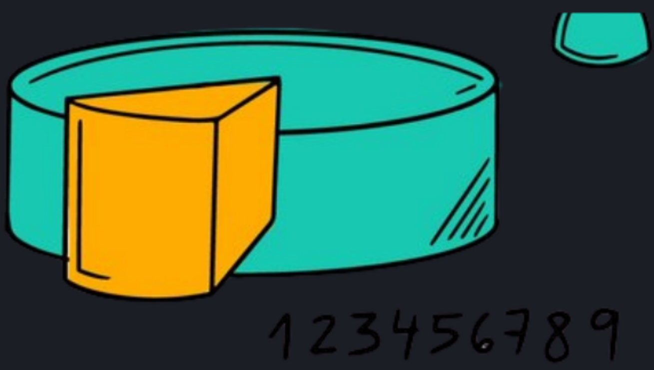
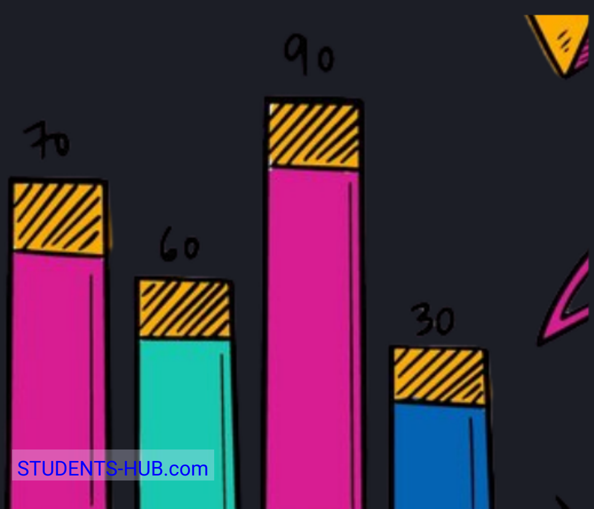
STATISTICS

By Rawan Alfares





Recorded lec 1

ch4 elementary statistics

We take a random Sample $(X_1, X_2, X_3, \ldots, X_n)$ of size (n) from a population. Some of Sample properties :-

1. Sample mean $\hat{\mu}_x$ is defined as :-

$$\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^{n} x_i$$

2. Sample variance $\hat{\sigma}_x^2$

→ when μ_x is known (1)
$$\hat{\sigma}_x^2 = \frac{1}{n} \sum (x_i - \mu_x)^2$$

→ when μ_x is unknown (2)
$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum (x_i - \hat{\mu}_x)^2$$

→ (3)
$$\hat{\sigma}_x^2 = \frac{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

3. Sample standard diviation :- $\hat{\sigma}_x = \sqrt{\hat{\sigma}_x^2}$

4. Sample Covariance between X and y :-

$$C_{xy} \triangleq \frac{1}{n-1} \sum (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y) = \hat{\mu}_{xy}$$

$$C_{xy} \triangleq \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n(n-1)}$$

5. Sample Correlation Coefficient

$$\hat{\rho}_{xy} = \frac{\sigma_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$$

Example 8-

X_i	75	80	90	85	95	99	70
y_i	80	78	92	83	90	95	75

$\Sigma X_i = 594$

$\Sigma y_i = 593$

Find ① $\hat{\mu}_x, \hat{\mu}_y$

② $\hat{\sigma}_x^2, \hat{\sigma}_y^2$

③ C_{xy}

④ $\hat{\rho}_{xy}$

1) $\hat{\mu}_x = \frac{594}{7} = 84.8$ | $\hat{\mu}_y = \frac{593}{7} = 84.714$

2) $\hat{\sigma}_x^2 = \frac{1}{n-1} \Sigma (x_i - \hat{\mu}_x)^2 = \frac{690.857}{6} = 115.143$

$\hat{\sigma}_y^2 = \frac{1}{n-1} \Sigma (y_i - \hat{\mu}_y)^2 = \frac{351.446}{6} = 58.574$

3) $= \frac{459.761}{6} = 76.627$

4) $= \frac{76.627}{82.138} = 0.933$

recorded lecture 2

Regression Techniques

$$\epsilon = \sum_{i=1}^{n} [y_i - (\alpha x_i + \beta)]^2$$ → least square error.

Curve بُعد الـ error عن الـ   for data fit in a staight line

→ we derevatie the equation two times, one to α the other to β.

$$\frac{\partial \epsilon}{\partial \alpha} = -2 \sum_{i=1}^{n} [y_i - (\alpha x_i + \beta)] x_i = 0 \rightarrow n\beta + \alpha \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \rightarrow ①$$

$$\frac{\partial \epsilon}{\partial \beta} = -2 \sum_{i=1}^{n} [y_i - (\alpha x_i + \beta)] = 0 \rightarrow \beta \sum_{i=1}^{n} x_i + \alpha \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \rightarrow ②$$

using Creamer's Rule

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \rightarrow Ⓐ$$

إما بتحفظ Ⓐ أو Ⓑ

$$\alpha = \frac{C_{xy}}{\hat{\sigma}_{xy}^2}, \quad \beta = \hat{\mu}_y - \alpha \hat{\mu}_x \rightarrow Ⓑ$$

Fitting a polynomial by the method of least squares

$y = \beta_1 + \beta_2 x + \beta_3 x^2$ → * نشتقها 3 مرات كل مرة لمتغير مختلف.

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$ → حفظ

تذكير:- احنا بنحل عشان نوجد α و β.

example:

x : Time (s)	y : Speed (m/s)	X_i^2	$X_i Y_i$	$X_i - \hat{\mu}_x$	$y_i - \hat{\mu}_y$
1	5.7	1	5.7		
1.3	6.3	1.69	8.19		
2.3	7.4	5.29	17.02		
3	8.4	9	25.2		
3.5	11.9	12.25	41.65		
4	13.7	16	54.8		
$\Sigma X_i = 15.1$	$\Sigma y_i = 53$	45.23	152.56		

find:-

1) $\hat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{6}(15.1) = 2.5166$

2) $\hat{\mu}_y = \frac{1}{6}(53) = 8.8333.$

3) $C_{xy} =$

4) $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} =$

$\Rightarrow$ find a linear equation $y = \alpha x + \beta$ that describes the speed of the vichle (y) vs. time (x) based on the above measurments:-

to find α and β

$$\begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}$$

$$\begin{bmatrix} 6 & 15.1 \\ 15.1 & 45.23 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} 53.4 \\ 152.56 \end{bmatrix}$$

$\alpha = 2.514$, $\beta = 2.573$

recorded lecture 3

Fitting a exponential by the method of least squares

$y = a e^{bx} \rightarrow \ln y = \ln a e^{bx}$

$\ln y = \ln a + \ln e^{bx}$

$\ln y = \ln a + bx \longrightarrow$ it seems to be linear equation

$y' = \beta' + \alpha' x \rightsquigarrow$ عند مقارنة المعادلتين نجد أنَّ

$y' = \ln y \quad , \quad \beta' = \ln a \quad , \quad \alpha' = b$

to Solve the prev. examble , to find α', β'

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \beta' \\ \alpha' \end{bmatrix} = \begin{bmatrix} \sum y_i new \\ \sum x_i \, y_i new \end{bmatrix}$$

B	C	D
yi	ynew(lny)	xi*yi new
5.7	1.740466	1.740466
6.3	1.84055	2.392715
7.4	2.00148	4.603404
8.4	2.128232	6.384695
11.9	2.476538	8.667884
13.7	2.617396	10.46958
53.4	12.80466	34.25875

$$\begin{bmatrix} 6 & 15.1 \\ 15.1 & 45.23 \end{bmatrix} \begin{bmatrix} \beta' \\ \alpha' \end{bmatrix} = \begin{bmatrix} 12.804 \\ 34.258 \end{bmatrix}$$

$\beta' = 1.4263$

$\alpha' = 0.28124$ * we want to find α and β.

$\alpha = e^{\beta'} = e^{1.42} = 4.137$

$\beta = \alpha' = 0.28$

$y = 4.137 \, e^{0.28x} \rightarrow$ المعادلة النهائية

2) find the speed at $x = 3.2$

$y = 4.137 \, e^{(0.28)(3.2)} = 10.13$

3) find the error , if the curve was $y = bx$

$E = \sum_{i=1}^{n} (y_i - b x_i)^2$

$\frac{\partial E}{\partial b} = -2 \sum_{i=1}^{n} (y_i - b x_i)(x_i) = 0 \Rightarrow b = \frac{\sum y_i x_i}{\sum x_i^2} = \frac{152.56}{45.23}$

example (5-2):-

$y = a x^{b}$, find the error for this Curve.

$y = \alpha X + \beta$

$\ln y = \ln(a x^{b}) = \ln a + b \ln x$

$y' = \alpha' + b x'$

$Y = \beta + \alpha X$

$Y = \ln y$, $X = \ln x$, $b = \alpha$, $\overset{\alpha'}{\beta = \ln a}$

$$\begin{bmatrix} 6 & \sum X_{i\,new} \\ \sum X_{i\,new} & \sum X_{i\,new}^{2} \end{bmatrix} \begin{bmatrix} \alpha' \\ b \end{bmatrix} = \begin{bmatrix} \sum Y_{i\,new} \\ \sum X_{i\,new} Y_{i\,new} \end{bmatrix}$$

$$\begin{bmatrix} 6 & 4.832943 \\ 4.832943 & 5.460749 \end{bmatrix} \begin{bmatrix} \alpha' \\ \beta \end{bmatrix} = \begin{bmatrix} 12.80466 \\ 11.219043 \end{bmatrix}$$

$6\alpha' + 4.832943\beta = 12.80466$

$4.832943\alpha' + 5.460749\beta = 11.219043$

$\alpha' = 1.669158992 = B$

$\beta = 0.5772271778 = \alpha$

$\ln a = B \Rightarrow a = e^{B} = 5.30770$

$Y = 5.30770\, X^{1.66916}$

EXAMPLE (5-3):

If $y = 1 - e^{\frac{-x^b}{a}}$

$$e^{\frac{-x^b}{a}} = 1 - y$$

$$e^{\frac{x^b}{a}} = \frac{1}{1-y} \Rightarrow \ln e^{\frac{x^b}{a}} = \ln \frac{1}{1-y}$$

$$\ln \frac{x^b}{a} = \ln \ln \frac{1}{1-y}$$

$$b \ln x - b \ln a = \ln \ln \frac{1}{1-y}$$

linear Regression

$$\beta' x' + \alpha' = y$$

$\alpha' = b \ln a$, $\beta' = b$, $x' = \ln x$, $y = \ln \ln \frac{1}{1-y}$

EXAMPLE (5-4):

If $y = \frac{L}{1 + e^{a+bx}}$

$$y + y e^{a+bx} = L$$

$$e^{a+bx} = \frac{L-y}{y} \Rightarrow a + bx = \ln\left(\frac{L-y}{y}\right)$$

$y' = \ln\left(\frac{L-y}{y}\right)$, $\beta' = a$, $\alpha' = b$

Recorded lecture 4

$y = c_1 x_1 + c_2 x_2$

$\hookrightarrow \mu_y = c_1 \mu_{x_1} + c_2 \mu_{x_2}$

$\hookrightarrow \sigma_y^2 = c_1^2 \sigma_{x_1}^2 + c_2^2 \sigma_{x_2}^2 + 2 c_1 c_2 \sigma_{x_1} \sigma_{x_2} \rho_{x_1 x_2}$

if x_1 and x_2 are indep = un carollated $\Rightarrow \rho_{x_1 x_2} = 0$

$\hookrightarrow \sigma_y^2 = c_1^2 \sigma_{x_1}^2 + c_2 \sigma_{x_2}^2$

Theorem: independent Gaussian Radom variables

- if x_1 and x_2 are gaussian distribution

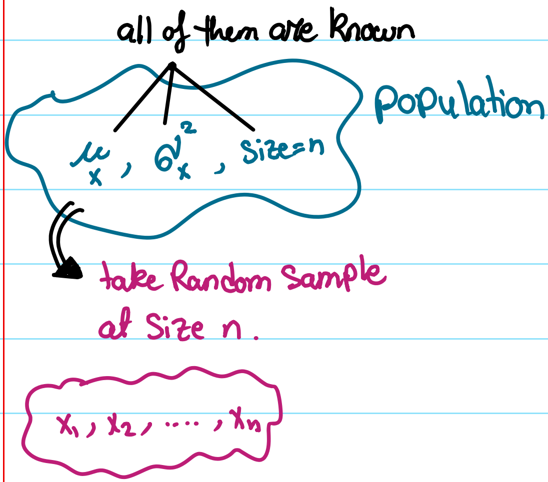
then y is also gaussian distribution.

$y = c_1 x_1 + c_2 x_2 + c_3 x_3$, find μ_y and σ_y^2.

$\mu_y = c_1 \mu_{x_1} + c_2 \mu_{x_2} + c_3 \mu_{x_3}$

$$\sigma_y^2 = c_1^2 \sigma_{x_1}^2 + c_2^2 \sigma_{x_2}^2 + c_3^2 \sigma_{x_3}^2 + \left. \begin{array}{l} 2 c_1 c_2 \sigma_{x_1} \sigma_{x_2} \rho_{x_1} \rho_{x_2} \\ + 2 c_2 c_3 \sigma_{x_2} \sigma_{x_3} \rho_{x_2} \rho_{x_3} \\ + 2 c_1 c_3 \sigma_{x_1} \sigma_{x_3} \rho_{x_1} \rho_{x_3} \end{array} \right\}$$

if x_1, x_2 and x_3 are indep then σ_y^2 is without this part. means $\rho_{x_1 x_2}, \rho_{x_2 x_3}, \rho_{x_1 x_3}$ are $\underline{0}$



examples: $\rho(\hat{\mu}_x \leq 80) \rightarrow$ we use "Central limit theorem" only for Sample mean

Assume that $y = \hat{\mu}_x = \frac{1}{n} \sum_{i=1}^{n} x_i$

$E[y] = E[\hat{\mu}_x] = \frac{1}{n} \hat{\mu}_{x_1} + \frac{1}{n} \hat{\mu}_{x_2} + \dots + \frac{1}{n} \mu_{x_n}$

$E[\hat{\mu}_x] = \mu_x \rightarrow$ ثبت

$Var[y] = \sigma_y^2 = Var[\hat{\mu}_x] = \left(\frac{1}{n}\right)^2 \sigma_{x_1}^2 + \left(\frac{1}{n}\right)^2 \sigma_{x_2}^2 + \dots + \left(\frac{1}{n}\right)^2 \sigma_{x_n}^2$

$= \cancel{n} \cdot \frac{1}{n^{\cancel{2}}} \sigma_x^2 = \boxed{\frac{\sigma_x^2}{n} = Var[\hat{\mu}_x]} \rightarrow$ حفظ

EXAMPLE (5-6):
Let X_1 and X_2 be two Gaussian random variables such that: $\mu_1 = 0$, $\sigma_1^2 = 4$, $\mu_2 = 10$, $\sigma_2^2 = 9$, $\rho_{1,2} = 0.25$. Define $Y = 2X_1 + 3X_2$

a. Find the mean and variance of Y
b. Find $P(Y \le 35)$.

a. $\mu_y = 2\mu_{x_1} + 3\mu_{x_2}$

$= (2)(0) + (3)(10) = 30$

$\sigma_y^2 = 4\sigma_{x_1}^2 + 9\sigma_{x_2}^2 + (2)C_1C_2\sigma_{x_1}\sigma_{x_2}\rho_{x_1x_2}$

$= (4)(4) + (9)(9) + (2)(2)(3)(2)(3)(0.25)$

$= 115$

$Y = 2X_1 + 3X_2$

so Also this gonna be gaussian

both are gaussian

$\Phi\left(\frac{35-30}{\sqrt{115}}\right) = \Phi(0.466) \rightarrow$ من الجدول

EXAMPLE (5-7):
Let X_1 and X_2 be two independent Gaussian random variables such that: $\mu_1 = 0$, $\sigma_1^2 = 4$, $\mu_2 = 10$, $\sigma_2^2 = 9$. Define $Y = 2X_1 + 3X_2$

c. Find the mean and variance of Y
d. Find $P(Y \le 35)$.

a. $\mu_y = 2\mu_{x_1} + 3\mu_{x_2}$

$= (2)(0) + (3)(10) = 30$

$\sigma_y^2 = 4\sigma_{x_1}^2 + 9\sigma_{x_2}^2$

$= (4)(4) + (9)(9) = 97$

b. Since its gaussian

$\Phi\left(\frac{35-30}{\sqrt{97}}\right) = \Phi(0.5077)$

EXAMPLE (5-8):
Soft-drink cans are filled by an automated filling machine. The mean fill volume is 330 ml and the standard deviation is 1.5 ml. Assume that the fill volumes of the cans are independent Gaussian random variables. What is the probability that the average volume of 10 cans selected at random from this process is less than 328 ml.

$\mu_x = 330$ ml. , $\sigma_x = 1.5$ ml , gaussian

Size = 10 → Random Sample.

(لا تقعدوا فراغاً فان الموت يطلبكم).

$\mu_{\hat{x}} = \mu_x = 330$

$\sigma_{\hat{x}}^2 = \frac{\sigma_x^2}{n} = \frac{(1.5)^2}{10} = 0.225$

$P(\hat{\mu} < 328) = \Phi\left(\frac{328-330}{\sqrt{0.225}}\right) = \Phi(-4.21) = 1 - \Phi(4.21)$ من الجدول

Recorded lecture 5

1)
$Y = X_1 + X_2 + X_3 + \dots + X_n$

$\hookrightarrow E[Y] = \mu_{X_1} + \mu_{X_2} + \mu_{X_3} + \dots + \mu_{X_n} = n\,\mu_X$

$\hookrightarrow Var[Y] = n\,\sigma_X^2$

2) $\hat{\mu}_X = Y \rightarrow$ average of the Sample $= \frac{1}{n} \sum X_i$

$\hat{\mu} = \frac{1}{n}[X_1 + X_2 + \dots + X_n]$

$\hookrightarrow E[\hat{\mu}_X] = \mu_X$

$\hookrightarrow Var[\hat{\mu}_X] = \frac{\sigma_X^2}{n}$

$\hookrightarrow STD = \frac{\sigma_X}{\sqrt{n}}$

Y Gaussian

Standard → mean $= 0$, var $= 1$

non-Standarization → $\Phi\left[\frac{\text{Value} - \mu_Y}{\sigma_Y}\right]$

EXAMPLE (5-10):

An electronic company manufactures resistors that have a mean resistance of 100 Ω and a standard deviation of 10 Ω. Find the probability that a random sample of n = 25 resistors will have an average resistance less than 95 Ω.

of the sample

$\mu = 100$, $STD = 10$, Size $= 25$

$P(\hat{\mu}_X < 95) = \Phi\left(\frac{95 - \hat{\mu}_X}{\hat{\sigma}_X}\right) = \Phi\left(\frac{95 - 100}{2}\right) = \Phi(-2.5)$

$= 1 - \Phi(2.5)$

① $\hat{\mu}_X = \mu_X = 100$

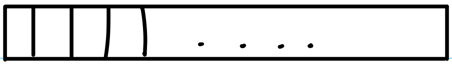
② $\hat{\sigma}_X^2 = \frac{\sigma_X^2}{n} = \frac{100}{25} = 4$

مهم

EXAMPLE (5-11):

μ_x

The lifetime of a special type of battery is a random variable with mean 40 hours and standard deviation σ 20 hours. A battery is used until it fails, then it is immediately replaced by a new one. Assume we have 25 such batteries, the lifetime of which are independent, approximate the probability that at least 1100 hours of use can be obtained.

 → total life time $25 * 40 = 1000$

$Y = X_1 + X_2 + \dots + X_{25}$

$\hookrightarrow \mu_Y = n\mu_X = (25)(40) = 1000.$

$\hookrightarrow \sigma_Y^2 = n\sigma_X^2 = (25)(400) = 10{,}000$

$\hookrightarrow \sigma_Y = 100.$

$$P(Y \geq 1100) = 1 - P(Y < 1100) = 1 - \Phi\left(\frac{1100 - 1000}{100}\right)$$
$$= 1 - \Phi(1)$$

EXAMPLE (5-12):

Suppose that the random variable X has a uniform distribution: over the interval $0 \leq X \leq 1$. A random sample of size 30 is drawn from this distribution.

a. Find the probability distribution of the sample mean $\hat{\mu}_X$

b. Find $P(\hat{\mu}_X) < 0.52$

uniform dis. $\rightarrow \mu_x = \frac{a+b}{2} = 0.5$

$\rightarrow \sigma_x^2 = \frac{(b-a)^2}{12} = 1/12$

a) $\mu_{\hat{x}} = \mu_x = 0.5$

$\sigma_{\hat{x}}^2 = \frac{\sigma_x^2}{n} = \frac{1/12}{30} = 1/360$

b) $P(\hat{\mu}_x < 0.52) = P\left(\hat{\mu}_x < \frac{0.52 - 0.5}{\sqrt{1/360}}\right) = \Phi(0.379)$

EXAMPLE (5-13):

Suppose that X is a discrete distribution which assumes the two values 1 and 0 with equal probability. A random sample of size 50 is drawn from this distribution.

a. Find the probability distribution of the sample mean $\hat{\mu}_X$

b. Find $P(\hat{\mu}_X) < 0.6$

$$pdf = f_X(x) = \begin{cases} 0.5, & x=1 \\ 0.5, & x=0 \\ 0, & o.w \end{cases}$$

$$① \ \mu_X = 0.5(1) + (0.5)(0) = 0.5$$

$$\sigma_X^2 = x\Sigma E[(X-\mu_X)^2] = (0.5)(1-0.5)^2 + (0-0.5)^2(0.5)$$

$$= 0.25$$

a) $\mu_X^{\wedge} = \mu_X = 0.5 \longrightarrow \sigma_X^{2\wedge} = \frac{\sigma_X^2}{50} = \frac{0.25}{50} = \frac{1}{200}$

b) $P(\mu_X^{\wedge} < 0.6) = \Phi\left(\frac{0.6 - 0.5}{\sqrt{1/200}}\right) = \Phi(1.414)$

ومن لمح فجر الأجر : هان عليه ظلام التكليف ،

ربنا تقبل منا إنك

أنت السميع العليم ..