# Chapter #4 Elementary Statistics

هو كمية إجمالية n مأخوذة من population كبيرة أصلية

## Basic Definitions about the random Sample:-

ما يميز الـ Sample عن الـ population هو الـ (^) or hat.

1 Sample mean :-

$$\boxed{\hat{\mu}_x = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

2 Sample Variance :- $S_x$ or $\hat{\sigma}_x^2$

$$\boxed{S_x = \hat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_x)^2}$$

$\mu_x$ is known (true mean is known).

OR

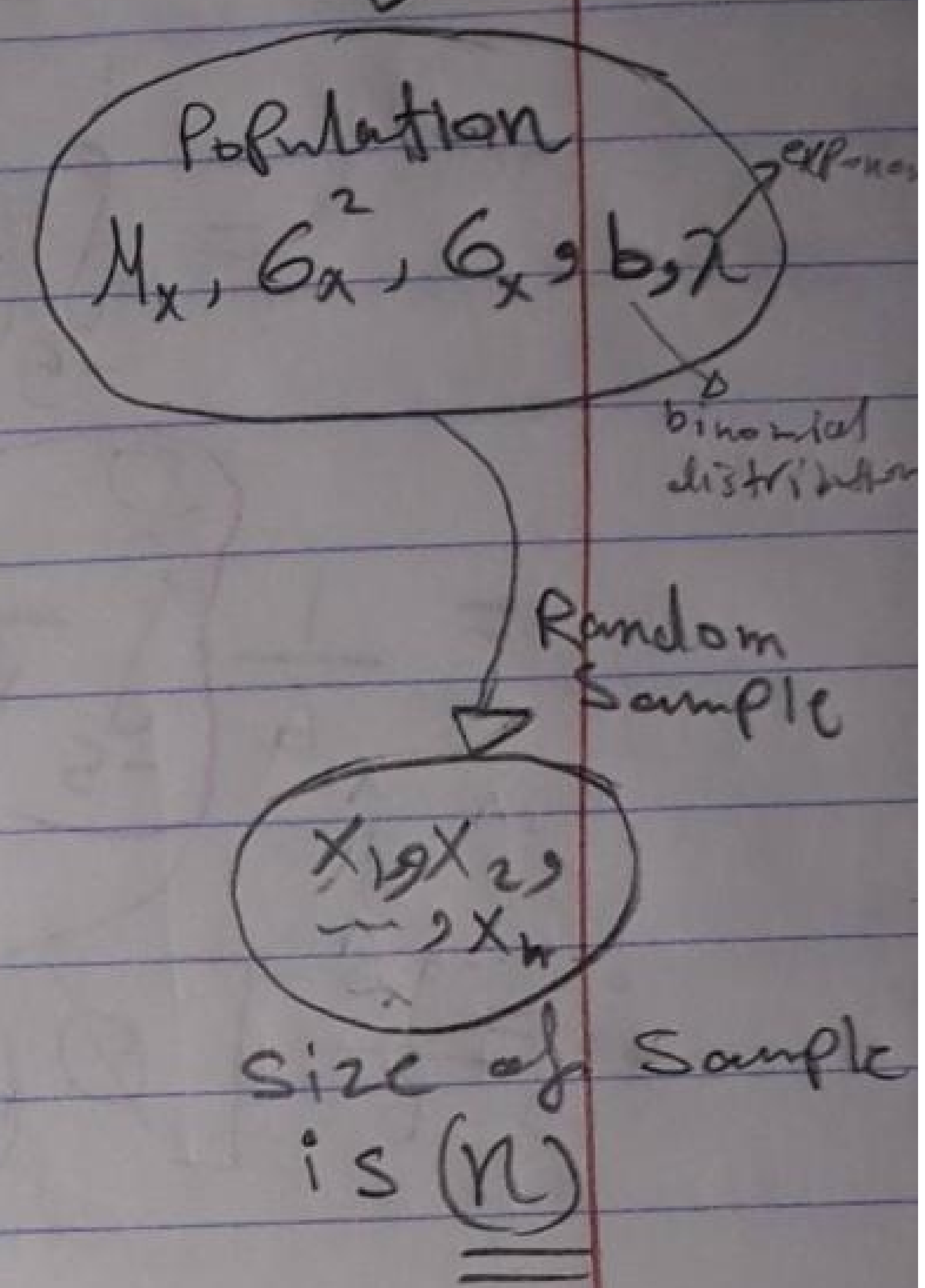
$$\boxed{S_x = \hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_x)^2}$$

True Mean is Unknown so we use $\boxed{\hat{\mu}_x}$

3 Sample Standard deviation $S_x$, $\hat{\sigma}_x$

$$\boxed{S_x = \hat{\sigma}_x = \sqrt{S_x^2} = \sqrt{\hat{\sigma}_x^2}}$$

زي ما بنعرف




إذا كان في 2 Random Variables بينهم Some Relation :-

Sample Covariance $= \hat{\mu}_{x,y} = C_{x,y}$ :-

$$\boxed{C_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}$$

Sample Correlation Coefficient $= r_{xy}$

$$\boxed{r_{xy} = \frac{C_{xy}}{S_x S_y}}$$

برضو زي ما أخدنا في الـ correlation coefficient $\rho_{xy}$ تشابتر 3.

بس هون اختلفت الرموز.

من معادلة ال variance $(\hat{S}_x^2)$ (التالية) بدنا نشتق معادلة جديدة :

$$S_x^2 = \hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)^2$$

فكينا التربيع ←

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[x_i^2 - 2x_i\hat{\mu}_x + \hat{\mu}_x^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - 2\underbrace{\hat{\mu}_x}_{\text{ثابت}}\sum_{i=1}^{n}x_i + \underbrace{\hat{\mu}_x^2}_{\text{ثابت}}\underbrace{\sum_{i=1}^{n}1}_{(1\times n)}\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - 2\hat{\mu}_x\, n\hat{\mu}_x + \hat{\mu}_x^2\, n\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - \cancel{2}n\hat{\mu}_x^2 + \cancel{n\hat{\mu}_x^2}\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - n\hat{\mu}_x^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2 - n\left(\frac{1}{n^2}\right)\left(\sum_{i=1}^{n}x_i\right)^2\right]$$

نضرب $\sum_{i=1}^{n}x_i^2$ بـ $\frac{n}{n}$ ثم نخرج $\frac{1}{n}$ عامل مشترك من المعادلة

$$\therefore\ S_x^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right]$$

← بنستخدمها إذا معيش لا $\mu_x$ ولا $\hat{\mu}_x$

أما هلأ فمن معادلة ال covariance بدي أشتق معادلة جديدة :

$$C_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[x_iy_i - x_i\hat{\mu}_y - y_i\hat{\mu}_x + \hat{\mu}_x\hat{\mu}_y\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_iy_i - \hat{\mu}_y\sum_{i=1}^{n}x_i - \hat{\mu}_x\sum_{i=1}^{n}y_i + \hat{\mu}_x\hat{\mu}_y\sum_{i=1}^{n}1\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}x_iy_i - \hat{\mu}_y\, n\hat{\mu}_x - \cancel{n\hat{\mu}_y\hat{\mu}_x} + \cancel{n\hat{\mu}_x\hat{\mu}_y}\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \cancel{n}\left(\frac{1}{\cancel{n}}\sum x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\right]$$

$$\therefore C_{xy} = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right]$$

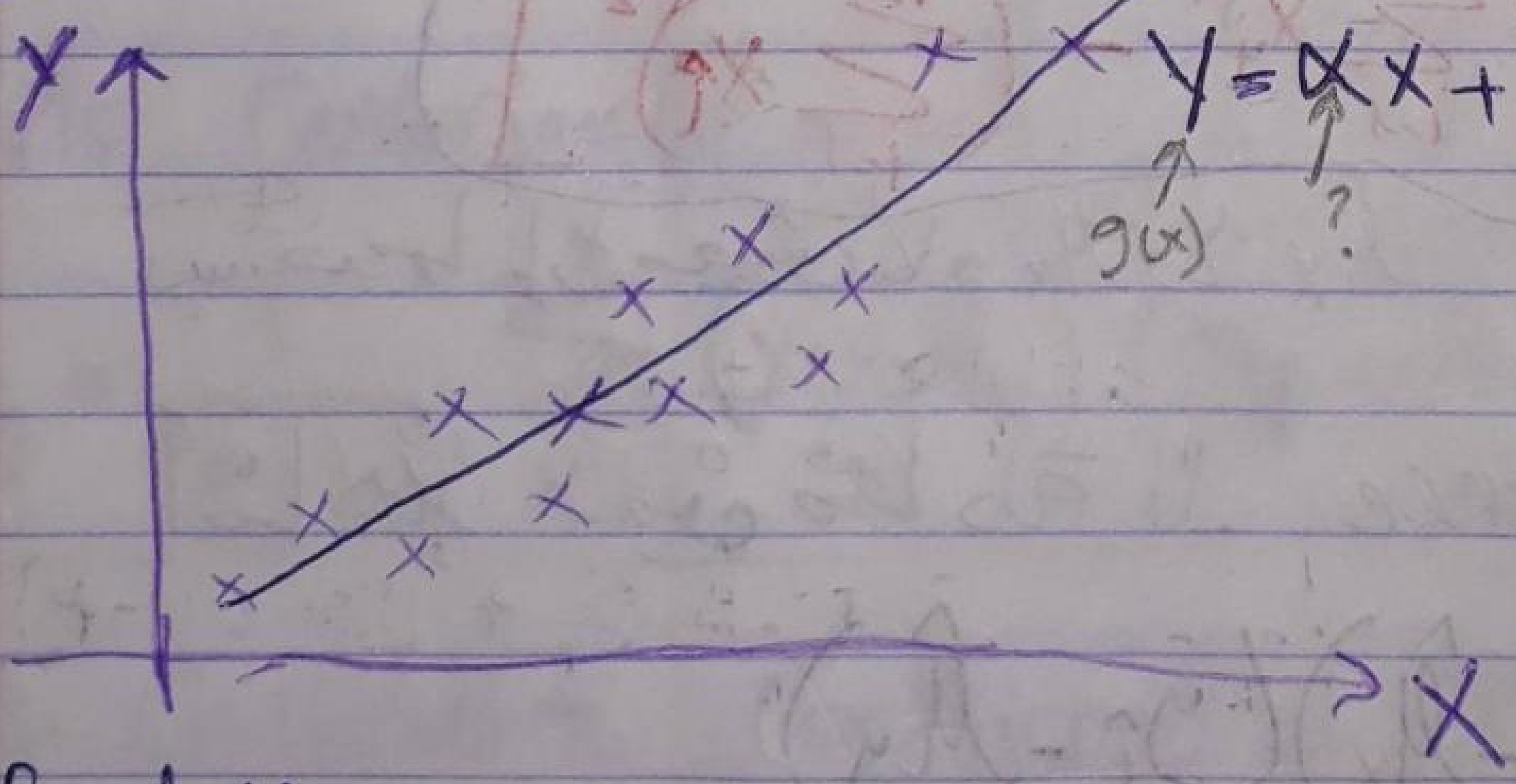
نلاحظ إنو إذا عوضت مكان y بـ x ، فبطلع نفس قانون الـ $S_x^2$
اللي اشتقيناه هلا .

## Regression Techniques :- two Random Variables إيجاد علاقة بين

| $X_i$ | $x_1$ | $x_2$ | $x_3$ | | | | $x_n$ |
|---|---|---|---|---|---|---|---|
| $y_i$ | $y_1$ | $y_2$ | $y_3$ | | | | $y_n$ ← Practical measurment |
| $Y = g(x)$ | $g(x_1)$ | $g(x_2)$ | $g(x_3)$ | | | | $g(x_n)$ ↳ theoritical |

Random Sample
$\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \ldots, \langle x_n, y_n \rangle$

So we try draw the plot out of these points to find the best $\boxed{\text{fit line}}$ passing through the points, and $\boxed{\text{Equation}}$.



$$\underset{g(x)}{Y} = \underset{?}{\alpha} X + \underset{?}{\beta}$$

to find the equation that describes the relationship between X and Y; a $\boxed{g(x) \text{ function } = Y}$ needs to be defined here.

وبدنا نحاول إنو تكون القيمة العملية أقرب مايكون للنظرية، وبالطبع رح يكون في نسبة خطأ ناتجة سواء من الـ measurments أو أي شيء آخر، و إحنا منهدف لتكون أصغر مايمكن، فبنعرف الـ Error Function :
هو الفرق بين النظري والعملي وبكون مش عند نقطة واحدة، إنما عند كل النقاط، عشان هيك حطينا $\sum$، وبعدها بناخذ الـ average وذلك بالقسمة على $n$، ولنتلاشى السوالب حطينا تربيع بالتالي :

$$E = \frac{1}{n} \sum_{i=1}^{n} \left( y_i - g(x_i) \right)^2$$ ← Mean Square error

From Equation $g(x_i) = \alpha x_i + \beta$ :-

$$E = \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \underset{?}{\alpha} x_i - \underset{?}{\beta} \right)^2$$

To find $\alpha$ and $\beta$, we derive the equation twice:

لأنهم عشان يطلع الأكبر قليل الـ minimum values

[1] In respect to $\beta$ :-

$$\frac{dE}{d\beta} = 0 \Rightarrow \frac{\cancel{1}}{n} \sum_{i=1}^{n} \cancel{2} \left( y_i - \alpha x_i - \beta \right) \underset{\text{معامل } \beta}{(-1)} = 0$$

$$\therefore -\sum y_i + \alpha \sum_{i=1}^{n} x_i + \beta n = 0$$

$$\therefore \beta n + \alpha \sum_{i=1}^{n} x_i = \sum y_i \quad \text{--- (1)}$$

نجيب جدول القيم فنبس نجمع جدول القيم

نفس الإشي هون، فضفيت بس بنعرف $\alpha$ و $\beta$.

راحوا عشان الصفر بالمعبر عن وجودهم وعدمه واحد.

[2] In respect to $\alpha$ :-

$$\frac{dE}{d\alpha} = 0 \Rightarrow \frac{\cancel{1}}{n} \sum_{i=1}^{n} \cancel{2} \left( y_i - \alpha x_i - \beta \right) \underset{\text{معامل } \alpha}{(-x_i)} = 0$$

$$-\sum_{i=1}^{n} y_i x_i + \alpha \sum_{i=1}^{n} x_i^2 + \beta \sum_{i=1}^{n} x_i = 0$$

$$\therefore \beta \sum_{i=1}^{n} x_i + \alpha \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \quad \text{--- (2)}$$

بنربط أول معادلتين مع بعض باستخدام الـ linear Algebra ← المصفوفات.

تكملة

المعادلة الأصلية :  $Y = \alpha X^1 + \beta X^0$

① :-  $$\beta n + \alpha \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

② :-  $$\beta \sum_{i=1}^{n} x_i + \alpha \sum x_i^2 = \sum_{i=1}^{n} x_i y_i$$

الإشي اللي مضروبة بـ β بالمعادلة الأولى

اللي فوق معاملة أول خطه أول

الإشي اللي مضروبة بـ α بالمعادلة ①

الإشي اللي مضروبة بـ β بمعادلة ②

الإشي اللي مضروبة بـ α بالمعادلة ②

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

ناتج معادلة ①

ناتج معادلة ②

بدلنا موقع β بالناتج

$$\beta = \frac{\begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix}}{\begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix}} = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

تكملة

$$\alpha = \frac{\begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix}}{\begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix}} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

نلاحظ أن : أرجع لمعادلة $C_{xy}$ و $S_x^2$

$$= \frac{\cancel{n(n-1)}\, C_{xy}}{\cancel{n(n-1)}\, S_x^2}$$

$$\therefore \boxed{\alpha = \frac{C_{xy}}{S_x^2}}$$

$$Y = \alpha X + \beta$$

بوزبط أعطي كمانه

$$\hat{\mu}_y = \alpha \hat{\mu}_x + \beta$$

$$\therefore \boxed{\beta = \hat{\mu}_y - \alpha \hat{\mu}_x}$$

Scanned by TapScanner

So we can rewrite the equation :- $y = \alpha X + \beta$ as follows :-

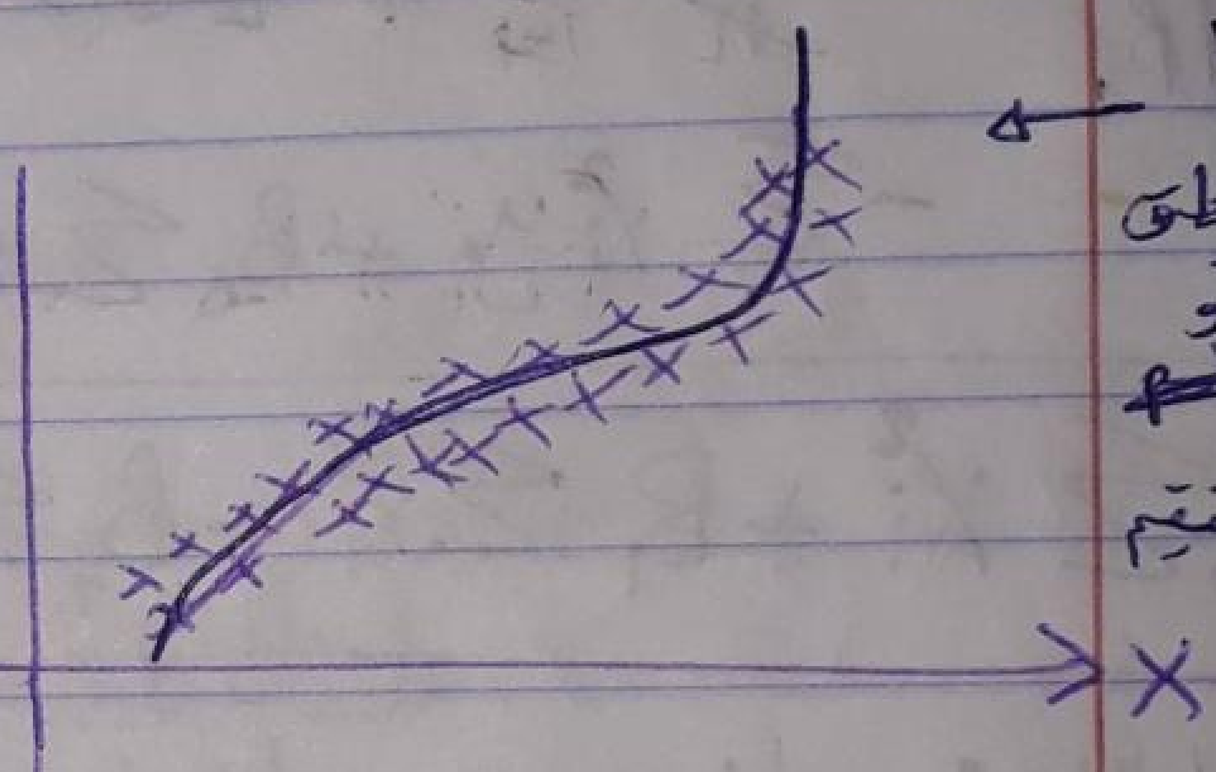
$$Y = \frac{C_{xy}}{S_x^2} X + \hat{\mu}_y - \frac{C_{xy}}{S_x^2} \hat{\mu}_x$$

$$Y - \hat{\mu}_y = \frac{C_{xy}}{S_x^2} (x - \hat{\mu}_x) \xrightarrow[\text{المعادلة على } S_y]{\text{نقسم طرفي}} \frac{Y - \hat{\mu}_y}{S_y} = \frac{C_{xy}}{S_x S_y} \left[ \frac{x - \hat{\mu}_x}{S_x} \right]$$

$$\therefore \boxed{\frac{Y - \hat{\mu}_y}{S_y} = r_{xy} \left( \frac{x - \hat{\mu}_x}{S_x} \right)}$$

## Polynomial Regression :-

$$Y = \underline{\underline{B_0}} X^0 + \underline{\underline{B_1}} X^1 + \underline{\underline{B_2}} X^2$$




إذا في هنا زي هيك نقاط فالمنطق يحكي إنو نرسم polynomial مش خط مستقيم

| | | | | |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | ------- | $x_n$ |
| $y_i$ | $y_1$ | $y_2$ | | $y_n$ |
| $g(x)$ | $g(x_1)$ | $g(x_2)$ | | $g(x_n)$ |

$$E = \frac{1}{n} \sum_{i=1}^{n} (y_i - g(x))^2$$

$$E = \frac{1}{n} \sum_{i=1}^{n} \left( y_i - [B_0 + B_1 x_i + B_2 x_i^2] \right)^2$$

$$\boxed{E = \frac{1}{n} \sum_{i=1}^{n} \left[ y_i - B_0 - B_1 x_i - B_2 x_i^2 \right]^2}$$

To find $B_0$, $B_1$ & $B_2$ $\longrightarrow$ DERIVE!

$$\frac{dE}{dB_0} = \cancel{\frac{1}{n}} \sum_{i=1}^{n} \cancel{2} [ y_i - B_0 - B_1 x_i - B_2 x_i^2 ) (-1) = 0$$

$$\Rightarrow -\sum_{i=1}^{n} y_i + n B_0 + B_1 \sum_{i=1}^{n} x_i + B_2 \sum_{i=1}^{n} x_i^2 = 0$$

$$\boxed{nB_0 + B_1\sum x_i + B_2\sum x_i^2 = \sum y_i} \quad \text{— ①}$$

Derive in Respect to $B_1$ :-

$$\frac{dE}{dB_1} = \cancel{\frac{n}{2}} \times \frac{1}{\cancel{n}} \sum_{i=1}^{n} \cancel{2}\left[y_i - B_0 - B_1 x_i - B_2 x_i^2\right](-x_i) = 0 \times \frac{n}{2}$$

$$\therefore \; -\sum x_i y_i + B_0\sum x_i + B_1\sum x_i^2 + B_2\sum x_i^3 = 0$$

$$\Rightarrow \boxed{B_0\sum x_i + B_1\sum x_i^2 + B_2\sum x_i^3 = \sum x_i y_i} \quad \text{— ②}$$

Derive in respect to $B_2$ :-

$$\frac{dE}{dB_2} = \frac{1}{\cancel{n}} \sum_{i=1}^{n} 2\left[y_i - B_0 - B_1 x_i - B_2 x_i^2\right](-x_i^2) = 0$$

$$-\sum x_i^2 y_i + B_0\sum x_i^2 + B_1\sum x_i^3 + B_2\sum x_i^4 = 0$$

$$\boxed{B_0\sum x_i^2 + B_1\sum x_i^3 + B_2\sum x_i^4 = \sum x_i^2 y_i} \quad \text{— ③}$$

Using linear algebra :-

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

وبعمل الحل باستخدام المحددات زي ما حلينا $\alpha$ و $\beta$ بالسؤال القبل .

# Fitting an Exponential by the Method of Least Squares

If we have the equation :- $Y = a e^{bx}$ , to solve it easily :-

We use the Linearization Method :-

إذن نحوّل لـ linear function :-

Take the (ln) for both sides :-

$$\ln[Y] = \ln[a e^{bx}]$$

$$= \ln[a] + \underbrace{\ln[e^{bx}]}_{bx}$$

$$\ln[Y] = \ln[a] + bx$$

$$Y_{new} = B + \alpha x$$ ⟵ نحرّكها الشمال عشان تصير خطي

$\Rightarrow$ $\boxed{B = \ln a}$ $\boxed{\alpha = b}$ $\boxed{Y_{new} = \ln[Y]}$

وبحل هاي المعادلة زي ما حلينا بالدرس السابق.

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

$$y(1+e^{a+bx}) = L \Rightarrow y + y e^{a+bx} = L$$
$$-y \qquad\qquad -y$$

$$e^{a+bx} = \frac{L-y}{y} \xrightarrow[\text{for both sides}]{\text{take (ln)}} a + bx = \underbrace{\ln\left(\frac{L-y}{y}\right)}_{y_{new}}$$

$$\boxed{y_{new} = \alpha + \beta x} \rightarrow \text{linear} \quad , \; y_{new} = \ln\left(\frac{L-y}{y}\right) \; , \alpha = a \; , \beta = B$$

Scanned by TapScanner

# Central Limit Theorem

بتحكيلي كيف بقدر أحسب probability معينة على sample مأخوذة من الpopulation

$P(\hat{\mu}_x \leq 11) = ??$

**Note:-** $Y = c_1 X_1 + c_2 X_2 + c_3 X_3$

$$\boxed{E\{Y\} = c_1 \mu_{x_1} + c_2 \mu_{x_2} + c_3 \mu_{x_3}}$$

$$Var\{Y\} = c_1^2 \sigma_{x_1}^2 + c_2^2 \sigma_{x_2}^2 + c_3^2 \sigma_{x_3}^2$$

$$+ 2 c_1 c_2 \sigma_{x_1} \sigma_{x_2} \rho_{x_1, x_2}$$ ← $X_1$ and $X_2$

$$+ 2 c_1 c_3 \sigma_{x_1} \sigma_{x_3} \rho_{x_1, x_3}$$ ← $X_1$ and $X_3$

$\rho$ → correlation coefficient

$$+ 2 c_2 c_3 \sigma_{x_2} \sigma_{x_3} \rho_{x_2, x_3}$$ ← $X_2$ and $X_3$

* if $x_1$ ~~and~~, $x_2$, and $x_3$ are S.I, then :-

$\rho_{x_1, x_2} = \rho_{x_1, x_3} = \rho_{x_2, x_3} = 0$ ← أمثلتها آخر ch3

$\boxed{Ex}$ 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 $Y = 2X_1 + 3X_2$.

$\boxed{a}$ Find the mean and variance of $Y$.

$\mu_Y = 2\mu_{x_1} + 3\mu_{x_2}$
$= 2(0) + 3(10)$
$= 30$

$\sigma_Y^2 = (2)^2 \sigma_{x_1}^2 + (3)^2 \sigma_{x_2}^2 + 2 \sigma_{x_1} \sigma_{x_2} \rho_{x_1 x_2}$
$= 4(4) + 9(9) + 2\sqrt{4}\sqrt{9}\,(0.25)$
$= 115$

$\boxed{b}$ Find $P(Y \leq 35)$. $Y$ is Gaussian ✓

$\mu_Y = 30$, $\sigma_y^2 = 115$

$\therefore \phi\left(\frac{35-30}{\sqrt{115}}\right) = \phi(0.466) = 0.6794$

$\boxed{Ex}$ 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 $Y = 2x_1 + 3x_2$.

$\boxed{\rho_{x_1,x_2} = 0}$

$\boxed{a}$ Find the mean and variance of $Y$.

$\mu_Y = 2\mu_{x_1} + 3\mu_2$

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

$= 30$

$\sigma_Y^2 = 4\sigma_{x_1}^2 + 9\sigma_{x_2}^2 + \cancel{2\sigma_{x_1}\sigma_{x_2}\rho_{x_1,x_2}}$

$= 4(4) + 9(9)$ ($16 + 81$)

$= 97$

$\boxed{b}$ $P(Y \leq 35)$ → Gaussian ✓, $\mu_Y = 30$, $\sigma_x^2 = 97$

$P(Y \leq 35) = \phi\left(\frac{35-30}{(\sqrt{97})}\right) = \phi(0.5077) = 0.6942$

$\boxed{Ex}$ Soft drink cans are filled by an automated filling machine. The mean fill volume is 330 ml, and the standard deviation is $\boxed{1.5\ ml}$ $\sigma^2$. Assume that the fill volume of the cans are $\boxed{\text{independent}}$ Gaussian $\rho_{x,x} =$ Random Variables. What is the probability that the average volume of 10 cans selected at random from this process is less than 328 ml.

$P(\hat{\mu}_x \leq 328) = ??$

→

$\mu_x = 330$ ml, $\sigma_x = 1.5$ — Population cans

$X_1, X_2, X_3, \ldots, X_{10}$ — $n = 10$

As $P(\hat{\mu}_x \leq 328) = ??$
we deal with $\hat{\mu}_x$ as if it is $Y$

$$\therefore \hat{\mu}_x = Y = \frac{1}{n} \sum X_i$$

$$= \frac{1}{n} X_1 + \frac{1}{n} X_2 + \frac{1}{n} X_3 + \ldots + \frac{1}{n} X_{10}$$

$$E\{Y\} = E\{\hat{\mu}_x\} = E\{\qquad\}$$

بنحل هادا ال — بتوزيع E

$$\therefore E\{\hat{\mu}_x\} = \frac{1}{n} E\{X_1\} + \frac{1}{n} E\{X_2\} + \frac{1}{n} E\{X_3\} + \ldots \frac{1}{n} E\{X_n\}$$

$$= \frac{1}{n}\mu_x + \frac{1}{n}\mu_x + \frac{1}{n}\mu_x + \ldots + \frac{1}{n}\mu_x$$

$$= \frac{1}{\cancel{n}}(\cancel{n})(\mu_x)$$

$$= \mu_x$$

$$= 330$$

عشان ال X هي مأخوذة من ال Population، فال E تبعتها ك X نفسها ال E الموجودة بال Population .

* بتقول ال Central Limit Theorem :- إذا باخذ Sample من ال Population فا ال expected value لل Sample Mean $E\{\hat{\mu}_x\}$ بساوي ال mean الحقيقي .

$$\boxed{E\{\hat{\mu}_x\} = \mu_x}$$

$$Var\{\hat{\mu}_x\} = \left(\frac{1}{n}\right)^2 \sigma_{x_1}^2 + \left(\frac{1}{n}\right)^2 \sigma_{x_2}^2 + \ldots + \left(\frac{1}{n}\right)^2 \sigma_{x_{10}}^2$$

نفس الإشي بما أنه كل X مأخوذة من ال Population فهي الها نفس $\sigma_x^2$

$$= \frac{1}{n^2}(n)\sigma_x^2 = \frac{\sigma_x^2}{n} = \frac{(1.5)^2}{10} = 0.225$$

$$\therefore \xrightarrow[\substack{\hat{\mu}_x = 330 \\ \hat{\sigma}_x^2 = 0.225}]{\text{Gaussian ✓}} P(\hat{\mu}_x < 328) = \phi\left(\frac{328 - 330}{\sqrt{0.225}}\right) = \phi\left(\frac{-2}{0.4743}\right)$$

$$= \phi(-4.2163)$$

وبطلع الجواب من الجدول .

**Ex 5-10** An Electric company manufactures resistors that have a mean resistence of $100\,\Omega$ and a standard deviation of $10\,\Omega$. Find the Probability that a random sample of $\boxed{n=25}$ resistors will have an average resistance less than $95\,\Omega$.

Random Sample → إزا هي Random Sample عطول لازم أعرف إنهم independent.

$\mu_x = 100\,\Omega$ , $\sigma_x = 10$ , $n = 25$

$P(\hat{\mu}_x < 95) = ??$ → هنا رح نتعامل مع $\hat{\mu}_x$ كمتغير عادي x أو y وهيك إلهم mean و Var.

Using the centeral theorem :-

$$\boxed{\underbrace{\hat{\mu}_x}_{\text{العينة}} = \underbrace{\mu_x}_{\text{pop. لل X}} = 100}$$

$$\hat{\sigma}_x^2 = \frac{\sigma_x^2}{n} = \frac{(10)^2}{25} = 4$$

$$\therefore P(\hat{\mu}_x < 95) = \phi\left(\frac{95-100}{\sqrt{4}}\right) = \phi\left(\frac{-5}{2}\right) = 0.00621$$

→ suppose it is a Gaussian ✓
$\hat{\mu}_x = 100$ ✓
$\hat{\sigma}_x^2 = 4$ ✓

**Ex** The lifetime of a special type of battery is a R.V with mean 40 hours and standard deviation 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 $\boxed{1100 \text{ hours}}$ of use can be obtained.

$\boxed{\mu_x = 40}$ $\boxed{\sigma_x = 20}$ $\boxed{n = 25}$

Let $X_1, X_2, X_3, \dots X_{25}$ be the lifetimes of batteries.
Let $Y = X_1 + X_2 + \dots + X_{25}$ be the overall lifetime of the system.
since $X_i$ are independent, Using Gaussian :-

$$\mu_Y = \mu_1 + \mu_2 + \dots \mu_{25} = 25\mu_x = 25 \times 40 = 1000$$

$$\sigma_Y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + \dots + \sigma_{x_{25}}^2 = 25(20)^2 = 10000$$

$$P(Y \geq 1100) = 1 - P(Y < 1100) = 1 - \phi\left(\frac{1100-1000}{\sqrt{10000}}\right) = 1 - \phi\left(\frac{100}{100}\right)$$

$$= 1 - \phi(1) = 0.158655$$

# Estimation of Parameters :-

ع جدول لصفحة 77 موجود في القوانين اللي أخذنا هيك .

Estimator :- is a function of the observable Sample data that is used to estimate an unknown Population parameter ($\mu_x$, $\sigma_x^2$, ...).

We consider two types of estimators:
1. Point Estimator 2. Interval Estimator.

## Point Estimator

Properties: [1] An estimator should be close to the true value of the unknown Parameters.

Def: A Point estimator ($\hat{\theta}$) is unbiased estimator of ($\theta$) if $\boxed{E(\hat{\theta}) = \theta}$.

* If the estimator is biased, then $\boxed{E(\hat{\theta}) - \theta = B}$ is called the bias of the estimator ($\hat{\theta}$)

* Let $\hat{\theta}_1, \hat{\theta}_2$ be unbiased estimators of ($\theta$), the one with the Smallest Variance is called the minimum Variance unbiased estimator (MVUE).

## In other words

$E\{\hat{\mu}_x\} = \mu_x \leftarrow \hat{\mu}_x$ is an unbiased estimator for the mean $\mu_x$.

$E\{\hat{P}\} = P \leftarrow \hat{P}$ is an unbiased estimator for the Probability of success P.

بنحسب $E\{\hat{\mu}_x\}$ حسبها عادي زي ما تعلمنا بـ chapter2 :-

$$E\{X\} \stackrel{D}{=} \sum_{x=-\infty}^{\infty} x\, P(X=x) \rightarrow \text{PMF}$$

$$\stackrel{C}{=} \int_{-\infty}^{\infty} x\, f_X(x)\, dx \rightarrow \text{pdf}$$

$\boxed{\text{Ex}}$ let $X_1$ and $X_2$ be a random sample of size two from a population with mean $\mu_X$ and variance $\sigma_X^2$. Two estimators for $\mu_X$ are proposed:-

$\mu_1 = \dfrac{X_1+X_2}{2}$ and $\mu_{X_2} = \dfrac{X_1+2X_2}{3}$. which estimator is better? and in what sense?

يعني الـ estimator الأول، مش أحسن واحد

$$E\{\mu_{X_1}\} = E\left\{\frac{X_1+X_2}{2}\right\} = E\left\{\frac{X_1}{2}\right\} + E\left\{\frac{X_2}{2}\right\} = \frac{1}{2}\mu_X + \frac{1}{2}\mu_X$$

$= \mu_X \Rightarrow$ as $E\{\hat{\mu}_{X_1}\} = \mu_1 \Rightarrow \therefore \hat{\mu}_{X_1} = \dfrac{X_1+X_2}{2}$ is unbiased.

For $\hat{\mu}_{X_2} \Rightarrow E\{\hat{\mu}_{X_2}\} = E\left\{\dfrac{X_1+2X_2}{3}\right\} = E\left\{\dfrac{X_1}{3}\right\} + \dfrac{2}{3}E\{X_2\}$

$= \dfrac{1}{3}\mu_X + \dfrac{2}{3}\mu_X = \mu_X \rightarrow \hat{\mu}_{X_2} = \dfrac{X_1+2X_2}{3}$ is unbiased

$\rightarrow$ Estimation إضافي مش 6 أشياء :- $\mu_{X_3} = \dfrac{X_1+X_2+1}{3}$

$$\therefore E\{\mu_{X_3}\} = E\left\{\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}\right\} = \frac{1}{3}E\{X_1\} + \frac{1}{3}E\{X_2\} + \frac{1}{3}$$

$= \dfrac{2}{3}\mu_X + \dfrac{1}{3} \rightarrow$ is a biased estimator

مقدار الـ biase هيك بنحسبه :-

$$\boxed{B = E\{\hat{\theta}\} - \theta}$$

$$= E\{\hat{\mu}_X\} - \mu_X$$

$$= \frac{2}{3}\mu_X + \frac{1}{3} - \mu_X = \frac{1}{3} - \frac{1}{3}\mu_X$$

$$\therefore \boxed{B = \frac{1-\mu_X}{3}}$$

Scanned by TapScanner

لما إنو طلع عندي 2 biased و 1 unbiased ، فأكيد رح أستثني الـ unbiased ، وهسا بدي أشوف أنو أفضل من بين الـ biased ، بأخذ الـ estimator اللي الـ variance تبعو أقل .

$$Var\{\hat{\mu}_{x,2}\} = Var\left\{\frac{1}{3}x_1 + \frac{2}{3}x_2\right\}$$

$$= \frac{1}{9}\sigma_x^2 + \frac{4}{9}\sigma_x^2$$

$$\boxed{\sigma^2_{\mu_{x,2}} = \frac{5}{9}\sigma_x^2}$$

حسب ما أخذنا بال central Theo. إذا كان $x_1, x_2$ SI و
$y = c_1x_1 + c_2x_2$

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

$$Var\{\mu_{x,1}\} = Var\left\{\frac{x_1 + x_2}{2}\right\}$$

$$= \frac{1}{4}\sigma_x^2 + \frac{1}{4}\sigma_x^2 = \frac{1}{2}\sigma_x^2$$

As $Var\{\hat{\mu}_{x,1}\} < Var\{\hat{\mu}_{x,2}\}$

$\therefore \hat{\mu}_{x,1}$ is the best estimator for $\mu_x$ .

**Example** Check whether the following estimator is biased or unbiased. Then try to modify the estimator to be unbiased if it is found to be biased.

$$\hat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)^2$$

$$E\{\hat{\sigma}_x^2\} = E\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)^2\right\} = E\left\{\frac{1}{n}\sum\left(x_i^2 - 2x_i\hat{\mu}_x + \hat{\mu}_x^2\right)\right\}$$

بتوزيع المجموع

$$= \frac{1}{n}E\left\{\sum_{x=1}^{n}x_i^2 - 2\hat{\mu}_x\sum_{x=1}^{n}x_i + n\hat{\mu}_x^2\right\}$$

$$= \frac{1}{n}E\left\{\sum_{x=1}^{n}x_i^2 - \cancel{2}\hat{\mu}_x\, n\hat{\mu}_x + \cancel{n\hat{\mu}_x^2}\right\}$$

حسب ما أخذنا

$$\frac{\sum x_i}{n} = \hat{\mu}_x$$ بتعريف

$$= \frac{1}{n} E\left\{\sum_{i=1}^{n} x_i^2 - n\hat{\mu}_x^2\right\}$$

$$= \frac{1}{n}\left[\sum E\{x_i^2\} - n E\{\hat{\mu}_x^2\}\right]$$

تكملة

* Note :- (1) $Var\{?\} = E\{?^2\} - (E\{?\})^2$

أعتبرها بتساوي 2 أطبع صابطة عالصفحة بجمعة مش نحتج عندما نوضحها.

$\therefore Var\{x\} = E\{x^2\} - (E\{x\})^2$

$\therefore \sigma_x^2 = E\{x^2\} - \mu_x^2$

$\therefore \boxed{E\{x^2\} = \sigma_x^2 + \mu_x^2}$ —— (1)

(2) $\therefore$ According to the Previous Note :-

$Var\{\hat{\mu}_x\} = E\{\hat{\mu}_x^2\} - (E\{\hat{\mu}_x\})^2$

According Central Limit theorem

$\frac{\sigma_x^2}{n} = E\{\hat{\mu}_x^2\} - \mu_x^2$

$\therefore \boxed{E\{\hat{\mu}_x^2\} = \frac{\sigma_x^2}{n} + \mu_x^2}$ —— (2)

$$\therefore = \frac{1}{n}\left(\sum \sigma_{x_i}^2 + \mu_x^2\right) - n\left(\frac{\sigma_x^2}{n} + \mu_x^2\right)$$

$$= \frac{1}{n}\left(n\sigma_x^2 + n\mu_x^2 - \sigma_x^2 - n\mu_x^2\right)$$

$$\therefore E\{\hat{\sigma}_x^2\} = \frac{1}{n}\left(n\sigma_x^2 - \sigma_x^2\right) = \frac{n-1}{n}\sigma_x^2 \Rightarrow \text{biased}$$

يكون طالع = $\sigma_x^2$ لكانها حسب تعريف الـ unbiased.

فنجرب بعكوس المقابل تبع $\sigma_x^2$ عشان يطلع unbiased هيك

$$\therefore E\left\{\frac{n}{n-1}\hat{\sigma}_x^2\right\} = E\left\{\frac{\cancel{n}}{n-1} * \frac{1}{\cancel{n}}\sum(x_i - \hat{\mu}_x)^2\right\} = \frac{\cancel{n}}{\cancel{n-1}} * \frac{\cancel{n-1}}{\cancel{n}}\sigma_x^2$$

$$\therefore \hat{\sigma}_{x_{new}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)^2 \Rightarrow E\{\hat{\sigma}_{x_{new}}^2\} = \sigma_x^2$$

إذا طلع العامل 20 بس نضرب الـ 19 بالمجموع وخلص فش داعي للـ 20 بالبسط عشان نشيل $\hat{\sigma}_x^2$