
MQG 811 - Automne 2015

Devoir 3

À remettre seul ou en équipe de 2 le **vendredi 20 novembre avant 16h.**

Reprenons le contexte de votre devoir 1 : la base de données `Devoir3_Cirque` contient des données fictives à propos d'un cirque qui donnerait des spectacles un peu partout sur la planète. Ce contexte est inspiré des données réelles accumulées par le Cirque du Soleil, et qui sont utilisées afin de mieux comprendre tout ce qui concerne les ventes de billets de leurs spectacles.

Chaque ligne de la base de données représente une *run*, c'est-à-dire le passage du cirque dans une ville pour donner des représentations d'un spectacle pendant une certaine période de temps. Plusieurs informations ont été collectées à propos de chacune des *runs*, allant du nombre d'habitants de la ville où elle a lieu, à des informations un peu plus complexes concernant la culture du pays. Nous considérons que l'ensemble des ces *runs* constitue un **échantillon**, et non l'ensemble complet des *runs* de ce cirque.

Pour la description complète des variables, voir l'énoncé du devoir 1.

Partie 1

Le but de cette première partie est de faire un premier modèle très simple pour prévoir le nombre de billets qui sont vendus lors d'une *run*. Utilisez les sorties pertinentes des figures 1, 2, 3, 4 et 5 pour répondre aux questions suivantes.

1. Commencez par une brève analyse descriptive des 4 variables concernées. Ne faites pas de rédaction, ne faites qu'examiner les sorties, elles pourraient vous être utiles pendant les analyses.
2. Parmi les variables `NombreJoursVente`, `NbPerformances` et `Marketing`, laquelle semble avoir le lien le plus fort avec le nombre de billets vendus ? Justifiez.
3. Est-ce que le lien entre le nombre de billets vendus et la variable identifiée en 2 semble linéaire ? Justifiez.
4. Quelle proportion de la variation du nombre de billets vendus est expliquée par la variable identifiée en 2 ?

5. Écrivez l'équation de la droite de régression qui a la variable identifiée en 2 comme variable explicative, et interprétez-la.
6. Utilisez l'équation écrite en 5 pour prévoir le nombre de billets vendus lorsque la variable explicative a une valeur de 150.

Statistiques descriptives					
	N	Minimum	Maximum	Moyenne	Ecart type
Billets	323	36915,00	316764,00	162443,1703	48146,88307
NombreJoursVente	323	150	314	225,37	22,038
NbPerformances	323	36	113	61,11	13,554
Marketing	323	513,33	2958,22	1220,6017	497,01178
N valide (liste)	323				

		Corrélations			
		Billets	NombreJoursVente	NbPerformances	Marketing
Billets	Corrélation de Pearson	1	,460**	,657**	,485**
	Sig. (bilatérale)		,000	,000	,000
	N	323	323	323	323
NombreJoursVente	Corrélation de Pearson	,460**	1	,515**	,115*
	Sig. (bilatérale)	,000		,000	,038
	N	323	323	323	323
NbPerformances	Corrélation de Pearson	,657**	,515**	1	,867**
	Sig. (bilatérale)	,000	,000		,000
	N	323	323	323	323
Marketing	Corrélation de Pearson	,485**	,115*	,867**	1
	Sig. (bilatérale)	,000	,038	,000	
	N	323	323	323	323

** La corrélation est significative au niveau 0,01 (bilatéral).

* La corrélation est significative au niveau 0,05 (bilatéral).

Figure 1 –

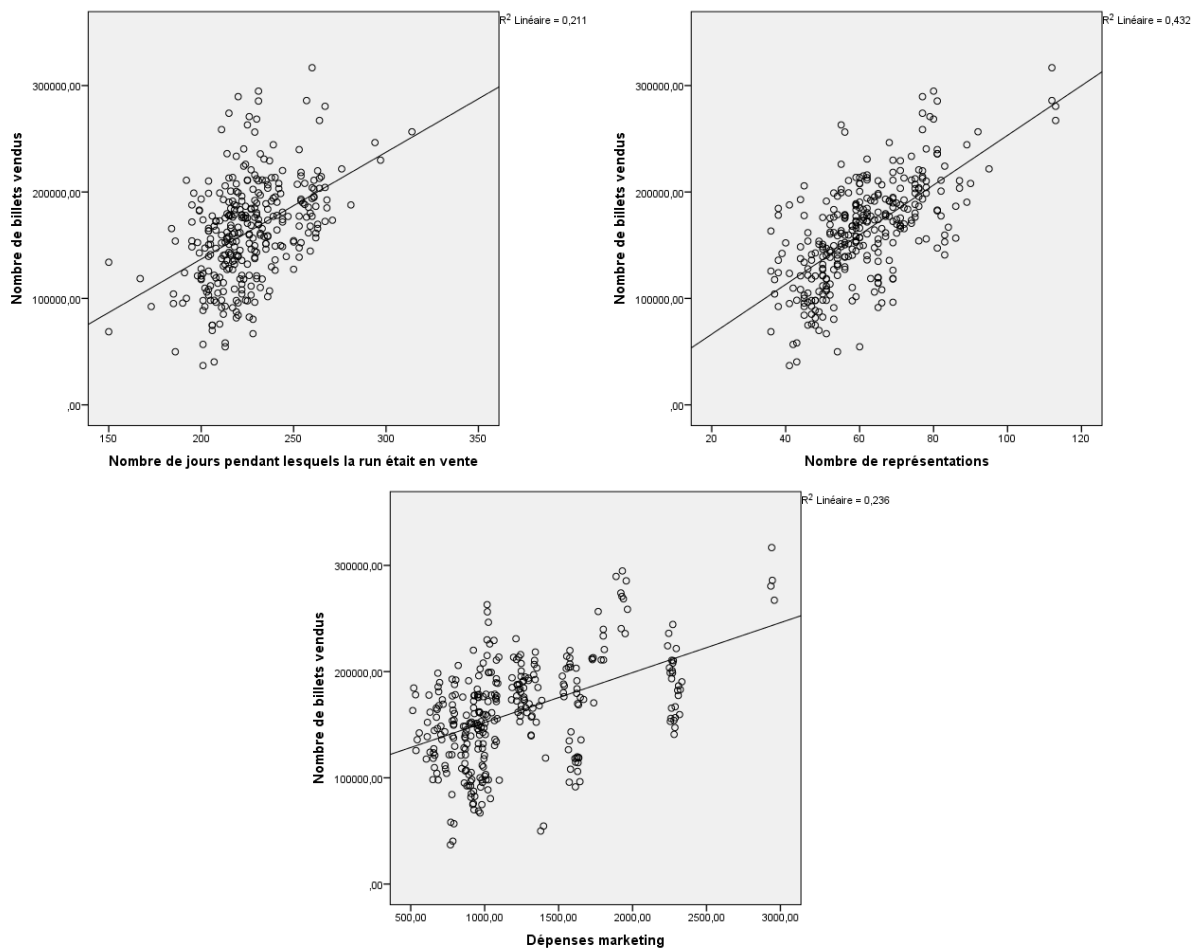


Figure 2 –

Récapitulatif des modèles^b

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,460 ^a	,211	,209	42825,16318

a. Prédicteurs : (Constante), NombreJoursVente

b. Variable dépendante : Billets

ANOVA^a

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	1,577E+11	1	1,577E+11	86,000	,000 ^b
	Résidu	5,887E+11	321	1833994601		
	Total	7,464E+11	322			

a. Variable dépendante : Billets

b. Prédicteurs : (Constante), NombreJoursVente

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	Intervalle de confiance à 95,0% pour B		Statistiques de colinéarité	
		B	Erreur standard	Bêta			Borne inférieure	Borne supérieure	Tolérance	VIF
1	(Constante)	-63880,022	24521,141		-2,605	,010	-112122,5	-15637,58		
	NombreJoursVente	1004,250	108,291	,460	9,274	,000	791,200	1217,300	1,000	1,000

a. Variable dépendante : Billets

Figure 3 –

Récapitulatif des modèles^b

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,657 ^a	,432	,430	36339,60087

a. Prédicteurs : (Constante), NbPerformances

b. Variable dépendante : Billets

ANOVA^a

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	3,225E+11	1	3,225E+11	244,239	,000 ^b
	Résidu	4,239E+11	321	1320566591		
	Total	7,464E+11	322			

a. Variable dépendante : Billets

b. Prédicteurs : (Constante), NbPerformances

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	Intervalle de confiance à 95,0% pour B		Statistiques de colinéarité	
		B	Erreur standard	Bêta			Borne inférieure	Borne supérieure	Tolérance	VIF
1	(Constante)	19757,689	9351,26		2,113	,035	1360,2	38155,181		
	NbPerformances	2335,077	149,415	,657	15,628	,000	2041,1	2629,033	1,000	1,000

a. Variable dépendante : Billets

Figure 4 –

Récapitulatif des modèles^b

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,485 ^a	,236	,233	42161,21259

a. Prédicteurs : (Constante), Marketing

b. Variable dépendante : Billets

ANOVA^a

Modèle		Somme des carrés	ddl	Carré moyen	F	Sig.
1	Régression	1,758E+11	1	1,758E+11	98,919	,000 ^b
	Résidu	5,706E+11	321	1777567847		
	Total	7,464E+11	322			

a. Variable dépendante : Billets

b. Prédicteurs : (Constante), Marketing

Coefficients^a

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	Intervalle de confiance à 95,0% pour B		Statistiques de colinéarité	
		B	Erreur standard	Bêta			Borne inférieure	Borne supérieure	Tolérance	VIF
1	(Constante)	105053,544	6228,865		16,866	,000	92798,990	117308,099		
	Marketing	47,017	4,727	,485	9,946	,000	37,717	56,318	1,000	1,000

a. Variable dépendante : Billets

Figure 5 –

Partie 2

On vous demande maintenant de faire un modèle un peu plus complexe et plus performant afin de prévoir le nombre de billets qui se vendent lors d'une *run*.

Vous commencez par faire un modèle de régression linéaire multiple ayant 4 variables explicatives : `NombreJoursVente`, `NbPerformances`, `LongTermOrientation` et `AppreciationShow`.

Faites les étapes suivantes au seuil $\alpha = 0,05$:

1. Générez et fournissez les statistiques descriptives des variables `AppreciationShow` et `LongTermOrientation`. Ne faites aucune rédaction, ne faites qu'examiner cette sortie, elle pourrait vous être utile pendant les analyses.
2. Générez (et fournissez) toutes les sorties du modèle de régression linéaire multiple avec les quatre variables explicatives, incluant les sorties concernant les résidus, et sauvegardez les résidus standardisés (ceci est nécessaire pour produire le test de normalité, qui fait partie des sorties concernant les résidus).
3. Procédons maintenant à l'analyse et l'utilisation du modèle généré en 2 :
 - (a) Faites l'analyse de la validité de ce modèle.
 - (b) Du point de vue de l'explication de la variation du nombre de billets vendus, quel est le gain de ce modèle relativement à celui étudié à la partie 1 ?
 - (c) Est-ce que cette régression multiple est significative ? Rédigez tous les détails pour répondre à cette question.
 - (d) Écrivez l'équation du modèle, et interprétez-la. Est-il possible d'interpréter la constante en tant que prévision du modèle ? Justifiez.
 - (e) Estimez ponctuellement et par intervalle de confiance de niveau 95 % le nombre de billets vendus en moyenne pour des *runs* ayant 100 performances, ayant obtenu une appréciation moyenne de 8 sur 10, dont les billets ont été en vente pendant 125 jours, et dans une région ayant un score de 60 pour le *Long Term Orientation*. Il n'est pas obligatoire de fournir une capture d'écran.
 - (f) Calculez le MAD et le MAPE de ce modèle, et interprétez-les (simplement expliquer ce qu'ils signifient).

-
- (g) (Bonus) Supposons que ce modèle soit jugé assez précis pour qu'un gestionnaire décide de l'utiliser afin de prévoir combien de billets pourront être vendus pour une *run* qui commence dans deux mois. Pensez-vous que ce modèle pourra être utilisé tel quel ? Justifiez.

Exigences de présentation du travail :

- ➡ Identifiez bien chaque partie du travail selon l'énumération de l'énoncé.
- ➡ Fournissez, en annexe, les sorties qui concernent les informations qui vous sont demandées. Numérotez-les, et faites référence à ces sorties dans votre rapport dès que vous utilisez une information contenue dans l'une d'elles.
- ➡ Pour copier-coller les sorties SPSS dans un document Word, cliquez avec le bouton droit de la souris sur la sortie (tableau ou graphique), sélectionnez *Copier spécial...*, et assurez-vous que seul le format Image (JPG, PNG) est coché. Ensuite cliquez sur OK, puis copiez l'image dans votre document. Ce format permet de facilement changer la dimension de la sortie.
- ➡ Les sorties et tableaux ne doivent pas chevaucher deux pages.