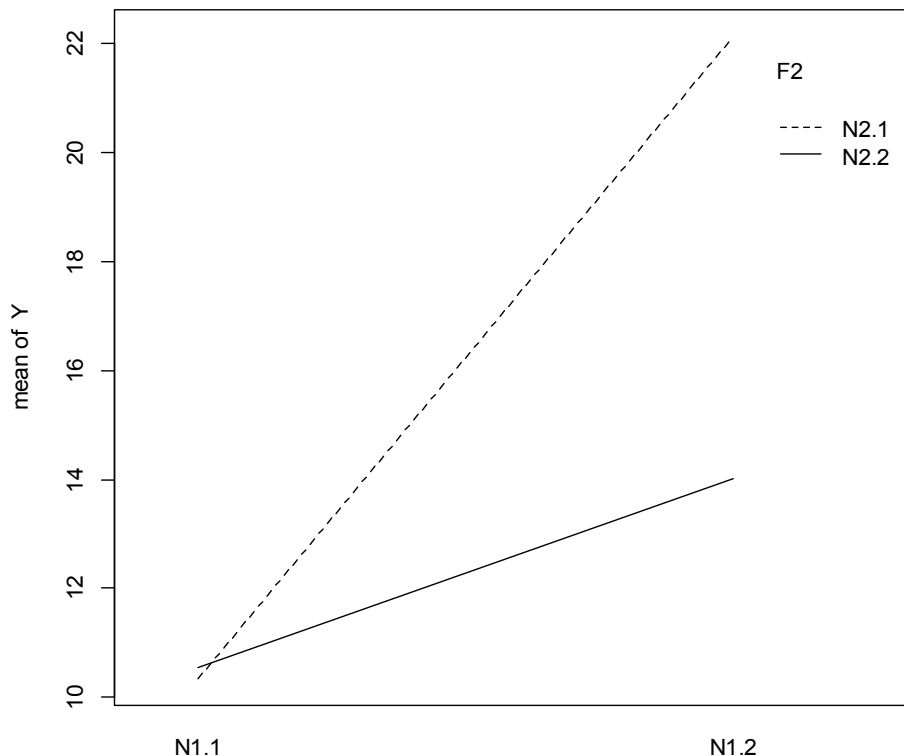


Conduire une analyse linéaire (1) : la table des effets

```
with( donnees, interaction.plot( F1, F2, Y)))
```

```
anova( lm( Y ~ F1 * F2, donnees))
```



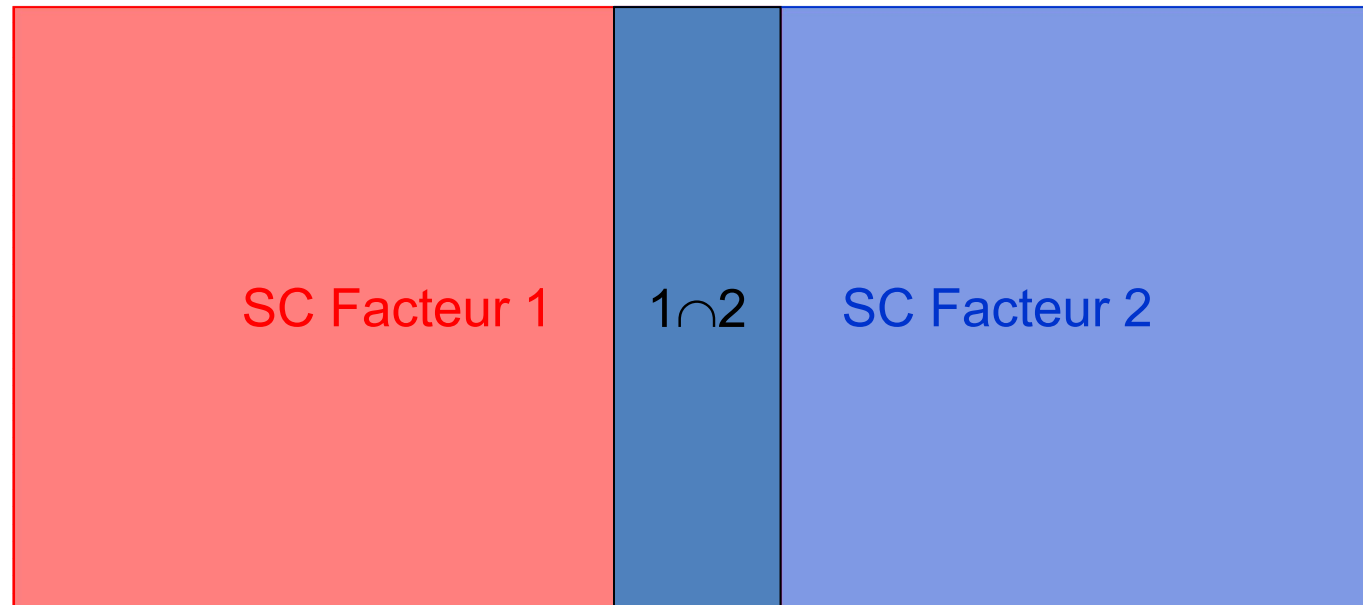
```
Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)
F1      1  320.87   320.87   657.39 < 2.2e-16 ***
F2      1  111.08   111.08   227.57 2.167e-12 ***
F1:F2    1  103.27   103.27   211.57 4.238e-12 ***
Residuals 20    9.76     0.49
---
```

```
anova( lm( Y ~ F2 * F1, donnees))
```

```
Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)
F2      1   81.08    81.08   166.12 3.809e-11 ***
F1      1  350.86   350.86   718.85 < 2.2e-16 ***
F2:F1    1  103.27   103.27   211.57 4.238e-12 ***
Residuals 20    9.76     0.49
---
```

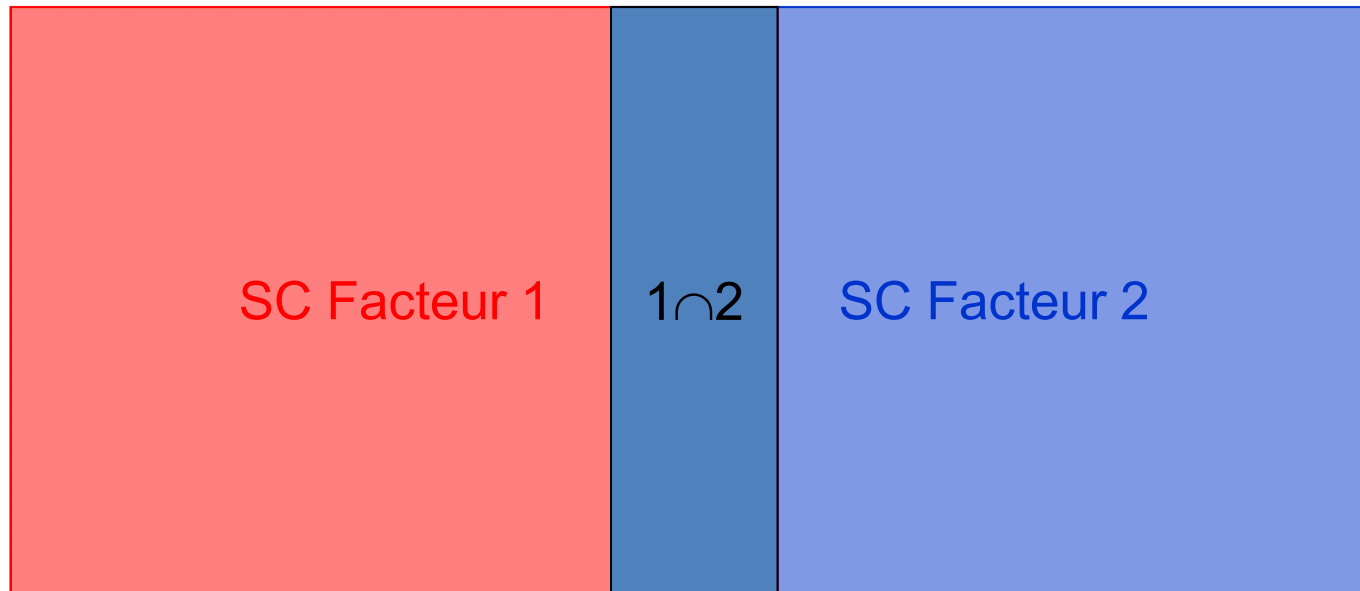
Résultats différents : pourquoi ?

Somme des carrés type I (fonction anova)



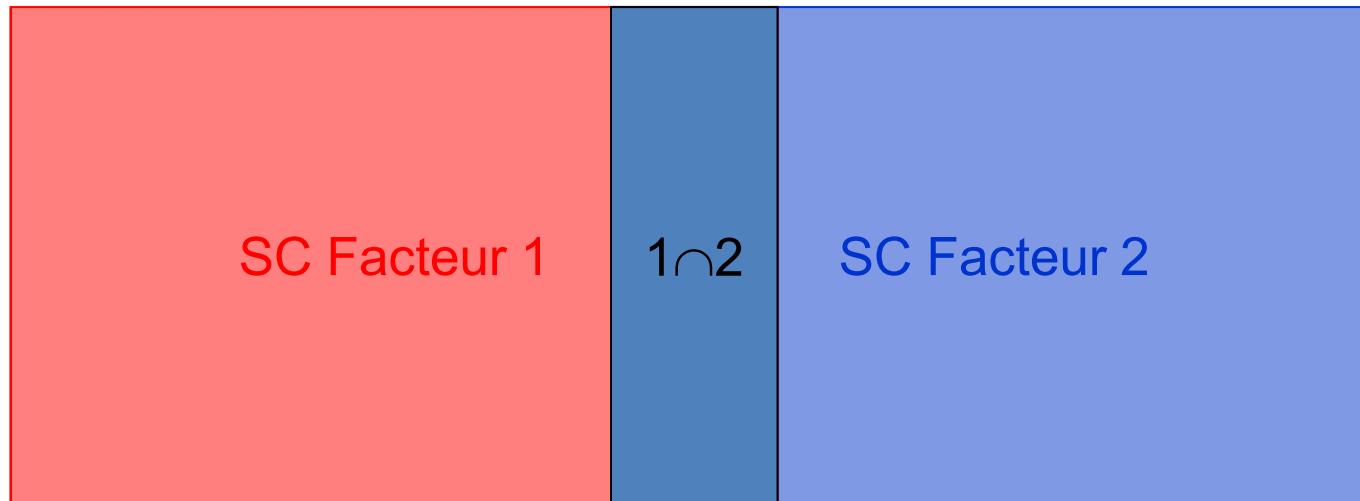
- 1) SC[Facteur 1] (rouge + bleu/gris)
 - 2) SC[Facteur 2] – SC[Facteur 2 ∩ Facteur 1] (bleu)
 - 3) Interaction : bleu/gris
- Sensible à l'ordre des facteurs
 - Groupes non équilibrés : sensible à la taille des groupes

Somme des carrés type II



- 1) SC[Facteur 1] (rouge + bleu/gris)
 - 2) SC[Facteur 2] (bleu + bleu/gris)
 - 3) Interaction : bleu/gris
- Indépendant de l'ordre des facteurs
 - Si interaction, SC[Facteur] peu d'intérêt

Somme des carrés type III



- 1) $SC[\text{Facteur 1}] - SC[\text{Facteur 2} \cap \text{Facteur 1}] - \dots$ (rouge)
- 2) $SC[\text{Facteur 2}] - SC[\text{Facteur 2} \cap \text{Facteur 1}] - \dots$ (bleu)
- 3) Interaction : bleu/gris

- Indépendant de l'ordre des facteurs
- $SC[\text{Facteur}]$: effet principal
- **Recommandé pour groupes non équilibrés**
- Défaut pour STATISTICA

La fonction Anova (package car)

Anova(lm.result, type=2)

```
Response: Y
      Sum Sq Df F value    Pr(>F)
F1      350.86  1  718.85 < 2.2e-16 ***
F2      111.08  1  227.57 2.167e-12 ***
F1:F2    103.27  1  211.57 4.238e-12 ***
Residuals    9.76 20
```

```
Response: Y
      Sum Sq Df F value    Pr(>F)
F2      111.08  1  227.57 2.167e-12 ***
F1      350.86  1  718.85 < 2.2e-16 ***
F2:F1    103.27  1  211.57 4.238e-12 ***
Residuals    9.76 20
```

Anova(lm.result, type=3)

```
Response: Y
      Sum Sq Df F value    Pr(>F)
(Intercept) 4810.3  1 9855.33 < 2.2e-16 ***
F1          345.5  1  707.77 < 2.2e-16 ***
F2           93.0  1  190.50 1.105e-11 ***
F1:F2       103.3  1  211.57 4.238e-12 ***
Residuals      9.8 20
```

```
Response: Y
      Sum Sq Df F value    Pr(>F)
(Intercept) 4810.3  1 9855.33 < 2.2e-16 ***
F2           93.0  1  190.50 1.105e-11 ***
F1          345.5  1  707.77 < 2.2e-16 ***
F2:F1       103.3  1  211.57 4.238e-12 ***
Residuals      9.8 20
```

Résultat de l'analyse : summary

- Résidus
- Paramètres estimés
 - Moyenne
 - Erreur type
 - Valeur du t
 - *p-value*
- Performances en termes d'explication (R^2 et R^2 ajusté)
- Significativité globale de l'analyse (F et *p-value*)

```
Call:
lm(formula = Y ~ AGE * GROUPE, data = df.out)

Residuals:
    Min       1Q   Median       3Q      Max
-29.2485  -0.1681  -0.0168   0.3838  13.5121

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -41.7191     5.9096  -7.060 2.76e-09 ***
AGE           0.6304     0.1203   5.240 2.52e-06 ***
GROUPE1      42.7323     5.9096   7.231 1.44e-09 ***
AGE:GROUPE1  -1.6308     0.1203 -13.556 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.975 on 56 degrees of freedom
Multiple R-squared:  0.9864,    Adjusted R-squared:  0.9857
F-statistic: 1356 on 3 and 56 DF,  p-value: < 2.2e-16
```

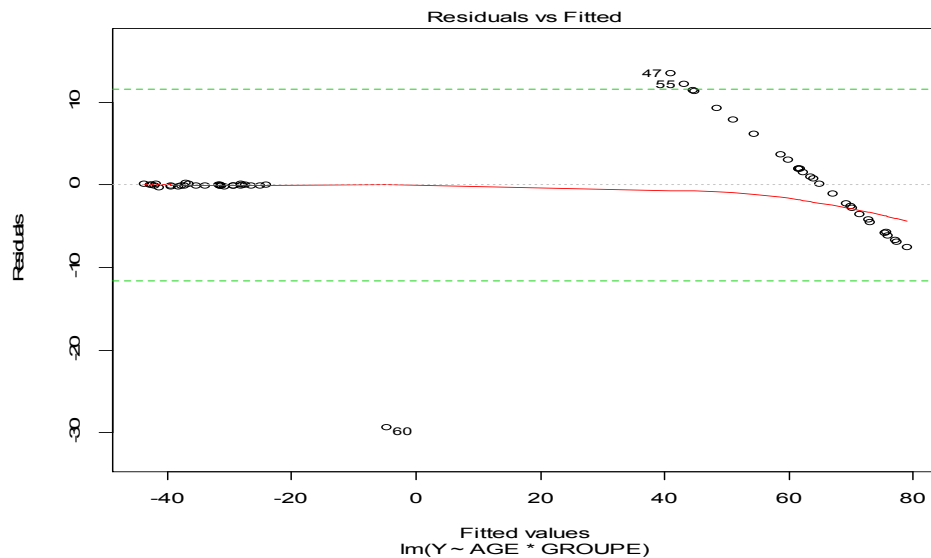
Interpréter le R^2

- Indice d'**ajustement** non de prédiction
- Variance expliquée par modèle / variance totale
 - R^2 (\sim ajustement à l'échantillon)
 - $\text{var}(\text{predict}(\text{res.lm})) / \text{var}(y)$
 - $\text{var}(y - \text{residuals}(\text{res})) / \text{var}(y)$
 - $1 - \text{var}(\text{residuals}(\text{res.lm})) / \text{var}(y)$
 - R^2 ajusté : R^2 corrigé des degrés de liberté (\sim ajustement à la population dont est issue l'échantillon)
- Qualité de l'ajustement
 - $R^2 < 0.5$: question sur la pertinence du modèle
 - R^2 ajusté > 0.5 : voir les autres paramètres de qualité

Analyse des résidus

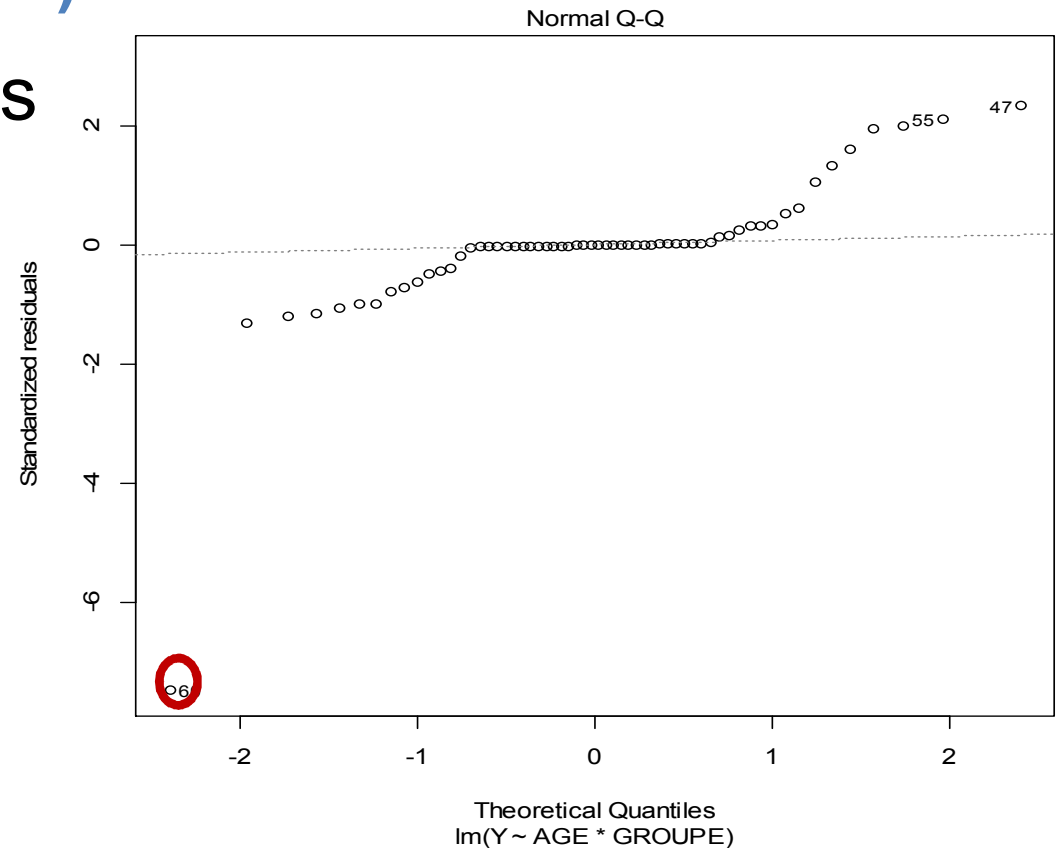
Résidus = f(valeurs ajustées)

- Points aberrants, mauvais modèle...



```
plot( lm.out, which=1)  
abline( h=2 * sd( residuals( lm.out)), col=3, lty=2)  
abline( h=-2 * sd( residuals( lm.out)), col=3, lty=2)
```

Normalité des résidus



```
plot( lm.out, which=2)  
ou fonction qqplot
```

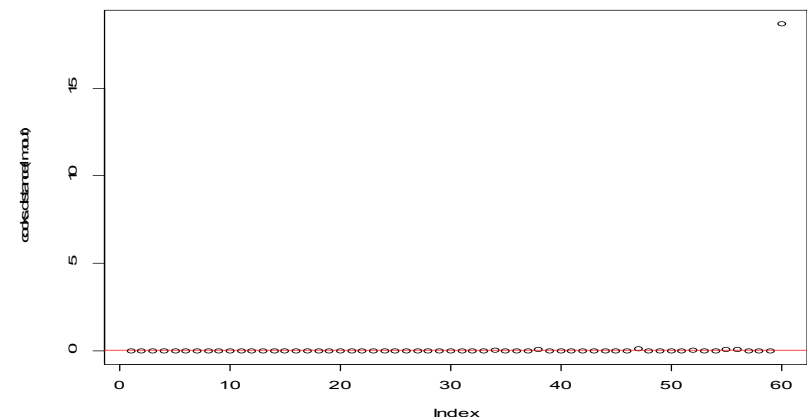

Influence

Colinéarité des prédicteurs

- Fonction `vif`
- > 3 à 5 : pb de colinéarité

Influence des points sur l'estimation des paramètres

- Distance de Cook : `cooks.distance`



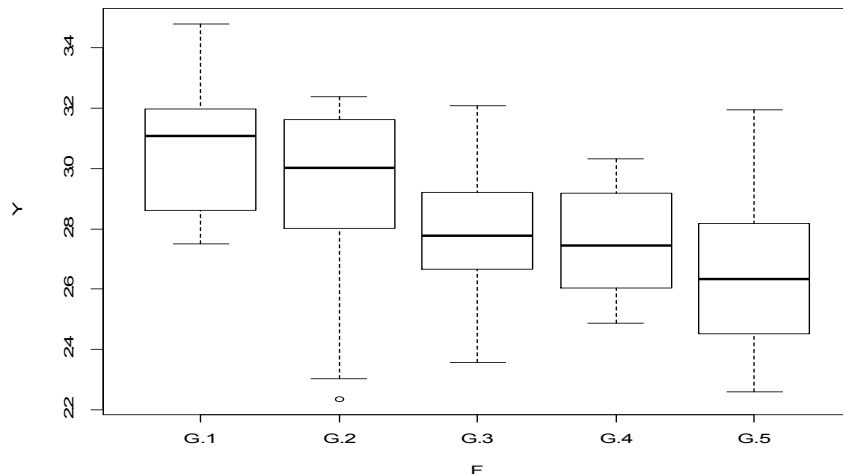
```
plot( cooks.distance( lm.out))  
abline( h=4 / (N - k), col=2)
```

```
plot( lm.out, which=4)
```

Conduire une analyse linéaire (2)

Question : Décroissance significative régulière fonction stade pathologique ?

`plot(Y ~ F, donnees)`



```
res.lm <- lm( Y ~ F, donnees)
anova( res.lm)
```

```
Response: Y
      Df Sum Sq Mean Sq F value    Pr(>F)
F       4  1.3214   0.33035   4.7968 0.002163 **
Residuals 55  3.7878   0.06887
```

`summary(res.lm)`

Résultats par défaut

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.08059     0.07576  40.664  < 2e-16 ***
F2            -0.17126     0.10714  -1.599  0.115650
F3            -0.31373     0.10714  -2.928  0.004950 **
F4            -0.33104     0.10714  -3.090  0.003139 **
F5            -0.42551     0.10714  -3.972  0.000209 ***
---
```

Que signifient F2, ... ?

Comparaisons planifiées (1)

- Les paramètres affichés sont les valeurs estimées par les comparaisons planifiées (tests *a priori*) définies au moment de l'analyse.
- Dans R, le choix des comparaisons planifiées se traduit par le choix du contraste appliqué au facteur considéré.

```
summary( res.lm)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.08059    0.07576  40.664  < 2e-16 ***
F2           -0.17126    0.10714  -1.599  0.115650
F3           -0.31373    0.10714  -2.928  0.004950 **
F4           -0.33104    0.10714  -3.090  0.003139 **
F5           -0.42551    0.10714  -3.972  0.000209 ***
---
```

=> *Intercept* : ordonnée à l'origine

=> Ligne 2 = comparaison planifiée n°1, ligne 3 = n°2, ...

Changer les contrastes dans R

- Session
 - Défaut R :
 - `options(contrasts=c("contr.treatment", "contr.poly"))`
 - Changer les défauts
 - `options(contrasts=c("contr.sum", "contr.poly"))`
- Pour une analyse : argument **contrasts**
 - de la fonction d'analyse (e.g., **lm**)
 - de la fonction **update** (`update(res.analyse, args...)`)
 - valeur liste nommée : `nom.facteur=contraste`
 contraste : chaînes de caractères, fonction, matrice

contr.treatment

Défaut pour les facteurs non ordonnés

Intercept : moyenne du 1^{er} niveau

F2 : différence entre 2^{ème} et 1^{er} niveau

F3 : différence entre 3^{ème} et 1^{er} niveau

...

summary(res.lm)

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.08059    0.07576  40.664  < 2e-16 ***
F2           -0.17126    0.10714  -1.599  0.115650
F3           -0.31373    0.10714  -2.928  0.004950 **
F4           -0.33104    0.10714  -3.090  0.003139 **
F5           -0.42551    0.10714  -3.972  0.000209 ***
---
```

Les comparaisons planifiées ne répondent pas vraiment à la question

contr.sum

```
summary( update( res.lm, contrasts=list( F=contr.sum)))
```

Intercept : grande moyenne

F1 : différence 1^{er} niveau à la moyenne des autres niveaux

F2 : différence 2^{ème} niveau à la moyenne des autres niveaux

...

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.83228	0.03388	83.599	< 2e-16	***
F1	0.24831	0.06776	3.665	0.000559	***
F2	0.07704	0.06776	1.137	0.260454	
F3	-0.06542	0.06776	-0.965	0.338537	
F4	-0.08273	0.06776	-1.221	0.227314	

Les comparaisons planifiées ne répondent pas à la question

contr.sdif (package MASS)

```
summary( update( res.lm, contrasts=list( F=contr.sdif)))
```

Hypothèse : facteurs ordonnés

Intercept : grande moyenne

Autres paramètres : différence entre deux niveaux consécutifs

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.83228	0.03388	83.599	<2e-16 ***
F2-1	-0.17126	0.10714	-1.599	0.116
F3-2	-0.14246	0.10714	-1.330	0.189
F4-3	-0.01731	0.10714	-0.162	0.872
F5-4	-0.09447	0.10714	-0.882	0.382

Pas de différences significatives entre niveaux consécutifs

Conclusion

- Effet significatif du facteur F (*cf.* anova)
- Décroissance régulière (différences négatives mais non significatives)

contr.poly

```
summary( update( res.lm, contrasts=list( F=contr.poly)))
```

Hypothèse : facteurs ordonnés (**contraste par défaut pour ce type**)

Intercept : grande moyenne

.L : tendance linéaire

.Q : tendance quadratique

.C : tendance cubique (?)

^n : tendance degré n (??)

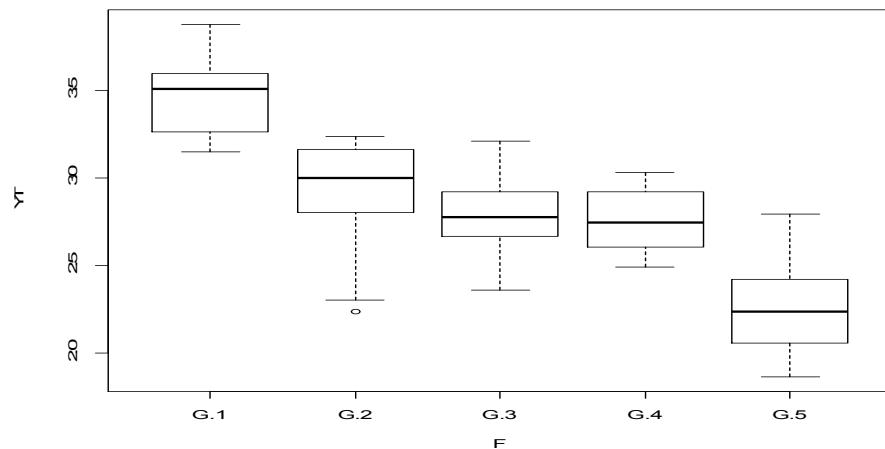
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.83228	0.03388	83.599	< 2e-16	***
F.L	-0.31964	0.07576	-4.219	9.24e-05	***
F.Q	0.07449	0.07576	0.983	0.330	
F.C	-0.03351	0.07576	-0.442	0.660	
F^4	-0.03570	0.07576	-0.471	0.639	

Conclusion

- Effet significatif du facteur F (*cf.* anova)
- Décroissance à tendance linéaire si points régulièrement espacés :
différence entre 2 niveaux consécutifs ~ constante
- Si groupes équilibrés, contrastes orthogonaux

Conduire une analyse linéaire (3)

Question : Décroissance significative régulière fonction stade pathologique ?



summary(res.lm)

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.83228    0.03388  83.599  < 2e-16 ***
F2-1         -0.57126    0.10714  -5.332  1.88e-06 ***
F3-2         -0.14246    0.10714  -1.330    0.189
F4-3         -0.01731    0.10714  -0.162    0.872
F5-4         -0.49447    0.10714  -4.615  2.39e-05 ***
```

Entre niveaux 2 et 4 plateau ou décroissance ?

```
res.lm <- lm( YT ~ F, donnees,
              contrasts=list( F=contr.sdif))
anova( res.lm)
```

```
Response: YT
          Df Sum Sq Mean Sq F value    Pr(>F)
F           4 9.2463  2.31157   33.565 3.571e-14 ***
Residuals 55 3.7878  0.06887
```


Résultats de l'analyse

- Calcul du contraste :

```
contr.mat <- ginv( mat.cp)
```

```
colnames( contr.mat) <- c( "(2-1)", "(4-2)", "(3-2=4-3)", "(5-4)")
```

```
lm.maT <- update( res.lm, contrasts=list( F=contr.mat))
```

```
summary( lm.maT)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.83228	0.03388	83.599	< 2e-16 ***
F(2-1)	-0.57126	0.10714	-5.332	1.88e-06 ***
F(4-2)	-0.15977	0.10714	-1.491	0.142
F(3-2=4-3)	-0.12515	0.18556	-0.674	0.503
F(5-4)	-0.49447	0.10714	-4.615	2.39e-05 ***

Conclusion : Compte-tenu des données, hypothèse du plateau plausible

Les contrastes *post-hoc*

- Testent l'**hypothèse nulle conjointe** :
parametre.1 = 0 ET parametre.2 = 0 ET ...
- N'ont de sens que si l'analyse montre un effet significatif lié au facteur
- À utiliser si paramètres testés conjointement sont liés :
 - Plus de paramètres que de degrés de liberté
 - Contrastes non orthogonaux
 - Groupes équilibrés : `contr.treatment`, `contr.sum`, `contr.sdif`
 - Groupes non équilibrés : en général, quelque soit le contraste

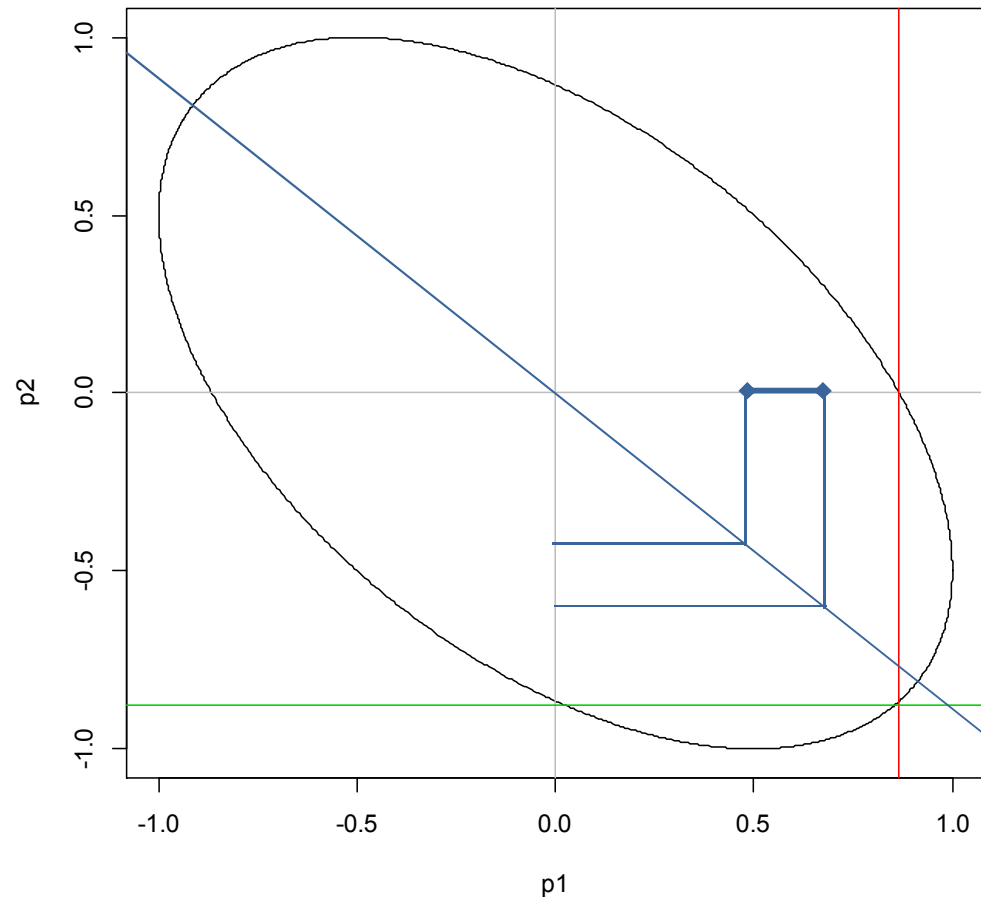
Contrastes non orthogonaux

Les comparaisons planifiées (i.e., la fonction `summary`) testent isolément les paramètres => déclareront significatives les valeurs de :

- p_1 à droite de la barre verticale rouge
- p_2 en dessous de la barre horizontale verte.

En réalité, compte-tenu de la corrélation des paramètres estimés (fonction `vcov`), certaines combinaisons de (p_1 , p_2) ne sont pas significatives dans ces régions.

=> Nécessité de la correction



Tests *post-hoc* (1)

- `LinearHypothesis` (package `car`) : conjonction d'hypothèses
 - `hypothesis.matrix` : combinaisons linéaires des paramètres estimés (matrice ou description symbolique)
 - `rhs` : valeur de la partie droite

```
linearHypothesis( update( res.lm,  
                        contrasts=list( F=contr.treatment)), "F4 = F2" )
```

```
Res.Df    RSS Df Sum of Sq    F Pr(>F)  
1      56 3.9409  
2      55 3.7878  1  0.15317 2.224 0.1416
```

Ici, qu'un test => pas besoin de correction des comparaisons multiples (*cf.* fonction `p.adjust`)

Pas toujours simple car dépend du contraste et des tests !

Tests *post-hoc* (2)

- fonction `glht` (package `multcomp`)
 - `linfct` : matrice, description symbolique ou fonction `mcp`
- package `lsmeans`
 - `specs` : formule décrivant le test
- `summary(glht(res.lm, linfct=mcp(F="Tukey")))`
- `lsmeans(res.lm, pairwise ~ F)`

Mêmes résultats :

- pas de différence significative
- perte de puissance

contrast	estimate	SE	df	t.ratio	p.value
G.1 - G.2	0.57126250	0.107136	55	5.332	<.0001
G.1 - G.3	0.71372518	0.107136	55	6.662	<.0001
G.1 - G.4	0.73103623	0.107136	55	6.823	<.0001
G.1 - G.5	1.22550805	0.107136	55	11.439	<.0001
G.2 - G.3	0.14246268	0.107136	55	1.330	0.6741
G.2 - G.4	0.15977373	0.107136	55	1.491	0.5724
G.2 - G.5	0.65424555	0.107136	55	6.107	<.0001
G.3 - G.4	0.01731104	0.107136	55	0.162	0.9998
G.3 - G.5	0.51178287	0.107136	55	4.777	0.0001
G.4 - G.5	0.49447182	0.107136	55	4.615	0.0002

Tests *post-hoc* complexes : multcomp

- Test : hypothèse de décroissance

```
res.glht <- glht( res.lm,  
  linfct=mcp( F=c( "G.2-G.1 <= 0", "G.3-G.2 <= 0",  
    "G.4-G.3 <= 0", "G.5-G.4 <= 0"))))
```

```
res.glht <- glht( res.lm,  
  linfct=mcp( F=rbind( "G.2-G.1"=c( -1, 1, 0, 0, 0),  
    "G.3-G.2"=c( 0, -1, 1, 0, 0),  
    "G.4-G.3"=c( 0, 0, -1, 1, 0), "G.5-G.4"=c( 0, 0, 0, -1, 1))),  
  alternative="greater")# statistique de test > valeur critique
```

```
summary( res.glht, test=adjusted( "none" ))
```


Tests *post-hoc* complexes : lsmeans

- Test : hypothèse de décroissance

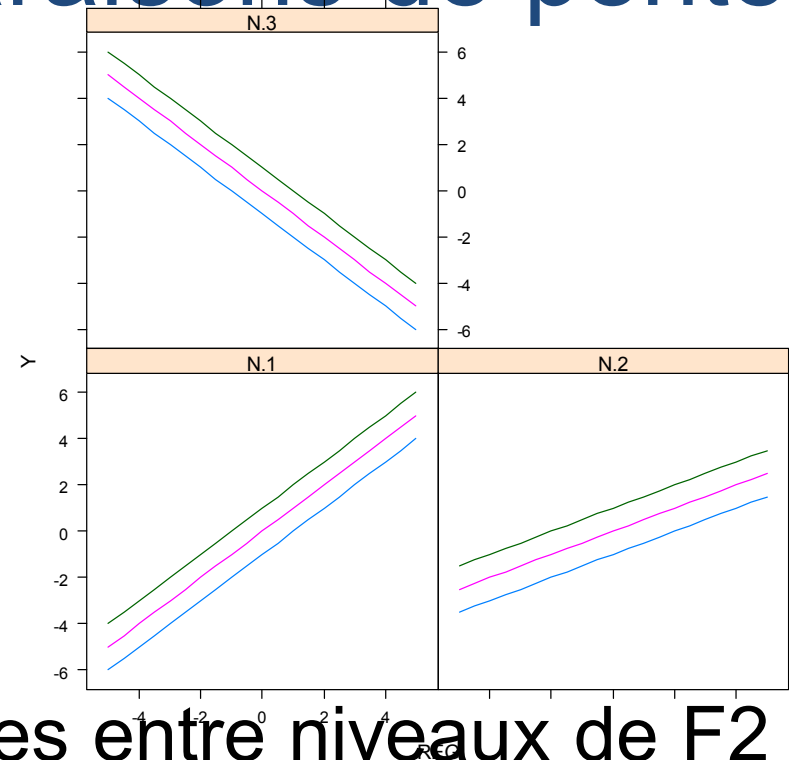
```
essai.lsmc <- function( levs, ...) {# levs : vecteur des noms des niveaux
  M <- as.data.frame(
    cbind( "G.2-G.1"=c( -1, 1, 0, 0, 0),"G.3-G.2"=c( 0, -1, 1, 0, 0),
          "G.4-G.3"=c( 0, 0, -1, 1, 0), "G.5-G.4"=c( 0, 0, 0, -1, 1)))
  attr( M, "desc") <- "decroissance"
  attr( M, "adjust") <- "none"
  M
}
# statistique de test > valeur critique
summary( lsmeans( lm.sd, essai ~ F), side=">")
```

Tests *post-hoc* : comparaisons de pentes

Package `lsmeans`

- Modèle : $Y \sim \text{REG} * F1 * F2$

- F1 et F2 : facteurs
- REG : covariable



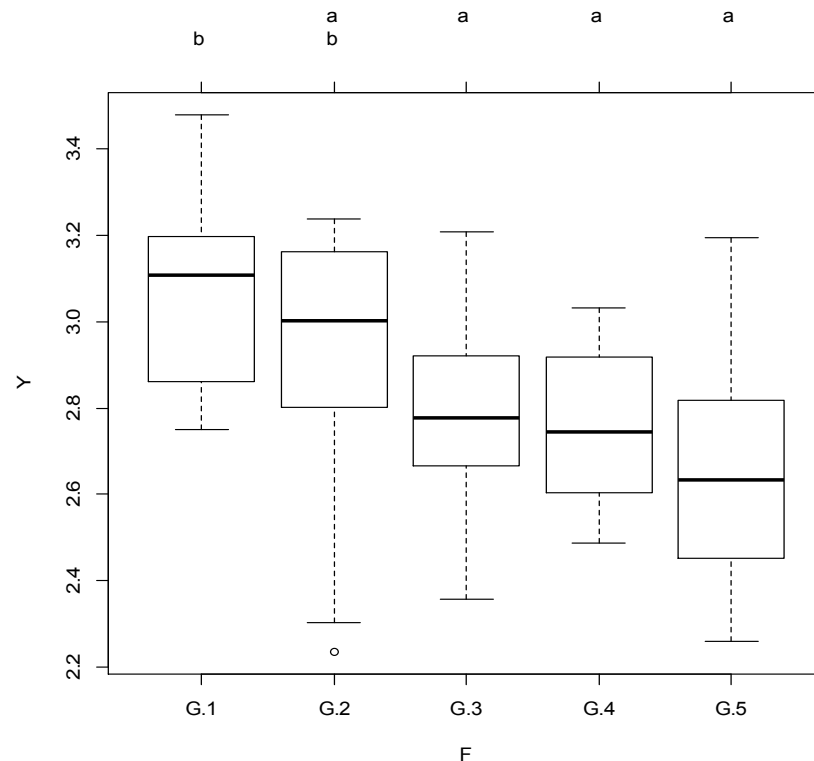
- Test du parallélisme des pentes entre niveaux de F2 pour chaque niveau de F1

```
lsmeans( res.lm, pairwise ~ F2 | F1, trend="REG")
```

```
lstrends( res.lm, pairwise ~ F2 | F1, var="REG" )
```

Tests *post-hoc* : groupes homogènes

Fonction `cld`



```
> cld( lsmeans( lm.sd, pairwise ~ F))  
F      lsmean      SE df lower.CL upper.CL .group  
G.5 2.655083 0.07575657 55 2.503264 2.806903 1  
G.4 2.749555 0.07575657 55 2.597736 2.901375 1  
G.3 2.766866 0.07575657 55 2.615047 2.918686 1  
G.2 2.909329 0.07575657 55 2.757509 3.061148 12  
G.1 3.080591 0.07575657 55 2.928772 3.232411 2  
  
Confidence level used: 0.95  
P value adjustment: tukey method for a family of 5 tests  
significance level used: alpha = 0.05
```

Package `multcomp`

```
plot( cld( glht( lm.sd, linfct=mcp( F="Tukey")))))
```

Package `lsmeans`

Contrastes de Tukey

- Comparaison deux à deux de tous les niveaux
 - N niveaux $\Rightarrow N(N - 1) / 2$ testsPlus de tests que de degrés de liberté
 - Les valeurs testées sont corrélées entre elles
- Implantation `lsmeans` : `pairwise`, `revpairwise`, `tukey`
- Implantation `multcomp` : `linfct=mcp(F="Tukey")`
- Permet de définir des classes homogènes de moyennes
Cf. fonction `cld` des packages `multcomp` et `lsmeans`

Contraste de Dunnett

- Comparaison des niveaux d'un facteur à un niveau de référence
 - Par exemple, test.1 vs. repos, test.2 vs. repos...
 - Paramètres corrélés à cause de la référence commune.
- Implantation `lsmeans` : `dunnett`, `trt.vs.ctrl`, `trt.vs.ctrl1`, `trt.vs.ctrlk`
- Implantation `multcomp` : `linfct=mcp(F="Dunnett")`
- Pourquoi les résultats sont-ils différents des comparaisons planifiées avec `contr.treatment` ?