

Analyse de la covariance

Pierre Gloaguen

22 novembre 2018

Objectifs

- ▶ Expliquer les variations d'une variable **quantitative**:
 - ▶ Un rendement, une abondance, un taux d'une substance. . .
- ▶ En fonctions d'autres variables:
 - ▶ Un fertilisant, une région, un apport chimique. . .

Avantages

- ▶ Formulation mathématique simple permettant de connaître ses propriétés.
- ▶ Bonne représentation (en première approximation) de nombreux phénomènes.

Cas d'étude: Rendement de maïs

- ▶ On souhaite expliquer le **rendement** de plants de maïs.
- ▶ On dispose de 288 parcelles (on a désormais enlevé 11 parcelles).
- ▶ Sur chaque parcelle, le maïs a un même *marqueur génétique*:
 - ▶ Soit un marqueur de type 1;
 - ▶ Soit un marqueur de type 2;
- ▶ Sur chaque parcelle, le maïs a une même *variété*:
 - ▶ Corn, European, Northern, Tropical.
- ▶ Sur chaque parcelle, on mesure différentes caractéristiques:
 - ▶ Le **rendement** de la parcelle;
 - ▶ La teneur moyenne en *huile* d'un grain de maïs;
 - ▶ La teneur moyenne en *protéine* d'un grain de maïs;
 - ▶ La teneur moyenne en *amidon* d'un grain de maïs;
 - ▶ Le nombre de degrés-jours moyen avant la *floraison* d'un plant de maïs;
 - ▶ Le nombre moyen de *feuilles* par plant de maïs;
- ▶ Quelles *variables explicatives* donnent des informations sur le **rendement**?

Principes du modèle linéaire

- ▶ Modèle mathématique décrivant le lien entre une variable explicative **quantitative** (le rendement) et des variables explicatives (la variété, la teneur en huile, . . .)
- ▶ Modèle décrit dans un cadre probabiliste décrivant l'aléa (part non prédite).

Principe d'application

1. Question biologique;
2. Ecriture du modèle;
3. Ajustement (estimation) du modèle grâce aux données;
4. Vérification de la validité des hypothèses faites dans le modèle;
5. Test de la pertinence du modèle linéaire par rapport à un modèle simple;
6. Test de la pertinence des différents éléments du modèle;
7. Critique du modèle;
8. Conclusion sur la question biologique.

Analyse de la covariance

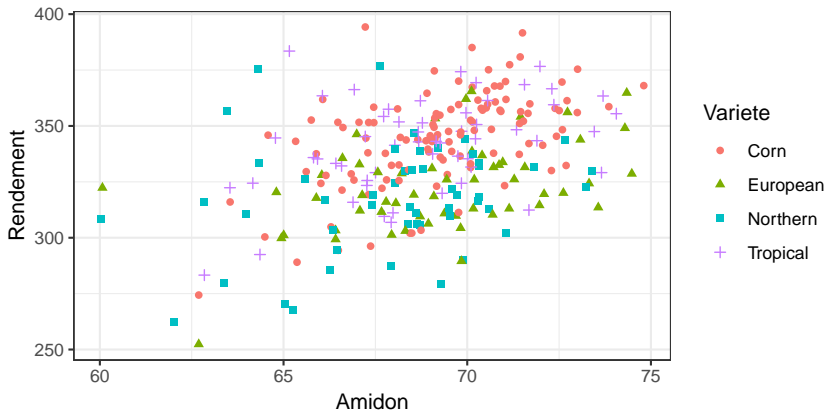
1) Question biologique

Question biologique: Le **rendement** d'une espèce peut-il être expliqué par sa *variété* et sa *teneur en amidon*?

- ▶ Le **rendement** est la variable à expliquer;
- ▶ La *variété* et sa *teneur en amidon* sont des variables explicatives. Une est *qualitative*, l'autre est *quantitative*.
- ▶ **Cadre de l'ANCOVA:** 1 **variable à expliquer**, quantitative, des *variable explicatives*, qualitatives et quantitatives.
- ▶ **Première étape:** Une approche descriptive.

Pour cet exemple, on a enlevé la variété Stiff Stalk

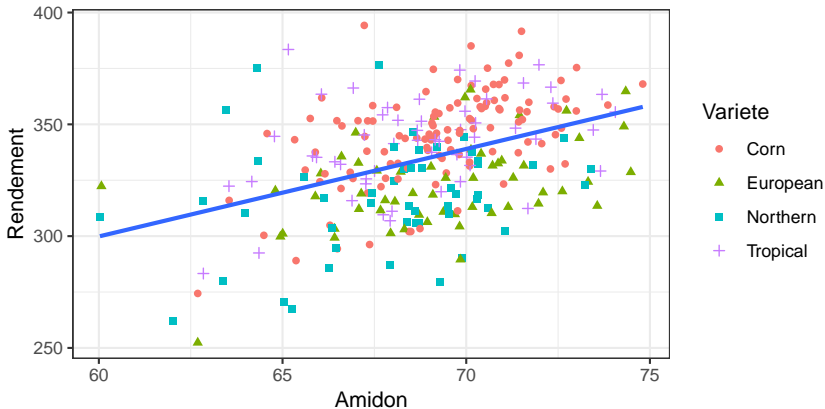
Visualisation des données



Regression simple Rendement en fonction du régresseur Amidon

On ajuste un modèle de régression simple:

$$\underbrace{\text{Rendement}}_{Y_k} = \beta_0 + \beta_1 \underbrace{\text{Amidon}}_{x_k} + E_k, \quad E_k \sim \mathcal{N}(0, \sigma^2)$$

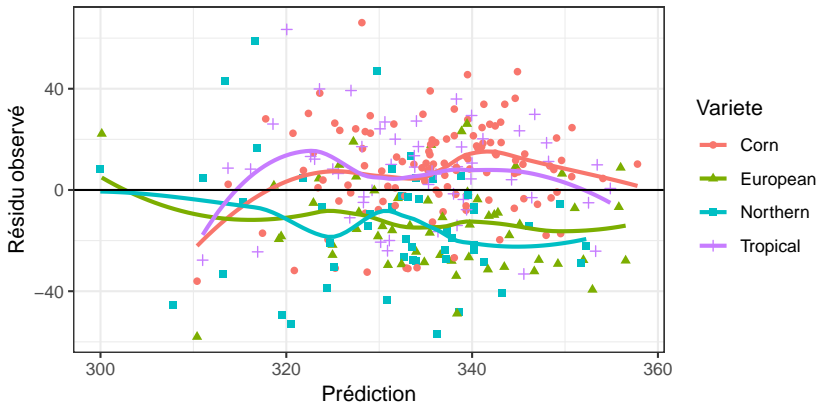


Régression simple:

On ajuste un modèle de régression simple:

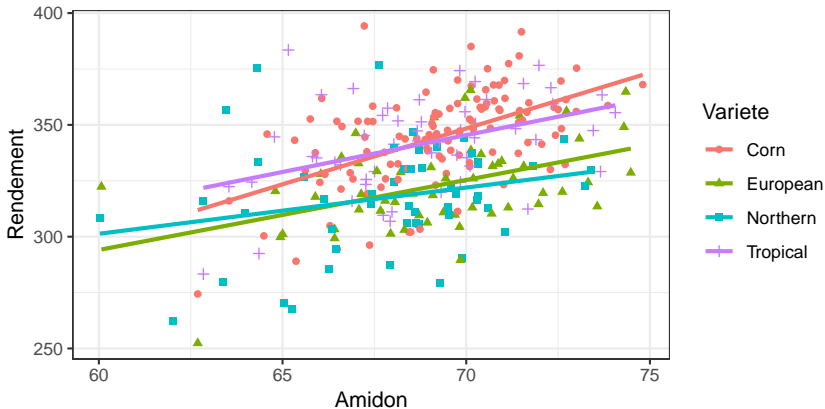
$$\underbrace{Y_k}_{\text{Rendement}} = \beta_0 + \beta_1 \underbrace{x_k}_{\text{Amidon}} + E_k, \quad E_k \sim \mathcal{N}(0, \sigma^2)$$

On voit cependant des résidus structurés par Variete!



Une régression linéaire simple par variété:

On ajuste 4 modèles de régression simple, un par Variete.



L'ANCOVA permet d'écrire ce modèle *globalement*, et de tester sa pertinence.

II) Ecriture du modèle (régulier)

Notations

On a $n = 277$ observations. Le facteur Variete a $I = 4$ niveaux avec: Corn ($i=1$), European ($i=2$), Northern ($i=3$), Tropical ($i=4$).

Pour chaque niveau i , on dispose de n_i observations avec $n_1 = 117, n_2 = 56, n_3 = 50, n_4 = 54$.

On note y_{ik} le rendement de la k -ième parcelle pour la Variete i et x_{ik} sa teneur en amidon ($1 \leq k \leq n_i$).

Modèle

On suppose que y_{ik} est la réalisation d'une V.A. Y_{ik} telle que:

$$Y_{ik} = \beta_{0,i} + \beta_{1,i}x_{ik} + E_{ik}, \quad 1 \leq i \leq I, \quad 1 \leq k \leq n_i$$

- ▶ $\beta_{0,i}$ est l'ordonnée à l'origine de la relation rendement/amidon pour la Variete i ;
- ▶ $\beta_{1,i}$ est la pente de la relation rendement/amidon pour la Variete i ;
- ▶ E_{ik} est le résidu (aléa) associé à l'observation Y_{ik} . $E_{ik} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

II) Ecriture du modèle (singulier)

Notations

On a $n = 277$ observations. Le facteur Variete a $I = 4$ niveaux avec: Corn ($i=1$), European ($i=2$), Northern ($i=3$), Tropical ($i=4$).

Pour chaque niveau i , on dispose de n_i observations avec $n_1 = 117, n_2 = 56, n_3 = 50, n_4 = 54$.

On note y_{ik} le rendement de la k -ième parcelle pour la Variete i et x_{ik} sa teneur en amidon ($1 \leq k \leq n_i$).

Modèle

On suppose que y_{ik} est la réalisation d'une V.A. Y_{ik} telle que:

$$Y_{ik} = (\mu + \alpha_i) + (\beta + \gamma_i)x_{ik} + E_{ik}, \quad 1 \leq i \leq I, \quad 1 \leq k \leq n_i$$

- ▶ μ et β sont les paramètres de la droite de référence. :
- ▶ α_i la modification de l'ordonnée à l'origine due à la variété i ;
- ▶ γ_i est la modification de la pente due à la variété i (**interaction**);
- ▶ E_{ik} est le résidu (aléa) associé à l'observation Y_{ik} . $E_{ik} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

III) Estimateurs

- ▶ **Estimateurs:** Variables aléatoires Pour chaque variété, on reprend les estimateurs de la régression simple:

$$\hat{\beta}_{1,i} = \frac{\sum_{k=1}^{n_i} (x_{ik} - x_{i\bullet})(Y_{ik} - Y_{i\bullet})}{\sum_{k=1}^n (x_{ik} - x_{i\bullet})^2}, \quad \hat{\beta}_{0,i} = Y_{i\bullet} - \hat{\beta}_{1,i} x_{i\bullet}$$

- ▶ **Estimations:** Réalisation sur les données:

$$\hat{\beta}_{1,i}^{obs} = \frac{\sum_{k=1}^{n_i} (x_{ik} - x_{i\bullet})(y_{ik} - y_{i\bullet})}{\sum_{k=1}^n (x_{ik} - x_{i\bullet})^2}, \quad \hat{\beta}_{0,i} = y_{i\bullet} - \hat{\beta}_{1,i}^{obs} x_{i\bullet}$$

Prédicteurs

- ▶ **Prédicteur** Le rendement prédit est:

$$\hat{Y}_{ik} = \hat{\beta}_{0,i} + \hat{\beta}_{1,i} X_{ik}$$

- ▶ **Prédiction** Réalisation sur les données:

$$\hat{y}_{ik} = \hat{\beta}_{0,i}^{obs} + \hat{\beta}_{1,i}^{obs} X_{ik}$$

III) Ajustement du modèle singulier

Il y a trop de paramètres de moyennes! ($2I + 2$). Il faut fixer 2 contraintes ($p + 1$ en ANCOVA à 1 facteur et plusieurs régresseurs).

On doit **poser 2 contraintes**, typiquement:

$$\alpha_1 = 0, \gamma_1 = 0$$

Les 1er niveau de facteur donne alors la **droite de référence**.

Les estimateurs des $\mu, \alpha_i, \beta, \gamma_i$ vont **dépendre** de la contrainte!

Pour la contrainte ci dessus, on a:

$$\hat{\mu} = \hat{\beta}_{0,1}, \hat{\alpha}_i = \hat{\beta}_{0,i} - \hat{\beta}_{0,1}, \hat{\beta} = \hat{\beta}_{1,1}, \hat{\gamma}_i = \hat{\beta}_{1,i} - \hat{\beta}_{1,1}$$

Le prédicteur est alors

$$\hat{Y}_{ik} = \hat{\mu} + \hat{\alpha}_i + (\hat{\beta} + \hat{\gamma}_i)x_{ik} = \hat{\beta}_{0,i} + \hat{\beta}_{1,i}x_{ik}$$

Le prédicteur **ne dépend pas** de la contrainte.

Estimateur et estimation de la variance σ^2

Résidus observés

$$\hat{e}_{ik} = y_{ik} - \hat{y}_{ik}, \quad 1 \leq i \leq I, \quad 1 \leq k \leq n_{ij}$$

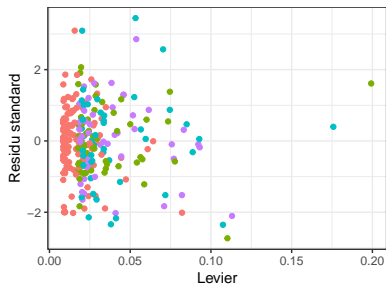
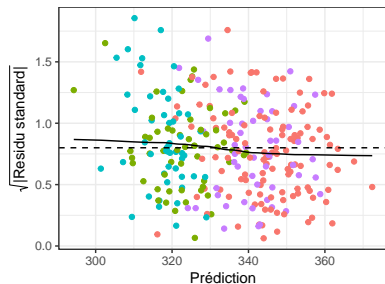
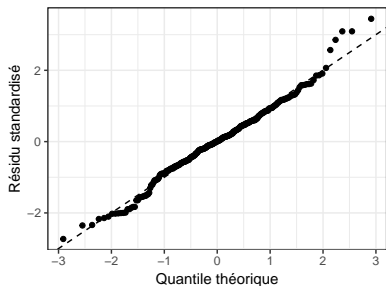
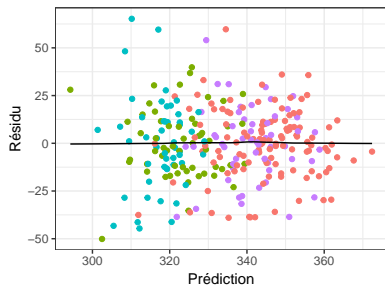
Estimateur

$$S^2 = \frac{\sum_{i=1}^I \sum_{k=1}^{n_i} (Y_{ik} - \hat{Y}_{ik})^2}{n - 2I}$$

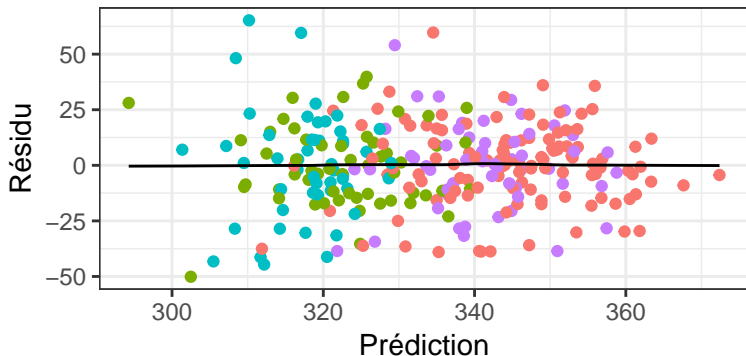
Estimation

$$\hat{\sigma}_{obs}^2 = \frac{\sum_{i=1}^I \sum_{k=1}^{n_i} (y_{ik} - \hat{y}_{ik})^2}{n - 2I} = \frac{\sum_{i=1}^I \sum_{k=1}^{n_i} \hat{e}_{ik}^2}{n - 2I} \stackrel{ici}{=} 378.7$$

IV) Validité des hypothèses



Distribution identique, espérance constante et nulle

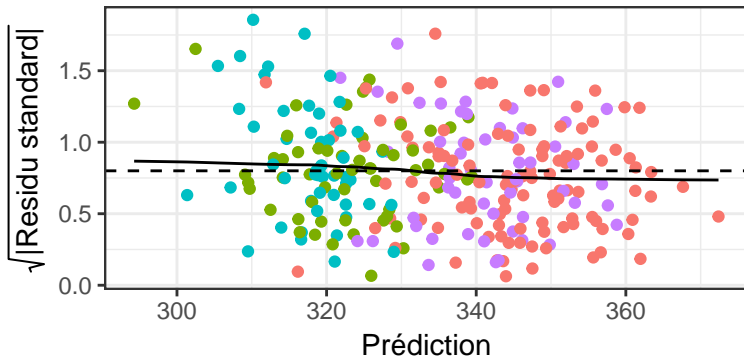


Ce qu'on regarde: Les résidus observés $\hat{\epsilon}_{ijk}$ en fonction des prédictions \hat{y}_{ijk} (équivalent à regarder en fonction des croisements de niveaux).

Ce qu'on voit: La distribution des résidus semble comparable dans toutes les variétés. L'espérance ne semble pas être structuré par la prédiction.

Ce qu'on conclut: L'hypothèse de distribution identique des résidus semble valable.

Distribution identique, variance constante

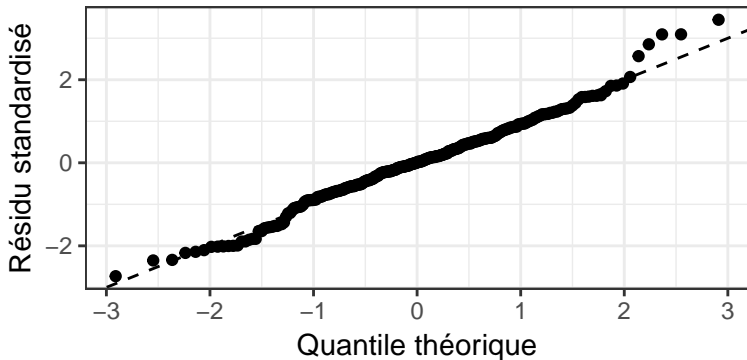


Ce qu'on regarde: La racine de la valeur absolue des résidus (standardisés) observés en fonction des prédictions \hat{y}_k (équivalent à regarder en fonction des croisements de niveaux).

Ce qu'on voit: La racine de la valeur absolue des résidus semble comparable dans toutes les variétés (autour de 0.8). La variance ne semble pas être structuré par la prédiction.

Ce qu'on conclut: L'hypothèse de variance identique des résidus semble valable.

Distribution normale

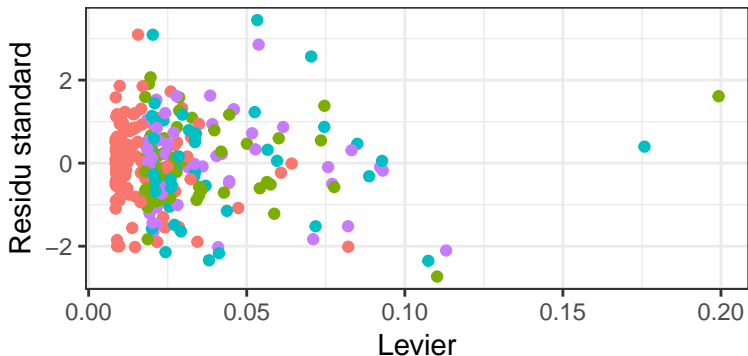


Ce qu'on regarde: La valeur des quantiles empiriques des résidus standardisés en fonction de la valeur quantiles théoriques d'une loi normale $\mathcal{N}(0, 1)$.

Ce qu'on voit: Les points sont globalement alignés sur la droite $y = x$.

Ce qu'on conclut: On peut valider l'hypothèse de distribution normale des résidus.

Résidus par variété

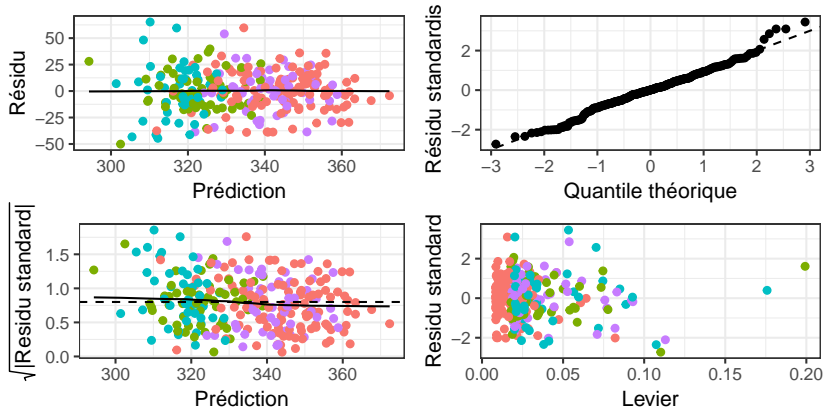


Ce qu'on regarde: Résidu standardisé en fonction du levier pour chaque observation.

Ce qu'on voit: Deux points ont un levier un peu conséquent. Mais leur distance de Cook n'est pas supérieure à 0.5.

Ce qu'on conclut: IL n'y a pas de point aberrant ou trop influent.

4 graphes et un oeil fin



Rien dans ces quatre graphiques ne permet de contredire les hypothèses faites sur les résidus.

V) Test du modèle

On veut tester si notre modèle impliquant Variete et Amidon du maïs explique mieux le **rendement** qu'un modèle simple, où le **rendement** est constant.

Hypothèses du test

On teste:

$$\begin{array}{ll} H_0 : & Y_{ik} = \mu + E_{ik} \quad \text{Modèle } M_0 \\ \text{contre } H_1 : & Y_{ik} = \mu + \alpha_i + (\beta + \gamma_i)x_{ik} + E_{ik}, \quad \text{Modèle } M_1 \end{array}$$

Pour tester cette hypothèse, on va décomposer la variabilité des données:

$$\sum_{i=1}^I \sum_{k=1}^{n_i} (Y_{ik} - \bar{Y})^2 \stackrel{SCT}{=} \sum_{i=1}^I \sum_{k=1}^{n_i} (\hat{Y}_{ik} - \bar{Y})^2 \stackrel{SCM}{=} \sum_{i=1}^I \sum_{k=1}^{n_i} (Y_{ik} - \hat{Y}_{ik})^2 \stackrel{SCR}{=}$$

Somme	ddl	Réalisation
$SCT:$	$n - 1$	$SCT_{obs} = \sum_{i,k} (y_{ik} - \bar{y})^2$
$SCM:$	$2I - 1$	$SCM_{obs} = \sum_{i,k} (\hat{y}_{ik} - \bar{y})^2$
$SCR:$	$n - 2I$	$SCR_{obs} = \sum_{i,k} (y_{ik} - \hat{y}_{ik})^2$

Test du modèle

Hypothèses du test

$$\begin{array}{ll} H_0 : & Y_{ik} = \mu + E_{ik} & \text{Modèle } M_0 \\ \text{contre } H_1 : & Y_{ik} = \mu + \alpha_i + (\beta + \gamma_i)x_{ik} + E_{ik}, & \text{Modèle } M_1 \end{array}$$

Statistique de test

On considère la statistique de test

$$F = \frac{SCM/ddl(SCM)}{SCR/ddl(SCR)}$$

Si H_0 est vraie, alors $F \stackrel{H_0}{\sim} \text{Fisher}(ddl(SCM), ddl(SCR))$.

Sur les données on observe

$$f_{obs} = \frac{SCM_{obs}/ddl(SCM)}{SCR_{obs}/ddl(SCR)}.$$

On rejette H_0 au risque de première espèce α si

$$\overbrace{\mathbb{P}(F > f_{obs})}^{\text{p-valeur}} < \alpha$$

Table d'analyse de la variance

<i>ddl</i>	Somme	<i>ddl</i>	Somme	Stat. test.	p-valeur
<i>ddl(SCT)</i>	SCT_{obs}				
<i>ddl(SCR)</i>	SCR_{obs}	<i>ddl(SCM)</i>	SCM_{obs}	f_{obs}	$\mathbb{P}(F > f_{obs})$

Analysis of Variance Table

Model 1: Rendement ~ 1

Model 2: Rendement ~ Variete * Amidon

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	276	163064				
2	269	101859	7	61206	23.091	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Conclusion du test: On rejette H_0 et on conclut que le modèle d'ANCOVA explique mieux les données qu'un modèle où le rendement est constant.

REMARQUE: $\hat{\sigma}^2 = S^2 = SCR/ddl(SCR)$

VI) Tests des effets

Chacun des effets dans le modèle est-il indispensable?

- ▶ Effet d'interaction (effet de la Variete sur la droite de régression)?
- ▶ Effet principal de la Variete?
- ▶ Effet principal de l' Amidon?

Test de l'effet d'interaction

On testera si l'ajout d'un effet d'interaction à un modèle avec les effets principaux apporte de l'information sur le rendement.

Test des effets principaux

2 manières de voir le problème?

- ▶ L'ajout de l'effet principal du facteur Variete est-il intéressant par rapport à un modèle constant? (Test de type I)
- ▶ L'ajout de l'effet principal du facteur Variete est-il intéressant par rapport à un modèle ayant l'effet Amidon? (Test de type II)

Notion de réduction

Notations

Modèle	Equation	SCM
M_μ	$Y_{ik} = \mu + E_{ik}$	SCM_μ
$M_{\mu,\alpha}$	$Y_{ik} = \mu + \alpha_j + E_{ik}$	$SCM_{\mu,\alpha}$
$M_{\mu,\beta}$	$Y_{ik} = \mu + \beta x_{ik} + E_{ik}$	$SCM_{\mu,\beta}$
$M_{\mu,\alpha,\beta}$	$Y_{ik} = \mu + \alpha_j + \beta x_{ik} + E_{ik}$	$SCM_{\mu,\alpha,\beta}$
$M_{comp} = M_{\mu,\alpha,\beta,\gamma}$	$Y_{ik} = \mu + \alpha_j + (\beta + \gamma_i)x_{ik} + E_{ik}$	$SCM_{\mu,\alpha,\beta,\gamma}$

Réduction

On appelle réduction associée α dans un modèle contenant μ la quantité:

$$R(\alpha|\mu) = SCM_{\mu,\alpha} - SCM_\mu$$

Son **degré de liberté** est le nombre de paramètres supplémentaires à estimé (après contraintes) dans $M_{\mu,\alpha}$ par rapport à M_μ , donc, $l - 1$ pour cet exemple.

On généralise cette définition pour tout modèle M_1 **emboîté** dans un modèle M_2 (noté $M_1 \subset M_2$), i.e. tel que M_2 consiste en l'**ajout** de paramètres à M_1 , en notant:

$$R(M_2|M_1) = SCM_2 - SCM_1$$

où M_2 est **au plus** le modèle complet M_{comp} .

Test associé à une réduction

Pour tester l'intérêt d'un modèle M_2 par rapport à un modèle M_1 ($M_1 \subset M_2$), on testera:

$$\begin{array}{l} H_0 : \text{ Le vrai modèle est } M_1 \\ \text{contre } H_1 : \text{ Le vrai modèle est } M_2 \end{array}$$

Statistique de test

On considère la statistique de test

$$F = \frac{R(M_2|M_1)/ddl(R(M_2|M_1))}{SCR/ddl(SCR)}$$

où SCR est la somme des carrés résiduelles du modèle **complet** M_{comp} .

Si H_0 est vraie, alors $F \overset{H_0}{\sim} Fisher(ddl(R(M_2|M_1)), ddl(SCR))$.

Sur les données on observe f_{obs} comme une réalisation de F .

On rejette H_0 au risque de première espèce α si $\overbrace{\mathbb{P}(F > f_{obs})}^{\text{p-valeur}} < \alpha$

Dans ce cas, on conclue que les effets de M_2 non présents dans M_1 apporte de l'information sur la variable explicative.

Tests de type I et II

Type I

Effet testé	H_0	H_1	Réduction
α	$Y_{ik} = \mu + E_{ik}$	$Y_{ik} = \mu + \alpha_i + E_{ik}$	$R(\alpha \mu)$
β	$Y_{ijk} = \mu + \alpha_i + E_{ik}$	$Y_{ik} = \mu + \alpha_i + \beta x_{ik} + E_{ik}$	$R(\beta \mu, \alpha)$
γ	$Y_{ik} = \mu + \alpha_i + \beta x_{ik} + E_{ik}$	$Y_{ijk} = \mu + \alpha_i + (\beta + \gamma_i)x_{ik} + E_{ik}$	$R(\gamma \mu, \alpha, \beta)$

Le test des effets principaux dépend de l'ordre d'entrée des facteurs!

Type II

Effet testé	H_0	H_1	Réduction
α	$Y_{ik} = \mu + \beta x_{ik} + E_{ik}$	$Y_{ik} = \mu + \alpha_i + \beta x_{ik} + E_{ik}$	$R(\alpha \mu, \beta)$
β	$Y_{ik} = \mu + \alpha_i + E_{ik}$	$Y_{ik} = \mu + \alpha_i + \beta x_{ik} + E_{ik}$	$R(\beta \mu, \alpha)$
γ	$Y_{ijk} = \mu + \alpha_i + \beta x_{ik} + E_{ik}$	$Y_{ik} = \mu + \alpha_i + (\beta + \gamma_i)x_{ik} + E_{ik}$	$R(\gamma \mu, \alpha, \beta)$

Test des effets Variete et Amidon (Type I)

Effet	<i>ddl</i>	Réduction	Ratio	Stat. test.	p-valeur
α	$ddl(R(\cdot))$	$R(\alpha \mu)$	F	f_{obs}	$\mathbb{P}(F > f_{obs})$
β	$ddl(R(\cdot))$	$R(\beta \mu, \alpha)$	F	f_{obs}	$\mathbb{P}(F > f_{obs})$
γ	$ddl(R(\cdot))$	$R(\gamma \mu, \alpha, \beta)$	F	f_{obs}	$\mathbb{P}(F > f_{obs})$
	$ddl(SCR)$	SCR	S^2		

Analysis of Variance Table

Response: Rendement

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Variete	3	36250	12083.3	31.9109	< 2.2e-16	***
Amidon	1	22737	22736.5	60.0453	1.905e-13	***
Variete:Amidon	3	2219	739.7	1.9536	0.1213	
Residuals	269	101859	378.7			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Au risque $\alpha = 0.01$

- ▶ Pas d'effet d'interaction;
- ▶ Effet principal de type I pour le facteur Variete;
- ▶ Effet du régresseur Amidon (à Variete connu);

Test des effets Variete et Amidon (Type II)

Effet	Réduction	ddl	Stat. test.	p-valeur
α	$R(\alpha \mu)$	$ddl(R(\cdot))$	f_{obs}	$\mathbb{P}(F > f_{obs})$
β	$R(\beta \mu, \alpha)$	$ddl(R(\cdot))$	f_{obs}	$\mathbb{P}(F > f_{obs})$
γ	$R(\gamma \mu, \alpha, \beta)$	$ddl(R(\cdot))$	f_{obs}	$\mathbb{P}(F > f_{obs})$
	SCR	$ddl(SCR)$		

Anova Table (Type II tests)

Response: Rendement

	Sum Sq	Df	F value	Pr(>F)	
Variete	30008	3	26.4166	5.239e-15	***
Amidon	22737	1	60.0453	1.905e-13	***
Variete:Amidon	2219	3	1.9536	0.1213	
Residuals	101859	269			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Au risque $\alpha = 0.01$

- ▶ Pas d'effet d'interaction;
- ▶ Effet principal de type II pour le facteur Variete;
- ▶ Effet du régresseur Amidon (à Variete connu);

VI) Test sur les paramètres de moyenne.

Hypothèses de test

Pour chaque paramètre de moyenne (**une fois les contraintes posées!**) on teste:

$$\begin{array}{l} H_0 : \text{ paramètre} = 0 \\ \text{contre } H_1 : \text{ paramètre} \neq 0 \end{array}$$

Ce test doit être interprété avec précaution car il dépend de la contrainte

Statistique de test

Exemple pour α_i , on utilise $T = \frac{\hat{\alpha}_i}{\sqrt{\widehat{\mathbb{V}}[\hat{\alpha}_i]}}$

où $\widehat{\mathbb{V}}[\hat{\alpha}_i]$ est l'estimateur de la variance de $\hat{\alpha}_i$ (calculé grâce à la réalisation de S^2).

Si H_0 est vraie, $T \stackrel{H_0}{\sim} \text{Student}(ddl(SCR))$.

On observe la réalisation t_{obs} de T . On rejette H_0 au risque α si:

$$\mathbb{P}(T < t_{obs}) < \alpha/2 \text{ ou } \mathbb{P}(T > t_{obs}) < \alpha/2$$

Ou, de manière équivalente: $\overbrace{2\mathbb{P}(T > |t_{obs}|)}^{\text{p-valeur dans R}} < \alpha$

Estimations du modèle

Call:

```
lm(formula = formule_ancov, data = donnees)
```

Residuals:

Min	1Q	Median	3Q	Max
-50.120	-11.451	0.084	11.296	65.223

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.178	55.645	-0.021	0.9831
VarieteEuropean	107.368	83.853	1.280	0.2015
VarieteNorthern	178.495	86.183	2.071	0.0393 *
VarieteTropical	115.903	88.904	1.304	0.1935
Amidon	4.994	0.803	6.219	1.9e-09 ***
VarieteEuropean:Amidon	-1.863	1.210	-1.539	0.1249
VarieteNorthern:Amidon	-2.927	1.257	-2.329	0.0206 *
VarieteTropical:Amidon	-1.698	1.288	-1.318	0.1885

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.46 on 269 degrees of freedom

Multiple R-squared: 0.3753, Adjusted R-squared: 0.3591

F-statistic: 23.09 on 7 and 269 DF, p-value: < 2.2e-16

VII) Critique du modèle

Comme attendu après ajustement des modèles précédents, un modèle comprenant à la fois la variété et la teneur en amidon explique mieux les données que des modèles plus simples. Par contre, il ne semble pas que la variété joue sur la corrélation Rendement/amidon.

Le modèle a toujours un faible pouvoir prédictif.

VIII) Conclusion biologique

La connaissance de la variété semble donner une information sur le rendement,
De même, il semble y avoir une corrélation positive entre amidon et rendement.