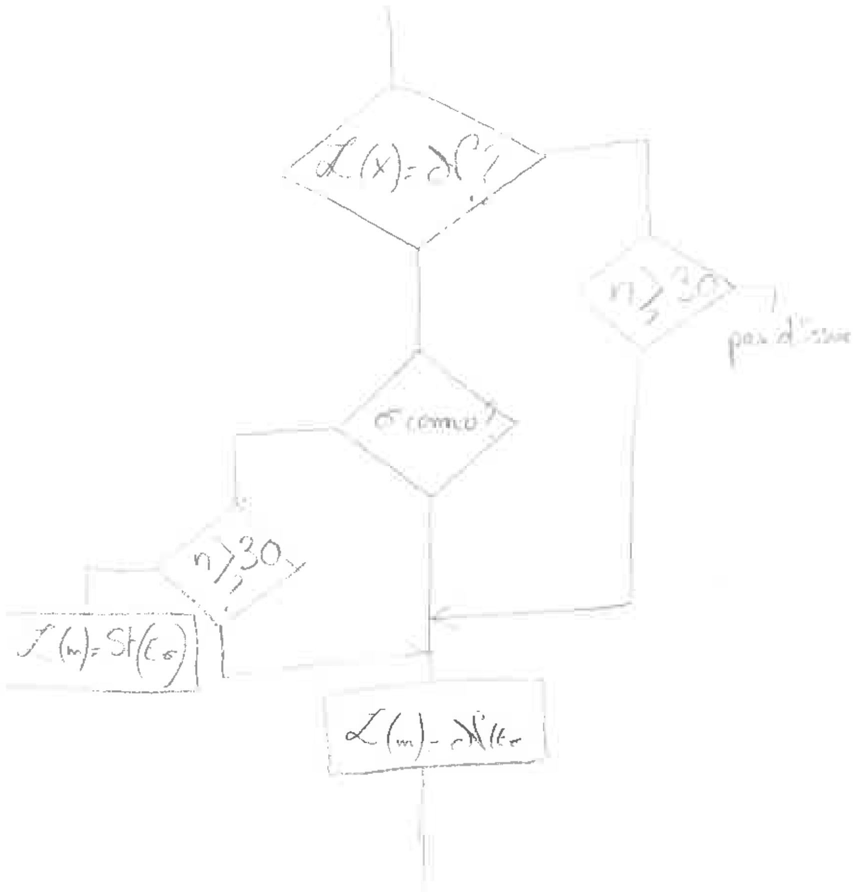
Statistiques

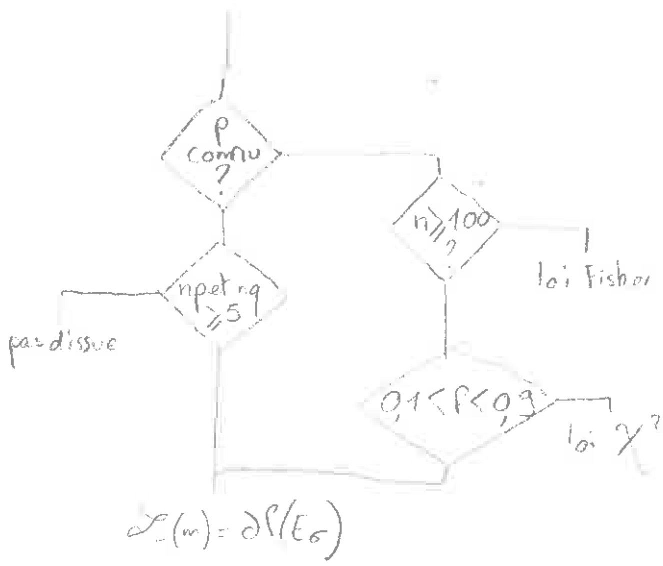
PARTIE 4
I Théorie de l'échantillonage
I 1. Distribution d'échantillonnage de la moyenne
2. Distribution d'échantillonnage des fréquences
p 6 — II Estimations
1. Définition et propriétés → ?
2. Estimations ponctuelles
3. Estimations par intervalle de confiance
3-1 Estimation d'une moyenne (variance connue)
3-2 Estimation d'une moyenne (variance inconnue)
3-3 Estimation d'une proportion
p 8 — 3-4 Estimation d'une variance (avec χ^2)
3-5 Estimation d'un écart type
3-6 Précision et taille d'un échantillon ⚠ À améliorer

p 7 — PARTIE 5 TESTS D'HYPOTHÈSE
I Généralités et définition
II Choix entre deux paramètres (= entre les paramètres de 2 pop différentes
1 Choix entre deux moyennes
2 Choix entre deux proportions
p 10 — III Comparaison d'une moyenne à une valeur donnée
1 La variance est connue
2 La variance est inconnue
IV Comparaison d'une proportion à une valeur donnée
p 14 → V Comparaison d'une variable à une valeur
VI Comparaison de 2 variances
p 16 — VII Comparaison de deux séries
p 18 → 1 Echantillons indépendants
p 19 → 2 Echantillons appariés
p 20 — VIII Comparaison de deux proportions
p 21 — IX Comparaison de plusieurs variances : Test Bartlett

QUANTITATIF



QUALITATIF



Alexia MONSAVOIR

variance
somme des carrés des écarts
SCE

Statistiques

13/1 Il serait judicieux de relire le cours de 1A, parce qu'on va faire la suite.

Programme : partie 4 : Echantillonage, estimation
partie 5 : Test → évaluation
partie 6 : ANOVA → évaluation
partie 7 : Regression multiple → évaluation

Utilisation de la calculatrice non programmable
partie 4 echantillonation

PARTIE 4 : Echantillonnage, estimations

* Nuances population / échantillon

Une population de taille N, avec une caractéristique X dont la moyenne est μ et à variance est σ_x^2

Un échantillon de taille n, caractère X, moyenne m et variance $S_x^2 = \frac{SCE_x}{n}$

→ Echantillonnage
Pop. de μ et σ_x^2 connus ⟶ on prévoit quelle sera la moyenne m.

→ Estimation
Echantillon de m connue ⟶ on suppose les valeurs de μ et σ_x^2

* Tirage avec ou sans remise

K, coef. d'exhaustivité $K = \frac{N-n}{N-1}$

Avec remise : le pop est "sans fin", on pourra tjrs repiocher
$N \to \infty$ et $\lim K = 1$

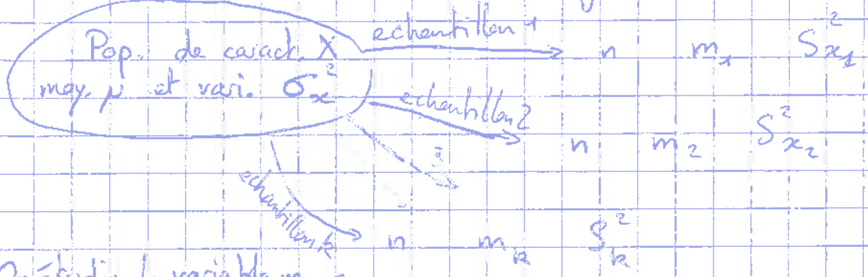
NB : si $\frac{n}{N} < 10\%$, on compte tjrs avec remise

si pas de précision → pop supposée grande → avec remise

Sans remise : $\frac{n}{N} > 10\%$ $K = \frac{N-n}{N-1}$

I.1. Distribution d'échantillonnage de la moyenne

I Théorie de l'échantillonnage

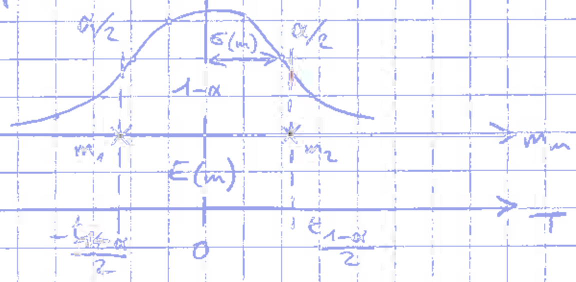


NB : tous les échantillons sont de taille n.

→ On étudie la variable m_n :
La moyenne des moyennes, c'est l'espérance des moyennes : $E(m)$
La variance des moyennes : $\sigma^2(m)$
→ On étudie la variable S_x^2 :
La moyenne des variances : $E(S^2)$
La variance des variances : $\sigma^2(S^2)$

Eh.. on peut pas échantillonner à l'infini.

Si la loi de la variable m_n suit la loi normale $N(E(m), \sigma(m))$
Alors on peut faire plein de choses sous réserve de maîtriser le cours 1A



$t = \frac{x - \bar{x}}{S_x}$ $\qquad$ $t_{\frac{1-\alpha}{2}} = \frac{m - E(m)}{\sigma(m)}$

① $P\left[E(m) - t_{\frac{1-\alpha}{2}}\,\sigma(m) < m_n < E(m) + \overbrace{t_{\frac{1-\alpha}{2}}\,\sigma(m)}^{\text{marge d'erreur absolue}}\right] = 1-\alpha$

De quoi on a besoin ?

→ $E(m) = E(x) = \mu$

→ $\sigma^2(m) = \frac{\sigma_x^2}{n} \times K$

→ $1-\alpha$ est <u>donné</u> ex : [illegible] 95 % [illegible]

NB : si un caractère suit une loi, les moyennes de ses échantillons suivent la même loi.
ex : Si $\mathcal{L}(x) = \mathcal{N}(\mu ; \sigma_x)$ alors $\mathcal{L}(m_n) = \mathcal{N}(E(m) ; \sigma(m))$

si $\mathcal{L}(X) \neq \mathcal{N}(\mu ; \sigma_x)$ alors il faut une taille d'échantillon $n \geq 30$ pour que $\mathcal{L}(m_n) = \mathcal{N}(E(m) ; \sigma(m))$

La formule ① devient :

$$P\Big[\underbrace{\mu - t_{\frac{1-\alpha}{2}} \frac{\sigma_x}{\sqrt{n}} \sqrt{K}}_{\text{limite inf.}} < m_n < \underbrace{\mu + \overbrace{t_{\frac{1-\alpha}{2}} \frac{\sigma_x}{\sqrt{n}} \sqrt{K}}^{\text{marge d'erreur ME}}}_{\text{limite sup}}\Big] = 1-\alpha$$

Ce qui signifie : $(1-\alpha)$ % des moyennes des échantillons de taille n sont comprises entre limite inf et limite sup de l'intervalle de prédiction.

On peut agir sur n ou α :

- si on augmente la taille des échantillons ($n \nearrow$), la marge d'erreur (ME) diminue. On est plus précis
- si on diminue α, l'intervalle $1-\alpha$ s'agrandit donc t augmente. On est moins précis.

13/01/2012

Démo StatisTICE
Dans une ville, divers déchets (papier, métaux, ménagers...)
Consigne : on doit estimer la qtté de déchets selon type... (en 3A)
Mais là, on étudie juste des échantillonnages

I.2 Distribution d'échantillonnage des fréquences

Une pop grande de N au caractère X qui est conforme (0) ou non conforme (1).

$k = \sum (X = 1)$

On étudie un échantillon de taille n_i, et sa fréquence f_i.
Ainsi
$f_1 = \frac{k_1}{n}$ $\quad f_2 = \frac{k_2}{n}$ $\quad f_3 = \frac{k_3}{n}$ etc.

On étudie then les $E(f)$, $\sigma^2(f)$, $\sigma(f)$

Si $\mathcal{L}(f_n) = \mathcal{N}(E(f), \sigma(f))$

alors $P\Big[\underset{\underset{p}{\parallel}}{E(f)} - t_{\frac{1-\alpha}{2}} \underset{\underset{\sqrt{\frac{pq}{n}K}}{\parallel}}{\sigma_f} < f_n < E(f) + t_{\frac{1-\alpha}{2}} \sigma_f\Big] = 1-\alpha$

II Les estimations

II 1 Définitions et propriétés

Je ne suis sûre à $\alpha = 0$

estimateur → quels paramètres je choisis dans mon échantillon pour en tirer une valeur inconnue ?

estimation → se note $\hat{\mu}$ = * valeur prise par... je sais pas, c'est bizarre... c'est une lettre mais c'est un nombre... *

Un estimateur a des propriétés : sans biais, convergent, absolument correct, efficace (cf. poly)

exemples : • sans biais → $E(m) = \mu$, m est estimateur sans biais
$E(\tilde{x}) = \mu$
$E(f) = p$, f est estimateur sans biais

• convergent → $\sigma^2(m) \searrow$ qd $n \nearrow$
$\sigma^2(\tilde{x}) \searrow$ qd $n \nearrow$
$\sigma^2(f) \searrow$ qd $n \nearrow$
$E(m) \to \mu$
$E(\tilde{x}) \to \mu$
$E(f) \to f$
→ m, $\tilde{x}$ et f sont convergents

• absolument correct = sans biais + $\sigma \searrow$ qd $n \nearrow$
→ m, $\tilde{x}$ et f

• efficace → le meilleur estimateur ... → falta

sans biais
↳ E → paramètre
convergent
↳ E → paramètre
↳ $\sigma \to 0$
absolu^{t} correct
↳ E = 0
↳ $\sigma \to 0$
Efficace
↳ E = 0
↳ $\sigma \to 0$
↳ σ minimale [...]

II 2 Estimations ponctuelle

estimation → moyenne $\hat{\mu} = m$
→ $\hat{p} = f = \frac{k}{n}$
→ $\hat{\sigma}^2_x = \frac{SCE_x}{n-1}$

NB : apparemment, $SCE_x = ns^2(x)$ ça pose pb aux élèves

II 3 Estimation par intervalle de confiance

Observations sur StatisTICE, poly

II 3-1 Estimation d'une moyenne (variance connue)

On part once again de $P[E_m - t_{\frac{1-\alpha}{2}}\, \sigma(m) < \mu < E_m + t_{\frac{1-\alpha}{2}}\, \sigma(m)] = 1-\alpha$

or $E(m) = \mu$ et comme on estime μ, on note $\hat{\mu} = m$

et $\sigma^2(m) = \frac{\sigma^2_x}{n}$ K (si σ^2_x connu) ou $\sigma^2_x = \frac{SCE}{n-1}$

On peut avoir $\mathcal{L}(m_x) = \mathcal{N}$

II 3-2 Estimation d'une moyenne (variance inconnue)

II 3-3 Estimation d'une proportion

Proportion en cas de loi normale : $P[f - t_{\frac{1-\alpha}{2}}\, \sigma(f) < p < f + t_{\frac{1-\alpha}{2}}\, \sigma(f)] = (1-\alpha)$

Application

Exercice 12

$N \to \alpha$ $n = 18$ $k = 9$ $f = \frac{9}{18} = 0,5$ $\alpha = 0,01$

On veut estimer une proportion dans un cas où $n < 0$

alors $P_1 = \dfrac{9}{9 + (18-9+1)\underbrace{F_{0,995}(20,18)}_{\text{loi de Fisher}}}$

$$P = \frac{k}{k \pm (n-k+1) F_{0,995}(\nu_1}$$

Avec les tables : $F_{0,995}(20;18) = 3,50$

$$P_1 = \frac{9}{9+(18-9+1) \times 3,50}$$

$$P_1 = 0,2045$$

$$\nu_1 = 2(18-9+1) = 20$$

$$\nu_2 = 2 \cdot 9 = 18$$

$$P_2 = \frac{(9+1) \times 3,50}{18-9+(9+1) \times 3,50} = 0,7955$$

Ça fait un encadrement $0,2045 < P < 0,7955$

Exercice 13

$N \to \alpha$ $n = 1000$ $k = 3$ $f = \frac{3}{1000} = 0,003$ $\alpha = \frac{1}{20} = 0,05$

On veut estimer une proportion dans le cas où $n > 100$ et $f < 0,1$

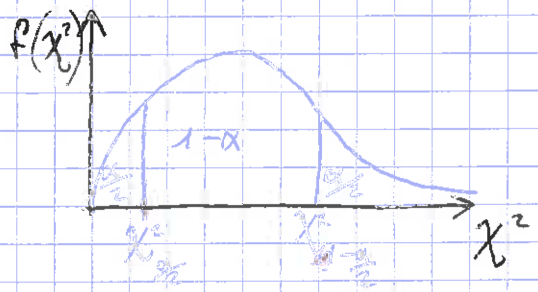
$$P_1 = \frac{1}{2 \times 1000} \times 1,24 = 0,00062 \quad \text{car } \chi^2_{0,025}(\nu) = 1,24$$

$$P_2 = \frac{1}{2 \times 1000} \times 17,5 = 0,00875 \quad \text{car } \chi^2_{0,025}(2k+2) = 17,5$$

II. 3. 4 Estimation d'une variance

La variance étant une somme de carré est forcément positive
→ pas de loi normale possible car celle-ci centrée sur 0, avec des + et −
→ variance étudiée avec χ^2

Illustrat° de la loi χ^2 :



Application

Exercice 11 et 14

$\alpha = 0,05$ $n = 20$ SCE (à la calculatrice) $= 24244,55$

$\chi^2_{0,975}(19) = 32,9$

$\chi^2_{0,025}(19) = 8,91$

Ainsi $P\left[\frac{24244,55}{32,9} < \sigma_x^2 < \frac{24244,55}{8,91}\right] = 0,95$

$\nu = 19$

$P[736,9 < \sigma_x^2 < 2721] = 0,95$

$P[\sqrt{736,9} < \sigma_x < \sqrt{2721}] = 0,95$

II - 3 - 6 - Précision et taille d'échantillon

↳ Si tu maîtrise pas, tu galères en 3A.

PARTIE 5 : Tests d'hypothèse

Quand on parle de "test", c'est qu'on compare une valeur à une autre ou bien à une valeur de référence

* Hypothèse nulle H_0 → les valeurs sont les m̂
* Hypothèse } alternative → les valeurs sont différentes
* Hypothèse } H_1 → une valeur est + grande que l'autre

<u>Approche</u> : H_0 → "innocent" → on suppose qu'il est aux normes
H_1 → "coupable" → on essaie de prouver qu'il n'est pas aux normes

risque de 1ère esp. α est le risque de se tromper en rejetant H_0,
"risque du fabiquant"

risque de 2ᵉ espèce β est le risque de se tromper en rejetant H_1
"risque du client"

NB : pas de relation entre α et β

<u>Illustration</u>

		Vérité	
		H_0 vraie	H_0 fausse
Décision	H_0 vraie	$1-\alpha$	β
	H_0 fausse	α	$1-\beta$

α : risque de première espèce
β : risque de seconde espèce
$1-\beta$ = puissance du test
$1-\alpha$: niveau de confiance

α et β peuvent varier, on va voir comment,
mais en sort, c'est imposé par le cahier des charges

II Choix entre deux paramètres

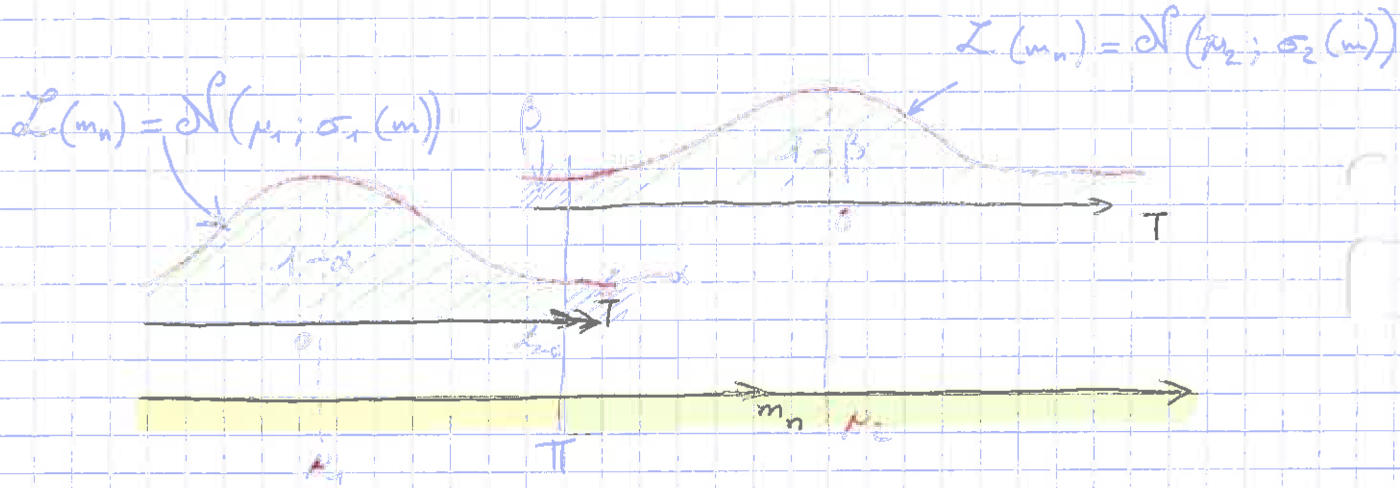
II-1- Choix entre deux moyennes

Dans une pop on étudie un échantillon n, m ... mais de quelle pop ?
→ 1. μ_1 $\sigma_1(x)$
→ 2. μ_2 $\sigma_2(x)$

On définit alors 2 cas : - H_0 : $\mu = \mu_1$

- H_1 : $\mu > \mu_1$
$\mu = \mu_2$

NB : on peut aussi étudier avec H : $\mu < \mu_1$, et $\mu = \mu_2$



Comment est-ce qu'on calcule π ?

$t_{1-\alpha} = \frac{\pi - \mu_1}{\sigma_1(m)}$ avec $\sigma_1(m) = \sqrt{\frac{\sigma_1^2(x)}{n} K}$

NB : very often $\sigma_1^2(x) = \sigma_2^2(x)$

$t_{\beta} = \frac{\pi - \mu_2}{\sigma_2(m)}$ avec $\sigma_2(m) = \sqrt{\frac{\sigma_2^2(x)}{n} K}$

↳ inconnues : α, β, n et π

: zone où on accepte H_0 et rejette H_1 avec β% de risque. $m < \pi$

: zone où on accepte H_1 et rejette H_0 avec α% de risque. $m > \pi$

Application

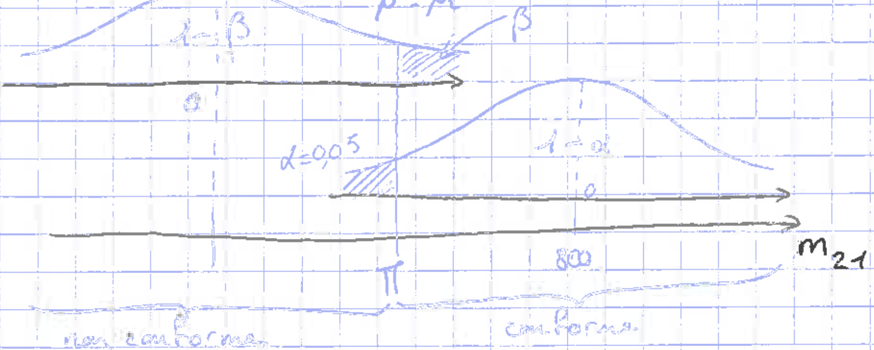
Exercice 1 $\sigma(m) = \sqrt{\frac{\sigma^2}{n}}$

Ici $\sigma_1 = \sigma_2 = 49$ $K = 1$ $\sigma(m) = \sqrt{\frac{49^2}{21}}$

Le fabriquant estime H_0 $\mu = 800$

il craint H_1 $\mu < 800$

$\mu = \mu_1$



On a $\alpha = 0,05$ et $n = 21$, donc on cherche π et β

$t_\alpha = t_{0,05} = -1,645 = \frac{\pi - 800}{\frac{49}{\sqrt{21}}}$ donc $\pi = 782,4$

$t_{1-\beta} = \frac{782,4 - 760}{\frac{49}{\sqrt{21}}} = 2,097$ donc $\beta = 0,0179$

I 2. Choix entre 2 proportions

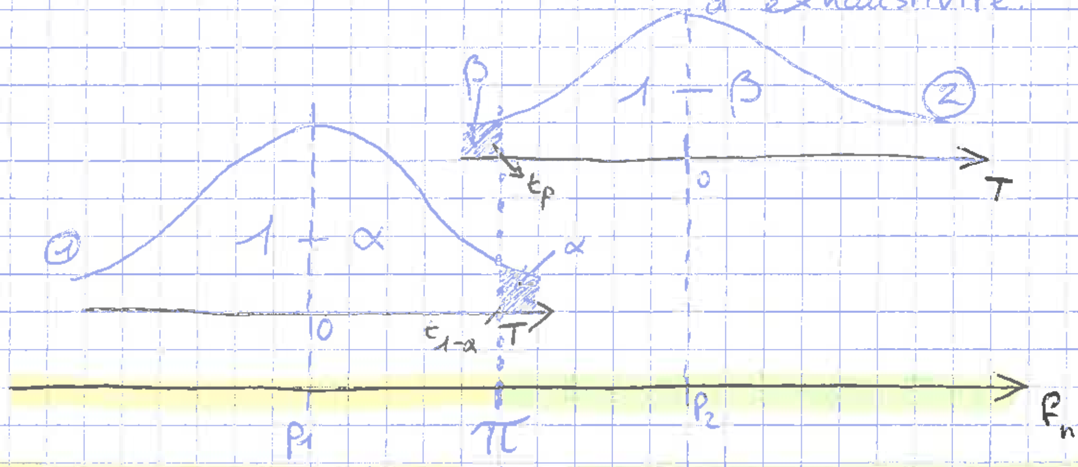
On étudie l'échantillon → taille n, k donc fréq. $f = \frac{k}{n}$ vient de ou Pop 1, taille N, propo° p_1 ; Pop 2, taille N, propo° p_2

Soit les hypothèses : $H_0 \rightarrow p = p_1$
$H_1 \rightarrow p = p_2$ ~~est~~ $p < p_1$ ou $p > p_1$

Nb : p_1 et p_2 sont des valeurs hypotétiques mais chiffrées.

On vérifie
- np_2 et $nq_2 \geq 5$ alors $\mathcal{L}(f_n) = \mathcal{N}(p_2, \sigma_2(f))$ ②
- np_1 et $nq_1 \geq 5$ alors $\mathcal{L}(f_n) = \mathcal{N}(p_1, \sigma_1(f))$ ①

avec $\sigma_i(f) = \sqrt{\frac{pq}{n}\,K}$ ⚠ ne pas confondre k, la propor° et K le coeft d'exhaustivité.



* ici on conserve H_0 et on rejette H_1 avec β% de risque *

* ici on conserve H_1 et on rejette H_0 avec α% de risque *

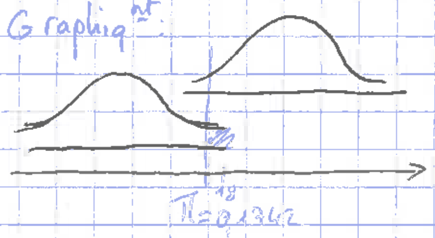
$$t_\beta = \frac{\pi - p_2}{\sigma_2(f)} \quad \text{et} \quad t_{1-\alpha} = \frac{\pi - p_1}{\sigma_1(f)}$$

Application

Exercice 3 : on choisit $p_1 = 0,10$ et $p_2 = 0,20$ car le risque de 1ère esp. α est le risque de juger à tort que le trait^{nt} est nécessaire, et $\alpha = 4\%$.

$$\text{On a } t_{0,96} = \frac{\pi - 0,1}{\sqrt{\frac{0,1 \times 0,9}{50}}} = 1,751 \quad \text{d'où } \pi = 0,1742$$

Graphiq^{nt} :



Avec 18% on obtient le risque α !

$\Delta a = 1,6 - 0,0153\,K$

18% > π, c'est mauvais signe

Risque β

$$t_\beta = \frac{0,1742 - 0,2}{\sqrt{\frac{0,2 \times 0,8}{50}}} = -0,456$$

$\beta = 0,3228$

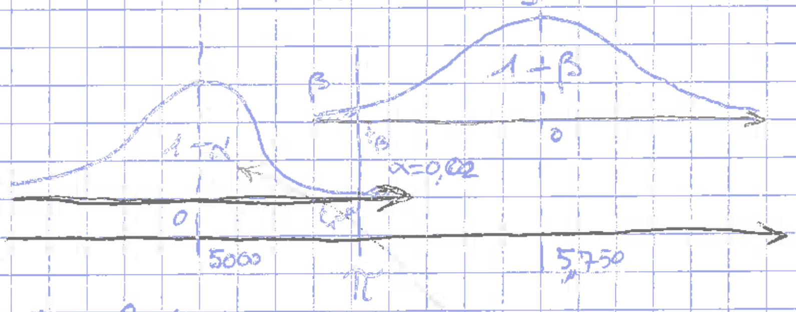
0,40 676

Exercice 2 Le but est donc d'obtenir $1,15 \times 5000 = 5750$
Pour l'instant on a $m = 5400$… on doit la classe soit dans 5000, soit dans 5750.

On pose les hypothèses $H_0 \Rightarrow \mu = 5000$ (pub nécessaire)
$H_1 \rightarrow \mu = 5750$ (pas de pub nécessaire)

On a $n = 9$ et $\sigma_m = \frac{\sigma_x}{\sqrt{n}} = \frac{\sqrt{360000}}{\sqrt{9}} = \frac{600}{3} = 200$

Graphiquement



On a $\mathcal{L}(X) = \mathcal{N}(5000 ; 600)$

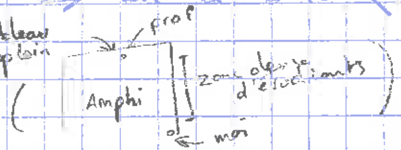
On cherche à prouver qu'on a besoin de pub (on espère que non)

$t_{0,98} = \frac{\pi - 5000}{200} = 2,054$ soit $\pi = 5410,74$

Risque β : $t_\beta = \frac{5410,74 - 5750}{200} = -1,7$ donc $\beta = 0,0446$

⇒ on a $5400 < \pi$ euh… pas on a réussi à prouvé qu'on avait besoin de pub, avec un risque de se tromper de 4,46 %

3/02/2012

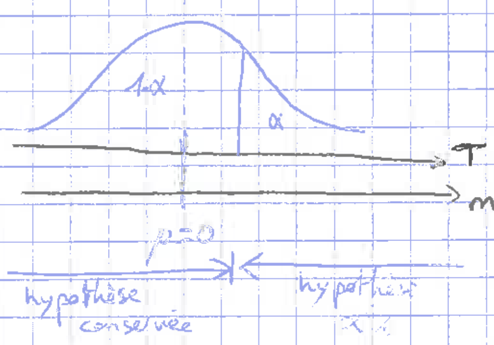


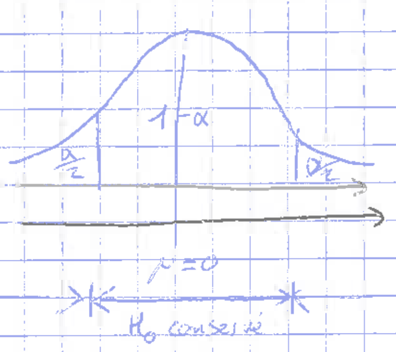
III Comparaison d'une moyenne à une valeur donnée

Dans H_1, le produit n'entre pas dans les normes (→ différence) $\mu \neq a$ ← test bilatéral
ou $\mu > a$ ou $\mu < a$ ← tests unilatéral
Dans H_0, pas de différence → $\mu = a$
Nouveau → p value et

1 échantillon + 1 variable (quali ou quanti) + 1 référence (cahier des charges)
n, m, SCE_x — moyenne — fréquence
issu de la population N, μ, σ_x

on a aucune indication sur β





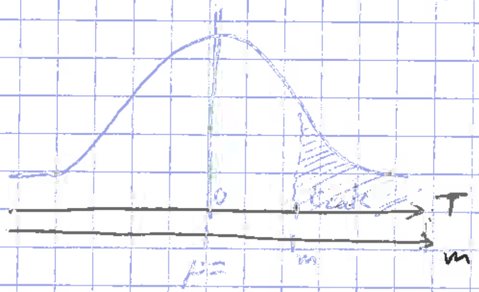
TEST UNILATERAL   TEST BILATERAL

La variable de l'étude c'est $\sigma a = m_n$
On veut trouver $\mathcal{L}(m_n) = ?$
On doit savoir si la variance σ est connue (σ_x) ou estimée $(\hat{\sigma}_x)$

1. La variance est connue

Alors la loi est $\mathcal{L}(m_n) = N(E_m ; \sigma_m)$ avec $E(m) = E(X) = \mu$

$\sigma(m) = \frac{\sigma_x}{\sqrt{n}} \sqrt{K}$



NB : on a pas α

alors on calcule $t_{calc} = \frac{m-\mu}{\sigma(m)}$

La p-value sera notre risque de se tromper en rejetant H_0 unilatéral
↳ p-value $= P(T > |t_{calc}|)$

Le risque est donc représenté sur graph par ////
et se nomme p-value !
Pour l'instant nous ne pourrons calculer la p-value qu'en loi normale → parfois la p-value nous sera donnée mais il faudra pouvoir l'interpréter.

⇓

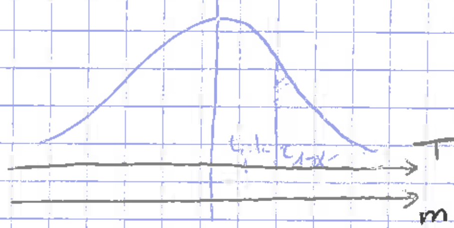
Si on rejette H_0, on a *p-value* % de risque de le faire à tord
(p-value = 0,03 alors 3% de risque de se tromper en prenant H_0)

Conventions

- p-value > 0,05 → la différence est non significative → H_0 conservé (NS)
- p-value ∈ [0,01 ; 0,05] → différence significative → H_0 rejeté (S ou *)
- p-value ∈ [0,001 ; 0,01] → différence très significative → H_1 mis en évidence (TS ou **)
- p-value ⩽ 0,001 → différence hautement signif → (HS ou ***) →

→ Si on ne peut pas calculer le risque p^{value}, c'est soit parce qu'il est donné (α fixé)

On va comparer le t_{calc} avec $\left(t_{1-\alpha} \text{ (unilatéral)} \right.$ ou $\left. t_{1-\frac{\alpha}{2}} \text{ (bilatéral)}\right)$ ← le $t_{théorique}$



* je vois RIEN *

NB = p-value est parfois appelé α calculé

2 La variance est inconnue

Alors $\mathcal{L}(m_n) = St(\nu)(E_m ; \sigma_m)$ avec $E_m = \mu = a$

$$\sigma_m = \frac{\sigma_x}{\sqrt{n}}\sqrt{K} = \frac{S_x}{\sqrt{n-1}}\sqrt{K} = \frac{\sqrt{SCE_x}}{\sqrt{n(\alpha-1)}} \times \sqrt{K}$$

$$t_{calc} = \frac{m-\mu}{\sigma_m} = \frac{m-a}{\sigma_m}$$

Comme nous sommes en loi Student (pas Normale), on ne peut calculer p. value

Alors on compare t_{calc} à $t_{théorique}$

Exemple On a un ddl de 10

on gradue un axe de t

α	$t_{théo}(10)$		
		non signif' (NS)	garde H_0
0,05	8,228	*	
0,01	3,169	** signif'	H_1 mis en évidence
0,001	4,587	***	

$\alpha \nearrow$ $t_{théo} \searrow$

Application : exercice 4 du poly

$N = 300$ $\mu = 6,34$ mm $\sigma_x = 0,25$
$n = 20$ $m = 6,50$

phrase clef → "si les normes sont respectés)
↳ on prend le complémentaire : $H_0 : \mu = 6,34$
$H_1 : \mu \neq 6,34$ (bilatéral)

v.a (variable aléat) m_{20}

D'abord : quelle loi suit $\mathcal{L}(m_n)$?

Ensuite : (situation sans remise) $n = 40$, $n > 30 \rightarrow \mathcal{L}_m = \mathcal{N}(E_m, \sigma_m)$

$E_m = \mu = 6,34$ $\sigma_m = \frac{\sigma_x}{\sqrt{n}}\sqrt{K} = 0,074$

ainsi $t_{calc} = \frac{6,50 - 6,34}{0,074}$ car $t_{calc} = \frac{m-\mu}{\sigma_m} = 2,16$

p. value = $P(T > 2,16) = 0,031$

→ j'ai pas réussi à lister la conclusion

→ pas aux normes

Exercice 8

1) On étudie X, le diamètre d'une pièce. $E(X) = \mu = 5$
$n=37$; $m = 5{,}5$; $S_x = 1{,}25$ $N \to \infty$ donc $K = 1$

$H_0 : \mu = 5$ $H_1 : \mu > 5$ variable : m_{37}

$n > 30$ donc $\mathcal{L}(m_n) = \mathcal{N}(E_m, \sigma_m)$ $E_m = \mu = 5$

$\sigma_m = \frac{s_x}{\sqrt{n}} \times 1 = 0{,}208$

$t_{calc} = \frac{5{,}5 - 5}{0{,}208} = 2{,}4$

p.value $= P(T > 2{,}4) = 0{,}0082$ p.value $= 0{,}82\%$ → TS ** → norme

On a pu montrer que les pièces ont un diamètre > norme de façon très significative ! avec un risque de 0,82% de se tromper

2) $\mathcal{L}(x) = \mathcal{N}$ mais σ inconnu et $n < 30$ → loi Student
pourquoi ?

$E_m = 5$ et $\sigma(m) = \frac{0{,}6}{\sqrt{24}} = 0{,}122$ puis $t_{calc} = \frac{5{,}3 - 5}{0{,}122} = 2{,}45$

On va dans les tables St avec ddl = 24

α	$t_{1-\alpha/2}(24)$
0,05	1,711
0,01	2,492
0,001	3,467

← 2,45 c'est là, entre 0,05 et 0,01

donc on est en signif *

IV Comparaison d'une proportion à une valeur donnée

On a un échantillon de paramètres n, k, $f = \frac{k}{n}$
issu d'une population de paramètres N, p

L'hypothèse $H_0 \to p = a$

H_1 → $p \neq a$ bilatéral
→ $p > a$ } unilatéral
→ $p < a$ }

On considère la variable aléatoire v.a. : f_n

Si np et $nq \geq 5$ alors $\mathcal{L}(f_n) = \mathcal{N}(E(f) ; \sigma(f))$

avec $E(f) = p$ et $\sigma(f) = \sqrt{\frac{pq}{n}} \cdot K$

dont le critère statistique calculé est $t_{calc} = \frac{f - E(f)}{\sigma(f)}$

et le critère théorique

- si α fixé : $|t_{calc}| < t_{théo}$ on conserve H_0
En conservant H_0,
on n'a pas pu prouver que l'échantillon était centré sur $p = a$
- si $|t_{calc}| > t_{thé}$ on garde H_1, on n'a pas pu montrer que l'échant. venait de

On peut aussi calculer p-valeur $\rightarrow$ unilatéral : $P(T > |t_{calc}|)$
$\hookrightarrow$ bilatéral : $2 \times P(T > |t_{calc}|)$

$\rightarrow$ pas de Student avec les fréquences

Exercice 9

$N \rightarrow 1000$ $\quad k = 840$ $\quad f = 0{,}84$ $\quad H_0 : p = 0{,}8$

Comme np et $nq > 5$ alors $\mathcal{L}(f) = \mathcal{N}\left(0{,}8 ; \frac{0{,}8 \times 0{,}2}{1000} \times 1\right)$

Ainsi $t_{calc} = \dfrac{0{,}84 - 0{,}8}{\frac{0{,}04}{\sqrt{1000}}} = 3{,}16$

On a $\alpha = 0{,}06$ donc $t_{0{,}94} = 1{,}555$ $\quad t_{calc} > t_{théo}$

$P(T > 3{,}16) = 0{,}0007$ soit $0{,}07\%$

V Comparaison d'une variable à une valeur

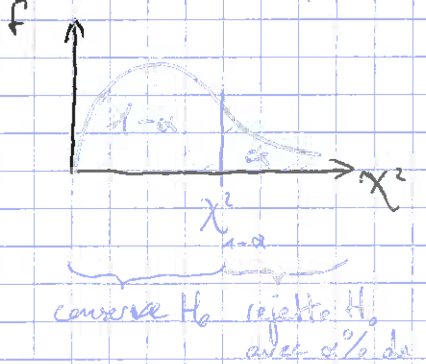
échantillon : n, m, $s_x^2 = \frac{SCE_x}{n}$ et pop N, μ, σ_x^2

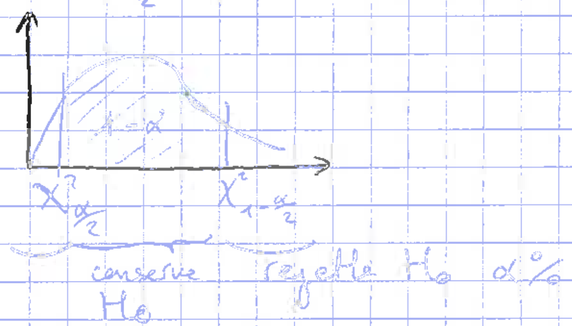
$H_0 : \sigma_x^2 = \sigma_0^2 \Leftarrow$ NB : [illegible]

$H_1 :$ $\sigma_x^2 \neq \sigma_0^2 \rightarrow$ bilatéral
$\left.\begin{array}{l}\sigma_x^2 > \sigma_0^2 \\ \sigma_x^2 < \sigma_0^2\end{array}\right\}$ unilatéral

Le critère statistique est donc $\chi^2 = \dfrac{SCE_x}{\sigma_0^2}$

Le critère statistique théorique $\chi^2 \rightarrow \chi^2_{1-\alpha}$ (unilat.)
$\hookrightarrow \chi^2_{1-\frac{\alpha}{2}}$ (bilat.)





Exercice 11

$n = 10$ $\quad s_x^2 = \frac{SCE_x}{n}$ et pop. N, μ $\quad \sigma_x^2 = 100$

$H_0 : \sigma_x^2 = 100$
$H_1 : \sigma_x^2 \neq 100$ (bilat)

$SCE = \sum x^2 - n\bar{x}^2 = \sum x^2 - n\left(\frac{\sum x}{n}\right)^2 = \sum x^2 - \frac{(\sum x)^2}{n} = 720$

critère stat calc : $\chi^2 = \dfrac{SCE_x}{\sigma_0^2} = \dfrac{720}{100} = 7{,}2$

$$\chi^2_{théo} \to \chi^2_{1-\alpha}(\nu) \text{ (unilat)} \qquad \nu = n-1$$

$$\to \begin{cases} \chi_{1-\frac{\alpha}{2}}(\nu) = \chi_{0,975}(9) = 19 \\ \chi_{\frac{\alpha}{2}}(\nu) = \chi^2_{0,025}(9) = 2,70 \end{cases} \text{ (bilat)}$$

On conserve H_0

VI Comparaison de 2 variances

Des moyennes ne sont comparables que si leur variance se vaut ≃
Il faut donc d'abord comparer les variances

Soient 2 échantillons $\begin{cases} n_1, m_1, s_1^2 \leftarrow Pop_1 : N_1, \mu_1, \sigma_1^2(x) \\ n_2, m_2, s_2^2 \leftarrow Pop_2 : N_2, \mu_2, \sigma_2^2(x) \end{cases}$

$H_0 : \sigma_1^2(x) / \sigma_2^2(x) = 1$

$H_1 \left(\quad " \quad \neq 1 \text{ (bilat)} \right.$

$\quad\quad\; " \quad <> 1 \text{ (unilat)}$ → on utilise la loi F,

→ on utilise des estim° $\hat{\sigma}$

Critère stat calc :

$$F_{calc} = \frac{\dfrac{\hat{\sigma}_1^2(x)}{\sigma_2^2(x)}}{\dfrac{\sigma_1^2(x)}{\sigma_2^2(x)}} = \frac{\hat{\sigma}_1^2(x)}{\hat{\sigma}_2^2(x)}$$

$$F_{calc} = \frac{SCE_{x(1)} / n_1 - 1}{SCE_{x(2)} / n_2 - 1} = \frac{n_1 s_1^2 / n_1 - 1}{n_2 s_2^2 / n_2 - 1}$$

⚠ La variance estimée la + petite au dénominateur

$$F_{théo} \to \begin{cases} F_{1-\alpha}(\nu_1 ; \nu_2) \\ F_{1-\frac{\alpha}{2}}(\nu_1 ; \nu_2) \end{cases}$$

avec $\nu_1 = n_1 - 1$, ddl du numérateur
$\nu_2 = n_2 - 1$, ddl du dénominateur

ν_1 → numérateur
ν_2 → dénominateur

Ce test vérifie l'invariance des variances (homogénéité des variances)
On dit que c'est un test d'homoscédasticité

⚠ Le test d'homoscédasticité pour comparer deux variances est un test de Fisher Snedecor, soit test F (utilisant le critère F).
Ce test pour comparer plusieurs variances est un test de Barclette (utilisant le critère χ^2)

→ vient de celle quelle pop
→ dans quelle ...
Test t → est-ce des ...

Exercice 16 (début)

Tu rentres les chiffres de la calculatrice → $\sigma_A^2 = 8,22$, $\sigma_B^2 = 4,5$
et en plus → $n_A = 10$ $n_B = 8$

$$F_{1-\frac{\alpha}{2}}(\nu_1 ; \nu_2) = F_{0,975}(9;7) \in [4,76 ; 4,90]$$

on a pas la valeur exacte … toujours !

$$F_{calc} < F_{théo}$$

donc je conserve H_0 : on a pas mis en evidence que les variables étaient différentes donc si besoin, on pourra calculer une estimation commune des deux variances.

Si on ne peut pas distinguer σ_A^2 et σ_B^2, on les note en une seule :

$$\hat{\sigma}_x^2 = \frac{SCE_1 + SCE_2}{n_1 - 1 + n_2 - 1} = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

VII Comparaison de deux séries

On a deux échantillons. Ils pourraient être indpts ou bien appariés (dits aussi appareillés)

1 Echantillons indépendants

Les échantillons → E1 : n_1, m_1, SCE_1 tirés des pops fantôme P1 $N_1, \mu_1, \sigma_1(x)$
↳ E2 : n_2, m_2, SCE_2 P2 $N_2, \mu_2, \sigma_2(x)$

H_0 : les échantillons sont pareils : $\mu_1 - \mu_2 = 0$
H_1 : les échantillons sont différents $\mu_1 - \mu_2 \neq 0$
$\mu_1 - \mu_2 > 0$
$\quad\quad\quad < 0$

On cherche la v.a, la loi, l'E et σ

→ va On imagine qu'on compare les moyennes de tous les échantillons de chaque pop. Et on étudie la différence entre ~~tous les~~ chaque échantillon d'une pop avec chaque éch de l'autre pop.
↳ on étudie $\Delta m = m_{x_1} - m_{x_2}$

→ loi Pour connaître la loi L, on relève les lois de chaque echantillon. Normalement, les lois de chaque éch sont les m̂. Sinon on peut pas étudier et c'est mort.

Ainsi Si $\mathcal{L}(m_{x_1})$ et $\mathcal{L}(m_{x_2})$ Alors $\mathcal{L}(\Delta m)$

$\mathcal{N}$	$\mathcal{N}$	→	$\mathcal{N}$
St (ν_1)	St (ν_2)	→	St (ν)
rien	rien	→	rien

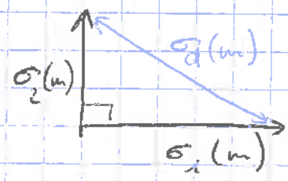
Comme $(\nu_1) = n_1 - 1$ $(\nu_2) = n_2 - 1$ et $(\nu) = \nu_1 + \nu_2 = n_1 + n_2 - 2$

on vérifie homoscédasticité

$\rightarrow E(m_{n_1} - m_{n_2}) = E(m_{n_1}) - E(m_{n_2}) = \mu_1 - \mu_2$

$\rightarrow \sigma(\Delta m) = \sigma_d(m)$

$\sigma_d = \sqrt{\sigma_2^2 + \sigma_1^2}$



CAS 1 $\mathcal{L}(X) = \mathcal{N}$ et $n \geq 30$ et σ_x connu $\mathcal{L}(m) = \mathcal{N}$ et $\mathcal{L}(\Delta m) = \mathcal{N}$

$\rightarrow \sigma_d(m) = \sqrt{\frac{\sigma_1^2(x)}{n_1} K_1 + \frac{\sigma_2^2(x)}{n_2} K_2}$

$\rightarrow$ crit. stat. calculé $t_{calc} = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{\sigma_d(m)} = \frac{m_1 - m_2}{\sigma_d(m)}$ $\quad (\mu_1 - \mu_2) \rightarrow = 0$ avec H_0

$\rightarrow$ crit. stat. théorique $t_{théo} = \begin{matrix} t_{1-\frac{\alpha}{2}} & \text{(bilat')} \\ t_{1-\alpha} & \text{(unilat')} \end{matrix}$

$\Rightarrow$ on compare t_{calc} et $t_{théo}$ puis on peut calculer p value

CAS 2 $n \geq 30$ et σ_x inconnus

$\rightarrow \sigma_d(m) = \sqrt{\frac{\sigma_1^2(x)}{n_1} K_1 + \frac{\sigma_2^2(x)}{n_2} K_2}$

$= \sqrt{\frac{S_1^2(x)}{n_1 - 1} K_1 + \frac{S_2^2(x)}{n_2 - 1} K_2}$

$= \sqrt{\frac{SCE_1}{n_1(n_1-1)} K_1 + \frac{SCE_2}{n_2(n_2-1)} K_2}$

$\rightarrow$ crit. stat calc $t_{calc} = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{\sigma_d(m)} = \frac{m_1 - m_2}{\sigma_d(m)}$

$\rightarrow$ crit. stat théo $t_{théo} = \begin{matrix} t_{1-\frac{\alpha}{2}} & \text{(bilat')} \\ t_{1-\alpha} & \text{(unilat')} \end{matrix}$

CAS 3 $\mathcal{L}(X) = \mathcal{N}$ et σ_x inconnu et $n < 30$ so $\mathcal{L}(m) = St$

$\hookrightarrow \mathcal{L}(\Delta m) = St(\nu)$ $\quad \nu = n_1 + n_2 - 2$

STEP 1 on vérifie l'homoscédasticité des variances (avec $\frac{\sigma_1^2(x)}{\sigma_2^2(x)} = 1$) → Il faut H_0 conservée

STEP 2 on détermine l'estimation de la variance intragroupe commune aux 2 échantillons. Elle représente les fluctuations naturelles entre les individus

$$\hat{\sigma}_x^2 = \frac{SCE_1(x) + SCE_2(x)}{n_1 + n_2 - 2} = \frac{n_1 S_1^2(x) + n_2 S_2^2(x)}{n_1 + n_2 - 2}$$

Comme $\sigma_1(x)$ et $\sigma_2(x)$ sont estimées et égales $\rightarrow \hat{\sigma}(x)$

$$\sigma_d(m) = \sqrt{\frac{\hat{\sigma}^2(x)}{n_1} K_1 + \frac{\hat{\sigma}^2(x)}{n_2} K_2}$$

Exercice 16 (fin)

échant A $n_A = 10$ $m_A = 220$ $SCE_A = 73,998$
échant B $n_B = 8$ $m_B = 218,75$ $SCE_B = 31,5$

STEP1 $F_{calc} = 1,83$
$F_{théo} = F_{0,975}(9;7) = 4,92$ $\Big\} \rightarrow$ on conserve $H_0 : \frac{\sigma_1^2(x)}{\sigma_2^2(x)} = 1$

STEP2 $\hat{\sigma}_x^2 = \frac{73,998 + 31,5}{10+8-2} = 6,59$

STEP3 $H_0 = \mu_1 - \mu_2 = 0$ avec remise $\left.\begin{matrix} N_1 \\ N_2 \end{matrix}\right\} \rightarrow \infty$ et $K = 1$
$H_1 : \mu_1 - \mu_2 \neq 0$

$\mathcal{L}(m_{10}) = St(9)$
$\mathcal{L}(m_8) = St(7)$ $\Big\} \rightarrow \mathcal{L}(\Delta m) = St(16)(\mu_1 - \mu_2 ; \sigma_d(m))$

$E(\Delta m) = \mu_1 - \mu_2 = 0$ $\sigma_d(m) = [\ldots] = 1,218$

$t_{calc} = \frac{220 - 218,75}{1,218} = 1,026$ $t_{théo} = 2,12$

$\Rightarrow$ On n'a pas pu mettre en évidence une différence significative

Exercice 13

Ech 1 $n_1 = 64$ $m_1 = 616$ Pop 1 $N_1 = 700$ μ_1 $\sigma_1^2(x) = 100$
Ech 2 $n_2 = 68$ $m_2 = 620$ Pop 2 $N_2 = 800$ μ_2 $\sigma_2^2(x) = 121$

$\rightarrow$ en prévision du t_{calc} où on veut un numérateur positif

$H_0 = \mu_2 - \mu_1 = 0$ $n \geqslant 30$ donc $\mathcal{L}(m) = \mathcal{N}$ et $\mathcal{L}(\Delta m) = \mathcal{N}$
$H_1 = \mu_2 - \mu_1 > 0$ avec remise $\rightarrow K_1 = K_2 = 1$

$$t_{calc} = \frac{620 - 616}{\sqrt{\frac{100}{64} \times 1 + \frac{121}{68} \times 1}} = 2,19$$

$t_{théo} = t_{0,995} = 1,96$ donc p value $= P(T > 2,19) = 0,0143$

Le 2nd engrais est significativement plus efficace, avec un risque de 1,43% de se tromper.

Exercice 15

$n > 30$ donc $\mathcal{L}(m) = \mathcal{N}$ et $\mathcal{L}(\Delta m) = \mathcal{N}$

Ho : $\mu_1 - \mu_2 = 0$
H_1 : $\mu_1 - \mu_2 \neq 0$

$$t_{calc} = \frac{90 - 87,3}{\sqrt{\frac{5,5^2}{37} \times \frac{300-38}{300-1} + \frac{4,5^2}{39} \times \frac{350-40}{350-1}}}$$

$t_{calc} = 2,488$

$t_{0,975} = 1,96$ p value $= P(T > 2,49) = 0,0123$

$\Rightarrow$ On a mis en évidence une différence significative entre les 2 régions, avec un risque d'erreur de 0,0128

18

2 Echantillons appariés

n°	x_i	x'_i	$d_i = x_i - x'_i$
1	x_1	x'_1	d_1
2	x_2	x'_2	d_2
⋮	⋮	⋮	⋮
i	x_i	x'_i	d_i
⋮	⋮	⋮	⋮
n	x_n	x'_n	d_n
			$\bar{d}$

⚠ la différence d_i peut être faite ds n'importe lequel des 2 sens. MAIS CE SENS EST GARDÉ

→ On étudie donc la variable aléatoire $\bar{d}_n$

$$\mathcal{L}(\bar{d}_n) = St(\nu)(E(\bar{d}); \sigma(\bar{d}))$$

$\nu = n-1$

→ La population est caractérisée par N, δ, $\sigma(d)$ et les éch. tirés :

$$\left.\begin{array}{l} n_1 \to \bar{d} \\ n_2 \to \bar{d}' \\ n_3 \to \bar{d}'' \\ \text{etc.} \end{array}\right\} \text{d'où } E(\bar{d}),\ \sigma(\bar{d})$$

* $E(\bar{d}) = \delta$
* $\sigma(\bar{d}) = \dfrac{\hat{\sigma}_d}{\sqrt{n}}\sqrt{K}$ avec $\hat{\sigma}_d = \sqrt{\dfrac{SCE_d}{n-1}}$

H_0 : $\delta = 0$

H_1 : $\delta \neq 0$

$\delta > 0$

$\delta < 0$

Le critère stat calculé $t_{calc} = \dfrac{\bar{d} - \delta}{\sigma(\bar{d})} = \dfrac{\bar{d}}{\sigma(\bar{d})}$

↓

Il est à comparé au critère stat théorique lu ds la table Student

test bilat → $t_{1-\frac{\alpha}{2}}(\nu)$

test unilat → $t_{1-\alpha}(\nu)$

La méthode des astérisques reste la m̂.

Exercice 17

Soit d_i = après − avant

$d_i = \{(0;1); (0,1;3); (0,2;5); (0,3;3)\}$ avec $n = 12$

$\Rightarrow \bar{d} = 0,1833$

En entrant la série sur calculatrice : $S_d = 0,08975$ donc $\hat{\sigma}_d = 0,0937$

On a donc : $E(\bar{d}) = \delta = 0$

$\sigma(\bar{d}) = \dfrac{\hat{\sigma}_d}{\sqrt{n}}\sqrt{K} = \dfrac{0,0937}{\sqrt{12}}\sqrt{1} = 0,0271$

(on aurait aussi pu faire $\dfrac{0,08975}{\sqrt{11}} = \sigma(\bar{d})$)

$t_{calc} = \dfrac{\bar{d} - \delta}{\sigma(\bar{d})} = \dfrac{0,1833}{0,0271} = 6,77$

$t_{théo} = t_{0,99}(11) \in [2,201; 2,718]$

⇒ On pût mettre en évidence l'améliorat$^{\circ}$ avec l'entraîne^{mt} à 2% de risque

→ haut^{mt} significative

NB : Mes δ ne sont pas des delta mais des sigma sur cette page

VIII Comparaison de deux proportions

Ech 1 : n_1 $f_1 = \frac{k_1}{n_1}$ ← Pop(fantôme) N_1 P_1
Ech 2 : n_2 $f_2 = \frac{k_2}{n_2}$ ← Pop(fantôme) N_2 P_2

$H_0 : P_1 = P_2$
$H_1 : P_1 \neq P_2$
$P_1 > P_2$
$P_1 < P_2$

Cette fois, la variable aléatoire est $\Delta F = F_{n_1} - F_{n_2}$

Si $\mathcal{L}(f_n) = \mathcal{N}$

alors $\mathcal{L}(\Delta f) = \mathcal{N}(E_d(f) ; \sigma_d(f))$

On estime une valeur $\hat{P}_0$ commune à P_1 et P_2

$$\hat{P}_0 = \frac{k_1 + k_2}{n_1 + n_2} \qquad \hat{q}_0 = 1 - \hat{p}_0$$

$$\sigma_d^2(f) = \sigma_1^2(f) + \sigma_2^2(f) = \frac{\hat{p}_0 \hat{q}_0}{n_1} K_1 + \frac{\hat{p}_0 \hat{q}_0}{n_2} K_2$$

Le critère stat calculé : $t_{calc} = \frac{f_1 - f_2}{\sigma_d(f)}$

Comme on est en loi normal, on peut utiliser la p.value pour caractériser le niveau.

Exercice 18

Ech 1 : $n_1 = 250$ $f_1 = \frac{175}{250} = 0,7$
Ech 2 : $n_2 = 300$ $f_2 = \frac{220}{300} = 0,73333$

$H_0 : P_1 = P_2$
$H_1 : P_2 > P_1$ v.a : $\Delta f = f_2 - f_1$

$$\hat{P}_0 = \frac{175 + 220}{250 + 300} = 0,7182 \qquad \hat{q}_0 = 1 - \hat{p}_0 = 0,2818$$

$$\hat{\sigma}_d(f) = \sigma_1^2(f) + \sigma_2^2(f) = 0,7182 \times 0,2818 \left(\frac{1}{250} + \frac{1}{300}\right)$$

$$t_{calc} = \frac{f_1 - f_2}{\sigma_d(f)} = \frac{0,7333 - 0,7}{0,0385} = 0,87$$

$t_{0,95} = 1,645$ $t_{théo} > t_{calc}$

→ On ne peut pas prouver que la 2^{nd} est mieux H_0 pas rejeté

Pour la fin
↳ p.value $P(T > 0,87) = 0,1922$

IX Comparaison de k variance : Test de Bartlett

$H_0 = \sigma_1^2(x) = \sigma_2^2(x) = \dots = \sigma_j^2(x) = \dots = \sigma_k^2(x)$

$H_1 =$ au moins <u>une</u> des k variances est $>$ à une autre

<u>Test unilatéral</u> $t_{théo} =$

critère statistique théorique : $\chi^2_{1-\alpha}(\nu)$ avec $\nu = k-1$

Ech.	n_j	SCE	$\nu_j = n_j - 1$	$\hat{\sigma}_j^2 = \frac{SCE_j}{\nu_j}$	$\frac{1}{\nu_j}$	$\nu_j \log \hat{\sigma}_j^2$
1	n_1	SCE_1	ν_1	$\hat{\sigma}_1^2$	$\frac{1}{\nu_1}$	$\nu_1 \log \hat{\sigma}_1^2$
2	n_2	SCE_2	ν_2	$\hat{\sigma}_2^2$	$\frac{1}{\nu_2}$	$\nu_2 \log \hat{\sigma}_2^2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
j	n_j	SCE_j	ν_j	$\hat{\sigma}_j^2$	$\frac{1}{\nu_j}$	$\nu_j \log \hat{\sigma}_j^2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	n_k	SCE_k	ν_k	$\hat{\sigma}_n^2$	$\frac{1}{\nu_k}$	$\nu_n \log \hat{\sigma}_n^2$
			$\nu = \sum \nu_j$		$\sum \frac{1}{\nu_j}$	$\sum \nu_j \log \hat{\sigma}_j^2$

Estimat° commune de la variance intragroupe :

$$\hat{\sigma}^2 = \frac{\sum SCE_j}{\sum \nu_j} = \frac{\sum \nu_j \hat{\sigma}_j^2}{\nu}$$

Exercice 20

$H_0 : \sigma_1^2(x) = \sigma_2^2(x) = \sigma_3^2(x)$

H_1 : une des 3 variances au moins est $>$ à une autre

Test unilat de $\chi^2_{0,99}(2) = 9,21$ $\nu = 3-1$

Ech	n_j	SCE	ν	σ	$\frac{1}{\nu}$	$\nu \log \hat{\sigma}^2$
1	6	$5 \times 59,6$	5	59,6	$\frac{1}{5}$	$5 \times \log 59,6$
2	12	$11 \times 154,97$	11	154,97	$\frac{1}{11}$	$11 \times \log 154,97$
3	9	$8 \times 548,53$	8	548,53	$\frac{1}{8}$	$8 \times \log 548,53$
		$= 390,89$	24		0,4159	54,88

$$\hat{\sigma}^2 = \frac{\sum SCE}{\sum \nu} = \frac{6390,89}{24} = 266,287$$

$$\chi^2_{calc} = \frac{2,3026}{1 + \frac{1}{3\times 2}\left(0,4159 - \frac{1}{24}\right)} \times \left[2 \log_{10} \times 266,287 - 54,88 \right] =$$

$\chi^2_{calc} < \chi^2_{théo} \Rightarrow$ on a pas pu mettre en évidence H_1.
On garde l'hypothèse d'homosédasticité

On va alors pouvoir estimer une variance intra-groupe commune reflétant les fluctuations naturelles entre les individus, qu'on appelera le carré moyen résiduel CM_e

Alexia MONSAVOIR
06.23.42.24.30

ISARA – 2ème année – STATISTIQUE
Exercices : Parties 4 et 5
Mme B. Bottollier-Lemallaz

Les corrigés de ces exercices sont consultables sur l'espace du cours de statistique.

TD 4EME PARTIE

Exercice 1

Supposons la série (xi; ni), représentative de la distribution d'une variable X dans une population de taille 9, suivante: ← 1 fois la valeur 10
{(2 ; 1); (4 ; 2); (6 ; 3); (8 ; 2); (10 ; 1)}
2 4 4 6 6 6 8 8 10

1) déterminer l'espérance mathématique et la variance de la variable aléatoire dans la population. à la calculatrice
2) établissez la liste des moyennes de tous les échantillons possibles de taille 2, avec remise, et calculer pour chaque échantillon la moyenne. → tableau à dble entrée
Déterminer l'espérance mathématique et la variance de la distribution des moyennes des échantillons de taille 2 dans le cas d'un tirage avec remise. → à la calculatrice
3) établissez la liste des moyennes de tous les échantillons possibles de taille 2, sans remise, et calculer pour chaque échantillon la moyenne. → on ne fait pas la diagonale
Déterminer l'espérance mathématique et la variance de la distribution des moyennes des échantillons de taille 2 dans le cas d'un tirage avec remise.

Exercice 2

Supposons qu'un caractère soit présent dans une proportion de 80% dans une population de 10 unités statistiques. Après avoir établi la distribution des fréquences*, déterminer :
1) l'espérance mathématique et la variance de la distribution des fréquences des échantillons de taille 2 dans le cas d'un tirage avec remise.
2) l'espérance mathématique et la variance de la distribution des fréquences des échantillons de taille 2 dans le cas d'un tirage sans remise

Exercice 3

Soit une population de 340 unités statistiques sur laquelle la variable étudiée suit une loi normale de moyenne 250 et de variance 256.
Entre quelles limites doit se trouver la moyenne d'un échantillon de 40 unités pour être représentatif de la population pour un niveau de confiance de 0,94?

Exercice 4

Supposons que sur une période de 4 ans on a étudié les rendements en blé de 370 parcelles. On a conclue que le rendement d'une parcelle suivait une loi normale de moyenne 590 et d'écart type 49.
1) Entre quelles limites doit se trouver la moyenne d'un échantillon de 25 parcelles pour être représentatif de la population pour un niveau de confiance de 0,90?
2) Quelle est la probabilité qu'un échantillon de 35 parcelles ait une moyenne comprise entre 580 et 610?
3) Quelle est la probabilité qu'un échantillon de 40 parcelles ait une moyenne comprise entre 570 et 600?

* pour n = 2

Exercice 5

Une entreprise fabrique des transistors utilisés dans un récepteur de haute qualité. Un contrôle régulier est effectué à l'aide d'un testeur électronique permettant de détecter d'une façon automatique les transistors défectueux.

1) Si le processus de fabrication produit, selon les normes, en moyenne une proportion de transistors défectueux de 3% entre quelles limites doit se trouver la fréquence de pièces défectueuses sur un échantillon de 200 pièces contrôlées pour un niveau de confiance de 0,90 pour que le processus de fabrication soit aux normes.

2) Afin d'améliorer la qualité de la fabrication on utilise sur la fabrication de 500 transistors un nouveau processus avec lequel on pense n'avoir que 2% de pièces défectueuses. Entre quelles limites doit se trouver la fréquence des pièces défectueuses sur un échantillon de 300 transistors contrôlés pour un niveau de confiance de 0,90?

Exercice 6 → à maîtriser à fond ! C'est un investis[illegible] pour toute la vie !

Si m est la moyenne d'un échantillon aléatoire de taille n tiré avec remise d'une population normale de moyenne μ et de variance 100, déterminer n tel que :

1) $P(\mu - 10 \leq m \leq \mu + 10) = 0{,}9544$
2) $P(\mu - 5 \leq m \leq \mu + 5) = 0{,}9544$
3) $P(\mu - 2 \leq m \leq \mu + 2) = 0{,}9544$
4) $P(\mu - 2 \leq m \leq \mu + 2) = 0{,}6826$
5) $P(\mu - 2 \leq m \leq \mu + 2) = 0{,}8664$

On supposera le tirage avec remise.

Quelles conclusions pouvez-vous en tirer ? Réaliser cet exercice sous forme d'un tableau.

Exercice 7

Le directeur des ressources humaines d'une entreprise de 200 employés veut estimer la dextérité manuelle de sa nouvelle main-d'œuvre. Des résultats aux tests de dextérité sur une ancienne main d'œuvre étaient distribués selon une loi normale de variance 25.

1) Les tests sur un échantillon de 40 employés sélectionné au hasard donnent une moyenne égale à 71. Donner les limites entre lesquelles se trouve la dextérité moyenne de tous les employés avec 1% de risque d'erreur. Calculer les marges d'erreur.

2) Si les tests avaient eu lieu avec un échantillon de 20 employés et une moyenne de 71, entre quelles limites se trouverait la dextérité moyenne de tous les employés ? Calculer les marges d'erreur.

Exercice 8

Des études sur 500 sportifs ont montré que le temps de réponse à un stimulus a une dispersion de 15μs.

1) Si on admet que la dispersion ne varie pas, et que pour un échantillon de 20 sportifs minimes tirés au hasard parmi 500, on a un résultat moyen de 71 μs. Estimer par intervalle de confiance le temps de réponse moyen de la catégorie minime avec un niveau de confiance de 0,95. Calculer les marges d'erreur.

2) Mêmes questions si sur un échantillon de taille 40 on observe un temps moyen de 71 μs.

3) Mêmes questions si sur un échantillon de taille 60 on observe un temps moyen de 71 μs.

Exercice 9 → juste avec $\mu = m$?

Le distributeur d'un appareil estime que celui-ci doit être vendu 50\$. Afin de vérifier si ce prix est respecté dans ses 300 magasins, il vérifie le prix moyen proposé par un échantillon de n magasins. On choisira un niveau de confiance de 0.95.

1) Sur 20 magasins le prix moyen est de 55\$ avec un écart type égal à 5\$. Quel doit être la conclusion du distributeur?

2) N'étant pas satisfait du résultat il complète son étude avec 20 autres magasins tirés au hasard. Sur les 40 magasins le prix moyen est de 56\$ avec un écart type égal à 6\$.. Quel doit être la conclusion du distributeur?

le

! On se moque de lui !

Exercice 10

On souhaite estimer le salaire mensuel moyen de 2 niveaux de cadres d'entreprises.
L'étude porte sur des **échantillons** et on a obtenu :

Cadres	effectif	Moyenne ($)	Ecart type ($)
Niveau 1	26	3521	338
Niveau 2	57	2756	245

Estimer, par intervalle de confiance à l'unité près, pour un niveau de confiance de 0,98 le salaire mensuel moyen pour chaque niveau. Calculer les marges d'erreur. On admettra que le salaire d'un cadre de niveau j suit une loi normale.

Exercice 11

Des essais en laboratoire sur 20 lampes témoins utilisées sur des panneaux de contrôle électroniques conduisent aux durées de vie en heures ci-dessous :

450	410	412	380	407
455	375	390	355	364
410	413	345	430	390
330	440	381	451	415

En admettant que la durée de survie d'une lampe suive une loi normale, estimer par intervalle de confiance la durée de vie moyenne de toute la production au niveau 0,98.

Exercice 12 → corrigé p.5

Sur un échantillon aléatoire de 18 prélèvements indépendants dans une population binomiale de taille infinie on observe 9 unités statistiques présentant une caractéristique étudiée. Déterminer une estimation par intervalle de confiance à 99% de la proportion d'unités présentant la caractéristique dans la population.

Exercice 13 → corrigé p.5

Un sondage mené auprès de 1000 adultes pour connaître le niveau de satisfaction d'une émission de télévision a donné les résultats suivants :
Satisfaits : 146 insatisfaits : 851 sans avis : 3
1) Entre quelles limites peut-on s'attendre de trouver la vraie proportion de personnes satisfaites au risque de se tromper 1 fois sur 20 ?
2) Même question pour les personnes sans avis.

Exercice 14

Déterminer un intervalle de confiance à 95% pour la variance et l'écart type de la population de l'exercice 11.

Maintenant il est temps d'aller vous entrainer dans la rubrique « Testez vos connaissances» de cette partie 4 sur http://ecampus.isara.fr cours STAT_S4
Il faut réussir à faire 22 questions en 45mn le jour de l'examen!!!!!

TD 5EME PARTIE

Exercice 1 choix entre 2 moyennes p. 8

Une entreprise fabrique des dispositifs électroniques dont la durée de vie suit une loi normale de moyenne égale à 800 heures et d'écart-type 49 heures. Pour vérifier la qualité des dispositifs on prélève régulièrement un échantillon aléatoire de 21 dispositifs que l'on soumet à des essais de fiabilité. Le risque que prend le fabriquant de déclarer un lot non acceptable alors qu'il l'est est de 5%.

1) Calculer la région critique à ne pas dépasser pour considérer le lot acceptable.
2) Calculer le risque du client, à savoir accepter à tort un lot qui n'est pas aux normes si le processus de fabrication est centré sur une durée de vie moyenne de :
 - 760 heures
 - 790 heures
3) Le responsable du contrôle modifie son plan en prélevant 36 dispositifs de la production. Quelle règle de décision doit-il alors adopter?

Exercice 2 Choix entre 2 prop. p 10

Au cours de l'année écoulée, les ventes hebdomadaires du rayon parfumerie de " Casimouth " ont pu être approximées par une loi normale de moyenne 5000$ et de variance 360000. Désirant développer ce rayon, " Casimouth " a procédé à un effort de promotion sur place (réaménagement du rayon, sélection des marques, démonstratrices...etc.) sans procéder à des campagnes publicitaires. Le but de la direction était d'obtenir un accroissement des ventes de 15% (avec l'idée de procéder à une action publicitaire si elle n'obtenait pas ce pourcentage d'augmentation). De plus elle a décidé de juger les résultats de son action en se basant sur les ventes des 9 premières semaines suivant la réorganisation du rayon. Sachant que la direction a fixé le risque de 1ère espèce à 2% (risque de ne pas procéder à l'action publicitaire alors qu'elle aurait été nécessaire) quelle aurait été votre conclusion si ce chiffre d'affaire enregistré sur ces 9 semaines avait été de 5400$?

Exercice 3 choix entre 2 prop. p. 9

Une culture soumise à des traitements curatif et préventif est habituellement contaminée par une maladie à raison d'un ratio de 10%. Pour des raisons de qualité de produit il est question de supprimer ces traitements si toutefois l'envahissement de la maladie n'atteint pas un taux moyen de 20%. L'étude porte sur un échantillon de 50 parcelles et le risque de première espèce (risque de juger à tort que le traitement est nécessaire) est fixé à 4%. En l'absence de tout traitement on a relevé un envahissement de 18% sur l'échantillon étudié.

1) Déterminer la région d'acceptation et celle de refus de la culture dans ces conditions
2) Calculer le risque de ne pas déceler à tort un envahissement de la maladie trop important.

Pour les exercices qui suivent, lorsque cela est possible, déterminer également la p-value et reformulez la conclusion.

Exercice 4 Comparaison d'une moy à une valeur donnée → σ^2 connue p. 12

Une entreprise achète à un fournisseur 300 pièces d'une matière isolante dont la norme exige qu'elle ait une épaisseur de 6.34 mm. Cette entreprise veut vérifier si les normes sont respectées et sachant que la variance de la variable est de 0.25 elle prélève un échantillon de 20 pièces elle trouve une épaisseur moyenne de 6.50 mm.

1) Que peut-on conclure sur le lot fourni au seuil de 5%?
2) Même question si l'échantillon avait été de taille 40?

Exercice 5

D'après des expériences antécédentes, il semble pour des individus de tranche d'âge comparable, que le temps moyen de réaction à un stimulus électrique soit de 76.26 ms avec un écart-type de 2 ms Parmi un groupe de 42 sujets de tranche d'âge supérieure on a trouvé un temps moyen de réponse à ce stimulus égal à 77 ms . Peut-on conclure au seuil de 1% que ce groupe à un moins bon temps de réaction?

Exercice 6

Pour qu'un certain type de détecteur de fumée soit efficace il faut qu'il se déclenche en moyenne dans les 1.9 secondes qui suivent le début de l'incendie. Afin de vérifier une nouvelle marque un service de contrôle en teste 12 sur les 200 fabriqués, en les submergeant de fumée. Les temps de réponse sont les suivants :

1,4	1,2	1,0	2,2	2,0	1,2	1,6	1,7	1,8	1,5	1,9	2,0

Un autre service en contrôle 36 et obtient :
{ (1,4; 3); (1,2; 4); (2,0; 3); (1,7; 2); (2,2; 5); (1,9; 12); (1,8; 7) }

Pour chaque contrôle, pouvez-vous conclure que la nouvelle marque soit plus performante que ce que nécessite la norme au seuil de 2.5% ?

Exercice 7

Pour une tranche d'âge donnée la fréquence cardiaque des individus doit être en moyenne de 72 battements par minute. On supposera que la variable aléatoire suit une loi normale.
1) On a isolé un groupe de 12 individus et l'on souhaite savoir si ce groupe représente la population centrée sur 72. Les résultats suivants ont été obtenus :

88	68	72	62	80	60	72	72	72	75	72	75

Que concluez-vous au seuil 4% ?
2) L'étude a porté sur un groupe de 53 individus où la fréquence cardiaque moyenne a été de 74 et la variance de 81. Pensez-vous que ce groupe représente une population centrée sur 72 au seuil de 4% ?

Exercice 8

Une usine fabrique des pièces circulaires dont il garantit le diamètre moyen à 5 cm.
1) Un acheteur affirme que sur un lot de 37 pièces il a trouvé un diamètre moyen de 5,5 cm avec un écart-type de 1,25 cm et donc que le diamètre des pièces fournies n'est pas aux normes avancées par le fabricant mais lui est supérieur. Quel est votre avis?
2) Un autre acheteur, partant de l'hypothèse que cette dimension suit une loi normale a vérifié le diamètre moyen sur un échantillon de 25 pièces et a trouvé un diamètre moyen de 5,3 cm avec un écart-type de 0,6 cm. Pensez-vous que le lot fournit à cet autre acheteur ait un diamètre supérieur à 5 cm au seuil de 5% ?

Exercice 9

Le directeur commercial d'un important quotidien affirme que plus de 80% des foyers du pays lisent ce quotidien. Un sondage est alors effectué auprès de 1000 foyers et indique que 840 foyers lisent ce quotidien. Est-ce que l'affirmation du directeur commercial est supportée par les résultats du sondage, au seuil de signification 0,06 ?

Exercice 10

Un laboratoire a mis au point un traitement contre une maladie et affirme que ce traitement a un taux de réussite de 79%. Dans un hôpital où cette maladie est traitée on administre ce traitement et on note que sur 200 malades traités, 150 sont guéris. Peut-on conclure que l'affirmation faite par le laboratoire est vraie au seuil de 5% ?

Exercice 11

Un échantillon aléatoire et indépendant de 10 individus est tiré d'une population où la variable étudiée (résultats à un test d'aptitude) suit une loi normale de variance égale à 100. Les résultats de ce test sont : 78; 60; 64; 82; 80; 66; 74; 61; 68; 57. Pensez-vous que la dispersion des valeurs est conforme à ce que l'on a habituellement dans la population, au risque de 5% ?

Exercice 12

Une étude comparative du prix d'un même produit de 2 marques différentes distribuées dans différents magasins a donné les résultats suivants :
Marque A : moyenne égale à 138$ sur 25 magasins
Marque B : moyenne égale à 135$ sur 22 magasins
On sait que la variance des prix est commune et est égale à 100 et que la distribution des prix de ces produits suit une loi normale. La différence entre ces prix moyens est-elle significative au seuil de 1% ?

Exercice 13

On veut comparer l'efficacité de 2 engrais sur le rendement de blé de parcelles en sachant que la variance liée au 1er engrais était de 100 et celle liée au 2ème engrais était 121. Les résultats suivants ont été obtenus :
Engrais 1 : moyenne égale à 616 sur 64 parcelles mesurées parmi les 700 traitées.
Engrais 2 : moyenne égale à 620 sur 68 parcelles mesurées parmi les 800 traitées.
Peut-on dire que l'on obtient un meilleur rendement avec le 2ème engrais au seuil 2,5% ?
+ Calculer la p.value

Exercice 14

Un cours de programmation en langage basic est offert à 2 groupes d'individus d'un même programme. Les résultats suivants ont été obtenus à l'évaluation commune en fin de session :

Groupe	nj	mj	sj
A	54	73	11
B	58	77	11.5

Peut- on conclure au seuil de signification 0,05 que le groupe B est supérieur au groupe A ?

Exercice 15

On veut comparer les dépenses hebdomadaires moyennes liées à l'alimentation de base de familles de 4 personnes dans 2 régions de même caractéristiques sociologiques. On a obtenu :

Région	Nj	nj	mj ($)	sj ($)
A	300	38	90	5.5
B	350	40	87.3	4.5

L'hypothèse selon laquelle les dépenses hebdomadaires moyennes pour l'alimentation ne diffèrent pas de façon significative entre les familles de ces 2 régions est-elle vraisemblable au seuil de 5%?

Exercice 16

Un laboratoire a vérifié la résistance à l'éclatement (Kg/cm2) d'un réservoir à essence fabriqué par 2 entreprises. On sait que cette variable suit une loi normale. Les résultats suivants ont été obtenus :

A	218	220	222	220	216	224	221	224	219	216
B	218	216	217	218	219	220	219	223		

Peut-on conclure, au seuil de signification de 5% que les 2 types de réservoirs présentent une résistance moyenne à l'éclatement identique?

Exercice 17

On a mesuré la consommation d'oxygène de 12 individus, avant et après un entraînement de 300 heures, en les soumettant à un effort physique intense pendant 30 minutes. Les résultats obtenus sont :

avant	2.4	2.5	2.6	2.7	2.6	2.5	2.4	2.6	2.5	2.5	2.6	2.7
après	2.6	2.6	2.8	2.9	2.7	2.8	2.7	2.9	2.6	2.5	2.8	2.9

18,14

Peut-on conclure à l'amélioration de la condition physique après entraînement au seuil de 2% ? On admet que la consommation d'oxygène d'un individu suit une loi normale.

Exercice 18

On veut évaluer l'efficacité de 2 insecticides notés 1 et 2. Avec le premier sur 250 insectes vaporisés 175 ont été éliminés alors qu'avec le deuxième sur 300 insectes vaporisés 80 ont pu survivre. Peut-on conclure, au seuil de 5%, que le premier insecticide est moins efficace que le deuxième?

Exercice 19

On veut comparer la variabilité de 2 procédés utilisés pour mesurer un caractère quantitatif qui suit une loi normale. Deux échantillons aléatoires indépendants donnent les résultats suivants :

Procédé	nj	mj	SCj
A	13	1.5	49
B	21	2	159

Pouvez-vous conclure que la variabilité du procédé B est plus grande que celle du procédé A, au seuil de 5% ?

Exercice 20

On a mesuré une caractéristique quantitative, distribuée normalement, dans 3 échantillons aléatoires et indépendants et on a obtenu :

Echantillon 1	53	61	42	59	63	52						
Echantillon 2	83	46	73	85	73	63	66	54	60	84	62	59
Echantillon 3		48	94	76	73	65	21	48	62	94		

Les variances sont-elles comparables entre ces 3 échantillons, au seuil de 1% ?

Maintenant il est temps d'aller vous entrainer dans la rubrique « Testez vos connaissances» de cette partie 5, http://ecampus.isara.fr cours STAT_S4

Il faut réussir à faire 22 questions en 45mn le jour de l'examen!!!!!

Stats → EXERCICES

Exercice 1

1)

x_i	n_i	$x_i n_i$	$x_i^2 n_i$
2	1	2	4
4	2	8	32
6	3	18	108
8	2	16	128
10	1	10	100
Total →	9	54	372

moyenne : $\mu = 6$
écart type : $\sigma_x = 2{,}309$
variance : $\sigma_x^2 = 5{,}333$

2)

	2	4	4	6	6	6	8	8	10
2	2	3	3	4	4	4	5	5	6
4	3	4	4	5	5	5	6	6	7
4	3	4	4	5	5	5	6	6	7
6	4	5	5	6	6	6	7	7	8
6	4	5	5	6	6	6	7	7	8
6	4	5	5	6	6	6	7	7	8
8	5	6	6	7	7	7	8	8	9
8	5	6	6	7	7	7	8	8	9
10	6	7	7	8	8	8	9	9	10

Tab. des moy. des éch. 2 par 2

m	n_i	$m \cdot n_i$	$m^2 \cdot n_i$
2	1	2	4
3	4	12	36
4	10	40	160
5	16	80	400
6	19	114	684
7	16	112	784
8	10	80	640
9	4	36	324
10	1	10	100
Total →	81	486	3132

moyenne : $\mu = 6$
écart type : $\sigma_m = 1{,}633$
variance : $\sigma_m^2 = 2{,}667$

$\Downarrow$

NB : $E(x) = E(m) = \mu$
$\sigma_m^2 = \frac{1}{2} \sigma_x^2$

3) * même tableau sans la diagonale *

m	n_i	$m \cdot n_i$	$m^2 n_i$
2	0	×	×
3	4	12	36
4	8	32	128
5	16	80	400
6	16	96	576
7	16	112	784
8	8	64	512
9	4	36	324
10	0	×	×

moyenne $\mu = 6$
écart type $\sigma_m = 1{,}528$
variance $\sigma_m^2 = 2{,}333$

$\Downarrow$

NB : $E(x) = E(m) = \mu$
$\sigma_m^2 = K \frac{\sigma_x^2}{n}$ (avec $K = \frac{N-n}{N-1}$)

ou $\sigma_x^2 = \frac{\sum (x_i^2 n_i)}{n} - \bar{x}^2$

Exercice 2

1) Soit $1 \rightarrow$ caractère présent $\qquad p = 0,8$
$0 \rightarrow$ caractère absent $\qquad q = 0,2$

On étudie la distribution des 100 fréquences de taille $n = 2$

	1	1	1	1	1	1	1	1	0	0
1	1	1	1	1	1	1	1	1	0,5	0,5
1	1	1	1	1	1	1	1	1	0,5	0,5
1	1	1	1	1	1	1	1	1	0,5	0,5
1	1	1	1	1	1	1	1	1	0,5	0,5
1	1	1	1	1	1	1	1	1	0,5	0,5
1	1	1	1	1	1	1	1	1	0,5	0,5
1	1	1	1	1	1	1	1	1	0,5	0,5
1	1	1	1	1	1	1	1	1	0,5	0,5
0	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0
0	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0

f	n_i	$f \cdot n_i$	$f^2 \cdot n_i$
0	4	0	0
0,5	32	16	8
1	64	64	64
Total $\rightarrow$	100	80	72

moyenne : $E(f) = 0,8$
écart-type : $\sigma(f) = 0,283$
variance : $\sigma^2(f) = 0,080$

NB : $E(f) = E(x) = 0,8 = p$
$\sigma^2(f) = \frac{qp}{n}$

2) * même tableau sans la diagonale *

f	n_i	$f \cdot n_i$	$f^2 \cdot n_i$
0	2	0	0
0,5	32	16	8
1	56	56	56
Total	90	72	64

moyenne : $E(f) = 0,8$
écart-type : $\sigma(f) = 0,267$
variance : $\sigma^2(f) = 0,071$

Exercice 3

Nous étudions une variable quantitative suivant une loi normale $\mathcal{L}(X) = \mathcal{N}$
$\mu = 250$ et $\sigma_x^2 = 256$ donc $\sigma_x = 16$ $\qquad N = 340 \qquad n = 40 \qquad \alpha = 0,06$
Nous avons donc toutes les données nécessaires à l'application de la loi :

$$P\left[\mu - t_{1-\frac{\alpha}{2}}\,\sigma(m) < m < \mu + t_{1-\frac{\alpha}{2}}\,\sigma(m)\right] = 1 - \alpha$$

$\mu = 250$
$t_{1-\frac{\alpha}{2}} = t_{0,97} = 1,881$

$$\sigma(m) = \frac{\sigma(x)}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\sigma(m) = \frac{16}{\sqrt{40}} \sqrt{\frac{300}{339}} = \frac{8}{\sqrt{10}} \times \sqrt{\frac{100}{113}} = \frac{8\sqrt{10}}{\sqrt{113}}$$

Ainsi $P\left[245,5 < m < 254,5\right] = 0,94$

$\Rightarrow$ PHRASE TYPE

Si on considère tous les échantillons exhaustifs de taille 40, issus d'une population centrée sur 250, avec un écart-type de 16, alors 94% des moyennes des échantillons doivent se trouver entre 245,5 et 254,5

Exercice 4

1) La variable suit une loi normale dont les paramètres μ et σ_m sont connus. Nous pouvons donc appliquer :

$$P\left[\mu - t_{1-\frac{\alpha}{2}}\sigma_m < m < \mu + t_{1-\frac{\alpha}{2}}\sigma_m\right] = (1-\alpha)$$

avec $\sigma_m = \frac{\sigma}{\sqrt{n}}\sqrt{K}$ ← K ? A car [illegible] < 10% alors K = 1

$$\sigma_m = \frac{49}{5} = 9,8$$

on a aussi $\alpha = 0,10$ ainsi $t_{1-\frac{\alpha}{2}} = t_{0,95} = 1,645$

Ainsi $P\left[573,879 < m < 606,121\right] = 0,90$

2) $P\left[580 < m_{35} < 610\right] = 1-\alpha$ $\sigma_m = \frac{49}{\sqrt{35}} = 8,283$

$P\left[(580-590)/8,283 < T < (610-590)/8,283\right] = 1-\alpha$

$P\left[-1,21 < T < 2,42\right]$

Exercice 5

1) $N = \infty$ $n = 200$ $\alpha = 10\%$ $p = 0,03$

$\sigma^2 = \frac{pq}{n}$ donc $\sigma = \sqrt{\frac{pq}{n}}$ $t_{0,95} = 1,645$

$p_1 = p - t_{0,95} \times \sqrt{\frac{qp}{n}} = 0,03 - 1,645 \times \sqrt{\frac{0,03 \times 0,97}{200}} = 0,0102$

$p_2 = p + t_{0,95} \times \sqrt{\frac{qp}{n}} = 0,0498$

2)

Exercice 10

On a les infos n, m et S_x

On a donc que X suit une loi normale, pas comme on a pas σ_x^2, on utilise la loi de Student car $n<30$: $\mathcal{L}(m_{26}) = St(ddl)(E(m) ; \sigma(m))$

$ddl = n-1$

Niveau 1 $\quad t_{0,99}(25) = 2,485 \qquad \sigma_{(m)} = \frac{S_x}{\sqrt{25}} \sqrt{1} = 67,6$

$n \ll N$ alors $K = 1$

petite erreure

$$P[3521 - 2,485 \times 67,6 < \mu < 3521 + 2,485 \times 67,6] = 0,98$$

$$P[3353 < \mu < 3689] = 0,98$$

$ME_{abs} = 2,485 \times 67,6 = 168 \qquad ME_{rel} = \frac{168}{3521} \times 100 = 4,8\%$

Niveau 2 $\quad t_{0,99}(56) = 2,3263 \qquad \sigma_{(m)} = \frac{445}{\sqrt{56}} \sqrt{1}$

$$P[2680 < \mu < 2832] = 0,98$$

Exercice 9

1) On a $\alpha = 5\%$, $n = 20$, $m = 55$, $S_x = 5$, $E(X) = \mu = 50$, $N = 300$ et X le prix d'une pièce

On ne sait pas si $\mathcal{L}(X) = \mathcal{N}$, et $n < 30$.

euh... donc on peu pas résoudre.

2) Cette fois $n = 40$, $m = 56$ et $S_x = 6$

$n > 30$, on peut donc appliquer $\mathcal{L}(m_{40}) = \mathcal{N}(E(m) ; \sigma(m))$

echantillonnage : $t_{1-\frac{\alpha}{2}} = t_{0,975} = 1,96 \qquad \sigma(m) = \frac{6}{\sqrt{39}} \sqrt{\frac{300-40}{300-1}} = 0,896$

$$P[50 - 1,96 \times 0,896 < m_{40} < 50 + 1,96 \times 0,896] = 0,95$$

$$P[48,24 < m_{40} < 51,76] = 0,95$$

→ les prix ne sont pas respectés

estimation : $P[56 - 1,96 \times 0,896 < m_{40} < 50 + 1,96 \times 0,896] = 0,95$

→ pas respecté non plus

Alexis MONSAVOIR

TD 1

Reprise du cours :
- on a cherché

étude 1. Objectif 1 : être capable de prévoir les caractéristiques d'un échantillon dont on connait les caractéristiques de la pop échantillonné.

Approxim° : on définit une fourchette (un intervalle) et un risque
Requis : les paramètres de la pop.

Objectif 2 : être capable d'estimer les caractéristiques d'une pop connait les paramètres d'un échantillon

Approxim° : on définit un intervalle et un risque
Requis : les paramètres de la péchantillon

Variable quantitative

$\mathcal{L}(x) = \mathcal{N}$?
- oui → σ_x connu ?
  - oui → $\mathcal{L}(m_n) = \mathcal{N}(E;\sigma)$
  - non → $n > 30$?
    - oui → $\mathcal{L}(m_n) = \mathcal{N}(E;\sigma)$
    - non → $\mathcal{L}(m_n) = St(E(m), \sigma(m))$
- non → $n > 30$?
  - oui → $\mathcal{L}(m_n) = \mathcal{N}(E;\sigma)$
  - non → $\mathcal{L}(m_n) = ?$ → pas d'issue

Variable qualitative

p connu ?
- oui → np et $nq > 5$
  - oui → $\mathcal{L}(f_n) = \mathcal{N}(E(f);\sigma(f))$
  - non → $\mathcal{L}(f_n) = ?$ pas d'issue
- non → $n > 100$?
  - non → loi f
  - oui → $0,1 < f < 0,9$
    - oui → $\mathcal{L}(f_n) = \mathcal{N}(E(f);\sigma(f))$
    - non → loi χ^2

cf p.7
→ autre cond°

$$P[\mu - t_{1-\frac{\alpha}{2}}\,\sigma(m) < m_n < \mu + t\,\sigma(m)] = 1-\alpha \qquad P[m - t_\sigma(m) < \mu < m + t\,\sigma(m)] = 1-\alpha$$

$$\sigma(m) = \frac{\sigma_x}{\sqrt{n}}\sqrt{K}$$

$$P[p - t_{1-\frac{\alpha}{2}}\,\sigma(f) < f_n < p + t\,\sigma(f)] = 1-\alpha \qquad P[f - t\,\sigma(f) < p < f + t\,\sigma(f)] = 1-\alpha$$

$$\sigma(f) = \sqrt{\frac{pq}{n}}\sqrt{K} \qquad \hookrightarrow \sqrt{\frac{f(1-f)}{n}}\sqrt{K}$$

↑
A SAVOIR TRES BIEN PAR ♡
REPRODUIRE VIIIIIITE

* Exercice 1

5 c g   6 a b d   7 a

NB "n−1" sous entend "estimé"

* Ex 2

1 c
2 b e
3 → pas fait en cours
4 a d
5 b c
6 a d e
7 a d
8 c

NB : niveau de confiance = $1-\alpha$

* Exercice 3

1 25

2 14 %

3 144

4 → $\mu + t\,\overbrace{\frac{\sigma_x}{\sqrt{n}}\sqrt{K}} = \text{lim sup} = 1200 + 2{,}536 \times \overbrace{\frac{100}{4} \times \sqrt{\frac{1000-16}{1000-1}}} = 1200 + 2{,}576 \times \overbrace{25 \times \sqrt{K}}$

$\approx$ ou $1/2$

lim sup = ~~1257,795~~ 1263,91

6 → ME absolue $= t\,\sigma(m) = 1{,}960 \times \frac{100}{4\sqrt{2}} \times \sqrt{\frac{1200-36}{1199}} = 32{,}09$

8 → $\text{lim sup} = \mu + t\,\frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 1200 + 2{,}576\,\frac{100}{4\sqrt{2}}\sqrt{\frac{100-32}{99}} = 1237{,}74$

10 → $\text{lim inf} = \mu - t\,\frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 15 - 2{,}326\,\frac{1{,}2}{6}\sqrt{\frac{1000-36}{999}} = 14{,}54$

12 → ME abs $= t\,\sigma(m) = 2{,}054 \times \frac{1{,}2}{\sqrt{72}} \times \sqrt{\frac{1000-72}{999}} = 0{,}28$

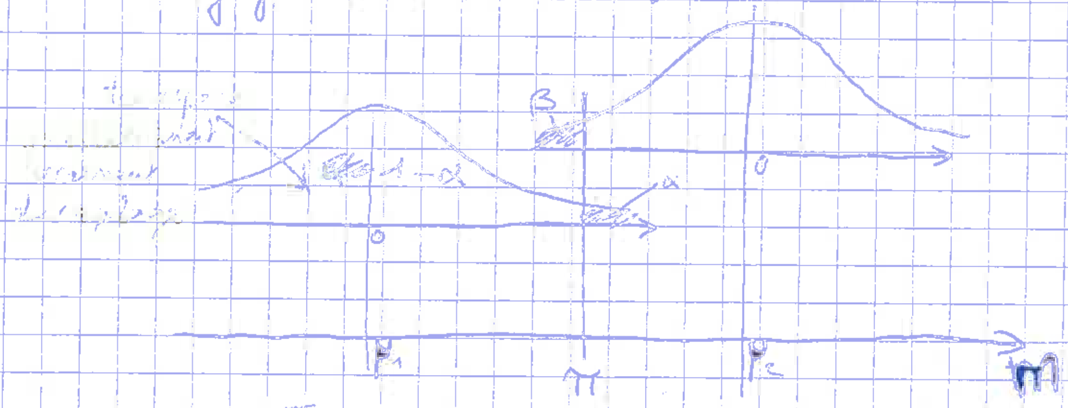
14 → $\sigma_m = \frac{\sigma_x}{\sqrt{n}}\sqrt{K} = \frac{100}{\sqrt{72}}\sqrt{\frac{300-72}{299}} = 10{,}29$

16 → ME abs =

Test paramétrique

① $N = 500$ $n = 60$ $\mu = 125g$ $\sigma_x = 1g$ surdosage $\to 0{,}7g$

déréglage à tard à ~~...~~ 5%



$$t_{0,95} = \frac{\Pi - \mu}{\sigma_m} \qquad t_{0,95} = \frac{\Pi - \mu}{\sigma_m} = 0{,}8289$$

$$\sigma_m = \frac{\sigma_x}{\sqrt{n}}\sqrt{K} \qquad \Pi = 0{,}8289 \times \sigma_m + \mu$$

$$\sigma_m = \frac{1}{\sqrt{60}} \times \sqrt{\frac{440}{499}} \qquad \boxed{\Pi = 125{,}1007}$$

on devrait trouver 125,20

$$\sigma_m = 0{,}1212$$

② $t_{0,90} = \dfrac{\Pi - \mu}{\sigma_m}$ $\sigma_m = \dfrac{2}{\sqrt{30}} \times 1$ $\Pi = t_{0,90} \times \dfrac{2}{\sqrt{30}} + 125$

0,8159

1,282

$$\boxed{\Pi = 125{,}468}$$

~~$t_\beta = \frac{125{,}23 - 126}{\frac{2}{\sqrt{30}}} = -1{,}9227$~~ ~~donc 0,9726~~

$$t_\beta = \frac{125{,}468 - 126}{\frac{2}{\sqrt{30}}} = -1{,}4566 \longrightarrow \text{d'où } \beta = 0{,}0721$$

car ds la table $\to$ 0,9265

et $1 - 0{,}9265 = 0{,}0721$

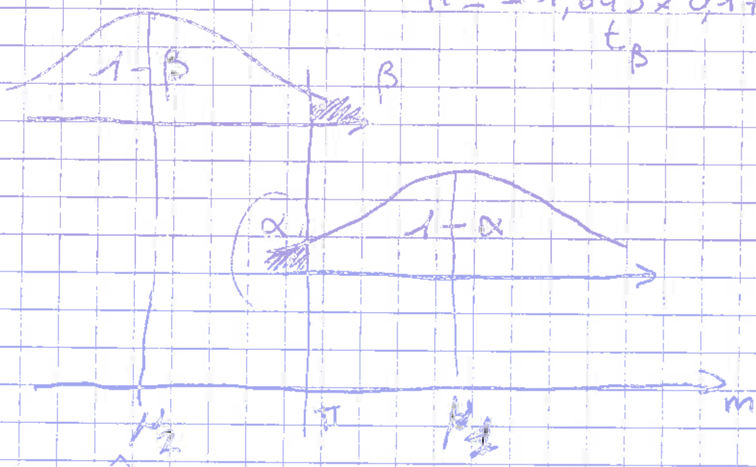
③ $t_{0,95} = \dfrac{\Pi - 125{,}7}{\sigma_m}$ $\sigma_m = \dfrac{1}{\sqrt{20}} \times \sqrt{\dfrac{480}{499}} = 0{,}179$

$t_{0,95} = 1{,}645$

$\Pi = 1{,}645 \times 0{,}179 + 125{,}7 = \cancel{126}{,}0678$

$\Pi = -1{,}645 \times 0{,}179 + 125{,}7 = 125{,}33$

t_β



6 → 0,0344

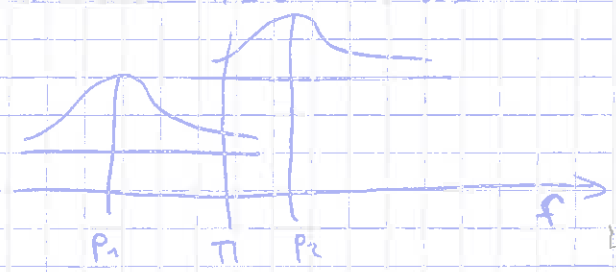
7 → 0,6062

3 → 125,33

σ_x $\hat{\sigma}_x$ S_x σ_m

"tout est possible avec moi"

⑤ $N = 1000$ $n = 60$ $\alpha = 5\%$



$t_{0,95} = \frac{\pi - p_1}{\sigma} = 0,8289$

$t_{0,95} = 1,645$ $t_{0,975} = 1,960$

$\pi = 0,8289 \times$

⑥ 20% défaut crainte → 40% $N = 1000$

$n = 60$

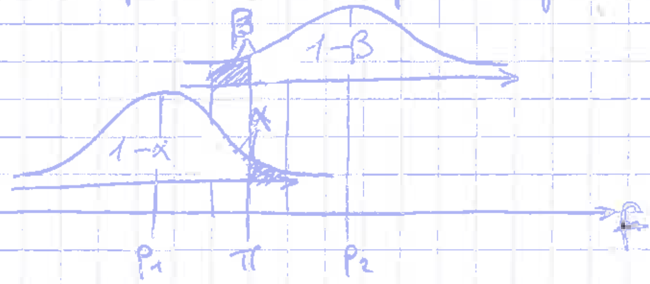
accepte ~~d'accepter~~ à tord 5% de détecter

→ Risque du client ?

2^e espèce → se tromper en rejetant H_1

$\alpha = 5\%$

np et nq >



$t_{1-\alpha} = \frac{\pi - p_1}{\sigma_1(f)} = 1,645$

$\pi = 1,645 \times \sigma_1(f) + p_1$

$\boxed{\pi = 0,28}$

$\sigma_1(f) = \sqrt{\frac{p_1 \times q_1}{n}}$

$=$

$t_\beta = \frac{\pi - p_2}{\sigma_2(f)}$

$\sigma_2(f) = \sqrt{\frac{p_2\, q_2}{n}}$

$t_\beta = -1,81914$

$\simeq -1,82$

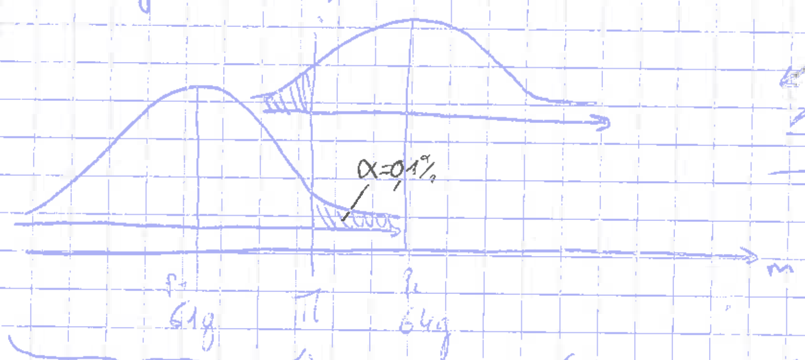
$\beta = 1 - 0,9656$

$\beta = 0,0344$

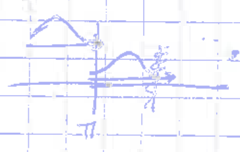
TD Stat

9 $\cancel{E_m} = 64g$ $\quad \sigma_1(m) = \frac{\sigma}{\sqrt{36}} = \frac{5}{\sqrt{36}} = 0,83$

$\mu_2 = 61g$ $\quad \sigma_2(m) =$



en fait ça devrait être



$t_{calc} = \frac{m - E_m}{\sigma_m} = \frac{61-64}{0,83} = -3,6$

11 $E_m = 64g$ $\quad \sigma(x) = 5$ $\quad \mathcal{L}(X) = \mathcal{N}(E_x ; \sigma_x)$ $\quad \sigma(m) = \frac{5}{\sqrt{18}} = 1,18$

$m = 61g$

12 $E_m = 64g$ $\quad \sigma(x) = 5$

$m = 61g$ $\quad t_{théo} = t_{1-\frac{\alpha}{2}} = t_{1-\frac{0,01}{2}} = t_{0,995} = 1,695$

$\alpha = 0,01$

$\boxed{t_{théo} = 2,576}$

15 $t_{théo} = t_{1-\frac{\alpha}{2}} = t_{0,99} = 2,326$

14 $\sigma(m) = \frac{\sigma(x)}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} =$

15 $E_m = 90g$ $\quad m = 92g$ $\quad \sigma_m = \frac{\sigma}{\sqrt{n}} = \frac{0,845}{\sqrt{36}}$ $\quad = 0,1344$ $\quad \sigma^2 = \frac{SCE}{n(n-1)}$

$t =$

13 $\sigma(m) = \sqrt{\frac{SCE}{n(n-1)}} = \sqrt{\frac{900}{36 \times 35}} = 0,845$

18

21 et 24 impossibles

STAT – TD 2