

Question sur la régression (30 points)

Un agronome a testé, deux fois pour chaque dosage (en Kg/ha), l'effet d'un engrais («Fertilizer») sur la récolte de tomates («Yield») dans un certain nombre de serres (N=14) choisies au hasard dans la même région. Il a obtenu, pour des dosages de 10 à 110 kg/ha des récoltes de 6kg à 58kg/are, pour une semaine de récolte pendant la période de maturité (2 mois après la plantation) des plants de la même variété plantés à la même date. La Fig. 1 montre graphiquement les résultats.

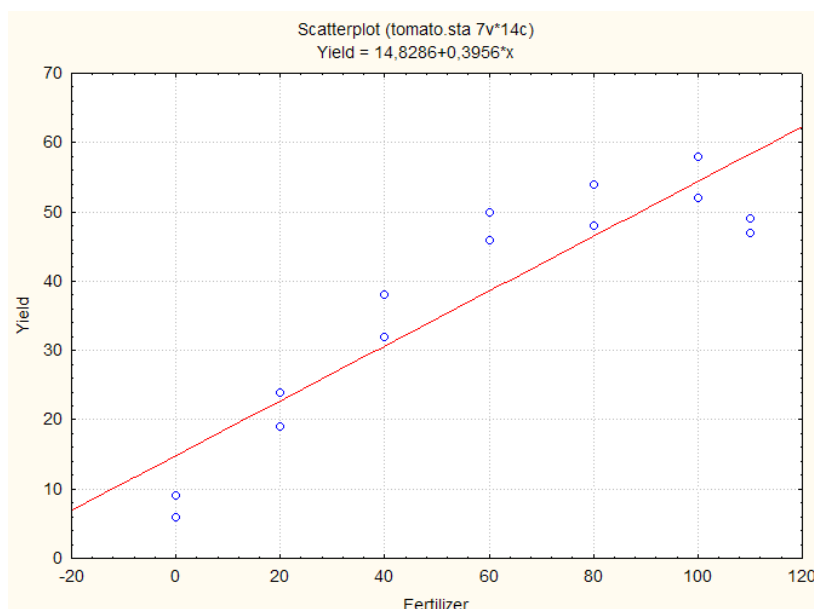


Fig.1 Rendement («Yield») en kg/are pour chaque parcelle de 100m² testée, pour des dosages d'engrais de 0 à 11kg d'engrais à l'hectare.

(a) (2 points) Au vu de ces données, pourriez-vous penser que le modèle linéaire est valide? Justifications requises.

Rép. On devrait penser à un modèle quadratique, i.e. utilisant le carré du dosage, vu la forme de la courbe prévisible.

Après divers essais de modèles, il fut décidé de garder le modèle dont les résultats sont les suivants (Tab. 1). Les résidus ont montré que le modèle serait valide.

Tab. 1 Tableau Anova et tableau des résultats (partiels) de la régression sur le modèle indiqué (Tableaux Statistica).

	Sums of Squares	Df	Mean Squares	F	p-level
Regress.	(a)	(b)	(c)	(d)	(e)
Residual	142,631	11	12,966		
Total	3820,000				

	Beta	Std.Err.	B	Std.Err.	t(11)	p-level
Intercept			5,864383	2,265743	2,58828	0,025215
Fertilizer	2,26123	0,221243	0,984030	0,096279	10,22059	0,000001
Fert2	-1,40162	0,221243	-0,005235	0,000826	(f)	(g)

(b) **(10 points)** Complétez les données manquantes dans ces deux tableaux. 7 items notés de (a) à (g). Indiquez aussi la valeur du R^2 de la régression. Les valeurs des probabilités de dépassement (e) et (g) sont forcément approximatives (mais suffisantes pour conclure).

Rép. (a) 3677,37 ; (b) 2 ; (c) 1638,68 ; (d) 141,80 ; (e) $p = 0,0000$; (f) -6,33; (g) $p = 0,00$; $R^2=0,96$.

(c) **(5 points)** Les valeurs estimées de β_1 et β_2 sont-elles significativement non nulles au seuil de 0,05 ?

Rép. Oui, car $p < 0,05$

(d) **(5 points)** Calculez à partir de l'équation du modèle, la dose (en kg/ha) qui maximise la récolte.

Rép. La première dérivée du modèle s'annule pour $-\beta_1/2\beta_2 = 93,986$, et c'est clairement un maximum (vu le signe de β_2).

(e) **(3points)** Calculez deux intervalles de confiance de niveau 99% pour, d'une part β_1 et, d'autre part β_2 .

Rép. On calcule à partir des résultats du tableau. Pour β_1 : (0,684; 1,2834); pour β_2 : (-0,007804; - 0,00267).

(f) **(2 points)** Obtenez à partir des calculs en (c) et (d) un intervalle de confiance à 99% pour la dose optimale.

Rép. Pour le minimum on calcule 82, et pour le maximum on calcule 128,1, à partir des résultats de (d) et (e).

(g) **(3 points)** Formulez la réserve principale à l'utilisation du modèle pour les résultats en (c) et (e) au vu des données de la Fig. 1, et des résultats précédents.

Rép. La dose optimale est dans une zone où il n'y a presque aucun échantillon.