

---

---

*EISTI - DÉPARTEMENT MATHÉMATIQUES*  
**EXAMEN DE STATISTIQUE COMPUTATIONNELLE**

10 janvier 2010 – **DURÉE 3h00**

*La consultation des documents et l'échange des documents et des calculatrices est interdit.*

*L'utilisation des 3 feuilles manuscrites recto-verso, format A4 est autorisée*

- Ne pas détacher les feuilles.
- Utiliser l'espace blanc pour vos réponses et le verso pour brouillon. Si vous avez besoin des feuilles supplémentaires pour brouillon, vous êtes priés de demander aux surveillant(e)s.
- Pensez à indiquer votre nom sur chaque feuille et votre orientation sur la 1ère page

NOM

ORIENTATION (GI, MI, SIE)

NOTE

**DÉTAIL**

	1	2	3	4	5	
Exercice 1.						
	<b>a</b>	<b>b</b>				
Exercice 2.1						
Exercice 2.2						
Exercice 2.3						
Exercice 2.4						
Exercice 2.5						
Exercice 2.6						
	<b>a</b>	<b>b</b>	<b>c</b>			
Exercice 2.7						

**Le corrigé se trouve sur Arel et sur <http://sifoci.eisti.fr>, rubrique Statistique computationnelle.**



NOM : .....

3. En prenant comme exemple la moyenne empirique justifier que les IDC sont plus étroits à mesure que la taille d'échantillon  $n$  est plus grande.

4. Expliciter la relation entre risque quadratique, variance et biais d'un estimateur. En déduire pourquoi, entre deux estimateurs sans biais, on choisira celui dont la variance est la plus petite.

NOM : .....

5. En régression linéaire, pourquoi, dans le cas de  $p = 1$  (une seule variable explicative), pour une même valeur du risque de 1ère espèce, le test (global) de Fisher et le test (de non nullité du coefficient) de Student donne le même résultat ?

NOM : .....

### Exercice 2

Depuis 2001, l'OCDE réalise tous les trois ans une enquête afin d'évaluer la capacité des jeunes de 15 ans à utiliser leurs connaissances et compétences pour relever les défis du monde réel. Cette approche reflète l'évolution des objectifs des programmes de cours : la priorité va désormais à ce que les élèves savent faire avec ce qu'ils ont appris à l'école plutôt qu'à la mesure dans laquelle ils ont assimilé des matières spécifiques. En 2009 cette enquête a été réalisée sur la totalité des 34 pays membres de l'OCDE ainsi que sur un échantillon de 31 autres pays. Elle a porté principalement sur les capacités des élèves concernant des textes écrits.

Du tableau de synthèse "Comparaison de la performance des pays et des économies" (p. 9 du document "Résultats du PISA 2009 : Synthèse", <http://www.oecd.org/home/>), nous avons extrait un tableau sur lequel pour le 65 pays, on a retenu six notes, à savoir :

- note globale de l'écrit (Ecri);
- note pour la localisation et extraction des textes (Extr);
- note pour intégrer et interpréter des textes (Intp);
- note pour réfléchir et évaluer (Eval);
- note globale pour la culture mathématique (Math), et
- note globale pour la culture scientifique (Sci).

Nous donnons ci-après les quatre premières lignes du tableau, ainsi que la ligne concernant la France.

Pays	OCDE (1 oui)	Écrit	Extract.	Interp.	Éval	Maths	Sciences
Shanghai (1)	2	556	549	558	557	600	575
Corée (2)	1	539	542	541	542	546	538
Finlande (3)	1	536	532	538	536	541	554
Hong-Kong (4)	2	533	530	530	540	555	549
...	...	...	...	...	...	...	...
France (22)	1	496	492	497	495	497	498

L'annexe fournit des tables avec des résultats que vous pouvez utiliser par la suite.

NOM : .....

TRAVAIL À FAIRE

1. On considère que la note globale de l'écrit (Ecri) est une variable aléatoire  $X_E$  d'espérance  $\mu_E$  et de variance  $\sigma_E^2$ .  
On utilise uniquement les résultats du groupe "autres pays" que vous trouverez en annexe.
  - (a) Quelle est la loi de l'estimateur que vous utilisez pour estimer  $\mu_E$ ? Faut-il émettre des hypothèses sur la variable aléatoire  $X_E$ ? Est-ce une loi exacte ou approchée? Justifier votre réponse.

NOM : .....

(b) Quelle est la valeur estimée de  $\mu_E$  sur l'échantillon ? Justifier votre réponse.

NOM : .....

2. Considérons maintenant les deux groupes des pays (pays de l'OCDE et autres pays) séparément. On prend au hasard un pays du 1er groupe on note  $X$  sa valeur pour la variable  $Ecr_i$ . On fait la même chose avec le 2e groupe et on note  $Y$  sa valeur pour la variable  $Ecr_i$ . On suppose que  $X$  et  $Y$  suivent deux lois normales différentes dont vous calculerez les paramètres.

Calculer la probabilité que la valeur de la note globale de l'écrit du pays du 1er groupe soit supérieure à la valeur correspondante du pays du 2e groupe.

NOM : .....

3. Construire pour la moyenne de la variable *Ecri* un intervalle de confiance avec seuil 95% pour les pays qui ne font pas partie de l'OCDE.

NOM : .....

4. À votre avis y a-t-il lieu de construire pour la moyenne de la variable  $E_{c,r,i}$ , un intervalle de confiance pour les pays de l'OCDE ? Justifier votre réponse.

NOM : .....

5. En parallèle avec les 31 pays qui n'appartiennent pas à l'OCDE, on choisit au hasard (tirage avec remise) 31 pays parmi le 34 de l'OCDE. Les résultats pour cet échantillon sont donnés en annexe. On cherche à savoir, avec 90% de confiance, si les deux groupes, de 31 pays chacun, sont équivalents en ce qui concerne leur moyenne.  
Nous dirons que les deux groupes ne sont pas équivalents si la différence des moyennes dépasse 15 points.  
Vérifier si les deux groupes sont équivalents.

NOM : .....

6. On souhaite étudier l'influence de la matière du couple écrit - maths sur les notes obtenues par tous les pays.  
On obtient le tableau d'analyse de variance suivant :

<b>Tableau de l'analyse de variance</b>		
<i>Variance</i>	<i>Somme des carrés</i>	<i>Degrés de liberté</i>
<i>Facteur A</i>	$SS_A = 1496$	$\nu_A = 1$
<i>Résiduelle</i>	$SS_R = 383403$	$\nu_R = 128$
<i>Totale</i>	$SS_T = 384899$	$\nu_T = 129$

Peut-on affirmer avec confiance 99% qu'il y a influence sur les notes ?

NOM : .....

7. Nous souhaitons avoir un modèle linéaire permettant d'estimer la note globale de l'écrit l' (Ecri), en fonction des variables Eval, Math et Sci. Le listing de l'annexe, fournit les résultats de toutes les régressions calculées sur tous les pays.
- (a) En examinant le graphique des résidus de la régression par rapport aux trois variables explicatives Eval, Math et Sci, donner votre avis sur la validité de cette régression. Justifier votre réponse.

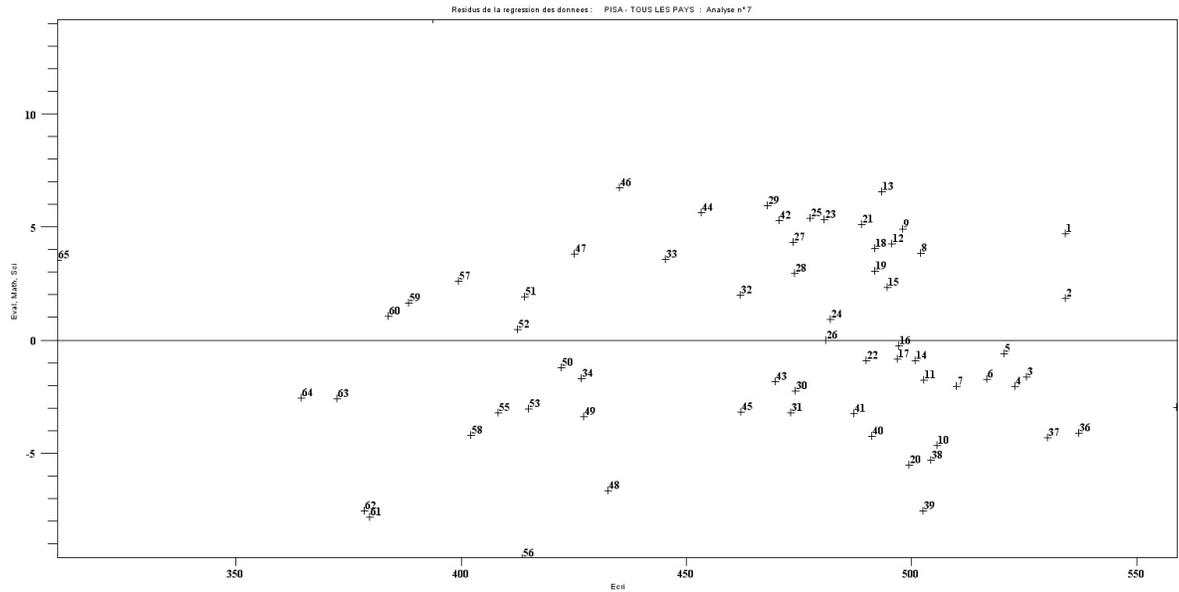


FIGURE 1 – Résidus de la régression de la note globale de l'écrit par rapport aux variables explicatives Eval, Math et Sci

NOM : .....

(b) Établir, en la justifiant, la meilleure régression (multi)-linéaire parmi celles proposées.

NOM : .....

- (c) On va utiliser les résultats de la régression de l' (Ecri) en fonction des variables Math et Sci. Calculer dans ce cas la note estimée de la France. Le résultat était-il attendu ? Justifier votre réponse.