REGRESSION

|  |  |  |
| --- | --- | --- |
| **QUESTION 1** | |  |
| **Pour la régression simple**, | Formule | |
| Total de la variabilité à expliquer pour Y= |  | |
| Écart-type de l’estimateur |  | |
|  |  | |
| **QUESTION 2** | | / |

Le fichier examRevision.xlsx contient 20 observations pour un examen. Chaque observation concerne un étudiant et reprend :

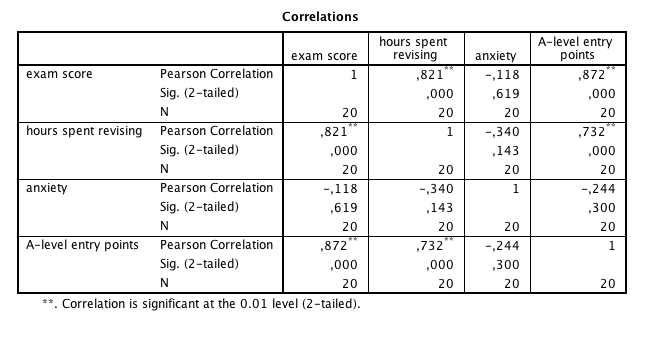
1) le score qu’il a obtenu à cet examen (cote sur 100)  
2) le nombre d’heures de révision   
3) le niveau d’anxiété  
4) des points représentant son niveau au départ.

A partir de ces données , les résultats ci-dessous ont été obtenus (cfr tableaux )

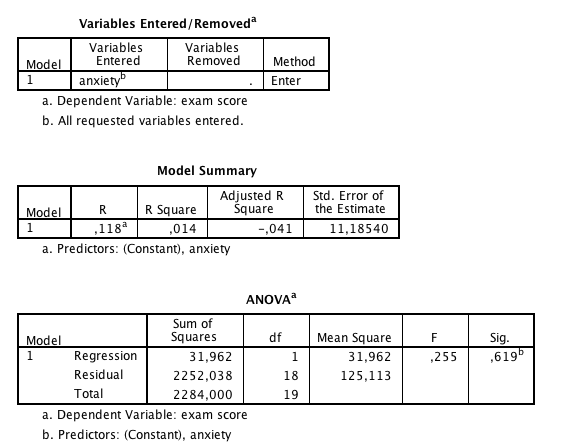
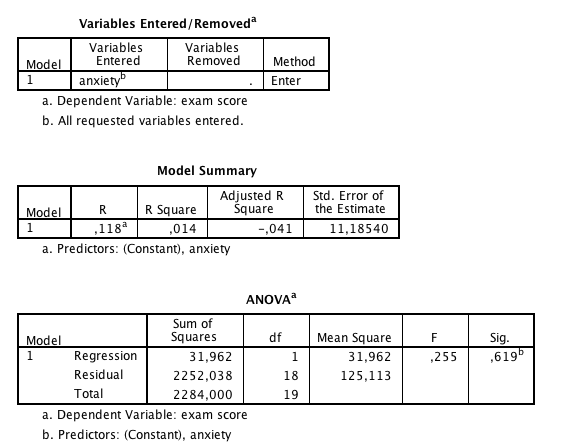
1. Quelle régression simple vous paraît la plus pertinente ? Justifiez.  
     
   Variable cible :   
   Variable explicative :   
   Justification :
2. Expliquez et interprétez contextuellement le paramètre R-square du tableau model summary de la régression score – anxiété.  
   Explication :

Interprétation contextuelle :

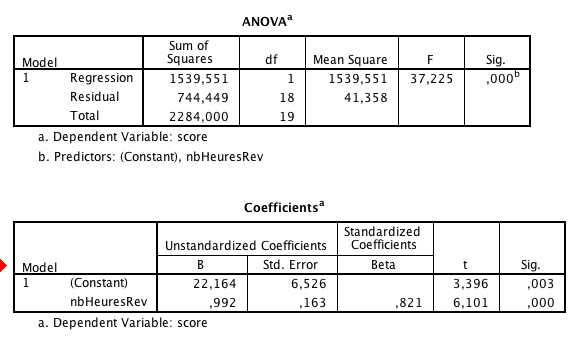
1. Pour cette même régression, expliquez et interprétez contextuellement le test F.
2. Quelle est l’équation de la régression simple score – nombre d’heures de révision ?
3. Selon le modèle proposé ci-dessous, combien d’heures de révision seraient nécessaires pour passer d’un score de 60 à un score de 70 ? Justifiez.



Quelques résultats score-anxiété



Quelques résultats score – nombre d’heures de révision



|  |  |
| --- | --- |
| **QUESTION 2 (régression linéaire et multiple)** |  |

Le fichier examRevision.xlsx contient 20 observations pour un examen. Chaque observation concerne un étudiant et reprend :

1) le score qu’il a obtenu à cet examen   
2) le nombre d’heures de révision   
3) le niveau d’anxiété  
4) des points représentant son niveau au départ.

A partir des données du fichier « examRevision.xlsx » et des résultats consignés dans le fichier Resultats-Reg.docx, on souhaite trouver un modèle de régression linéaire

* quelle sera la variable cible ?
* quelles pourraient être les variables explicatives ? Justifiez
* pour chacune des variables explicatives, testez la pertinence de la régression correspondante en calculant les paramètres les plus adéquats. Notez vos solutions et argumentez ;
* trouvez la régression qui vous paraît la plus significative et argumentez;
* calculez le SSE de cette régression et précisez ce qu’il signifie ;
* que signifie l’indicateur s (std error of the estimate) dans une régression linéaire simple ?
* combien de point(s) supplémentaire(s) à l’examen peut-on espérer en étudiant cette matière une heure de plus ? Justifiez.
* quelle est la différence principale entre l’information donnée par les tests T et celle donnée par le test F ?

|  |  |
| --- | --- |
| **QUESTION 2 (régression linéaire)** | / 10 |

A partir des données du fichier « interro.xlsx », on souhaite trouver un modèle de régression linéaire  
  
- quelle sera la variable cible ?

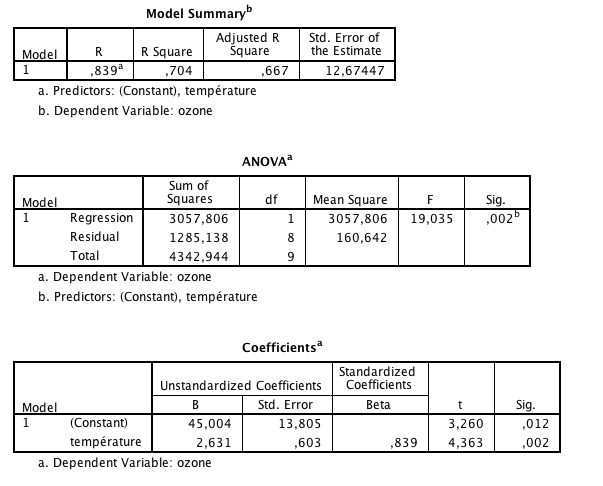
- quelles pourraient être les variables explicatives ?  
  
  
- pour chacune des variables explicatives, testez les différentes régressions possibles en calculant ;   
notez ici les solutions obtenues :

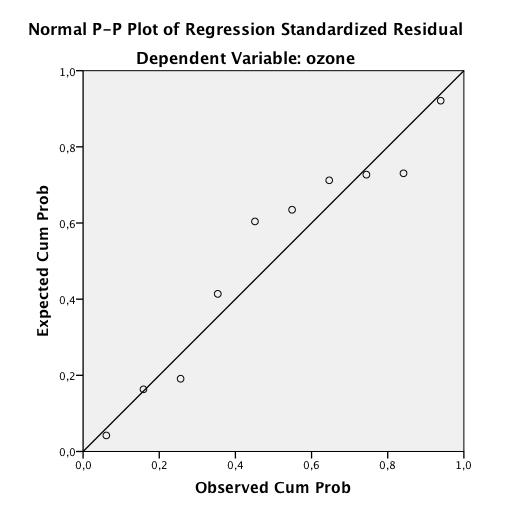
- trouvez celle qui vous paraît la plus significative et justifiez ;

- quel indicateur supplémentaire faudrait-il ajouter pour comparer les régressions ? Expliquez.

|  |  |
| --- | --- |
| **QUESTION 3 (régression linéaire)** | / 10 |

Repartons de l’exercice qui expliquait la concentration d’ozone en fonction de la température (les données se trouvent aussi dans le fichier interro.xlsx). Voici l’analyse de la régression proposée par un logiciel statistique :





* Expliquez de manière générale ce que représentent les valeurs fléchées (il y en a 3) ainsi que le graphique ci-dessus et interprétez-les dans le contexte de l’exercice température-ozone
* Les données température-ozone présente une valeur qui pourrait avoir un levier élevé, laquelle ? Expliquez.