Tests multiples

EIDD, Filière GB, Biostatistiques
4 octobre 2024

Iqraa Meah

iqraa.meah@inserm.fr

Déroulement du cours aujourd'hui

Session du matin : 3 heures plutôt théoriques

- Rappels sur les tests simples
- Définition du contexte des tests multiples
- Procédures classiques en tests multiples

Session de l'après-midi : 2 heures pratiques

Réviser et comprendre les notions du cours avec un TP en R

Objectifs de la journée : de façon théorique et pratique

- se (re) familiariser avec les tests d'hypothèses
- connaître le contexte des tests multiples
- connaître les principes de base en tests multiples

Déroulement du cours aujourd'hui

Session du matin : 3 heures plutôt théoriques

- Rappels sur les tests simples
- Définition du contexte des tests multiples
- Procédures classiques en tests multiples

Session de l'après-midi : 2 heures pratiques

Réviser et comprendre les notions du cours avec un TP en R

Objectifs de la journée : de façon théorique et pratique

- se (re) familiariser avec les tests d'hypothèses
- connaître le contexte des tests multiples
- connaître les principes de base en tests multiples

Déroulement du cours aujourd'hui

Session du matin : 3 heures plutôt théoriques

- Rappels sur les tests simples
- Définition du contexte des tests multiples
- Procédures classiques en tests multiples

Session de l'après-midi : 2 heures pratiques

Réviser et comprendre les notions du cours avec un TP en R

Objectifs de la journée : de façon théorique et pratique

- se (re) familiariser avec les tests d'hypothèses
- connaître le contexte des tests multiples
- connaître les principes de base en tests multiples

Rappels sur les tests d'hypothèse (simples) en statistiques

Exemple de problème

→ Comprendre l'effet d'un gène A sur un caractère phénotypique T

Etude in vivo sur des souris

- Deux populations, l'une contrôle et l'autre traitée
- Ablation du gène A pour la population traitée au stade embryonnaire
- Étude de la différence du caractère T entre les deux populations

→ Vrai consortium https://www.mousephenotype.org/

Exemple de problème

Dans ce contexte

- Hypothèse nulle H_0 : le gène A n'a pas d'effet sur T
- Hypothèse alternative H_1 : le gène A a un effet sur T

Objectif idéal : rejeter à juste titre l'hypothèse nulle
$\rightarrow$ Rejeter H_0 c'est faire une découverte

Exemple de problème

Dans ce contexte

- Hypothèse nulle H_0 : le gène A n'a pas d'effet sur T
- Hypothèse alternative H_1 : le gène A a un effet sur T

Objectif idéal : rejeter à juste titre l'hypothèse nulle
$\rightarrow$ Rejeter H_0 c'est faire une découverte

Type d'erreurs possible

Pour éviter les confusions, raisonner seulement en termes de

"j'ai rejeté ou pas H_0" + "fallait-il vraiment rejeter ou pas H_0"

Erreur de type I

Rejeter à tort H_0.
Conclure qu'il y a un lien entre A et T alors qu'en vrai non
→ On fait une fausse découverte (=faux positif)

Erreur de type II

Ne pas rejeter à juste titre H_0.
Conclure qu'il n'y a pas de lien entre A et T alors qu'en vrai oui
→ On loupe une vraie découverte

Priorité sur l'erreur de type I : penser à la présomption d'innocence

Type d'erreurs possible

Pour éviter les confusions, raisonner seulement en termes de
"j'ai rejeté ou pas H_0" + "fallait-il vraiment rejeter ou pas H_0"

Erreur de type I

Rejeter à tort H_0.
Conclure qu'il y a un lien entre A et T alors qu'en vrai non
→ On fait une fausse découverte (=faux positif)

Erreur de type II

Ne pas rejeter à juste titre H_0.
Conclure qu'il n'y a pas de lien entre A et T alors qu'en vrai oui
→ On loupe une vraie découverte

Priorité sur l'erreur de type I : penser à la présomption d'innocence

Type d'erreurs possible

Pour éviter les confusions, raisonner seulement en termes de
"j'ai rejeté ou pas H_0" + "fallait-il vraiment rejeter ou pas H_0"

Erreur de type I

Rejeter à tort H_0.
Conclure qu'il y a un lien entre A et T alors qu'en vrai non
$\rightarrow$ On fait une fausse découverte (=faux positif)

Erreur de type II

Ne pas rejeter à juste titre H_0.
Conclure qu'il n'y a pas de lien entre A et T alors qu'en vrai oui
$\rightarrow$ On loupe une vraie découverte

Priorité sur l'erreur de type I : penser à la présomption d'innocence

Type d'erreurs possible

Pour éviter les confusions, raisonner seulement en termes de
"j'ai rejeté ou pas H_0" + "fallait-il vraiment rejeter ou pas H_0"

Erreur de type I

Rejeter à tort H_0.
Conclure qu'il y a un lien entre A et T alors qu'en vrai non
→ On fait une fausse découverte (=faux positif)

Erreur de type II

Ne pas rejeter à juste titre H_0.
Conclure qu'il n'y a pas de lien entre A et T alors qu'en vrai oui
→ On loupe une vraie découverte

Priorité sur l'erreur de type I : penser à la présomption d'innocence

Modélisation du problème

- $X_1, \ldots, X_n$ mesures du caractère T sur n souris traités
- Supposons que $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$ i.i.d
- $H_0 : \mu = 0$ vs $H_1 : \mu > 0$
- Rejeter H_0 c'est faire une découverte
  $\rightarrow$ *Les données contiennent du signal, elles diffèrent des données de référence* $\sim \mathcal{N}(0, 1)$
- $\alpha \in (0, 1)$: niveau de tolérance d'erreur du type I fixé à l'avance
  La probabilité de rejeter à tort H_0 *est borné par* α

Modélisation du problème

- $X_1, \ldots, X_n$ mesures du caractère T sur n souris traités
- Supposons que $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$ i.i.d
- $H_0 : \mu = 0 \quad$ vs $\quad H_1 : \mu > 0$
- Rejeter H_0 c'est faire une découverte
  $\rightarrow$ *Les données contiennent du signal, elles diffèrent des données de référence* $\sim \mathcal{N}(0,1)$
- $\alpha \in (0,1)$: niveau de tolérance d'erreur du type I fixé à l'avance
  La probabilité de rejeter à tort H_0 *est borné par* α

Modélisation du problème

- $X_1, \ldots, X_n$ mesures du caractère T sur n souris traités
- Supposons que $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$ i.i.d
- $H_0 : \mu = 0 \quad$ vs $\quad H_1 : \mu > 0$
- Rejeter H_0 c'est faire une découverte
  $\rightarrow$ *Les données contiennent du signal, elles diffèrent des données de référence* $\sim \mathcal{N}(0, 1)$
- $\alpha \in (0, 1)$: niveau de tolérance d'erreur du type I fixé à l'avance
  La probabilité de rejeter à tort H_0 est borné par α

Modélisation du problème

- $X_1, \ldots, X_n$ mesures du caractère T sur n souris traités
- Supposons que $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$ i.i.d
- $H_0 : \mu = 0$ vs $H_1 : \mu > 0$
- Rejeter H_0 c'est faire une découverte
  $\rightarrow$ *Les données contiennent du signal, elles diffèrent des données de référence* $\sim \mathcal{N}(0, 1)$
- $\alpha \in (0, 1)$: niveau de tolérance d'erreur du type I fixé à l'avance
  La probabilité de rejeter à tort H_0 est borné par α

Construction du test

Test de statistique : $T(X) = \sqrt{n}\frac{\bar{X}-\mu_0}{\sigma}$

Avec $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ la moyenne empirique.

Ici, on a posé $\mu_0 = 0$, $\sigma^2 = 1 \rightarrow$, on a donc $T(X) = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}$

Décision

- Intuition : $T(X)$ à une valeur qui serait peu probable sous $H_0 \rightarrow$ on rejette H_0
- Rigoureusement : comparer la valeur de $T(X)$ au quantile d'ordre $1-\alpha$ de $\mathcal{N}(0,1)$.
  Si $T(X) \geq q_{(1-\alpha)}$ on rejette H_0.

Construction du test

Test de statistique : $T(X) = \sqrt{n}\frac{\bar{X}-\mu_0}{\sigma}$

Avec $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ la moyenne empirique.

Ici, on a posé $\mu_0 = 0$, $\sigma^2 = 1 \rightarrow$, on a donc $T(X) = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}$

Décision

- Intuition : $T(X)$ à une valeur qui serait peu probable sous $H_0 \rightarrow$ on rejette H_0
- Rigoureusement : comparer la valeur de $T(X)$ au quantile d'ordre $1-\alpha$ de $\mathcal{N}(0,1)$.
  Si $T(X) \geq q_{(1-\alpha)}$ on rejette H_0.

Construction du test

Test de statistique : $T(X) = \sqrt{n}\frac{\bar{X}-\mu_0}{\sigma}$

Avec $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ la moyenne empirique.

Ici, on a posé $\mu_0 = 0$, $\sigma^2 = 1 \rightarrow$, on a donc $T(X) = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}$

Décision

- Intuition : $T(X)$ à une valeur qui serait peu probable sous $H_0 \rightarrow$ on rejette H_0
- Rigoureusement : comparer la valeur de $T(X)$ au quantile d'ordre $1-\alpha$ de $\mathcal{N}(0,1)$.
  Si $T(X) \geq q_{(1-\alpha)}$ on rejette H_0.

Interprétation graphique du test simple

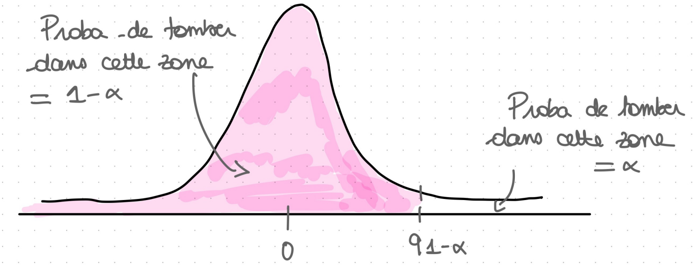

Figure: Zone de rejet (partie à droite du quantile) pour le test sur la moyenne d'une loi normale où la variance est connue

Que signifie "tester au niveau α" ?

Borne sur la probabilité de faire une erreur de type I
$\hookrightarrow$ Généralement "petit", de manière classique on choisit $\alpha = 0.05$

Interprétation graphique du test simple

$\hookrightarrow$ α borne la probabilité de faire une erreur de type I

Pourquoi comparer au quantile $q_{(1-\alpha)}$?

Probabilité que $T(X)$ ait une valeur au-dessus du quantile est $\leq \alpha$ sous H_0

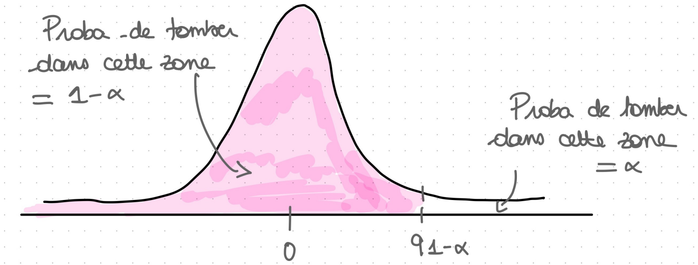

Figure: Zone de rejet (partie à droite du quantile) pour le test sur la moyenne d'une loi normale où la variance est connue

P-valeur

C'est la probabilité que $T(X)$ ait une valeur égale ou plus extrême à la valeur observée sous l'hypothèse nulle.

Pourquoi rejeter lorsque que la p-valeur est $\leq \alpha$?

Rappel : α borne le fait d'observer des choses extêmes sous H_0.
Si p-valeur $\leq \alpha$ on peut dire qu'il y a peu de chance d'observer cette valeur pour $T(X)$ si H_0 était vraie $\rightarrow$ on rejette H_0

P-valeur

C'est la probabilité que $T(X)$ ait une valeur égale ou plus extrême à la valeur observée sous l'hypothèse nulle.

Pourquoi rejeter lorsque que la p-valeur est $\leq \alpha$?

Rappel : α borne le fait d'observer des choses extêmes sous H_0.
Si p-valeur $\leq \alpha$ on peut dire qu'il y a peu de chance d'observer cette valeur pour $T(X)$ si H_0 était vraie $\rightarrow$ on rejette H_0

Puissance

Ne jamais rejeter H_0 = pas d'erreur de type I MAIS aucune découverte intéressante possible $\rightarrow$ erreur de type II

La puissance est la probabilité de rejeter à juste titre l'hypothèse nulle
$\hookrightarrow$ capacité à faire de vraies découvertes = capacité à ne pas faire d'erreur de type II

Puissance

Ne jamais rejeter H_0 = pas d'erreur de type I MAIS aucune découverte intéressante possible $\rightarrow$ erreur de type II

La puissance est la probabilité de rejeter à juste titre l'hypothèse nulle
$\hookrightarrow$ capacité à faire de vraies découvertes = capacité à ne pas faire d'erreur de type II

Le problème de la multiplicité

Exemples d'applications

Biologie moléculaire : IMPC

- Comprendre comment le génotype est associé au phénotype
- Tester l'effet de nombreux gènes sur un trait phénotypique

→ voir https://www.mousephenotype.org/

Neuroscience

- Faire une carte du cerveau
- Tester l'activation de régions du cerveau (avec des outils d'imagerie cérébrale) lors d'exécution d'une tâche particulière

→ voir https://doi.org/10.3389/fninf.2015.00008

Le problème de la multiplicité

$\boxed{\text{Tests multiples = contexte où } m \geq 1 \text{ tests considérés simultanément}}$

$\hookrightarrow$ $\boxed{\text{En moyenne, on va faire } m\alpha \text{ erreurs}}$

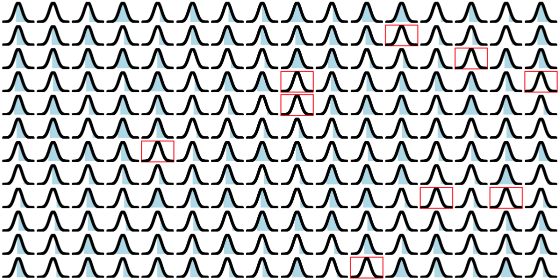

Figure: Illustration du problème de la multiplicité. Ici toutes les hypothèses sont de vraies nulles, mais on fait quand même un certain nombre de fausses découvertes (encadrées en rouge).

Erreurs en tests multiples

Erreur de type I : on aimerait

- Tests individuellement valides

  $+$
- Évaluation collective de toutes les décisions prises

- Bornes sur les proba de rejeter à tort chaque hypothèse nulle individuellement
- Borne le # total d'erreurs de type I

Erreur de type II : on aimerait

Faire le plus de vraies découvertes possibles

Erreurs en tests multiples

Erreur de type I : on aimerait

- Tests individuellement valides

  $+$
- Évaluation collective de toutes les décisions prises

- Bornes sur les proba de rejeter à tort chaque hypothèse nulle individuellement
- Borne le # total d'erreurs de type I

Erreur de type II : on aimerait

Faire le plus de vraies découvertes possibles

Erreurs en tests multiples

Erreur de type I : on aimerait

- Tests individuellement valides

  $+$

- Évaluation collective de toutes les décisions prises

- Bornes sur les proba de rejeter à tort chaque hypothèse nulle individuellement
- Borne le # total d'erreurs de type I

Erreur de type II : on aimerait

Faire le plus de vraies découvertes possibles

L'enjeu en tests multiples

Résumé facile

Tests multiples = Problème d'allocation avec des contraintes

Pour ce faire : pour tous les tests effectués, fixer un budget d'erreur $\alpha \in (0, 1)$ qu'il faut

- allouer à chaque hypothèse pour faire des tests individuellement valides
- tout en cherchant à optimiser le $\#$ de vraies découvertes

L'enjeu en tests multiples

Résumé facile

Tests multiples = Problème d'allocation avec des contraintes

Pour ce faire : pour tous les tests effectués, fixer un budget d'erreur $\alpha \in (0,1)$ qu'il faut

- allouer à chaque hypothèse pour faire des tests individuellement valides
- tout en cherchant à optimiser le # de vraies découvertes

L'enjeu en tests multiples

Résumé facile

Tests multiples = Problème d'allocation avec des contraintes

Pour ce faire : pour tous les tests effectués, fixer un budget d'erreur $\alpha \in (0,1)$ qu'il faut

- allouer à chaque hypothèse pour faire des tests individuellement valides
- tout en cherchant à optimiser le $\#$ de vraies découvertes

Procédures classiques de tests multiples

Procédures de tests multiples

Sortie d'une procédure : ensemble $\mathcal{R}$ qui indique les hypothèses nulles à rejeter

Objectifs d'une procédure de tests multiples

Contrôler le # total d'erreur

$+$

Permettre de faire autant de vraies découvertes que possible

$\rightarrow$ Permettre de faire de vraies découvertes = puissance de procédure

$\rightarrow$ Contrôle du # total d'erreur = allocation du budget d'erreur $\alpha \in (0, 1)$

$\hookrightarrow$ Stratégie d'allocation ? Dépend de comment on compte le # total d'erreurs

Procédures de tests multiples

Sortie d'une procédure : ensemble $\mathcal{R}$ qui indique les hypothèses nulles à rejeter

Objectifs d'une procédure de tests multiples

Contrôler le # total d'erreur

$+$

Permettre de faire autant de vraies découvertes que possible

$\rightarrow$ Permettre de faire de vraies découvertes = puissance de procédure

$\rightarrow$ Contrôle du # total d'erreur = allocation du budget d'erreur $\alpha \in (0, 1)$

$\hookrightarrow$ Stratégie d'allocation ? Dépend de comment on compte le # total d'erreurs

Procédures de tests multiples

Sortie d'une procédure : ensemble $\mathcal{R}$ qui indique les hypothèses nulles à rejeter

Objectifs d'une procédure de tests multiples

Contrôler le # total d'erreur

$+$

Permettre de faire autant de vraies découvertes que possible

$\rightarrow$ Permettre de faire de vraies découvertes = puissance de procédure

$\rightarrow$ Contrôle du # total d'erreur = allocation du budget d'erreur $\alpha \in (0, 1)$

$\hookrightarrow$ Stratégie d'allocation ? Dépend de comment on compte le # total d'erreurs

Procédures de tests multiples

Sortie d'une procédure : ensemble $\mathcal{R}$ qui indique les hypothèses nulles à rejeter

Objectifs d'une procédure de tests multiples

Contrôler le # total d'erreur

$+$

Permettre de faire autant de vraies découvertes que possible

$\rightarrow$ Permettre de faire de vraies découvertes = puissance de procédure

$\rightarrow$ Contrôle du # total d'erreur = allocation du budget d'erreur $\alpha \in (0, 1)$

$\hookrightarrow$ Stratégie d'allocation ? Dépend de comment on compte le # total d'erreurs

Notations usuelles

- On teste $H_{0,1}, \ldots, H_{0,m}$, m hypothèses nulles
- Chaque hypothèse nulle $H_{0,i}$ est représentée par une p-valeur $p_i \in (0,1)$
- $\mathcal{H}_0$ contient les indices des vraies hypothèses nulles
- $\overline{\mathcal{H}_0} = \mathcal{H}_1$ contient les indices des fausses hypothèses nulles
- $\mathcal{R}$ contient les indices des hypothèses nulles rejetées
- $|\mathcal{H}_0 \cap \mathcal{R}|$ donne le # de fausses découvertes
- $|\mathcal{H}_1 \cap \mathcal{R}|$ donne le # de vraies découvertes

Notations usuelles

- On teste $H_{0,1}, \ldots, H_{0,m}$, m hypothèses nulles
- Chaque hypothèse nulle $H_{0,i}$ est représentée par une p-valeur $p_i \in (0,1)$
- $\mathcal{H}_0$ contient les indices des vraies hypothèses nulles
- $\overline{\mathcal{H}_0} = \mathcal{H}_1$ contient les indices des fausses hypothèses nulles
- $\mathcal{R}$ contient les indices des hypothèses nulles rejetées
- $|\mathcal{H}_0 \cap \mathcal{R}|$ donne le # de fausses découvertes
- $|\mathcal{H}_1 \cap \mathcal{R}|$ donne le # de vraies découvertes

Notations usuelles

- On teste $H_{0,1}, \ldots, H_{0,m}$, m hypothèses nulles
- Chaque hypothèse nulle $H_{0,i}$ est représentée par une p-valeur $p_i \in (0,1)$
- $\mathcal{H}_0$ contient les indices des vraies hypothèses nulles
- $\overline{\mathcal{H}_0} = \mathcal{H}_1$ contient les indices des fausses hypothèses nulles
- $\mathcal{R}$ contient les indices des hypothèses nulles rejetées
- $|\mathcal{H}_0 \cap \mathcal{R}|$ donne le # de fausses découvertes
- $|\mathcal{H}_1 \cap \mathcal{R}|$ donne le # de vraies découvertes

Notations usuelles

- On teste $H_{0,1}, \ldots, H_{0,m}$, m hypothèses nulles
- Chaque hypothèse nulle $H_{0,i}$ est représentée par une p-valeur $p_i \in (0,1)$
- $\mathcal{H}_0$ contient les indices des vraies hypothèses nulles
- $\overline{\mathcal{H}_0} = \mathcal{H}_1$ contient les indices des fausses hypothèses nulles
- $\mathcal{R}$ contient les indices des hypothèses nulles rejetées
- $|\mathcal{H}_0 \cap \mathcal{R}|$ donne le # de fausses découvertes
- $|\mathcal{H}_1 \cap \mathcal{R}|$ donne le # de vraies découvertes

Notations usuelles

- On teste $H_{0,1}, \ldots, H_{0,m}$, m hypothèses nulles
- Chaque hypothèse nulle $H_{0,i}$ est représentée par une p-valeur $p_i \in (0,1)$
- $\mathcal{H}_0$ contient les indices des vraies hypothèses nulles
- $\overline{\mathcal{H}_0} = \mathcal{H}_1$ contient les indices des fausses hypothèses nulles
- $\mathcal{R}$ contient les indices des hypothèses nulles rejetées
- $|\mathcal{H}_0 \cap \mathcal{R}|$ donne le # de fausses découvertes
- $|\mathcal{H}_1 \cap \mathcal{R}|$ donne le # de vraies découvertes

Notations usuelles

- On teste $H_{0,1}, \ldots, H_{0,m}$, m hypothèses nulles
- Chaque hypothèse nulle $H_{0,i}$ est représentée par une p-valeur $p_i \in (0,1)$
- $\mathcal{H}_0$ contient les indices des vraies hypothèses nulles
- $\overline{\mathcal{H}_0} = \mathcal{H}_1$ contient les indices des fausses hypothèses nulles
- $\mathcal{R}$ contient les indices des hypothèses nulles rejetées
- $|\mathcal{H}_0 \cap \mathcal{R}|$ donne le $\#$ de fausses découvertes
- $|\mathcal{H}_1 \cap \mathcal{R}|$ donne le $\#$ de vraies découvertes

Notations usuelles

- On teste $H_{0,1}, \ldots, H_{0,m}$, m hypothèses nulles
- Chaque hypothèse nulle $H_{0,i}$ est représentée par une p-valeur $p_i \in (0,1)$
- $\mathcal{H}_0$ contient les indices des vraies hypothèses nulles
- $\overline{\mathcal{H}_0} = \mathcal{H}_1$ contient les indices des fausses hypothèses nulles
- $\mathcal{R}$ contient les indices des hypothèses nulles rejetées
- $|\mathcal{H}_0 \cap \mathcal{R}|$ donne le $\#$ de fausses découvertes
- $|\mathcal{H}_1 \cap \mathcal{R}|$ donne le $\#$ de vraies découvertes

FWER et procédure de Bonferroni

Family Wise Error Rate (FWER)

Probabilité de faire au moins une fausse découverte parmi les hypothèses rejetées

$$\mathrm{P}(|\mathcal{H}_0 \cap \mathcal{R}| \geq 1)$$

Procédure de Bonferroni

L'ensemble des hypothèses nulles rejetées est donné par

$$\mathcal{R} = \{1 \leq i \leq m : p_i \leq \alpha/m\}$$

- Contrôle du FWER = contrôle du # d'erreurs strict : 0 erreur avec grand proba : $\mathrm{P}(|\mathcal{H}_0 \cap \mathcal{R}| \geq 1) \leq \alpha$
- Puissance de la procédure faible : $\alpha/m \to 0$ lorsque $m \to \infty$

FWER et procédure de Bonferroni

Family Wise Error Rate (FWER)

Probabilité de faire au moins une fausse découverte parmi les hypothèses rejetées

$$\mathrm{P}(|\mathcal{H}_0 \cap \mathcal{R}| \geq 1)$$

Procédure de Bonferroni

L'ensemble des hypothèses nulles rejetées est donné par

$$\mathcal{R} = \{1 \leq i \leq m : p_i \leq \alpha/m\}$$

- Contrôle du FWER = contrôle du # d'erreurs strict : 0 erreur avec grand proba : $\mathrm{P}(|\mathcal{H}_0 \cap \mathcal{R}| \geq 1) \leq \alpha$
- Puissance de la procédure faible : $\alpha/m \to 0$ lorsque $m \to \infty$

FWER et procédure de Bonferroni

Family Wise Error Rate (FWER)

Probabilité de faire au moins une fausse découverte parmi les hypothèses rejetées

$$\mathrm{P}(|\mathcal{H}_0 \cap \mathcal{R}| \geq 1)$$

Procédure de Bonferroni

L'ensemble des hypothèses nulles rejetées est donné par

$$\mathcal{R} = \{1 \leq i \leq m : p_i \leq \alpha/m\}$$

- Contrôle du FWER = contrôle du # d'erreurs strict : 0 erreur avec grand proba : $\mathrm{P}(|\mathcal{H}_0 \cap \mathcal{R}| \geq 1) \leq \alpha$
- Puissance de la procédure faible : $\alpha/m \to 0$ lorsque $m \to \infty$

FDR et procédure de Benjamini et Hochberg

False discovery Rate (FDR)

L'espérance du taux de fausses découvertes

$$\mathrm{E}[\frac{|\mathcal{H}_0 \cap \mathcal{R}|}{|\mathcal{R}|}]$$

Procédure de Benjamini et Hochberg

L'ensemble des hypothèses nulles rejetées est donné par

$$\mathcal{R} = \{1 \leq i \leq m : p_{(i)} \leq p_{(\hat{k})}\}$$

avec $\hat{k} = \max\{1 \leq i \leq m : p_{(i)} \leq \alpha i / k\}$ un point de coupure

FDR et procédure de Benjamini et Hochberg

False discovery Rate (FDR)

L'espérance du taux de fausses découvertes

$$\mathsf{E}[\frac{|\mathcal{H}_0 \cap \mathcal{R}|}{|\mathcal{R}|}]$$

Procédure de Benjamini et Hochberg

L'ensemble des hypothèses nulles rejetées est donné par

$$\mathcal{R} = \{1 \leq i \leq m : p_{(i)} \leq p_{(\hat{k})}\}$$

avec $\hat{k} = \max\{1 \leq i \leq m : p_{(i)} \leq \alpha i / k\}$ un point de coupure

FDR et procédure de Benjamini et Hochberg

False discovery Rate (FDR)

L'espérance du taux de fausses découvertes

$$\mathrm{E}[\frac{|\mathcal{H}_0 \cap \mathcal{R}|}{|\mathcal{R}|}]$$

Procédure de Benjamini et Hochberg

L'ensemble des hypothèses nulles rejetées est donné par

$$\mathcal{R} = \{1 \leq i \leq m : p_{(i)} \leq p_{(\hat{k})}\}$$

avec $\hat{k} = \max\{1 \leq i \leq m : p_{(i)} \leq \alpha i/k\}$ un point de coupure

FDR et procédure de Benjamini et Hochberg

Interprétation de la procédure

- On ordonne les p-valeurs en ordre croissant
- On rejette les $\hat{k}$ plus petites p-valeurs
  $\hookrightarrow$ $\hat{k}$ indique le $\#$ de p-valeurs qu'on peut rejeter tout en gardant $\mathrm{E}[\frac{|\mathcal{H}_0 \cap \mathcal{R}|}{|\mathcal{R}|}] \leq \alpha$

- Contrôle du FDR = contrôle du $\#$ d'erreurs plus libéral
  $\hookrightarrow$ plus on rejette plus on autorise à faire des erreurs
- Puissance de la procédure beaucoup plus forte que celle de Bonferroni

FDR et procédure de Benjamini et Hochberg

Interprétation de la procédure

- On ordonne les p-valeurs en ordre croissant
- On rejette les $\hat{k}$ plus petites p-valeurs
  $\hookrightarrow$ $\hat{k}$ indique le # de p-valeurs qu'on peut rejeter tout en gardant $E[\frac{|\mathcal{H}_0 \cap \mathcal{R}|}{|\mathcal{R}|}] \leq \alpha$

- Contrôle du FDR = contrôle du # d'erreurs plus libéral
  $\hookrightarrow$ plus on rejette plus on autorise à faire des erreurs

- Puissance de la procédure beaucoup plus forte que celle de Bonferroni

Visualisation des deux procédures

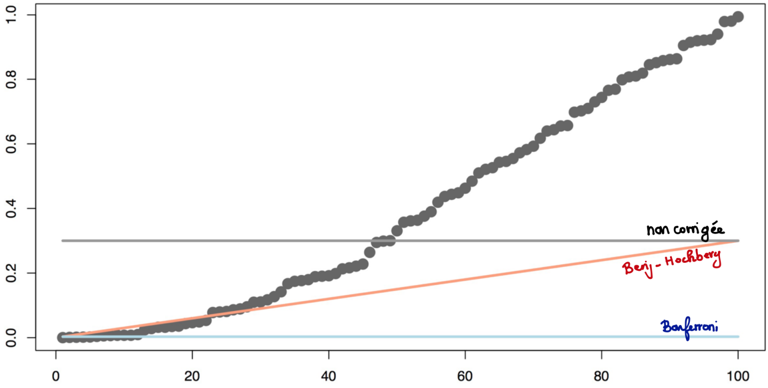

Figure: Comparaison visuelle des procédures de tests multiples avec la procédure de test simple (i.e. tout tester au même niveau α)

Visualisation des deux procédures

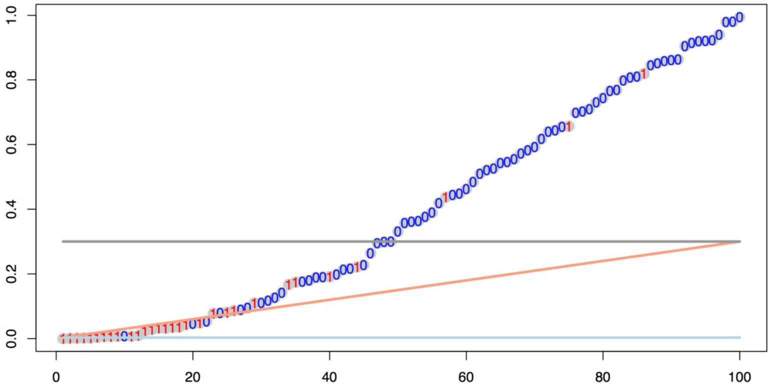

Figure: Ici on indique en plus les vraies hypothèses nulles (avec un 0) et les vraies hypothèses alternatives (avec un 1).

Note de fin

Bon à savoir

- FWER plutôt utilisé dans les phases finales d'essais cliniques
- FDR plutôt utilisé dans les études exploratoires

A retenir

- Le problème généré par la multiplicité
- Comment apporter une solution à ce problèmes
- FWER et FDR : définitions + interprétations

Note de fin

Bon à savoir

- FWER plutôt utilisé dans les phases finales d'essais cliniques
- FDR plutôt utilisé dans les études exploratoires

A retenir

- Le problème généré par la multiplicité
- Comment apporter une solution à ce problèmes
- FWER et FDR : définitions + interprétations