

STATISTIQUES

Regroupement du 10 mars 2007

Précisions concernant l'examen :

- ✓ Tous les documents sont autorisés (cours, notes, etc...)
- ✓ La maîtrise de la calculatrice est obligatoire
- ✓ Avoir « l'arbre de décision » le jour de l'examen (*page 295 et 296 du fascicule reçu en dernier et contenant les sujets d'examens et des corrigés*)
- ✓ Le « t » et le « chi2 » sont très présents à l'examen.

Les 3 échelles de mesure

Nominale

Ici on travaille sur des effectifs, pas de hiérarchie ni d'écart quantifiable.

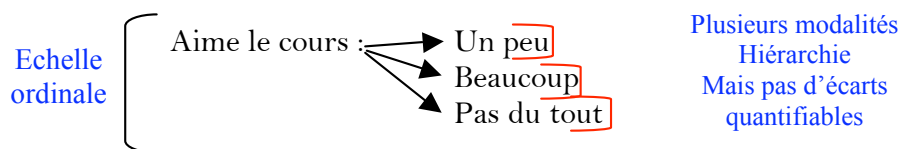
Ordinale

Hiérarchie entre les modalités mais pas d'intervalles égaux.

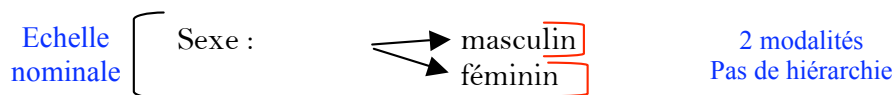
d'Intervalle

Les intervalles sont égaux (km, litre, sec. âge, degrés)

Exemples :



Une interro avec notation \longrightarrow échelle ordinale ! (pas d'écart de niveau égal entre deux élèves qui ont 3 et 4 et deux élèves qui ont 15 et 16) Mais dans certains énoncés il est explicitement précisé qu'il faut utiliser l'échelle d'intervalles. Dans ce cas pas de soucis.



Il faut utiliser les outils propres à chaque échelle.

Nominale

Chi 2
%
permet de faire des graphiques

Ordinale

U
Médiane
quantiles

d'Intervalle

appelle le « t » de student
intervalles de confiance
loi normale
moyenne
écart type
variance

Quelle que soit l'échelle, au départ on considère toujours l'hypothèse nulle. (exemple de formulation de H_0 dans les exercices du fascicule d'exercices).

Exemple : « t » de student

Gr A
Moyenne 12

Gr B
Moyenne 13

On ne peut pas considérer que $12 < 13$ sans test. Le test du « t » de student permet de dire si 12 est presque égal ou inférieur à 13.

Dans tous les exercices la bonne formulation de H_0 rapporte des points (cf le prof).

La 1^{ère} chose à faire dans les exercices est de formuler H_0 . (H zéro)

On estime le seuil de prise de risque à .05 (point zéro 5) en général. Cela signifie que l'on a moins de 5 chances sur 100 de se tromper.

Le « t » de Student

On calcule le « t » (calculé) pour le comparer au « t » théorique.

Cheminement à suivre :

- **Présentation de H_0**
- **Calcul du « t » calculé** (formule indiquée dans l'arbre pour chaque situation)
- **Calcul du ddl** (indiqué dans l'arbre pour chaque situation)
- **Recherche du « t » théorique** (cf 1^{ère} table numérique dans cours) : le v de la 1^{ère} colonne correspond au ddl. Le « t » théorique est la valeur entre la colonne .05 et la ligne du ddl.
- **Comparer les « t »**
- **Accepter ou rejeter H_0**

Pour éviter les risques d'erreurs il faut écrire les mots (écart type, variance, ...) et non les sigles car en fonction des livres les sigles ne sont pas toujours identiques !!

Si « t » calculé < « t » théorique alors H_0 est acceptée.

Si « t » calculé > « t » théorique sur .05 alors le résultat est significatif et H_0 est rejetée.

Si « t » calculé > « t » théorique sur .05 et .01 alors H_0 est rejetée mais on dit que le résultat est très significatif. (*Si ce cas se présente il faut absolument le dire car cela rapporte des points à l'examen*)

Echantillons appariés

Ce sont des paires de données pour la même personne, ce qui veut dire que le même individu est mesuré 2 fois. Par exemple dans le cas où il y a un test avant l'expérience et un re-test après l'expérience pour tester l'influence de la variable ; dans ce cas pour chaque personne on a une paire d'échantillon (avant + après).

F de Snedecor

C'est un des nœud de décision dans l'arbre de l'échelle d'intervalle.

Il teste l'homogénéité de la variance.

Pour plus de compréhension du F de Snedecor, se reporter à l'exercice corrigé page 297 et 298 du fascicule d'exercices.

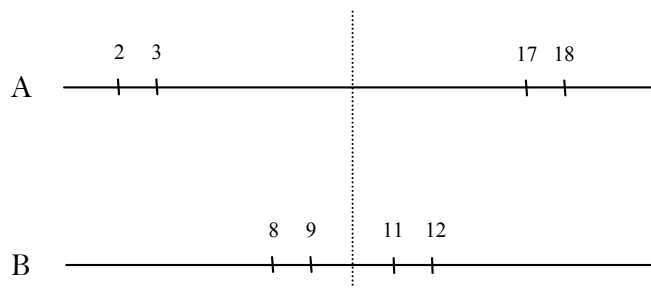
A savoir :

$$\text{Variance} = \frac{(\text{somme des écarts})^2}{\text{nombre d'individus} - 1}$$

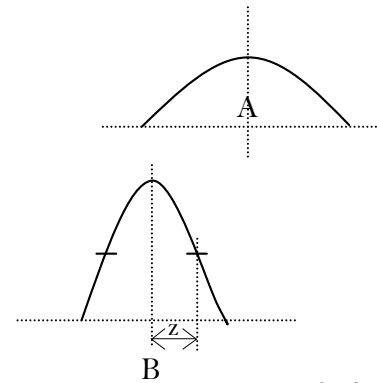
$$\text{Ecart type} = \sqrt{\text{variance}}$$

$$F \text{ calculé} = \frac{\text{Variance la plus forte}}{\text{Variance la plus faible}}$$

Exemple d'homogénéité de variance :



Dans ce cas les variances de A et B ne sont pas homogènes.



$z = \text{nombre d'écarts type}$

Dans le cas d'échantillons indépendants il faut :

- Donner H_0
- Calculer F
- Ensuite calculer le « t » correspondant à l'homogénéité de la variance. (formule B ou C sur l'arbre)

Dans la formule de calcul de F (cf arbre) le S^2 n'est pas « S au carré » mais « S deux ».

Cheminement à suivre dans le cas où l'on passe par le nœud du F de Snedecor :

- **Présentation de H_0 (t)** (càd le H_0 général pour l'exercice)
 - **Présentation du H_0 (F)** il s'agit de l'hypothèse nulle concernant l'homogénéité de la variance.
 - **Calcul du F calculé**
 - **Calcul du ddl 1 et du ddl 2**
 - **Recherche du F théorique dans la table** (pages 14 et 15 des tables numériques). Le F théorique se trouve à l'intersection de L1 et L2 sur la table. Le L1 est le ddl de l'échantillon ayant la plus grande variance, L2 étant le ddl de l'échantillon ayant la plus petite variance.
 - **Comparaison du F calculé et du F théorique**
 - **Acceptation ou rejet de H_0** . Cela détermine la formule à utiliser pour calculer « t ».
- **Calcul du « t » calculé** (formule indiquée dans l'arbre pour chaque situation)
- **Calcul du ddl** (indiqué dans l'arbre pour chaque situation)
- **Recherche du « t » théorique** (cf 1^{ère} table numérique dans cours) : le v de la 1^{ère} colonne correspond au ddl. Le « t » théorique est la valeur entre la colonne .05 et la ligne du ddl.
- **Comparer les « t »**
- **Accepter ou rejeter H_0**

Chi 2 (x2) - Echelle nominale - échantillons indépendants

Il y a des effectifs et jamais de % à l'intérieur d'un chi 2.

Dans un chi2 on se demande si une variable n'a pas une incidence sur l'autre.

Exemple de calcul :

H0 : à .05 les performances en statistiques ne sont pas significativement différentes en fonction de la boisson bue.



Les H0 des chi 2 sont les plus difficiles.

Il faut bien faire attention à citer : les variables et le « en fonction de »

1
Tu bois :
Bordeaux
Gaillac
Condrieux

2
Résultats en stats :
Bon
Mauvais

Les résultats donnent un « tableau de contingence » :

	B	G	C	
BON	-10.3 12 22.3	-2.7 12 14.7	+13 47 34	0 71 71
MAUVAIS	+10.3 26 15.7	+2.7 13 10.3	-13 11 24	0 50 50
	0 38 38	0 25 25	0 58 58	121 121 121

En vert : les valeurs théoriques

La valeur théorique c'est l'image de H0.

Pour calculer la valeur théorique d'une cellule :

(total ligne X total colonne) / nombre total de sujets

exemple 1^{ère} case : (71 X 38)/121 = 22.3

En noir : les valeurs observées

En rouge : l'écart entre valeur observée et valeur théorique

1 case = 1 croisement de variable.

Chaque individu n'est compté qu'une fois dans le tableau.

Dans le tableau on voit l'écart mais il faut démontrer à quel point l'écart est important, significatif. On calcule donc le chi 2. (Σ)

