MOUR HAMER LEILA

Statistique

I. Statistique Descriptive :

Mesure de Tendence centrale et De Dispersion.

↳ Collecte Des Donneés ↳ traitement Des donneés collectés

↳ Interprétat des données.

Effectif : c'est le Nombre total des élts constituant cette populat° "N"

Frequence : C'est le nbre d'individus possédant ce caractère divisé par l'effectif total de la populat° : Ni

	Échantillon	Populat°
Moyenne	$\bar{x}$	μ
Variance	s^2	σ^2
Ecart-type	s	σ
Covariance	s_{xy}	σ_{xy}
corrélat°	r_{xy}	ρ_{xy}

Échantillon

Population :

Moyenne :

$$\bar{x} = \frac{\Sigma x_i}{n}$$

$$\mu = \frac{\Sigma x_i}{N}$$

Médiane : Il s'agit de la valeur centrale de l'ensemble des données, classés en ordre croissant.

Mode : Défini comme la valeur de l'observat° la plus fréquente.

Percentile : au moins p pour cent des Observat° ont cette valeur

quartile : 25e, 50e, 75e => Mesure de Tendance centrale

Mesure De Dispersion :

Etendue : égale à la Différence entre la plus grande et la plus petite valeurs.

Etendue Interquartile : (EIQ) : égale à la Différence entre le 3ème et la 1ère quartile :

Variance : basé sur les écarts au carré des Observat° par rapport à La moyenne :

$$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$$

$$\sigma^2 = \frac{\Sigma (x_i - \mu)^2}{N}$$

Ecart type :

$$s = \sqrt{s^2}$$

$$\sigma = \sqrt{\sigma^2}$$

Coefficient de variation : Mesure de Dispersion relative, égale au rapport de l'écart type à la moyenne, multiplié par 100,

$$\frac{\text{Ecart type}}{\text{Moyenne}} \times 100$$

Valeur singulière : Observat anormalement Grande ou petite.

Degré d'asymétrie : + Des données biaisées à gauche sont caractérisé par Un degré d'asy négatif

+ Des Données bi à dte sont " " " positive.

Variable centrée réduite z_i : $z_i = \frac{x_i - \bar{x}}{s}$

Variance = 4 + 100 + 4 + 4 + 144

$= \frac{256}{4} = 64$

Ecart type = $\sqrt{s} = 8$

	Ecart par rap à la Moyenne	V. de la vari centré réduite
46	0,25	0,25
54	10/8	1,25
42	-2/8	-0,25
46	2/8	0,25
32	12/8	- 1,50

Théorème de Chebyshev

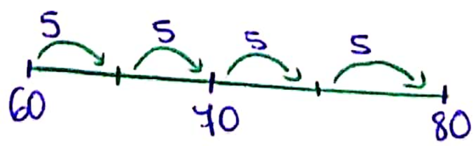
Théo utilisé pour réduire le pourcentage d'Observat qui se situe dans un Interwalle de z écart type de part et d'autre de la Moyenne.

⇒ Au Moins $\left(1 - \frac{1}{3^2}\right)$ des Observat doivent se situer au plus à $|z|$ écart type de part et d'autres de la Moyenne

càd dans l'intervalle $[\bar{x} - zs, \bar{x} + zs]$) avec $z > 1$ ».

Exple : Supposons que la Moyenne des notes de 100 Étudiants de l'ENGT soit égale à 70 et que l'écart type = 5,

1/ combien d'étudiants ont Obtenu une note entre 60 et 80 ?



$[70 - 2s ; 70 + 2s] \Rightarrow 1 - \frac{1}{2^2} = 75\%$

75 % des étudiants ont Obtenus une note entre 60 et 80.

⇒ ⚠ On peut pas l'utiliser 1s ;

Scanné avec CamScanner

Régle Empirique : Régle qui donne le % d'Observat situées dans des Intervalles 1, 2, 3 écarts type autour de la Moyenne, pour une Distribut en forme de cloche (Distribut normale).

- Environ 68% des Observat se situe dans $[\bar{x} - S ; \bar{x} + S]$
- Environ 95% des Observat se situe dans $[\bar{x} - 2S ; \bar{x} + 2S]$
- presque toutes les Observat se situent dans $[\bar{x} - 3S ; \bar{x} + 3S]$

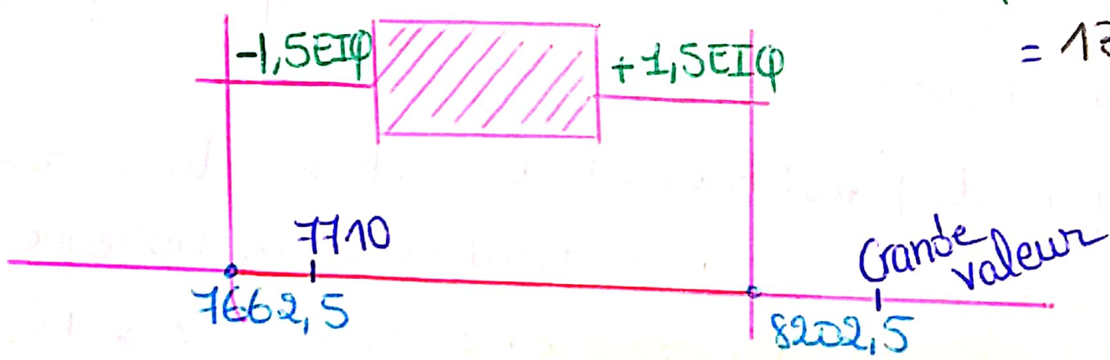
Résumé en 5 chiffres : Technique d'analyse Exploratoire des données qui utilise 5 chiffres pour résumer les données, la plus Grande valeur, le 1^{er} quartile, la médiane, le $3^{ē}$ quartile et La plus Grande valeur, Exple :

7710 - 7755 7850 7880 7880 7890 7920 7940 7950 8050 8130 8325

1/ 7710 2/ $Q_1 = 7865$ 3/ $Q_2 = 7905$

4/ 8000 5/ 8325

$EIQ = Q_3 - Q_1$
$= 135$



$Q_1 - 1{,}5EIQ = 7662{,}25$

$Q_3 + 1{,}5EIQ = 8202{,}5$

Donc 8325 est Une Valeur Singulière

II_ Statisque Bivariée :

Nuage des pts :

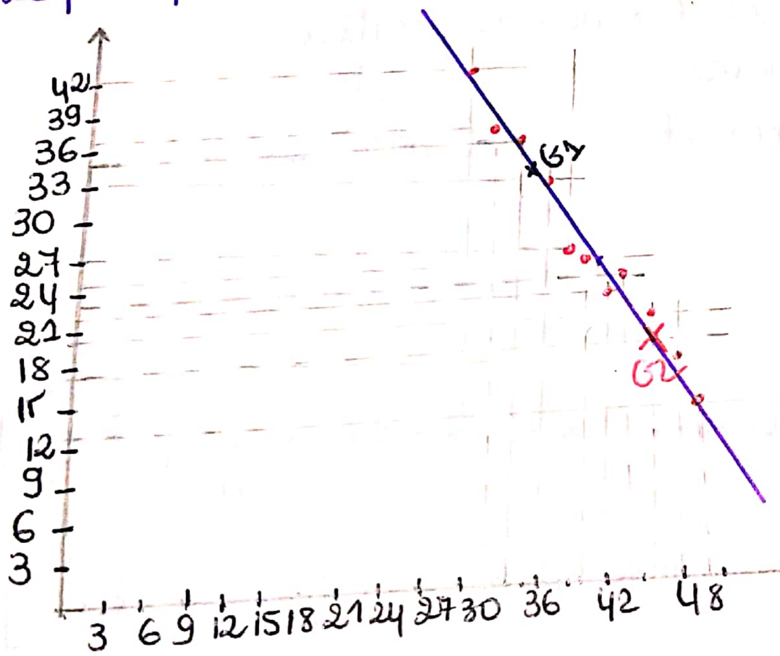
Exple : Une chaîne franchisé Désire Déterminer le prix idéal, puis il a proposé les prix Différents ... à relever le % des clts interessé par cette préstation.

Prix proposé	30	32	34	36	38	40	42	44	46	48	50
% des clts interessé	42	37	35	33	27	25	22	23	20	17	13

1/ Présenter Graphiquement puis tracer la dte d'ajustement.

2/ Le prix pour avoir 30 %. des clts interessés ?

Mthde De Mayer



$G_1 = (35 , 33,16)$ $G_2(46, 19)$

$y = ax + b$

$$a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{19 - 33,16}{46 - 35} = -1,28$$

$b = y - a x \Rightarrow 33,16 - (-1,28 \times 35) = 78,25$

$y \Rightarrow -1,28x + 78,25$

d'où $x = \dfrac{30 - 78,25}{-1,28} = 37,69 \Rightarrow$ Le prix pour avoir 30 % des Clts interessé est : 37,69 dh.

Scanné avec CamScanner

Mesure par la covariance :

Mesure de la relat linéaire entre deux variables.

Covariance populat : $\sigma_{xy} = \frac{\Sigma(x_i - \mu_x)(y_i - \mu_y)}{N}$

Covariance échantillon : $S_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1}$

Coefficient de corrélat : Mesure de la relat linéaire entre deux variables, dont les valeurs sont comprises entre −1 et +1

$r_{xy} = \frac{S_{xy}}{S_x S_y}$ ou $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

- proche de +1 ⇒ forte relat linéaire positive
- " −1 ⇒ forte relat " négative
- proche de 0 ⇒ absence de relat linéaire

Estimation ponctuelle :

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Estimation de la Moyenne

↗ σ connu $\mu \in [\bar{x} - z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} ; \bar{x} + ..]$

↘ σ inconnu $\mu \in [\bar{x} - t_{\alpha/2} \times \frac{}{\sqrt{n}} ; \bar{x} + ..]$

Distribution d'échantillonnage par $\bar{p}$

population finie → $\sqrt{\frac{N-n}{N-1}} \times \sqrt{\frac{P(1-P)}{n}}$

population infinie → $\sqrt{\frac{P(1-P)}{n}}$

→ si $\frac{N}{n} < 0{,}05$

La marge d'erreur : $n = \frac{(z_{\alpha/2})^2 \times \sigma^2}{E^2}$

Scanné avec CamScanner

Chapitre III : Regréssion linéaire simple :

$$y = \beta_0 + \beta_1 x + \varepsilon$$

y : variable à expliquer

x : variable explicative (indépendante).

B_0 et B_1 sont les paramètres du modèle

ε est une aléatoire appelée : terme D'erreur, ce terme prend en compte la variabilité de y qui n'est pas expliquée par la relation linéaire entre x et y, ce terme regroupe 3 erreures :

→ E. de spécification : le fait que la seule variable explicative n'est pas suffisante pour rendre compte de la totalité du phénomène expliqué.

→ E de Mesure : les données ne représente pas exactement le phénomène.

→ E. fluctuation d'échantillonnage : d'un échantillon à l'autre les Observations, et donc les éstimations sont légérement différentes.

$$E(y) = \beta_0 + \beta_1 x.$$

En pratique la valeur du paramètre n'est pas connue et doit être estimé en utilisant les données d'un échantillon : $\hat{y} = b_0 + b_1 x$.

Méthode Des Moindres carrées :

est une procédure qui permet d'utiliser les données de l'échantillon pour estimer l'équation de la régression (b_0 et b_1). Elle consiste à minimiser la somme des écarts au carré :

$$\min \Sigma (y_i - \hat{y}_i)^2$$

Or : $b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$ et $b_0 = \bar{y} - b_1\bar{x}$

$r = \frac{cov}{V}$

Exple : Restaurant :

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$\hat{y}_i = 60 + 5x_i$	$(y_i - \hat{y})$	$(y_i - \hat{y})^2$	$(x_i - \bar{x})^2$
2	58	-12	-72	70	-12	144	144
6	105	-8	-25	90	15	225	64
8	88	-6	-42	100	-12	114	36
8	118	-6	-12	100	18	324	32
12	117	-2	-13	120	-3	9	4
16	137	2	7	140	-3	9	4
20	157	6	27	160	-3	9	36
20	169	6	39	160	9	81	36
22	149	8	19	170	-21	441	64
26	202	12	72	190	12	144	144

$\bar{x} = \frac{140}{10} = 14$ et $\bar{y} = 130$; $b_1 = 5$; $b_0 = 60$

donc $\hat{y}_i = 60 + 5x_i$

+ $SC_{res} = \Sigma(y_i - \hat{y}_i)^2 \rightarrow$ erreur
+ $SC_{reg} = \Sigma(\hat{y}_i - \bar{y})^2 \rightarrow$ la partie expliqué

$SCT = \Sigma(y_i - \bar{y})^2$

$SCT = SC_{rés} + SC_{reg}$

Coeff détermination : $\frac{SC_{reg}}{SCT} = 1 \Rightarrow$ Meilleure situation

Coeff de corrélation : $r_{xy} = (\text{signe de } b_1)\sqrt{r^2}$

On a $r = \frac{SC_{reg}}{SCT} = \frac{14200}{15700} = 0,9027$

en d'autre terme 90,27% de la variation des ventes mensuelle peut s'expliquer par la relation linéaire de la pop et ventes, Une telle adéquation de l'équation éstimée de la reg est satisfaisante

V_ hypothèse du Modèle

H_0 = Hypothèse nulle → il faut l'a rejeté.

H_a = hyp altérnative

→ test uni infé (1)

$H_0 = \mu \geq 2$

$H_a = \mu < 2$

test uni supé (2)

$H_0 = \mu \leq 2$

$H_a = \mu > 2$

Test Bilaterle (3)

$H_0 = \mu = 2$

$H_a = \mu \neq 2$.

Seuil de signification : est la probabilité de faire une erreur de 1ère espèce lorsque l'hypothèse nulle est vraie et satisfaite avec égalité.

$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$; Rejet de H_0 si

(1) $Z \leq -Z_\alpha$

(2) $Z \geq Z_\alpha$

(3) $Z \leq -Z_{\alpha/2}$ ou $Z \geq Z_{\alpha/2}$

I_ Estimation du σ^2

$S^2 = MC_{res} = \frac{SC_{res}}{n-2} \Rightarrow S = \sqrt{\frac{SC_{res}}{n-2}}$

Espérance : $E(b_1) = B_1$ et $\sigma_{b_1} = \frac{\sigma}{\sqrt{\sum(x_i - \bar{x})^2}}$

$MC_{reg} = \frac{SC_{reg}}{\text{Nbr de V. Indép}}$

$F = \frac{MC_{reg}}{MC_{res}}$; suit une loi de Fisher avec 1 ddl en numé et $N-2$ ddl en Déno