)
- 9.5 Statistical Testing (Hypothesis Testing)
  - 9.5.1 Slope Parameter Test
  - 9.5.2 Correlation Test
  - 9.5.3 Regression Test (F-test)

c1. Descriptive Statistics

1.1 Graphical Representation.

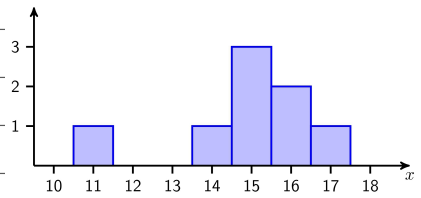
1. Dot plot.

{11, 14, 15, 16, 17, 15, 16, 15}

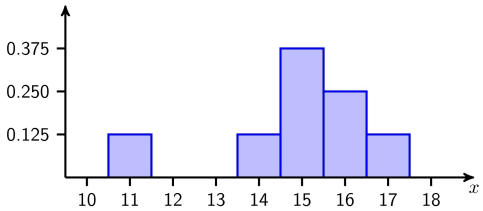


2. Histogram.

{11, 14, 15, 16, 17, 15, 16, 15}

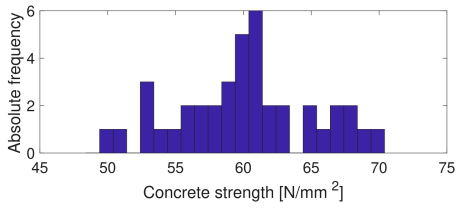


(absolute freq.)

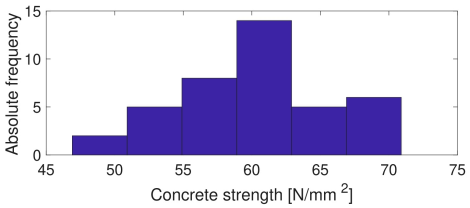


(relative freq.)

} different in frequency type



(bin size = 1)



(bin size = 4)

} different in bin size.

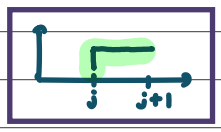
- bin size does not have to be identical in a single plot!

3. CDF for disrete (more on CDF in c3)

(cummulative relative frequency)

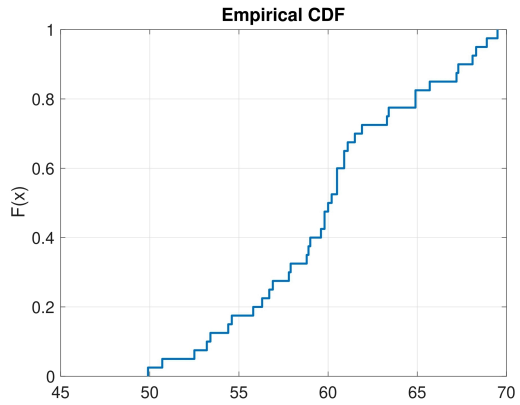
formula is given by:

$$F(x) = \begin{cases} 0 & \text{for } x < x_{(1)} \\ j/n & \text{for } x_{(j)} \le x < x_{(j+1)} \text{ and } 1 \le j \le n-1 \\ 1 & \text{for } x \ge x_{(n)} \end{cases}$$



all the green parts takes all the frequency before (j+1)

it's like bump up the cumulative frequency only once you reach exactly the x.



eg1:

The current speed in the River Thames has been measured every 3 hours over the course of a day in m/s as given below:

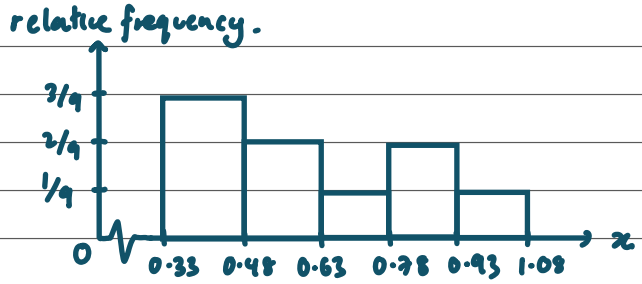
$$\{0.54;\ 0.82;\ 0.71;\ 0.92;\ 0.41;\ 1.07;\ 0.41;\ 0.56;\ 0.37\}$$

(a) Draw a histogram of the dataset with 0.15 m/s bin widths, starting at 0.33 m/s.

(b) Plot the empirical cumulative distribution frequency of the current speed.

(a)

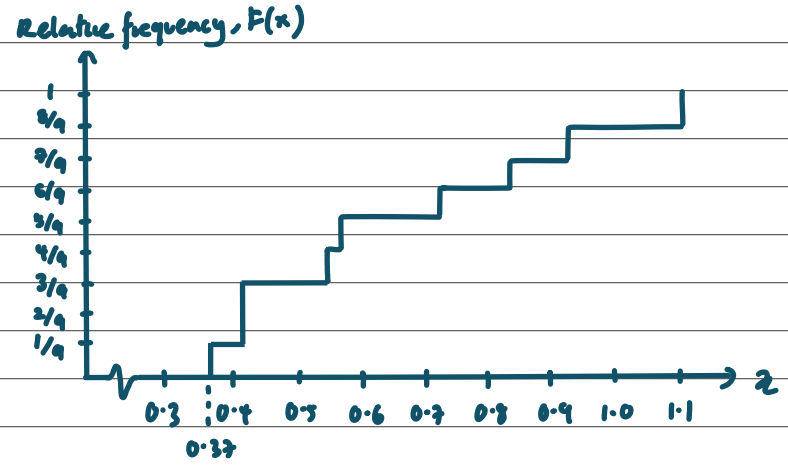
Histogram



Bin	f
0.33 - 0.48	3
0.48 - 0.63	2
0.63 - 0.78	1
0.78 - 0.93	2
0.93 - 1.08	1

$\Sigma f = 9$, $1/9 = 0.\overline{11}$

(b) CDF



j	$x_{(j)}$	F(x)
1	0.37	1/9
2	0.41	2/9
3	0.41	3/9
4	0.54	4/9
5	0.64	5/9
6	0.71	6/9
7	0.82	7/9
8	0.92	8/9
9	1.07	1

$n = 9$

HOW TO KNOW TO USE SAMPLE OR POPULATION?

Real data → use sample.
eg. "given these numbers...", "from this survey..."

Theoretical distributions → population.
eg. "X follows normal distribution...",

∴ Basically if experiment is conducted with real data, use sample.

1.2 Statistical Descriptors.

1.2.1 Measure of Central Tendancy.

Describe around which central value the data cluster.
Describe with: mean, mode, median.

1. Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\Sigma x}{n}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i = \frac{\Sigma f x}{\Sigma f}$$

2. Median.

$$\text{median (if } n \to \text{odd)} = x\left(\frac{n+1}{2}\right)$$

$$\text{median (if } n \to \text{even)} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$$

3. Mode.

mode = x with the highest frequency.
(can have more than one mode → multi-modal)

4. Geometric Mean.

$$\underset{\text{(geometric mean)}}{\overset{*}{x}} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

1·2·2 Measure of Dispersion.

Describe how spread out is the data.

Described with: range, interquartile range, variance, standard deviation, mean absolute deviation, coeff. of variance.

1. Range.

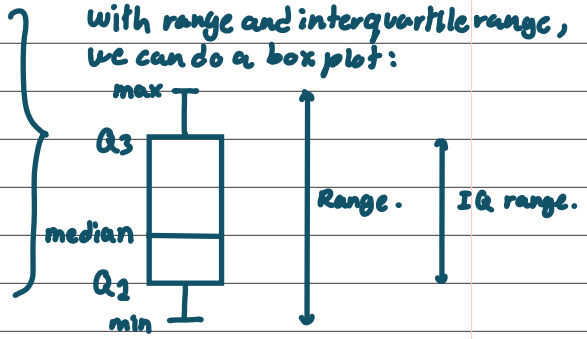
$\text{range} = \max(x) - \min(x)$.

$= x_{(n)} - x_{(1)}$

2. Interquartile Range.

$\text{I·Q Range} = q_{75}(x) - q_{25}(x)$

with range and interquartile range, we can do a box plot:



how to find quantile?

Method 1:

for $q_p(x)$, find j such that:

$$\frac{j-0.5}{n} < \frac{p}{100} \leq \frac{j+0.5}{n}$$

and take mean values of x_j and x_{j+1}.

eg. $x = \{1,4,6,6,7\}$, find $q_{25}(x)$

test $j=1$:

$\frac{1-0.5}{5} = 0.1$; $\frac{1+0.5}{5} = 0.3$ $\rightarrow$ $0.1 \leq 0.25 \leq 0.3$ $\rightarrow$ $j=1$ works!

$q_{25}(x) = \frac{x_{(1)} + x_{(1+1)}}{2} = \frac{1+4}{2} = 2.5$

Method 2:

find $k = \frac{p}{100}(n+1)$

if k lies on an integer, then $x_{(k)} = q_p(x)$

if k lies between two number (eg. 2.25), then $\frac{x_{(2)} + x_{(2)}}{2} = q_p(x)$

3. Variance. → population variance, σ^2 (population variance is biased low (under estimates true variance))
→ sample variance, s^2

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2}_{\text{easier to use.}} \quad \left(\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2\right) \text{ with frequencies version.}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$ } no simplified version formula

→ $\frac{1}{n-1}\sum f(x-\bar{x})^2$ (with frequencies version)

4. Standard Deviation.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \left(\sqrt{\frac{1}{n}\sum f(x-\bar{x})^2}\right) \text{ with frequencies}$$

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \left(\sqrt{\frac{1}{n-1}\sum f(x-\bar{x})^2}\right) \text{ with frequencies}$$

5. Coefficient of Variation.

used to compare the degree of spread across multiple datasets.

$$V = \frac{s}{\bar{x}}$$

6. Mean Absolute Deviation.

Better version of standard deviation.

because less influenced by extreme values.

$$d = \frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}| = \frac{1}{n}\sum|x-\bar{x}|$$

1.2.3 Measure of Asymmetry.

population skewness. currently use this.

sample skewness can use both of these

1. Biased skewness.

$$g_1(x) = \frac{\frac{1}{n}\sum(x-\bar{x})^3}{\sigma^3}$$

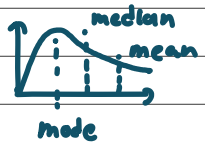
2. Unbiased skewness.

$$\hat{g}_1(x) = \frac{\sqrt{n(n-1)}}{n-2} g_1(x)$$

$$\sigma^3 = \left(\frac{\sum(x-\bar{x})^2}{n}\right)^{3/2}$$

we will use more biased skewness than unbiased.

i. $g_1(x) > 0$ (positive skewness)



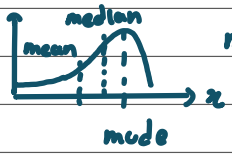
mode < median < mean.

ii $g_1(x) \approx 0$ (symmetric)

mode = median = mean

mode, mean, median.

iii $g_1(x) < 0$ (negative skewness)



mode > median > mean.

1.2.4 Measure Relation between Datasets. (to check whether there is any LINEAR correlation between x_i and y_i)

1. SAMPLE covariance.

(problem about this is $cov_{x,y}$ has units. if x and y has unit [m] then $cov_{x,y}$ has unit [m^2])

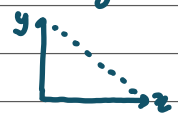
$$cov_{x,y} = \frac{1}{n-1}\sum(x-\bar{x})(y-\bar{y})$$

so to non-dimensionalise...

2. SAMPLE correlation coefficient.

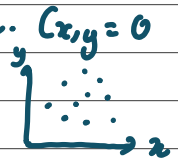
$$C_{x,y} = \frac{cov_{x,y}}{S_x S_y}, \quad \text{bounded by: } -1 \le C_{x,y} \le 1$$

i. $C_{x,y} = -1$:



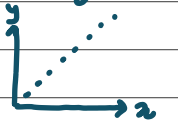
(perfect negative linear correlation).

ii. $C_{x,y} = 0$



(completely no correlation)

iii $C_{x,y} = 1$



(perfect positive linear correlation)

if population:

1. Population covariance

$$Cov(X,Y) = \frac{1}{n}\sum(x-\mu_x)(y-\mu_y) = E(XY) - E(X)E(Y)$$

2. Population correlation coefficient.

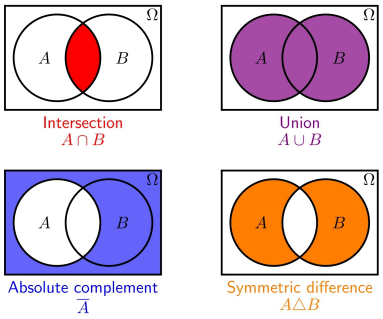
$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

c2. Probability

2.1 Probability.

Venn Diagram

Consider two sets A and B in the universal set Ω



Types of event.

same row (do something once)

1. Mutually Exclusive: events cannot happen at the same time. $P(A \cap B) = 0$
2. Mutually Inexclusive: events that can happen at the same time. $P(A \cap B) \neq 0$

two rows (do two things)

3. Independent: event will not be affected by any events. $P(A|B) = P(A)$; $P(B|A) = P(B)$
   $\downarrow$
   $P(A \cap B) = P(A) \times P(B)$
4. Dependent: event that will be affected by other events.

3 Theorems of Probability.

1. $P(\bar{A}) = 1 - P(A)$
2. if $A_i \cap A_j = \emptyset \ \forall i,j$: $P(A_1 \cup A_2 \cup \dots \cup A_k) = \sum_{i=1}^{k} P(A_i) \rightarrow P(A \cup B) = P(A) + P(B)$
3. if $A \cap B \neq \emptyset$: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

${}^nP_k$ is permutation (keyword: in order, sequence is important)

${}^nC_k$ is combination (keyword: choose, without order, sequence is not important)

$${}^nP_k = \frac{n!}{(n-k)!} \quad ; \quad {}^nC_k = \frac{n!}{(n-k)!k!}$$

Addition Rule

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

if mutually exclusive:

$P(A \cup B) = P(A) + P(B)$

Multiplication Rule.

$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

if independent:

$P(A \cap B) = P(A) \times P(B)$

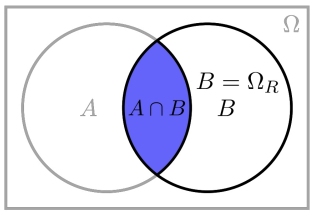
Bayes' Law

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

2.2 Conditional Probability.

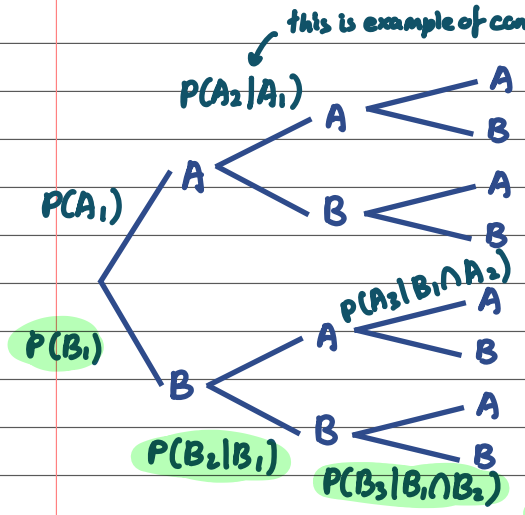
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

"Probability of A given B (already happened)"



sample space from Ω reduced to Ω_R.

TREE DIAGRAM (I will explain all the rule of Conditional Probability with Tree Diagram)



this is example of condition probability. 'Probability of 2nd is A given 1st is A'

IF the whole thing is INDEPENDENT

$P(A_2|A_1) = P(A_2) = P(A)$

(probability of getting A is constant throughout)

example of conditional probability.

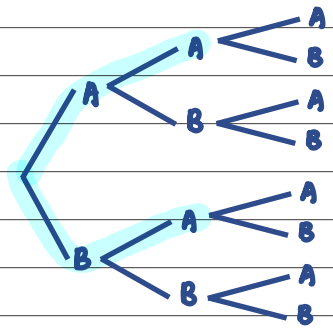
"Probability of 3rd is A given 1st is B, 2nd is A"

Chain Rule:

$$P(B_1 \cap B_2 \cap B_3) = P(B_1) \times P(B_2|B_1) \times P(B_3|B_1 \cap B_2)$$

or more generally:

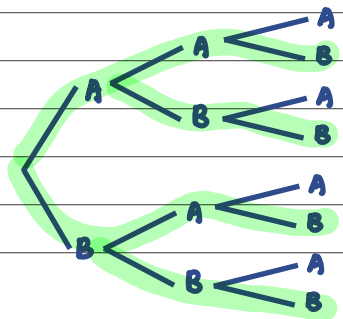
$$P(A \cap B \cap C \cap \dots \cap N) = P(A) \times P(B|A) \times P(C|B \cap A) \times P(D|C \cap B \cap A) \times \dots \times P(N|M \cap L \cap K \cap J \cap \dots \cap C \cap B \cap A)$$



Law of Total Probability: $P(B) = \sum_{i=1}^{k} P(B|A_i)\,P(A_i)$ (chain rule.)

eg1. $P(A_2) = P(A_1 \cap A_2) + P(B_1 \cap A_2)$

$= P(A_2|A_1)P(A_1) + P(A_2|B_1)P(B_1)$



eg2. $P(B_3) = P(A_1 \cap A_2 \cap B_3) + P(A_1 \cap B_2 \cap B_3) + P(B_1 \cap A_2 \cap B_3) + P(B_1 \cap B_2 \cap B_3)$

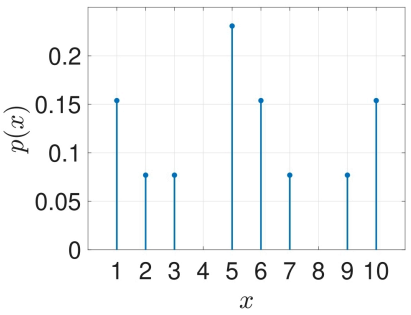
$= P(B_3|A_1 \cap A_2)P(A_2|A_1)P(A_1) + P(B_3|A_1 \cap B_2)P(B_2|A_1)P(A_1) + \dots$

c3. Random Variables.

3.1 Random Variables

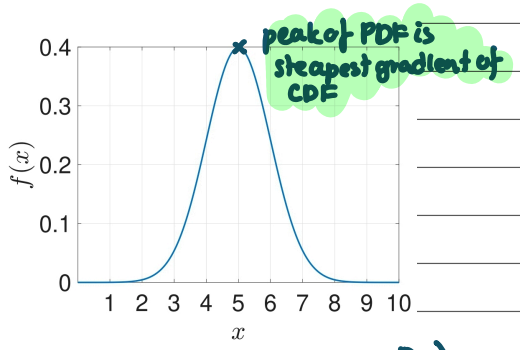
Probability Mass Function (PMF)

- tell us the probability of each DISCRETE random variable.
- $p(x) = P(X = k)$
  → $P(X = k) = p(x)$
- $\sum_x p(x) = 1$



Probability Density Function (PDF)

- tell us the probability of CONTINUOUS random variable within a certain range.
- $f(x) = \frac{dF(x)}{dx}$
  → $P(a \le X \le b) = \int_a^b f(x)\,dx$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$



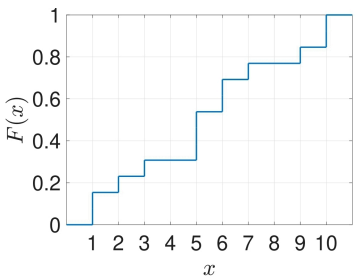
note that saying $P(X = a)$ for PDF will mean nothing, we need a RANGE
$P(X = a) = 0$
$P(a \le X \le b) \neq 0$

Cummulative Distribution Function (CDF)

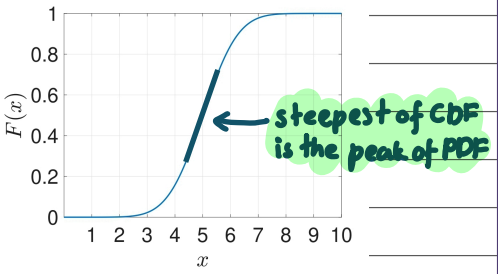
d/dx ∫dx

- $F(k) = P(X \le k)$

for discrete data:



for continuous data:

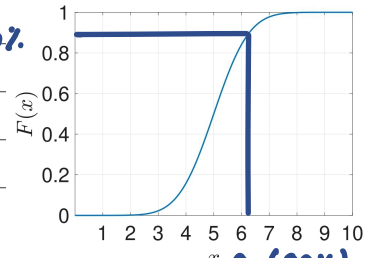


quantile function, $q_p(x)$

- quantile function is the inverse of CDF.

$q_p(p/100) = u \quad P(X \le u) = p/100 \quad F(u) = p/100 \} \; q(x) = F^{-1}(x)$ ☆

$p\% = 90\%$



$q_p(90\%) = 6.25$

☆ note that:
$q_p(x) = F^{-1}(x)$ is exactly as $q_p\left(\frac{p}{100}\right) = F^{-1}\left(\frac{p}{100}\right)$
we usually write the second form as $x = p/100$

eg. $f(x) = \frac{x}{20}\sin\left(\frac{\pi x}{10}\right)$ for $0 \le x \le 10$, find 90^{th} percentile.

Step 1: Find $F(x)$

from $0 \le x \le 10$:

$$F(x) = P(X \le x) = \int_0^x \frac{t}{20}\sin\left(\frac{\pi t}{10}\right)dt$$

$$= \frac{x}{20}\left[-\frac{10}{\pi}\cos\left(\frac{\pi t}{10}\right)\right]_0^x$$

$$= \frac{1}{2} - \frac{1}{2}\cos\left(\frac{\pi x}{10}\right), \quad x \in [0, 10]$$

Step 2: find $q_p(x)$

$$q_p\left(\tfrac{p}{100}\right) = F^{-1}\left(\tfrac{p}{100}\right)$$

let $F^{-1}(x) = y$

$F(y) = x$

$$\frac{1}{2} - \frac{1}{2}\cos\left(\frac{\pi y}{10}\right) = x$$

$$y = \frac{10}{\pi}\cos^{-1}(1 - 2x)$$

$$F^{-1}(x) = \frac{10}{\pi}\cos^{-1}(1 - 2x)$$

$$\therefore q_p\left(\frac{p}{100}\right) = \frac{10}{\pi}\cos^{-1}\left(1 - \frac{p}{50}\right)$$

Step 3: Sub $p = 90$:

$$q_{90}\left(\tfrac{90}{100}\right) = \frac{10}{\pi}\cos^{-1}\left(1 - \frac{90}{50}\right) = 7.952 \text{ //}$$

NOT in Y2!

3-2 Function of Random Variable (Conversion of PMF, PDF and CDF)

1. PMF conversion.

let $Y = h(X)$

$p_Y(y) = p_X(h^{-1}(y))$

$$\begin{aligned} p_Y(y) &= P(Y=y) \\ &= P(h(X)=y) \\ &= P(X=h^{-1}(y)) \\ &= p_X(h^{-1}(y)) \end{aligned}$$

eg. if we have:

$$p_X(x) = \begin{cases} x/3, & x=0 \\ 1-x/3, & x=1 \end{cases}$$

and $Y = 2X$

$$\begin{aligned} p_Y(y) &= P_X(Y=y) \\ &= P_X(2X=y) \\ &= P_X(X=y/2) \\ &= \begin{cases} y/6, & y=0 \\ 1-y/6, & y=2 \end{cases} \end{aligned}$$

2. PDF conversion.

let $Y = h(X)$

$$f_Y(y) = f_X(x)\left|\frac{dh(x)}{dx}\right|^{-1}$$

or usually written as.

$$f_Y(y)|dy| = f_X(x)|dx|$$

(absolute values are dropped, but remember dx dy considered positive)

not in syllabus, included to complete

3. CDF conversion

let $Y = h(X)$

$$F_Y(y) = P(Y \le y)$$

$$= P(h(X) \le y)$$

~~~ depends on $h(X)$ to continue. as there is an inequality.

eg. $h(X) = X+2$

$$F_Y(y) = P(X+2 \le y) = P(X \le y-2) = F_X(y-2)$$

eg. $h(X) = -X+3$

$$F_Y(y) = P(-X+3 \le y) = P(X \ge 3-y) = 1 - P(X < 3-y) = 1 - F_X(3-y)$$
~~~

3.3 Moments

1. Moment (order of m)

$$E(X^m) = \sum_x x^m p(x) \quad \text{(discrete)} \qquad \text{or} \qquad E(X^m) = \int_{-\infty}^{\infty} x^m f(x)\,dx \quad \text{(continuous)}$$

replace X with $X - E(X)$

2. Centred Moment (order of m)

$$E[(X-E(X))^m] = \sum_x (x-E(x))^m p(x) \quad \text{(discrete)} \qquad \text{or} \qquad E[(X-E(X))^m] = \int_{-\infty}^{\infty} (x-E(x))^m f(x)\,dx \quad \text{(continuous)}$$

1st moment is actually just mean:

$$E(X^1) = E(X)$$

2nd centred moment is actually just variance:

$$E[(X-E(X))^2] = Var(X)$$

simplified to $= E(X^2) - (E(X))^2$

3rd centred moment, divided by s^3 is the skewness.

$$g_1(X) = \frac{E[(X-E(X))^3]}{(s(X))^3}$$

eg. $E[(h(X)-E(h(X)))^3]$

☆ Moments of $Y = h(X)$ (just sub $Y = h(X)$!) → same with central moment.

$$E(Y^m) = E(h(X)^m) = \sum_x \underbrace{h(x)^m}_{Y=h(X)} \underbrace{p(x)}_{\text{PMF of } X} \quad \text{(discrete)} \qquad \text{or} \qquad E(Y^m) = E(h(X)^m) = \int_{-\infty}^{\infty} \underbrace{h(x)^m}_{Y=h(X)} \underbrace{f(x)\,dx}_{\text{PDF of } X} \quad \text{(continuous)}$$

not the same!

3.4 Expectation Algebra.

$E(a) = a$	$Var(a) = 0$
$E(aX) = aE(X)$	$Var(aX) = a^2 Var(X)$
$E(a \pm X) = a \pm E(X)$	$Var(a \pm X) = Var(X)$
$E(a \pm bX) = a \pm bE(X)$	$Var(a \pm bX) = b^2 Var(X)$
$E(X \pm Y) = E(X) \pm E(Y)$	$Var(X \pm Y) = Var(X) + Var(Y)$

$E(aX+bY+c) = aE(X) + bE(Y) + c$

only hold if X and Y are independent!

$Var(aX+bY+c) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$ if independent!

c4. Probability Distribution.

4.1 Disrete Distribution.

1. Bernoulli Distribution. (1 trial, 2 outcomes)

$X \sim B(1,p)$

$P(X=k) = p^k(1-p)^{1-k}$ for $k \in \{0,1\}$ } since n=1, k can only be 0 or 1 (either happen or no)

$E(X)=p$; $Var(X)=p(1-p)$

increase no. of trials ↓

2. Binomial Distribution (n trial, each trial INDEPENDENT, 2 outcomes)

$X \sim B(n,p)$

$P(X=k) = {}^nC_k\, p^k(1-p)^{n-k}$, for $k \in \{0,1,\dots,n\}$

$E(X)=np$; $Var(X)=np(1-p)$

as n approaches infinity. ↓

3. Poisson Distribution. (number of INDEPENDENT occurence over a reference time)

$X \sim Po(\mu)$ (μ is the average number of occurence over a reference time)

$P(X=k) = \frac{e^{-\mu}\mu^k}{k!}$, for $k \in \{0,1,\dots\}$

$E(X)=\mu$, $Var(X)=\mu$

4. Geometric Distribution. (infinite trial, each trial INDEPENDENT, 2 outcomes only, sequences of failure followed by one success)

$X \sim Geo(p)$

$P(X=k) = p(1-p)^k$ (if k is number of failure till success) or $P(X=k) = p(1-p)^{k-1}$ for $k \in \{0,1,\dots\}$ (if k is number of trials till success)

$P(X>k) = (1-p)^k$

$E(X) = \frac{1-p}{p}$; $Var(X) = \frac{1-p}{p^2}$

$E(X) = \frac{1}{p}$; $Var(X) = \frac{1-p}{p^2}$

Both are correct!

* This is the version given in data sheets.

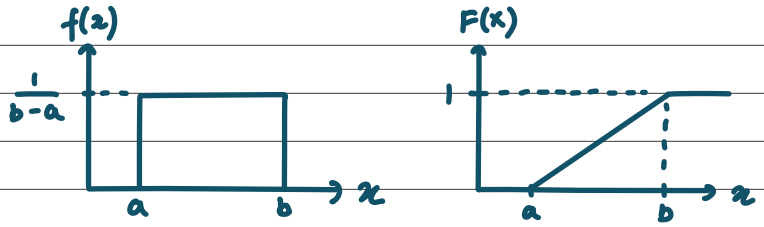
4.2 Continuous Distribution.

1. Uniform Distribution. (same probability over a given interval)

$X \sim U[a,b]$

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } x \in [a,b] \\ 0, & \text{otherwise.} \end{cases}$$

$$E(X) = \frac{a+b}{2} \; ; \; Var(X) = \frac{(b-a)^2}{12}$$



(PDF) (CDF)

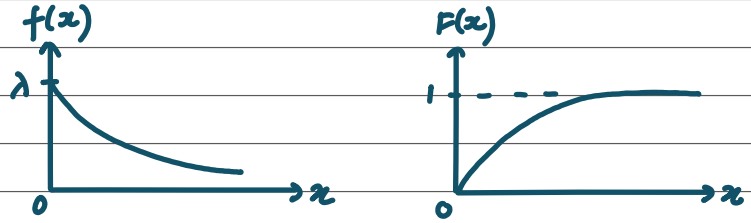
★ similar to geometric

2. Exponential Distribution. (time taken till INDEPENDENT events occur at rate λ, λ is no. of occurence per unit time)

$X \sim Exp(\lambda)$

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

$$E(X) = \frac{1}{\lambda} \; ; \; Var(X) = \frac{1}{\lambda^2}$$



(PDF) (CDF)

$X \sim Po(\mu)$, μ is no. of occurence per reference time, so $\mu = \lambda x$!

$$P(X=0) = \frac{e^{-\lambda x}(\lambda x)^0}{0!} = e^{-\lambda x}$$

scale λ to μ

$X \sim Exp(\lambda)$

$P(X > x) = 1 - F(x)$

since $F(x) = \int \lambda e^{-\lambda x} dx$

$= e^{-\lambda x}$

same.

$X \sim Po(\lambda t)$, $Y \sim Exp(\lambda)$

$P(X=0) = P(Y > t)$

★ $X \sim Po(\mu)$ and $X \sim Exp(\lambda)$

μ and λ difference:

eg μ: study three times in 8 days. ← in reference time frame.

λ: study at rate $0.375\ day^{-1}$

3. Gamma Distribution.

$X \sim Gamma(\alpha, \beta)$ (α and β are shape and scale parameters)

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

★ when $\alpha = 1$, $X \sim Gamma(1, \beta) = X \sim Exp(\frac{1}{\beta})$

$$E(X) = \alpha\beta \; ; \; Var(X) = \alpha\beta^2$$

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad , \alpha \in \mathbb{R}$

for $\alpha \in \mathbb{Z}$: $\Gamma(\alpha) = (\alpha-1)!$

} basically $\Gamma(\alpha)$ function extends factorial to real domain.

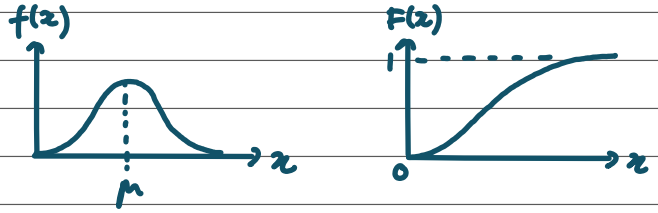
Gamma distribution graph varies a lot as α and β determine its shape.
for example, when $\alpha = 1$, $\beta = 1/\lambda$, $X \sim \text{Gamma}(1, 1/\lambda) = X \sim \text{Exp}(\lambda)$!
$X \sim \text{Gamma}(k/2, 2) = X \sim \chi^2(k)$!

4. Normal Distribution (symmetry about μ, deviation with σ among datas)

$X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \forall x \in \mathbb{R}$$

$E(X) = \mu$; $\text{Var}(X) = \sigma^2$



However, $f(x)$ for $X \sim N(\mu, \sigma^2)$ is complicated to compute so:

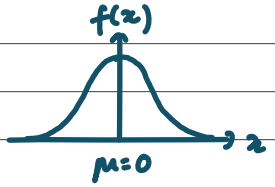
Standard Normal Distribution.

from $X \sim N(\mu, \sigma^2)$

let $z = \frac{X-\mu}{\sigma}$

to $z \sim N(0,1)$ or $\left(\frac{X-\mu}{\sigma}\right) \sim N(0,1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}, \forall z \in \mathbb{R}$$



still have to compute? the CDF, $F(z)$ or known as $\Phi(z)$ can be referred from 'Normal Distribution Table':

NORMAL CUMULATIVE DISTRIBUTION FUNCTION

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

eg. $\Phi(0.56) = 0.7123$.

eg 1. Exponential Distribution Question (can be done with Poisson!)

During the period 1836 – 1961, 16 earthquakes of magnitude 6 or more on the Richter scale happened in San Francisco. What is the probability that an earthquake of magnitude 6 or more does not occur in the next 10 years?

Exponential Distribution

$X \sim Exp(\lambda)$

λ is the rate of occurrence ie: (number of occurrence per unit time)

$$\lambda = \frac{16}{1961-1836} = 0.128 \text{ per year}$$

$$P(X>10) = 1 - P(X \leq 10)$$
$$= 1 - \underbrace{F(10)}_{CDF}$$

'X in exponential distribution is time till success'

X>10 means at least 10 years till success (earthquake)

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

for $x \geq 0$,

$$F(x) = \int_0^x \lambda e^{-\lambda t}\, dt = \left[-e^{-\lambda t}\right]_0^x$$
$$= 1 - e^{-\lambda x}$$

$$P(X>0) = 1 - F(10)$$
$$= 1 - (1 - e^{-0.128(10)})$$
$$= 0.278$$

Poisson Distribution.

$X \sim Po(\mu)$

μ is number of occurence per reference time.

$\mu = 0.128 \times 10 = 1.28$ per ten year.

$$P(X=0) = \frac{e^{-\mu}\mu^0}{0!}$$

'X in Poisson is x occurence in reference time'

$$= \frac{e^{-1.28}\,1.28^0}{0!}$$
$$= 0.278$$

X=0 means 0 occurence (earthquake) in reference time (10 years)

4.3 Derivation of E(X) and Var(X) for all distribution

4.3.1 Discrete Distribution.

$E(X) = \sum_{\text{ALL } i} x_i P(X = x_i)$, $\text{Var}(X) = \sum_{\text{ALL } i} x_i^2 P(X = x_i) - [\sum_{\text{ALL } i} x_i P(X = x_i)]^2$ ← $E(X^2) - [E(X)]^2$

← ALWAYS KNOW YOUR DISTRIBUTION DOMAIN!

1. Bernoulli Distribution.

$X \sim B(1, p)$

$P(X = k) = p^k (1-p)^{1-k}$ for $k \in \{0, 1\}$

$$E(X) = \sum_{k=0}^{1} k P(X = k)$$

since $k \in \{0, 1\}$,

$E(X) = 0 P(X=0) + 1 P(X=1)$

$= 0 \cdot p^0 (1-0)^{1-0} + 1 \cdot p^1 (1-p)^{1-1}$

$= p$

$$\text{Var}(X) = \sum_{k=0}^{1} k^2 P(X = k) - E(X)^2$$

$= 0^2 P(X=0) + 1^2 P(X=1) - p^2$

$= p - p^2$

$= p(1-p)$

2. Binomial Distribution

$X \sim B(n, p)$

$P(X = k) = {}^nC_k \, p^k (1-p)^{n-k}$, for $k \in \{0, 1, \ldots, n\}$

$$E(X) = \sum_{k=0}^{n} k P(X = k)$$

but $k=0$ doesn't contribute ($0 \times$ anything $= 0$), so why not start at 1.

$$= \sum_{k=1}^{n} k \, {}^nC_k \, p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} k \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$ ← $k! = k(k-1)!$

$$= \sum_{k=1}^{n} \frac{n! \leftarrow n(n-1)!}{(n-k)!(k-1)!} p^k (1-p)^{n-k}$$

look a bit like ${}^{n-1}C_{k-1} = \frac{(n-1)!}{(n-1-k+1)!(k-1)!} = \frac{(n-1)!}{(n-k)!(k-1)!}$

$$= \sum_{k=1}^{n} n \, {}^{n-1}C_{k-1} \, p^k (1-p)^{n-k}$$

← since till this step, everything got (-1) why not $p^k = p \, p^{k-1}$?

bring the 'p' out

last to modify: $n-k=(n-1)-(k-1)$

$$= np \sum_{k=1}^{n} {}^{n-1}C_{k-1}\, p^{k-1}(1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} {}^{n-1}C_{k-1}\, p^{k-1}(1-p)^{(n-1)-(k-1)}$$

$k=n \Rightarrow k-1=n-1$, $k=1 \Rightarrow k-1=0$

$$= np \sum_{k-1=0}^{n-1} {}^{n-1}C_{k-1}\, p^{k-1}(1-p)^{(n-1)-(k-1)}$$

binomial formula $\rightarrow [p+(1-p)]^{n-1} = 1^{n-1} = 1$

$$= np$$ //

$$\text{Var}(X) = E(X^2) - E(X)^2$$

$$= \sum_{k=0}^{n} k^2\, {}^{n}C_k\, p^k (1-p)^{k-1} - (np)^2$$

same steps as before.

$$= np \sum_{k=1}^{n} k\, {}^{n-1}C_{k-1}\, p^{k-1}(1-p)^{(n-1)-(k-1)}$$

this is a little bit hard...

3. Poisson Distribution.

$X \sim Po(\mu)$

$P(X=k) = \frac{e^{-\mu}\mu^k}{k!}$, for $k \geq 0$

$e^{-\mu}$: constant.

$$E(X) = \sum_{k=0}^{\infty} k \frac{e^{-\mu}\mu^k}{k!} \quad (k! = k(k-1)!)$$

$$= e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu \cdot \mu^{k-1}}{(k-1)!}$$

$k=0$ doesn't contribute.

$$= \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!}$$

let $n = k-1$

$$\sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = \sum_{k-1=0}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = \sum_{n=0}^{\infty} \frac{\mu^n}{n!} = e^{\mu}$$

through taylor's series of $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$

$$= \mu e^{-\mu} e^{\mu}$$

$$= \mu \quad //$$

$$Var(X) = \sum_{k=0}^{\infty} k^2 \frac{e^{-\mu}\mu^k}{k!} - \mu^2$$

$k = k-1+1$

$$= \mu e^{-\mu} \sum_{k=1}^{\infty} k \frac{\mu^{k-1}}{(k-1)!} \quad \mu^{k-1} = \mu^{k-2}\mu \quad - \mu^2$$

$$= \mu e^{-\mu} \left(\sum_{k=2}^{\infty} (k-1) \frac{\mu^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} \right) - \mu^2$$

$k-1=0$, $k=1$: when $k=1$, the whole thing is zero. can ignore. so start from 2.

$$= \mu^2 e^{-\mu} \sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!} + \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} - \mu^2$$

$$= \mu^2 e^{-\mu} e^{\mu} + \mu e^{-\mu} e^{\mu} - \mu^2$$

$$= \mu \quad //$$

4. Geometric Distribution.

$X \sim Geo(p)$

$P(X=k) = p(1-p)^{k-1}$ for $k \in \{0, 1, \dots\}$ (k is number of trials till success)

using k as number of trial not failure! depends on case to case.

$$E(X) = \sum_{k=0}^{\infty} kp(1-p)^{k-1}$$

$$= p \sum_{k=1}^{\infty} k(1-p)^{k-1}$$

looks like: $\frac{d}{dp}[-(1-p)^k]$

$$= p \sum_{k=1}^{\infty} \frac{d}{dp}[-(1-p)^k]$$

$$= -p\frac{d}{dp}\sum_{k=1}^{\infty}(1-p)^k \qquad S_\infty = \frac{a}{1-r} = \frac{1-p}{1-(1-p)} = \frac{1-p}{p}$$

$$= -p\frac{d}{dp}\left(\frac{1-p}{p}\right)$$

$$= -p\left(-\frac{1}{p^2}\right)$$

$$= \frac{1}{p} \;//$$

$$\text{Var}(X) = \sum_{k=0}^{\infty} k^2 p(1-p)^{k-1} - \frac{1}{p^2}$$

$$= p\sum_{k=1}^{\infty} k^2(1-p)^{k-1} - \frac{1}{p^2}$$ ← it is like $\frac{d}{dp}[-k(1-p)^k]$

$$= -p\frac{d}{dp}\sum_{k=1}^{\infty} k(1-p)^k - \frac{1}{p^2}$$ ← this looks like $E(X) = \sum kp(1-p)^{k-1}$

$$= -p\frac{d}{dp}\sum_{k=1}^{\infty} kp(1-p)^{k-1}\frac{(1-p)}{p} - \frac{1}{p^2}$$

bring it out.

$$= -p\frac{d}{dp}\left(\frac{1-p}{p}\left(\frac{1}{p}\right)\right) - \frac{1}{p^2}$$

$$= -p\frac{d}{dp}\left(\frac{1-p}{p^2}\right) - \frac{1}{p^2}$$

$$= -p\frac{d}{dp}\left(p^{-2} - p^{-1}\right) - p^{-2}$$

$$= -p\left(-2p^{-3} + p^{-2}\right) - p^{-2}$$

$$= 2p^{-2} - p^{-1} - p^{-2}$$

$$= p^{-2} - p^{-1}$$

$$= \frac{1}{p^2} - \frac{1}{p}$$

$$= \frac{1-p}{p^2} \;//$$

4.3.2 Continuous Distribution.

$$E(X) = \int_{\text{bies}} xf(x)\,dx \quad , \quad \text{Var}(X) = \int_{\text{bies}} x^2 f(x)\,dx - \left[\int_{\text{bies}} xf(x)\,dx\right]^2$$

C5 · Fitting Probability Distribution

5.1 Method of Moment (1 out of 2 method covered in y2)

Step 1. Think which distribution we want to fit into our sample. ✡ (usually told in exam)

eg. $X \sim B(n,p)$, $X \sim N(\mu, \sigma^2)$, ...

Step 2. Equate population moment (our fitted distribution's) to the unbiased sample moment (the original data sets)

$E(X) = \bar{x}$ (moment 1)

$Var(X) = s_x^2$ (centred moment 2)

$g_1(X) = \hat{g}_1(x)$ (centred moment 3)

⋮ ⋮ ⋮

population (have to memorise) | sample (have to find)

✡ IF NUMBER OF EQUATION AVAILABLE IS GREATER THAN THE NUMBER OF PARAMETER OF OUR FITTED DISTRIBUTION, WE NEED TO DO MINIMISATION WITH WEIGHTAGE INSTEAD

do this instead if no of moment eqn > no of parameters.

✡ if in exams, was not told how many eqn of moment required, choose exactly the same:

$X \sim B(n,p)$: $E(X) = np = \bar{x}$

$Var(X) = np(1-p) = s_x^2$

2 parameters ↔ 2 eqns

Step 3. Minimisation process with weightage.

$$\underset{\sim}{\text{Min}} \sum_{k=1}^{p} w_k (M_k - m_k)^2 = \underset{\sim}{\text{Min}} \{ w_1 (E(X) - \bar{x})^2 + w_2 (Var(X) - s_x^2)^2 + \dots \}$$

M_k: population moments; m_k: sample moments

✡ w (weightage) are usually given in exams.

But if not, smaller moment should have higher weightage.

eg if there is w_1, w_2 and w_3:

$w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$

✡ Why when no of moment eqn = no of parameters is enough but sometimes we use more no of moment eqn? (which is a lot harder to do)

cause the more no of moment eqn we use, the more accurate the fit is!

eg1. (let us use no of moment eqns = no of parameters)

Using the method of moments, fit a uniform distribution to the sample below

$$\{1.0,\ 1.1,\ 1.2,\ 1.1,\ 1.3,\ 1.0,\ 1.7,\ 2.0\}$$

Step 1. Determine the fitted probability distribution

$X \sim U[a,b]$ ← 2 parameters

1st moment : $E(X) = \frac{a+b}{2}$

2nd moment : $Var(X) = \frac{(b-a)^2}{12}$

Step 2. Equate population moment to sample moment.

$\bar{x} = \frac{1.0+1.1+1.2+1.1+1.3+1.0+1.7+2.0}{8} = 1.3$ } 2 eqns of moment

(unbiased) $s_x^2 = 0.3625$

$\frac{a+b}{2} = 1.3 \quad \frac{(b-a)^2}{12} = 0.3625$

$\therefore a = 0.2572,\ b = 2.3428$

$X \sim U[0.2572, 2.3428]$

eg2. (Let us try no of moment eqn > no of parameters)

Using the method of moments, fit a Poisson distribution to the sample below of numbers of monthly nuclear alerts at a power plant:

$$\{0,\ 1,\ 1,\ 0,\ 0,\ 2,\ 0,\ 1\}$$

Step 1. Determine the fitted probability distribution

$X \sim Po(\mu)$

1st moment: $E(X) = \mu$

2nd moment: $Var(X) = \mu$

$\bar{x} = \frac{5}{8}$, $s_x^2 = 0.5536$

WE SKIP STEP TWO CAUSE:

NO OF MOMENT EQN > NO OF PARAMETERS!

Step 3. Minimisation process with weightage.
(for this question, we use $w_1 = 1$, $w_2 = 1$)

let $M(\mu) = 1(\mu - \frac{5}{8})^2 + 1(\mu - 0.5536)^2$

to find min of M: $\frac{dM}{d\mu} = 0 \left(\frac{d^2M}{d\mu^2} > 0\right)$

$\frac{dM}{d\mu} = 2(\mu - \frac{5}{8}) + 2(\mu - 0.5536)$

$0 = 2(\mu - \frac{5}{8}) + 2(\mu - 0.5536)$

$\therefore \mu = 0.585 \qquad X \sim Po(0.585)$

(just to check $\mu = 0.585$ will result is M as minimum and not maximum:
$\frac{d^2M}{d\mu^2} = 2 + 2 = 4 > 0$ ✓)

eg3. (Let us use 3 eqn of moments) ✡ IGNORE THIS QUESTION not in year 2 stats!
Fit, with method of moment: (binomial)

[2, 1, 3, 2, 4, 0, 2, 3, 1, 2]

$E(X) = np$ $\qquad \bar{x} = 2$

$Var(X) = np(1-p)$ $\qquad s_x^2 = \frac{12}{9}$

$g_1(X) = \frac{1-2p}{\sqrt{np(1-p)}}$ $\qquad \hat{g}_1(x) = 0$

let $S(n,p) = w_1(np - \bar{x})^2 + w_2(np(1-p) - s^2)^2 + w_3\left(\frac{1-2p}{\sqrt{np(1-p)}} - \hat{g}_1(x)\right)^2$
let $w_1, w_2, w_3 = 1$ (for simplicity)
to minimise a multivariable function, it is hard (partial derivative)
(alt way: test every possible n and p):

$\therefore n = 5$, $p = 0.4$ will result in minimal S

$\therefore X \sim B(5, 0.4)$

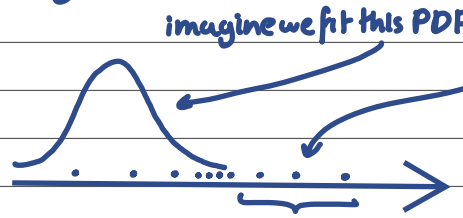
5.2 Maximum Likelihood Method (2 out of 2 method covered in y2)

basically just: $\underset{\tilde{\Theta}}{Max}\left\{\mathcal{L}(\tilde{\Theta})=\prod_{i=1}^{m} f_x(x_i|\tilde{\Theta})\right\}$ (continuous)

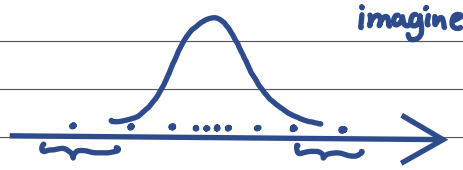
$$\underset{\tilde{\Theta}}{Max}\left\{\mathcal{L}(\tilde{\Theta})=\prod_{i=1}^{m} p_x(x_i|\tilde{\Theta})\right\} \quad \text{(discrete)}$$

looks complicated?
Basically it means, we want to find the parameter of a fitted distribution (eg. for $X\sim B(n,p)$, $\tilde{\Theta}$ is n and p) that will result a maximum $\mathcal{L}(n,p)$:
$\mathcal{L}(n,p) = p(X=x_1)\times p(X=x_2)\times p(X=x_3)\times\dots$

Why?



probability getting these dots are so small,
when multiply all the probability you can feel that
the $\mathcal{L}(\Theta)$ is very small!



although the probability on the sides are still small,
but the majority is in the center, which results in higher probability!

So, now to find $\tilde{\Theta}$ that correspond to $\mathcal{L}(\Theta)$ maximum?
if one parameter only, if more than one parameter,

$\frac{d\mathcal{L}(\tilde{\Theta})}{d\tilde{\Theta}}=0$, $\frac{d^2\mathcal{L}(\tilde{\Theta})}{d\tilde{\Theta}}<0$ (maxima) $\frac{\partial\mathcal{L}(\tilde{\Theta})}{\partial\tilde{\Theta}}=0$ for every $\tilde{\Theta}$, $\frac{\partial^2\mathcal{L}(\Theta)}{\partial\tilde{\Theta}}<0$

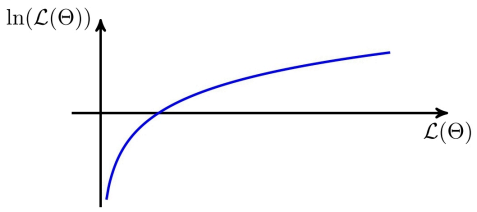
However, $\mathcal{L}(\Theta)=\prod_{i=1}^{n} f_x(x_i|\tilde{\Theta})$ is hard to be differentiated (product rule too many products)

hence we find $\underset{\tilde{\Theta}}{Max}\left\{\ln\mathcal{L}(\Theta)=\ln\left(\prod_{i=1}^{m} f_x(x_i|\tilde{\Theta})\right)=\sum_{i=1}^{m}\ln(f_x(x_i|\tilde{\Theta}))\right\}$

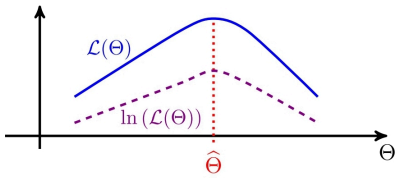
how does product become sum? $\ln(ABC)=\ln A+\ln B+\ln C$

we can do that cause $\ln(\mathcal{L}(\Theta))$ and $\mathcal{L}(\Theta)$ shares the same maximum!

Log is a monotonically increasing function



This means that the *log* likelihood is maximum at the same point as the likelihood



eg1. (I didn't do $\sum \ln(f_x(x|\Theta))$ cause exponent are easy to multiply)

Using the maximum likelihood method, fit a Poisson distribution to the sample below of numbers of monthly nuclear alerts at a power plant:

$$\{0,\ 1,\ 1,\ 0,\ 0,\ 2,\ 0,\ 1\}$$

$X \sim Po(\mu)$ $\quad \underset{\Theta}{\text{Max}} \left\{ \mathcal{L}(\Theta) = \prod_{i=1}^{n} p_x(x_i|\Theta) \right\}$ (discrete)

$p_x(x_i|\mu) = \frac{\mu^x e^{-\mu}}{x}$, $\mathcal{L}(\mu) = \prod p_x(x_i|\mu)$

inserting the datasets from sample,

$$\mathcal{L}(\mu) = P(X=0)P(X=1)P(X=1)P(X=0)P(X=0)P(X=2)P(X=0)P(X=1)$$
$$= P(X=0)^4 P(X=1)^3 P(X=2)^1$$
$$= \left(\frac{\mu^0 e^{-\mu}}{0!}\right)^4 \times \left(\frac{\mu^1 e^{-1}}{1!}\right)^3 \times \left(\frac{\mu^2 e^{-2}}{2!}\right)^1$$
$$= \frac{1}{2}\mu^5 e^{-8\mu}$$

to find max of $\mathcal{L}(\mu)$:

$$\frac{d\mathcal{L}(\mu)}{d\mu} = \frac{1}{2}(5\mu^4 e^{-8\mu} - 8\mu^5 e^{-8\mu})$$
$$0 = \frac{1}{2} e^{-8\mu}(5\mu^4 - 8\mu^5)$$
$$\mu = \frac{5}{8}$$

$\therefore X \sim Po\left(\frac{5}{8}\right)$

$\frac{d^2\mathcal{L}(\mu)}{d\mu^2} < 0$ (check passed)

2021q2c(ii)

☆ Usually max likelihood method in this course will only ask for distribution that has one unknown parameter:
eg. $X \sim Exp(\lambda)$, $X \sim Po(\mu)$, $X \sim B(50, p)$ given.

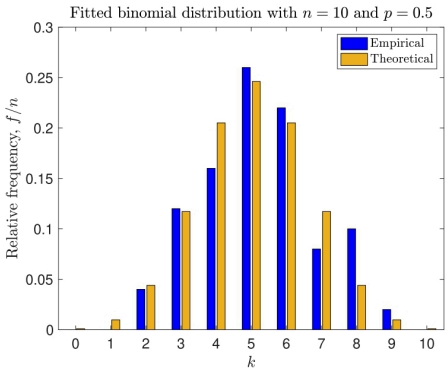
but there is an exception!
$X \sim U(a, b)$, width $= b - a$
$\mathcal{L}(w) = \prod \frac{1}{w}$
if width = range of sample, $\mathcal{L}(w) \to$ max!

5.3 Goodness of Fit

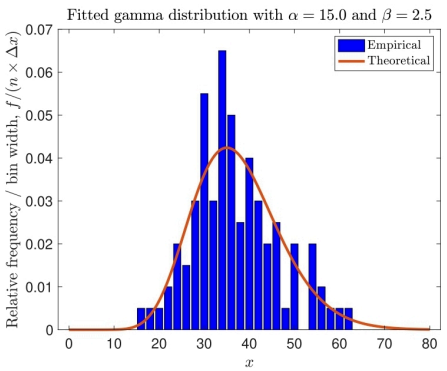
After fitting a distribution onto a sample, we should check how "good" of a fit the distribution is.
3 methods to do so:
1. Compare sample's histogram to distribution's PDF (PMF if discrete)
2. Compare sample's CDF to distribution's CDF
3. Create a Quantile-Quantile Plot (QQ plot)

1. Compare sample's histogram to distribution's PDF (PMF if discrete)

Discrete X: use relative frequency



Continuous X: use relative frequency divided by bin width

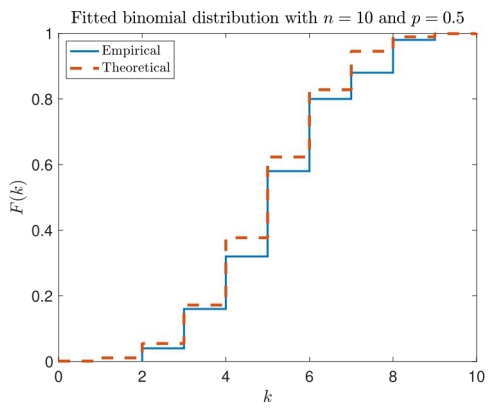


empirical means sample's, theoretical means fitted population distribution

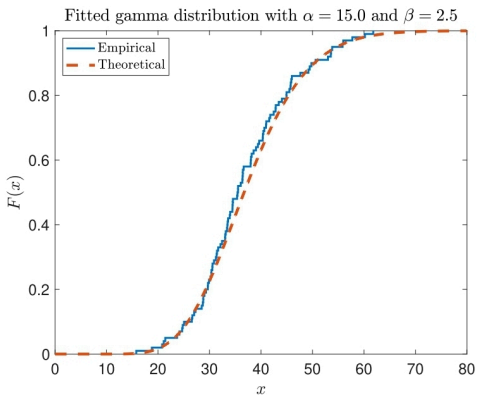
histogram with PMF (discrete); histogram with PDF (continuous)

2. Compare sample's CDF to distribution's CDF

Discrete distribution



Continuous distribution



empirical means sample's, theoretical means fitted population distribution

3. Create a Quantile-Quantile Plot (QQ plot)

Having a QQ Plot let us able to compare quantile of the sample and the quantile of the fitted distribution (and check if they match)

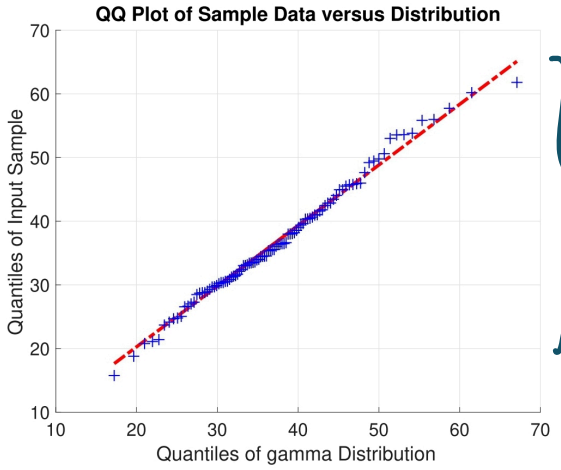
To plot QQ plot:

- vertical axis (y-coordinate) is just your sample, sorted from first to last.
- horizontal axis (x-coordinate), use:

$$qp\left(\frac{p}{100}\right) = F^{-1}\left(\frac{p}{100}\right), \text{ where } \frac{p}{100} = \frac{j-0.5}{n}, \text{ for } j = 1, 2, 3, \ldots, n$$

so we need to find CDF (of the fitted distribution), then inverse it

Continuous X: Gamma distribution



if it assembles a straight line, it is a good fit.

C6. Multiple Variables

→ This chapter have a lot of info to memorise, Hence, I'll highlight all the part that should be memorise. (Other part that is not highlight is the explaination)

6.1 Covariance and Correlation

6.1.1 Covariance

$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$ ✱ $Cov(X,X) = E[XX] - E[X]E[X]$

$= E[X^2] - E[X]^2$

$= Var(X)$

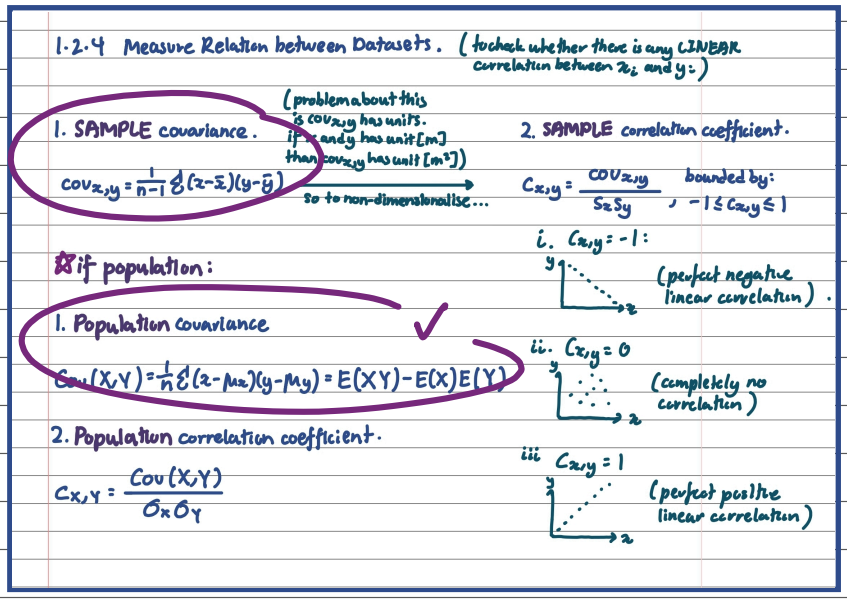
$Cov(X,Y) = E[XY] - E[X]E[Y]$ ← used to prove

$Var(X+Y) = Cov(X+Y, X+Y)$

$= Cov(X,X) + Cov(X,Y) + Cov(X,Y) + Cov(Y,Y)$

$= Var(X) + 2Cov(X,Y) + Var(Y)$

} using the fact that: $Cov(X,X) = Var(X)$



1.2.4 Measure Relation between Datasets. (to check whether there is any LINEAR correlation between x_i and y_i)

1. SAMPLE covariance.

$cov_{x,y} = \frac{1}{n-1}\sum(x-\bar{x})(y-\bar{y})$

(problem about this is $cov_{x,y}$ has units. If x and y has unit [m] then $cov_{x,y}$ has unit [m^2]) → so to non-dimensionalise...

2. SAMPLE correlation coefficient.

$C_{x,y} = \frac{cov_{x,y}}{s_x s_y}$, bounded by: $-1 \le C_{x,y} \le 1$

if population:

1. Population covariance ✓

$Cov(X,Y) = \frac{1}{n}\sum(x-\mu_x)(y-\mu_y) = E(XY) - E(X)E(Y)$

2. Population correlation coefficient.

$C_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

i. $C_{x,y} = -1$: (perfect negative linear correlation)

ii. $C_{x,y} = 0$ (completely no correlation)

iii. $C_{x,y} = 1$ (perfect positive linear correlation)

in c1.2.4, we looked at both sample covariance and population covariance. In this chapter we only look at population covariance as X and Y are random variables.

6.1.2 Properties of Covariance.

1. $Cov(cX,Y) = cCov(X,Y)$
$Cov(X,cY) = cCov(X,Y)$ ← constant can be brought out!

} $Cov(aX,bY) = abCov(X,Y)$

2. $Cov(X,Y) = Cov(Y,X)$ ← random variable can be swapped! (known as symmetric property)

3. $Cov(X,c) = 0$
$Cov(c,Y) = 0$
} if any random variable is 'Cov' with a constant. the Covariance is 0!

4. $Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$ ← don't remember the formula!

looks complicated, but just remember: $Cov(aX_1 + bX_2, cY_1 + dY_2)$ (do step by step)

$= Cov(aX_1, cY_1) + Cov(aX_1, dY_2) + Cov(bX_2, cY_1) + Cov(bX_1, dY_2)$

$= acCov(X_1, Y_1) + adCov(X_1, Y_2) + bcCov(X_2, Y_1) + bdCov(X_1, Y_2)$

COV is not the same as Cov.
COV is Covariance MATRIX and Cov is just Covariance (a number)

not as important as the first 4. (can skip)

5. $COV(X_1, X_2, X_3, \ldots, X_n)$

$$= \begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Var(X_2) & & Cov(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & Var(X_n) \end{pmatrix}$$

things to take note: 1. $Cov(X_1,X_2) = Cov(X_2,X_1)$, hence COV is symmetric about the main diagonal.

2. The main diagonal are all $Cov(X,X) = Var(X)$!

eg.

Let X and Y be two independent $N(0,1)$ random variables and

$$Z = 1 + X + XY^2,$$
$$W = 1 + X.$$

Find $Cov(Z,W)$.

$Cov(Z,W) = Cov(1+X+XY^2, 1+X)$

$= Cov(1,1) + Cov(1,X) + Cov(X,1) + Cov(X,X) + Cov(XY^2,1) + Cov(XY^2,X)$

$Cov(1,1)$: 0; $Cov(1,X)$: 0; $Cov(X,1)$: 0; $Cov(XY^2,1)$: 0

$\{ Cov(X,X) = Var(X)$

$= Var(X) + Cov(XY^2, X)$

after simplifying with all covariant property, we use the covariance formula:

$Cov(X,Y) = E[XY] - E[X]E[Y]$

$= Var(X) + E(XY^2X) - E(XY^2)E(X)$

$= Var(X) + E(X^2Y^2) - E(XY^2)E(X)$

now we finish simplifying into only expectation and variance,

we can now use the fact that $X \sim N(0,1)$, $Y \sim N(0,1)$ ← from question.

$E(X) = 0$ $E(Y) = 0$

$Var(X) = 1$ $Var(Y) = 1$

... and also independent means:

$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ ← will be covered later in C6.

$= Var(X) + E(X^2)E(Y^2) - E(X)E(Y^2)E(X)$

$= 1 + 1 \times 1 + 0 \times 1 \times 0$

$= 2$

$Var(Y) = E(Y^2) - E(Y)^2$

$E(Y^2) = Var(Y) + E(Y)^2$

$= 1 + 0^2$

$= 1$

$E(X^2) = Var(X) + E(X)^2$

$= 1$

→ ☆only measures LINEAR correlation!

6.1.3 Correlation Coefficient.

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

1.2.4 Measure Relation between Datasets. (to check whether there is any LINEAR correlation between x_i and y_i)

1. SAMPLE covariance. (problem about this is $cov_{x,y}$ has units. if x and y has unit [m] than $cov_{x,y}$ has unit [m^2]) so to non-dimensionalise...

$$cov_{x,y} = \frac{1}{n-1}\sum(x-\bar{x})(y-\bar{y})$$

2. SAMPLE correlation coefficient.

$$C_{x,y} = \frac{cov_{x,y}}{S_x S_y}, \text{ bounded by: } -1 \le C_{x,y} \le 1$$

☆if population:

1. Population covariance

$$Cov(X,Y) = \frac{1}{n}\sum(x-\mu_x)(y-\mu_y) = E(XY) - E(X)E(Y)$$

2. Population correlation coefficient.

$$C_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

i. $C_{x,y} = -1$: (perfect negative linear correlation)

ii. $C_{x,y} = 0$ (completely no correlation)

iii. $C_{x,y} = 1$ (perfect positive linear correlation)

in c1.2.4, we looked at both sample correlation coefficient and population correlation coefficient. In this chapter we only look at population correlation coeff. X and Y are random variables.

6.1.4 Properties of Correlation Coefficient

1. $-1 \le \rho(X,Y) \le 1$ (correlation coeff. is always between -1 and 1)

2. $\rho(aX+b, cY+d) = \rho(X,Y)$ (multiply random variable with scalar or addition with constant has NO effect)

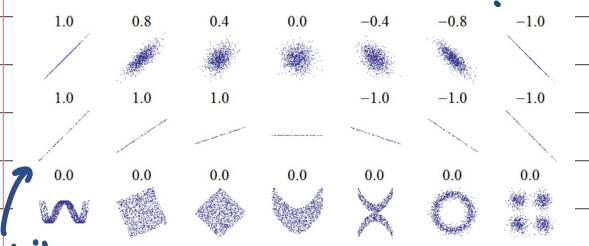
3. If $\rho(X,Y) = 0$, uncorrelated

If $\rho(X,Y) > 0$, positively correlated

If $\rho(X,Y) < 0$, negatively correlated.

If $\rho(X,Y) = 1$, $Y = aX+b$, $a>0$

If $\rho(X,Y) = -1$, $Y = aX+b$, $a<0$



note that it doesn't depend on slope

these are example of no correlation

☆ uncorrelated is not same as independent!
if independent → uncorrelated ✓
but if uncorrelated, it doesn't mean it is independent!
(there's some special case like if both random variable are normally distributed)

$\rho(X,Y) = 0$ means uncorrelated but doesn't means independent!

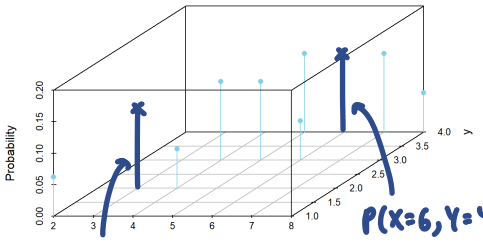
any property independent have → uncorrelated have (but not the other way)
eg. $Var(X+Y) = Var(X) + Var(Y)$
this is true if X and Y are independent,
if independent → uncorrelated.
hence it is true if X and Y are uncorrelated!

6.2 Joint Distribution

It is just probability distribution in 2D! PMF for discrete usually given in table; PDF for continuous usually just given as a function $f(x,y)$

Joint Distribution

Discrete



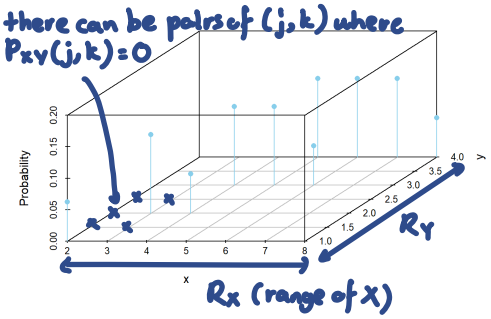
$P(X=3, Y=2)$

$P(X=6, Y=4)$ or $P_{XY}(6,4)$

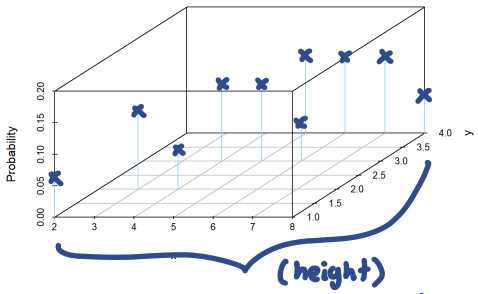
no need to memorise

Joint PMF: $P_{XY}(j,k) = P(X=j, Y=k)$

there can be pairs of (j,k) where $P_{XY}(j,k)=0$



we usually define range as a rectangle, $R_X \times R_Y$ in size.

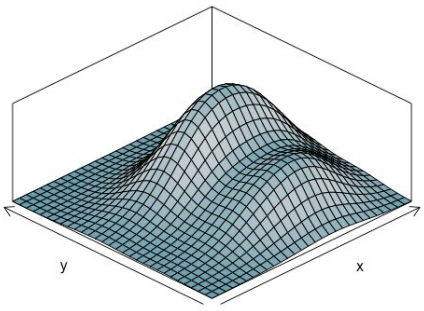


summing all the probability in the domain

$$\sum_{(j,k)\in R_{XY}} P_{XY}(j,k) = 1$$

$$\sum_{k\in R_Y} \sum_{j\in R_X} P_{XY}(j,k) = 1$$

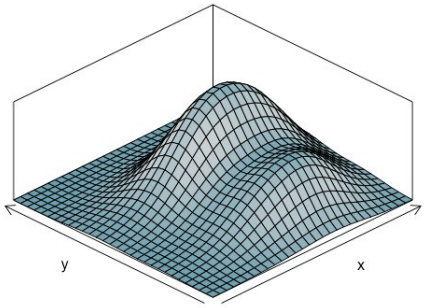
Continuous



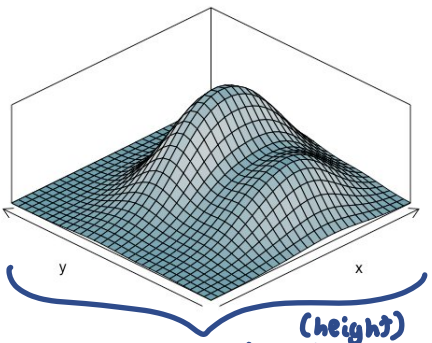
Joint PDF:

$$f_{XY}(x,y) = \frac{P(x \le X \le x+dx, y \le Y \le y+dy)}{dx\,dy}$$

no need to memorise.



continuous, range is from $-\infty$ to ∞ for both X and Y. (And can be trimmed into and shape, eg. Range: $x^2+y^2 < 9$)



Summing all the probability

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

Joint Distribution

Discrete

	$Y=0$	$Y=1$	$Y=2$
$X=0$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$
$X=1$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$

PMF usually expressed in a table

Continuous

$$f_{XY}(x,y) = \begin{cases} c(x^2+y^2) & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

PDF usually expressed as a multivariable function $f(x,y)$

for CDF:

$$F_{XY}(j,k) = P(X \le j, Y \le k)$$

for CDF: remember CDF ∫∫

$$F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(u,v)\,du\,dv$$

u and v are just dummy variables

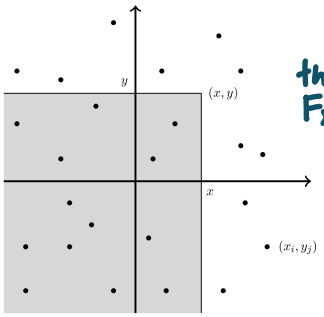
this also means that differentiating CDF twice can get PDF too:

$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$$

important results:

$0 \le F_{XY}(x,y) \le 1$

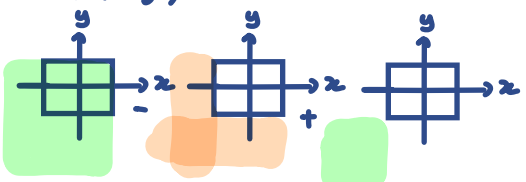
$F_{XY}(\infty,\infty) = 1$
$F_{XY}(-\infty,y) = 0$ for any y
$F_{XY}(x,-\infty) = 0$ for any x



this is $F_{XY}(x,y)$

$$P(x_1 \le X \le x_2, y_1 \le Y \le y_2)$$
$$= F_{XY}(x_2,y_2) - F_{XY}(x_1,y_2) - F_{XY}(x_2,y_1) + F_{XY}(x_1,y_1)$$

DO NOT MEMORISE understand it.



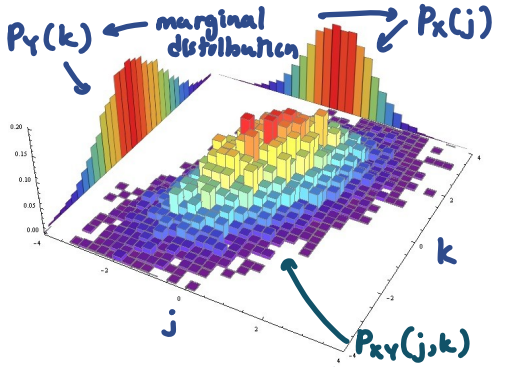
$F_{XY}(x_1,y_1)$ is minused twice → so have to add one back

6.3 Marginal Distribution

Marginal Distribution

basically it's when the distributions of the random variable considered seperately

Discrete



$$P_X(j) = \sum_{k=0}^{\infty} P_{XY}(j,k)$$

for each j, sum all the k

eg. $P_X(1) = \sum_{k=0}^{\infty} P_{XY}(1,k)$

$P_X(2) = \dots$

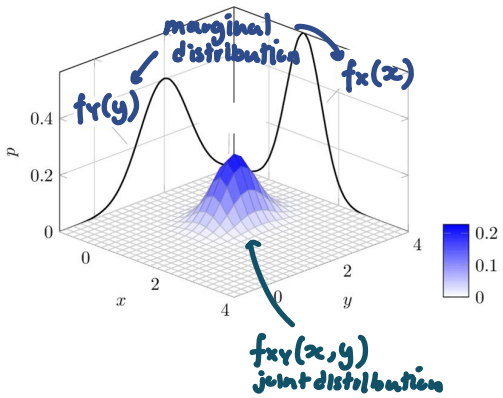
⋮

all of these put in a table and you get the distribution.

similarly,

$$P_Y(j) = \sum_{j=0}^{\infty} P_{XY}(j,k)$$

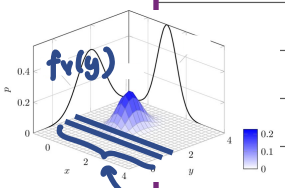
Continuous



$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$



integrate all these small strip so dx.

☆ FIRST WAY TO PROOF INDEPENDENT :

IF DISCRETE:

$P_{XY}(j,k) = P_X(j)P_Y(k)$

joint distribution. marginal dist.

IF CONTINUOUS :

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

joint dist. marginal dist.

instead of PDF, if continuous we can use CDF :

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

6.4 Conditional Distribution.

Conditional Distribution

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \xrightarrow{\text{2 vars}} P(X\in C|Y\in D)=\frac{P(X\in C, Y\in D)}{P(Y\in D)}$$

C, D are range of X and Y

Discrete	Continuous
$P_{X\|Y}(j,k)=P(X=j\|Y=k)=\frac{P(X=j,Y=k)}{P(Y=k)}$	$f_{X\|Y}(x\|y)=\frac{f_{XY}(x,y)}{f_Y(y)}$

$\therefore$ basically, Conditional $=\frac{\text{Joint}}{\text{Marginal}}$

SECOND WAY TO PROOF INDEPENDENT:

IF DISCRETE:	IF CONTINUOUS:
$P_{X\|Y}(x\|y)=P_X(x)$	$f_{X\|Y}(x\|y)=f_X(x)$
$P_{Y\|X}(y\|x)=P_Y(y)$	$f_{Y\|X}(y\|x)=f_Y(y)$

recall that this is similar to:

$P(A|B)=P(A)$ proof that A and B are independent event.

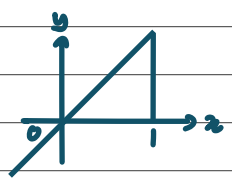
eg. Find CDF of X given PDF of XY: $f_{XY}=\begin{cases} xy, & 0\le y\le x\le 1 \\ 0, & \text{otherwise}\end{cases}$

step 1. Find PDF of X, $f_X(x)$

$$f_X(x)=\int_{\text{by}\in R_Y} f_{XY}(x,y)\,dy$$

$$=\int_{y_0=0}^{y_1=x} xy\,dy$$

$$=\frac{1}{2}x^3$$



step 2: using $f_X(x)$, find $F_X(x)$

$$F_X(x)=\int_0^x\left(\int_v^x \frac{1}{2}u^3\,du\right)dv$$

$$=\int_0^x\left[\frac{1}{8}u^4\right]_v^x dv$$

$$=\frac{1}{8}\int_0^x x^4-v^4\,dv$$

$$=\frac{1}{8}\left[x^4v-\frac{1}{5}v^5\right]_0^x$$

$$=\frac{1}{8}\left(\frac{4}{5}x^5\right)=\frac{1}{10}x^5$$

6.5 Important Identity

1. Conditional Expectation:

$$E[X|Y=y_j] = \sum_{x_i \in R_X} x_i P_{X|Y}(x_i|y_j)$$ from: $E(X) = \sum_{x_i \in R_X} x_i P(X=x_i)$

$X \to X|Y=y_j$, $x=x_i \to x=x_i|Y=y_j$

2. Conditional Variance

$$Var(X|Y=y) = E[X^2|Y=y] - E[X|Y=y]^2$$ from: $Var(X) = E(X^2) - E(X)^2$

$X \to X|Y=y$, $X^2 \to X^2|Y=y$

3. Law of Total Probability.

$$P_X(x) = \sum_{y_j \in R_Y} P_{XY}(x,y_j) = \sum_{y_j \in R_Y} P_{X|Y}(x|y_j) P_Y(y_j)$$ from: $P(A) = \sum_{B_i} P(A \cap B_i) = \sum_{B_i} P(A|B_i) P(B_i)$

$A \to X=x$, $B_i \to Y=y_j$

4. Law of Total Expectation.

$$E[X] = \sum_{y_j \in R_Y} E[X|Y=y_j] P_Y(y_j)$$

$P_X(x) = \sum P_{X|Y}(x|y_j) P_Y(y_j)$ ← no change.

$P_X(x) \to E[X]$, $P_{X|Y}(x|y_j) \to E[X|Y=y_j]$

5. Law of Total Variance

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$$

☆ 6. Law of the Unconscious Statistician.

$$E[g(X,Y)] = \sum_{y_j \in R_Y} \sum_{x_i \in R_X} g(x_i,y_j) P_{XY}(x_i,y_j)$$ from: $E[g(X)] = \sum g(x) P(X=x)$

$g(X) \to g(X,Y)$, $g(x) \to g(x,y)$, $P(X=x) \to P(X=x, Y=y)$

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\, dx\, dy$$

7. Law of Iterated Expectation.

$$E[X] = E[E[X|Y]]$$ ← This is kinda useful when you can find $f_{X|Y}$ but not f_X

☆ CHANGE PMF to PDF, and SUMMATION to INTEGRAL

you will get continuous version of all the laws.

eg. $E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$

If X and Y are INDEPENDENT variables,

$E[XY] = E(X)E(Y)$; the converse is ~~not true~~ not always true.

→ E[XY] = E[X]E[Y] does not imply independence. (BUT IT DOES IMPLY UNCORRELATED!)

$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

$E[X|Y] = EX$;

$E[g(X)|Y] = E[g(X)]$.

✡ all these are true for uncorrelated variables as independent → uncorrelated!

Let $X_1, X_2, .., X_i$ be independent random variables and also independent of N. Let

$$Y = \sum_{i=1}^{N} X_i.$$

We have

$$E[Y] = E[X]E[N]$$
$$Var(Y) = E[N]Var(X) + (E[X])^2 Var(N)$$

} just memorise, quite hard.

6.6 Sum of Two Random Variables.

6.1 Method of Convolution

If we have two distributions, eg. $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ and we want to find the distribution of $X+Y$, eg. $X+Y \sim N(....)$ (more specific case check 6.6.3)

we need to first find the PDF of $X+Y$: ✡ remember, PDF of $X+Y$ is NOT just $f_x(x) + f_y(y)$!

let $Z = X+Y$, ✡ w is just a dummy variable!

If don't know X and Y are independent or not:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(w, z-w)\,dw \quad \text{or} \quad f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z-w, w)\,dw$$

If know X and Y are independent:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\,dw \quad \text{or} \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(z-w) f_Y(w)\,dw$$

✡ given in the data sheets:

Assuming independence of the added distributions:

$$K = I + J \Rightarrow p_K(k) = \sum_{j=0}^{k} p_I(k-j)p_J(j); \quad Z = X + Y \Rightarrow f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy$$

discrete cases

eg. $X \sim Exp(\lambda)$, find the distribution of $2X$ (assume independent)

Step 1. Find the PDF of $2X$:

$$Z = X + Y \Rightarrow f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy$$

let $Z = 2X$,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-w) f_X(w)\, dw$$

we know that if $X \sim Exp(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x}$$

☆ change limit as $X \sim Exp(\lambda)$, $f_X(x) = \lambda e^{-\lambda x}$ is defined for $x \geq 0$

$z - w \geq 0$; $w \geq 0$

$w \leq z$; $w \geq 0$

$$= \int_0^z \lambda e^{-\lambda(z-w)} \cdot \lambda e^{-\lambda w}\, dw$$

$$= \lambda^2 \int_0^z e^{-\lambda z + \lambda w} \cdot e^{-\lambda w}\, dw$$

$$= \lambda^2 \int_0^z e^{-\lambda z}\, dw$$

$$= \lambda^2 \left[w e^{-\lambda z} \right]_0^z = \lambda^2 z e^{-\lambda z}$$

3. Gamma Distribution.

$X \sim Gamma(\alpha, \beta)$ (α and β are shape and scale parameters)

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$E(X) = \alpha\beta$; $Var(X) = \alpha\beta^2$

$\beta = \frac{1}{\lambda}$, $\alpha = 2$: $f(x) = \frac{x^{2-1} e^{-x/1/\lambda}}{\frac{1}{\lambda^2}(2-1)!} = \lambda^2 x e^{-\lambda x}$

$\therefore 2X \sim Gamma(2, \frac{1}{\lambda})$

6.6.2 Moments.

X and Y doesn't have to be independent for this to be true.

$$E[aX+bY+c] = aE[X] + bE[Y] + c$$

$$Var(aX+bY+c) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$$

if only X and Y are independent, $Cov(X,Y) = 0$ and hence:

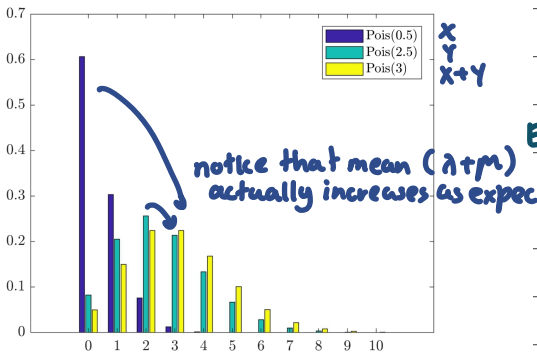
$$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y)$$

6.6.3 Distributions.

1. Sum of Poisson Variables

If $I \sim Po(\lambda)$, $J \sim Po(\mu)$ are two independent discrete variables, then:

☆ $I + J \sim Po(\lambda + \mu)$



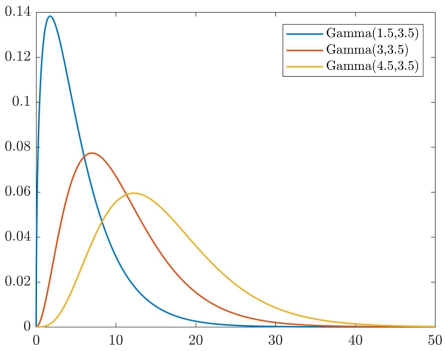
notice that mean ($\lambda+\mu$) actually increases as expected.

$E(X) = \lambda + \mu$

2. Sum of Gamma Variables

If $X \sim \text{Gamma}(d_1, \beta)$ and $Y \sim \text{Gamma}(d_2, \beta)$ are two independent variables, then:

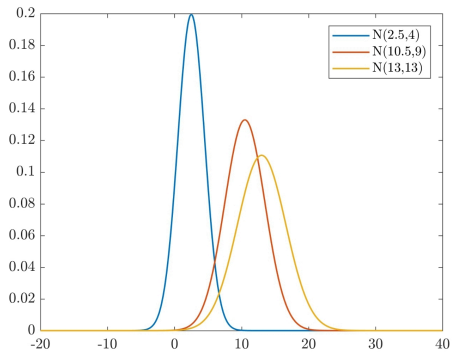
★ $X+Y \sim \text{Gamma}(d_1+d_2, \beta)$



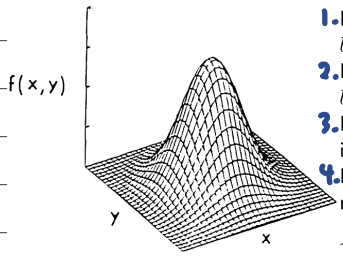
3. Sum of Normal Variables

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are two independent variables, then:

★ $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$



6.7 Bivariate Normal Distribution (Jointly Normal with 2 independent vars.)



1. If X and Y are bivariate normal, then by letting $a = 1$, $b = 0$, we conclude X must be normal.
2. If X and Y are bivariate normal, then by letting $a = 0$, $b = 1$, we conclude Y must be normal.
3. If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent, then they are jointly normal.
4. If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are jointly normal, then

$$X+Y \sim N\left(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2+2\rho(X,Y)\sigma_X\sigma_Y\right).$$

properties of bivariate normal) NOT IMPORTANT.

1 and 2 basically just says X and Y are normal.

3 says X and Y are independent.

★ $$X+Y \sim N\left(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2+2\rho(X,Y)\sigma_X\sigma_Y\right).$$

This one is important! Memorise this instead of next page.

$aX+bY \sim N\left(E(aX+bY), \text{Var}(aX+bY)\right)$

$\sim N\left(aE(X)+bE(Y), a^2\text{Var}(X)+b^2\text{Var}(Y)+2ab\,\text{Cov}(X,Y)\right)$

1. Bivariate Normal Distribution. ($\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y > 0$, $\rho \in (-1,1)$ and $\rho \neq 0$)

Definition: Two random variables X and Y are said to have a **bivariate normal distribution** with parameters μ_X, σ_X^2, μ_Y, σ_Y^2, and ρ, if their joint PDF is given by

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

where $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y > 0$ and $\rho \in (-1,1)$ are all constants.

Don't memorise! not important.

{ set $\mu_X, \mu_Y = 0$ and $\sigma_X, \sigma_Y = 1$ } $X \sim N(0,1)$, $Y \sim N(0,1)$

2. Standard Bivariate Normal Distribution with Correlation Coefficient

Two random variables X and Y are said to have the **standard bivariate normal distribution with correlation coefficient** ρ if their joint PDF is given by

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}[x^2 - 2\rho xy + y^2]\}$$

where $\rho \in (-1,1)$.

{ set $\rho = 0$

3. Standard Bivariate Normal Distribution

If $\rho = 0$, then we just say that X and Y have the standard bivariate normal distribution.

$f_{XY}(x,y) = \frac{1}{2\pi} \exp\{-\frac{1}{2}(x^2+y^2)\}$

Useful theorems:

1. Let X and Y be two bivariate normal random variables. Then there exist independent standard normal random variables Z_1 and Z_2 such that

$$\begin{cases} X &= \sigma_X Z_1 + \mu_X \\ Y &= \sigma_Y(\rho Z_1 + \sqrt{1-\rho^2} Z_2) + \mu_Y \end{cases}$$

2. Suppose X and Y are jointly normal random variables with parameters μ_X, σ_X^2, μ_Y, σ_Y^2, and ρ. Then, given $X = x$, Y is normally distributed with

$$E[Y|X=x] = \mu_Y + \rho\sigma_Y \frac{x-\mu_X}{\sigma_X},$$
$$Var(Y|X=x) = (1-\rho^2)\sigma_Y^2.$$

3. If X and Y are bivariate normal and uncorrelated, then they are independent.

not important can skip

☆ we always say if independent then it is correlated and NOT THE OTHER WAY but...

THIS IS A SPECIAL CASE.
only bivariate normal and uncorrelated does mean independence!

C7 Confidence Interval

(No notes on t-distribution and χ^2-distribution cause we just need to know when to use it and how, and not the details of the distributions)

7.0 Pre-requisites.

Before we start with this chapter, let's be clear about samples and population.

Population is the entire data that you are interested / studying.
Samples is a subset of population that you are interested / studying.

In simple term, population is the "entire cake" and sample is just a few slices of it. (The slices of cakes, a.k.a the sample, can be used to make inference about the whole cake, a.k.a the population)

☆ This chapter, is mainly about relating samples to population... which is very USEFUL! (we can't always know about the population, eg: you can't measure the average height of human population in the entire world, but you can measure some of it and make an inference about the entire population!)

7.1 Introduction to $E(\bar{x})$ and $Var(\bar{x})$

$X_1, X_2, X_3, \ldots, X_n$ are INDIVIDUAL OBSERVATIONS!
eg. height of one human from our sample.
and each height, are independent from one another
i.e. my height does not depends on yours.

$E(X_i) = E(X)$ since they are taken from population

if this is too complicated, can skip this part as it is not important

for expectation, $E(\bar{x})$:

$$\bar{x} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$E(\bar{x}) = \frac{E(X_1) + E(X_2) + \ldots + E(X_n)}{n} = \frac{E(X) + E(X) + \ldots + E(X)}{n} = \frac{nE(X)}{n} = E(X)$$

this 'n' is kind of confusing. It's number of individual observation from ONE sample and not the number of sample.

$$\boxed{E(\bar{x}) = E(X)}$$

☆ WHAT IS THE MEANING OF THIS? ($E(\bar{x})$)
This says that the mean, of mean of a sample is equal to the mean of population! ← $E(X)$

Same for variance:

$$Var(\bar{x}) = \frac{Var(X_1 + X_2 + \ldots + X_n)}{n^2} = \frac{Var(X_1) + Var(X_2) + \ldots + Var(X_n)}{n^2} = \frac{nVar(X)}{n^2} = \frac{Var(X)}{n}$$

remember these

$$\boxed{Var(\bar{x}) = \frac{Var(X)}{n}}$$

← This says that the variance, of the mean of a sample is equal to the variance of the population divided by n.
($Var(\bar{x})$ ↙ $Var(X)$)

7.2 Limit Theorems.

as the name suggest, 'limit' theorems are theorem when $\lim_{n\to\infty}$, where sample size approaches infinity, which is the population!

1. Law of Large Numbers.

not really important BUT it is beneficial to understand

Theorem: Weak Law of Large Numbers
Let $X_1, X_2, \ldots, X_n$ be iid random variables with a finite expected value $E(X_i) = \mu$. Then, for any $\epsilon > 0$

$$\lim_{n\to+\infty} P(|\bar{x}-\mu| \geq \varepsilon) = 0$$

"limit as n approaches ∞ of probability of sample mean minus population mean is greater than any positive numbers is 0"

in simple terms, the difference, between sample mean, and population mean is going to be equal to zero, WHEN the sample size, n approaches ∞

"$\bar{x} \longrightarrow \mu$ if $n \to \infty$"

2. Central Limit Theorem. (distribution of sample mean is ALWAYS NORMAL if sample size is big)

(ONE SAMPLE'S MEAN DISTRIBUTION)

don't even bother reading this, I put it as aesthetic lol

Definition:
Let $X_1, X_2, ..., X_n$ be **iid random variables** with expected value $E(X_i) = \mu < \infty$ and variance $0 < Var(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + ... + X_n - n\mu}{\sqrt{n\sigma^2}}$$

converges to the standard normal random variable:

$$\lim_{n\to+\infty} P(Z_n \leq x) = \Phi(x), \quad \text{for all } x \in R$$

where $\Phi(x)$ is the standard normal CDF.

if sample size of each sample, n is big enough (around 30) then the sample distribution will approximate be a normal distribution with population mean and population variance.

basically it just means:

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$$

VERY IMPORTANT

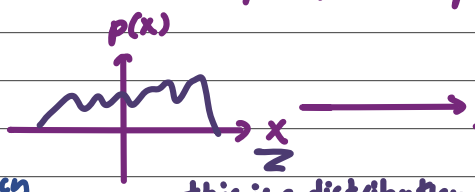
after this, we can just treat it as a simple normal distribution question from c4.

like we can standardized it to make finding probability easier where:

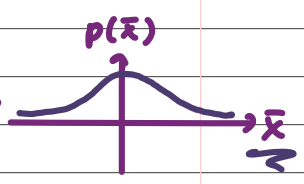
$$z = \frac{X - E(X)}{\sqrt{Var(X)}} \quad X \to \bar{x},\ E(X) \to \mu,\ Var(X) \to \sigma^2/n$$

$$= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

what is $\bar{x} \sim N(\mu, \sigma^2)$ or sample distribution?



this is a distribution of a population. It can be anything.



if we take sample from the population and find the sample mean, and repeat, plot it on a graph, if n is big, we will get a normal distribution!

DOESN'T MATTER WHAT DISTRIBUTION THE POPULATION IS!

7.3 Confidence Interval

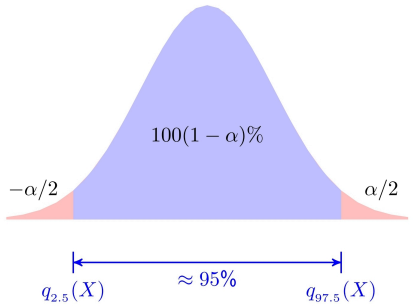
Confidence Interval of p% means you are p% confident that the outcome will be in your interval).

eg. if we say 95% confidence interval of random variable X is between 5 and 12, it means 95% of the time, X will be in between 5 and 12.

Although we can use confidence interval for a lot of case like (random varible X), we will usually find CI of sample mean, $\bar{X}$; and CI of sample variance, $\hat{S}^2$.
The reason being in real life we usually don't know the distribution of the random variable X but we can estimate the distribution of $\bar{X}$, or $\hat{S}^2$ ($\sim N$, $\sim t$, $\sim \chi^2$ and etc.)

There is two type of CI, one-sided CI and two-sided CI:

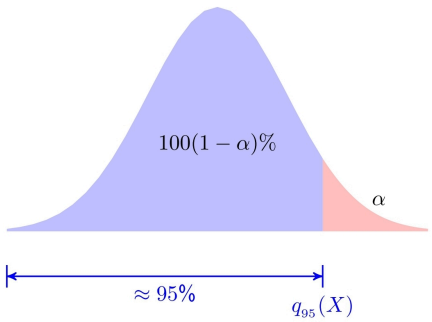
Two-sided interval:



$$P(C_l \leq X \leq C_u) = 1-\alpha \quad 0 < \alpha < 1$$

where C_l and C_u are the lower and upper confidence limits

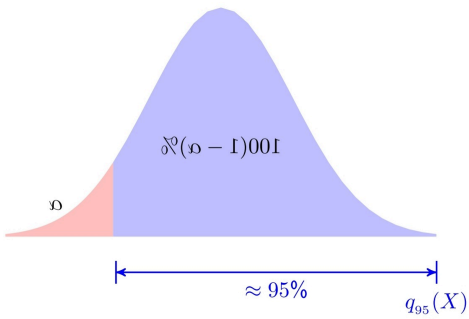
One-sided interval:



$$P(X \leq C_u) = 1-\alpha$$

where $0 < \alpha < 1$

One-sided interval:



$$P(C_l \leq X) = 1-\alpha$$

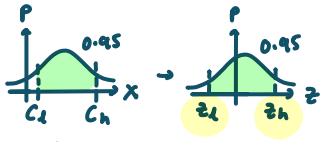
where $0 < \alpha < 1$

eg.

Confidence interval for a random variable.

Consider that the concentration of a substance X [mg/l] in a river is distributed as $N(3, 0.25)$. What is the value of the random variable X at a level of confidence 95%?

$X \sim N(3, 0.25)$



if doesn't mention anything → two-sided CI

$P(C_l \leq X \leq C_h) = 0.95$

$P(X \leq C_h) = 0.975$ $\quad P(X \leq C_l) = 0.025$

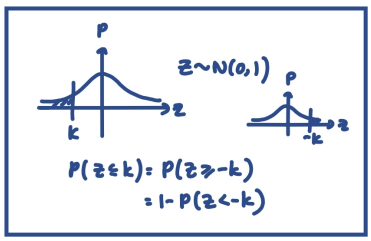
$P\left(z \leq \frac{C_h - 3}{\sqrt{0.25}}\right) = 0.975$ $\quad P\left(z \leq \frac{C_l - 3}{\sqrt{0.25}}\right) = 0.025$

$\frac{C_h - 3}{\sqrt{0.25}} = 1.96$ $\quad \frac{C_l - 3}{\sqrt{0.25}} = -1.96$

$C_h = 1.96\sqrt{0.25} + 3$ $\quad C_l = -1.96\sqrt{0.25} + 3$

$C_h = 3.98$ $\quad = 2.02$

$\therefore X \in (2.02, 3.98)$



7.4 Sampling Distribution.

Sampling Distribution → distribution of sample statistic ($\bar{x}$, $\hat{s}^2$ ...)

Sampling Distribution can be used to estimate population statistic (μ, σ^2)

☆ we know that from Central Limit Theorem, $\bar{X} \sim N(\mu, \sigma^2/n)$, if size of each sample, $n \to \infty$. This is a type of sampling distribution! Distribution of Sample Mean! However, $\bar{X} \sim N(\mu, \sigma^2/n)$ is only when $n \to \infty$, what if n is small? what if σ^2 is not known? what distribution will we use?

7.4.1 Sampling Distribution of Sample Mean.

basically we will try to use ~N
If σ^2 is known, if not known, try to estimate it
ONLY IF not possible to estimate (n small) we will use t-test.

Sample Mean

Standard Deviation (Population) is known

→ if was told X is normally distributed, no matter n is big (>30) or small, $\bar{x} \sim N(\mu, \sigma^2/n)$

→ if n is big (>30), $\bar{x} \sim N(\mu, \sigma^2/n)$

☆ $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ is equivalent to $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Standard Deviation (Population) is not known.

→ if n is small (n<30) or was told not to estimate σ, $\frac{\bar{x} - \mu}{\hat{s}_x/\sqrt{n}} \sim t(n-1)$

☆ WHAT IS T-DISTRIBUTION? CHECK EXAMPLE 3 BELOW!

→ if n is big (n>30), and is possible to estimate σ (given dist. of $\bar{x}$), $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ with estimated σ

☆ for year 2 — statistics, don't worry about:
- s.d. known, X is not normally dist. AND n is small. (technically we use t-test)

(two-sided, population s.d. known)

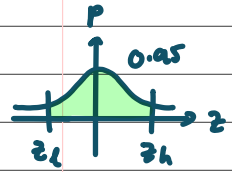
eg. Consider the compressive strengths of 40 test cubes of concrete. They have sample mean and standard deviation 60.14 and 5.02 N/mm^2. Assuming a normal distribution and that the population standard deviation is well estimated by the sample standard deviation, estimate the 95% confidence level for the population mean.

✿ no hint at one-sided or two-sided
→ choose two-sided

σ^2 known, $\sigma^2 = \hat{s}^2 = 5.02$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \rightarrow \bar{X} \sim N(\mu, \frac{5.02}{40}) \rightarrow z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

μ is unknown, that's why we need to find the CI of μ! we want to know for 95% confidence, what range (interval) will μ be at.



$P(z_l < \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} < z_h) = 0.95$

from table, $P(-1.96 < z < 1.96) = 0.95$:

$P(-1.96 < \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} < 1.96) = 0.95$

$P(\bar{x} - 1.96\sqrt{\sigma^2/n} < \mu < \bar{x} + 1.96\sqrt{\sigma^2/n}) = 0.95$

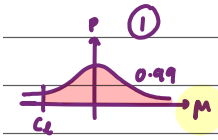
substituting $\bar{x} = 60.14$, $\sigma^2 = 5.02$ and $n = 40$,

$P(58.59 < \mu < 61.71) = 0.95$

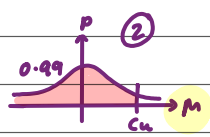
∴ CI of 95% of $\mu = (58.59, 61.71)$

(one-sided, population s.d. unknown but can be estimated)

eg.2 Consider the measurement of time-intervals $x_i (i = 1, 2, \ldots, 204)$ between vehicles at an observation point. The sample mean was found to be $\bar{x} = 0.551$ mins and the exponential distribution provides a good fit to the data. The transport experts' concern is for small mean inter-arrival times, since they lead to congestion. Obtain a one-sided 99% confidence interval for the population mean.



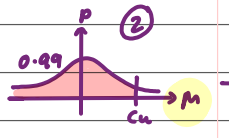
or ?

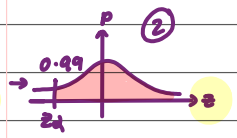


→ we concern the small (worse case)
→ we choose the second one

→ s.d. (population) unknown, but able to estimate σ^2
cause given $X \sim Exp(\lambda)$, so treat this question as s.d. known.
→ n big, → $\bar{x} \sim N(\mu, \sigma^2/\sqrt{n})$

BUT! when convert to z...
since $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, hence...

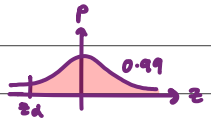




It will flip!
hence we find
$P(z > z_d) = 0.99$!

how to estimate σ^2?
if $X \sim Exp(\lambda)$
$E(X) = 1/\lambda$
$Var(X) = 1/\lambda$
} $E(X) = Var(X)$

we know $\bar{x} = 0.551$
$E(X) \approx \bar{x} = 0.551 = 1/\lambda$ $\quad Var(X) = \sigma^2 = \frac{1}{0.551}$
$\lambda = 1/0.551$



$z = \frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}}$ $\quad P(z > z_d) = 0.99$ $\quad P(\mu < \bar{x} - z_d\sqrt{\frac{\sigma^2}{n}}) = 0.99$

$P(\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} > z_d) = 0.99$ $\quad P(\mu < 0.658) = 0.99$

two-sided, population s.d. unknown

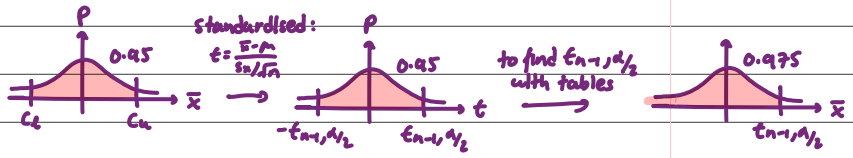
$\bar{x} = 60.14$ $\quad \hat{s}_x = 5.02$

Consider the compressive strengths of 40 test cubes of concrete. They have sample mean and standard deviation 60.14 and 5.02 N/mm^2. Assuming a normal distribution and that the population standard deviation is well estimated by the sample standard deviation, estimate the 95% confidence level for the population mean.

Reconsider Example 43 with the compressive strengths of 40 test cubes of concrete. Without assuming that σ is known estimate the 95% confidence level for the population mean.

→ use t-test!

$$T = \frac{\bar{x}-\mu}{\hat{s}_x/\sqrt{n}} \sim t(n-1)$$



$$P(-t_{n-1,\alpha/2} < T < t_{n-1,\alpha/2}) = 0.95$$

$$P\left(-t_{n-1,\alpha/2} < \frac{\bar{x}-\mu}{\hat{s}_x/\sqrt{n}} < t_{n-1,\alpha/2}\right) = 0.95$$

$$P\left(\bar{x} - t_{n-1,\alpha/2}\frac{\hat{s}_x}{\sqrt{n}} < \mu < \bar{x} + t_{n-1,\alpha/2}\frac{\hat{s}_x}{\sqrt{n}}\right) = 0.95$$

$$P\left(60.14 - 2.023 \times \frac{5.02}{\sqrt{40}} < \mu < 60.14 + 2.023 \times \frac{5.02}{\sqrt{40}}\right) = 0.95$$

$$P(58.53 < \mu < 61.76) = 0.95$$

$\therefore$ CI of 95% = (58.53, 61.76)

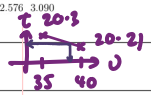
How to get t-values from t-distribution table?

STUDENT'S t Table

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

$v = n-1$
$= 40-1$
$= 39$

interpolate for $v = 39$:



$$20.30 - \frac{39-35}{40-35} \times (20.30 - 20.21)$$

$$= 20.228 \approx 20.23$$

7.4.2 Sampling Distribution of Sample Variance

unlike sample mean, sample variance is more straightforward:

$$V = \frac{(n-1)\hat{s}_x^2}{\sigma^2} \sim \chi^2(n-1)$$, no matter n big or small...

7.4.1 Sampling Distribution of Sample Mean.

basically we will try to use $\bar{x}$
If σ^2 is known, if not known, try to estimate it
ONLY IF not possible to estimate (n small) we will use t-test.

Sample Mean

Standard Deviation (Population) is known	Standard Deviation (Population) is not known.
→ if was told X is normally distributed, no matter n is big (>30) or small, $\bar{x} \sim N(\mu, \sigma^2/n)$	→ if n is small (n<30) or was told not to estimate σ, $\frac{\bar{x}-\mu}{\hat{s}_x/\sqrt{n}} \sim t(n-1)$ WHAT IS T-DISTRIBUTION? CHECK EXAMPLE 3 BELOW!
→ if n is big (>30), $\bar{x} \sim (\mu, \sigma^2/n)$	→ if n is big (n>30), and is possible to estimate σ (given dist. of $\bar{x}$), $\bar{x} \sim N(\mu, \frac{\sigma^2}{\sqrt{n}})$ with estimated σ

$\bar{x} \sim N(\mu, \sigma^2/n)$ is equivalent to $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

for year 2 — statistics, don't worry about:
- s.d. known, X is not normally dist. AND n is small. (technically we use t-test)

compared to sample mean...

eg.

Reconsider Example 39 with the compressive strengths of 40 test cubes of concrete. Estimate the 99% confidence level for the population variance.

given that (in example 39), $\bar{x} = 60.14$; $\hat{S}_x = 5.02$

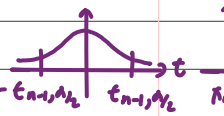
notice that here we don't write $P(-\chi^2_{n-1,\alpha/2} < V < \chi^2_{n-1,\alpha/2})$ cause χ^2 is not a symmetric dist. like N or t

$P(\chi^2_{n-1,1-\alpha/2} < V < \chi^2_{n-1,\alpha/2}) = 0.99$ (two-sided CI)

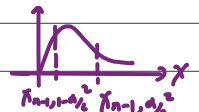
$P(\chi^2_{n-1,1-\alpha/2} < \frac{(n-1)\hat{S}_x^2}{\sigma^2} < \chi^2_{n-1,\alpha/2}) = 0.99$

$P(\frac{n-1}{\chi^2_{n-1,\alpha/2}}\hat{S}_x^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\hat{S}_x^2) = 0.99$

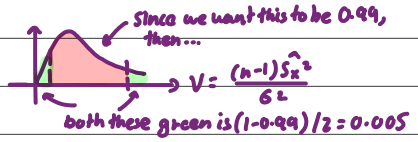
standardised normal

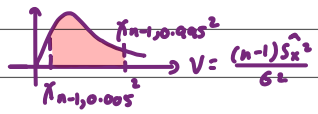


standardised t-dist with $\upsilon = n-1$ (degree of freedom)



standardised χ^2-dist with $\upsilon = n-1$ (degree of freedom)





(by the way, degree of freedom, $\upsilon = n-1$ in this case, $\upsilon = 40-1 = 39$)

* I am taking $\upsilon = 40$ cause I am lazy to interpolate (it's fine)

to find $\chi^2_{39,0.005}$

CHI-SQUARED Table - 1/2

ν	0.1%	0.5%	1.0%	2.5%	5.0%	10.0%	12.5%	20.0%	25.0%	33.3%	50.0%
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337
26	9.222	11.160	12.198	13.844	15.379	17.292	18.021	19.820	20.843	22.399	25.336
27	9.803	11.808	12.879	14.573	16.151	18.114	18.861	20.703	21.749	23.339	26.336
28	10.391	12.461	13.565	15.308	16.928	18.939	19.704	21.588	22.657	24.280	27.336
29	10.986	13.121	14.256	16.047	17.708	19.768	20.550	22.475	23.567	25.222	28.336
30	11.588	13.787	14.953	16.791	18.493	20.599	21.399	23.364	24.478	26.165	29.336
35	14.688	17.192	18.509	20.569	22.465	24.797	25.678	27.836	29.054	30.894	34.336
40	17.916	20.707	22.164	24.433	26.509	29.051	30.008	32.345	33.660	35.643	39.335
45	21.251	24.311	25.901	28.366	30.612	33.350	34.379	36.884	38.291	40.407	44.335
50	24.674	27.991	29.707	32.357	34.764	37.689	38.785	41.449	42.942	45.184	49.335
55	28.173	31.735	33.570	36.398	38.958	42.060	43.220	46.036	47.610	49.972	54.335
60	31.738	35.534	37.485	40.482	43.188	46.459	47.680	50.641	52.294	54.770	59.335

to find $\chi^2_{39,0.995}$

CHI-SQUARED Table - 2/2

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	59.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607

$\therefore \chi^2_{40,0.005} = 20.707$, $\chi^2_{40,0.995} = 66.766$

$P(\frac{n-1}{\chi^2_{n-1,\alpha/2}}\hat{S}_x^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\hat{S}_x^2) = 0.99$

$P(\frac{39}{66.766}5.02^2 < \sigma^2 < \frac{39}{20.707}5.02^2) = 0.99$

$P(14.72 < \sigma^2 < 47.46) = 0.99$

$\therefore$ CI of 99% for $\sigma^2 = (14.72, 47.46)$

$\chi^2_{n-1,\alpha/2}$ is the "right" one and $\chi^2_{n-1,1-\alpha/2}$ is the "left" one.

(in this case $\chi^2_{n-1,\alpha/2} = \chi^2_{39,0.995}$, $\chi^2_{n-1,1-\alpha/2} = \chi^2_{39,0.005}$)

summary :

H_0 (null), H_1 (alt)

population

statistic

$P(\text{statistic} \mid H_0 \text{ true})$

p-value

p-value $< \alpha \rightarrow$ reject H_0

p-value $\geq \alpha \rightarrow$ accept H_0

c8 Statistical Testing (Hypothesis Testing)

8.1 Introduction

We do statistical testing when we have a hypothesis in mind, and want to check if our hypothesis is true or not. (eg. mean of height of year 2's student = 170 cm)

To do hypothesis testing, we will implement the following steps (let μ = mean of height of year 2's student)

1. null hypothesis, H_0 : $\mu = 170$

2. alternative hypothesis, H_1 :
   $\mu \neq 170$
   (since we want to test if μ is really 170, and not like whether it is greater or smaller than 170, it is a TWO-SIDED hypothesis test)

(null hypothesis doesn't necessary is our initial claim. In this case, our initial claim is "$\mu = 170$". What if my initial claim is "Year 2 average height is greater than Year 1"? Then my null hypothesis, H_0 will be: $\mu_2 = \mu_1$, and alternative hypothesis will be $\mu_2 > \mu_1$. This prove that initial claim is not always H_0 or H_1. Null hypothesis is just always the "neutral" one, and is always treated as true, unless proven false)

3. Define a statistical test

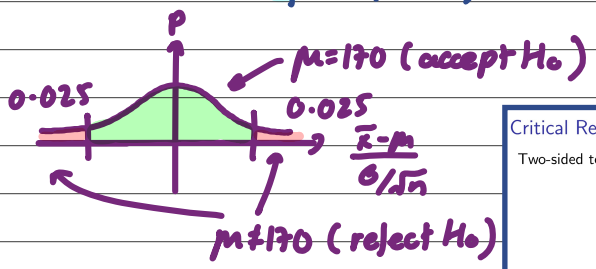
eg. If mean height of year 2's student is normal distribution (this is true if σ is known and $n > 30$):

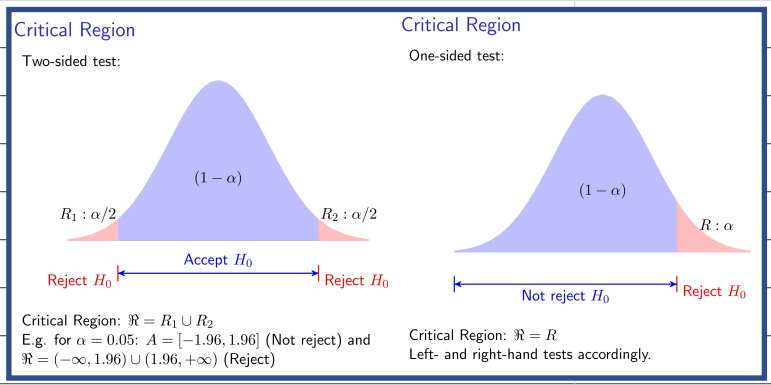
$$\bar{X} \sim N(\mu, \sigma^2)$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

4. Define critical region OR find the probability of alternative hypothesis.

eg. if I want to test at 95% confidence (5% significant) whether $\mu = 170$, we need to check if $P(\mu \neq 170) < 0.05$, if yes, we reject H_0 (H_1 is true, i.e. $\mu \neq 170$)





eg. (just an example on how hypothesis testing procedure works)

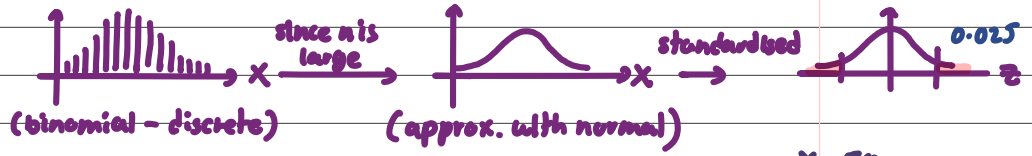
Check if a coin is fair. Toss it 100 times and answer with 95% confidence. (5% significant)

1. Define H_0: Coin is fair (i.e. $P(H)=0.5$)
   ← probability of getting heads = 0.5 (mathematical term)
2. Define H_1: Coin is not fair (i.e. $P(H)\neq 0.5$)

3. Define a statistical test.

☆ AT THIS STAGE (when defining a statistical test) WE MUST TREAT H_0 AS TRUE!
i.e. $P(H)=0.5$!

→ since tossing coins ($n=100$) times, and each toss is ($p=0.5$), this is BINOMIAL.
→ $X\sim B(100,0.5)$
→ to reject H_0,



$X\sim B(100,0.5)$ → $X\sim N(50,25)$ → $\frac{X-50}{\sqrt{25}}\sim N(0,1)$

$E(X)=np=50$
$Var(X)=np(1-p)=25$

we will reject H_0 if $P(Z<-z_a)=0.025$
$P(Z>z_a)=0.025$

$P(Z<-z_{a/2})=0.025$	$P(Z>z_{a/2})=0.025$
$P(\frac{X-50}{5}<-1.96)=0.025$	$P(\frac{X-50}{5}>1.96)=0.025$
$P(X<40.2)=0.05$	$P(X>59.8)=0.025$

hence we will reject H_0 (coin is not fair), if:
we get 40 or less head out of 100 throws, or
60 or more head out of 100 throws
we will accept H_0 (coin is fair), if:
we get 41 to 59 head out of 100 throws.

8.2 Asymmetric Case (One-sample Analysis) and Symmetric Case (Two-sample Analysis)

Statistical Testing (Hypothesis Testing)

Asymmetric Case (one-sample)
Compare ONE SAMPLE's statistic to a known population statistic
(this is a constant.)

Symmetric Case (two-sample)
Compare TWO SAMPLE's statistic

eg. testing whether Year 2's average height is greater than world's average height

$H_0: \mu_2 = 170$ $(\mu_2 = \mu_{world} = 170)$
$H_1: \mu_2 > 170$ $(\mu_2 > \mu_{world} = 170)$

notice population mean is constant. That is why we call it one-sample analysis because only one sample, the other one is population and is constant.

eg. testing whether Year 2's average height is greater than Year 1's average height.

$H_0: \mu_2 = \mu_1$
$H_1: \mu_2 > \mu_1$

notice that both sample mean is NOT constant. That is why we called it two sample analysis.

suppose A be the population with known mean:

1. σ_B is known ($\sigma_B = \sigma_A = 0$)

$H_0: \mu_B = \mu_A$ ← μ_A is constant
$H_1: \mu_B > \mu_A$ (one-sided)
$\mu_B \neq \mu_A$ (two-sided)

$$z = \frac{\bar{x}_B - \mu_B}{\sigma_B/\sqrt{n}} \sim N(0,1)$$

2. σ_B is not known (and can't be estimated)

$H_0: \mu_B = \mu_A$ ← μ_A is constant.
$H_1: \mu_B > \mu_A$ (one-sided)
$\mu_B \neq \mu_A$ (two-sided)

$$T = \frac{\bar{x}_B - \mu_B}{\hat{s}_{x_B}/\sqrt{n}} \sim t(n-1)$$

1. σ_B is known, σ_A is known

$H_0: \mu_B = \mu_A$ ← μ_B and μ_A both NOT constant
$H_1: \mu_B > \mu_A$ (one-sided)
$\mu_B \neq \mu_A$ (two-sided)

($\mu_A - \mu_B = 0$)

$$z = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim N(0,1)$$

2. σ_B and σ_A is not known but ($\sigma_A = \sigma_B$)

$H_0: \mu_B = \mu_A$ ← μ_B and μ_A both NOT constant!
$H_1: \mu_B > \mu_A$ (one-sided)
$\mu_B \neq \mu_A$ (two-sided)

$$t = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{s_p^2}{n_A} + \frac{s_p^2}{n_B}}} \sim t(n_A + n_B - 2)$$

$$\left(\hat{s}_p^2 = \frac{(n_A-1)\hat{s}_{x_A}^2 + (n_B-1)\hat{s}_{x_B}^2}{n_A + n_B - 2}\right)$$

$\hat{s}_p^2$: pooled variance

3. σ_B and σ_A is not known and ($\sigma_A \stackrel{?}{=} \sigma_B$)

same as (2) but

$$\sim t(\upsilon) \text{ where } \upsilon = \frac{\left(\frac{\hat{s}_{x_A}^2}{n_A} + \frac{\hat{s}_{x_B}^2}{n_B}\right)^2}{\frac{1}{n_A - 1}\left(\frac{\hat{s}_{x_A}^2}{n_A}\right)^2 + \frac{1}{n_B - 1}\left(\frac{\hat{s}_{x_B}^2}{n_B}\right)^2}$$

8.3 Test Error

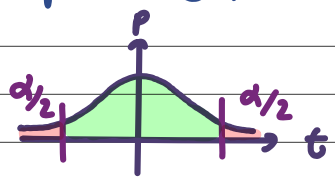
Two types of errors when testing:

Condition / Decision	*accept* H_0	reject H_0
H_0 true	$1-\alpha$	α
H_1 true	β	$1-\beta$

Type 1

Type 2

Type 1 error is when H_0 (Null Hypothesis) is TRUE, but we reject it.
The probability of this happening is just α (significant level)



we reject H_0 when our test statistic lands in α.
And if H_0 is true, only α% we get it wrong

Type 2 error is when H_0 (Null Hypothesis) is FALSE, but we accept it.
The probability of this happening is β

$\therefore \alpha = P(\text{reject } H_0 \mid H_0)$
$\therefore \beta = P(\text{accept } H_0 \mid H_1)$

Two types of errors when testing:

Condition / Decision	*accept* H_0	reject H_0
H_0 true	$1-\alpha$	α
H_1 true	β	$1-\beta$

confidence level

significance

power

VERY HARD EXAMPLE (one-sample, type 2 error, two-sample)

3. *Samples of concrete beams*

The densities of 30 concrete beams produced by a given supplier A have been tested. In standard units, the sample mean is 20 and the sample standard deviation (using the unbiased formula) is 1.8.

$n_A = 30,\ \bar{x}_A = 20,\ \hat{s}_{x_A} = 1.8$

A set of 10 beams is found in a warehouse, and the mean density of this new sample is 18 standard units.

$n = 10,\ \bar{x} = 18$

c - Assuming that the sample mean and standard deviation found for the first sample of 30 beams can be used as reliable estimates of, respectively, the population mean and standard deviation of beams produced by supplier A, test with 5% significance the hypothesis that these 10 beams, although their mean density is smaller than 20, are produced by supplier A.

$\mu_A = \bar{x}_A = 20$

$\sigma_A = \hat{s}_{x_A} = 1.8$

one-tailed test

Step 1. Define H_0 and H_1

$H_0: \mu = \mu_A\ (=20)$

$H_1: \mu < \mu_A\ (=20)$

at this stage, we will assume null hypothesis, H_0 IS TRUE (until proven false otherwise)

$\therefore \mu = \mu_A = 20$

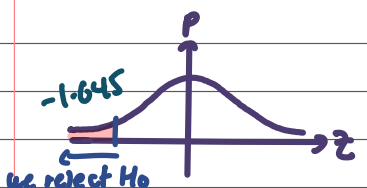
Step 2. Define a statistical test

- using z-test as σ^2 known ($\sigma^2 = \sigma_A^2$)

$$z = \frac{\bar{x} - \mu}{\hat{s}_{\bar{x}}/\sqrt{n}} \sim N(0,1)$$

$\hat{s}_x^2$ in this question is not given! However, we can treat $\hat{s}_x^2 = \sigma_A^2$ as we supposed H_0 as true, and H_0 says that sample is from A, hence: $\hat{s}_x^2 = \sigma_A^2$!

$$\frac{\bar{x} - 20}{1.8/\sqrt{10}} \sim N(0,1)$$



we reject H_0 if t-value is to the left of $-t_{9,0.05}$

to find what $\bar{x}$ will give critical z-value ($-z_{0.05}$)

$P(z < -z_{0.05}) = 0.05$

$P(z > z_{0.05}) = 0.05$

$P(z < z_{0.05}) = 0.95$

$z_{0.05} = 1.645$

$$\therefore P\left(\frac{\bar{x} - 20}{1.8/\sqrt{10}} < -1.645\right) = 0.05$$

$$P(\bar{x} < 19.06) = 0.05$$

since $\bar{x} = 18 < 19.06$, reject H_0! (these 10 beams are NOT produced by supplier A)

d - The suspicion arises that these 10 beams are actually produced by another supplier, B, for whom the population mean is estimated to be 19 standard units (we assume for this question that the population standard deviation is the same as for supplier A, i.e. 1.8). If the test carried out in (c) is now considered as a test of supplier A (null hypothesis) against supplier B (alternative hypothesis), what is the power of this test?

from part (c): $\mu_A = 20,\ \sigma_A = 1.8$

from part (d): $\mu_B = 19,\ \sigma_B = 1.8$

$H_0: \mu = \mu_A\ (=20)$

$H_1: \mu = \mu_B\ (=19)$

power $= 1 - \beta = P(\text{reject } H_0 \mid H_1)$

power = $P(\text{reject } H_0 \mid H_1)$

from part (c) we know we reject H_0 when $\bar{x} < 19.0637$

power = $P(\bar{x} < 19.0637 \mid \mu = 19)$

$= P\left(z < \frac{19.0637 - 19}{1.8/\sqrt{10}}\right)$

$= P(z < 0.1119)$

$= 0.5445$

e - It is now no longer assumed that the original sample of 30 beams provides a reliable estimate of the population statistics of beam densities from supplier A. Given that the sample standard deviation (unbiased) for the new sample of size 10, is 1.75 standard units, test with 5% significance whether this sample of 10 beams has a different mean from the original sample of 30 beams (in this test, the alternative hypothesis is, as in

(d), that the sample of 10 beams is from supplier B). Note that the two sample standard deviations are very similar.

$n_A = 30, \bar{x}_A = 20, \hat{s}_{x_A} = 1.8$

$n_B = 10, \bar{x}_B = 19, \hat{s}_{x_B} = 1.75$

$H_0: \mu_A = \mu_B$
$H_1: \mu_A \neq \mu_B$
} both μ_A, μ_B are unknown (two-sample analysis)
σ_A^2 and σ_B^2 unknown but $(\sigma_A^2 = \sigma_B^2)$

$\mu_A - \mu_B = 0$

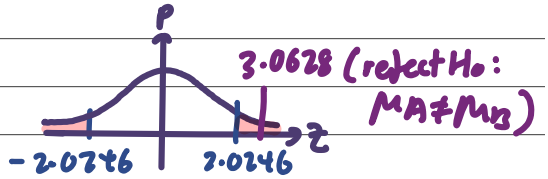
$$t = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{s_p^2}{n_A} + \frac{s_p^2}{n_B}}} \sim t(n_A + n_B - 2)$$

$$\left(\hat{s}_p^2 = \frac{(n_A - 1)\hat{s}_{x_A}^2 + (n_B - 1)\hat{s}_{x_B}^2}{n_A + n_B - 2}\right)$$

$\hat{s}_p^2$: pooled variance

$$\hat{s}_p^2 = \frac{(30-1) \times 1.8^2 + (10-1) \cdot 1.75^2}{30+10-2} = 3.198$$

$$t = \frac{(20-18) - (0)}{\sqrt{\frac{3.198}{30} + \frac{3.198}{10}}} = 3.0628 \sim t(30+10-2)$$



$P(T > t_{38, 0.025}) = 0.025$

$P(T < t_{38, 0.025}) = 0.975$

STUDENT'S t Table

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307

$$2.030 - (2.030 - 2.021) \times \frac{38-35}{40-35}$$

$= 2.0246$

8.4 Goodness of Fit Test.

Goodness of fit test is to check how well a distribution fit onto data sample.
In CS3 Goodness of Fit, we covered the following test

Tools we already have to assess the goodness of fit:

- qq-plots
- Compare CDFs
- Compare PDFs (use of histograms)

All these tools are visual ~ qualitative. → not very good.

Quantitative tests: → better!

- Decide with a level of confidence $1-\alpha$ if distribution D provides a good fit to the data

two type of test covered in year 2:

Feature	Chi-Squared	Kolmogorov-Smirnov (KS)
Data Type	Binned Continuous	Continuous
Basis	Observed vs. Expected Frequencies	ECDF vs. Theoretical CDF
Comparison Metric	Sum of squared relative differences (χ^2)	Maximum absolute difference (D)

8.4.1 Chi-squared Test (χ^2)

χ^2 test is always a one-sided (specifically right-sided) test!

1.	Null Hypothesis	H_0 : Distribution is a good fit
2.	Alternative Hypothesis	H_1 : Distribution is not a good fit
3.	Test statistic	$C_n = \sum_{i=1}^{l} \frac{(O_i - E_i)^2}{E_i}$
4.	Distribution of test statistic	$C_n \sim \chi^2(l-1-k)$
5.	Critical Region $\Re$	$C_n > \chi^2_{l-1-k,1-\alpha}$ (e.g. $\alpha = 5\%$)
6.	Evaluate C_n under H_0	if $c_{n0} > \chi^2_{l-1-k,\alpha} \Rightarrow$ Reject H_0
7.	$p-value$	$p(C_n > c_{n0})$

} H_0 and H_1 will be always like this!

Expected frequencies:

$$E_i = (F_X(x_{i+1}) - F_X(x_i)) \times n \quad \forall i \in [2, l-1]$$

where F_X is the CDF under test.
For class 1: $E_1 = F_X(x_2) \times n$
For class l: $E_l = (1 - F_X(x_l)) \times n$
Since most distributions extend to infinity ($\pm$)
Careful when selecting own classes (bins) - avoid empty bins. Best practice $E_i \geq 5$ (combine classes to form)

} important! for continuous distribution

eg1.

100 random values of x are taken from a distribution X which can take values 0 , 1 , 2 , 3 , 4 & 5, shown in the table. Test, at 1 % significance level, whether X~B(5 , 0.1) is a good fit.

x	0	1	2	3	4	5
Frequency	60	21	15	2	1	1

← This is observed frequency (data sample)

1.	Null Hypothesis	H_0 : $X \sim B(5, 0.1)$ is a good fit
2.	Alternative Hypothesis	H_1 : $X \sim B(5, 0.1)$ is NOT a good fit
3.	Test statistic	$C_n = \sum_{i=1}^{l} \frac{(O_i - E_i)^2}{E_i}$

To find expected frequency, E_i: $E_i = n \times P(X = x_i)$ ← for $X \sim B(n,p)$, $P(X=x) = {}^nC_x \times p^x \times (1-p)^{n-x}$

$E_0 = 100\,P(X=0) = 100\left({}^5C_0 \times 0.1^0 \times 0.9^5\right) = 59.049$

$E_1 = 100\,P(X=1) = 100\left({}^5C_1 \times 0.1^1 \times 0.9^4\right) = 32.805$

and so on... $E_2 = 7.29$, $E_3 = 0.81$, $E_4 = 0.045$, $E_5 = 0.001$

x	0	1	2	3	4	5
O_i	60	21	15	2	1	1
E_i	59.049	32.805	7.29	0.81	0.045	0.001

these three $E_i < 5$, we need to combine till no $E_i < 5$

so combine 2,3,4,5 (combining 3,4,5 won't make $E_i \geq 5$!)

x	0	1	2 to 5
O_i	60	21	19
E_i	59.049	32.805	8.146

$$C_n = \sum_{i=1}^{l} \frac{(O_i - E_i)^2}{E_i} = \frac{(60-59.049)^2}{59.049} + \frac{(21-32.805)^2}{32.805} + \frac{(19-8.146)^2}{8.146}$$

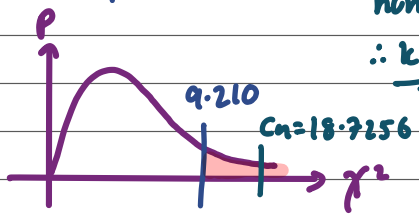
$\therefore C_n = 18.7256$

4. Distribution of test statistic $C_n \sim \chi^2(l-1-k)$
5. Critical Region $\Re$ $C_n > \chi^2_{l-1-k,1-\alpha}$ (e.g. $\alpha = 1\%$)

$C_n \sim \chi^2(3-1-0)$

$C_n \sim \chi^2(2)$

k is number of estimated parameter.
In this question, we were told $X \sim B(n,p)$
non of the "n" and "p" are estimated.
$\therefore k = 0$



CHI-SQUARED Table - 2/2

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515

$P(\chi^2 > \chi^2_{2,0.01}) = 0.01$

$P(\chi^2 < \chi^2_{2,0.01}) = 0.99$

$\chi^2_{2,0.01} = 9.210$

since $C_n = 18.7256 > \chi^2_{2,0.01} = 9.210$ (reject H_0, $X \sim B(5,0.1)$ is NOT a good fit)

eg2.

150 values of x are taken from a continuous population, shown in the table. Test whether N(? , 40) is a good fit at 10 % significance level.

x	f
– 19	0
20 – 29	22
30 – 34	54
35 – 44	65
45 – 49	9
50 –	0

there is one unknown parameter

$C_n \sim \chi^2(l-1-k)$, $k=1$

Step 1. Estimate the unknown parameter

$X \sim N(\mu, \sigma^2)$

mean, so find mean of the sample.

$$\mu \approx \bar{x} = \frac{22(24.5)+54(32)+65(39.5)+9(47)}{150} = 35.05$$

Step 2. Define H_0 and H_1

H_0 : $X \sim N(35.05, 40)$ is a good fit

H_1 : $X \sim N(35.05, 40)$ is NOT a good fit

Step3. Find expected frequency, E_i

x	f
– 19	0
20 – 29	22
30 – 34	54
35 – 44	65
45 – 49	9
50 –	0

$E_i = 150\,P(x_l \le X < x_u)$

$150\,P(X<19.5) = 1.046$

$150\,P(19.5 \le X < 29.5) = 27.469$

$150\,P(29.5 \le X < 34.5) = 41.288$

$150\,P(34.5 \le X < 44.5) = 70.063$

$150\,P(44.5 \le X < 49.5) = 8.466$

$150\,P(X \ge 49.5) = 1.675$

example how to find $P(19.5 \le X < 29.5)$

$= P(X<29.5) - P(X<19.5)$

$= P\left(z < \frac{29.5-35.05}{\sqrt{40}}\right) - P\left(z < \frac{19.5-35.05}{\sqrt{40}}\right)$

$= P(z<-0.878) - P(z<-2.459)$

$= (1-P(z<0.878)) - (1-P(z<2.459))$

$= P(z<2.459) - P(z<0.878)$

$= 0.9931 - 0.8106$

$= 0.1825$

Step 4. Combine classes such that $E_i \ge 5$

x	f	E_i
– 19	0	1.046 } 28.515
20 – 29	22	27.469 }
30 – 34	54	41.288
35 – 44	65	70.063
45 – 49	9	8.466 } 10.141
50 –	0	1.675 }

→

x	O_i	E_i
– 19 }	22	28.515
20 – 29 }		
30 – 34	54	41.288
35 – 44	65	70.063
45 – 49 }	9	10.141
50 – }		

Step 5. Find $C_n = \sum_i^l \frac{(O_i - E_i)^2}{E_i}$

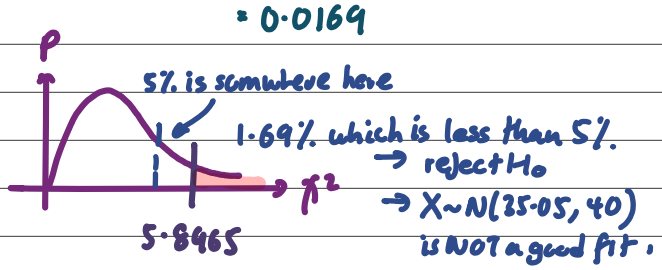
$$C_n = \frac{(22-28.515)^2}{28.515} + \frac{(54-41.288)^2}{41.288} + \frac{(65-70.063)^2}{70.063} + \frac{(9-10.141)^2}{10.141} = 5.8465$$

Step 6. Define teststatistic and find critical region.

$C_n \sim \chi^2(l-1-k) = \chi^2(4-1-1) = \chi^2(2)$

(last example I find critical region, this question I try another approach — find p-value)

$P(\chi^2 \geq 5.8965) = 1 - P(\chi^2 < 5.8965)$
$= 1 - 0.9831$
$= 0.0169$



CHI-SQUARED Table - 2/2

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458

$0.975 + \frac{5.8965 - 5.024}{6.635 - 5.024} \times (0.99 - 0.975)$

≈ 0.9831

⚠ ONE THING I NOTICE IS FOR YEAR 2'S STATS, IF $E_i < 5$ WE DON'T COMBINE CLASSES. WE ALSO ALWAYS MINUS NUMBER OF PARAMETER FOR $C_n \sim \chi^2(l-1-k)$ NO MATTER IF THEY ARE ESTIMATED OR NOT!
2

8.4.2 Kolmogorov-Smirnov (KS) Test

1.	Null Hypothesis	H_0 : Distribution is a good fit
2.	Alternative Hypothesis	H_1 : Distribution is not a good fit
3.	Test statistic	$D_n = \max \lvert F_n(x) - F_X(x) \rvert$
4.	Distribution of test statistic	$\lim_{n\to\infty} P(\sqrt{n} D_n \leq z)$ is tabulated
5.	Critical Region $\Re$	$D_n > d_{n,1-\alpha}$ (e.g. $\alpha = 5\%$)
6.	Evaluate D_n under H_0	if $d_{n0} > d_{n,1-\alpha} \Rightarrow$ Reject H_0
7.	$p - value$	$p(D_n > d_{n0})$

same eg from 8.4.1

150 values of x are taken from a continuous population, shown in the table. Test whether N(? , 40) is a good fit at 10 % significance level.

x	f
-19	0
$20-29$	22
$30-34$	54
$35-44$	65
$45-49$	9
$50-$	0

this is how we write H_0 and H_1 for ks-test! we are comparing empirical CDF to theoretical CDF

Step 1. Define H_0 and H_1

H_0: $F(x) = F_T(x)$ for ALL values of x
H_1: $F(x) \neq F_T(x)$ for at least one value of x

the reason for "one" value is because we only compare the max difference.

Step 2. Find ECDF and Theoretical CDF at boundaries.

(PDF) (CDF) $ECDF = \frac{CDF}{\text{Tf}}$ $D = |ECDF - Theo.CDF|$

Class Interval	Upper Boundary	Freq	Cum. Obs. Freq	ECDF	Z-score	Theo. CDF	D
$\leqslant 19$	19.5	0	0	0.0000	-2.459	0.0070	0.0070
20 – 29	29.5	22	22	0.1467	-0.878	0.1899	0.0432
30 – 34	34.5	54	76	0.5067	-0.0875	0.4651	0.0416
35 – 44	44.5	65	141	0.9400	1.494	0.9324	0.0076
45 – 49	49.5	9	150	1.0000	2.285	0.9888	0.0112
$\geqslant 50$	$+\infty$	0	150	1.0000	$+\infty$	1.0000	0.0000

$z = \frac{\text{upper boundary} - \mu}{\sigma} = \frac{\text{upper boundary} - 35.05}{\sqrt{40}}$

Theoretical CDF = $\Phi(z)$

eg. for $z = 1.494$

NORMAL CUMULATIVE DISTRIBUTION FUNCTION

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.93	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

Step 3. Define test statistic

$D_n = \max(D) = 0.0432$

Kolmogorov-Smirnov Test (level of significance a)

n	0.2	0.1	0.05	0.02	0.01
1	0.9000	0.9500	0.9750	0.9900	0.9950
2	0.6838	0.7764	0.8419	0.9000	0.9293
3	0.5648	0.6360	0.7076	0.7846	0.8290
4	0.4927	0.5652	0.6239	0.6889	0.7342
5	0.4470	0.5094	0.5633	0.6272	0.6685
6	0.4104	0.4680	0.5193	0.5774	0.6166
7	0.3815	0.4361	0.4834	0.5384	0.5758
8	0.3583	0.4096	0.4543	0.5065	0.5418
9	0.3391	0.3875	0.4300	0.4796	0.5133
10	0.3226	0.3687	0.4092	0.4566	0.4889
11	0.3083	0.3524	0.3912	0.4367	0.4677
12	0.2958	0.3382	0.3754	0.4192	0.4490
13	0.2847	0.3255	0.3614	0.4036	0.4325
14	0.2748	0.3142	0.3489	0.3897	0.4176
15	0.2659	0.3040	0.3376	0.3771	0.4042
16	0.2578	0.2947	0.3273	0.3657	0.3920
17	0.2504	0.2863	0.3180	0.3553	0.3809
18	0.2436	0.2785	0.3094	0.3457	0.3706
19	0.2373	0.2714	0.3014	0.3369	0.3612
20	0.2316	0.2647	0.2941	0.3287	0.3524
21	0.2262	0.2586	0.2872	0.3210	0.3443
22	0.2212	0.2528	0.2809	0.3139	0.3367
23	0.2165	0.2475	0.2749	0.3073	0.3295
24	0.2120	0.2424	0.2693	0.3010	0.3229
25	0.2079	0.2377	0.2640	0.2952	0.3166
26	0.2040	0.2332	0.2591	0.2896	0.3106
27	0.2003	0.2290	0.2544	0.2844	0.3050
28	0.1968	0.2250	0.2499	0.2794	0.2997
29	0.1935	0.2212	0.2457	0.2747	0.2947
30	0.1903	0.2176	0.2417	0.2702	0.2899
31	0.1873	0.2141	0.2379	0.2660	0.2853
32	0.1844	0.2108	0.2342	0.2619	0.2809
33	0.1817	0.2077	0.2308	0.2580	0.2768
34	0.1791	0.2047	0.2274	0.2543	0.2728
35	0.1766	0.2018	0.2242	0.2507	0.2690
36	0.1742	0.1991	0.2212	0.2473	0.2653
37	0.1719	0.1965	0.2183	0.2440	0.2618
38	0.1697	0.1939	0.2154	0.2409	0.2584
39	0.1675	0.1915	0.2127	0.2379	0.2552
40	0.1655	0.1891	0.2101	0.2349	0.2521
> 40	$1.07/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.52/\sqrt{n}$	$1.63/\sqrt{n}$

$n = 150$

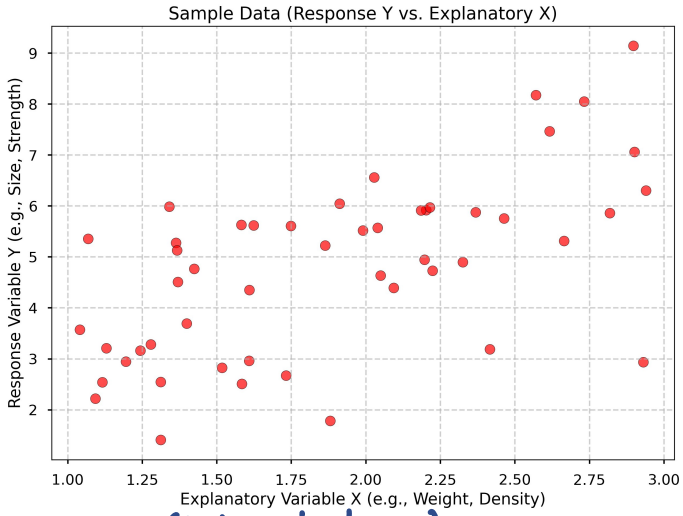
$d_{150,0.9} = \frac{1.22}{\sqrt{150}} = 0.0996$

Step 4. Conclude

$D_n = 0.0432 < d_{150,0.9} = 0.0996$ (do not reject $H_0 \rightarrow X \sim N(35.05, 40)$ is a good fit)

c9. Linear Regression

9.1 Introduction



(dependant var.)

(independant var.)

Linear regression is to fit a straight line onto a dataset of dependant vs. independant

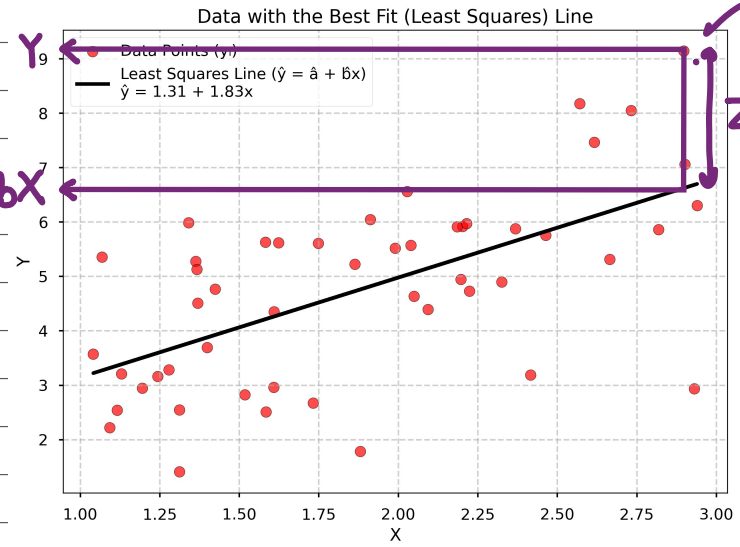
The equation of Linear Regression is :

$$Y = a + bX + Z$$

Z is an error term : (residual)

1. Z is an independent term. It is independent of X!

a and b are called regression coefficient.



our example reference point.

this is only "Z" of one of the point from the samples. For n points in a samples, there is n number of X, Y and Z

cool thing to note is for all n number of Z, the mean of all the Z is 0!

→ $E[Z] = 0$!

this infer that the mean of all the Y is a+bX!

→ $E[Y] = a + bX$

about a and b:

in real world scenario, we can't measure population, but we can obtain sample to estimate population...

eg. we can't find the mean of human height (there is 8 billion human → μ)

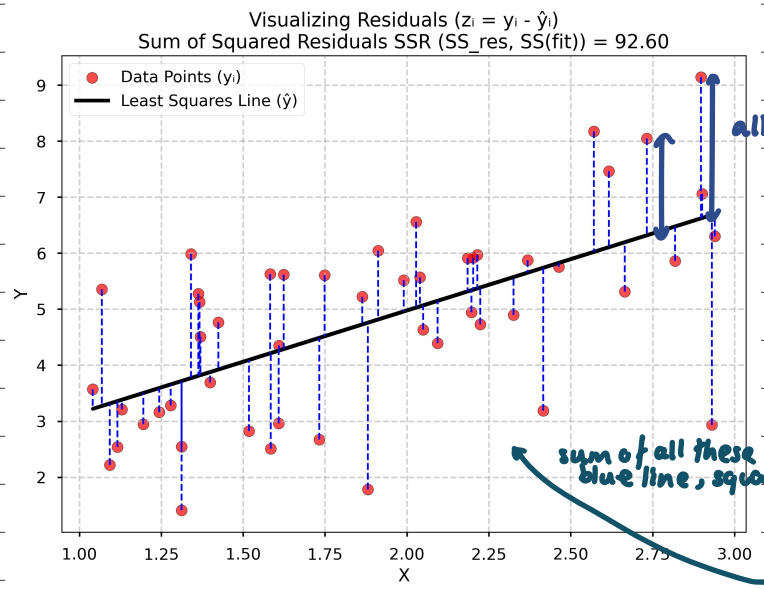
but we can obtain a sample and get the sample mean → $\bar{x}$

same for a and b. We can't get the entire population data to fit a linear line, but we can obtain a sample and fit a linear line.

Hence, the "a" and "b" we get for our line is named "$\hat{a}$" and "$\hat{b}$"

(same idea as "$\bar{x}$" for "μ" and "$\hat{s}_x^2$" for "σ_x^2")

So how do we obtain $\hat{a}$ and $\hat{b}$ that will give the best fit?



all these blue lines are z (residual)
the idea is: line of best fit should have small residual.

Hence: we will try to obtain

$$\min\left\{\sum_{i=1}^{N}[y_i-(a+bx_i)]^2\right\}$$

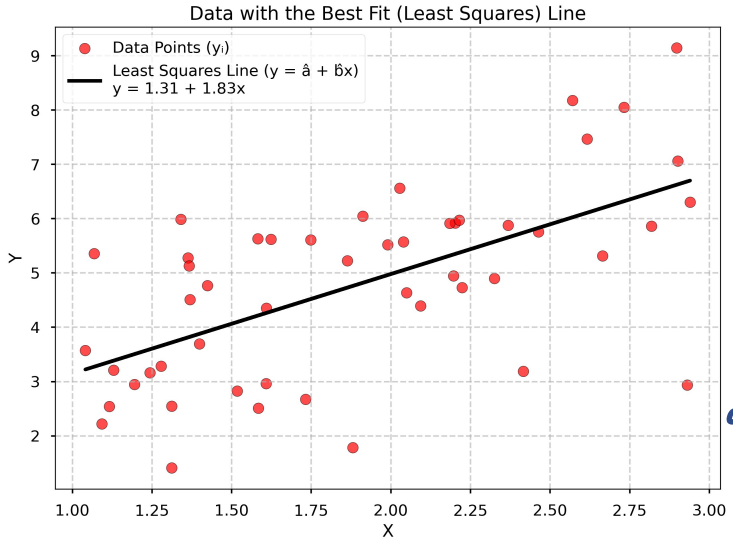
$y_i-(a+bx_i)=z_i$
cause $y=a+bx+z$

sum of all these blue line, squared

we call this sum of squared residual
cause $\min\{\sum z^2\}$

For any sets of sample, like the one shown in the graph above, will have ONE sets of $\hat{a}$ and $\hat{b}$ that gives line of best fits.

In this example, $\hat{a}=1.31$ and $\hat{b}=1.83$:



There is a shortcut in solving

$$\min\left\{\sum_{i=1}^{N}[y_i-(a+bx_i)^2]\right\}$$

instead of plotting hundreds of graph and find the best $\hat{a}$ and $\hat{b}$:

$$\hat{b}=\frac{\text{cov}_{x,y}}{\text{var}_x},\quad \hat{a}=\bar{y}-\hat{b}\bar{x}$$

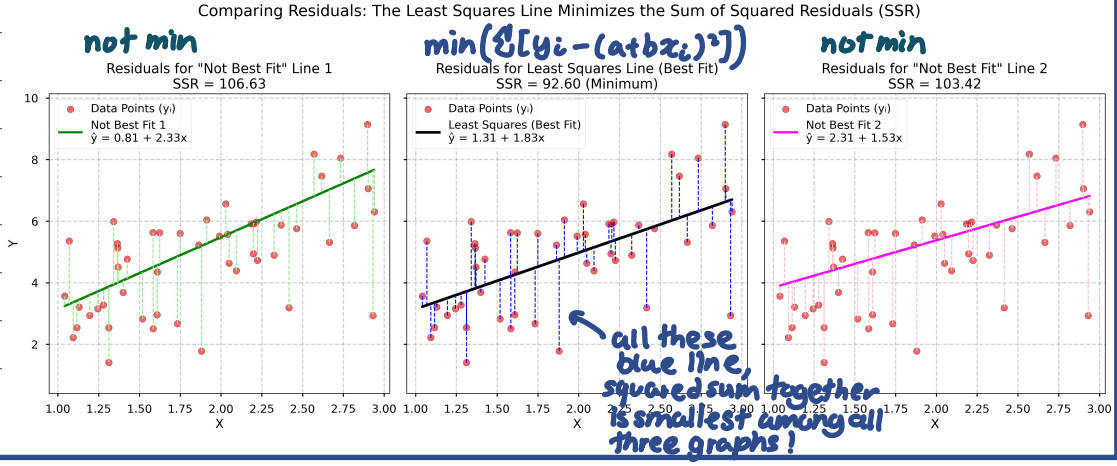
sample covariance

sample variance x

sample mean y

sample mean x

this graph shows different $\sum[y_i-(a+bx_i)^2]$ for same sample:



$\min\left\{\sum_{i=1}^{n}[y_i-(a+bx_i)^2]\right\}$ is only for determining $\hat{a}$ and $\hat{b}$ for the line of best fit. However, it doesn't tell us how good of a fit, compare to other samples.

The goodness-of fit of the regression model is measured with the **coefficient of determination** r^2:

$$r^2 = \frac{var_{\hat{a}+\hat{b}x}}{var_y} = c_{xy}^2 \leq 1$$

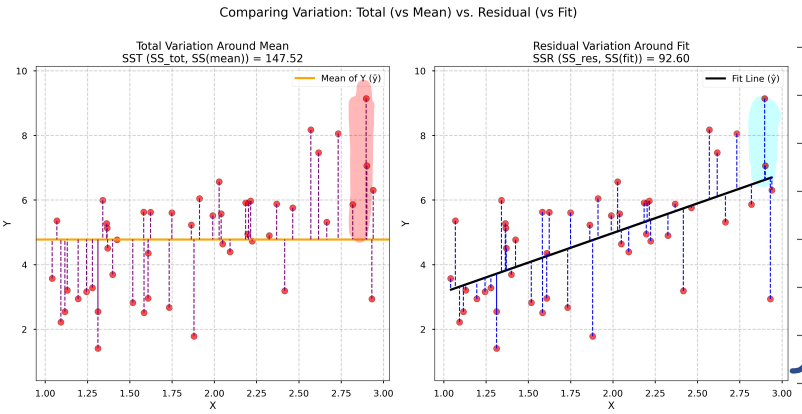
$r^2 = c_{xy}^2$ or $r^2 = \frac{cov_{xy}^2}{var_x var_y}$

JUST REMEMBER THE FORMULA

Don't confuse with correlation coefficient ($c_{xy} = \frac{cov_{xy}}{s_x s_y}$)

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{a}+\hat{b}x_i-(\hat{a}+\hat{b}\bar{x}))^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = \left(\frac{cov_{xy}}{var_x}\right)^2 \frac{var_x}{var_y} = \frac{cov_{xy}^2}{var_x var_y}$$

THIS IS NOT IMPORTANT DONT NEED UNDERSTAND



this is how to visualise r^2: it's the blue line distance minus its corresponding red line distance

Basically, population parameter (eg. μ, σ^2) are unknown.
To estimate we first find (eg. $\bar{x}$ and $\hat{s}^2$)
Then refering to CT, we can find CI of μ and σ^2

9.2 Properties of Parameter Estimates

Remember in CT, we know we can find $\bar{x}$, then estimate for μ:

7.4.1 Sampling Distribution of Sample Mean.

basically we will try to use σ
If σ^2 is known, if not known, try to estimate it
ONLY IF not possible to estimate (n small) we will use t-test.

Sample Mean

Standard Deviation (Population) is known	Standard Deviation (Population) is not known.
→ if was told X is normally distributed, no matter n is big (>30) or small, $\bar{x} \sim N(\mu, \sigma^2/n)$	→ if n is small (n<30) or was told not to estimate σ, $\frac{\bar{x}-\mu}{\hat{s}_x/\sqrt{n}} \sim t(n-1)$ ☆ WHAT IS T-DISTRIBUTION? CHECK EXAMPLE 3 BELOW!
→ If n is big (>30), $\bar{x} \sim (\mu, \sigma^2/n)$	→ If n is big (n>30), and is possible to estimate σ (given dist. of $\bar{x}$), $\bar{x} \sim N(\mu, \frac{\hat{\sigma}^2}{n})$ with estimated σ

☆ $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ is equivalent to $\bar{x} \sim N(\mu, \sigma^2/n)$

☆ for year 2 — statistics, don't worry about:
- s.d. known, X is not normally dist. AND n is small. (technically we use t-test)

and we can find $\hat{s}_x^2$, the we can estimate for σ^2:

7.4.2 Sampling Distribution of Sample Variance

unlike sample mean, sample variance is more straightforward:

$$U = \frac{(n-1)\hat{s}_x^2}{\sigma^2} \sim \chi^2(n-1) \quad \text{, no matter n big or small...}$$

the estimates is just an interval, eg. 95% confidence that μ is between (5, 10) given $\bar{x}$ is 7.5

but to do this estimation, we need to know:

1. what is the distribution of the sample statistic: eg. $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$
2. we need to find the parameter of the distribution: eg. μ and $\frac{\sigma^2}{n}$

☆ In c9, we are estimating for a and b with $\hat{a}$ and $\hat{b}$
From 9.1 Introduction, we know we can find $\hat{a}$ and $\hat{b}$ but not a and b:

$$\hat{b} = \frac{cov_{x,y}}{var_x} \quad \text{and} \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

Not going into too much details (cause it's very complicated), you just need to know that:

$$\hat{a} \sim N(E(\hat{a}), Var(\hat{a})) \; ; \; \hat{b} \sim N(E(\hat{b}), Var(\hat{b}))$$

So what we need to find? $E(\hat{a})$, $Var(\hat{a})$, $E(\hat{b})$, $Var(\hat{b})$

☆ $$E(\hat{a}) = a \quad \text{and} \quad E(\hat{b}) = b$$

} we can say that $\hat{a}$ and $\hat{b}$ are unbiased estimator of a and b.
same like $\bar{x}$ because $E(\bar{x}) = \mu$.

$$Var(\hat{a}) = \left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right) Var(Z), \quad \text{and} \quad Var(\hat{b}) = \frac{1}{nvar_x} Var(Z)$$

However, Var(Z) itself is a **POPULATION PARAMETER** that needs to be estimated:

$$Var(Z) \approx \widehat{Var(Z)} = \frac{1}{n-2}\left(nvar_y - n\frac{cov_{xy}^2}{var_x}\right)$$

this is mathematically equivalent to:

$$\Sigma[y_i - (\hat{a} + \hat{b}x_i)^2]$$

☆ because of this n-2, the var_x, var_y, and $covar_{xy}$ should be population (biased)

small info on what is Var(Z) ← population parameter

Var(Z) is 'true error variance'
which is the sum of square residual divided by n, where n is size of population.

But since it is population, we can only estimate it with $var(\hat{z})$ = sum of square residual divided by n-2, where n is size of sample.

So Var(Z) is actually very related to the residual (the blue line on the graph shown just now)

9.3 Confidence Interval (Population Parameters: a and b Estimate)

$\hat{a} \sim N\left(E(\hat{a}), Var(\hat{a})\right)$, $\hat{b} \sim N\left(E(\hat{b}), Var(\hat{b})\right)$

$$\hat{a} \sim N\left(a, \left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right) Var(Z)\right), \qquad \hat{b} \sim N\left(b, \frac{Var(Z)}{nvar_x}\right).$$

both of these requires population statistic of Var(Z),

Two possible scenario:

1. If question straight away give us Var(Z): USE NORMAL DISTRIBUTION

$\frac{X - E(X)}{\sqrt{Var(X)}}$

$$\frac{\hat{a} - a}{\sqrt{Var(Z)\left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right)}} \sim N(0,1)$$

☆ straightaway use the given Var(Z)!

$$P\left(-z_{\alpha/2} < \frac{\hat{a} - a}{\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right) Var(Z)}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\hat{a} - z_{\alpha/2}\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right) Var(Z)} < a < \hat{a} + z_{\alpha/2}\sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right) Var(Z)}\right) = 1 - \alpha$$

$$\frac{\hat{b} - b}{\sqrt{Var(Z)/nvar_x}} \sim N(0,1)$$

$$P\left(-z_{\alpha/2} < \frac{\hat{b} - b}{\sqrt{Var(Z)/nvar_x}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\hat{b} - z_{\alpha/2}\sqrt{\frac{Var(Z)}{nvar_x}} < b < \hat{b} + z_{\alpha/2}\sqrt{\frac{Var(Z)}{nvar_x}}\right) = 1 - \alpha$$

2. If question does not give us Var(Z): USE T DISTRIBUTION (with estimated Var(Z))

$$\widehat{Var(Z)} = \frac{1}{n-2}\left(nvar_y - n\frac{cov_{xy}^2}{var_x}\right)$$

☆ Estimate Var(Z) with $\widehat{Var}(Z)$ first!

☆ remember: ALWAYS "n-2"!

$$\frac{\hat{a} - a}{\sqrt{\widehat{\text{Var}}(Z)\left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right)}} \sim t(n-2)$$

$$P\left(-t_{n-2,\alpha/2} < \frac{\hat{a} - a}{\sqrt{\widehat{\text{Var}}(Z)\left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right)}} < t_{n-2,\alpha/2}\right) = 1 - \alpha$$

$$P\left(\hat{a} - t_{n-2,\alpha/2}\sqrt{\widehat{\text{Var}}(Z)\left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right)} < a < \hat{a} + t_{n-2,\alpha/2}\sqrt{\widehat{\text{Var}}(Z)\left(\frac{1}{n} + \frac{\bar{x}^2}{nvar_x}\right)}\right) = 1 - \alpha$$

$$\frac{\hat{b} - b}{\sqrt{\widehat{\text{Var}}(Z)/nvar_x}} \sim t(n-2)$$

$$P\left(-t_{n-2,\alpha/2} < \frac{\hat{b} - b}{\sqrt{\frac{\widehat{\text{Var}}(Z)}{nvar_x}}} < t_{n-2,\alpha/2}\right) = 1 - \alpha$$

$$P\left(\hat{b} - t_{n-2,\alpha/2}\sqrt{\frac{\widehat{\text{Var}}(Z)}{nvar_x}} < b < \hat{b} + t_{n-2,\alpha/2}\sqrt{\frac{\widehat{\text{Var}}(Z)}{nvar_x}}\right) = 1 - \alpha$$

(An example for this 9.3 will be given at the very end of this chapter)

when told to find an estimate value for Y for a given X, should find both $E(Y|X=x_0)$ and $Y(X=x_0)$'s interval!

9.4 Confidence Interval for $E(Y|X=x_0)$ and $Y_{X=x_0}$

1. Confidence Interval for the Mean Value of Y at $X = x_0$

- **Question:** If we observed *many, many* Y values when X is *exactly* x_0, what is the likely range for the *average* of all those Y values?

understand what are we finding!

average of all those Y values: $E(Y \mid X = x_0) = a + bx_0$

$$\frac{(\hat{a} + \hat{b}x_0) - (a + bx_0)}{\sqrt{\widehat{\text{Var}}(Z)\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{nvar_x}\right)}} \sim t(n-2)$$

$$P\left(-t_{n-2,\alpha/2} < \frac{(\hat{a} + \hat{b}x_0) - (a + bx_0)}{\sqrt{\widehat{\text{Var}}(Z)\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{nvar_x}\right)}} < t_{n-2,\alpha/2}\right) = 1 - \alpha$$

$$P\left((\hat{a} + \hat{b}x_0) - t_{n-2,\alpha/2} \cdot \sqrt{\widehat{\text{Var}}(Z)\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{nvar_x}\right)} < a + bx_0 < (\hat{a} + \hat{b}x_0) + t_{n-2,\alpha/2} \cdot \sqrt{\widehat{\text{Var}}(Z)\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{nvar_x}\right)}\right) = 1 - \alpha$$

2. Prediction Interval for an Individual Value of Y at $X = x_0$

same as confidence interval

- **Question:** If we observe *one single new* Y value when X is x_0, what is the likely range for that *specific* Y value?

understand what we are finding!

new Y value when X is x_0: $Y_{(X=x_0)} = a + bx_0 + Z$

$$\frac{(a + bx_0 + Z) - (\hat{a} + \hat{b}x_0)}{\sqrt{\widehat{\text{Var}}(Z)\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{nvar_x}\right)}} \sim t(n-2)$$

$$P\left(-t_{n-2,\alpha/2} < \frac{(a + bx_0 + Z) - (\hat{a} + \hat{b}x_0)}{\sqrt{\widehat{\text{Var}}(Z)\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{nvar_x}\right)}} < t_{n-2,\alpha/2}\right) = 1 - \alpha$$

$$P\left((\hat{a} + \hat{b}x_0) - t_{n-2,\alpha/2} \cdot \sqrt{\widehat{\text{Var}}(Z)\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{nvar_x}\right)} < (a + bx_0 + Z) < (\hat{a} + \hat{b}x_0) + t_{n-2,\alpha/2} \cdot \sqrt{\widehat{\text{Var}}(Z)\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{nvar_x}\right)}\right) = 1 - \alpha$$

basically same steps as c8. Looks more complicated, but is just the same.

9.5 Statistical Testing (Hypothesis Testing)

Three different test will be examined

(a) **Test on the slope parameter** b
Most commonly this is used to test whether there is a linear relation or not. Therefore, the *null* hypothesis will be $H_0 : b = 0$. Generally, we consider

$$H_0 : b = b^* \quad \text{and} \quad H_1 : b \neq b^*$$

(so commonly $H_0: b=0$, $H_1: b\neq 0$)

(b) **Test on the correlation**
The sample correlation coefficient can be used to decide whether two variables X and Y are linearly related (i.e. population $Corr(X, Y) \neq 0$).

(c) **Test on the regression - *F-test***
This test assess the quality of the regression using the F distribution.

9.5.1 Slope Parameter Test

(a) **Test on the slope parameter** b

1.	Null Hypothesis	$H_0 : b = b^*$
2.	Alternative Hypothesis	$H_1 : b \neq b^*$
3.	Test statistic	$T = \frac{\hat{b}-b}{\sqrt{\frac{\widehat{Var(Z)}}{nvar_x}}}$
4.	Distribution of test statistic	$T \sim t(n-2)$
5.	Critical Region $\Re$	$\lvert T\rvert > t_{n-2,\alpha/2}$
6.	Evaluate T under H_0	if $\lvert t_0\rvert > t_{n-2,\alpha/2} \Rightarrow$ Reject
7.	$p - value$	$p(\lvert T\rvert > t_0)$

This can easily be transformed to one-sided test where the $alternative$ hypothesis could be $H_1 : b > b^*$

(Example for this question will be provided at the very end of this chapter)

9.5.2 Correlation Test (between X and Y)

(b) **Test on the correlation**

1.	Null Hypothesis	$H_0 : Corr(X,Y) = 0$
2.	Alternative Hypothesis	$H_1 : Corr(X,Y) \neq 0$
3.	Test statistic	$V = \frac{\sqrt{n-3}}{2} \ln\left[\frac{(1+c_{xy})(1-Corr(X,Y))}{(1-c_{xy})(1+Corr(X,Y))}\right]$
4.	Distribution of test statistic	$V \sim N(0,1)$
5.	Critical Region $\Re$	$\lvert V\rvert > z_{\alpha/2}$
6.	Evaluate V under H_0	if $\lvert v_0\rvert > z_{\alpha/2} \Rightarrow$ Reject H_0
7.	$p - value$	$p(\lvert V\rvert > v_0)$

where c_{xy} is the sample correlation coefficient.
In practice, the correlation test does not require the use of the regression coefficients.

9.5.3 Regression Test (F-test)

(c) **Test on the regression - *F-test***

1.	Null Hypothesis	$H_0 : b = 0$
2.	Alternative Hypothesis	$H_1 : b \neq 0$
3.	Test statistic	$F = \frac{ss_{reg}}{ss_{res}/n-2}$
4.	Distribution of test statistic	$F \sim F(1, n-2)$
5.	Critical Region $\Re$	$F > f_{1,n-2,\alpha}$
6.	Evaluate F under H_0	if $f_0 > f_{1,n-2,\alpha} \Rightarrow$ Reject
7.	$p - value$	$p(F > f_0)$

The test statistic can also be written as:

$$F = \frac{\sum_{i=1}^{n}\left(\widehat{a} + \widehat{b}x_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - (\widehat{a} + \widehat{b}x_i)\right)^2/(n-2)}$$

$$ss_{reg} = \sum_{i=1}^{n}\left(\widehat{a} + \widehat{b}x_i - (\widehat{a} + \widehat{b}\bar{x})\right)^2 = nvar_{\widehat{a}+\widehat{b}x}$$

$$ss_{res} = \sum_{i=1}^{n}\left(y_i - (\widehat{a} + \widehat{b}x_i)\right)^2 = (n-2)\widehat{Var(Z)}$$

$$ss_{tot} = ss_{reg} + ss_{res} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = nvar_y$$

not important, but if you want to know what is f-distribution...

If $U \sim \chi^2(v_1)$ and $V \sim \chi^2(v_2)$

$$\frac{U/v_1}{V/v_2} \sim F(v_1, v_2)$$

f-distribution is ratio of two chi-squared distribution.

eg.

4. Relation between applied force and length of a wire

Ten steel wires of diameter 0.5 mm and length 2.5 m were extended in a laboratory by applying vertical forces of varying magnitudes. The results are as follows:

Force (kg)	15	19	25	35	42	48	53	56	62	65
Length increase (mm)	1.7	2.1	2.5	3.4	3.9	4.9	5.4	5.7	6.6	7.2

a - We wish to explore the correlation between Force and Length increase. Test the hypothesis that the population correlation is equal to 0. (significant level α is not given, let us use $\alpha = 5\%$.)

(b) Test on the correlation

1.	Null Hypothesis	$H_0 : Corr(X,Y) = 0$
2.	Alternative Hypothesis	$H_1 : Corr(X,Y) \neq 0$
3.	Test statistic	$V = \frac{\sqrt{n-3}}{2} \ln\left[\frac{(1+c_{xy})(1-Corr(X,Y))}{(1-c_{xy})(1+Corr(X,Y))}\right]$
4.	Distribution of test statistic	$V \sim N(0,1)$
5.	Critical Region $\Re$	$\lvert V\rvert > z_{\alpha/2}$
6.	Evaluate V under H_0	if $\lvert v_0\rvert > z_{\alpha/2} \Rightarrow$ Reject H_0
7.	$p - value$	$p(\lvert V\rvert > v_0)$

Step 1. Define H_0 and H_1

$H_0 : Corr(X,Y) = 0$, $H_1 : Corr(X,Y) \neq 0$

Step 2. Define Test Statistic

3. Test statistic $V = \frac{\sqrt{n-3}}{2} \ln\left[\frac{(1+c_{xy})(1-Corr(X,Y))}{(1-c_{xy})(1+Corr(X,Y))}\right]$ (need to find sample's correlation coefficient)

1.2.4 Measure Relation between Datasets. (to check whether there is any LINEAR correlation between x_i and y_i)

1. SAMPLE covariance. (problem about this is $cov_{x,y}$ has units. if x and y has unit [m] then $cov_{x,y}$ has unit [m^2])

$$cov_{x,y} = \frac{1}{n-1}\sum(x-\bar{x})(y-\bar{y})$$

→ so to non-dimensionalise...

2. SAMPLE correlation coefficient.

$$c_{x,y} = \frac{cov_{x,y}}{s_x s_y}$$ bounded by: $-1 \leq c_{x,y} \leq 1$

From the sample,

$n = 10$

$\bar{x}$ (mean of Force) $= 42.0$

$\bar{y}$ (mean of Length Increase) $= 4.34$

$\sum (x-\bar{x})(y-\bar{y}) = 305.7$

$cov_{x,y} = \frac{305.7}{10-1} = 33.9667$

$s_x = 17.8823$ ($s \rightarrow$ sample's s.d.)

$s_y = 1.9156$

$c_{x,y} = \frac{33.9667}{17.8823 \times 1.9156} = 0.9916$

Test statistic, $V = \frac{\sqrt{10-3}}{2} \ln\left(\frac{(1+0.9916)(1-0)}{(1-0.9916)(1+0)}\right)$

$= 7.2341$

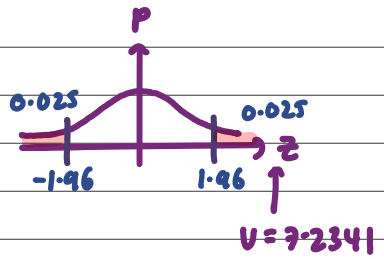
* $V = \frac{\sqrt{n-3}}{2} \ln\left[\frac{(1+c_{xy})(1-Corr(X,Y))}{(1-c_{xy})(1+Corr(X,Y))}\right]$ ← $Corr(X,Y) = 0$ cause at this stage, we assume H_0 is true!

Step 3. Define critical region

$P(Z > z_{\alpha/2}) = 0.025$

$P(Z < z_{\alpha/2}) = 0.975$

$z_{\alpha/2} = 1.96$



$V = 7.2341$

$\therefore V = 7.2341 > z_{\alpha/2} = 1.96$

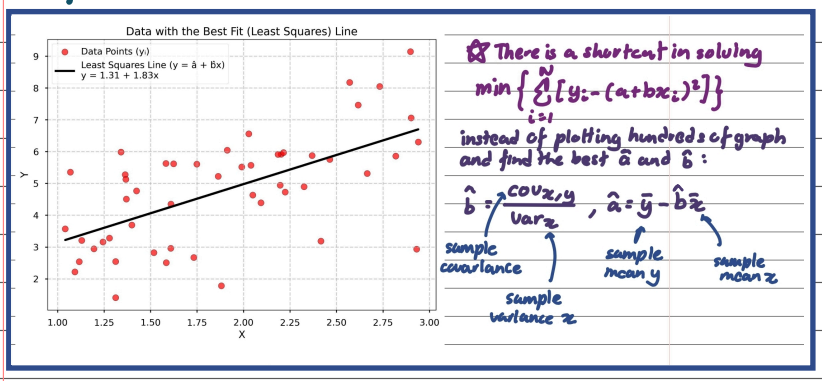
Reject $H_0 \rightarrow Corr(X, Y) \neq 0$

(find $\hat{a}$ and $\hat{b}$)

b - Estimate the parameters of a simple linear regression model with Force as an explanatory variable, indicating the coefficient of determination (find r^2)

independent variable (x)

To find $\hat{a}$ and $\hat{b}$:



$\hat{b} = \frac{cov_{x,y}}{var_x} = \frac{33.9667}{319.7767} = 0.1062$

$\hat{a} = \bar{y} - \hat{b}\bar{x} = 4.34 - 0.1062(42.0) = -0.1204$

$\hat{y} = -0.1204 + 0.1062\hat{x}$

To find r^2:

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{a} + \hat{b}x_i - (\hat{a} + \hat{b}\bar{x}))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \left(\frac{cov_{xy}}{var_x}\right)^2 \frac{var_x}{var_y} = \frac{cov_{xy}^2}{var_x var_y}$$

$r^2 = \frac{33.9667^2}{319.7767 \times 3.6695} = 0.9832$

(This r^2 result indicates 98.32% of Y can be explained with a linear relationship with X)

c - Find 95% confidence limits for the two parameters

2. If question does not give us Var(Z): USE T DISTRIBUTION (with estimated Var(Z))

$$\widehat{Var}(Z) = \frac{1}{n-2}\left(nvar_y - n\frac{cov_{xy}^2}{var_x}\right)$$

* Estimate Var(Z) with $\widehat{Var}(Z)$ first!

* remember: ALWAYS "n-2"!

$$\frac{\hat{b}-b}{\sqrt{\widehat{Var}(Z)/nvar_x}} \sim t(n-2)$$

$$\frac{\hat{a}-a}{\sqrt{\widehat{Var}(Z)\left(\frac{1}{n}+\frac{\overline{x^2}}{nvar_x}\right)}} \sim t(n-2)$$

$$P\left(-t_{n-2,\alpha/2} < \frac{\hat{b}-b}{\sqrt{\frac{\widehat{Var}(Z)}{nvar_x}}} < t_{n-2,\alpha/2}\right) = 1-\alpha$$

$$P\left(-t_{n-2,\alpha/2} < \frac{\hat{a}-a}{\sqrt{\widehat{Var}(Z)\left(\frac{1}{n}+\frac{\overline{x^2}}{nvar_x}\right)}} < t_{n-2,\alpha/2}\right) = 1-\alpha$$

$$P\left(\hat{b}-t_{n-2,\alpha/2}\sqrt{\frac{\widehat{Var}(Z)}{nvar_x}} < b < \hat{b}+t_{n-2,\alpha/2}\sqrt{\frac{\widehat{Var}(Z)}{nvar_x}}\right) = 1-\alpha$$

$$P\left(\hat{a}-t_{n-2,\alpha/2}\sqrt{\widehat{Var}(Z)\left(\frac{1}{n}+\frac{\overline{x^2}}{nvar_x}\right)} < a < \hat{a}+t_{n-2,\alpha/2}\sqrt{\widehat{Var}(Z)\left(\frac{1}{n}+\frac{\overline{x^2}}{nvar_x}\right)}\right) = 1-\alpha$$

Step 1. Find Var($\hat{Z}$)

$$Var(\hat{Z}) = \frac{1}{10-2}\left(10\times 7.6695 - 10\times\frac{33.9667^2}{319.7767}\right) = 0.0769$$

Step 2. Find critical value of t

$P(T > t_{n-2,\alpha/2}) = 0.025$

$P(T < t_{8,0.025}) = 0.975$

$t_{8,0.025} = 2.306$

STUDENT'S t Table

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297

Step 3. Find CI of $\hat{a}$ and $\hat{b}$

$$\frac{\hat{a}-a}{\sqrt{\widehat{Var}(Z)\left(\frac{1}{n}+\frac{\overline{x^2}}{nvar_x}\right)}} \sim t(n-2)$$

$$P\left(-t_{n-2,\alpha/2} < \frac{\hat{a}-a}{\sqrt{\widehat{Var}(Z)\left(\frac{1}{n}+\frac{\overline{x^2}}{nvar_x}\right)}} < t_{n-2,\alpha/2}\right) = 1-\alpha$$

$$P\left(\hat{a}-t_{n-2,\alpha/2}\sqrt{\widehat{Var}(Z)\left(\frac{1}{n}+\frac{\overline{x^2}}{nvar_x}\right)} < a < \hat{a}+t_{n-2,\alpha/2}\sqrt{\widehat{Var}(Z)\left(\frac{1}{n}+\frac{\overline{x^2}}{nvar_x}\right)}\right) = 1-\alpha$$

$$\text{CI of } \hat{a} \text{ at } 95\% = \left(-0.1204 \pm 2.306\sqrt{0.0769\left(\frac{1}{10}+\frac{42^2}{10\times 319.7767}\right)}\right)$$

$$= (-0.6366, 0.3958)$$

$$\frac{\hat{b}-b}{\sqrt{\widehat{Var}(Z)/nvar_x}} \sim t(n-2)$$

$$P\left(-t_{n-2,\alpha/2} < \frac{\hat{b}-b}{\sqrt{\frac{\widehat{Var}(Z)}{nvar_x}}} < t_{n-2,\alpha/2}\right) = 1-\alpha$$

$$P\left(\hat{b}-t_{n-2,\alpha/2}\sqrt{\frac{\widehat{Var}(Z)}{nvar_x}} < b < \hat{b}+t_{n-2,\alpha/2}\sqrt{\frac{\widehat{Var}(Z)}{nvar_x}}\right) = 1-\alpha$$

$$\text{CI of } \hat{b} \text{ at } 95\% = \left(0.1062 \pm 2.306\sqrt{\frac{0.0769}{10\times 319.7767}}\right)$$

$$= (0.0949, 0.1175)$$

d - Test the hypothesis that the slope is zero with significance 5%

(a) **Test on the slope parameter** b

1.	Null Hypothesis	$H_0 : b = b^*$
2.	Alternative Hypothesis	$H_1 : b \neq b^*$
3.	Test statistic	$T = \frac{\hat{b}-b}{\sqrt{\frac{\widehat{Var(Z)}}{nvar_x}}}$
4.	Distribution of test statistic	$T \sim t(n-2)$
5.	Critical Region $\Re$	$\lvert T\rvert > t_{n-2,\alpha/2}$
6.	Evaluate T under H_0	if $\lvert t_0\rvert > t_{n-2,\alpha/2} \Rightarrow$ Reject
7.	$p - value$	$p(\lvert T\rvert > t_0)$

This can easily be transformed to one-sided test where the *alternative* hypothesis could be $H_1 : b > b^*$

Step 1. Define H_0 and H_1

$H_0 : b = 0$

$H_1 : b \neq 0$

Step 2. Define Test Statistic

$$T = \frac{\hat{b} - b}{\sqrt{\frac{Var(\hat{z})}{nvar_x}}}$$

← at this stage, assume H_0 is true, $b=0$!

$$= \frac{0.1062}{\sqrt{\frac{0.0769}{10 \times 319.7767}}}$$

$$= 21.6563$$

Step 3. Find Critical Region

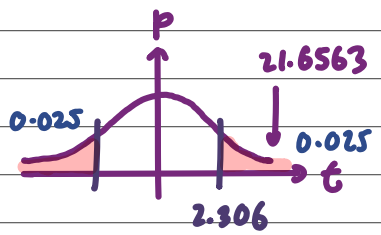
$P(T > t_{n-2,\alpha/2}) = 0.025$

$P(T < t_{8,0.025}) = 0.975$

$t_{8,0.025} = 2.306$

STUDENT'S t Table

ν	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297



reject H_0 as $T = 21.6563 > t_{8,0.025} = 2.306$

The slope is not zero!