Rapport de TP

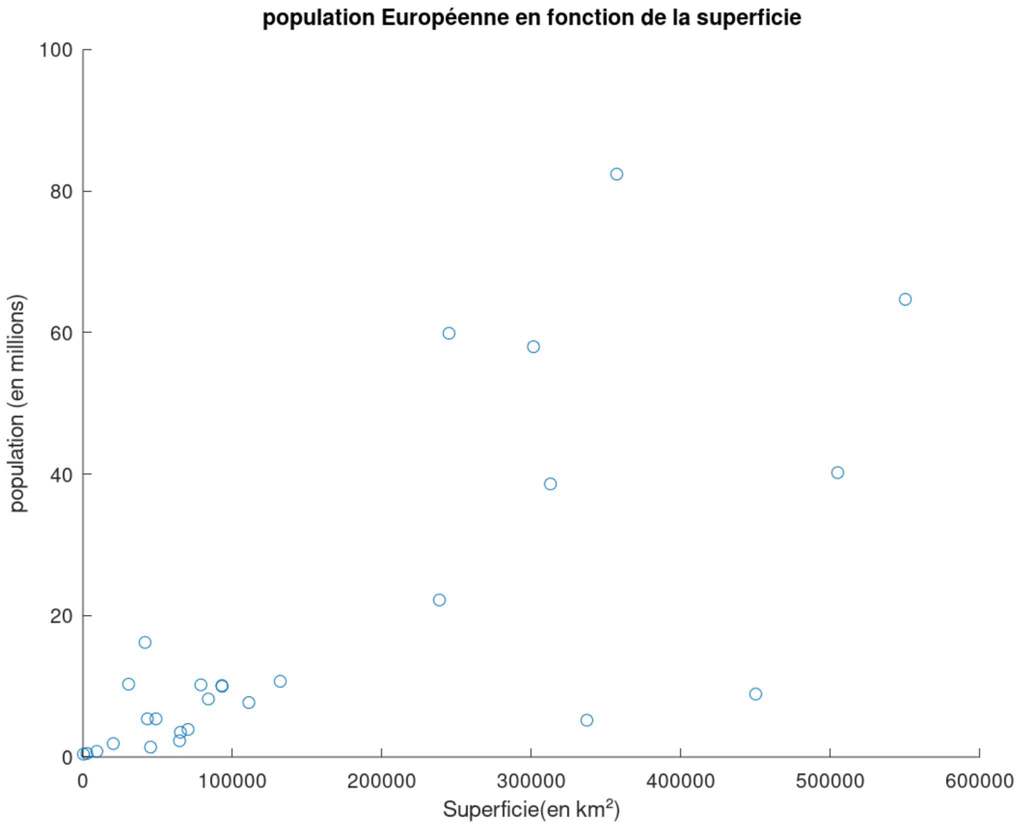
MODELISATION ET ETUDE DE CAS :
PROBABILITES ET STATISTIQUES

TERENCE SURENDRA Darvin |2021/2022

Dans le cadre des TP de modélisation et étude de cas, nous avons eu plusieurs cas de figures de régressions linéaires. Les différents cas sont les suivants : l'étude de la population européenne, l'estimation de la droite de régression, le calcul de résidus et enfin un les résultats d'une expérimentation.

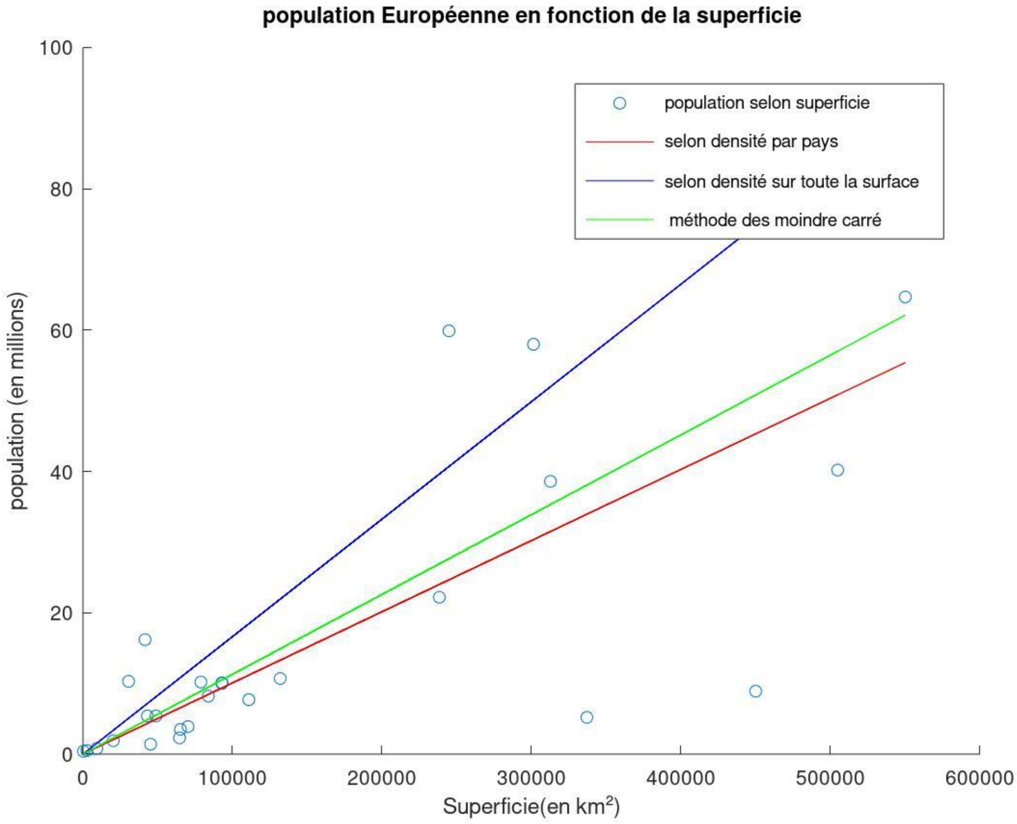
1) Etude de la population européenne

A partir du tableau donné, nous avons eu le nuage de point suivant :



Le postulat du modèle de régression linéaire n'est pas respecté, nous avons des points isolés dont la superficie n'impacte pas la population.

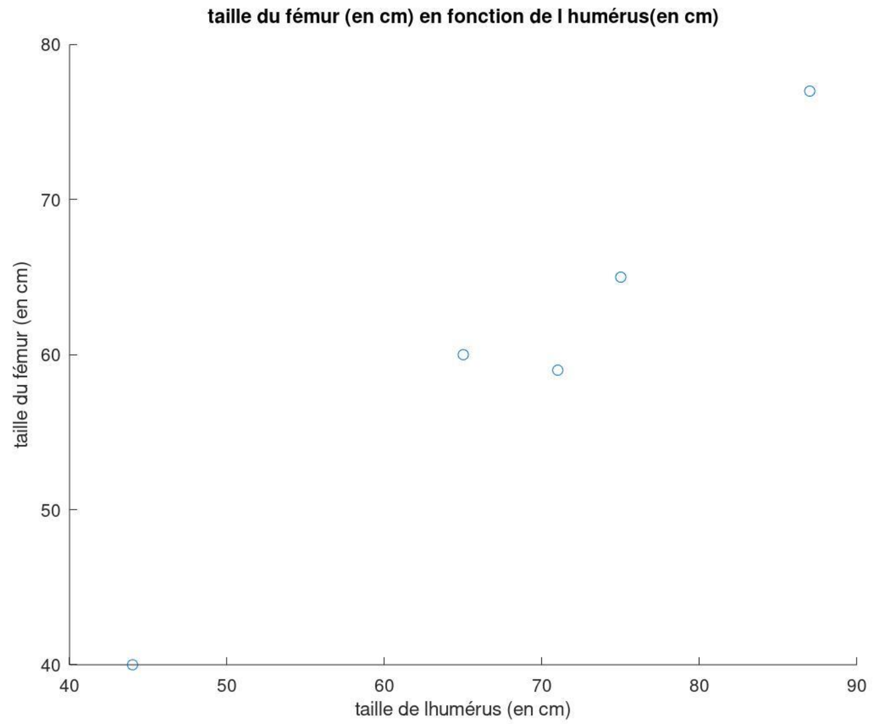
Après le calcul de chacune de ces moyennes, nous avons pu avoir ce graphique qui nous donne selon l'approche une proportionnalité par rapport à la superficie.



La meilleure approche semble être la méthode du moindre carré qui passe par le plus de point du nuage de point initial.

2)Estimation de la droite de régression

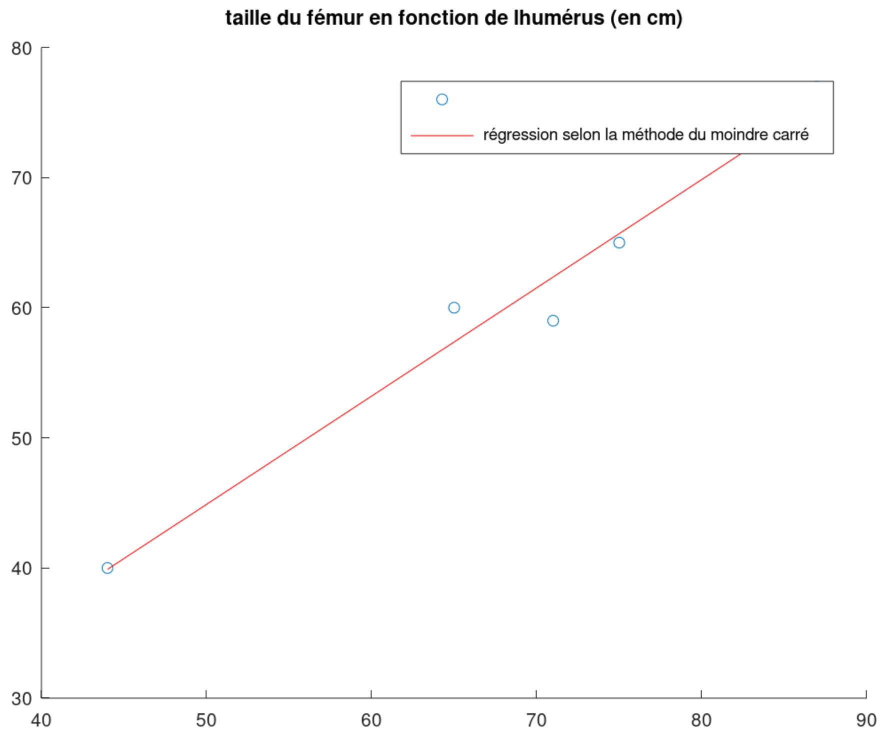
A partir des données du tableau, nous avons le nuage de point suivant :



On peut avoir une courbe qui approche tous ces points, donc une régression linéaire est envisageable.

Nous avons var(x) = 250.8 ; Var(y) = 178.70 et cov(x,y) = 208.65.

Et nous avons la régression suivante en utilisant la méthode des moindres carré :



Pour savoir si elle passe par G , on calcul a * 68.4 +b , où a = cov(x,y)/var(x) ;
et b = E[y] – a * E[x] ;

On obtient donc a * 68.4 + b = 60.2 , on en déduit que notre courbe passe bien par le point G.

Par définition, le coefficient de corrélation est r = cov(x,y)/ sqrt(var(x)var(y)).

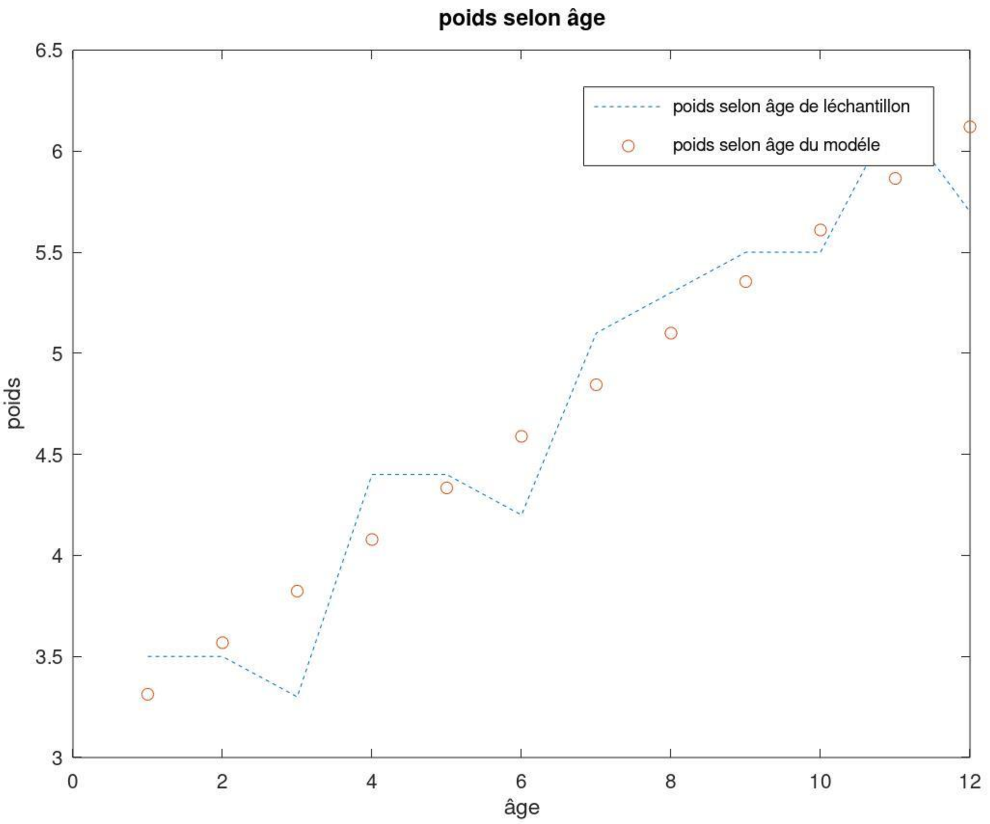
On a r = 0.98558, r est proche de 1 on en déduit qu'il y a une forte relation positive entre les deux variables.

Pour déterminer la longueur d'un fémur pour un humérus de 55cm, on calcul a * 55 + b = 49.05 .

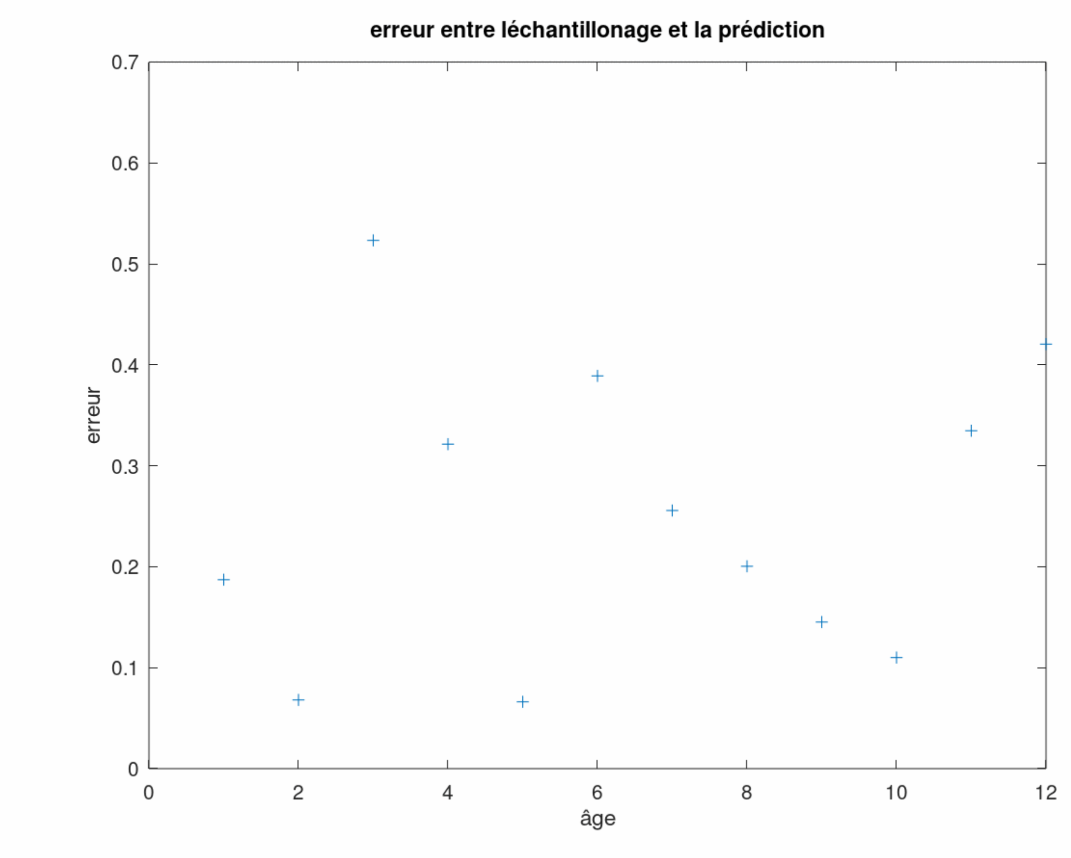
3)Calcul de résidus

A partir du tableau donnée, nous pouvons déterminer a et b. Nous avons a=cov(x,y)/var(x) et b = E[y] – a * E[x], nous avons par le calcul a=0.25524 et b = 3.0576, nous pouvons donner raison au statisticien. Toutefois, son modèle n'est guère pas pertinent si l'âge de la population qui sert d'échantillon est plus grand.

Grâce à la formule a*x+b, nous allons déterminer un poids selon l'âge par rapport à notre modèle. Nous avons le graphique suivant :



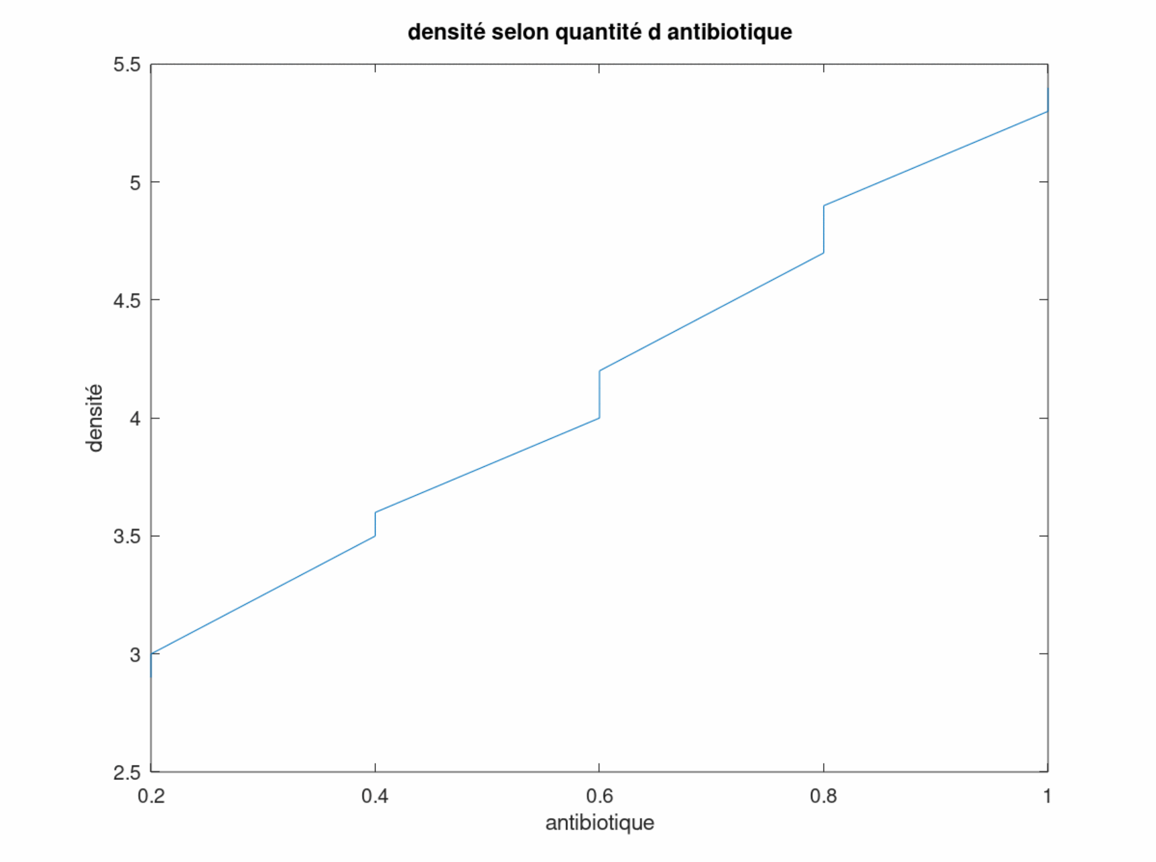
Le graphique suit les valeurs de l'échantillon.



Nous avons l'erreur ci-dessus, on remarque que les points sont dispersés donc une erreur qui est aléatoire et ne grandira donc pas par rapport à l'âge.

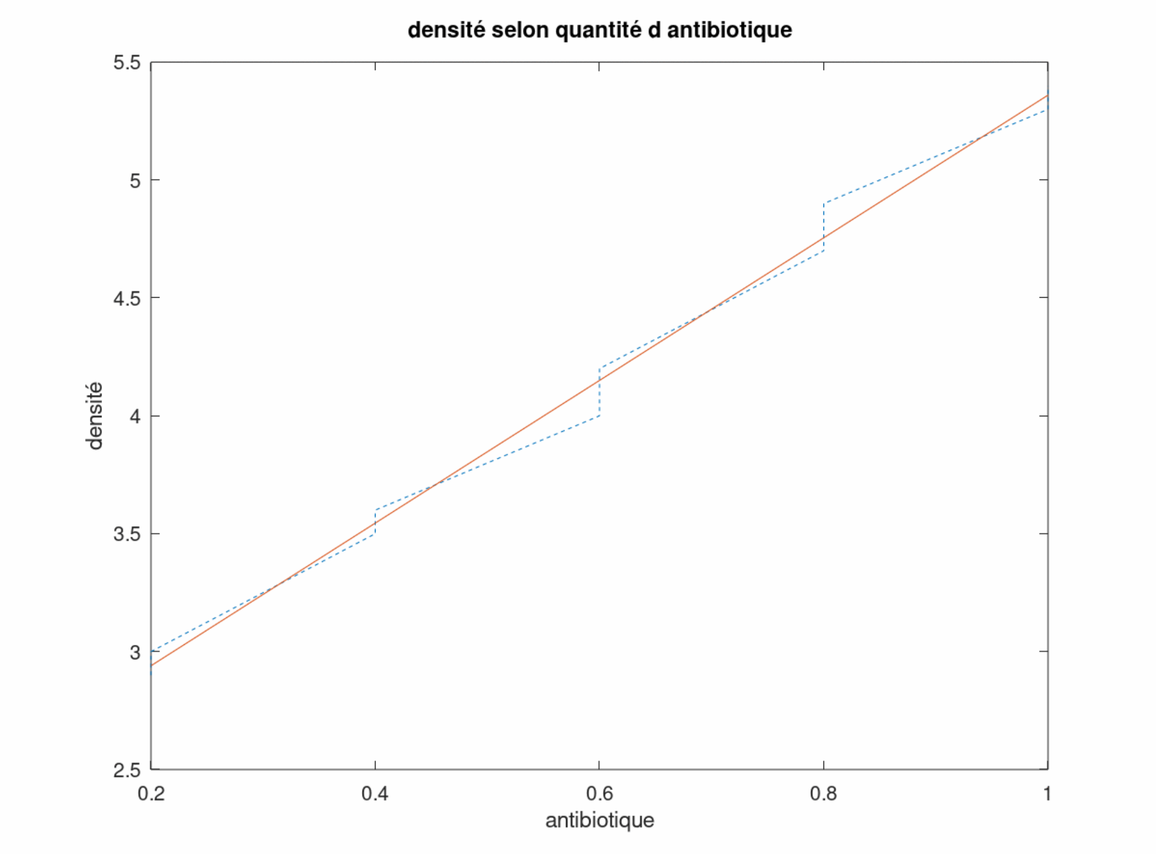
4) Résultat d'expérimentation

A partir des données du tableau nous avons le graphique suivant :



Nous pouvons déterminer l'équation de la droite si on détermine a, a=(4.15-2.355)/0.6=3.025 , nous avons donc l'équation suivante y=3.025*x+2.355 .

Nous avons le graphique suivant :



Pour déterminer Beta_0 et Beta_1 nous appliquons les formules, nous avons après calcul, Beta_0=0.0666 et Beta_1= 0.100 .

Pour vérifier que la pente soit supérieur à 3 au seuil 5%, nous vérifions la borne inf de l'intervalle où il pourrait être. Après calcul, nous avons Intervalle_min=2.8453 cela implique que la pente n'est pas nécessairement supérieure à 3.

La variation de la densité optique est proportionnel à la quantité d'antibiotique.

Le 11ème tube sera dans un intervalle entre [4.4961 ;5.6189] à 5% près.