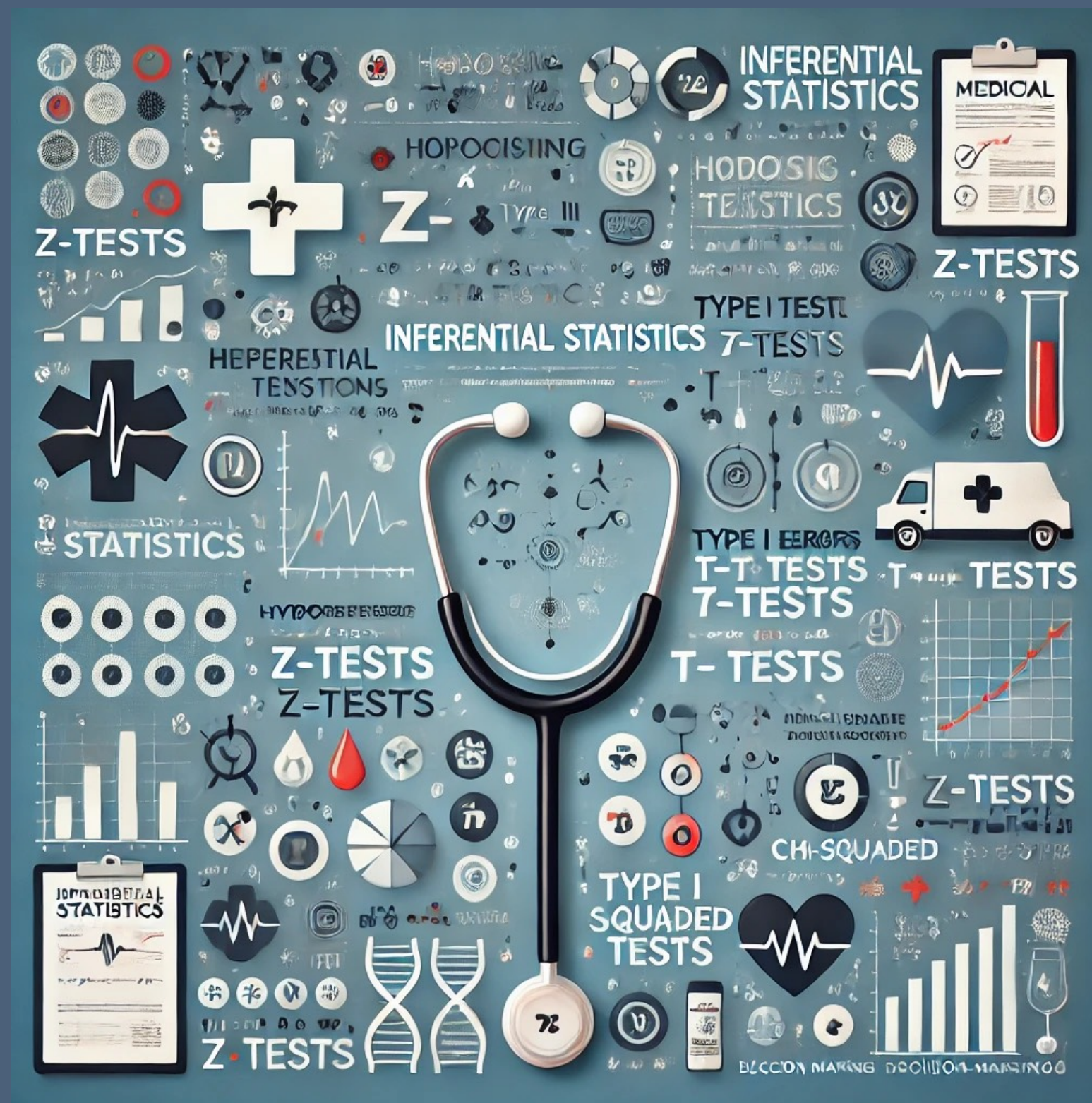


Les statistiques déductives

aka les goats de la biostat'



Définition des hypothèses

H0 et H1

- **Hypothèse nulle** (H0) : suppose qu'il **n'y a pas de différence significative** ou de relation particulière entre les groupes ou les variables étudiées. = les variations observées sont dues au hasard.
- **Hypothèse alternative** (H1) : suppose qu'il **existe une différence significative** ou une relation.
- **Objectif** : décider si l'hypothèse nulle (H0) doit être rejetée ou non. ++

6 étapes clés du test d'hypothèse



- 1. Définir H_0 et H_1 ++
- 2. Choisir le test statistique adapté (variables qualitatives ou quantitatives ?)
- 3. Fixer le risque d'erreur alpha : en général **5% +++**, correspond à la probabilité de rejeter H_0 alors qu'elle est vraie.
- 4. Recueillir les données
- 5. Comparer Z_c (calculé, toujours donné dans l'énoncé) et Z_t (théorique, à trouver dans la table) **+++**, si $Z_c > Z_t$ alors on rejette H_0
- 6. Interprétation des résultats

Risques α et β

- **Risque α** (erreur de type 1) : on rejette H_0 alors qu'elle est vraie.

exemple : on conclut que le médicament est efficace alors qu'il ne l'est pas réellement.

- **Risque β** (erreur de type 2) : on accepte H_0 alors qu'elle est fautive.

exemple : on conclut que le médicament n'est pas efficace alors qu'il l'est en réalité

- La puissance du test, qui est la probabilité de rejeter H_0 si H_1 est vraie, est donnée par **$1-\beta$** .



Tests pour des variables qualitatives

- Test de comparaison des pourcentages ++
- Test du chi carré (X^2) ++

Test de comparaison des pourcentages

- Z correspond à l'écart réduit ε

- Méthode pour trouver ε dans la table :

1. Déterminer le risque α : donné par l'énoncé sinon prendre comme valeur 5% soit 0,05.

2. On regarde le dixième de α sur les lignes et le centième de α sur les colonnes. ε se trouve à l'intersection.

exemple : si $\alpha = 0,05$ alors 0 (dixième) pour la ligne et 5 (centième) pour la colonne, donc $\varepsilon = 1,96$.

		0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	∞	2,576	2,326	2,17	2,054	1,96	1,881	1,812	1,751	1,695
0,1	1,645	1,598	1,555	1,514	1,476	1,44	1,405	1,372	1,341	1,311
0,2	1,282	1,254	1,227	1,2	1,175	1,15	1,126	1,103	1,08	1,058
0,3	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,86
0,4	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,69
0,5	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,6	0,524	0,51	0,496	0,482	0,468	0,454	0,44	0,426	0,412	0,399
0,7	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,8	0,253	0,24	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,9	0,126	0,113	0,1	0,088	0,075	0,063	0,05	0,038	0,025	0,013

Test du chi carré (X^2)

À partir de 2 lignes et 2 colonnes

- Méthode pour trouver X^2 t dans la table :

1. Déterminer le risque α : donné par l'énoncé sinon prendre comme valeur 5% soit 0,05.

2. Calculer le DDL (degré de liberté) avec la formule suivante : (nombre de lignes -1) * (nombre de colonnes -1)

X^2 t se trouve à l'intersection entre α et le DDL.

Exemple : exposition au benzène et leucémie

	Leucémie	Non leucémie	Total
Expo	15	485	500
Non expo	20	980	1000
Total	35	1465	1500

exemple : exposition au benzène et leucémie

nombre de lignes = 2

nombre de colonnes = 2

(On ne prends pas le total en compte !!)

$$DDL = (2-1) * (2-1) = 1$$

X^2 t se trouve à l'intersection entre le DDL = 1

et $\alpha = 0,05$ donc X^2 t = 3,84

		α								
		0,9	0,5	0,3	0,2	0,1	0,05	0,02	0,01	0,001
ddl	1	0,016	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
	2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,21	13,815
	3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
	4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
	5	1,61	4,351	6,064	7,289	9,236	11,07	13,388	15,086	20,515
	6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
	7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
	8	3,49	7,344	9,524	11,03	13,362	15,507	18,168	20,09	26,125
	9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
	10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
	11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
	12	6,304	11,34	14,011	15,812	18,549	21,026	24,054	26,217	32,909
	13	7,042	12,34	15,119	16,985	19,812	22,362	25,472	27,688	34,528
	14	7,79	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
	15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
	16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32	39,252
	17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,79

Tests pour des variables quantitatives en lien avec des variables qualitatives



- Test de comparaison des moyennes ++
- Test T de student ++

Test de comparaison des moyennes

Pour des grands échantillons (n_1 et $n_2 > 30$)

- **Z correspond à l'écart réduit ε**

Exemple : On cherche à comparer le taux de T3 libre chez les femmes prenant un contraceptif oral et celles qui n'en prennent pas. Après tirage au sort on obtient :

- Femmes sans c.o : **$n_1 = 50$** ; $m_1 = 2$ nmol
- Femmes avec c.o : **$n_2 = 33$** ; $m_2 = 2,5$ nmol
- **Variable 1** : prise ou non de la pilule : qualitatif
- **Variable 2** : dosage de T3 : quantitatif
- **n_1 et $n_2 > 30$**
- → Test de comparaison de moyennes
- $\alpha = 5\%$ (valeur de référence) donc $\varepsilon_t = 1,96$ (d'après la table)
- $\varepsilon_c = 6,94$ (la valeur d' ε_c vous sera toujours donnée dans l'énoncé) **$\varepsilon_c > \varepsilon_t$ donc rejet de H_0** , donc la prise de contraceptifs oraux augmentent le taux de T3 libre chez les femmes.

Test T de student

Pour des petits échantillons (n_1 ou $n_2 < 30$)

- Z correspond à T
- **Tt** : lu dans la table du T de student
- même raisonnement que pour le test du chi carré (X^2) avec X^2_t pour trouver **Tt** :

1. Déterminer le risque d'erreur α

2. Calculer DDL donné par la formule suivante : $(n_1-1) + (n_2-1)$ **ATTENTION** pas la même formule que le DDL du chi carré !!

Exemple : Soient 15 femmes obèses et 12 femmes de poids normal. On mesure le taux de corticoïde sanguin moyen dans chaque groupe. L'obésité a-t-elle une influence sur le taux de corticoïde ?

n1 = 15 ; $m_1 = 6,3$

n2 = 12 ; $m_2 = 4,5$

H_0 : m_1 et m_2 ne sont pas différents dans les 2 groupes

Variable 1 : obèse ou non : qualitatif

Variable 2 : taux de corticoïde : quantitatif

n 1 ou n 2 < 30 à T de student

$\alpha = 5\%$ (valeur de référence) ; $DDL = (15-1) + (12-1) = 25$ donc $T_t = 2,06$ (valeur qui correspond à l'intersection sur la table entre α et DDL)

$T_c = 2,92$ (valeur donnée par l'énoncé), **$T_c > T_t$ donc on rejette H_0 au seuil 5%**, il existe donc une relation claire entre l'obésité et le taux de corticoïde au niveau de cet échantillon.

Lien entre deux variables quantitatives

définitions à connaître 

- **corrélation** : évaluation de la liaison entre 2 variables quantitatives
- **!!** corrélation \neq causalité **!!**

corrélation : il existe un lien : l'âge et le cholestérol sont liés

causalité : l'un est la conséquence de l'autre : l'âge cause le cholestérol

- **régression** : méthode mathématique permettant d'expliquer les relations entre les variables observées

Tests paramétriques et non paramétriques

- **Test paramétrique** : test à forte contrainte, car il n'est fiable que si les données suivent une distribution selon une loi normale.
- **Test non paramétrique** : test qui ne précise pas les conditions que doivent remplir les paramètres de la population dont a été extrait l'échantillon.
- On utilise obligatoirement un test **non paramétrique ++** quand les **effectifs** sont très **faibles ($4 < n < 12$) +++**
- Pour les variables quantitatives, on utilise obligatoirement un test non paramétrique si les effectifs sont inférieurs à 5 car les populations ne sont plus distribuées normalement
- **Variables qualitatives** : Comparaison des pourcentages, Chi carré
- **Variables quantitatives** : r' de Spearman
- **Variables qualitatives-quantitatives** : U de Mann et Whitney

Tableau récap des tests !!++

	Variables quantitatives	Variables qualitatives	Variables qualitative - quantitative
4 < n < 12 (non paramétrique)	r' de Spearman	Comparaison des pourcentages X^2	U de Mann et Whitney
12 ≤ n < 30	Coefficient de corrélacion r' de Spearman	Comparaison des pourcentages X^2	T de student U de Mann et Whitney
30 ≤ n	Coefficient de corrélacion r' de Spearman	Comparaison des pourcentages X^2	Comparaison des moyennes T de student U de Mann et Whitney

Fin du cours 🎉🎉

Pour toute question vous pouvez me contacter sur :

-messenger : Paolina Lozato

-forum et Discord : Neuro'lina

Surtout n'hésitez pas, on est là pour vous aider !!

Force, courage et honneur