

Examen de modèle linéaire gaussien général

Thomas Laurent

18 Juin 2017

Exercice 1

Question 1

Dans ce plan d'expérience, on souhaite étudier le score de performance en fonction de 3 facteurs: sexe (2 modalités), classe d'âge (4 modalités) et le type d'études (2 modalités). Au total, il y a 16 cellules ($2 \times 4 \times 2$) pour ce plan d'expérience.

Question 2

Question 2.a

Pour le plan contenant 76 observations (commerciaux), toutes les cellules contiennent plus d'une observation mais le nombre d'observations par cellule n'est pas égal entre les différentes cellules.

Pour le plan contenant 16 observations, chaque cellule contient une seule observation, ce qui correspond bien à un plan équilibré et non répété.

Pour le plan contenant 10 observations, certaines cellules ne contiennent aucune observation d'où le caractère incomplet et non répété du plan.

Question 2.b

Dans cette étude, étant donné que la variable réponse (score) est une variable unidimensionnelle et que l'on dispose de trois facteurs, il est possible de mettre en place pour les trois plans une ANOVA.

Question 2.c

Pour le plan présentant 76 observations, le plan étant complet et répété, il y est possible de tester les effets principaux, les interactions d'ordre 2 et les effets d'interactions d'ordre maximal, ici de degré 3 puisque le plan d'expérience contient 3 facteurs.

Pour le plan présentant 20 observations, le plan étant complet mais non répété, il y est possible de tester les effets principaux, les interactions d'ordre 2 mais pas les effets d'interactions d'ordre 3.

Pour le plan présentant 10 observations, le plan étant incomplet et non répété, il est seulement possible de tester les effets principaux des facteurs.

Exercice 2

Question 1

Question 1.a)

Etant donné que l'on s'intéresse à l'effet du traitement, l'effet du temps et à l'influence du temps sur l'effet du traitement, on choisit de mettre en place un modèle ANOVA pour données répétées.

Question 1.b)

Table 1: Plan d'expérience

Traitement	Observations
0	17
1	15
2	15
3	13

Table 2: Moyenne de la clairance de la créatinine par groupe de traitement

Traitement	1 mois	2 mois	3 mois	4 mois	5 mois	6 mois
0	27.18824	27.26471	27.69412	28.72941	28.72353	29.27647
1	29.12000	29.60000	30.54000	30.93333	31.55333	31.90000
2	26.00667	27.99333	28.30000	30.35333	29.84667	31.87333
3	29.48462	32.40000	36.20769	40.25385	43.06923	46.23846

On remarque que le plan d'expérience est bien répété. Ensuite, les moyennes par groupe de la clairance de la créatinine en fonction du temps montrent qu'il y a une faible augmentation en fonction du temps pour les traitements 0 et 1, légèrement plus marquée pour le traitement 2, et une forte augmentation pour le dernier traitement. On met ensuite le modèle en place pour tester l'effet global du traitement et l'effet du traitement au cours du temps.

Table 3: Tests unidimensionnels

Statistique	Response cc1	Response cc2	Response cc3	Response cc4	Response cc5	Response cc6
R2	0.0387608	0.0570295	0.1310440	0.1915319	0.2169158	0.2582718
P-values	0.5254101	0.3452371	0.0473784	0.0073489	0.0031730	0.0007495

Tout d'abord, pour conclure sur l'effet global du traitement, les tests unidimensionnels montrent que l'effet du traitement est significatif à partir de 3 mois à un seuil de 5%. Ensuite, les tests multidimensionnels permettent de montrer que l'effet du traitement est significatif globalement.

Multivariate Tests: trt

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
Pillai	3	0.1520427	3.347021			3	56	0.025366	*	
Wilks	3	0.8479573	3.347021			3	56	0.025366	*	
Hotelling-Lawley	3	0.1793047	3.347021			3	56	0.025366	*	

```
Roy                3 0.1793047 3.347021      3      56 0.025366 *
```

```
---
```

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

Ensuite si l'on s'intéresse à l'effet principal du temps et à l'interaction entre le facteur temps et le traitement, il faut vérifier si l'hypothèse de sphéricité n'est pas rejetée. Or le test est rejeté à un seuil de 5%.

Test de sphéricité

```

      Test statistic    p-value
temps          0.012344 1.3941e-42
trt:temps      0.012344 1.3941e-42
```

Pour conclure sur l'effet principal du temps et l'interaction temps x traitement, il faut regarder les sorties relatives aux tests de Greenhouse-Geisser et Huynh-Feldt. Ces tests étant significatifs, on rejette l'hypothèse d'absence d'effet du temps et l'absence de l'interaction temps x traitement.

Table 4: Tests de Greenhouse-Geisser et Huynh-Feldt

	GG_stat	GG_pval	HF_stat	HF_pval
temps	3.31514e-01	1.715455e-21	3.403657e-01	5.341152e-22
trt:temps	3.31514e-01	4.358439e-12	3.403657e-01	2.358561e-12

Question 1.c)

On désigne les effets intra-sujets comme les effets liés au temps, c'est-à-dire l'effet principal du temps et les interactions liées au temps. En revanche, les effets inter-sujets correspondent aux effets principaux liés aux autres facteurs que le temps, et les interactions ne faisant pas intervenir le temps. Dans notre modèle, l'effet du traitement est l'effet inter-sujets, et l'effet du temps et l'interaction temps x traitement sont les effets intra-sujets.

Question 1.d)

L'effet du traitement étant significatif, on peut s'intéresser à une ANOVA réalisée sur les moyennes temporelles des observations.

Table de l'ANOVA

```

      Df Sum Sq Mean Sq F value Pr(>F)
trt      3      823   274.29    3.347 0.0254 *
Residuals 56    4589    81.95
```

```
---
```

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

Coefficients du modèle ANOVA sur la moyenne temporelle des observations

```

      Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.942      2.511   15.112  <2e-16 ***
trt0         -9.796      3.335   -2.937   0.0048 **
trt1         -7.335      3.430   -2.138   0.0369 *
trt2         -8.880      3.430   -2.589   0.0123 *
```

```
---
```

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

Residual standard error: 9.053 on 56 degrees of freedom

Multiple R-squared: 0.152, Adjusted R-squared: 0.1066

F-statistic: 3.347 on 3 and 56 DF, p-value: 0.02537

Le tableau de l'ANOVA permet de conclure que l'effet du traitement est significatif. D'après les coefficients de ce modèle, on en déduit que le traitement 3 semble le plus efficace pour augmenter la clairance de la créatinine. Cela est en accord avec les résultats d'analyse descriptive donnés en question 1.b).

Question 1.e)

Etant donné que le test de sphéricité a été rejeté, on ne peut pas prendre une structure de covariance de type "compound symmetry".

Table 5: Comparaison du critère BIC entre les différentes structures de covariance

	BIC
Autoregressif	1792.510
Sans structure	1824.652

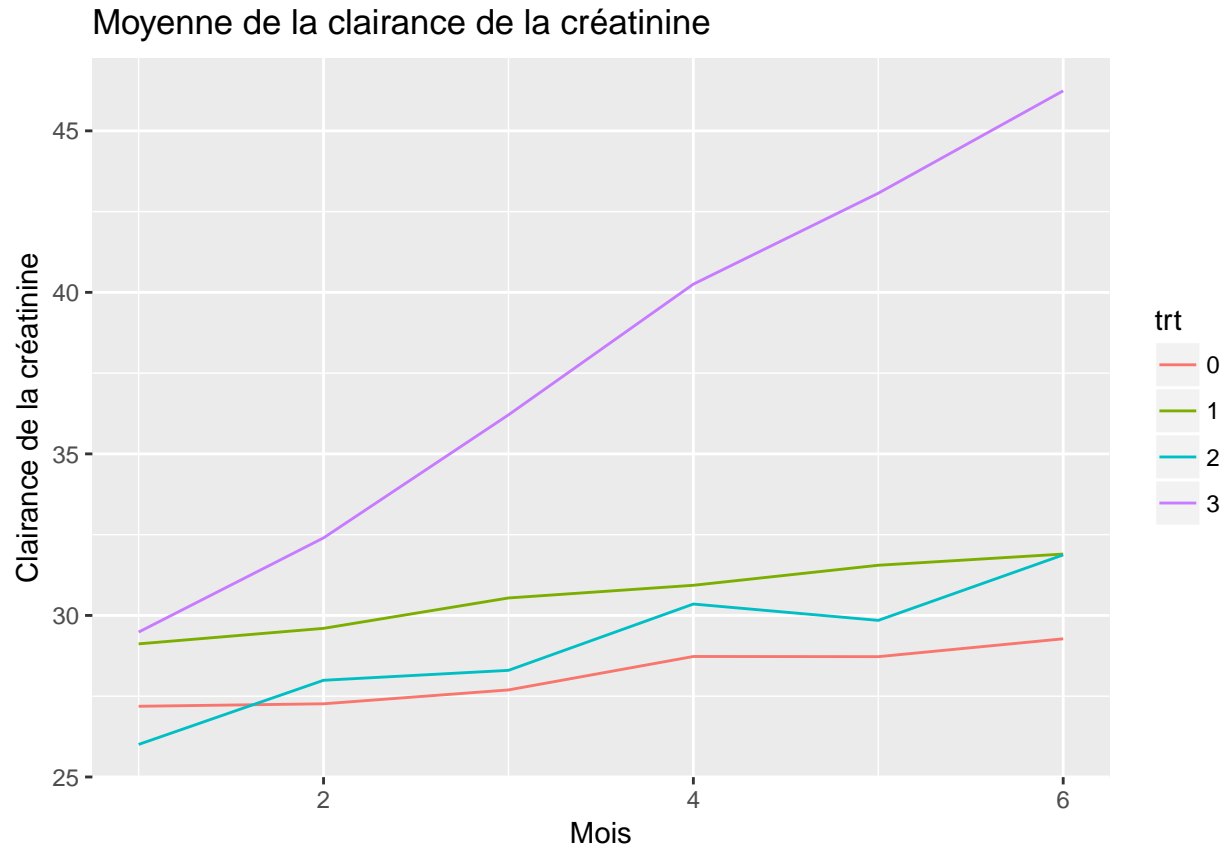
Le critère BIC étant plus faible pour la structure auto-régressive d'ordre 1. La particularité de la structure auto-régressive est que la covariance diminue d'autant plus que l'intervalle entre les deux instants considérés (i et i') est grand puisque la covariance correspond à la variance multipliée par un facteur $\rho^{i-i'}$, ρ étant compris entre 0 et 1. Si le choix se porte sur la structure indépendante, cela revient à effectuer une ANOVA avec un facteur traitement et un facteur temps (facteurs considérés comme des effets fixes).

Question 1.f)

Afin de conclure sur les effets mis en évidence, on s'intéresse à la fois aux coefficients du modèle estimé et aux moyennes en fonction du temps et du traitement.

	Value	Std.Error	t-value	p-value
(Intercept)	46.238461	2.6116734	17.704535	0.0000000
trt0	-16.961991	3.4694052	-4.889020	0.0000016
trt1	-14.338461	3.5682251	-4.018374	0.0000724
trt2	-14.365128	3.5682251	-4.025847	0.0000702
tempssc1	-16.753846	1.1851799	-14.136121	0.0000000
tempssc2	-13.838461	1.0657379	-12.984864	0.0000000
tempssc3	-10.030769	0.9279204	-10.809946	0.0000000
tempssc4	-5.984615	0.7617340	-7.856569	0.0000000
tempssc5	-3.169231	0.5415457	-5.852195	0.0000000
trt0:tempssc1	14.665611	1.5744194	9.314933	0.0000000
trt1:tempssc1	13.973846	1.6192640	8.629752	0.0000000
trt2:tempssc1	10.887180	1.6192640	6.723536	0.0000000
trt0:tempssc2	11.826697	1.4157500	8.353662	0.0000000
trt1:tempssc2	11.538462	1.4560751	7.924359	0.0000000
trt2:tempssc2	9.958462	1.4560751	6.839250	0.0000000
trt0:tempssc3	8.448416	1.2326702	6.853753	0.0000000
trt1:tempssc3	8.670769	1.2677806	6.839329	0.0000000
trt2:tempssc3	6.457436	1.2677806	5.093496	0.0000006
trt0:tempssc4	5.437557	1.0119044	5.373587	0.0000001
trt1:tempssc4	5.017949	1.0407267	4.821582	0.0000022
trt2:tempssc4	4.464615	1.0407267	4.289902	0.0000234

	Value	Std.Error	t-value	p-value
trt0:tempscc5	2.616290	0.7194014	3.636759	0.0003194
trt1:tempscc5	2.822564	0.7398922	3.814831	0.0001622
trt2:tempscc5	1.142564	0.7398922	1.544230	0.1234738



D'après le graphique, on observe qu'au début de l'expérience dans le groupe traitement 2 et le groupe traitement 3, la clairance de la créatinine est respectivement minimale (valeur égale à 26) et maximale (valeur égale à 29). D'après les tests, on sait que l'effet du traitement est significatif et que les coefficients associés le sont aussi d'après le tableau précédent. Les coefficients liés aux traitements (effets principaux) sont négatifs et indiquent que l'attribution du traitement 3 est associée avec une clairance de la créatinine supérieure aux autres traitements (entre 14 et 17 environ). Les coefficients associés au temps (effet principal) sont tous significatifs et négatifs, et la valeur augmente en fonction du temps. Aussi, d'après le graphique, on peut vérifier pour tous les traitements que la clairance de la créatinine augmente au cours du temps. Cependant, l'augmentation de la clairance est remarquable pour le groupe traitement 3 dont la valeur atteint environ 46 lors du dernier mois. De plus, l'effet d'interaction temps x traitement et les coefficients étant significatifs, on en déduit que le temps a un effet sur le traitement et que le traitement 3 semble être le traitement à privilégier.

Question 2.a)

On envisage de mettre en place le modèle suivant:

$$Clairance_{ij} = \beta Age_i + Traitement_j + \alpha_j Age_i + U_{ij}$$

où $Clairance_{ij}$ correspond au vecteur aléatoire de dimension T de la clairance, β est un vecteur contenant les paramètres liés à l'âge, $Traitement_j$ est un vecteur contenant les paramètres liés au traitement j et α_j un vecteur des paramètres liés aux interactions âge x traitement (j indiquant le traitement j). Enfin U_{ij} est le vecteur aléatoire de loi $N_T(0, R)$, où R est la matrice de variance-covariance. Dans ce modèle, tous les vecteurs sont de dimension 6 (nombre d'instants d'observation).

Question 2.b)

Les tests unidimensionnels montrent que les effets principaux (age et traitement) et l'effet d'interaction age x traitement sont significatifs.

Table 7: Tests unidimensionnels

Statistique	1 mois	2 mois	3 mois	4 mois	5 mois	6 mois
R2	0.7582031	0.7544585	0.8103539	0.8326315	0.8472135	0.8722236
Pval-trt	0.0502352	0.0119332	0.0000045	0.0000000	0.0000000	0.0000000
Pval-Age	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Pval-Age x Trt	0.7892716	0.7376625	0.1085915	0.0061134	0.0020486	0.0008391

Tests multidimensionnels - Traitement

Multivariate Tests: trt

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
Pillai	3	0.2302900	5.185971			3		52	0.0032851	**
Wilks	3	0.7697100	5.185971			3		52	0.0032851	**
Hotelling-Lawley	3	0.2991907	5.185971			3		52	0.0032851	**
Roy	3	0.2991907	5.185971			3		52	0.0032851	**

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

Tests multidimensionnels - Age

Multivariate Tests: age

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
Pillai	1	0.5598206	66.13365			1	52	7.8719e-11	***	
Wilks	1	0.4401794	66.13365			1	52	7.8719e-11	***	
Hotelling-Lawley	1	1.2718010	66.13365			1	52	7.8719e-11	***	
Roy	1	1.2718010	66.13365			1	52	7.8719e-11	***	

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

Tests multidimensionnels - Age x Traitement

Multivariate Tests: trt:age

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
Pillai	3	0.1473216	2.994768			3		52	0.039034	*
Wilks	3	0.8526784	2.994768			3		52	0.039034	*
Hotelling-Lawley	3	0.1727750	2.994768			3		52	0.039034	*

```
Roy          3 0.1727750 2.994768      3      52 0.039034 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On peut vérifier que les tests multidimensionnels correspondant à l'effet du traitement, de l'âge et de leur interaction sont tous significatifs et on rejette l'absence de ces effets. On peut souligner la forte significativité des tests liés à l'âge.

Question 2.c)

D'après les résultats de la question précédente, on remarque que la covariable age est très significative ($P\text{-value}=7.8719\text{e-}11$). De plus, le coefficient R^2 pour les modèles unidimensionnels varie entre 0.75 et 0.87, ce qui est bien supérieur aux modèles unidimensionnels mis en place dans la question 1 qui varient entre 0.04 et 0.26. Ainsi, la prise en compte de l'âge dans le modèle permet de bien expliquer la variabilité dans les données, et on conclut qu'il est fortement intéressant d'ajouter cette covariable.

Question 2.d)

Ensuite, on réalise l'analyse complète en incluant la covariable age. On reprend le modèle mis en oeuvre dans la question précédente. Les effets liés au traitement et à l'âge (effets principaux et interactions) étant significatifs, on garde ces effets fixes dans le modèle. Dans une deuxième étape, il faut vérifier si la structure de la covariance est "compound symmetry".

Test de sphéricité (modèle incluant la covariable age)

	Test statistic	p-value
temps	0.16231	2.5325e-13
trt:temps	0.16231	2.5325e-13
age:temps	0.16231	2.5325e-13
trt:age:temps	0.16231	2.5325e-13

Or le test de sphéricité étant rejeté, il est nécessaire de choisir une structure différente et de tester les effets liés au temps avec les tests corrigés. Les tests de Greenhouse-Geisser et Huynh-Feldt étant très significatifs, l'absence d'effets liés au temps est rejetée.

Table 8: Tests de Greenhouse-Geisser et Huynh-Feldt

	GG_stat	GG_pval	HF_stat	HF_pval
temps	6.005461e-01	4.626912e-36	6.414427e-01	2.431097e-38
trt:temps	6.005461e-01	1.846227e-20	6.414427e-01	1.054748e-21
age:temps	6.005461e-01	1.171438e-18	6.414427e-01	9.154767e-20
trt:age:temps	6.005461e-01	4.170015e-11	6.414427e-01	1.003615e-11

Table 9: Comparaison BIC entre les différentes structures de covariance

	BIC
Autoregressif	1844.943
Sans structure	1879.755

D'autre part, lorsque l'on compare différentes structures de covariance, le critère BIC étant le plus faible pour la structure autorégressive d'ordre 1, on retient celle-ci pour le modèle. On peut s'intéresser à vérifier à

nouveau les tests pour chaque effet dans ce modèle.

Table 10: Test des effets dans le modèle

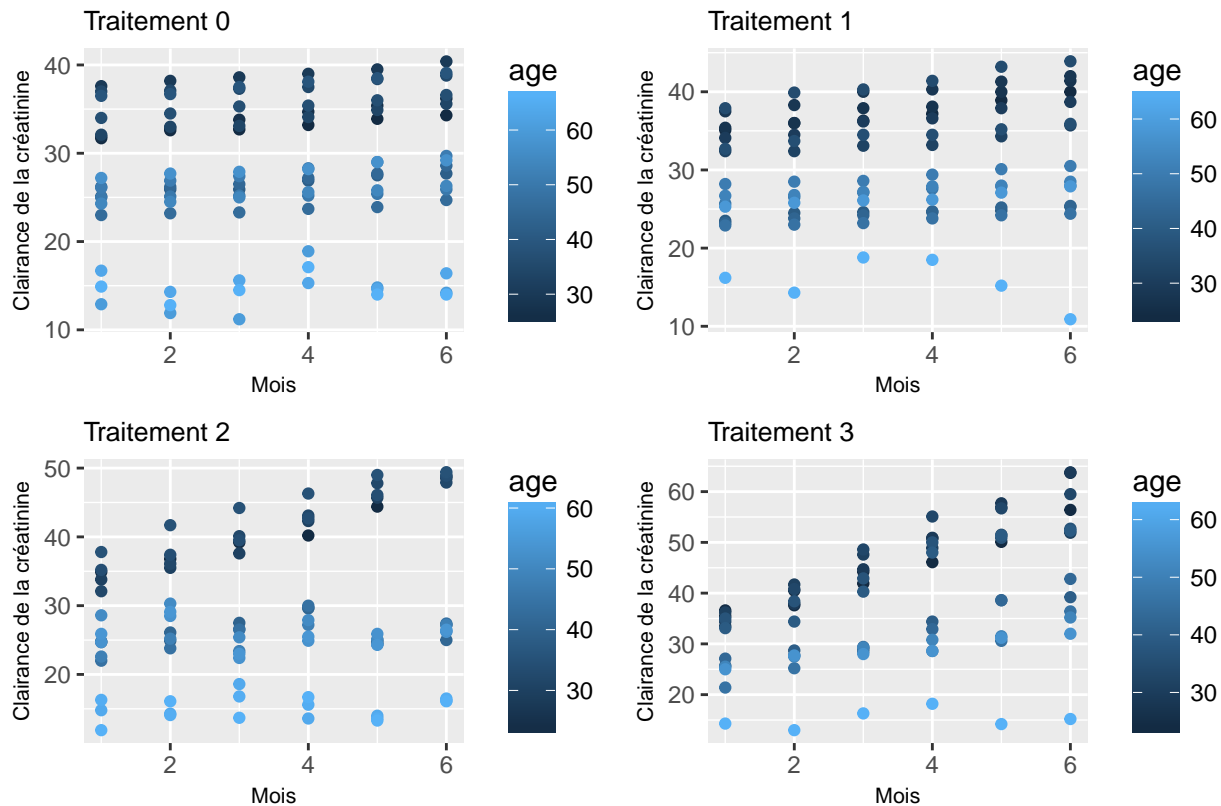
	numDF	F-value	p-value
(Intercept)	1	3673.511736	0.0000000
trt	3	17.160229	0.0000000
age	1	227.031702	0.0000000
temps	5	41.360061	0.0000000
trt:age	3	3.105765	0.0268009
trt:temps	15	11.716181	0.0000000
age:temps	5	13.681902	0.0000000
trt:age:temps	15	2.917790	0.0002415

La table de l'anova montre que tous les effets sont très significatifs mis à part l'interaction traitement x age qui est bien significative mais avec une p-value plus élevée. Ainsi, on retient donc ce modèle dont on estime les paramètres (indiqués ci-dessous).

	Value	Std.Error	t-value	p-value
(Intercept)	91.7713664	4.4105819	20.8070880	0.0000000
trt0	-36.3043068	5.9536059	-6.0978686	0.0000000
trt1	-34.1650537	6.0026283	-5.6916824	0.0000000
trt2	-16.0158600	6.2726310	-2.5532922	0.0111465
age	-1.1471468	0.1068708	-10.7339572	0.0000000
tempsec1	-41.2987309	3.6015130	-11.4670503	0.0000000
tempsec2	-34.1497453	3.2832336	-10.4012537	0.0000000
tempsec3	-25.8926170	2.8988292	-8.9320948	0.0000000
tempsec4	-16.4901158	2.4137145	-6.8318420	0.0000000
tempsec5	-7.0798147	1.7409996	-4.0665229	0.0000605
trt0:age	0.5620745	0.1372777	4.0944330	0.0000540
trt1:age	0.5211814	0.1431843	3.6399353	0.0003191
trt2:age	0.1573234	0.1445655	1.0882497	0.2773248
trt0:tempsec1	35.7284413	4.8614876	7.3492816	0.0000000
trt1:tempsec1	31.1783952	4.9015174	6.3609680	0.0000000
trt2:tempsec1	16.5756296	5.1219913	3.2361690	0.0013416
trt0:tempsec2	31.3223716	4.4318594	7.0675464	0.0000000
trt1:tempsec2	26.1230331	4.4683516	5.8462349	0.0000000
trt2:tempsec2	12.8805047	4.6693414	2.7585271	0.0061489
trt0:tempsec3	23.1036493	3.9129727	5.9043728	0.0000000
trt1:tempsec3	18.0399840	3.9451923	4.5726501	0.0000070
trt2:tempsec3	11.0269840	4.1226500	2.6747320	0.0078732
trt0:tempsec4	12.4405970	3.2581426	3.8183096	0.0001621
trt1:tempsec4	9.3511294	3.2849703	2.8466405	0.0047111
trt2:tempsec4	7.3277107	3.4327307	2.1346593	0.0335679
trt0:tempsec5	6.2362472	2.3500812	2.6536305	0.0083710
trt1:tempsec5	3.6771978	2.3694319	1.5519322	0.1216922
trt2:tempsec5	4.7666502	2.4760107	1.9251331	0.0551216
age:tempsec1	0.6183789	0.0872666	7.0860858	0.0000000
age:tempsec2	0.5117184	0.0795546	6.4322955	0.0000000
age:tempsec3	0.3996202	0.0702402	5.6893356	0.0000000
age:tempsec4	0.2646735	0.0584856	4.5254446	0.0000086
age:tempsec5	0.0985225	0.0421854	2.3354649	0.0201527

	Value	Std.Error	t-value	p-value
trt0:age:tempscc1	-0.5405932	0.1120958	-4.8226009	0.0000022
trt1:age:tempscc1	-0.4396369	0.1169188	-3.7601902	0.0002027
trt2:age:tempscc1	-0.1930458	0.1180467	-1.6353341	0.1029874
trt0:age:tempscc2	-0.4934985	0.1021894	-4.8292512	0.0000022
trt1:age:tempscc2	-0.3722692	0.1065862	-3.4926574	0.0005473
trt2:age:tempscc2	-0.1194799	0.1076144	-1.1102588	0.2677425
trt0:age:tempscc3	-0.3726656	0.0902250	-4.1304030	0.0000466
trt1:age:tempscc3	-0.2415204	0.0941070	-2.5664438	0.0107411
trt2:age:tempscc3	-0.1449067	0.0950148	-1.5250952	0.1282483
trt0:age:tempscc4	-0.1864319	0.0751260	-2.4815904	0.0136060
trt1:age:tempscc4	-0.1143735	0.0783583	-1.4596207	0.1454004
trt2:age:tempscc4	-0.0922884	0.0791142	-1.1665204	0.2442949
trt0:age:tempscc5	-0.0920302	0.0541880	-1.6983501	0.0904387
trt1:age:tempscc5	-0.0241081	0.0565195	-0.4265449	0.6700050
trt2:age:tempscc5	-0.0920601	0.0570647	-1.6132587	0.1076993

Clairance de la créatinine



Globalement, les coefficients estimés sont significatifs, et en particulier les coefficients liés aux effets principaux sont tous significatifs. Qualitativement, les interprétations pour les effets principaux du traitement et du temps sont les mêmes que le modèle mis en oeuvre sans la covariable age. Le coefficient lié à l'âge est égal à -1.1 environ et cet effet négatif de l'âge sur la clairance de la créatinine est appuyé par les graphiques. En effet, quel que soit le traitement ou l'instant considéré, un âge plus élevé est lié à une clairance plus faible. Le traitement 3 semble le plus efficace puisque la clairance de la créatinine augmente pour un âge équivalent mis à part pour les personnes très âgées pour lesquelles l'augmentation est faible. Pour les traitements 0 et 1, pour un âge équivalent, la clairance de la créatinine augmente peu. En revanche, pour le traitement 2, une tendance différente entre les patients âgés et les patients moins âgés est observable puisque la clairance

augmente fortement seulement pour les patients moins âgés. Ainsi, ce traitement semble être efficace pour la population plus jeune. En conclusion, la prise en compte du facteur age a permis de mettre en évidence une efficacité du traitement 2 pour les jeunes patients même si globalement le traitement 3 semble être meilleur (d'un point de vue de l'efficacité).

Exercice 3

Question 1

Question 1.a)

En choisissant de mettre en oeuvre une régression linéaire multiple, on suppose tout d'abord qu'il existe une relation linéaire entre la concentration plasmatique de rétinol et les variables explicatives, c'est-à-dire l'age, le poids, les graisses consommées par jour et le nombre de verres d'alcool consommés par semaine. Ensuite, on suppose que les observations sont indépendantes et que la concentration plasmatique de rétinol a une distribution normale.

Question 1.b)

Dans le premier modèle, il y a au total 5 paramètres estimés (la constante et un paramètre pour chaque variable explicative). La constante correspond à la valeur moyenne si toutes les variables sont nulles. Les coefficients liés à l'age et à l'alcool étant significatifs à 5 %, on interprète ceux-ci. Pour l'age, le coefficient étant égal à 2.9 environ, une augmentation d'une année s'accompagne en moyenne d'une augmentation de 2.9 ng/mL de la concentration plasmatique de rétinol, toutes choses égales par ailleurs. De manière identique, si on regarde la valeur du coefficient pour l'alcool, on interprète que la consommation d'un verre d'alcool s'accompagne d'une augmentation moyenne de 10.0 ng/mL de rétinol plasmatique.

Question 2.a)

Dans le premier modèle, on a observé que les coefficients liés au bmi et graisses étaient non significatifs. On peut penser qu'une démarche descendante a été mise en oeuvre et que ces deux variables n'ont pas été incluse dans le deuxième modèle dû à leur non significativité à chaque étape.

Question 2.b)

La valeur du R^2 ajusté est équivalent entre les deux modèles qui est de l'ordre de 0.09 et les tests globaux des coefficients sont très significatifs. Ainsi, on peut penser que ces modèles sont équivalent en terme de variabilité expliquée. On peut remarquer que les estimations pour les variables communes sont très proches (age et alcool). Cependant, le deuxième modèle contenant moins de variables explicatives sera retenu car plus simple à interpréter.

Question 3

Pour le deuxième modèle, les p-values pour les coefficients age et alcool étant largement inférieures au seuil de 5 % et les valeurs des coefficients étant positives, l'age et la consommation d'alcool semble avoir un effet positif sur la concentration plasmatique de rétinol. En moyenne, l'augmentation de l'âge d'une année et l'augmentation du nombre de verres d'alcool consommés par semaine sont associés à un accroissement de la concentration de rétinol égal à 3.1 ng/mL et 9.3 ng/mL, respectivement.

Question 4

Pour valider ce modèle, il faudrait tout d'abord bien vérifier l'hypothèse de normalité des résidus avec un test de shapiro-wilk, par exemple, et visualiser leur distribution visuellement afin de vérifier s'il n'y a pas de point atypique, par exemple en utilisant un Q-Q plot. Ensuite, il conviendrait de vérifier l'influence de potentiels points atypiques en évaluant l'effet levier et la distance de cook pour chaque observation.

Annexe

Les codes utilisés pour l'exercice 2 sont indiqués ci-dessous.

```
#Package utilise
library(dplyr)
library(nlme)
library(magrittr)
library(doby)
library(knitr)
library(car)
library(tidyr)
library(nlme)
library(gridExtra)
library(ggplot2)

#Question 1

#Import des données
rein=read.table("creatinine.csv",header=TRUE,sep=" ")
rein=rein %>% mutate(trt=factor(trt))

#Plan d'experience
data.frame(table(rein$trt)) %>% set_names(.,c("Traitement","Observations")) %>%
  kable(.,caption="Plan d'expérience")

#Moyenne par groupe
#Analyse descriptive en fonction du facteur traitement
rein%>% select(-age,-num) %>% group_by(trt)%>%summarise_all(mean) %>%
  set_names(.,c("Traitement","1 mois","2 mois","3 mois","4 mois","5 mois","6 mois")) %>%
  kable(.,caption="Moyenne de la clairance de la créatinine par groupe de traitement")

#Mise en place de l'ANOVA pour donnees repetees
cc=rein %>% select(starts_with("cc")) %>% as.matrix()
attach(rein)
options(contrasts = c("contr.SAS","contr.SAS"))
modele_rep=lm(cc~trt)
summary_lin=summary(modele_rep)
ana_manova=manova(modele_rep)
H=summary.aov(ana_manova)

##Formattage des sorties
R2=sapply(summary_lin,function(x)x$r.squared)
p_values=sapply(H,function(x)x$Pr[1])
rbind(R2,p_values) %>% tbl_df() %>% add_rownames(.,var="Statistique") %>%
  mutate(Statistique=c("R2","P-values")) %>%
  kable(.,caption="Tests unidimensionnels")
```

```

#Modele de donnees repetees
##Vecteur correspondant aux 6 repetitions des mesures
vect_rep=data.frame(temps=as.factor(1:6))
##Mise en place du modele
modele_rep=Anova(modele_rep,idata=vect_rep,idesign=~temps,type="III",test="Pillai")
summary=summary(modele_rep)$multivariate.tests
cat(capture.output(summary)[48:55], sep = "\n")

#Effet global du temps
cat("Test de sphéricité \n")
summary(modele_rep)$sphericity.test

#ANOVA sur la moyenne temporelle des observations
data_moy=rein %>% mutate(moyenne=(cc1+cc2+cc3+cc4+cc5+cc6)/6)
modele_moy=lm(moyenne~trt,data=data_moy)
summary=summary(modele_moy)
anova_moy=aov(moyenne~trt,data=data_moy)
cat("Table de l'ANOVA")
summary.aov(anova_moy)
cat("Coefficients du modèle ANOVA sur la moyenne temporelle des observations")
cat(capture.output(summary)[10:20], sep = "\n")

#Choix de la structure de covariance
##Changement des donnees en format long
rein_long=gather(rein,key="temps",value="cc",-num,-trt,-age) %>%
  mutate(temps=as.factor(temps)) %>% arrange(num,trt,age)
detach(rein)
attach(rein_long)

#Mis en place du modele avec structure de covariance auto-regressive d'ordre 1
modele_rep_autoreg=glsc(cc~trt*temps,data=rein_long,method="REML",
  correlation=corAR1(form=~1|num))

#Mis en place du modele avec matrice de variance-covariance sans structure particuliere
modele_rep_unstr=glsc(cc~trt*temps,data=rein_long,method="REML",
  correlation=corSymm(form=~1|num))

#Comparaison de BIC

result=rbind(c(BIC(modele_rep_autoreg)),c(BIC(modele_rep_unstr)))
as.data.frame(result,row.names=c("Autoregressif","Sans structure")) %>%
  tbl_df() %>% setNames(.,c("BIC")) %>%
  kable(.,
    caption="Comparaison du critère BIC entre les différentes structures de covariance")

#Estimation du modèle
coefficients=as.data.frame(summary(modele_rep_autoreg)$tTable)%>% tbl_df()
kable(coefficients)

#Graphique des moyennes en fonction du temps et du traitement
##Calcul des moyennes par traitement et par instant considere
moyennes_sum=rein_long %>% group_by(trt,temps) %>% summarise(mean=mean(cc)) %>%
  mutate(tps=as.numeric(substr(temps, 3, 3)))

```

```
##Graphique
ggplot(data = moyennes_sum,aes(x=tps,y=mean)) +
  geom_line(aes(color=trt)) +
  ylab("Clairance de la créatinine")+
  xlab("Mois")+
  ggtitle("Moyenne de la clairance de la créatinine")+
  scale_x_continuous(limits=c(1,6))
```

#Question 2

#Modele incluant la covariable age

```
attach(rein)
modele_age=lm(cc~trt*age)
ana_manova=manova(modele_age)
summary_lin=summary(modele_age)
H=summary.aov(ana_manova)
```

##Formatage des sorties

```
R2=sapply(summary_lin,function(x)x$r.squared)
p_values=sapply(H,function(x)x$Pr[1:3])
rbind(R2,p_values) %>% tbl_df() %>% add_rownames(.,var="Statistique") %>%
  setNames(.,c("Statistique","1 mois","2 mois","3 mois","4 mois","5 mois","6 mois")) %>%
  mutate(Statistique=c("R2","Pval-trt","Pval-Age","Pval-Age x Trt")) %>%
  kable(.,caption="Tests unidimensionnels")
```

#Modele de donnees repetees

```
##Vecteur correspondant aux 6 repetitions des mesures
vect_rep=data.frame(temps=as.factor(1:6))
##Mise en place du modele
modele_age_rep=Anova(modele_age,idata=vect_rep,idesign=~temps,type="III",test="Pillai")
##Formatage des sorties
summary=summary(modele_age_rep)$multivariate.tests
cat("Tests multidimensionnels - Traitement \n")
cat(capture.output(summary)[48:55], sep = "\n")
cat("\n \n \nTests multidimensionnels - Age \n")
cat(capture.output(summary)[76:83], sep = "\n")
cat("\n \n \nTests multidimensionnels - Age x Traitement")
cat(capture.output(summary)[103:112], sep = "\n")
```

#Test sphericite pour le modele incluant la covariable age

```
cat("Test de sphéricité (modèle incluant la covariable age)\n")
summary(modele_age_rep)$sphericity.test
cat("\n")
```

#Effet global du temps

```
data.frame(summary(modele_age_rep)$pval.adjustments) %>% format(.,scientific = TRUE) %>%
  set_names(.,c("GG_stat","GG_pval","HF_stat","HF_pval")) %>%
  kable(.,caption="Tests de Greenhouse-Geisser et Huynh-Feldt")
```

```

#Choix de la structure de covariance non "compound symmetry"
attach(rein_long)

#Mis en place du modele avec structure de covariance auto-regressive d'ordre 1
modele_age_ar=glS(cc~trt*age*temps,data=rein_long,method="REML",
                  correlation=corAR1(form=~1|num))

#Mis en place du modele avec matrice de variance-covariance sans structure particuliere
modele_age_unstr=glS(cc~trt*age*temps,data=rein_long,method="REML",
                    correlation=corSymm(form=~1|num))

#Comparaison de BIC
result=rbind(c(BIC(modele_age_ar)),c(BIC(modele_age_unstr)))
as.data.frame(result,row.names=c("Autoregressif","Sans structure")) %>%
  tbl_df() %>% setNames(.,c("BIC")) %>%
  kable(.,caption="Comparaison BIC entre les différentes structures de covariance")

#Test des differents effets pour le modele incluant la covariable age (structure ar)
as.data.frame(anova(modele_age_ar)) %>% kable(.,caption="Test des effets dans le modèle")

#Estimation du modèle
coefficients=as.data.frame(summary(modele_age_ar)$tTable)%>% tbl_df()
kable(coefficients)

#Graphique des moyennes en fonction du temps et du traitement
##Graphique
graph=function(a,b){
rein_trt=rein_long %>% filter(trt==a) %>%
  mutate(tps=as.numeric(substr(temps, 3, 3)))
ggplot(data = rein_trt,aes(x=tps,y=cc)) +
  geom_point(aes(color=age))+
  ylab("Clairance de la créatinine")+
  xlab("Mois")+
  ggtitle(b)+
  theme(axis.title=element_text(size=8),
        plot.title=element_text(size=10))
}
grid.arrange(graph(a=0,b="Traitement 0"),graph(a=1,b="Traitement 1"),
              graph(a=2,b="Traitement 2"),
              graph(a=3,b="Traitement 3"),ncol=2,nrow=2,
              top = "Clairance de la créatinine")

```