Traitement de données et modélisation avec

https://elearn.univ-tlemcen.dz/

SCAN ME

tewfik.mahdjoub@univ-tlemcen.dz

SCAN ME

Chapitre 2: Tests statistiques

Section 1: Tests de conformité

1. Position du problème.

Un laboratoire pharmaceutique indique que chacun de ses comprimés doit contenir 2.5g de substance active.
Un échantillon de100 comprimés est analysé. Les résultats sont contenus dans le fichier suivant : **Exemple_1.csv.**
Avec un risque $\alpha = 5\%$, peut-on dire que le médicament respecte les spécifications.

Pour répondre à la question, on réalise un test statistique.

2. Test statistique classique.

<u>2.1. Hypothèse nulle **H₀**</u>
Un test statistique est une méthode qui permet d'accepter ou pas une hypothèse appelée **hypothèse nulle notée H_0.**

	A	B	C	D	E
1	Comprime	Echantillon_1	Echantillon_2	Echantillon_3	
2	1	2,44	2,54	1,95	
3	2	2,24	2,78	2,6	
4	3	2,67	2,24	2,76	
5	4	2,39	2,94	2,97	
6	5	2,39	1,74	2,2	
7	6	2,17	1,91	2,96	
8	7	2,11	3,22	2,55	
9	8	2,34	2,49	2,92	
10	9	2,54	2,14	2,58	
11	10	2,98	2,95	2,33	
12	11	2,1	2,36	2,52	
13	12	2,44	2,01	3,07	
14	13	2,73	1,76	2,39	
15	14	2,13	1,83	2,95	
16	15	3,41	2,9	2,81	
17	16	2,22	2,94	1,87	
18	17	2,04	2,96	2,92	
19	18	2,63	2,21	2,91	
20	19	2,82	2,53	2,58	
21	20	1,96	2,52	3,17	
22	21	2,96	2,18	3,28	
23	22	3,57	3,39	2,08	
24	23	3,01	2,04	2,97	
25	24	2,49	2,37	2,43	
26	25	2,35	2,37	3,21	

Exemple_1

L'hypothèse concerne les paramètres statistiques de la population d'où est extrait l'échantillon.

Dans ce cas, l'échantillon 1 de 100 comprimés est extrait d'une population de moyenne μ et d'écart-type σ, tous deux inconnus.

On va comparer la moyenne μ à la moyenne de référence $\mu_0 = 2.5$.
On pose alors :

$$\mathbf{H_0} : \mu = 2.5$$

<u>2.2. Test de conformité.</u>
Lorsqu'on compare la moyenne d'une population à une moyenne de référence, le test est appelé **un test de conformité**.

2.3. Risque d'erreur α.
Après les calculs, il est possible qu'on va conclure que l'échantillon analysé ne soit pas extrait d'une population de moyenne $\mu = 2.5$ alors qu'il est extrait de cette population.
Le risque de commettre cette erreur est noté α : c'est **le risque d'erreur $\boldsymbol{\alpha}$.**
Dans ce cas $\alpha = 5\%$, : on a 5% de chances de rejeter **H_0** alors qu'elle est vraie.

2.4. Hypothèse alternative **H_1**.
Si on doit rejeter **H_0**, on doit alors accepter une autre hypothèse, dite **hypothèse alternative** notée **H_1**

L'hypothèse alternative H_1 n'est pas nécessairement la négation de l'hypothèse nulle H_0

Dans ce cas, on peut avoir :

$$\mathbf{H_1 :}\ \mu \neq 2.5$$

ou

$$\mathbf{H_1 :}\ \mu > 2.5$$

ou

$$\mathbf{H_1 :}\ \mu < 2.5$$

2.5 Test bilatéral/unilatéral.
$\boldsymbol{\mu} \neq 2.5$ signifie 2 cas possibles : $\boldsymbol{\mu} > 2.5$ ou $\boldsymbol{\mu} < 2.5$. Dans ce cas, le risque alfa est partagé en deux : on a $\frac{\alpha}{2}$ **%** de chances de rejeter **H_0** en affirmant que $\boldsymbol{\mu} > 2.5$ et $\frac{\alpha}{2}$ **%** de chances de rejeter **H_0** en affirmant que $\boldsymbol{\mu} < 2.5$. **Le test est bilatéral**.
Si dans **H_1** il n'y a qu'une seule inégalité stricte, **le test est unilatéral.**

Par la suite, on prendra les hypothèses :

$$\mathbf{H_0 :}\ \mu = 2.5$$
$$\mathbf{H_1 :}\ \mu \neq 2.5$$

Le test qu'on souhaite faire est un test bilatéral de conformité.

2.6. Statistique du test.
On calcule :

$$S = \frac{\bar{x} - \mu_0}{\frac{s_e}{\sqrt{n}}}$$

Pour chaque échantillon analysé, une valeur de S est obtenue**. S suit une loi de Student à (n-1) degré de liberté**.

On calcule alors la moyenne et l'écart-type de l'échantillon 1 :

Commandes 1 :

```
> Ex1 = read.csv ("~/Exemple_1.csv", header=T, sep==";", dec==",")

> (x_barre = mean (Ex1$Echantillon_1))
[1] 2.6091
```

```
> (variance = var(Ex1$Echantillon_1) *99/100)
[1] 0.1636482

> (se=sqrt (variance))
[1] 0.4045345

> (S= (x_barre-2.5) / (se/sqrt(100)))
[1] 2.696927
```

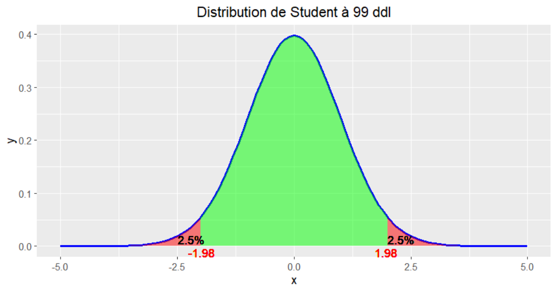
Donc S=2.696927≈ 2.69

2.7. Zone d'acceptation.
La zone d'acceptation pour un test bilatéral avec un risque $\alpha = 5\%$ se calcule par :

Commande 2 :

```
> qt (0.025,99)
[1] -1.984217
```

On a donc :



2.8. Conclusion.

$$S \approx 2.69 \notin [-1.98\,, 1.98]$$

On rejette donc **H₀**. On ne peut pas dire que l'échantillon choisi est extrait d'une population de moyenne 2.5.

3. Test bilatéral de confirmité sous R.

Sous R, ce test se fait par la commande :

Commandes 3 :

```
> t.test (Ex1$Echantillon_1, mu=2.5,
alternative="two.sided", conf.level =
0.95)

        One Sample t-test

data:  Ex1$Echantillon_1
t = 2.6834, df = 99, p-value = 0.008543
alternative hypothesis: true mean is not
equal to 2.5
95 percent confidence interval:
 2.528427 2.689773
sample estimates:
mean of x
  2.6091
```

✓ Le test de Student s'écrit **t.test.**
✓ **mu :** moyenne de référence.
✓ **alternative="two.sided" :** test bilateral.
✓ **conf.level = 0.95 :** risque d'erreur α.

Le test retourne les informations suivantes :

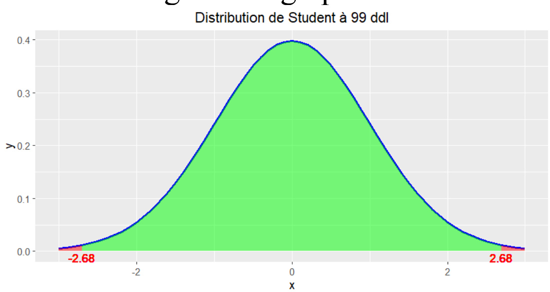
One Sample t-test
C'est un test à 1 seul échantillon

data: Ex1$Echantillon_1
Les données sont dans la colonne Echantillon_1 du tableau Ex1.

t = 2.6834
C'est la statistique calculée.

df = 99
C'est le degré de liberté de la loi de Student (degree of freedom)

p-value = 0.008543
L'aire en rouge sur le graphe suivant :



Commandes 4 :

```
> 2*pt(-2.6834, 99)
[1] 0.008542769
```

95 percent confidence interval:
2.528427 2.689773
L'échantillon analysé appartient à une population dont la moyenne est dans l'intervalle [2.528427 , 2.689773] pour un risque de 5%.

sample estimates:
mean of x
2.6091
C'est la Moyenne calculée de l'échantillon.

De façon générale, pour conclure un test statistique on regarde la p-value :

si p-value $< \alpha$, on rejette **H₀.**

si p-value $\geq \alpha$, on accepte **H₀**.

4. Exercice

Que peut-on conclure pour les échantillons 2 et 3?