

STAT' AIRWAYS ✈️ : ESCALE DE 30'
VEUILLEZ DETACHER VOTRE CEINTURE

Statistiques Descriptives

PAR TRYPTOPHRANE



Programme TTR (pas complet)



I- Variables

II- Representation de variables quantitatives

III- Estimation statistique

IV- Loi de Gauss

Introduction

Biostatistique : appliquer des théories statistiques au domaine du vivant → observation épidémiologique → analyse descriptive puis déductive (autre cours)

Stat descriptive

DESCRIPTION D'UNE SITUATION À L'AIDE DE PARAMÈTRES (MOYENNE, MÉDIANE, QUARTILES...)

Stat déductive

CONCLUSIONS À PARTIR D'OBSERVATIONS ET DE MESURES → HASARD OU PAS ?



DEFINITIONS PART 1

FLIGHT NO.

DESTINATION

TIME

STATUS

GATE

- **Donnée** : résultat de l'observation d'un individu, par l'utilisation d'un instrument de mesure ou par le sens de l'observateur. Il existe une variabilité des données entre individus qui peut être due au hasard ou physiologique.
- **Population** : série exhaustive de tous les individus étudiés, sur lesquels on veut inférer des décisions.
- **Echantillon** : ensemble fini et d'effectif limité, extrait de la population.

1- Variables

Qualitative binaire	sexe
Qualitative nominale	couleur des yeux
Qualitative ordinale (ou pseudo quantitative ++)	douleur articulaire (absente, modérée, intense)
Quantitative discrète	nombre de dents saines, nombre de médecins agréés en France...
Quantitative continue	poids de bébés à la naissance



QRU TIME :



A propos des définitions en statistiques descriptives, indiquer la proposition exacte :

- A) On peut facilement étudier une population
- B) L'échantillon est inconnu
- C) La population est connue
- D) Les résultats sont inférés de la population à l'échantillon
- E) Les propositions A, B, C et D sont fausses

QRU TIME :



A propos des définitions en statistiques descriptives, indiquer la proposition exacte :

- A) On peut facilement étudier une population
- B) L'échantillon est inconnu
- C) La population est connue
- D) Les résultats sont inférés de la population à l'échantillon
- E) Les propositions A, B, C et D sont fausses

II- Représentation de variables quantitatives



Poids (g)	Nb bébés
2200	2
2300	2
2400	4
2500	5
...	
3100	121
3200	150
3300	162
3400	170

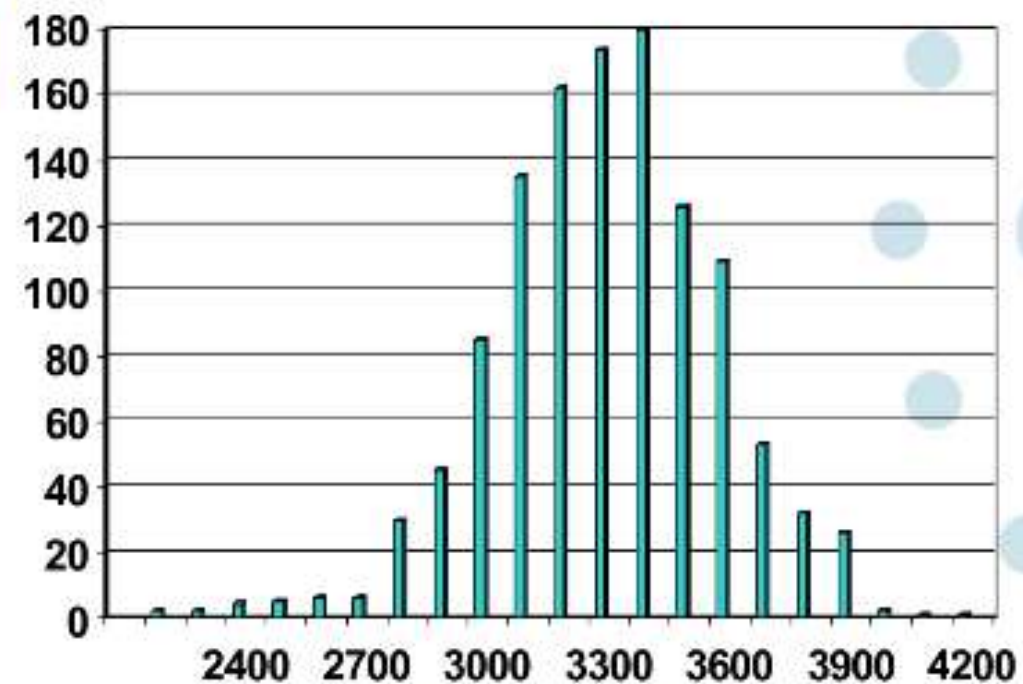


Tableau ou histogramme des effectifs
Mais pas seulement

MOYENNE :

$$m = \frac{\sum_{i=1}^n x_i}{n}$$

VARIANCE :

paramètre indiquant la dispersion des données autour de la moyenne

MÉDIANE :

valeur centrale si les données sont rangées en ordre croissant

QUARTILES :

ils partagent la série en 4 groupes de même effectif (série en ordre croissant) ex : 1e quartile = 25% de la série est < à cette valeur





Exemple : Soit une série de poids de bébés (n = 15 valeurs)

1 3400
2 2570
3 3210
4 4070
5 3840
6 4180
7 3480
8 3990
9 2640
10 3000
11 3830
12 1890
13 2350
14 2140
15 2470

Valeurs
rangées
par ordre
croissant

1 1890
2 2140
3 2350
4 2470
5 2570
6 2640
7 3000
8 **3210**
9 3400
10 3480
11 3830
12 3840
13 3990
14 4070
15 4180

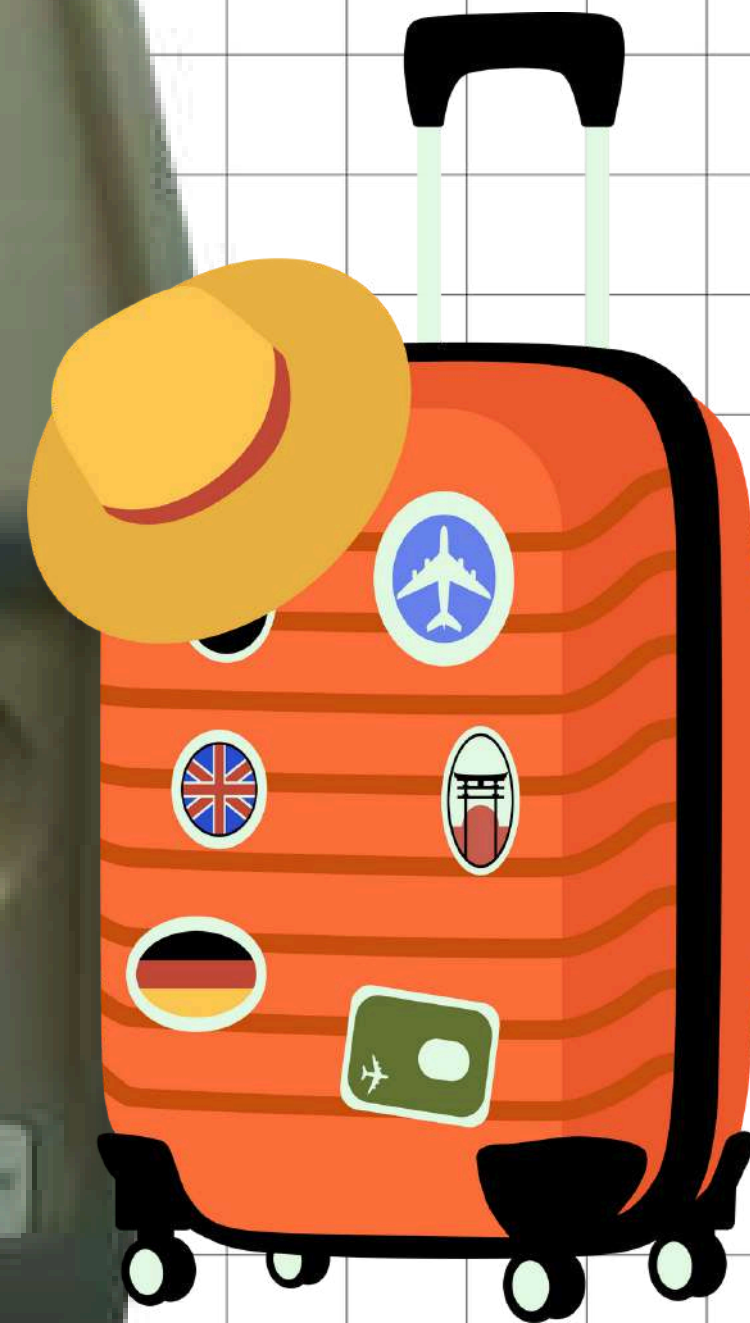
➡ **Médiane** : $n=15$ donc $(n+1)/2 = 8$
8^{ème} valeur = 3210

Si n pair, médiane située
entre les n° $n/2$ et $n/2+1$. **Moyenne**
des 2 valeurs correspondantes

➡ **3^{ème} quartile** $Q_{75} = 0,75 \times 15 = 11,25$

Le 3^{ème} quartile est situé entre le n°11 et
le n°12 soit $(3830+3840)/2 = 3835$

➡ **Moyenne de la série = 3137,3**



III- Estimation statistique

LE PROBLÈME EST LE SUIVANT : DÉTERMINER UN PARAMÈTRE AU NIVEAU D'UNE POPULATION À PARTIR D'OBSERVATIONS RÉALISÉES SUR UN ÉCHANTILLON DE CETTE POPULATION. APRÈS L'ÉTUDE, ON RÉFLÉCHIT À LA LÉGITIMITÉ DES RÉSULTATS ET À LEUR EXTRAPOLATION À LA POPULATION.

EXEMPLE, COMMENT CONNAÎTRE LA DURÉE DE SÉJOUR MOYENNE DES PATIENTS HOSPITALISÉS EN FRANCE, POUR UNE PATHOLOGIE DONNÉE ?

Estimation ponctuelle



= valeur jugée la meilleure à un instant t (peu fiable)
Elle est plus précise mais moins juste que l'estimation par intervalle +++

Estimation par intervalle



= intervalle de valeurs contenant la valeur recherchée (beaucoup plus fiable)



DEFINITIONS PART 2

FLIGHT NO.

DESTINATION

TIME

STATUS

GATE

- Intervalle de confiance (IC) : c'est l'estimation de la moyenne vraie μ à partir de la moyenne m calculée sur l'échantillon. L'IC est aussi appelé intervalle au risque α . On donne un intervalle auquel μ appartient :

$$\mu \in \left[m \pm \frac{\varepsilon s}{\sqrt{n}} \right] \rightarrow \text{l'intervalle de confiance IC au risque } \alpha$$

il s'agit de l'indice de précision i et cette valeur représente la largeur de l'IC ++

- Risque α : c'est le risque d'erreur dans l'estimation de μ (le risque que notre IC ne contienne pas μ). On prend en général un risque $\alpha = 5\%$ (on a 95% de chance que la moyenne vraie soit dans notre IC).



DEFINITIONS PART 2

FLIGHT NO.

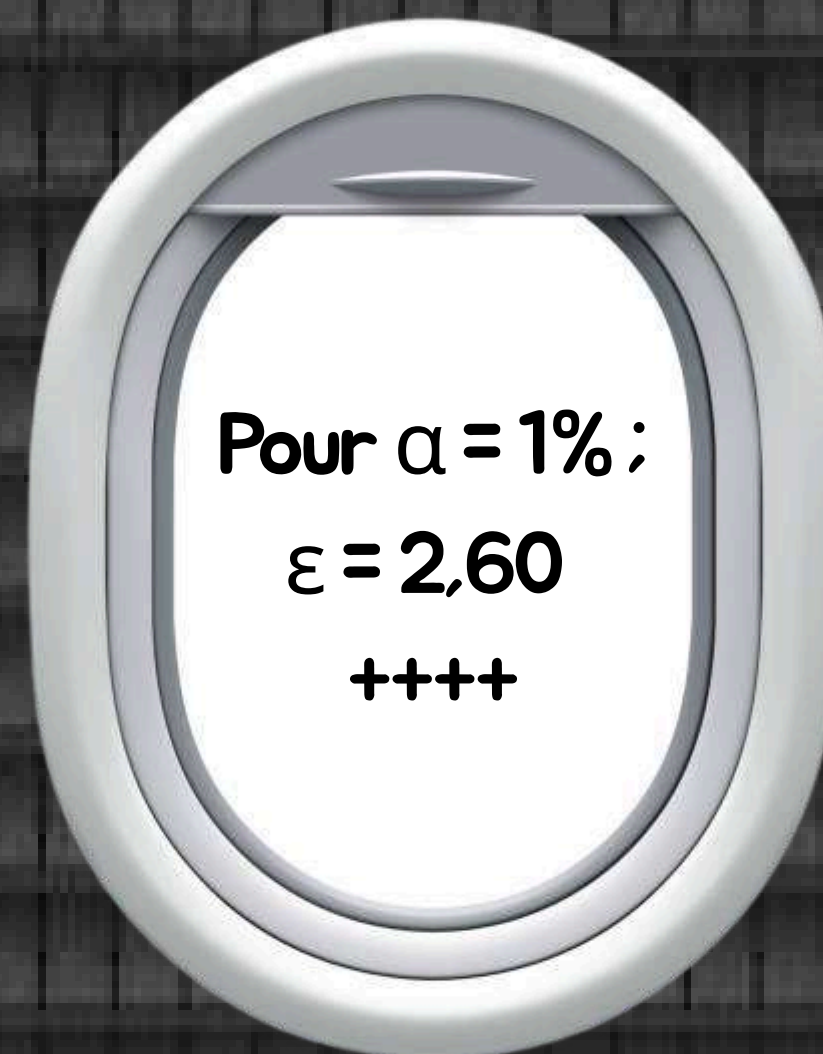
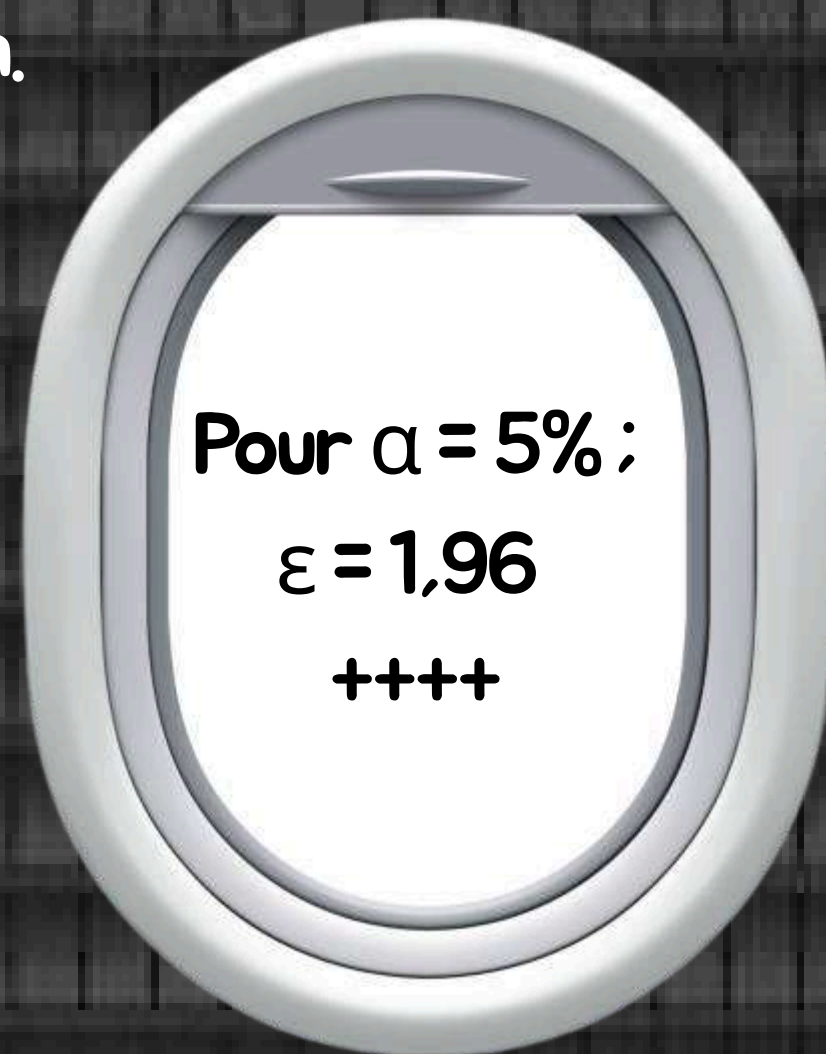
DESTINATION

TIME

STATUS

GATE

- Ecart réduit ε : c'est une valeur qui dépend du risque α : ils varient en sens inverse, si α augmente, ε diminue. Un écart-réduit mesure de combien d'écart-types une observation particulière est éloignée de la population.



QRU TIME :

On prend un échantillon de 36 patients atteints du cancer du côlon dont la moyenne d'âge est de 28 ans, l'écart-type est de 2. A propos de l'intervalle de confiance, indiquez la proposition exacte :

A) $IC_{99\%} = \mu \in \left[36 \pm \frac{1,96 \times 2}{\sqrt{28}} \right]$

C) $IC_{99\%} = \mu \in \left[28 \pm \frac{1,96 \times 2}{\sqrt{36}} \right]$

B) $IC_{95\%} = \mu \in \left[28 \pm \frac{1,96 \times 2}{\sqrt{36}} \right]$

D) $IC_{95\%} = \mu \in \left[28 \pm \frac{2,6 \times 2}{\sqrt{36}} \right]$

E) Les propositions A, B, C et D sont fausses

QRU TIME :

$$\mu \in \left[m \pm \frac{\varepsilon s}{\sqrt{n}} \right]$$

On prend un échantillon de 36 patients atteints du cancer du côlon dont la moyenne d'âge est de 28 ans, l'écart-type est de 2. A propos de l'intervalle de confiance, indiquez la proposition exacte :

A) $IC_{99\%} = \mu \in \left[36 \pm \frac{1,96 \times 2}{\sqrt{28}} \right]$

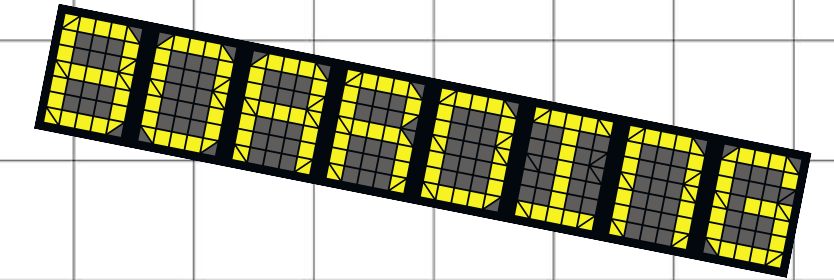
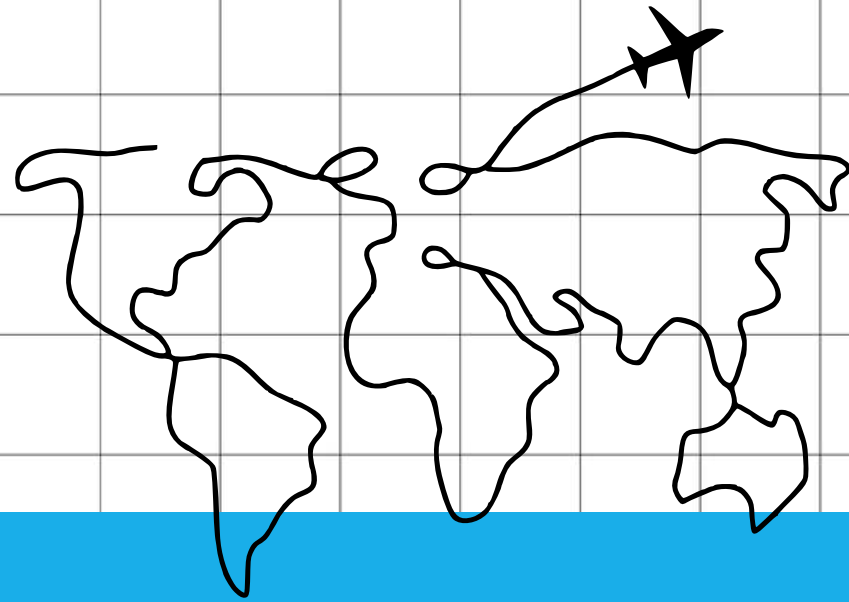
C) $IC_{99\%} = \mu \in \left[28 \pm \frac{1,96 \times 2}{\sqrt{36}} \right]$

B) $IC_{95\%} = \mu \in \left[28 \pm \frac{1,96 \times 2}{\sqrt{36}} \right]$

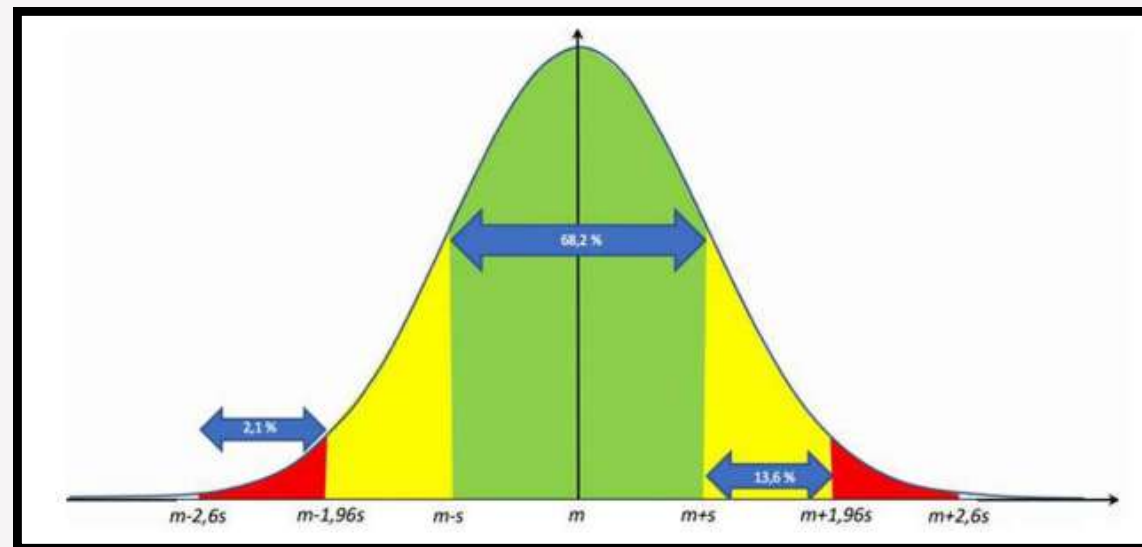
D) $IC_{95\%} = \mu \in \left[28 \pm \frac{2,6 \times 2}{\sqrt{36}} \right]$

E) Les propositions A, B, C et D sont fausses

N°- Loi de Gauss ou loi normale



- En abscisse $[m \pm \epsilon_s]$ donc l'IC
- En ordonnée : l'effectif pour chaque valeur
- L'aire sous la courbe, le % de la population concernée



A partir de la loi normale (= loi de Gauss), on précise les intervalles de confiance :

- ° $[m - 1 s ; m + 1s]$ contient 68,2% de la population ++
- ° $[m - 1,96 s ; m + 1,96s]$ contient 95,4% de la population ++
- ° $[m - 2,6 s ; m + 2,6s]$ contient 99,6% de la population ++



Conclusion



- ★ L'IC C'EST L'ESTIMATION DE LA MOYENNE VRAIE À PARTIR DE LA MOYENNE CALCULÉE SUR L'ÉCHANTILLON. IL EST AUSSI APPELÉ 'INTERVALLE AU RISQUE α '.
- ★ LE RISQUE α C'EST LE RISQUE D'ERREUR DANS L'ESTIMATION DE μ .
- ★ ε REPRÉSENTE L'ÉCART-RÉDUIT.
- ★ LES VARIATIONS DU RISQUE α DÉTERMINENT LA PRÉCISION DE L'ESTIMATION
- ★ I REPRÉSENTE LA LARGEUR DE L'IC
- ★ $IC = [M \pm I]$
- ★ SI $n \nearrow$, $I \searrow$ DONC L'IC \searrow DONC LA PRÉCISION $\nearrow ++$
- ★ SI $\alpha \nearrow$ ALORS $\varepsilon \searrow$ DONC $I \searrow$ DONC L'IC SE RESSERRE DONC LA PRÉCISION $\nearrow ++$



STAT' AIRWAYS vous souhaite un excellent voyage, en espérant vous revoir très vite à bord de notre compagnie !

