Régression multiple

Généralités

La procédure de régression multiple de l'Assistant est ajustée aux modèles linéaires et quadratiques avec cinq prédicteurs maximum (X) et une réponse continue (Y) à l'aide de l'estimation des moindres carrés. L'utilisateur choisit le type du modèle et l'Assistant en sélectionne les termes. Dans ce document, nous expliquons les critères que l'Assistant utilise pour sélectionner le modèle de régression.

En outre, nous examinons plusieurs facteurs importants pour obtenir un modèle de régression valide. Tout d'abord, l'échantillon doit être assez grand pour conférer suffisamment de puissance au test et pour fournir une précision suffisante pour l'estimation de l'importance de la relation entre X et Y. Ensuite, il est important d'identifier les données aberrantes susceptibles d'influer sur les résultats de l'analyse. Nous étudions également l'hypothèse selon laquelle le terme d'erreur suit une loi normale et évaluons l'impact de la non-normalité sur les tests d'hypothèse du modèle global.

Sur la base de ces facteurs, l'Assistant effectue automatiquement les contrôles suivants sur vos données et répertorie les résultats dans le rapport :

- Quantité de données
- Données aberrantes
- Normalité

Dans ce document, nous étudions l'importance pratique de ces facteurs dans l'analyse de régression et nous décrivons la façon dont nous avons établi notre méthode de contrôle de ces facteurs dans l'Assistant.

Méthodes de régression

Sélection du modèle

L'analyse de régression de l'Assistant est ajustée à un modèle comportant une réponse continue et deux à cinq prédicteurs. L'un des prédicteurs peut être un prédicteur de catégorie. Il existe deux types de modèles parmi lesquels choisir :

- Linéaire : $F(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$
- Quadratique : $F(x) = \beta_0 + \sum_i \beta_i X_i + \sum_i \beta_{ii} X_i^2 + \sum_{i<j} \beta_{ij} X_i X_j$

L'Assistant sélectionne les termes du modèle linéaire ou quadratique complet.

Objectif

Nous souhaitions examiner différentes méthodes susceptibles d'être utilisées pour la sélection du modèle afin de déterminer lequel utiliser dans l'Assistant.

Méthode

Nous avons examiné trois types différents de sélections de modèles : descendante, ascendante et pas à pas. Ces types de sélections de modèles comprennent plusieurs options que nous avons également examinées, notamment :

- Les critères utilisés pour ajouter des termes au modèle ou en supprimer.
- La nécessité ou non de forcer l'intégration de certains termes dans le modèle ou d'inclure certains termes dans le modèle initial.
- La hiérarchie des modèles.
- La normalisation des variables X dans le modèle.

Nous avons examiné ces options, étudié leur effet sur les résultats de la procédure et nous sommes intéressés aux méthodes qui avaient les faveurs des spécialistes.

Les résultats

Voici la procédure que nous avons utilisée pour sélectionner les termes du modèle dans l'Assistant :

- La sélection de modèle pas à pas est utilisée. Souvent, un ensemble de variables X potentielles sont corrélées, de telle sorte que l'effet d'un terme dépendra des autres termes également présents dans le modèle. La sélection pas à pas représente sans doute la meilleure approche dans ces conditions, car elle permet de saisir les termes lors d'une étape, et de les supprimer ultérieurement, selon les autres termes inclus dans le modèle.
- La hiérarchie du modèle est maintenue à chaque étape et plusieurs termes peuvent être inclus dans le modèle lors de la même étape. Par exemple, si le terme le plus significatif est X_1^2, il est inclus avec X_1, que X_1 soit significatif ou non. La hiérarchie est souhaitable car elle permet de traduire le modèle d'unités normalisées en unités non

normalisées. De plus, la hiérarchie permettant d'inclure plusieurs termes dans le modèle à chaque étape, il est possible d'identifier un terme d'interaction ou quadratique important, même si le terme linéaire associé n'est pas étroitement lié à la réponse.

- Les termes sont saisis ou supprimés du modèle en fonction de la valeur $\alpha = 0{,}10$. L'utilisation de $\alpha = 0{,}10$ rend la procédure plus sélective que la procédure pas à pas dans Minitab, qui utilise la formule $\alpha = 0{,}15$.
- A des fins de sélection des termes du modèle, l'utilisateur normalise les prédicteurs en leur soustrayant la moyenne et en divisant par l'écart type. Le modèle final s'affiche en unités de X non normalisés. La normalisation des X supprime la plupart de la corrélation entre les termes linéaires et quadratiques, ce qui diminue les risques d'ajouter inutilement des termes d'ordre supérieur.

Vérifications de données

Quantité de données

La puissance fait référence à la probabilité qu'un test d'hypothèse rejette l'hypothèse nulle, lorsqu'elle est fausse. Dans le cas de la régression, l'hypothèse nulle stipule l'absence de relation entre X et Y. Si l'ensemble de données est trop petit, la puissance du test peut ne pas être suffisante pour détecter une relation entre X et Y qui existe réellement. Par conséquent, l'ensemble de données doit être assez grand pour détecter une relation importante d'un point de vue pratique, avec une probabilité élevée.

Objectif

Nous souhaitions déterminer l'impact de la quantité de données sur la puissance du test F global de la relation entre X et Y, ainsi que sur la précision de R^2_{ajust}, qui représente l'estimation de l'importance de la relation entre X et Y. Ces informations sont essentielles. Elles permettent de déterminer si l'ensemble de données est assez grand pour considérer l'importance de la relation observée dans les données comme un indicateur fiable de l'importance sous-jacente réelle de la relation. Pour plus d'informations sur R^2_{ajust}, reportez-vous à l'Annexe A.

Méthode

Nous avons adopté une approche similaire pour déterminer l'effectif d'échantillon recommandé que nous avons utilisé pour une régression simple. Nous avons examiné la variabilité des valeurs R^2_{ajust} pour déterminer l'effectif d'échantillon recommandé pour que la valeur R^2_{ajust} soit proche de ρ^2_{ajust}. Nous avons également vérifié que l'effectif d'échantillon recommandé conférait une puissance raisonnable même lorsque l'importance de la relation entre les variables Y et X est relativement faible. Pour plus d'informations sur les calculs, reportez-vous à l'Annexe B.

Les résultats

Comme dans le cas de la régression simple, nous recommandons un échantillon assez grand pour que vous puissiez être sûr à 90 % que la valeur observée de R^2_{ajust} se trouvera dans un écart type de 0,20 à partir de ρ^2_{ajust} . Nous avons constaté que l'effectif d'échantillon nécessaire augmente à mesure que vous ajoutez des termes au modèle. Par conséquent, nous avons calculé l'effectif d'échantillon nécessaire pour chaque taille de modèle. L'effectif recommandé est arrondi au multiple de 5 le plus proche. Par exemple, si le modèle compte huit coefficients en plus de la constante, notamment quatre termes linéaires, trois termes d'interaction et un terme quadratique, l'effectif d'échantillon minimum nécessaire pour répondre au critère est n = 49. L'Assistant arrondit ceci à un effectif d'échantillon recommandé de n = 50. Pour plus d'informations sur les recommandations d'effectif d'échantillon spécifiques en fonction du nombre de termes, reportez-vous à l'Annexe B.

Nous avons également vérifié que les effectifs d'échantillons recommandés offrent une puissance suffisante. Nous avons constaté que, pour des relations relativement faibles, ρ^2_{ajust} = 0,25, la puissance est en général d'environ 80 % minimum. Par conséquent, si vous suivez les recommandations de l'Assistant concernant l'effectif d'échantillon, vous aurez la garantie de bénéficier d'une puissance relativement satisfaisante et d'une bonne précision pour l'estimation de l'importance de la relation.

Sur la base de ces résultats, l'Assistant affiche les informations suivantes dans le Rapport lorsqu'il vérifie la quantité de données :

Etat	Condition
	**Effectif d'échantillon < recommandation** L'effectif d'échantillon n'est pas assez grand pour permettre une estimation très précise de l'importance de la relation. Les mesures de l'importance de la relation, telles que R carré et R carré (ajusté), peuvent varier énormément. Pour obtenir une estimation précise, il est nécessaire d'utiliser des échantillons plus grands pour un modèle de cette taille. **Effectif d'échantillon ≥ recommandation** L'échantillon est assez grand pour permettre une estimation précise de l'importance de la relation.

Données aberrantes

Dans la procédure de régression de l'Assistant, nous définissons les données aberrantes comme des observations ayant des valeurs résiduelles normalisées importantes ou des valeurs à effet de levier importantes. Ces mesures permettent en général d'identifier les données aberrantes dans l'analyse de régression (Neter et al., 1996). Les données aberrantes pouvant avoir une forte influence sur les résultats de l'analyse, il peut être nécessaire de corriger les données pour que l'analyse soit valide. Toutefois, les données aberrantes peuvent également résulter de la variation naturelle du procédé. Par conséquent, il est important d'identifier la cause du comportement aberrant afin de déterminer comment traiter ces points de données.

Objectif

Nous souhaitions déterminer l'importance des valeurs résiduelles normalisées et des valeurs à effet de levier nécessaire pour qu'un point de données aberrant puisse être signalé.

Méthode

Nous avons élaboré nos indications concernant l'identification d'observations aberrantes sur la base de la procédure de régression standard de Minitab (**Stat** > **Régression** > **Régression**).

Les résultats

VALEURS RESIDUELLES NORMALISEES

La valeur résiduelle normalisée est égale à la valeur résiduelle, e_i, divisée par une estimation de son écart type. Habituellement, une observation est considérée comme aberrante si la

valeur absolue de la valeur résiduelle normalisée est supérieure à 2. Néanmoins, cette valeur est quelque peu prudente. En général, environ 5 % de la totalité des observations répondent à ce critère du simple fait du hasard (si les erreurs sont distribuées normalement). Par conséquent, il est important d'étudier l'origine du comportement aberrant pour déterminer si une observation est réellement aberrante.

VALEUR A EFFET DE LEVIER

Les valeurs à effet de levier sont uniquement liées à la valeur X d'une observation et ne dépendent pas de la valeur Y. Une observation est déterminée comme aberrante si la valeur à effet de levier est égale à plus de 3 fois le nombre de coefficients de modèle (p), divisée par le nombre d'observations (n). Encore une fois, il s'agit d'une valeur limite fréquemment utilisée, bien que certains manuels utilisent $\frac{2 \times p}{n}$ (Neter et al., 1996).

Si vos données comprennent des points à effet de levier élevés, déterminez s'ils ont une influence excessive sur le modèle sélectionné pour ajuster les données. Par exemple, une valeur X extrême unique peut générer une sélection d'un modèle quadratique au lieu d'un modèle linéaire. Vous devez déterminer si la courbure observée dans le modèle quadratique correspond à votre compréhension du procédé. Si ce n'est pas le cas, ajustez un modèle plus simple aux données ou regroupez des données supplémentaires pour étudier de façon plus approfondie le procédé.

Lors du test des données aberrantes, l'Assistant affiche les indicateurs d'état suivants dans le rapport :

Etat	Condition
	Il n'y a pas de points de données aberrants.
	Il existe au moins une valeur résiduelle normalisée élevée ou au moins un point à effet de levier élevé.

Normalité

Dans le cas de la régression, l'hypothèse générale stipule que les erreurs aléatoires (ε) sont distribuées normalement. L'hypothèse de normalité est importante lors de la réalisation de tests d'hypothèse des estimations des coefficients (β). Heureusement, même lorsque les erreurs aléatoires ne sont pas distribuées normalement, les résultats de test sont en général fiables lorsque l'échantillon est assez grand.

Objectif

Nous souhaitions déterminer l'effectif d'échantillon nécessaire pour obtenir des résultats fiables avec la loi normale. Nous souhaitions déterminer dans quelle mesure les résultats de test réels correspondaient au seuil de signification cible (alpha ou taux d'erreur de 1ère espèce) pour le test, c'est-à-dire, si le test rejetait incorrectement l'hypothèse nulle plus souvent ou moins souvent que prévu pour les différentes lois non normales.

Méthode

Pour estimer le taux d'erreur de 1ère espèce, nous avons effectué plusieurs simulations à l'aide de lois asymétriques, à queues lourdes et à queues légères qui s'écartent sensiblement de la loi normale. Nous avons effectué des simulations à l'aide d'un effectif d'échantillon de 15. Nous avons examiné le test F global pour plusieurs modèles.

Pour chaque condition, nous avons effectué 10 000 tests. Nous avons généré des données aléatoires afin que pour chaque test, l'hypothèse nulle soit vraie. Nous avons ensuite effectué les tests en utilisant un seuil de signification cible de 0,10. Nous avons compté le nombre de fois, sur 10 000, où les tests avaient rejeté l'hypothèse nulle, puis nous avons comparé cette proportion au seuil de signification cible. Si le test est adéquat, les taux d'erreur de 1ère espèce doivent être très proches du seuil de signification cible. Pour plus d'informations sur les simulations, reportez-vous à l'Annexe C.

Les résultats

Pour le test F global, la probabilité d'obtenir des résultats significatifs sur le plan statistique ne diffère pas considérablement pour chacune des lois non normales. Les taux d'erreur de 1ère espèce se trouvent tous entre 0,08820 et 0,11850, relativement près du seuil de signification cible de 0,10.

Les tests étant correctement effectués avec des échantillons relativement petits, l'Assistant ne teste pas la normalité des données. En revanche, l'Assistant vérifie l'effectif de l'échantillon et signale les effectifs d'échantillon inférieurs à 15. L'Assistant affiche les indicateurs d'état suivants dans le rapport de la régression :

Etat	Condition
	L'effectif d'échantillon est d'au moins 15. La normalité n'est donc pas un problème.
	L'effectif d'échantillon étant inférieur à 15, la normalité peut être un problème. Vous devez interpréter la valeur de p avec la plus grande vigilance. Pour les échantillons réduits, l'exactitude de la valeur de p est sensible aux erreurs résiduelles non normales.

Références

Neter, J., Kutner, M.H., Nachtsheim, C.J. et Wasserman, W. (1996), *Applied linear statistical models*, Chicago : Irwin.

Annexe A : modèle et statistiques

Un modèle de régression reliant un prédicteur X à une réponse Y prend la forme suivante :

$Y = f(X) + \varepsilon$

la fonction f(X) représentant la valeur attendue (moyenne) de Y en fonction de X.

L'Assistant propose deux formes possibles de fonction f(X) :

Type de modèle	f(X)
Linéaire	$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$
Quadratique	$\beta_0 + \sum_i \beta_i X_i + \sum_i \beta_{ii} X_i^2 + \sum_{i<j} \beta_{ij} X_i X_j$

Les valeurs des coefficients β sont inconnues et doivent être estimées à partir des données. La méthode d'estimation est celle des moindres carrés, qui minimise la somme des valeurs résiduelles quadratiques dans l'échantillon :

$$\min \sum_{i=1}^{n} \left(Y_i - \hat{f}(X_i)\right)^2.$$

Une valeur résiduelle correspond à la différence entre la réponse observée Y_i et la valeur ajustée $\hat{f}(X_i)$ en fonction des coefficients estimés. La valeur minimisée de cette somme des carrés est la SCE (somme des carrés d'erreur) d'un modèle donné.

Test F global

Cette méthode teste le modèle global (linéaire ou quadratique). Pour la forme sélectionnée de fonction de régression f(X), elle teste :

$H_0 : f(X) est\ constant$

$H_1 : f(X)\ n'est\ pas\ constant$

R^2 ajusté

Le R^2 (R^2_{ajust}) ajusté mesure le degré de variabilité de la réponse attribué à X par le modèle. Il existe deux moyens fréquents de mesurer l'importance de la relation observée entre X et Y :

$$R^2 = 1 - \frac{SCE}{STC}$$

Et

$$R^2_{ajust} = 1 - \frac{SCE/(n-p)}{STC/(n-1)}$$

où

$$\text{STC} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

STC désigne la somme totale des carrés, qui mesure la variation des réponses autour de leur moyenne globale $\bar{Y}$. La valeur SCE mesure leur variation autour de la fonction de régression f(X). L'ajustement de la valeur R^2_{ajust} correspond au nombre de coefficients (p) dans le modèle complet, ce qui laisse n – p degrés de liberté pour estimer la variance de ε. La valeur R^2 ne diminue jamais lorsque davantage de coefficients sont ajoutés au modèle. Toutefois, du fait de l'ajustement, la valeur R^2_{ajust} peut diminuer lorsque les coefficients supplémentaires n'améliorent pas le modèle. Ainsi, si l'ajout d'un autre terme au modèle n'explique pas la variance supplémentaire dans la réponse, la valeur R^2_{ajust} diminue, indiquant que le terme supplémentaire n'est pas utile. Par conséquent, la mesure ajustée doit être utilisée à des fins de comparaison de modèles de différentes tailles.

Relation entre le test F et R^2_{ajust}

La statistique F pour le test du modèle global peut être exprimée en termes de SCE et de STC, qui sont également utilisées dans le calcul de R^2_{ajust}:

$$F = \frac{(STC - SCE)/(p-1)}{SCE/(n-p)}$$

$= 1 + \left(\frac{n-1}{p-1}\right)\frac{R^2_{ajust}}{1-R^2_{ajust}}$. Les formules ci-dessus montrent que la statistique F est une fonction croissante de R^2_{ajust}. Ainsi, le test rejette l'hypothèse H_0 si et seulement si la valeur R^2_{ajust} est supérieure à une valeur spécifique déterminée par le seuil de signification (α) du test.

Annexe B : quantité de données

Dans cette section, nous examinons l'influence de la valeur n, le nombre d'observations, sur la puissance du test de modèle global et sur la précision de R^2_{ajust}, l'estimation de l'importance du modèle.

Pour quantifier l'importance de la relation, nous introduisons une nouvelle quantité, ρ^2_{ajust}, en tant qu'équivalent de la population de la statistique d'échantillon R^2_{ajust}. Souvenez-vous que

$$R^2_{ajust} = 1 - \frac{SCE/(n-p)}{STC/(n-1)}$$

Nous définissons donc

$$\rho^2_{ajust} = 1 - \frac{E(SCE|X)/(n-p)}{E(STC|X)/(n-1)}$$

L'opérateur E(·|X) indique la valeur attendue, ou la moyenne d'une variable aléatoire en fonction de la valeur de X. En supposant que le modèle correct est $Y = f(X) + \varepsilon$ avec une valeur ε indépendante distribuée de façon identique, nous avons

$$\frac{E(SCE|X)}{n-p} = \sigma^2 = Var(\varepsilon)$$

$$\frac{E(STC|X)}{n-1} = \sum_{i=1}^{n} \frac{(f(X_i) - \bar{f})^2}{(n-1)} + \sigma^2$$

$\bar{f} = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$.

D'où,

$$\rho^2_{ajust} = \frac{\sum_{i=1}^{n} \left(f(X_i) - \bar{f}\right)^2/(n-1)}{\sum_{i=1}^{n} \left(f(X_i) - \bar{f}\right)^2/(n-1) + \sigma^2}$$

Signification du modèle global

Lorsque nous testons la signification statistique du modèle global, nous supposons que les erreurs aléatoires ε sont indépendantes et distribuées normalement. Ensuite, sous l'hypothèse nulle selon laquelle la moyenne de la valeur Y est constante ($f(X) = \beta_0$), la statistique du test F a une distribution $F(p-1, n-p)$. Dans l'hypothèse alternative, la statistique F a une distribution $F(p-1, n-p, \theta)$ non centrale avec un paramètre de non-centralité :

$$\theta = \sum_{i=1}^{n} \left(f(X_i) - \bar{f}\right)^2/\sigma^2$$

$$= \frac{(n-1)\rho^2_{ajust}}{1 - \rho^2_{ajust}}$$

La probabilité de rejet de l'hypothèse H_0 augmente avec le paramètre de non-centralité, dont les valeurs n et ρ^2_{ajust} augmentent.

Importance de la relation

Comme nous l'avons démontré avec la régression simple, une relation significative sur le plan statistique au niveau des données n'indique pas nécessairement une forte relation sous-jacente entre X et Y. C'est pourquoi de nombreux utilisateurs comptent sur des indicateurs tels que R^2_{ajust} pour connaître l'importance réelle de la relation. Si nous considérons R^2_{ajust} comme une estimation de ρ^2_{ajust}, nous voulons avoir la certitude que l'estimation est relativement proche de la valeur ρ^2_{ajust} réelle.

Pour chaque taille de modèle possible, nous avons déterminé un seuil approprié pour les effectifs d'échantillons acceptables en identifiant la valeur minimale de n pour laquelle la probabilité que des différences absolues $|R^2_{ajust} - \rho^2_{ajust}|$ soient supérieures à 0,20 ne dépasse pas 10 %. Ceci est indépendant de la valeur réelle de ρ^2_{ajust}. Les effectifs d'échantillons recommandés n(T) sont récapitulés dans le tableau ci-dessous, où T correspond au nombre de coefficients du modèle autres que le coefficient constant.

T	n(T)
1-3	40
4-6	45
7-8	50
9-11	55
12-14	60
15-18	65
19-21	70
22-24	75
25-27	80
28-31	85
32-34	90
35-38	95
39-41	100
42-45	105
46-48	110
49-52	115

T	n(T)
53-56	120
57-59	125
60-63	130
64-67	135
68-70	140
71-73	145

Nous avons évalué la puissance du test F global du modèle pour une valeur relativement faible de $\rho^2_{ajust} = 0{,}25$, pour vérifier que la puissance au niveau des effectifs d'échantillons recommandés est suffisante. Les tailles de modèles figurant dans le tableau ci-dessous représentent l'hypothèse la plus pessimiste pour chaque valeur de n(T). Les modèles plus petits ayant la même valeur n(T) auront plus de puissance.

T	n(T)	Puissance à $\rho^2_{ajust} = 0,25$
3	40	0,902791
6	45	0,854611
8	50	0,850675
11	55	0,831818
14	60	0,820592
18	65	0,798003
21	70	0,796425
24	75	0,796911
27	80	0,798856
31	85	0,789861
34	90	0,794367
38	95	0,788625
41	100	0,794511
45	105	0,790864
48	110	0,797487
52	115	0,79525

T	n(T)	Puissance à $\rho^2_{ajust} = 0,25$
56	120	0,793698
59	125	0,800982
63	130	0,800230
67	135	0,799906
69	140	0,814664

Annexe C : normalité

Les modèles de régression utilisés dans l'Assistant présentent tous la même forme :

$Y = f(X) + \varepsilon$

L'hypothèse générale autour des termes aléatoires ε stipule que ce sont des variables aléatoires normales indépendantes et distribuées de façon identique avec une moyenne nulle et une variance commune σ^2. Les estimations par la méthode des moindres carrés des paramètres β constituent tout de même les meilleures estimations non biaisées linéaires, même si nous renonçons à l'hypothèse selon laquelle les valeurs ε sont distribuées normalement. L'hypothèse de normalité devient seulement importante lorsque nous tentons de lier les probabilités à ces estimations, comme nous le faisons dans les tests d'hypothèse autour de f(X).

Nous souhaitions déterminer la valeur n nécessaire pour que les résultats d'une analyse de régression en fonction de l'hypothèse de normalité soient fiables. Nous avons procédé à des simulations pour comparer les taux d'erreur de 1ère espèce des tests d'hypothèse avec différents lois de distribution non normales de l'erreur.

Le tableau 1 ci-dessous montre la proportion de simulations, parmi 10 000, dans lesquelles le test F global était significatif au niveau d'$\alpha = 0{,}10$ pour plusieurs distributions de la valeur ε dans trois modèles différents. Dans ces simulations, l'hypothèse nulle, qui indique l'absence de relation entre X et Y, était vraie. Les valeurs X ont été générées sous forme de variables normales multivariées par la commande RANDOM de Minitab. Nous avons utilisé un effectif d'échantillon de n = 15 pour tous les tests. Tous les modèles impliquaient cinq prédicteurs continus. Le premier modèle était le modèle linéaire avec les cinq variables X. Le deuxième modèle comportait l'ensemble des termes quadratiques et linéaires. Le troisième modèle possédait tous les termes linéaires et sept des interactions à 2 facteurs.

Tableau 1 Taux d'erreur de 1ère espèce pour les tests F globaux avec n = 15 pour les lois non normales

Loi de distribution	Linéaire	Linéaire + quadratique	Linéaire + 7 interactions
Normale	0,09910	0,10270	0,10060
t(3)	0,09840	0,1185	0,118
t(5)	0,09980	0,10010	0,10430
Laplace	0,09260	0,09400	0,09650
Uniforme	0,10630	0,10080	0,09480
Bêta(3, 3)	0,09980	0,10120	0,10020
Exponentielle	0,08820	0,09500	0,09960
Khi(3)	0,09890	0,114	0,10970

Loi de distribution	Linéaire	Linéaire + quadratique	Linéaire + 7 interactions
Khi(5)	0,09730	0,10590	0,10330
Khi(10)	0,10150	0,09930	0,10360
Bêta(8, 1)	0,09870	0,10230	0,10490

Les résultats de la simulation montrent que la probabilité d'obtenir des résultats statistiquement significatifs ne diffère pas considérablement de la valeur nominale de 0,10 pour chacune des distributions d'erreurs. Les taux d'erreur de 1ère espèce observés se trouvent tous entre 0,08820 et 0,11850.