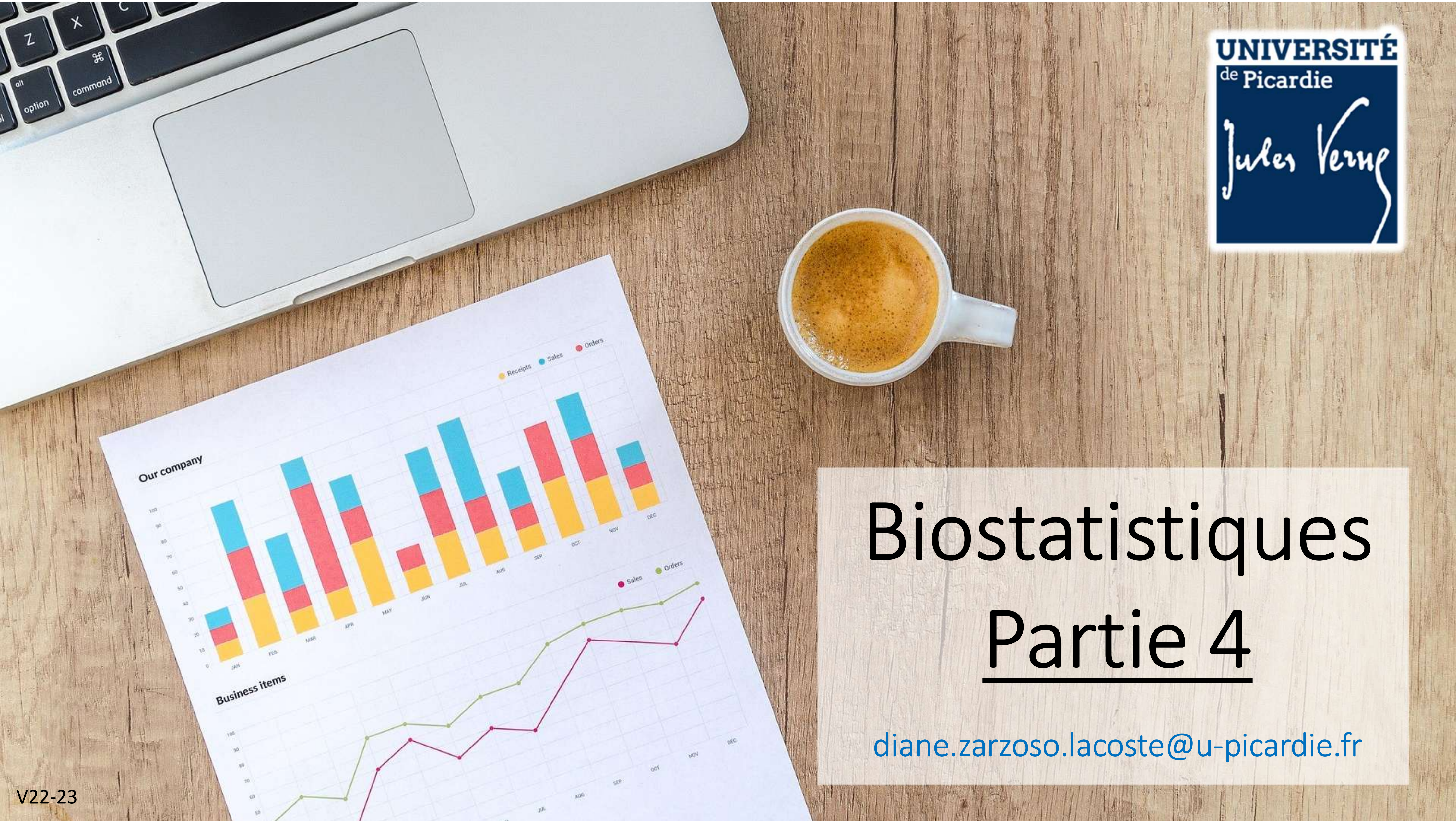
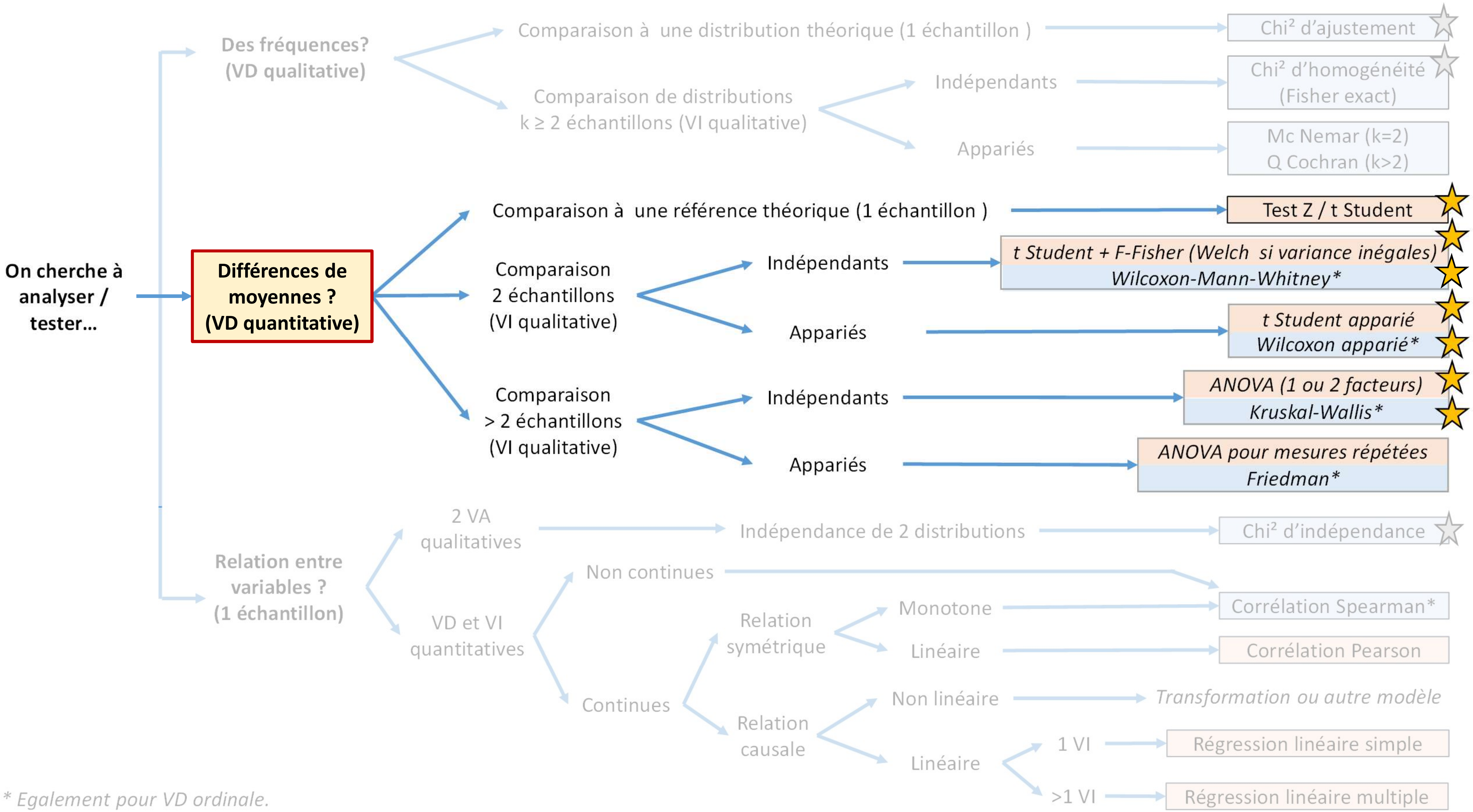


UNIVERSITÉ de Picardie Jules Verne

Biostatistiques
Partie 4

diane.zarzoso.lacoste@u-picardie.fr

V22-23



On cherche à analyser / tester…
Des fréquences? (VD qualitative)
Comparaison à une distribution théorique (1 échantillon)
Chi² d'ajustement
Comparaison de distributions k ≥ 2 échantillons (VI qualitative)
Indépendants
Chi² d'homogénéité (Fisher exact)
Appariés
Mc Nemar (k=2) Q Cochran (k>2)
Différences de moyennes ? (VD quantitative)
Comparaison à une référence théorique (1 échantillon)
Test Z / t Student
Comparaison 2 échantillons (VI qualitative)
Indépendants
t Student + F-Fisher (Welch si variance inégales)
Wilcoxon-Mann-Whitney*
Appariés
t Student apparié
Wilcoxon apparié*
Comparaison > 2 échantillons (VI qualitative)
Indépendants
ANOVA (1 ou 2 facteurs)
Kruskal-Wallis*
Appariés
ANOVA pour mesures répétées
Friedman*
Relation entre variables ? (1 échantillon)
2 VA qualitatives
Indépendance de 2 distributions
Chi² d'indépendance
VD et VI quantitatives
Non continues
Corrélation Spearman*
Continues
Relation symétrique
Monotone
Linéaire
Corrélation Pearson
Relation causale
Non linéaire
Transformation ou autre modèle
Linéaire
1 VI
Régression linéaire simple
>1 VI
Régression linéaire multiple
* Egalement pour VD ordinale.

Plan du cours

Partie 4. Les tests sur les différences de moyennes

1. Test de conformité : moyenne observée vs théorique

2. Test d'homogénéité : Comparer les moyennes de plusieurs échantillons

2.1. Comparaison de deux échantillons indépendants

2.1.1. Test paramétrique : t de Student (Welch)

(Comparaison de variances : test F de Fisher)

2.1.2. Test non paramétrique : Mann-Whitney

2.3. Comparaison de deux échantillons appariés

2.3.1. Test paramétrique : t de Student apparié

2.3.2. Test non paramétrique : Wilcoxon apparié

2.4. Comparaison de trois échantillons ou plus

2.4.1. Test paramétrique : ANOVA 1 facteur

2.4.2. Test non paramétrique : Kruskal-Wallis

Test de conformité : moyenne observée vs théorique

Les données : On mesure 1 VA X sur 1 échantillon de n individus.

ECQ la moyenne $\overline{x}$ de X dans l'échantillon est compatible avec la moyenne théorique μ de la population statistique de référence ?

Les hypothèses :

$H_0 : \overline{x} = \mu$

La différence observée entre $\overline{x}$ et μ est suffisamment faible pour être explicable par les fluctuations d'échantillonnage.

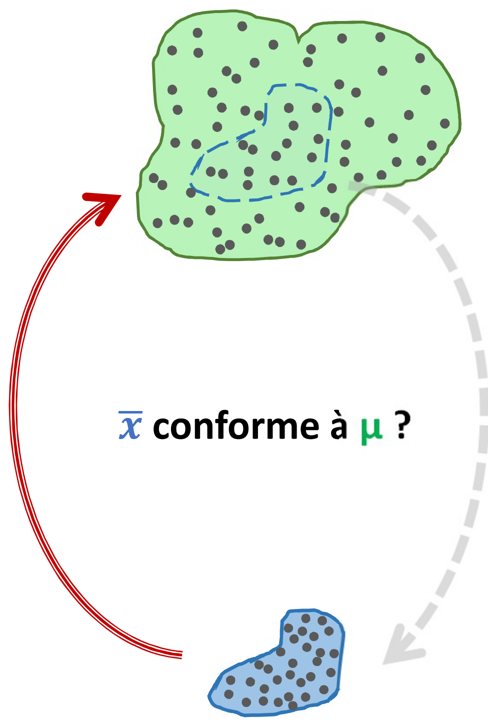
→ *L'échantillon appartient vraisemblablement à la population de référence.*

$H_1 : \overline{x} \neq \mu$ *(ou unilatéral : $\overline{x}$ > ou < μ)*

La différence est trop importante pour que le hasard à lui seul puisse l'expliquer.

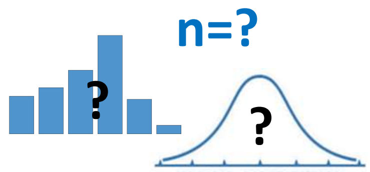
→ *L'échantillon provient vraisemblablement d'1 autre population statistique.*

Population de moyenne théorique μ connue



1 échantillon
moyenne $\bar{x}$
variance S^2

Choix du test ~ taille et distribution de l'échantillon →



Test de conformité : moyenne observée vs théorique

Infos sur la population :

Infos sur l'échantillon :

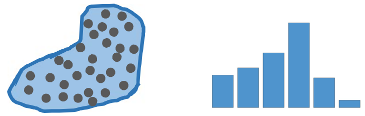
Choix du test :

μ est **connu mais**
σ^2 est **inconnue**

+

Petit (n≤30),
Distri. ~ loi Normale

Cas 1 :
Approximation par
les lois de Student :
Test t de Student



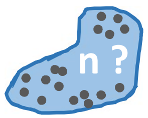
Grand (n>30),
Distri. quelconque

Cas 2 :
Approximation par la
loi Normale CR (TLC) :
Test Z

OU

μ **<u>et</u>** σ^2 sont **connues**

+



Taille n quelconque

Cas 3 :
Approximation par
la loi Normale CR :
Test Z

Test de conformité : moyenne observée vs théorique

Quel que soit le test, la méthode reste la même :

Calculer la valeur d'une statistique de test à partir des observations de l'échantillon (**St_{obs}**)

et voir dans quelle mesure elle s'éloigne de la valeur théorique attendue sous H_0 (**$St_{théo}$**).

<u>Rappel</u> *: la/les valeur(s) critique(s) de **$St_{théo}$** délimite(nt) la/les borne(s) des zones de rejet et d'acceptation de H_0 pour un seuil de risque α choisi.*

La statistique de test : St quantifie l'écart entre l'observation $\bar{x}$ et la référence μ.

$$St = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

t ou Z en fonction du test (à chercher dans la table Student ou LNCR)

$\bar{x}$: moy. éch.

μ : moy. pop.

n : effectif éch.

σ : écart type pop.

Si **σ** inconnu, on le remplace par son **estimation** à partir de l'échantillon **S* (écart type corrigé).**

Test de conformité : moyenne observée vs théorique

Règle de décision :

Pour un **seuil α** fixé, on **comparera** ensuite **|St_{obs}| :**

- **|t_{obs}|** → aux valeurs de la table de **Student** ($t_{théo}$) à **n-1** degré de liberté **si n < 30,**
- Ou **|Z_{obs}|** → aux valeurs de la table de la Loi **Normale Centrée Réduite** ($Z_{théo}$) si **n ≥ 30.**

→ **On rejettera H₀** acceptera H_1 **si** :

- $|St_{obs}| \geq |St_{théo}|$
- p-value < seuil α

***NB* : Plus la valeur de $|St_{obs}|$ sera grande et ≥ 2, plus la différence entre $\overline{x}$ et μ sera importante (éloignée de 0), et donc moins H_0 sera crédible.**

Test de conformité : moyenne observée vs théorique

Conditions d'application :

- **Test t de Student :** La distribution de la VA X ~ Normale (symétrique unimodale)
  - ✓ Vérifier graphiquement (histogramme, density et/ou qq plot)
  - ✓ Tester un défaut de normalité : Shapiro-Wilk

- **Test Z :** aucune (à part grand effectif ou variance population connue)!

NB : → Tests très peu restrictifs en terme de conditions.

→ Dans le doute, optez toujours pour un test *t*-Student car les lois de Student tendent vers la loi Normale pour les grands (n> 30) échantillons (TLC).

Test de conformité : moyenne observée vs théorique

Application :

On sait que la population de grande musaraigne (*Blarina brevicauda*) possède une masse moyenne de 23.3 g.

On a échantillonné aléatoirement 16 individus et déterminé leur masse :

20, 30, 26, 28, 33, 28, 15, 18, 32, 33, 32, 29, 24, 35, 34, 23

On suppose ici que la variable Masse suit une loi Normale.

D'un point de vue statistique, ECQ l'échantillon appartient à la population de référence ?

→ *Revoir la réalisation du test « à la main » en vidéo sur Moodle* ←

Test de conformité : moyenne observée vs théorique

Solution :

Données:

μ = 23.3

n = **16**, $\overline{x}$ = 27.5

S²* = 36.4, **S***= 6.03

Hypothèses :

$H_0 : \bar{x} = \mu$

$H_1 : \bar{x} \neq \mu$

Statistique :

$$t = \frac{\bar{X} - \mu}{S^* / \sqrt{n}}$$

Table de la loi de Student :

→ $t_{théo\_0.025\_15}$ = 2.131

α/2 (test bilatéral) = 0.025

ddl = 16 – 1 = 15

n \ α	0.45	0.3	0.2	0.1	0.05	0.025	0.01	0.001
1	0.158	0.727	1.376	3.078	6.314	12.706	31.821	318.309
14	0.128	0.537	0.868	1.345	1.761	2.145	2.624	3.787
15	0.128	0.536	0.866	1.341	1.753	$\boxed{2.131}$	2.602	3.733
16	0.128	0.535	0.865	1.337	1.746	2.120	2.583	3.686

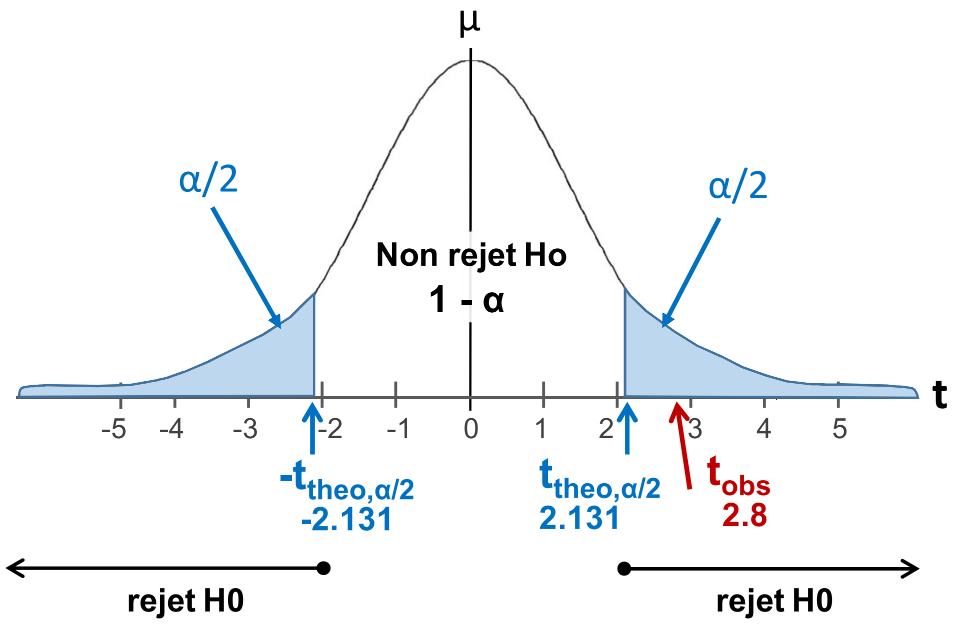
Test de conformité : moyenne observée vs théorique

Solution :

- Distribution théorique de t sous H_0 : Zones de rejet délimitées par les valeurs de $\mathbf{t_{théo.\alpha/2}}$.

- Calcul de $\mathbf{t_{obs}}$ à partir des données :

$$t_{obs} = \frac{27.5 - 23.3}{6.03/\sqrt{16}} = 4.2 / 1.5 = 2.8$$



- Comparaison $\mathbf{t_{obs}}$ à la distribution théorique de t sous H_0

Règle de décision :

- $\mathbf{t_{obs}}$ est dans la (les) zone(s) de rejet de H_0
- $|\, \mathbf{t_{obs}} \,| = 2.8 \geq |\mathbf{t_{théo}}| = 2.13$

Rejet de H_0 au risque α de se tromper

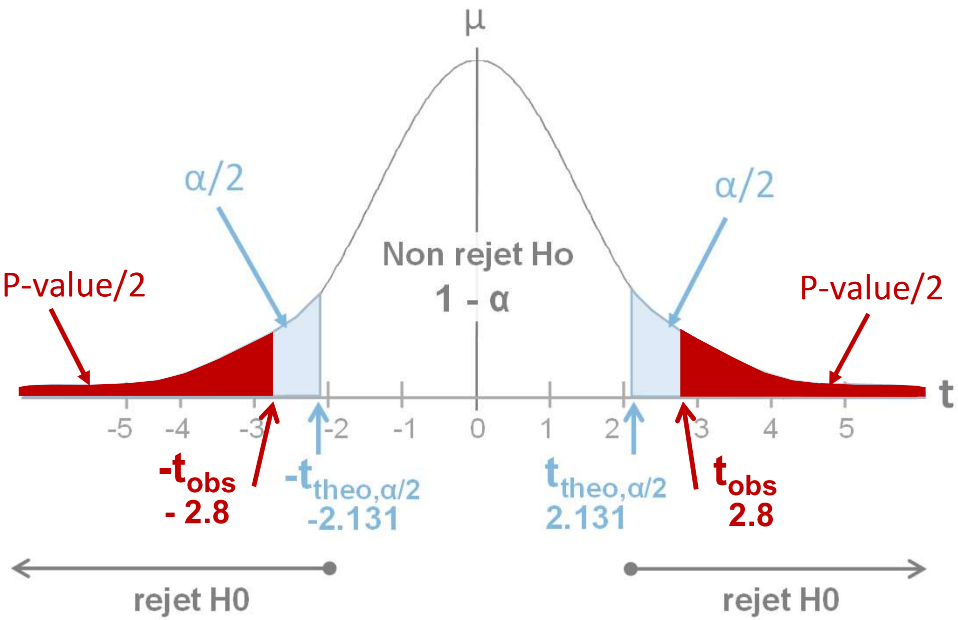
L'échantillon n'appartient pas à la population statistique de référence

Test de conformité : moyenne observée vs théorique

Et la p-value ??

- Risque exact de « rejeter à tort H_0 », calculé **d'après les observations** (échantillon)!!

- Aire sous la courbe (rouge) → valeurs de t au moins aussi extrêmes que t_{obs}.



NB: La répartition de cette aire dépend de H1

- Test bilatéral (H_1: $\bar{x} \neq \mu$) , ½ de p-value de part et d'autre de l'intervalle [$-t_{obs}$, t_{obs}]

- Test unilatéral (H_1 : $\bar{x} > ou < \mu$) → p-value entièrement à gauche ou droite de t_{obs}