

SÉANCE 2

LES TESTS STATISTIQUES

Objectifs à atteindre

- Test de significativité d'un paramètre
- Calcul d'un intervalle de confiance d'un paramètre
- Calcul d'une prévision et de son intervalle de confiance
- Calcul du coefficient de détermination et du coefficient de détermination ajusté
- Test de significativité globale

Rappels de cours

Test de Student

Test de significativité d'un coefficient avec un seuil de risque de $\alpha\%$:

$$H_0 : a_1 = 0 \text{ contre } H_1 : a_1 \neq 0$$

Statistique de test :

$$t^* = \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} \sim t(T - k)$$

Règle de décision :

$$\text{Rejet de } H_0 \text{ si } |t^*| > t_{T-k}^{\alpha/2}$$

Interprétation :

Lorsque l'on rejette H_0 , on dit que le coefficient est statistiquement significatif. La variable explicative associée a bien un impact significatif sur la variable à expliquer.

Coefficient de détermination

Le coefficient de détermination, noté R^2 , est utilisé pour mesurer la qualité de l'ajustement réalisé au moyen des estimateurs. R^2 est défini comme la part de la variance expliquée dans la variance totale :

$$R^2 = \frac{V(\hat{y})}{V(y)} = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$

On a aussi :

$$\bar{R}^2 = 1 - \frac{SCR/(T - k)}{SCT/(T - 1)}$$

Test de Fisher

Test de significativité globale du modèle :

$$H_0 : a_1 = \dots = a_k = 0 \text{ contre } H_1 : \exists a_i \neq 0$$

Statistique de test :

$$F = \frac{R^2}{1 - R^2} \frac{T - k}{k - 1} \sim F(k - 1, T - k)$$

Règle de décision :

Rejet de H_0 si $F^* > F(k - 1, T - k)$.

Intervalles de confiance

L'intervalle de confiance du paramètre a_1 est tel que :

$$a_1 = \hat{a}_1 \pm t_{T-k}^{\alpha/2} \times \hat{\sigma}_{\hat{a}_1}$$

L'intervalle d'une prévision est donné par :

$$y_t = \hat{y}_t \pm t_{T-k}^{\alpha/2} \times \hat{\sigma}_{e_t}$$

Exercices

Exercice 1 : Tests statistiques - Cas du modèle de régression simple

Soit la variable à expliquer (ou variable endogène) y_t et la variable explicative (ou variable exogène) x_t connues sur 12 périodes. Les valeurs sont reportées dans le tableau 1.

t	y_t	x_t
1	20	54
2	19	53
3	21	59
4	21	66
5	23	63
6	20	62
7	25	65
8	24	60
9	28	59
10	27	65
11	31	70
12	33	65

TABLE 1 – Données

On a estimé le modèle suivant :

$$y_t = a_0 + a_1 x_t + \epsilon_t \quad t = 1, \dots, n.$$

On vous donne les résultats :

$$\hat{y}_t = -11,017 + 0.572 x_t$$

Avec : $\hat{\sigma}_{\hat{a}_0} = 14,03$ et $\hat{\sigma}_{\hat{a}_1} = 0,226$

1. Tester, à un seuil de 5%, l'hypothèse $H_0 : a_1 = 0$ contre l'hypothèse $H_1 : a_1 \neq 0$. Pourquoi cette hypothèse est-elle très importante à tester ?
2. Donner un intervalle de confiance pour le coefficient a_1 à 95%.
3. Tester l'hypothèse suivante à un seuil de 5% : $H_0 : a_1 = 0,5$ contre l'hypothèse $H_1 : a_1 \neq 0,5$.
4. Calculer la somme des carrés totale $SCT = \sum_{t=1}^n (y_t - \bar{y})^2$, la somme des carrés des résidus $SCR = \sum_{t=1}^n e^2$ et la somme des carrés expliqués $SCE = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2$. Calculer le coefficient de détermination et interpréter.
5. Pour les périodes $t = 13$ et $t = 14$, nous connaissons les valeurs prévues de x_t , $x_{13} = 72$ et $x_{14} = 62$. Calculer les prévisions et les intervalles de confiances associés.

Exercice 2 : Les tests statistiques - Cas du modèle de régression multiple

Soit la variable à expliquer y_t et deux variables explicatives $x_{1,t}$ et $x_{2,t}$ observées sur 10 périodes. Les valeurs des variables sont reportées dans le tableau ci-dessous.

t	y_t	$x_{1,t}$	$x_{2,t}$
1	12	7	48
2	21	9	40
3	24	11	18
4	24	12	28
5	13	7	40
6	17	9	32
7	21	12	31
8	26	14	24
9	31	19	22
10	30	21	25

TABLE 2 – Données

On a estimé le modèle suivant : $y_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t} + \epsilon_t$

On a obtenu les résultats suivants : $\hat{y} = 18,87 + 0,902x_{1,t} - 0,256x_{2,t}$

1. Dresser le tableau d'analyse de la variance.
2. Calculer le coefficient de détermination et le coefficient de déterminant ajusté.
3. Effectuer le test de significativité globale de la régression à un seuil de 5%.
4. Calculer les prévisions de y_t pour les périodes 11 et 12, connaissant les valeurs prévues des variables explicatives : $x_{1,11} = 24$ et $x_{2,11} = 21$, $x_{1,12} = 12$ et $x_{2,12} = 30$.

Exercice 3 : L'effet du contrôle continu

Une université souhaite mettre en place une étude de l'impact du contrôle continu sur la réussite des étudiants en première année de licence d'économie et gestion. Cette mesure est mise en place sur la base du volontariat, les étudiants peuvent participer à un contrôle continu dans plusieurs enseignements, qui dépendent de l'option choisie par l'étudiant (économie et gestion ou AES).

Nous disposons d'un échantillon de 267 étudiants, pour lesquels nous avons les informations suivantes :

- le sexe, $femme=1$ s'il s'agit d'une femme ;
- la moyenne générale sur l'année ;
- la réussite en L1 avec deux modalités, $russite=1$ si l'étudiant a réussi son année ;
- la série du baccalauréat avec 4 modalités, $bac=1$ si l'étudiant a fait un bac S, $bac=2$ si l'étudiant a fait un bac ES, $bac=3$ si l'étudiant a fait un bac Pro, et $bac=4$ si l'étudiant a fait un autre bac ;
- la mention obtenue au baccalauréat avec 3 modalités, $mention=1$ si l'étudiant a eu la mention passable, $mention=2$ si l'étudiant a eu la mention assez-bien, et $mention=3$ si l'étudiant a eu la mention bien ;
- le campus sur lequel l'étudiant suit ses cours, $campus=1$ si l'étudiant est localisé à Aix, et $campus=0$ si c'est à Marseille ;

- si l'étudiant a plus d'une année de retard dans son cursus depuis le collège, $retard=1$ si oui, et $retard=0$ sinon ;
- si l'étudiant a effectué une réorientation depuis son baccalauréat, $reorientation=1$ si oui, et $reorientation=0$ sinon ;
- l'option choisie par l'étudiant en L1, $option=1$ si l'étudiant a choisi AES, et $option=0$ si l'étudiant a choisi économie et gestion ;
- si l'étudiant a participé ou non au contrôle continu, $controle_continu=1$ si oui, et $controle_continu=0$ sinon.

Le modèle économétrique estimé est le suivant (modèle (1)) :

$$\begin{aligned} Moyenne_i = & \beta_0 + \beta_1 femme_i + \beta_2 bac_si + \beta_3 bac_es_i + \beta_4 bac_pro_i + \beta_5 mention_ab_i + \\ & \beta_6 mention_bien_i + \beta_7 retard_i + \beta_8 reorientation_i + \beta_9 option_aes_i + \\ & \beta_{10} controle_continu_i + \varepsilon_t \end{aligned}$$

L'estimation par MCO nous donne les résultats suivants :

FIGURE 1 –

Source	SS	df	MS	Number of obs	=	267
Model	953.823086	10	95.3823086	F(10, 256)	=	
Residual	1955.97781	256	7.64053832	Prob > F	=	
Total	2909.8009	266	10.9391011	R-squared	=	
				Adj R-squared	=	
				Root MSE	=	2.7642

moyenne_finale	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
femme	.7429867	.3428307			.0678592 1.418114
baccalaureat_s	1.685185	.6234278	2.70	0.007	.4574849 2.912885
baccalaureat_es	.1634863	.4989145			-.8190129 1.145986
baccalaureat_pro	-.7696393	.6204518	-1.24	0.216	-1.991479 .4522003
mention_ab	1.05559	.4229212	2.50	0.013	.2227429 1.888438
mention_bien	1.487554	.6375993	2.33	0.020	.2319464 2.743162
retard	-.706009	.4592866	-1.54	0.125	-1.61047 .1984522
reorientation	.3062411	.4828929			
aes	.1504356	.3696537	0.41	0.684	-.5775138 .878385
controle_continu	3.369687	.4555728			2.47254 4.266835
_cons	4.594816	.6301789	7.29	0.000	3.353821 5.835811

1. Calculer le coefficient de détermination et le coefficient de détermination ajusté. Quelle est la différence entre ces deux coefficients ? Que pouvez-vous dire de la qualité d'ajustement du modèle ?
2. Donner une estimation de la variance résiduelle.
3. Le modèle est-il globalement significatif ?
4. Tester la significativité des variables *femme*, *bac_es* et *controle_continu*. Interpréter les résultats. En particulier, quelles sont vos préconisations concernant le contrôle continu ?
5. Déterminer l'intervalle de confiance (à 95%) pour le coefficient associé à la variable *reorientation* et donner une interprétation.

6. Donner une prévision de la moyenne pour un homme, ayant un bac S sans mention, sans année de retard mais ayant effectué une réorientation, avec l'option AES et n'ayant pas choisi le contrôle continu.
7. On s'intéresse à la variable de mention au baccalauréat.
- Pourquoi ne pas avoir introduit une variable indicatrice correspondant à l'obtention d'un baccalaureat autre que S, ES, ou professionnel ?
 - La présence des variables de mention au baccalauréat dans le modèle est-elle pertinente ? Faites le test correspondant (hypothèses, statistique de test, règle de décision, etc.).
Note : SCR dans le modèle sans variable de mention au baccalauréat vaut 2031,231.
8. On se pose maintenant la question d'un éventuel effet de campus sur le rôle du contrôle continu. Pour cela, on introduit une variable croisée dans le modèle (*campus * controle_continu*). Les résultats du modèle sont les suivants (modèle (2)) :

FIGURE 2 –

Source	SS	df	MS				
Model	993.833494	11	90.3484995	Number of obs =	267		
Residual	1915.9674	255	7.51359766	F(11, 255) =	12.02		
Total	2909.8009	266	10.9391011	Prob > F =	0.0000		
				R-squared =	0.3415		
				Adj R-squared =	0.3131		
				Root MSE =	2.7411		

moyenne_finale	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
femme	.7965654	.3407628	2.34	0.020	.1254977 1.467633
baccalaureat_s	1.739792	.61868	2.81	0.005	.5214189 2.958165
baccalaureat_es	.2801189	.4973275	0.56	0.574	-.6992734 1.259511
baccalaureat_pro	-.6229355	.6185518	-1.01	0.315	-1.841056 .5951852
mention_ab	1.025652	.4195939	2.44	0.015	.199341 1.851962
mention_bien	1.585752	.6337109	2.50	0.013	.3377781 2.833725
retard	-.5534024	.4602314	-1.20	0.230	-1.459741 .3529362
reorientation	.2196649	.4803321	0.46	0.648	-.7262582 1.165588
aes	.0201397	.3708932	0.05	0.957	-.7102643 .7505436
controle_continu	2.755806	.5242781	5.26	0.000	1.72334 3.788272
aix_controle_continu	.9353037	.4053126	2.31	0.022	.1371173 1.73349
_cons	4.519349	.6257772	7.22	0.000	3.286999 5.751698

- Expliquer la démarche.
- Le modèle est-il globalement significatif ? Justifier.
- Interpréter la variable croisée.

9. On vous donne maintenant les résultats suivants :

FIGURE 3 – Estimation sur l'échantillon du campus d'Aix

Source	SS	df	MS	Number of obs	=	156
Model	560.13406	10	56.013406	F(10, 145)	=	8.82
Residual	921.058372	145	6.3521267	Prob > F	=	0.0000
Total	1481.19243	155	9.55608021	R-squared	=	0.3782
				Adj R-squared	=	0.3353
				Root MSE	=	2.5203

moyenne_finale	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
femme	.9049505	.4100236	2.21	0.029	.0945553 1.715346
baccalaureat_s	1.59305	.7443415	2.14	0.034	.1218892 3.064211
baccalaureat_es	.7746565	.5801522	1.34	0.184	-.3719908 1.921304
baccalaureat_pro	-1.18415	.8065079	-1.47	0.144	-2.778181 .4098798
mention_ab	1.667799	.4925657	3.39	0.001	.6942623 2.641335
mention_bien	2.188599	.7958074	2.75	0.007	.6157182 3.76148
retard	-.0219553	.5673116	-0.04	0.969	-1.143224 1.099313
reorientation	-.2134275	.5679379	-0.38	0.708	-1.335934 .9090787
aes	.7315844	.4471901	1.64	0.104	-.1522688 1.615438
controle_continu	4.314972	.7157051	6.03	0.000	2.90041 5.729534
_cons	2.975397	.9356308	3.18	0.002	1.126161 4.824634

FIGURE 4 – Estimation sur l'échantillon du campus de Marseille

Source	SS	df	MS	Number of obs	=	111
Model	396.677411	10	39.6677411	F(10, 100)	=	4.50
Residual	881.619345	100	8.81619345	Prob > F	=	0.0000
Total	1278.29676	110	11.6208796	R-squared	=	0.3103
				Adj R-squared	=	0.2413
				Root MSE	=	2.9692

moyenne_finale	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
femme	.7594982	.60128	1.26	0.209	-.4334241 1.952421
baccalaureat_s	1.718496	1.062356	1.62	0.109	-.389189 3.826181
baccalaureat_es	-.6960872	.9256545	-0.75	0.454	-2.532559 1.140385
baccalaureat_pro	-.0833525	.9824135	-0.08	0.933	-2.032433 1.865728
mention_ab	.202251	.7632757	0.26	0.792	-1.312066 1.716568
mention_bien	1.696625	1.063724	1.59	0.114	-.413774 3.807024
retard	-1.039128	.7909634	-1.31	0.192	-2.608376 .5301213
reorientation	.7072664	.8611309	0.82	0.413	-1.001193 2.415726
aes	-.902029	.6728078	-1.34	0.183	-2.236861 .4328025
controle_continu	2.727034	.6625698	4.12	0.000	1.412515 4.041554
_cons	5.673792	.9327887	6.08	0.000	3.823166 7.524418

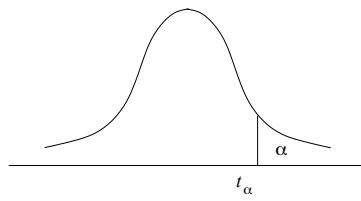
- (a) Expliquer la démarche.
(b) Que pouvez-vous conclure sur les différences entre le campus d'Aix et le campus de

Marseille ?

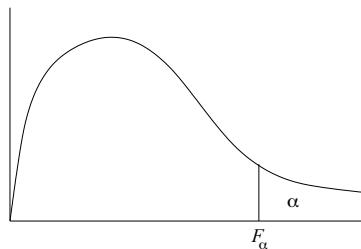
- (c) Tester la stabilité des coefficients du modèle (1) entre le campus d'Aix et le campus de Marseille. Quelles sont les hypothèses sous-jacentes à l'application du test ? Quelles conclusions littéraires pouvez-vous en tirer ?

Tables statistiques

Loi de Student

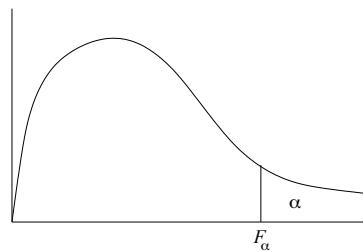


d.d.l.	α	0.10	0.05	0.025	0.01	0.005
1		3.078	6.314	12.706	31.821	63.657
2		1.886	2.920	4.303	6.965	9.925
3		1.638	2.353	3.182	4.541	5.841
4		1.533	2.132	2.776	3.747	4.604
5		1.476	2.015	2.571	3.365	4.032
6		1.440	1.943	2.447	3.143	3.707
7		1.415	1.895	2.365	2.998	3.499
8		1.397	1.860	2.306	2.896	3.355
9		1.383	1.833	2.262	2.821	3.250
10		1.372	1.812	2.228	2.764	3.169
11		1.363	1.796	2.201	2.718	3.106
12		1.356	1.782	2.179	2.681	3.055
13		1.350	1.771	2.160	2.650	3.012
14		1.345	1.761	2.145	2.624	2.977
15		1.341	1.753	2.131	2.602	2.947
16		1.337	1.746	2.120	2.583	2.921
17		1.333	1.740	2.110	2.567	2.898
18		1.330	1.734	2.101	2.552	2.878
19		1.328	1.729	2.093	2.539	2.861
20		1.325	1.725	2.086	2.528	2.845
21		1.323	1.721	2.080	2.518	2.831
22		1.321	1.717	2.074	2.508	2.819
23		1.319	1.714	2.069	2.500	2.807
24		1.318	1.711	2.064	2.492	2.797
25		1.316	1.708	2.060	2.485	2.787
26		1.315	1.706	2.056	2.479	2.779
27		1.314	1.703	2.052	2.473	2.771
28		1.313	1.701	2.048	2.467	2.763
29		1.311	1.699	2.045	2.462	2.756
∞		1.282	1.645	1.960	2.326	2.576

Loi de Fisher : $\alpha = 0.05$ 

	Degrés de liberté du numérateur : ν_1								
	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Loi de Fisher : $\alpha = 0.05$ (suite)



	Degrés de liberté du numérateur : ν_1									
	10	12	15	20	24	30	40	60	120	∞
Degrés de liberté du dénominateur : ν_2	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.30
1	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
2	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
3	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
4	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
5	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
6	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
7	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
8	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
9	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
10	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
11	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
12	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
13	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
14	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
15	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
16	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
17	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
18	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
19	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
20	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
21	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
22	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
23	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
24	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
25	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
26	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Exercice 2.

© Théo Jalabert

1) SC d.d.l Variance

Modèle	345,515 ①	2	172,76	SCE $\sum_{k=1}^m (y_i - \bar{y})^2$
Résidus	31,469 ②	7	15,73	SCR $\sum_{k=1}^m \epsilon_i^2$
Tchol	376,9 ③	9	41,88	SCT $\sum_{i=1}^m (y_i - \bar{y})^2$

$$\text{Ici } \bar{y} = 21,9$$

$$\hat{y} = 18,87 + 0,902x_{1,t} - 0,256x_{2,t}$$

$$③: SCT = \sum_{i=1}^m (y_i - \bar{y})^2$$

$$\text{Modèle: } y_t = a_0 + a_1 x_{1,t} + a_2 x_{2,t} + \epsilon_t$$

$$SCE = \sum_{i=1}^m (\hat{y}_i - \bar{y})^2 = 345,515$$

$$SCR = \sum_{i=1}^m (y_i - \hat{y}_i)^2 = 31,469$$

$$SCT = \sum_{i=1}^m (y_i - \bar{y})^2 = 376,9$$

$$2) R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 0.91673$$

$$\bar{R}^2 = 1 - \frac{SCR/m-k}{SCT/m-1} = 0.89265$$

3) Test de Significativité globale $\alpha = 5\%$.

$$H_0: a_1 = a_2 = 0 \quad \text{Contre } H_1: \exists a_i \neq 0 \quad i \in \{1, 2\}$$

$$\text{Statistique de test: } F^* = \frac{R^2}{1-R^2} \frac{T-k}{k-1} = 38.53 \sim F(2,7)$$

Le modèle permet d'expliquer environ 92% des variantes de y .

Règle de décision: rejeter H_0 si $F^* > F_{\alpha}(2,7) = 4,74$

On a $F^* = 38,53 > F_{\alpha}(2,7) = 4,74$ donc on rejette H_0

Le modèle est globalement significatif.

$$4) \text{ On a } x_{1,11} = 24 \quad x_{2,11} = 21 \quad \Rightarrow \quad \hat{y}_{11} = 35,142 \\ x_{1,12} = 12 \quad x_{2,12} = 30 \quad \hat{y}_{12} = 22,014$$

Exercice 3.

$$1) R^2 = \frac{SCE}{SCT} = \frac{953.823086}{2909.809} = 0.3278$$

$$\bar{R}^2 = 1 - \frac{SCR/T-k}{SCT/T-1} = 1 - \frac{1955.97781/256}{2909.809/256} = 0.3015$$

R^2 ne prend pas en compte le nbr de variables explicatives et augmente mécaniquement avec elles

\bar{R}^2 prend en compte le nbr de variables explicatives et permet ainsi la comparaison du modèle de complexités différentes.

La qualité d'ajustement du modèle est faible car le R^2 est bas.

$$2) \hat{\sigma}_\epsilon^2 = \frac{\sum_1^m \epsilon_i^2}{m-k} = \frac{SCR}{m-k} = \frac{1955.97781}{256} = 7,6405 \text{ déjà dans le tableau}$$

3) Test de significativité globale

$$F^* = \frac{R^2}{1-R^2} \frac{T-k}{k-1} = 12,484$$

$$F_\alpha(10, 256) = 1,83$$

$$\text{Donc } F^* > F_\alpha(10, 256)$$

\Rightarrow rejet de H_0

au seuil $\alpha=5\%$, $\exists \beta_i, \beta_i \neq 0$

Le modèle est globalement significatif.

$$4) b_{femme}^* = \frac{\beta_1}{\hat{\sigma}_{\beta_1}} = \frac{0.7429867}{0.3428307} = 2,17 > 1,96 \Rightarrow \text{significative}$$

$$b_{vacances}^* = \frac{0.1634863}{0.49383155} = 0,328 < 1,96 \Rightarrow \text{non significative}$$

$$b_{Généralisation}^* = \frac{3.369687}{0.4555728} = 7,39 > 1,96 \Rightarrow \text{significative}$$

$$t_{256}^{\alpha/2} = t_{256}^{0,025} = 1,96$$

Le Généralisé continu est significatif et améliore les résultats.

$$5) IC_{\text{réelisat}^{\circ}} = [\hat{\beta}_8 \pm r_{256}^{0,025} \times \hat{\sigma}_{\hat{\beta}_8}]$$

$$= [0,3062411 - 1,96 \times 0,4828929; 0,3062411 + 1,96 \times 0,4828929]$$

$$= [-0,640; 1,253]$$

L'intervalle de confiance est très étendu relativement à la valeur estimée du coeff.

De +, O E à IC. On ne peut pas tirer de grandes conclusions si ce n'est la volatilité du coeff et donc s'interroger sur sa place dans le modèle.

$$6) \text{moyenne} = \beta_0 + \beta_2 + \beta_8 + \beta_9 \\ = 6,74$$

7) a) La série du bac comporte 4 modalités: S, ES, Pro, autre

Donc 3 autres variables indicatrices ont été introduites par soucis de colinéarité parfaite.

Le groupe de référence choisi correspond aux individus ayant fait un bac autre donc on n'a pas introduit de va indicatrice pour cette modalité.

En d'autres termes, si un individu a fait un bac autre, on égale les autres va indicatrices à 0.

b) Test de contraintes linéaires de Fisher:

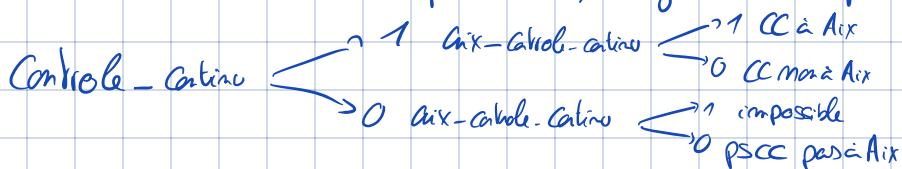
$$H_0: \beta_5 = \beta_8 = 0 \quad \text{contre} \quad H_1: \beta_5 \neq 0 \text{ ou } \beta_8 \neq 0$$

$$F^* = \frac{SCR_c - SCR_g}{c} \cdot \frac{m-k}{SCR_g} = \frac{2031,231 - 1955,97781}{2} \times \frac{256}{1955,97781} = 4,92 > F_{\alpha}(2,256)$$

\Rightarrow On rejette H_0

8)

a) On sait que le Contrôle continu a un impact sur la moyenne finale
On veut déterminer si cet impact est \neq en fonction du campus.



$$5) F^* \sim F(11, 255) = 12,02 > 1,83$$

\Rightarrow globalement significatif

g) La diff entre ceux qui ont suivi le CC à Aix et ceux qui n'ont pas suivi le CC est de 3,69

TCEEP, les étudiants qui suivent le CC à Aix ont une note finale supérieure d'en 0,94 par rapport à ceux qui suivent le CC à Marseille.

La diff entre ceux qui ont suivi le CC à Marseille et ceux qui ne l'ont pas suivi est de 3,756.

g)a) Teste si l'ensemble de va explicatives ont des effets \neq la fact° du campus

b) Fait au dessus

c) Teste la stabilité des paramètres : Test de Chow

$$H_0: \beta^T = \beta^A = \beta^M \text{ contre } H_1: \beta^A \neq \beta^M$$

$$F^* = \frac{(SCR_T - (SCR_A + SCR_M)) / k}{(SCR_A + SCR_M) / (N - 2k)} \sim F(k, N - 2k)$$

$$\text{Si } F^* > F_{\alpha}(k, N - 2k) \quad \underline{\text{Rejet}}$$

$$F^* = \frac{(1955,98 - (921,08 + 881,62)) / 11}{(921,08 + 881,62) / (262 - 2 \times 11)} = 1,89$$

$$F_{\alpha}(k, N - 2k) = 1,79 \Rightarrow F^* > F_{\alpha}(k, N - 2k) \Rightarrow \text{rejet de } H_0$$

On rejette l'hypothèse de stabilité des paramètres

Il y a effectivement une diff significative entre les paramètres estimés pour le sous-échantillon d'Aix et le sous-échantillon de Marseille.