

Nom de la SAE	Expliquer ou prédire une variable quantitative à partir de plusieurs facteurs		semestre / Période	Semestre 4
volume horaire consacré par l'étudiant	avec enseignant	6h	en autonomie	6h
coéquipiers :	Melissa Djenadi		Zara Zulfikar	
	Safae Sebay			

Sujet spécifique	Sur des données d'immunologie, étudier le lien entre les signaux produits par les cellules dendritiques et les signaux produits par les lymphocytes T, ainsi que les perturbateurs immunologiques.
Objectifs	Voir s'il existe un lien entre l'émission de IL12p70 (signal produit par des cellules dendritiques) et l'émission d'IFN γ (signal produit par des lymphocytes T) et expliquer la production d'IFN γ en fonction des perturbateurs immunologiques.
Livrables	Rapport

Bilan de la SAE

(reproduire le tableau autant de fois que de compétences mobilisées dans la SAE)

Compétence	Modéliser les données dans un cadre statistique
Apprentissages critiques sollicités	Comprendre l'impact du type de données sur le choix de la modélisation à mettre en œuvre
	Apprécier les limites de validité et les conditions d'application d'un modèle
	Réaliser l'importance de la mise en œuvre d'une procédure de test statistique pour valider ou non une hypothèse
Composantes essentielles à respecter	En choisissant le modèle adapté à la situation
	En maîtrisant la qualité du modèle
	En s'adaptant aux spécificités (données, enjeux, méthodes) d'un domaine d'application particulier (santé, marketing, assurance, qualité, socio-démographie...)

Compétence	Analyser statistiquement les données
Apprentissages critiques sollicités	Prendre conscience de la différence entre modélisation statistique et analyse exploratoire
	Appréhender l'idée de confronter une hypothèse avec la réalité pour prendre une décision
	Apprécier les limites de validité et les conditions d'application d'une analyse
Composantes essentielles à respecter	En mettant en évidence les grandes tendances et les informations principales
	En identifiant et en mettant en œuvre les techniques adaptées aux attentes du client ou de l'instance décisionnaire

Compétence	Valoriser une production dans un contexte professionnel
Apprentissages critiques sollicités	Saisir la nécessité de choisir des indicateurs pertinents pour communiquer sur les résultats
	Savoir défendre ses choix d'analyses
	Prendre conscience de la rigueur requise dans ses productions et dans la communication à leur propos
Composantes essentielles à respecter	En interprétant et contextualisant les résultats (citations, vérification des sources, esprit critique)
	En s'adaptant au niveau d'expertise, à la culture et au statut du destinataire
	En s'exprimant correctement, aussi bien en français que dans une langue étrangère, à l'oral comme à l'écrit

Ma démarche

Savoirs / connaissances	Savoir-faire	Savoir-être
Utilisation de Rstudio afin de comparer plusieurs groupes de variables et modéliser le lien entre des variables afin d'expliquer une variable d'intérêt.	<ul style="list-style-type: none"> -Mettre en place un modèle linéaire multivarié à partir d'un échantillon pour expliquer les variations d'une variable quantitative à partir de plusieurs variables explicatives qualitatives ou quantitatives -Identifier différentes possibilités de modèles en vue de sélectionner le plus pertinent -Choisir un modèle adapté aux types des variables (régression linéaire, ANOVA, analyse de covariance...) -Valider le modèle, notamment via l'étude des résidus et l'interprétation de certains coefficients (coefficient de détermination R^2, p-value, ...) -Sélectionner les variables explicatives les plus pertinentes (via les procédures de sélection, le critère AIC, ...) 	<ul style="list-style-type: none"> Communication au sein d'une équipe Respect des délais

Evaluation du résultat

- Ce que je trouve bien réalisé, pourquoi ?

Notre méthode rigoureuse pour tester l'hypothèse que les moyennes des groupes sont nulles était efficace. La clarté de nos modèles statistiques, nos hypothèses claires et l'utilisation correcte des résultats de R ont rendu l'interprétation des données facile. Nous avons découvert des différences significatives dans la production d'IFNg par certains perturbateurs, ce qui est crucial. Cela démontre l'effet marqué de ces variables et valide l'utilité de l'ANOVA dans notre étude.

- Ce que je n'ai pas bien compris ; ce qui serait à améliorer pour une prochaine fois : pourquoi ? comment ?

Un point peu clair dans notre projet était comment les différents perturbateurs interagissent entre eux. Nous avons analysé chaque perturbateur séparément pour voir son effet sur la production d'IFNg. Toutefois, dans la réalité, les interactions entre plusieurs perturbateurs pourraient aussi influencer la production d'IFNg. Nous n'avons pas étudié en profondeur comment ces interactions affectent les résultats, ce qui aurait pu nous donner plus d'informations sur la complexité des réponses immunitaires. Pour de futures recherches, il serait utile d'étudier les effets combinés de différents facteurs pour mieux comprendre leur impact collectif sur les résultats.

Eléments de preuve, ce que je peux montrer

- 1) Preuve pour les apprentissages critiques « **Modéliser les données dans un cadre statistique** » et « **Apprécier les limites de validité et les conditions d'application d'un modèle** » :

```
21 Dans notre cas on va mettre une contrainte telle que l'intercept sera nul, qui nous permettra de
    savoir si la moyenne d'un groupe est égale à 0 ou non.
22
23 Les hypothèses du test sont les suivantes :
24
25  $H_0 : Y_i = E_i$ 
26  $H_1 : Y_{i,j} = \mu_j + E_i$ 
27
28 On établit le modèle suivant :
29
30 ```{r}
31 mod_1 <- lm(IFNg~ Perturbators - 1, data = immuno)
32 summary(mod_1)
33 ```
```

Ce modèle illustre la compétence en modélisation statistique et notre capacité à adapter le modèle aux données spécifiques de l'étude, en l'occurrence l'utilisation d'un modèle sans intercept pour examiner l'effet des perturbateurs sur la variable IFNg. Ici, nous discutons des contraintes imposées par la fixation de l'intercept à zéro dans notre modèle ANOVA. Cette discussion illustre notre capacité à évaluer quand et comment un modèle statistique est approprié pour l'analyse des données et reconnaît les implications de ces choix méthodologiques.

2) Preuve pour les apprentissages critiques « Analyser statistiquement les données » et « Valoriser une production dans un contexte professionnel » :

On veut maintenant tester au risque 5% si les espérances de IFNg sont égales dans chaque groupe ou non (=voir si tous les perturbateurs produisent autant de IFNg ou pas) :

```
##
## Call:
## lm(formula = IFNg ~ C(Perturbators, base = 46), data = immuno)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -65393 -14661  -4548   8923 130361
```

```
##
## Coefficients:
##                                     Estimate
## (Intercept)                        28114.4
## C(Perturbators, base = 46)Alum (200ug/ml) -16997.0
## C(Perturbators, base = 46)Alum (200ug/ml) + Flu (1X) -9140.2
## C(Perturbators, base = 46)Curdlan (10ug/ml) -2423.1
##
##                                     Std. Error
## (Intercept)                        3745.8
## C(Perturbators, base = 46)Alum (200ug/ml) 14826.5
## C(Perturbators, base = 46)Alum (200ug/ml) + Flu (1X) 14826.5
## C(Perturbators, base = 46)Curdlan (10ug/ml) 7624.3
##
##                                     t value Pr(>|t|)
## (Intercept)                        5.370 1.42e-07
## C(Perturbators, base = 46)Alum (200ug/ml) -1.146 0.252396
## C(Perturbators, base = 46)Alum (200ug/ml) + Flu (1X) -0.616 0.537969
## C(Perturbators, base = 46)Curdlan (10ug/ml) -0.318 0.750812
##
## (Intercept)                        ***
## C(Perturbators, base = 46)Alum (200ug/ml)
## C(Perturbators, base = 46)Alum (200ug/ml) + Flu (1X)
## C(Perturbators, base = 46)Curdlan (10ug/ml)
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24850 on 358 degrees of freedom
## Multiple R-squared:  0.2943, Adjusted R-squared:  0.1583
## F-statistic: 2.164 on 69 and 358 DF, p-value: 2.768e-06
```

Conclusion du test : la p-valeur du test de Fisher est égale à 2.768e-06, ce qui est plus petit que 5%. On rejette donc l'hypothèse selon laquelle il n'y a pas de différence entre les moyennes des groupes. En pratique, on a alors mis en évidence qu'au moins une des moyennes des perturbateurs autres que "PerturbatorsMed" diffère de celle de "PerturbatorsMed".

Ici, on a montré les coefficients estimés de l'ANOVA en utilisant une variable de référence.

On a donc démontré l'analyse statistique en examinant si les moyennes de la variable IFNg diffèrent significativement parmi les différents groupes de perturbateurs.

On tire également des conclusions de la section ANOVA où la p-valeur du test de Fisher est discutée, reflétant la compétence à contextualiser et valoriser les résultats de l'analyse dans un rapport professionnel. Ce passage montre aussi l'importance de choisir les bons indicateurs pour communiquer efficacement sur les résultats.