UNIVERSITÉ DE NANTES

# Chapitre 1 :

# Comparaison d'une variable quantitative sur *J* groupes indépendants (ANOVA à un facteur sur groupes indépendants)

Module HPS5-42
Année 2021-22

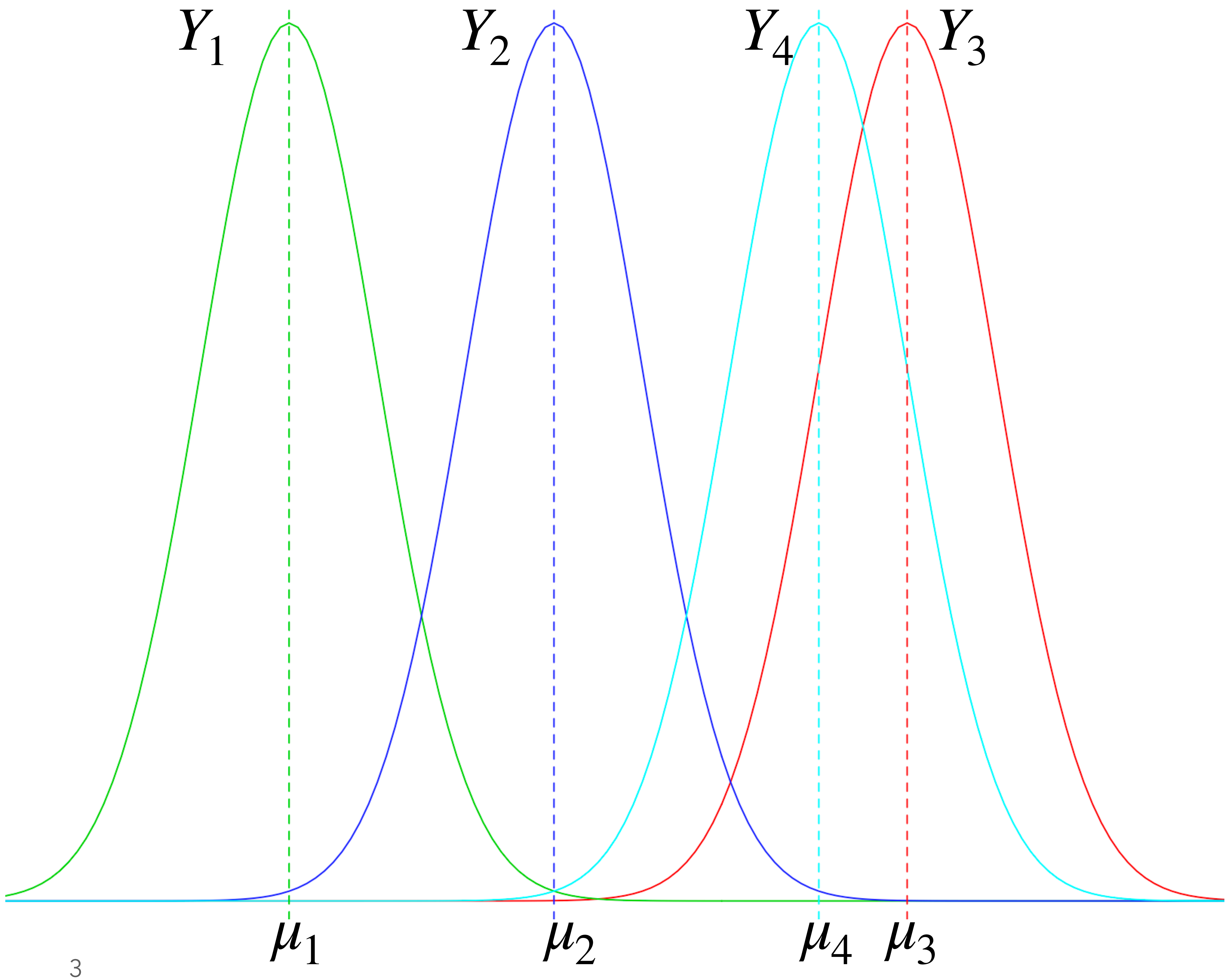
GALHARRET J-M,
Laboratoire de Mathématiques Jean Leray
Faculté de Psychologie.

# Test de Fisher

## (Test OMNIBUS de l’ANOVA)

# Généralités

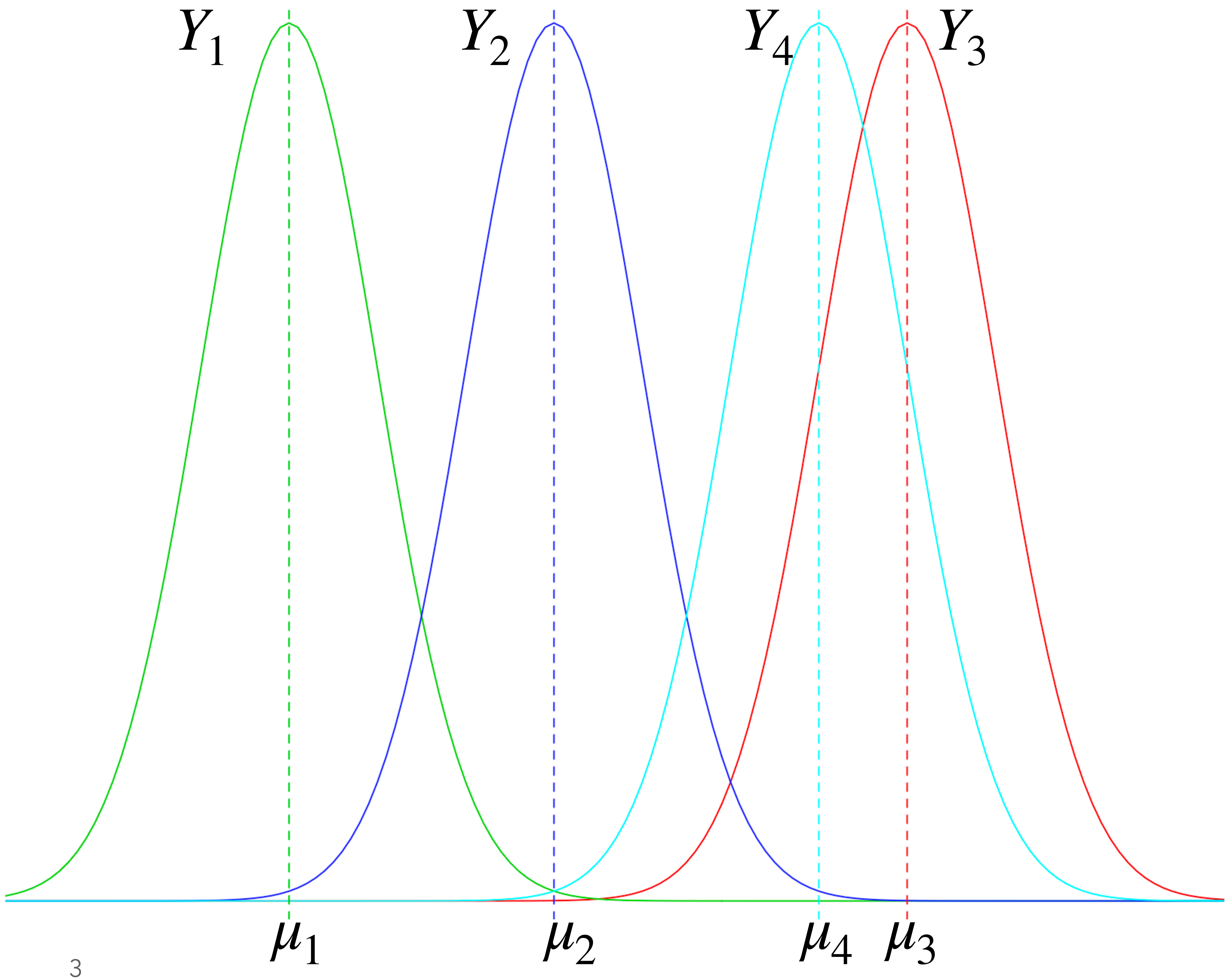
## Conditions d'application de l'ANOVA



# Généralités

## Conditions d'application de l'ANOVA

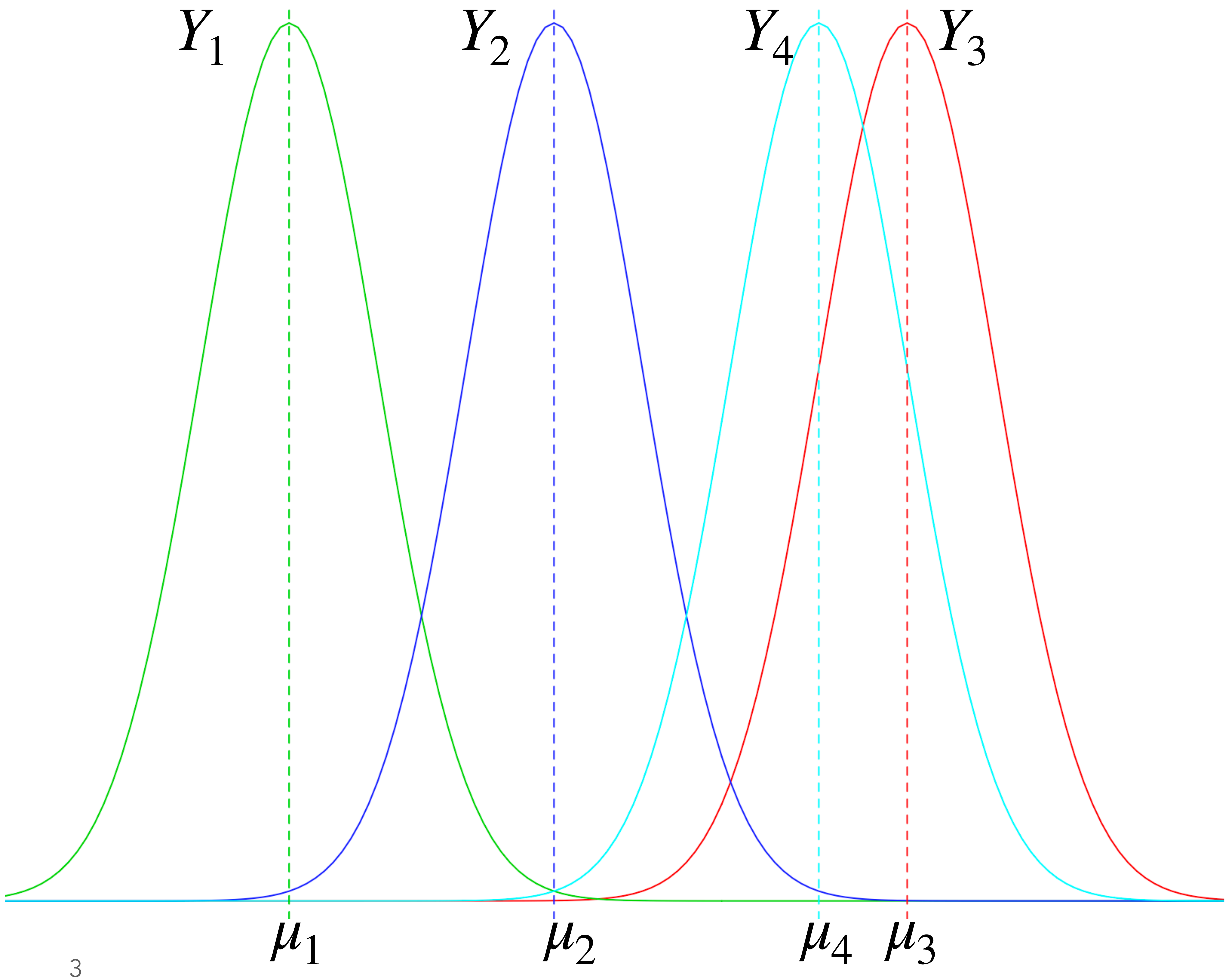
- On considère X une variable indépendante catégorielle à J modalités (J>2, on parle de facteur) et une variable dépendante Y quantitative.



# Généralités

## Conditions d'application de l'ANOVA

- On considère X une variable indépendante catégorielle à J modalités (J>2, on parle de facteur) et une variable dépendante Y quantitative.
- On suppose que la distribution de la variable $Y$ est connue sur la modalité $j$ et que :

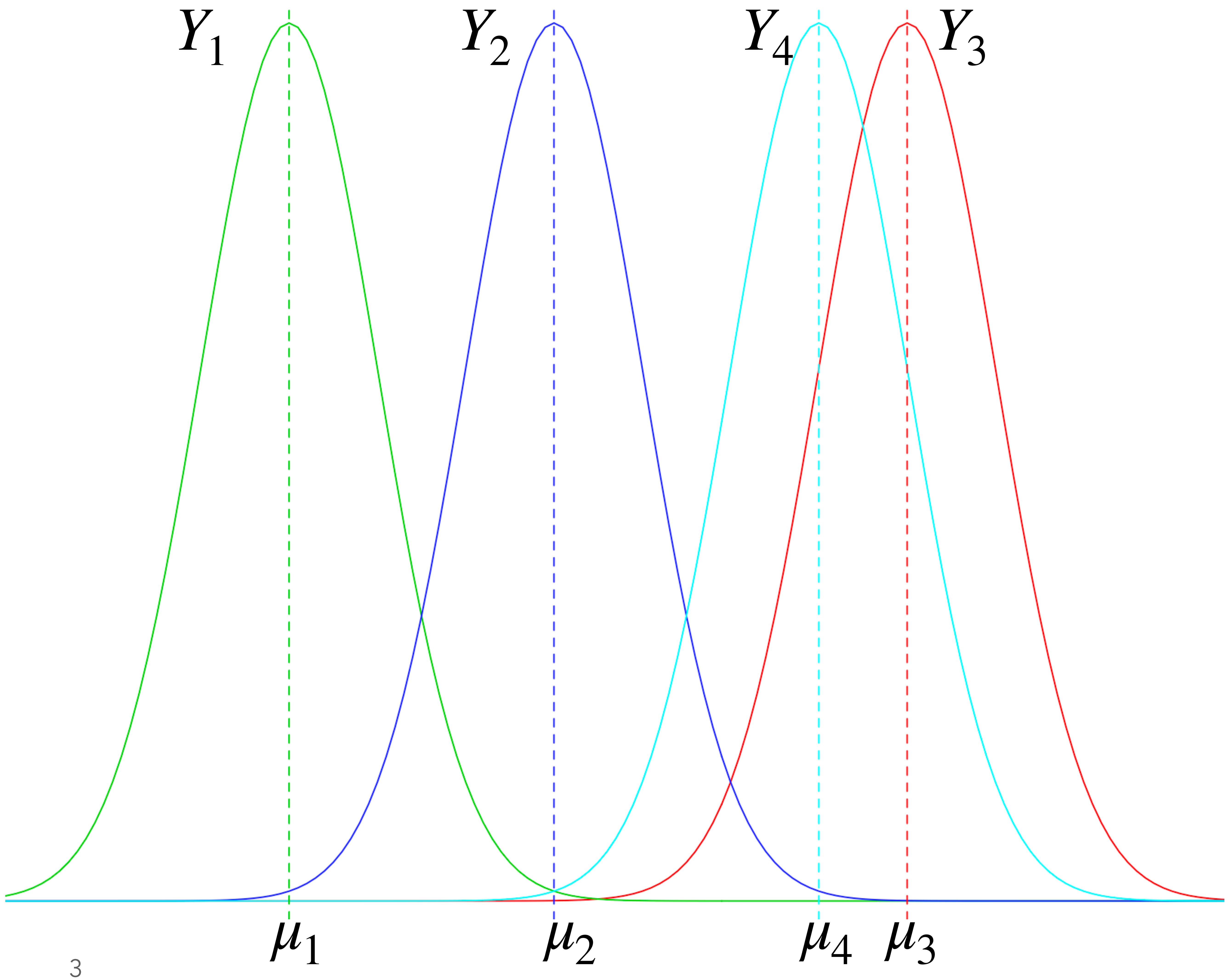


# Généralités

## Conditions d'application de l'ANOVA

- On considère X une variable indépendante catégorielle à J modalités (J>2, on parle de facteur) et une variable dépendante Y quantitative.
- On suppose que la distribution de la variable $Y$ est connue sur la modalité $j$ et que :

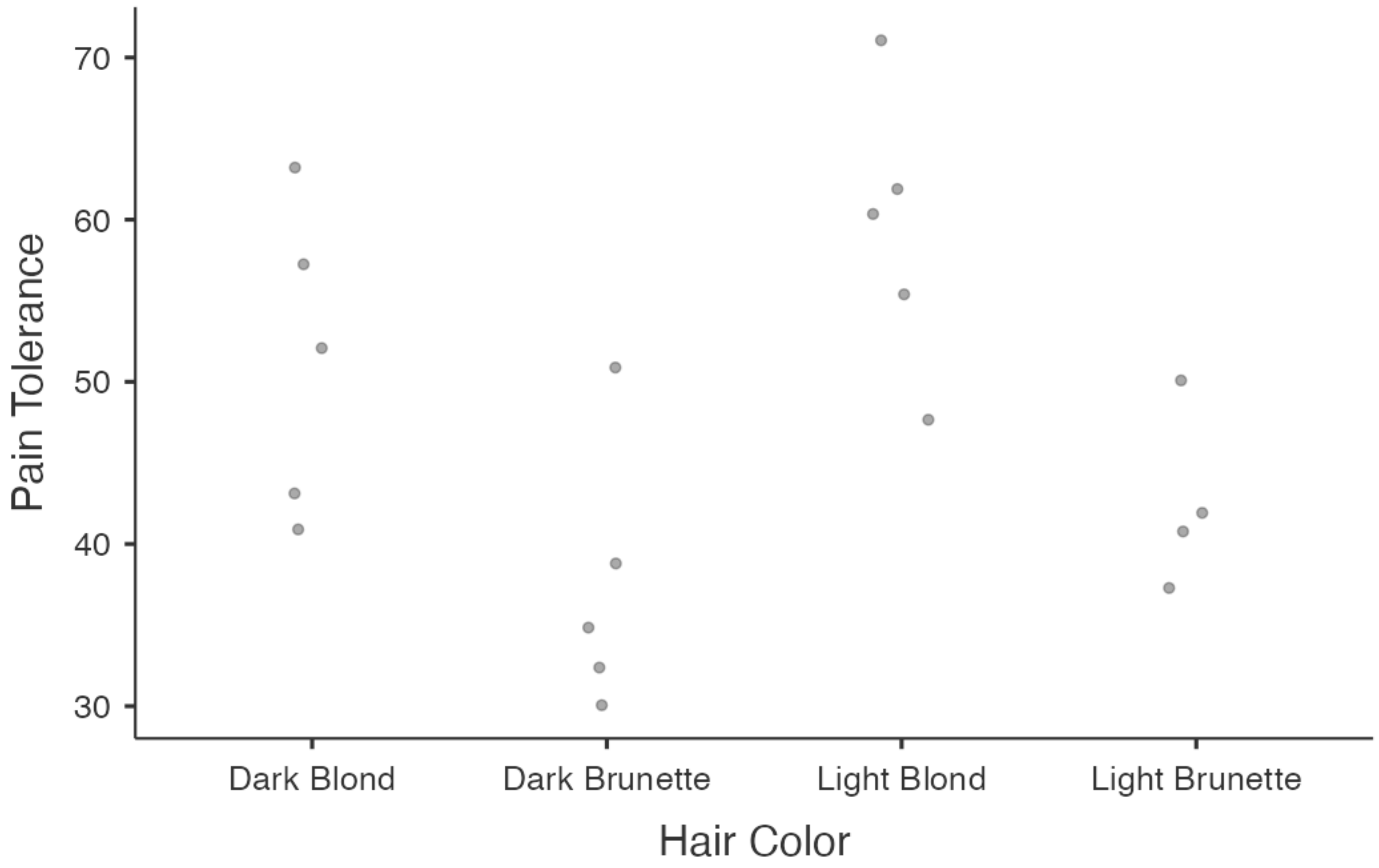
$$Y_j \sim \mathcal{N}(\mu_j, \sigma^2)$$



# Exemple sensibilité à la douleur/couleur de cheveux

McClave, J. T. and Dietrich, F. H. (1991). Statistics. San Francisco: Dellen publishing.

| Identifica... | Hair Color | Pain Tole... |
|---|---|---|
| 1 | Light Blond | 62 |
| 2 | Light Blond | 60 |
| 3 | Light Blond | 71 |
| 4 | Light Blond | 55 |
| 5 | Light Blond | 48 |
| 6 | Dark Blond | 63 |
| 7 | Dark Blond | 57 |
| 8 | Dark Blond | 52 |
| 9 | Dark Blond | 41 |
| 10 | Dark Blond | 43 |
| 11 | Light Brunette | 42 |
| 12 | Light Brunette | 50 |
| 13 | Light Brunette | 41 |
| 14 | Light Brunette | 37 |
| 15 | Dark Brunette | 32 |
| 16 | Dark Brunette | 39 |
| 17 | Dark Brunette | 51 |
| 18 | Dark Brunette | 30 |
| 19 | Dark Brunette | 35 |



# Statistiques Descriptives

# Statistiques Descriptives

| Descriptives - Pain Tolerance | | | |
|---|---|---|---|
| **Hair Color** | **Mean** | **s** | **n** |
| DBl | 51,2 | 9,284 | 5 |
| DBr | 37,400 | 8,325 | 5 |
| LBl | 59,200 | 8,526 | 5 |
| LBr | 43,400 | 5,128 | 5 |

# Statistiques Descriptives

| Descriptives - Pain Tolerance | | | |
|---|---|---|---|
| **Hair Color** | **Mean** | **s** | **n** |
| DBl | 51,2 | 9,284 | 5 |
| DBr | 37,400 | 8,325 | 5 |
| LBl | 59,200 | 8,526 | 5 |
| LBr | 43,400 | 5,128 | 5 |

La moyenne des N=20 individus est :

# Statistiques Descriptives

| **Descriptives - Pain Tolerance** | | | |
|---|---|---|---|
| **Hair Color** | **Mean** | **s** | **n** |
| DBl | 51,2 | 9,284 | 5 |
| DBr | 37,400 | 8,325 | 5 |
| LBl | 59,200 | 8,526 | 5 |
| LBr | 43,400 | 5,128 | 5 |

La moyenne des N=20 individus est :

$$m = \frac{51.2 + 36.4 + 59.2 + 43.4}{4} = 47.8$$

# Variabilité de la variable dépendante Y

Somme des carrés des écarts

# Variabilité de la variable dépendante Y

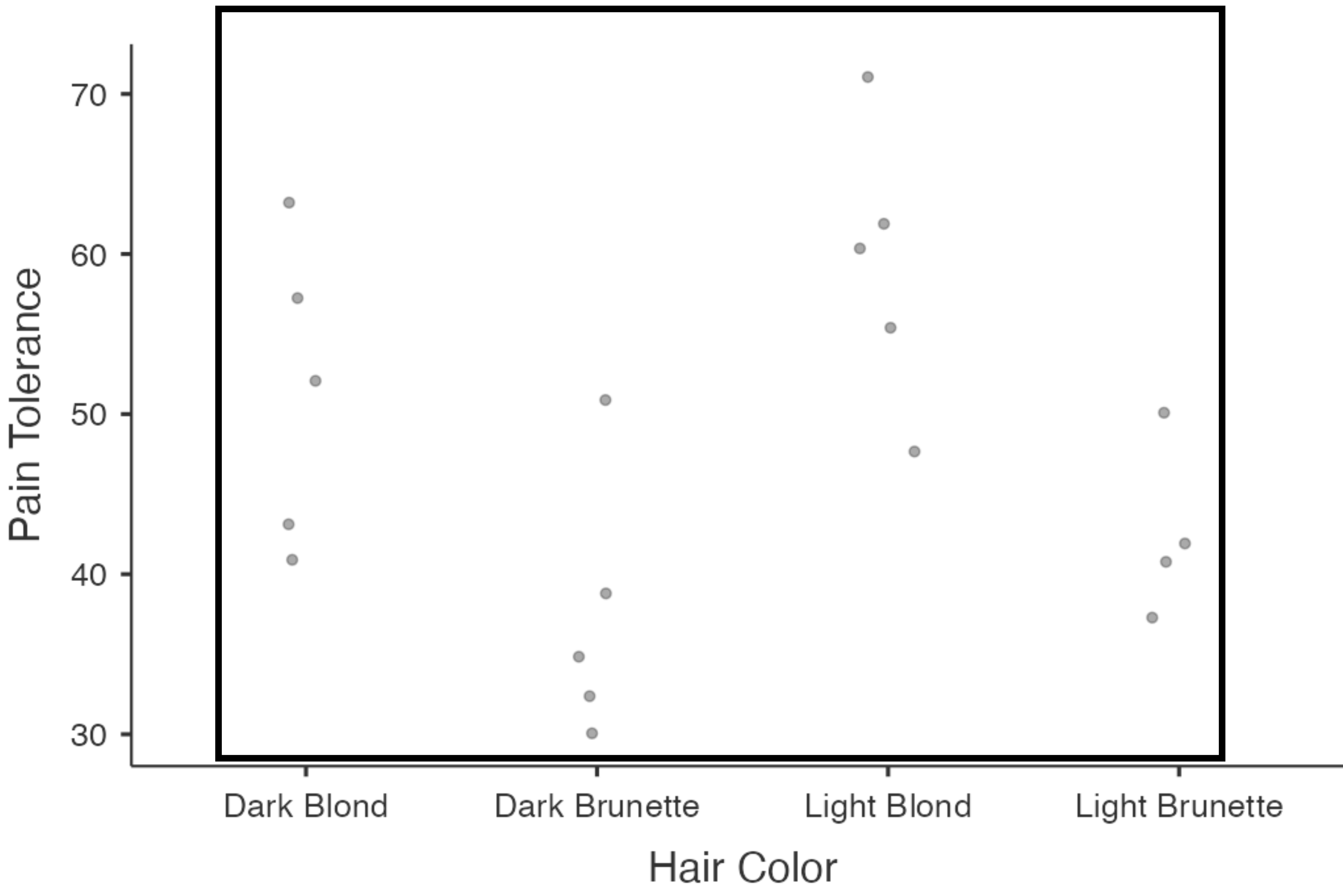
## Somme des carrés des écarts

La variabilité de Y (Pain tolerance) est mesurée par SCE (somme carrée des écarts) :

# Variabilité de la variable dépendante Y

## Somme des carrés des écarts

La variabilité de Y (Pain tolerance) est mesurée par SCE (somme carrée des écarts) :

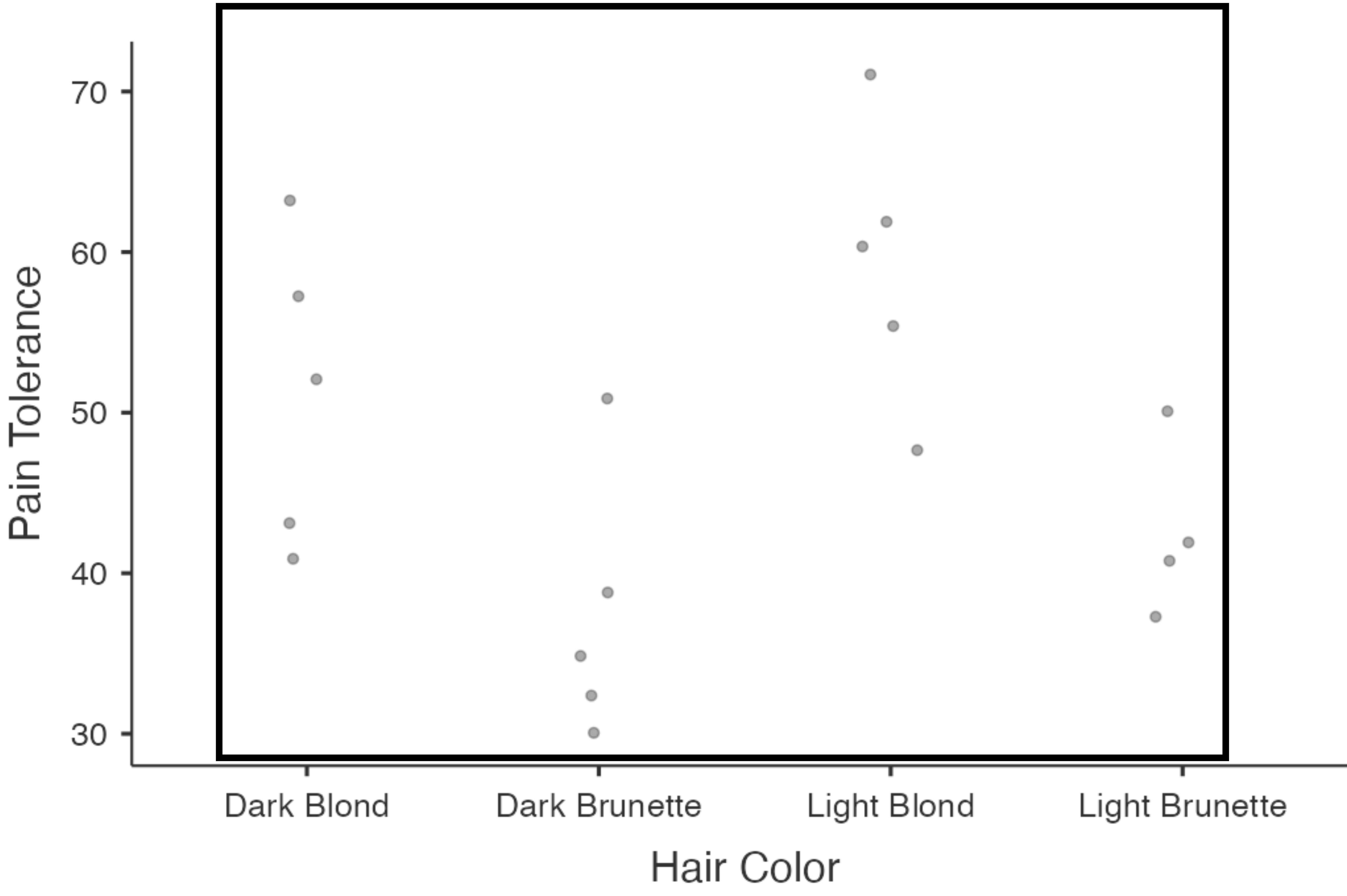


# Variabilité de la variable dépendante Y

## Somme des carrés des écarts

La variabilité de Y (Pain tolerance) est mesurée par SCE (somme carrée des écarts) :

$$
\begin{aligned}
SCE_Y &= N\sigma_Y^2 \\
&= (62 - \mathbf{47.8})^2 + (60 - \mathbf{47.8})^2 + \ldots \\
&= 2363.2
\end{aligned}
$$



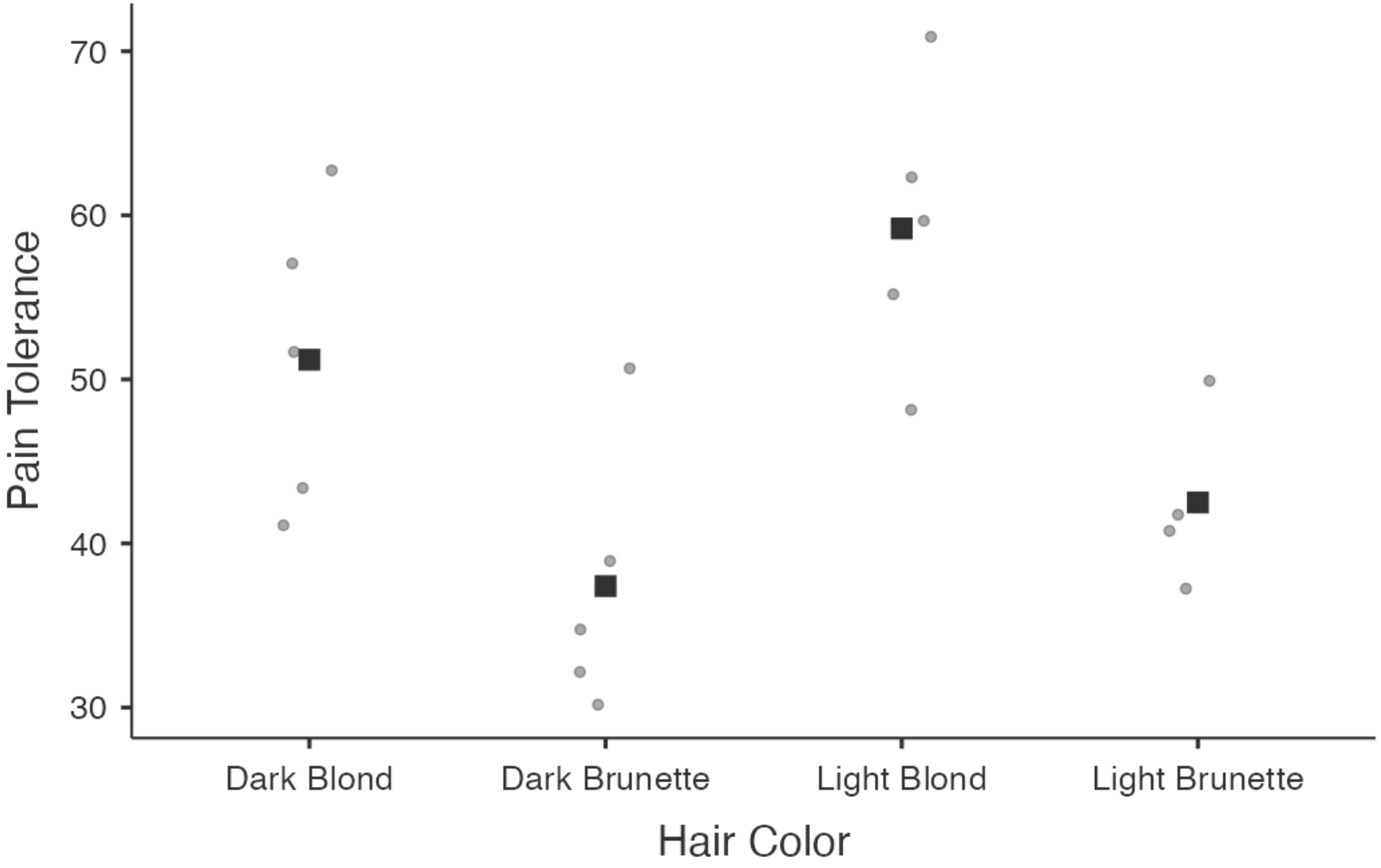
# Variabilité inter-groupes (factorielle)

# Variabilité inter-groupes (factorielle)

La variabilité entre les 4 groupes est l'écart entre les moyennes des groupes et la moyenne totale

# Variabilité inter-groupes (factorielle)

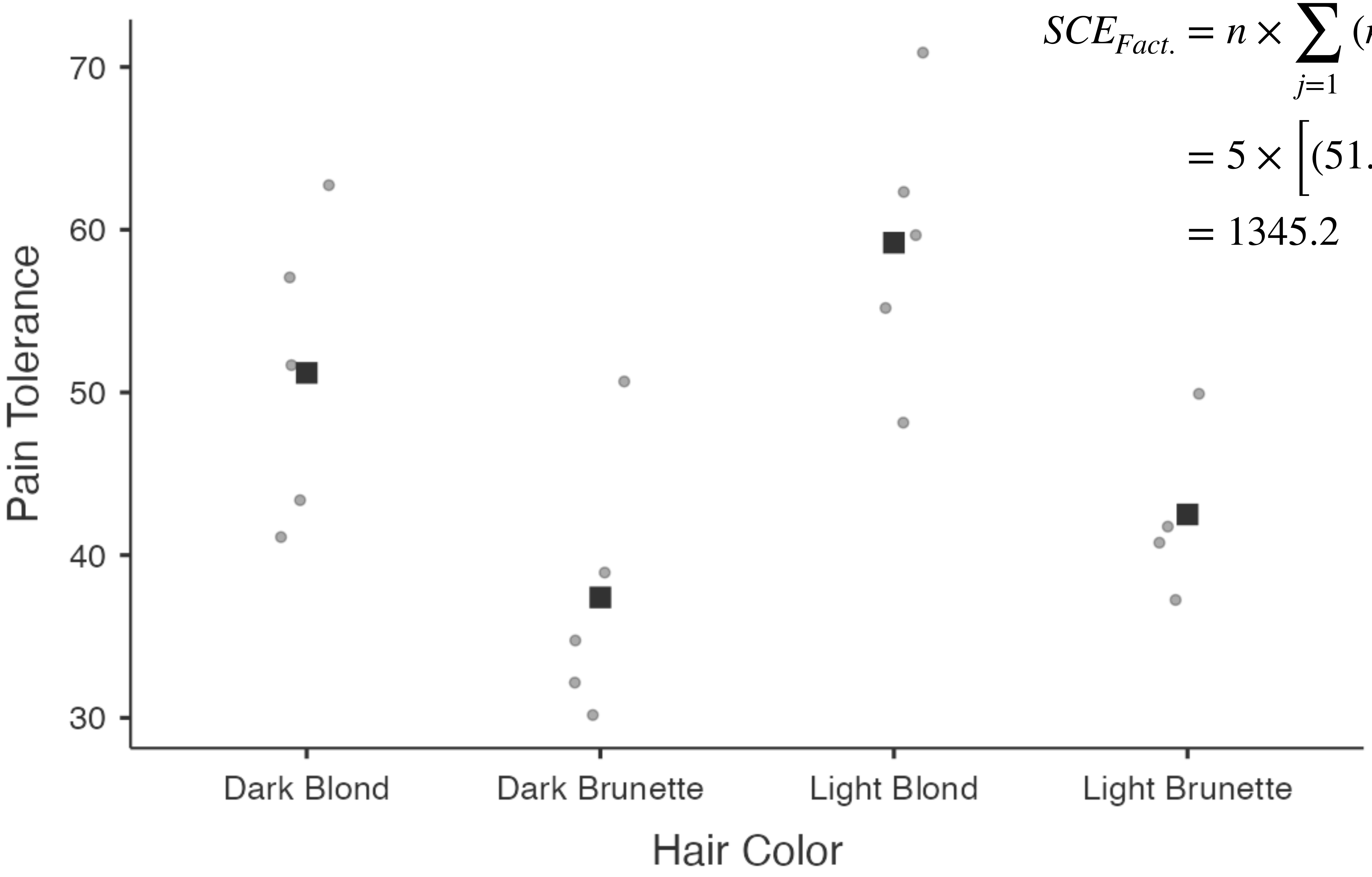
La variabilité entre les 4 groupes est l'écart entre les moyennes des groupes et la moyenne totale



# Variabilité inter-groupes (factorielle)

La variabilité entre les 4 groupes est l'écart entre les moyennes des groupes et la moyenne totale

$$
\begin{aligned}
SCE_{Fact.} &= n \times \sum_{j=1}^{J} (m_j - m)^2 \\
&= 5 \times \left[(51.2 - 47.8)^2 + (37.4 - 47.8)^2 + (59.2 - 47.8)^2 + (43.4 - 47.8)^2\right] \\
&= 1345.2
\end{aligned}
$$



# Variabilité résiduelle

## Somme des Carrés des Ecarts Résiduels

# Variabilité résiduelle

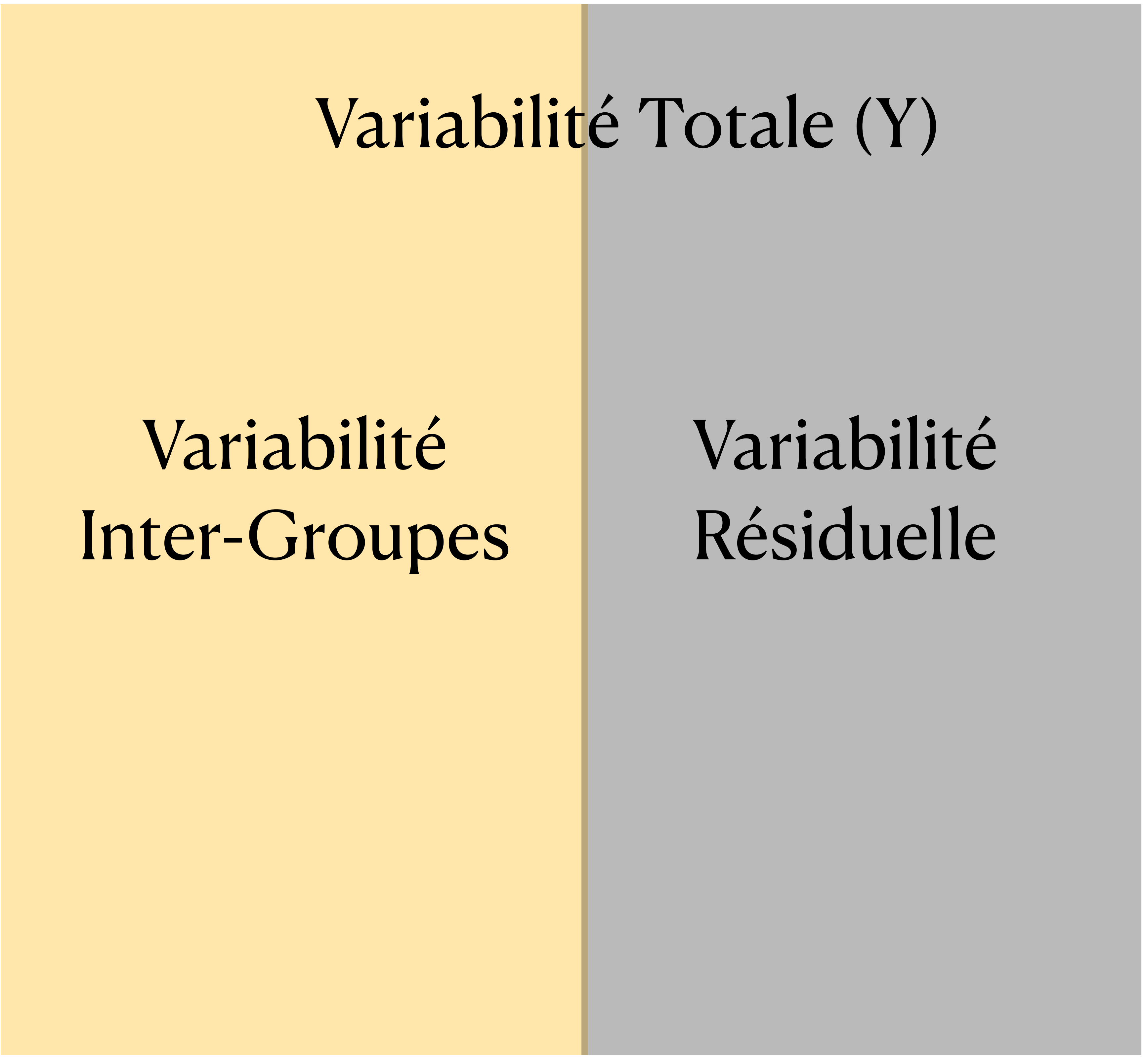
**Somme des Carrés des Ecarts Résiduels**

- La variabilité résiduelle est la variabilité de Y qui n'est pas associée au fait d'appartenir aux groupes c'est à dire

# Variabilité résiduelle

## Somme des Carrés des Ecarts Résiduels

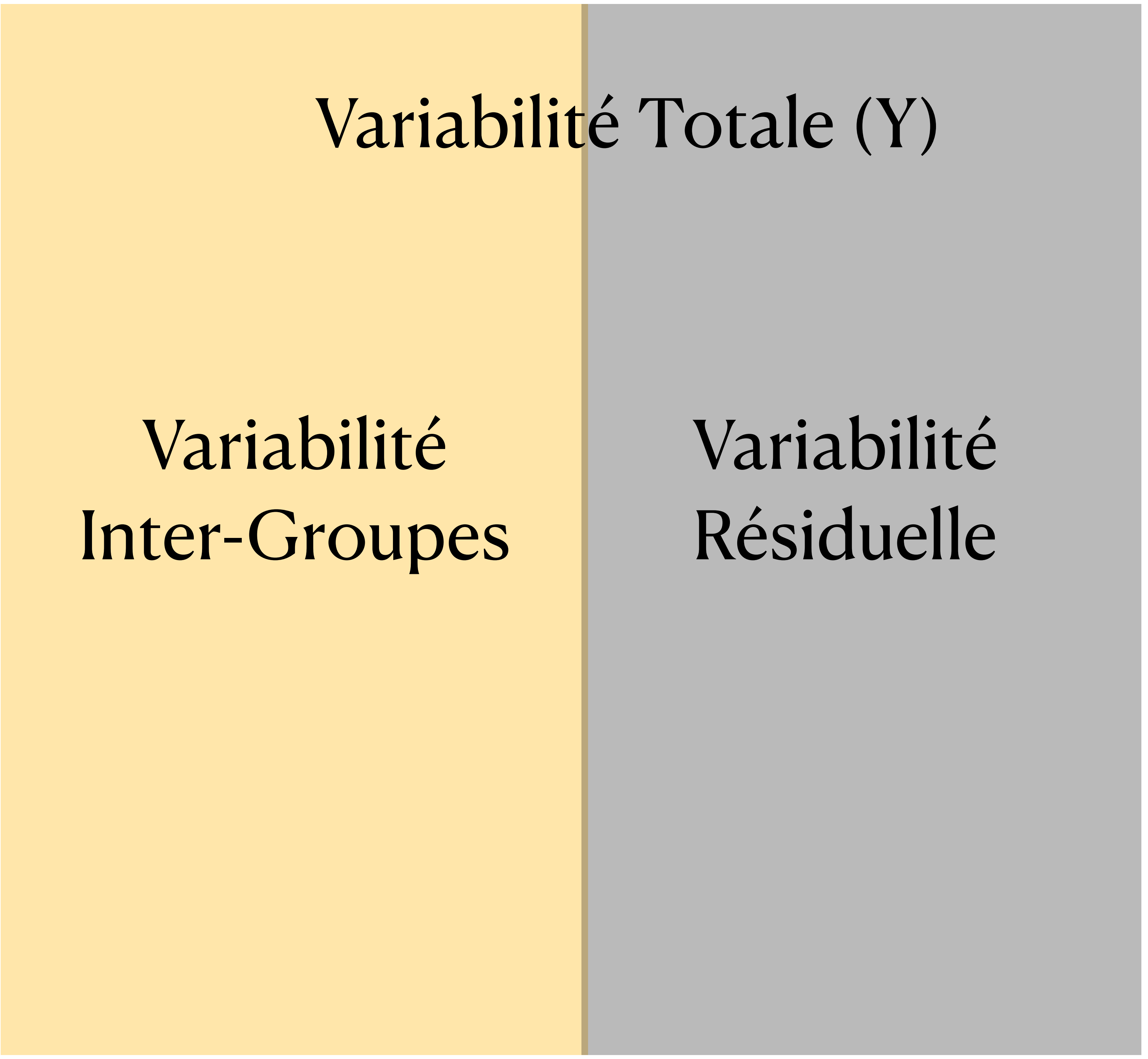
- La variabilité résiduelle est la variabilité de Y qui n'est pas associée au fait d'appartenir aux groupes c'est à dire



# Variabilité résiduelle

## Somme des Carrés des Ecarts Résiduels

- La variabilité résiduelle est la variabilité de Y qui n'est pas associée au fait d'appartenir aux groupes c'est à dire




$$SCE_Y = SCE_{Fact} + SCE_{Res}$$

# Les ddl et les variances associées

## Calculs

# Les ddl et les variances associées

## Calculs

- Le ddl factoriel est égal à J-1 et la variance factorielle à $s^2_{Fact.} = \dfrac{SCE_{Fact.}}{J-1}$,

# Les ddl et les variances associées

## Calculs

- Le ddl factoriel est égal à J-1 et la variance factorielle à $s^2_{Fact.} = \dfrac{SCE_{Fact.}}{J-1}$,
- Le ddl résiduel est égal à N-J er la variance résiduelle à $s^2_{Res.} = \dfrac{SCE_{Res.}}{N-J}$,

# Les ddl et les variances associées

## Calculs

- Le ddl factoriel est égal à J-1 et la variance factorielle à $s^2_{Fact.} = \dfrac{SCE_{Fact.}}{J-1}$,
- Le ddl résiduel est égal à N-J er la variance résiduelle à $s^2_{Res.} = \dfrac{SCE_{Res.}}{N-J}$,
- La variance estimée de Y est égale à $s^2_Y = \dfrac{SCE_Y}{N-1}$ (formule habituelle).

# Le pourcentage de variation associée au facteur

**ETA carré ($\eta^2$)**

# Le pourcentage de variation associée au facteur

**ETA carré ($\eta^2$)**

- Le pourcentage de variance associée au facteur F est définie par : $\eta^2 = \dfrac{SCE_{Fact.}}{SCE_Y} \times 100$

# Le pourcentage de variation associée au facteur

**ETA carré ($\eta^2$)**

- Le pourcentage de variance associée au facteur F est définie par : $\eta^2 = \frac{SCE_{Fact.}}{SCE_Y} \times 100$
- Dans l'exemple $\eta^2 = \frac{1345.2}{2363.2} \times 100 = 56.9$. Ce qui signifie que 56.9% de la variabilité de la tolérance à la douleur est associée à la couleur des cheveux.

# Le pourcentage de variation associée au facteur

**ETA carré ($\eta^2$)**

- Le pourcentage de variance associée au facteur F est définie par : $\eta^2 = \dfrac{SCE_{Fact.}}{SCE_Y} \times 100$
- Dans l'exemple $\eta^2 = \dfrac{1345.2}{2363.2} \times 100 = 56.9$. Ce qui signifie que 56.9% de la variabilité de la tolérance à la douleur est associée à la couleur des cheveux.

| $\eta^2$ | [0.01,0.06[ | [0.06,0.14[ | [0.14,1] |
|---|---|---|---|
| Effet | Faible | Modéré | Fort |

# Hypothèse testée

# Hypothèse testée

- $H_0 : \mu_1 = \ldots = \mu_J$

# Hypothèse testée

- $H_0 : \mu_1 = \ldots = \mu_J$
- $H_0 : \eta^2 = 0$

# Hypothèse testée

- $H_0 : \mu_1 = \ldots = \mu_J$
- $H_0 : \eta^2 = 0$
- $H_0$ : Les variations de $Y$ ne sont pas significativement associées au facteur considéré

# Hypothèse testée

- $H_0 : \mu_1 = \ldots = \mu_J$
- $H_0 : \eta^2 = 0$
- $H_0$ : Les variations de $Y$ ne sont pas significativement associées au facteur considéré
- $H_0$ : Il n'existe pas de différence significative au niveau de $Y$ entre les $J$ groupes considérés

# Test de Fisher

## Test Omnibus de l'ANOVA

# Test de Fisher

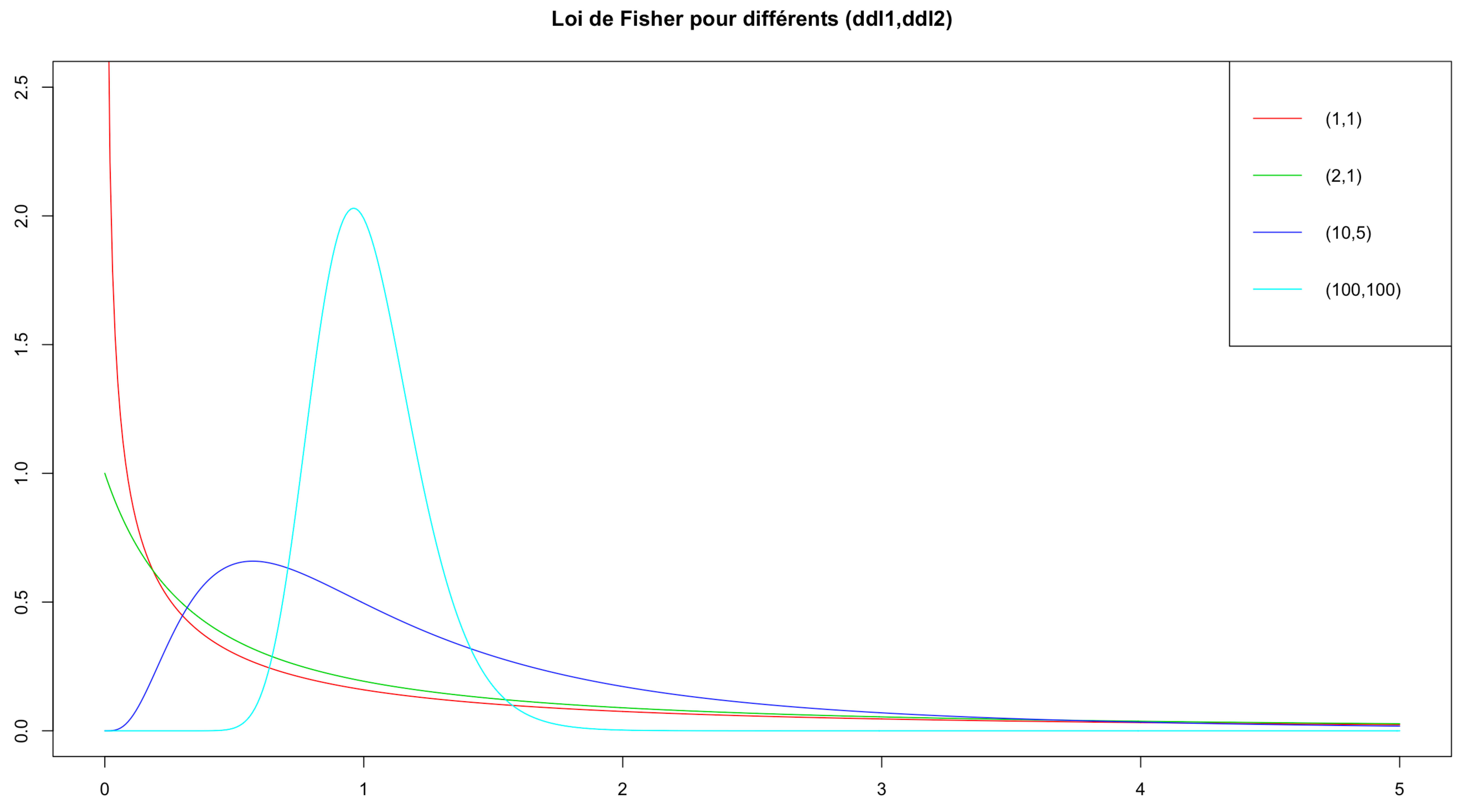
## Test Omnibus de l'ANOVA

Sous $H_O : \mu_1 = \ldots = \mu_J$ la statistique $F = \dfrac{s^2_{Fact.}}{s^2_{Res.}}$ suit une loi de Fisher à $(J-1, N-J)$ degrés de liberté.

# Test de Fisher

## Test Omnibus de l'ANOVA

Sous $H_O : \mu_1 = \ldots = \mu_J$ la statistique $F = \dfrac{s^2_{Fact.}}{s^2_{Res.}}$ suit une loi de Fisher à $(J-1, N-J)$ degrés de liberté.

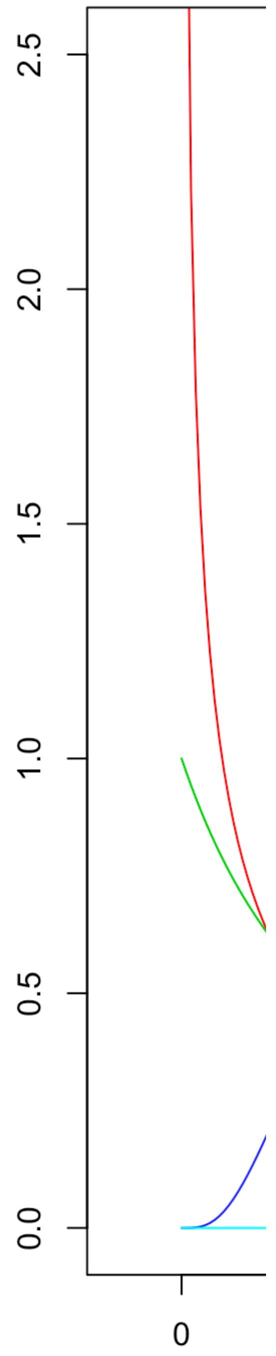


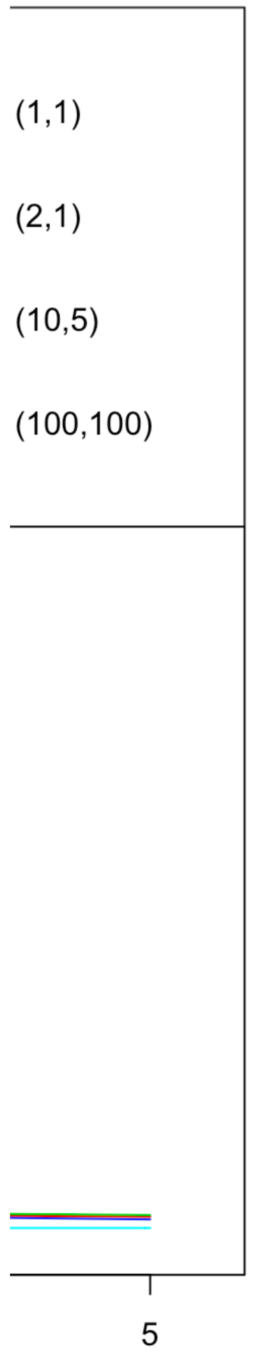
# Test d

## Test Omnibu

Sous $H_O : \mu_1 = \ldots = \mu_J$ la statistique $F = \dfrac{s^2_{Fact.}}{s^2_{Res.}}$ suit une loi de Fisher à $(J-1, N-J)$ degrés de liberté.

| ddl résiduel | p_value 0,1 | 0,05 | 0,01 | 0,001 |
|---|---|---|---|---|
| 10 | 2,728 | 3,708 | 6,552 | 12,553 |
| 11 | 2,660 | 3,587 | 6,217 | 11,561 |
| 12 | 2,606 | 3,490 | 5,953 | 10,804 |
| 13 | 2,560 | 3,411 | 5,739 | 10,209 |
| 14 | 2,522 | 3,344 | 5,564 | 9,729 |
| 15 | 2,490 | 3,287 | 5,417 | 9,335 |
| 20 | 2,380 | 3,098 | 4,938 | 8,098 |
| 25 | 2,317 | 2,991 | 4,675 | 7,451 |
| 30 | 2,276 | 2,922 | 4,510 | 7,054 |
| 35 | 2,247 | 2,874 | 4,396 | 6,787 |
| 40 | 2,226 | 2,839 | 4,313 | 6,595 |
| 50 | 2,197 | 2,790 | 4,199 | 6,336 |
| 60 | 2,177 | 2,758 | 4,126 | 6,171 |
| 120 | 2,130 | 2,680 | 3,949 | 5,781 |
| inf | 2,084 | 2,605 | 3,782 | 5,422 |



# Résultat Test de Fisher

## ANOVA à un facteur sur groupes indépendants

# Résultat Test de Fisher

## ANOVA à un facteur sur groupes indépendants

**ANOVA - Pain Tolerance**

| Cases | Sum of Squares | df | Mean Square | F | p |
|---|---|---|---|---|---|
| Hair Color | 1345.2 | 3 | 448.4 | 7.05 | 0.003 |
| Residuals | 1018.0 | 16 | 63.6 | | |

# Résultat Test de Fisher

## ANOVA à un facteur sur groupes indépendants

**ANOVA - Pain Tolerance**

| Cases | Sum of Squares | df | Mean Square | F | p |
|---|---|---|---|---|---|
| Hair Color | 1345.2 | 3 | 448.4 | 7.05 | 0.003 |
| Residuals | 1018.0 | 16 | 63.6 | | |

Une ANOVA à un facteur sur groupes indépendants a permis de mettre en évidence une relation entre la couleur des cheveux et la sensibilité à la douleur des individus de l'échantillon considéré, $F(3,16) = 7.05$, $p = .003$.

**Attention :** on parle de relation, d'association on ne parle pas de causalité ! On peut aussi dire « expliqué par » même si c'est moins exact.

# Test Post-Hoc

# Position du problème

## Comparaison groupes à groupes

# Position du problème

## Comparaison groupes à groupes

- Lorsque le test de Fisher conduit au rejet de Ho, c'est à dire qu'il existe une différence significative entre les moyennes des différents groupes alors on peut se poser la question de savoir quels sont les groupes qui sont différents les uns des autres.

# Position du problème

## Comparaison groupes à groupes

- Lorsque le test de Fisher conduit au rejet de Ho, c'est à dire qu'il existe une différence significative entre les moyennes des différents groupes alors on peut se poser la question de savoir quels sont les groupes qui sont différents les uns des autres.
- On va comparer deux à deux les groupes mais on n'utilise pas le test de Student qui induirait un problème de niveau de la procédure.

# Position du problème

## Comparaison groupes à groupes

- Lorsque le test de Fisher conduit au rejet de Ho, c'est à dire qu'il existe une différence significative entre les moyennes des différents groupes alors on peut se poser la question de savoir quels sont les groupes qui sont différents les uns des autres.
- On va comparer deux à deux les groupes mais on n'utilise pas le test de Student qui induirait un problème de niveau de la procédure.
- On va utiliser le test de Tukey qui est une version corrigée du test de Student.

# Test Post-Hoc de Tukey

## Mise en oeuvre du test

# Test Post-Hoc de Tukey

## Mise en oeuvre du test

- Soient A et B deux groupes parmi les J groupes que l'on veut comparer. On note leurs moyennes $m_A, m_B$.

# Test Post-Hoc de Tukey

## Mise en oeuvre du test

- Soient A et B deux groupes parmi les J groupes que l'on veut comparer. On note leurs moyennes $m_A, m_B$.
- On teste l'hypothèse $H_0 : \mu_A = \mu_B$ contre $H_1 : \mu_A \neq \mu_B$

# Test Post-Hoc de Tukey

## Mise en oeuvre du test

- Soient A et B deux groupes parmi les J groupes que l'on veut comparer. On note leurs moyennes $m_A, m_B$.
- On teste l'hypothèse $H_0 : \mu_A = \mu_B$ contre $H_1 : \mu_A \neq \mu_B$
- Sous Ho la statistique $t_{A-B} = \dfrac{m_A - m_B}{\sqrt{\dfrac{2 \times s^2_{Res.}}{n}}}$ suit une loi de Tukey à $(J, \mathrm{ddl}_{Res.})$ degrés de liberté.

# Exemple de calcul

## Test de Tukey

# Exemple de calcul

## Test de Tukey

**ANOVA - Pain Tolerance**

| Cases | Sum of Squares | df | Mean Square | F | p |
|---|---|---|---|---|---|
| Hair Color | 1345.2 | 3 | 448.4 | 7.05 | 0.003 |
| Residuals | 1018.0 | 16 | 63.6 | | |

**Descriptives - Pain Tolerance**

| Hair Color | Mean | s | | n |
|---|---|---|---|---|
| DBl | 51,2 | 9,284 | 8,303 | 5 |
| DBr | 37,400 | 8,325 | 7,446 | 5 |
| LBl | 59,200 | 8,526 | 7,625 | 5 |
| LBr | 43,400 | 5,128 | 4,586 | 5 |

# Exemple de calcul

## Test de Tukey

**ANOVA - Pain Tolerance**

| Cases | Sum of Squares | df | Mean Square | F | p |
|---|---|---|---|---|---|
| Hair Color | 1345.2 | 3 | 448.4 | 7.05 | 0.003 |
| Residuals | 1018.0 | 16 | 63.6 | | |

**Descriptives - Pain Tolerance**

| Hair Color | Mean | s | | n |
|---|---|---|---|---|
| DBl | 51,2 | 9,284 | 8,303 | 5 |
| DBr | 37,400 | 8,325 | 7,446 | 5 |
| LBl | 59,200 | 8,526 | 7,625 | 5 |
| LBr | 43,400 | 5,128 | 4,586 | 5 |

$$t_{DBl-DBr} = \frac{51.2 - 37.4}{\sqrt{\frac{2 \times 63.6}{5}}} = 2.736$$

# Retour sur l'exemple

## Table Résumé

# Retour sur l'exemple

## Table Résumé

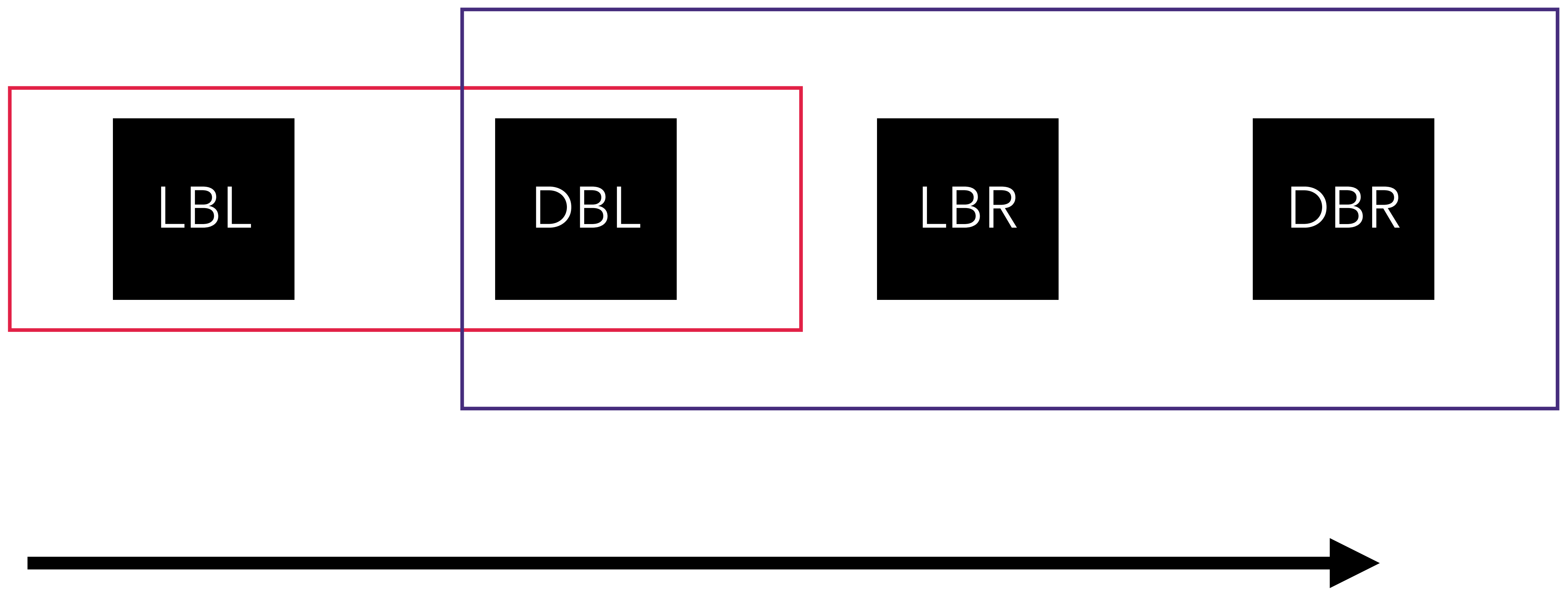
| Post Hoc Comparisons - Hair Color | | | | | |
|---|---|---|---|---|---|
| | | **Mean Diff.** | **SE** | **t** | **p tukey** |
| DBr | LBr | 6.000 | 5.045 | 1.189 | 0.642 |
| | DBl | 13.800 | 5.045 | 2.735 | 0.063 |
| | LBl | 21.800 | 5.045 | 4.321 | 0.003 |
| LBr | DBl | 7.800 | 5.045 | 1.546 | 0.435 |
| | LBl | 15.800 | 5.045 | 3.132 | 0.030 |
| DBl | LBl | 8.000 | 5.045 | 1.586 | 0.414 |

# Retour sur l'exemple

## Table Résumé

**Post Hoc Comparisons - Hair Color**

| | | **Mean Diff.** | **SE** | **t** | **p tukey** |
|---|---|---|---|---|---|
| DBr | LBr | 6.000 | 5.045 | 1.189 | 0.642 |
| | DBl | 13.800 | 5.045 | 2.735 | 0.063 |
| | LBl | 21.800 | 5.045 | 4.321 | 0.003 |
| LBr | DBl | 7.800 | 5.045 | 1.546 | 0.435 |
| | LBl | 15.800 | 5.045 | 3.132 | 0.030 |
| DBl | LBl | 8.000 | 5.045 | 1.586 | 0.414 |




On classe les groupes en fonction de leur moyenne (de la + grande à la plus petite ou le contraire !

# Vérification des conditions d’application de l’ANOVA

# Conditions d'application

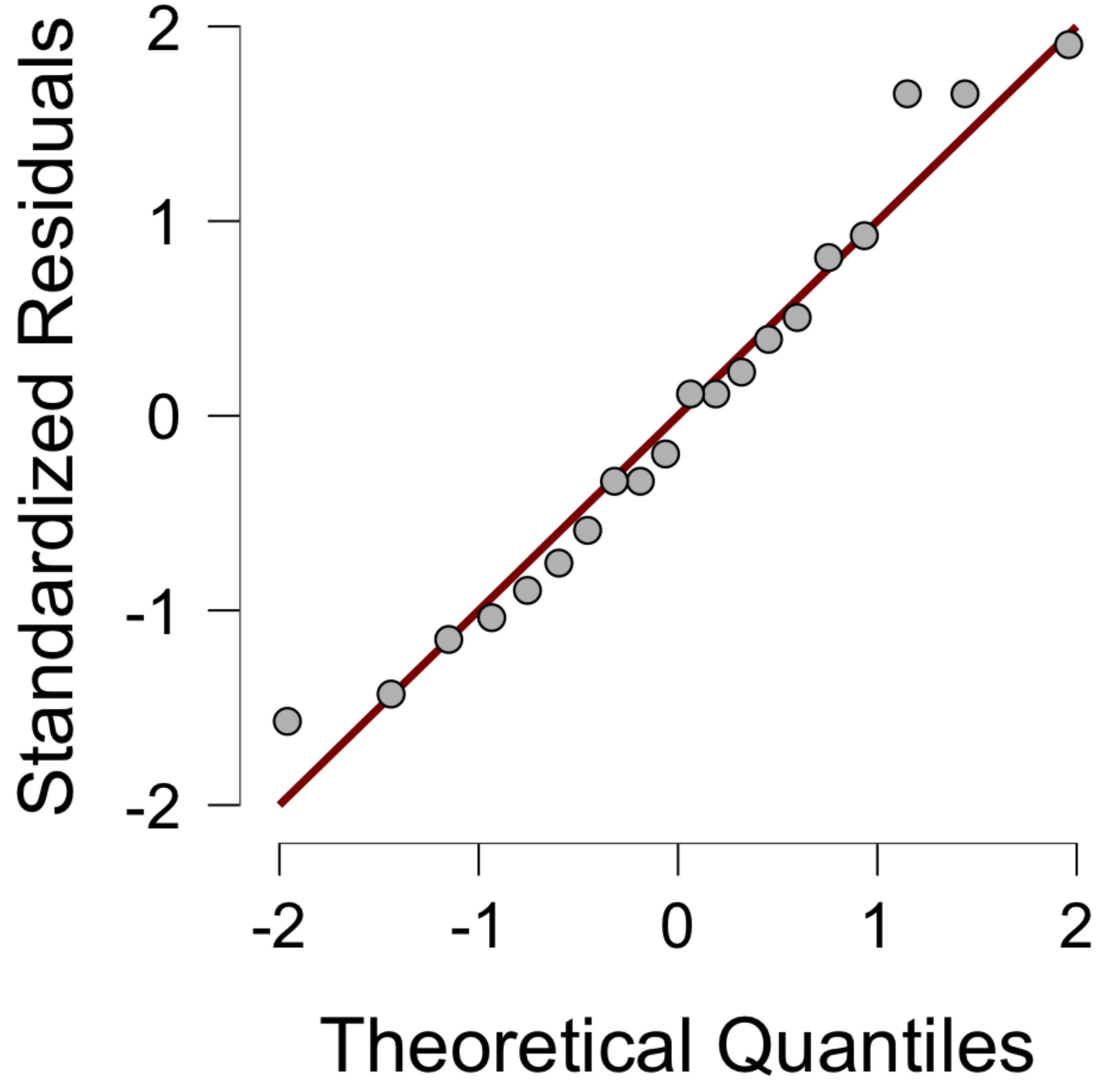
## Normalité et homogénéité des variances

- La condition de normalité requise n'est en fait pas que $Y_j \sim \mathcal{N}(\mu_j, \sigma^2)$ mais que les résidus du modèle soit normaux. Pour le vérifier il suffit de faire le test de Shapiro-Wilk ou bien de regarder le Q-Q plot.
- La condition d'homogénéité des variances c'est à dire que $\sigma_1^2 = \ldots = \sigma_J^2$ se vérifie avec le test de Levene.

# Normalité des résidus

## Test de Shapiro-Wilk

On teste Ho les résidus standardisés du modèle suivent une loi normale contre H1 les résidus ne sont pas distribués normalement. Ici la condition de normalité ne peut pas être rejetée, $W = .96$, $p = .553$.



# Homogénéité des variances

## Test de Levene

- On teste $H_0 : \sigma_1^2 = \ldots = \sigma_J^2$ contre $H1$ l'une au moins des variance est significativement différentes des autres.
- On considère pour ce faire le test de Levene (voir table ci-dessous) : on ne peut pas considérer au risque de 5% que l'une des variances est significativement différentes des autres, $F(3,16) = 0.511,\ p = .681$

**Test for Equality of Variances (Levene's)**

| F | df1 | df2 | p |
|---|---|---|---|
| 0.511 | 3.000 | 16.000 | 0.681 |