NOTATIONS ET FORMULAIRE

PROTOCOLE SUR UN ECHANTILLON

Ensemble des n sujets de l'échantillon $S = \{ s_1 ; s_2 ; \ldots ; s_n \}$ (1)

Variable aléatoire $S \rightarrow X$ (2)

Modalité observée de X pour un sujet i : x_i (3)

Protocole: ensemble des observations $\{(s_1, x_1);(s_2, x_2);\ldots;(s_n, x_n)\}$ (4)

Protocole équipondéré: les observations ne sont pas regroupées, x_i.a un poids absolu = 1

Protocole pondéré: distribution d'effectifs, la valeur x_i est observée n_i fois (poids absolu n_i)

PROTOCOLE UNIVARIE CATEGORISE (variable à k modalités)

Effectif observé pour modalité n° k n_k (5)

Fréquence de la modalité n°k : $f_k = n_k / n$ (6)

PROTOCOLE UNIVARIE NUMERIQUE DONNE PAR INTERVALLES

Variable donnée par intervalles :

Amplitude de l'intervalle [a ; b] : ***b – a*** (7)

Densité d'un intervalle : ***densité = effectif / amplitude*** (8)

PROTOCOLE UNIVARIES NUMERIQUES DISCRETS PONDERES OU NON

Rang du quantile q_α d'ordre α (une proportion α de sujets a un score ≤ q_α)

Rang = α n + ½ (9)

Médiane :α = ½ q_1 = 1^er^ quartile α= ¼ q_3 =3^ème^ quartile α= ¾:

Poids relatif ou fréquence de la valeur x_i

$p_i = \frac{1}{n}$ en équipondéré (série de valeurs) souvent noté aussi $f_i = \frac{1}{n}$ (10)

$p_i = \frac{n_i}{n}$ en pondéré (distribution d'effectifs) souvent noté aussi $f_i = \frac{n_i}{n}$ (11)

Moyenne d'un protocole

$$m = \overline{x} = \sum_{i \in I} p_i x_i = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n \qquad (12)$$

Etendue = Max – Min (13)

Ecart inter quartile $EIQ = q_3 - q_1$ (14)

EAM $EAM = \sum_{i \in I} p_i |x_i - m|$ (15)

Contribution absolue de i à la variance

$Cta_i = p_i (x_i - m)^2$ (16)

Variance « variance population »)

$$Var = \sigma^2 = \sum_{i \in I} Cta_i = \sum_{i \in I} p_i (x_i - m)^2 \qquad (17)$$

Contribution relative à la Variance

$Ctr_i = Cta_i / Var$ (18)

Ecart-type ("Ecart-type population")

$\sigma = \sqrt{Var}$ (19)

ECART A LA MOYENNE, ECART CENTRE-REDUIT

Ecart à la moyenne de l'individu i (variable centrée)

$$\boldsymbol{E_i = x_i - m} \qquad (20)$$

Ecart-réduit de l'individu (variable centrée-réduite)

$$z_i = \frac{x_i - m}{\sigma} \qquad (21)$$

COMBINATOIRE

Coefficients binomiaux

$$C_n^p = \binom{n}{p} = \frac{n!}{p!(n-p)!} = nCp \quad \text{par la calculette} \qquad (22)$$

PROBABILITES

Probabilité fréquentiste: P(A) = Nombre de cas favorables / Nombre de cas possibles (23)

Probabilité conditionnelle $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$ (24)

Evènements indépendants $P(A \cap B) = P(A) \times P(B)$ (25)

C'est-à-dire $P(A/B) = P(A)$ (26)

TEST DE TYPICALITE – DISTRIBUTIONS D'ECHANTILLONAGE

Distribution d'échantillonnage d'une moyenne (DEM)– Propriétés

	Effectif	Moyenne	Ecart-type
POPULATION DE REFERENCE	N	μ	σ
Echantillon observé	n	m	Ety
DEM (variable M)	Eff (M)	Moy (M)	Ety (M)

Moyenne: $$Moy(M) = \mu \qquad (27)$$

(Théorème des 3 moyennes: "La moyenne de la distribution d'échantillonnage de la statistique Moyenne est égale à la moyenne parente")

Ecart-type (N fini): $$Ety(M) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad (28)$$

Ecart-type (N infini ou $\frac{n}{N}$ petit) $$Ety(M) = \frac{\sigma}{\sqrt{n}}$$ ***(erreur-type de la moyenne)*** ***(29)***

Forme – Théorème de la Limite Centrale

Pour une distribution de référence quelconque, la forme de la distribution d'échantillonnage de la moyenne se rapproche de la forme **normale** lorsque *n* croît, et cela **d'autant plus** que la distribution de référence est symétrique et est proche d'une distribution normale.

Distribution d'échantillonnage d'une proportion (ou fréquence) (DEF)– Propriétés

	Effectif	Proportion	Moyenne	Ecart-type
POPULATION DE REFERENCE	N	φ		
Echantillon observé	n	f		
DEF (variable F)	Eff (F)	F	Moy(F)	Ety (F)

Moyenne: $$Moy(F) = \varphi \qquad (30)$$

Ecart-type (N fini): $$Ety(F) = \sqrt{\frac{\varphi(1-\varphi)}{n}}\sqrt{\frac{N-n}{N-1}} \qquad (31)$$

Ecart-type (N infini ou $\frac{n}{N}$ petit) $$Ety(F) = \sqrt{\frac{\varphi\ (1-\varphi)}{n}} \qquad (32)$$

Forme

La proportion F de l'échantillon, suit approximativement une loi normale de moyenne φ et d'écart type Ety(F).

TEST Z - INFERENCE SUR UNE MOYENNE (σ CONNU)

$$Ety(M) = \frac{\sigma}{\sqrt{n}} \qquad (33)$$

$$z_{obs} = \frac{m_{obs} - \mu}{Ety(M)} \qquad (34)$$

TEST Z - INFERENCE SUR UNE FREQUENCE (OU PROPORTION)

$$Ety\ F = \sqrt{\frac{\varphi \quad (1-\varphi)}{n}} \qquad (35)$$

$$z_{obs} = \frac{f_{obs} - \varphi}{Ety \quad F} \qquad (36)$$

ESTIMATION PONCTUELLE – VARIANCE ET ECART-TYPE CORRIGES

Notations de la DEM

	Effectif	Moyenne	Variance	Ecart-type
POPULATION PARENTE	N	μ	σ^2	σ
Echantillon observé	n	m	Var	Ety
DEM (variable M)	Eff (M)	Moy (M)		Ety (M)

Estimateur sans biais de μ la moyenne parente :m la moyenne de l'échantillon (moyenne empirique) *(37)*

Estimateur sans biais de σ^2 la variance de la population parente:

la variance corrigée: $s^2 = \frac{n}{n-1} Var$ ***où Var est la variance de l'échantillon*** *(38)*

d'où l'écart-type corrigé $s = \sqrt{\frac{n}{n-1}} Ety$ ***où Ety est l'écart-type de l'échantillon*** *(39)*

INFERENCE SUR UNE MOYENNE (σ INCONNU)

On estime *σ par l'écart-type corrigé s de l'échantillon*

Si n≥300 on utilise le *test Z*

Si n <300 on utilise un test de Student avec comme degré de liberté ddl = ν = n-1

$$Ety(M) = \frac{s}{\sqrt{n}} \qquad (40)$$

$$t_{obs} = \frac{m_{obs} - \mu}{Ety(M)} \qquad (41)$$

ESTIMATION PAR INTERVALLE DE CONFIANCE DE LA MOYENNE μ [RESP. DE LA PROPORTION φ] DE LA POPULATION PARENTE

Notations de la DEM [resp. de la DEF]

Exemple:

Dire que l'on attribue au fait que μ [resp. φ] appartienne à [0,56 ; 0,64]

le niveau de confiance $1 - \alpha = 0{,}95 = 95\%$ signifie :

que pour ces valeurs de μ [resp. φ], l'échantillon est typique au seuil bilatéral $\alpha = 0{,}05$, que l'hypothèse nulle selon laquelle μ [resp φ] est dans cet intervalle ne serait pas rejetée, que le test (Z ou t selon les cas) de cette hypothèse nulle serait non significatif

Intervalle de confiance au seuil α bilatéral de la moyenne μ, **(σ connu)**

$z_{\alpha/2}$ désigne la valeur de z au seuil unilatéral α/2

$$[m - z_{\alpha/2}.\ \sigma/\sqrt{n}\ ;\ m + z_{\alpha/2}.\ \sigma/\sqrt{n}\ ] \qquad (42)$$

Intervalle de confiance au seuil α bilatéral de la proportion φ

$z_{\alpha/2}$ désigne la valeur de z au seuil unilatéral α/2

$$[f - z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\,;\, f + z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}] \qquad (43)$$

Intervalle de confiance au seuil α bilatéral de la moyenne μ, **(σ inconnu estimé par s)**

$t_{\frac{\alpha}{2},n-1}$ désigne la valeur de t au seuil unilatéral α/2, degré de liberté n-1

$$[m - t_{\alpha/2,\,n-1}.\ s/\sqrt{n}\ ;\ m + t_{\alpha/2,\,n-1}.\ s/\sqrt{n}\ ] \qquad (44)$$

INFERENCE SUR UNE REPARTITION DE FREQUENCES (OU PROPORTIONS)

Distribution d'effectifs observés : $(n_k)_{k \in K}$ sur un ensemble à *K* modalités

Effectif total $$n = \sum_{k=1 \grave{a} K} n_k \qquad (45)$$

Fréquences observées $$f_k = \frac{n_k}{n} \qquad (46)$$

Distribution de fréquences théoriques: $$(\varphi_k)_{k=1\ \grave{a}K} \qquad (47)$$

Effectifs théoriques: $$\hat{n}_k = n\,\varphi_k \qquad (48)$$

Test : $$\chi^2_{obs} = \sum_{k=1 \grave{a} K} \frac{(n_k - \hat{n}_k)^2}{\hat{n}_k} = \sum \frac{(obs - th\acute{e}o)^2}{th\acute{e}o} \qquad (49)$$

***Degré de liberté**:* *ddl = K - 1* (50)