Engineering statistics "ENEE2307"

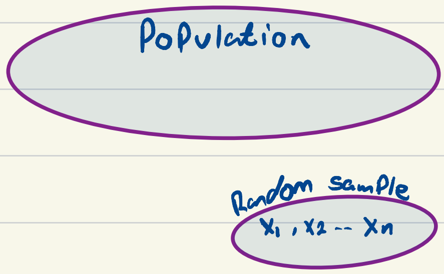
Chapter 4

By : Jibreel Bornat

Notes, questions and forms

Chapter 4
Two or more Random Variables

- it's like i have a population and i want to take a sample of size n in order to calculate something.



Note:-

anything with a hat above it like $\hat{\mu}_x$ stands for the mean of the chosen sample " The Experimental mean"

Rules :-

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

Sample Variance : $\hat{\sigma}_x^2 = S_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)^2$ True mean

نفس الشيء

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

Sample standard deviation : $\hat{\sigma}_x = S_x = \sqrt{\hat{\sigma}_x^2}$ or $\sqrt{S_x^2}$

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

Sample Correlation Coefficient : $r_{xy} = \frac{C_{xy}}{S_x S_y}$

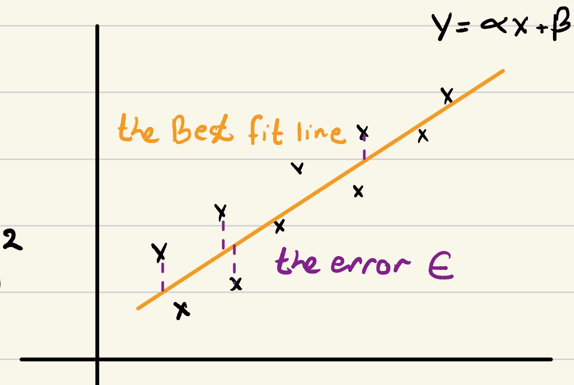
Also $\hat{\sigma}_x^2 = S_x^2 = \frac{1}{n(n-1)} \left[n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right]$

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

Regression Techniques

Error or Cost: $\epsilon = \frac{1}{n} \sum_{i=1}^{n} (y_i - y(x_i))^2$

$$= \frac{1}{n} \sum_{i=1}^{n} (y_i - \alpha x_i - \beta)^2$$



Scalter plat

Note :-

We have to differentiate ϵ with respect to α and β to get the final equation that we will solve .

$$\boxed{\beta \sum_{i=1}^{n} x_i + \alpha \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i} \text{ —— } \#$$

How Can we Solve it ?

Using Crammer Rule

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

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

(تبديل العامود الأول بعامود النواتج)

$\Longrightarrow$ بطلع الحدود OR $\beta = \hat{\mu}_y - \alpha \mu_x$

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

(تبديل العامود الثاني بعامود النواتج)

$\Longrightarrow$ بطلع الحدود OR $\alpha = \frac{C_{xy}}{\sigma_x^2}$

$$\boxed{y = \hat{\mu}_y - \frac{C_{xy}}{\sigma_x^2} \hat{\mu}_x + \frac{C_{xy}}{\sigma_x^2} x}$$

Polynomial Regression

$$y = \beta + \alpha x + \gamma x^2$$

Note:

we have to differentiate E with respect to β, α and γ to get the equation we want to fit in Crammer's rule

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \\ \gamma \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

Fitting Exponential

Example

$$y = a e^{bx} \implies \ln y = \ln(a e^{bx})$$

$$\ln y = \ln a + bx$$

$$\text{new}y = \ln a + bx$$

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

x	y	new y

بعوض قيمة y
في معادلة newy
عشان أقدر أحل

Example

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

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

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

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

$$\text{new}y = a + bx$$

The speed of a vehicle, measured at different times, is given in the table

Time(s) x	1	1.3	2.3	3	3.5	4	x
Speed (m/s) y	5.7	6.3	7.4	8.4	11.9	13.7	y

We suspect that the data follow a linear function of the form $y = \alpha x + \beta$.

a. Set up the necessary equations needed to determine α and β.

b. Solve the equations in Part a for α and β.

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

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

$$\beta = \frac{\begin{vmatrix} 53.4 & 15.1 \\ 152.56 & 45.23 \end{vmatrix}}{\begin{vmatrix} 6 & 15.1 \\ 15.1 & 45.23 \end{vmatrix}} = \frac{2415.282 - 2303.656}{271.38 - 228.01}$$

$$\beta = \frac{111.626}{43.37} \Rightarrow \boxed{\beta = 2.573}$$

$$\frac{\begin{vmatrix} 6 & 53.4 \\ 15.1 & 152.56 \end{vmatrix}}{\begin{vmatrix} 6 & 15.1 \\ 15.1 & 45.23 \end{vmatrix}} = \frac{915.36 - 806.34}{271.38 - 228.01}$$

$$\alpha = \frac{109.02}{43.37} \Rightarrow \boxed{\alpha = 2.513}$$

Question#4 [16 Points]

The data of an experiment is collected and found to be as shown in the table below. For these values find the best fitting curve equation $y = e^{ax}$ that describes the experiment results (y) versus the input (x).

x_i	-1	-1.2	-0.6	0.9	0.5	-0.1	-0.8	1	-0.2	1.3
y_i	2.7	3.3	1.8	0.4	0.6	1.1	2.2	0.3	1.2	0.2
$\ln(y_i)$	0.99	1.19	0.58	-0.916	-0.51	0.095	0.788	-1.2	0.18	-1.6

Hint: For the linear model $y = \alpha x + \beta$, we have

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

$$y = e^{ax} \Rightarrow \ln y = ax \Rightarrow \text{new } y = \ln y = ax$$

$$\begin{bmatrix} 10 & -0.2 \\ -0.2 & 7.24 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} -0.404 \\ -7.831 \end{bmatrix}$$

Delete because we don't have β in the eq.

$$7.24\alpha = -7.831 \Rightarrow \boxed{\alpha = a = -1.08}$$

Central Limit Theorem

Population

$\mu_x = 3$, $\sigma_x^2 = 2$

random sample

$X_1, X_2 \dots \dots X_n$

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

let $Y = C_1X_1 + C_2X_2 + C_3X_3$

then $\sigma_y^2 = C_1^2 X_1^2 + C_2^2 X_2^2 + C_3^2 X_3^2$

$+ 2C_1C_2\, \sigma_{x_1} \sigma_{x_2}\, \rho_{x_1,x_2}$ ← Between X_1 and X_2

$+ 2C_1C_3\, \sigma_{x_1} \sigma_{x_3}\, \rho_{x_1,x_3}$ ← Between X_1 and X_3

$+ 2C_2C_3\, \sigma_{x_2} \sigma_{x_3}\, \rho_{x_2,x_3}$ ← Between X_2 and X_3

Note :

if X_1 , X_2 , X_3 are independent $\Rightarrow \sigma_y^2 = C_1^2 X_1^2 + C_2^2 X_2^2 + C_3^2 X_3^2$

A lot of examples :

Example 1

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.

SOLUTION:

$\hat{\mu}_X$ is approximately normal with:

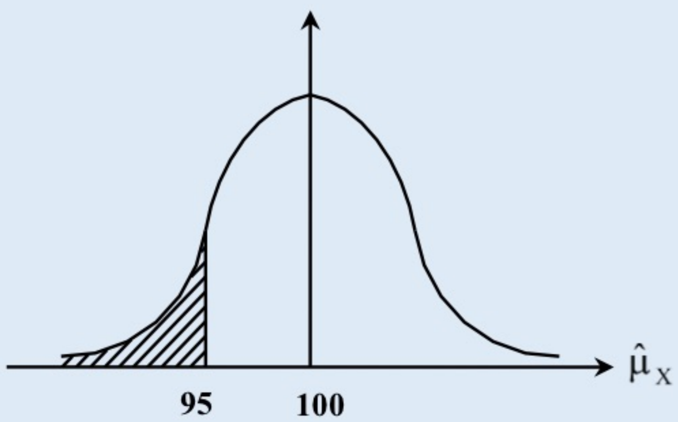
mean = $E(\hat{\mu}_X) = 100\ \Omega$.

$$\mathrm{Var}(\hat{\mu}_X) = \hat{\sigma}_X^2 = \frac{\sigma_X^2}{n} = \frac{10^2}{25}$$

$$\hat{\sigma}_X = \sqrt{\frac{\sigma_X^2}{n}} = \sqrt{\frac{10^2}{25}} = 2$$

$$P\{\hat{\mu}_X < 95\} = P\{Z < \frac{95-100}{2}\}$$

$$= \Phi(-2.5) = 0.00621$$



Example 2

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 $P(Y \leq 35)$.

SOLUTION:

$\mu_Y = 2\mu_1 + 3\mu_2 = 2(0) + 3(10) = 30$

$\sigma_Y^2 = 4\sigma_1^2 + 9\sigma_2^2 = = 4(4) + 9(9) = 97$

$$P(Y < 35) = \Phi\left(\frac{35-30}{\sqrt{97}}\right) = \Phi(0.5077) = 0.6942.$$

Example 3

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.

SOLUTION:

$\hat{\mu} = (X_1 + X_2 + \ldots X_n)/n$

$E\{\hat{\mu}\} = (\mu + \mu + \ldots \mu)/n = \mu = 330$

$Var(\hat{\mu}) = \sigma^2 / n = (1.5)^2 / 10 = 0.225$

$\hat{\mu}$ is Gaussian with mean 330 and variance 0.225.

$$P(\hat{\mu} < 328) = \Phi\left(\frac{328-330}{\sqrt{0.225}}\right) = \Phi(-4.21) = 1.2769\text{e-}005.$$

Example 4

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 $P(Y \leq 35)$.

SOLUTION:

$\mu_Y = 2\mu_1 + 3\mu_2 = 2(0) + 3(10) = 30$

$\sigma_Y^2 = 4\sigma_1^2 + 9\sigma_2^2 + 2(2)(3)(\sigma_1)(\sigma_2)\rho_{1,2} = 4(4) + 9(9) + 2(2)(3)(2)(3)(0.25) = 115$

$$P(Y < 35) = \Phi\left(\frac{35-30}{\sqrt{115}}\right) = \Phi(0.466) = 0.6794$$

Example 5

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

SOLUTION:

Since n=50 > 30, then we can approximate the sample mean by a normal distribution with:

$E(\hat{\mu}_X) = E(X) = 0*1/2 + 1*1/2 = 1/2.$

$$Var(\hat{\mu}_X) = \hat{\sigma}_X^2 = \frac{\sigma_X^2}{n} = \frac{(0-1/2)^2 * 1/2 + (1-1/2)^2 * 1/2}{50} = \frac{1}{200}$$

$$P\{\hat{\mu}_X < 0.6\} = P\{Z < \frac{0.6-0.5}{\sqrt{1/200}}\} = \Phi(1.414) = 0.92073$$

Example 6

The lifetime of a special type of battery is a random variable 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 1100 hours of use can be obtained.

SOLUTION:

Let $X_1, X_2, \ldots, X_{25}$ be the lifetimes of the batteries.

Let $Y = X_1 + X_2 + \ldots\ldots + X_{25}$ be the overall lifetime of the system

Since X_i are independent, then Y will be approximately normal with mean and variance:

$$\mu_Y = \mu_1 + \mu_2 + \ldots + \mu_{25} = 25\mu = 25 * 40 = 1000$$

$$\sigma_Y^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{25}^2 = 25\sigma_X^2 = 25 * (20)^2 = 10000$$

$$P(Y > 1100) = P(Z > \frac{1100-1000}{\sqrt{10000}}) = P(Z > 1) = 1 - \Phi(1) = 0.158655$$

وَقُل رَّبِّ زِدْنِي عِلْمًا

By : Jibreel Bornat