

L'analyse de la variance

Joël Quinqueton

Université Paul Valéry

Master 1 IDS

But du test

- Une variable qualitative et une variable quantitative sont-elles indépendantes?
- Variable qualitative:
 - Ensemble fini de valeurs possibles
 - Partage la population en groupes
- Variable quantitative:
 - Résumée, sur une population, par sa moyenne et son écart type

Exemples

- Choix d'un itinéraire
 - Temps de trajet
 - Itinéraire choisi
- Comparaison d'exploitations agricoles
 - Production annuelle de lait de chaque vache
- Performances boursières
 - Performance boursière d'une action
 - Secteur d'activité

Hypothèse nulle

- La variable qualitative partage la population en groupes
- La variable quantitative suit une loi normale:
 - Dispersion due au « hasard » autour d'une valeur centrale
 - La valeur centrale est estimée par la moyenne et a dispersion par la variance
- La moyenne et la variance sont les mêmes pour chaque groupe

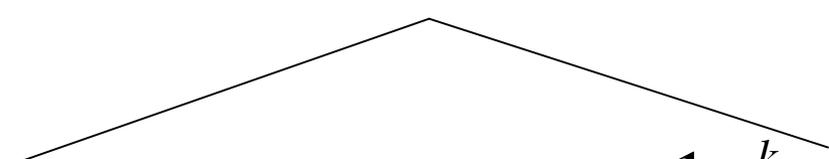
Tester l'hypothèse nulle

- Tester l'égalité des moyennes:
 - test de Fisher Snedecor
- Tester l'égalité des variances:
 - Il existe un test basé sur le Khi-2
 - Non étudié dans ce cours
 - On suppose donc que les variances sont égales

Décomposition de la variance

- Théorème: la variance est la somme de la variance des moyennes et de la moyenne des variances

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_i^j - \bar{X})^2$$


$$\sigma_{\text{int}}^2 = \frac{1}{n} \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 \quad \sigma_R^2 = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_i^j - \bar{X}_i)^2$$

Test de Fisher Snedecor

- Statistique sur le rapport entre les composantes de la variance

The diagram illustrates the Fisher-Snedecor F-test formula, $F = \frac{\sigma_{\text{int}}^2 / (k-1)}{\sigma_R^2 / (n-k)}$, with callouts explaining its components:

- Variance des moyennes**: Points to the numerator's variance component, σ_{int}^2 .
- Nombre de groupes**: Points to the denominator of the numerator, $(k-1)$.
- d.d.l. « colonne »**: Points to the denominator of the numerator, $(k-1)$.
- Moyenne des variances**: Points to the denominator's variance component, σ_R^2 .
- Taille de l'échantillon**: Points to the denominator of the denominator, $(n-k)$.
- d.d.l. « ligne »**: Points to the denominator of the denominator, $(n-k)$.

$$F = \frac{\sigma_{\text{int}}^2 / (k-1)}{\sigma_R^2 / (n-k)}$$

Comparaison à la table

- Une table pour chaque taux de risque
 - Donnée en cours: 5% et 1%
- Lire la valeur $F(n-k, k-1)$ de la ligne $(n-k)$ et la colonne $(k-1)$ de la table choisie
- Si $F > F(n-k, k-1)$ alors l'hypothèse nulle est rejetée, sinon elle est retenue.

Interprétation du test

- Hypothèse nulle: les moyennes des groupes sont égales
- Alternative: au moins une moyenne est différente des autres de manière significative

Exemple du choix d'un itinéraire

- Mesure du temps de trajet (en minutes)

jour	ancien itinéraire	itinéraire bis	itinéraire ter
lundi	11	11	11
mardi	17	11	7
mercredi	15	20	12
jeudi	11	16	8
vendredi	17	11	10
lundi	18	19	8
mardi	12	17	8
mercredi	15	16	15
jeudi	14	15	15
vendredi	13	13	11
lundi	17	13	6
mardi	12	15	9
mercredi	14	16	11
jeudi	14	12	15
vendredi	11	17	6

Exemple: Mise en œuvre du test

- Calcul des moyennes et variances:
 - Échantillon complet: moyenne 13, variance 11,8
 - Groupe 1: moyenne 14,1, variance 5,8
 - Groupe 2: moyenne 14,8, variance 8,3
 - Groupe 3: moyenne 10,1, variance 9,7
- Taille: $n = 45$, $k = 3$

Exemple: Résultat du test

- Calcul de F:
 - Moyenne des variances = 7,9
 - Variance des moyennes = 3,9
 - Degrés de liberté: $n-k = 42$, $k-1 = 2$
 - $F = (3,9/2)/(7,9/42) = 10,4$
- Comparaison à la table
 - Table à 1%: $F(42,2) = F(40,2) = 5,18$
- L'hypothèse nulle est rejetée, donc au moins une des moyennes est différente des autres: on peut donc choisir un itinéraire