



Mathématiques	Rattrapage de statistiques	ING 2 -GI
2017 - 2018		30/01/18 (2 h)

Calculatrice autorisée.
4 feuilles manuscrites R/V autorisées.

Exercice 1. QCM :

Indiquer, sur votre copie, la ou les lettres correspondant à la bonne ou aux bonnes réponses.

1. Le but d'un test de l'analyse de la variance, ANOVA, est de comparer :
 - A) des effectifs
 - B) des moyennes
 - C) des variances
 - D) des probabilités
2. Pour juger des qualités de deux estimateurs d'un paramètre θ , on compare :
 - A) leur biais
 - B) leur fonction de densité
 - C) leur risque quadratique

Exercice 2. Intervalle de confiance :

Une clinique a proposé une nouvelle opération chirurgicale, et a connu 40 échecs sur 200 tentatives. On note p le pourcentage de réussite de cette nouvelle opération.

1. Donner une estimation ponctuelle de p .
2. Donner un intervalle de confiance pour p de niveau de confiance 0,95.
3. Combien d'opérations la clinique devrait-elle réaliser pour connaître le pourcentage de réussite avec une précision de plus ou moins 1 %, au niveau de confiance 0,95 ?
On conservera la même estimation de la variance que précédemment.

Exercice 3. Test d'hypothèses et risques :

1. Donner la définition des risques de première et seconde espèce α et β , dans la théorie des tests statistiques.
2. Un test d'hypothèses, portant sur la moyenne μ d'une loi normale d'écart-type $\sigma = 5$, est effectué pour départager :

$$\begin{aligned}H_0 : & \quad \mu = \mu_0 = 12; \\H_1 : & \quad \mu = \mu_1 = 15.\end{aligned}$$

En se basant sur la moyenne observée \bar{x} d'un échantillon de taille $n = 25$, on décide d'appliquer la règle de décision suivante :

$$\begin{aligned} \text{si } \bar{x} < 14, & \text{ on décide } H_0 \\ \text{si } \bar{x} \geq 14, & \text{ on décide } H_1 \end{aligned}$$

3. Représenter graphiquement la région critique et les risques α et β .
4. Calculer α et β .
5. Quel devrait être l'effectif de l'échantillon pour que $\alpha = \beta = 0,01$? Déterminer alors la région d'acceptation de H_0 .

Exercice 4. khi-deux :

On s'interroge sur l'indépendance de deux variables : la catégorie socioprofessionnelle et le style d'éducation familiale.

On suppose que la population a été répartie en 3 catégories socioprofessionnelles (CSP), et on distingue entre éducation faible, souple ou rigide.

On dispose pour décider des tableaux de contingence observés et théoriques suivants :

Distribution observée :

	CSP1	CSP2	CSP3	Total
Faible	9	8	6	23
Souple	40	21	8	69
Rigide	10	22	31	63
Total	59	51	45	155

Distribution théorique :

	CSP1	CSP2	CSP3	Total
Faible	8,75	7,57	6,68	23
Souple	26,26	22,70	20,03	69
Rigide	23,98	20,73	18,29	63
Total	59	51	45	155

1. Préciser sous quelle hypothèse, (H_0), est obtenue la distribution théorique.
2. Expliquer comment est obtenue l'une des valeurs intérieures du tableau théorique (par exemple 22,70).
3. La variable de décision de ce test est la distance du khi-deux dont la valeur ici est de : $D = 18,47$.
Expliquer comment est calculée cette distance.
4. Quelle est la loi suivie par D sous l'hypothèse (H_0) ?
5. Effectuer le test à l'aide de la valeur de D et de la table de la loi adéquate, et conclure quant à l'indépendance des deux variables qui nous intéressent.

Exercice 5. RLM :

Afin d'étudier l'impact des publicités télé, radio et presse sur les ventes d'une marque de chaussures, une régression linéaire multiple a été effectuée pour expliquer la variable Sales(Ventes) par les variables TV, Radio et Newspaper. Les résultats obtenus grâce au logiciel R, sont regroupés dans le tableau suivant :

```

Call:
lm(formula = Sales ~ TV + Radio + Newspaper, data = ventes)

Residuals:
    Min       1Q   Median       3Q      Max
-8.8277 -0.8908  0.2418  1.1893  2.8292

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.938889   0.311908   9.422  <2e-16 ***
TV           0.045765   0.001395  32.809  <2e-16 ***
Radio        0.188530   0.008611  21.893  <2e-16 ***
Newspaper    -0.001037   0.005871  -0.177    0.86
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.686 on 196 degrees of freedom
Multiple R-squared:  0.8972,    Adjusted R-squared:  0.8956
F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16

```

1. Pouvez-vous retrouver la taille de l'échantillon grâce à ces résultats ?
2. Donner la formule permettant de prédire la variable Sales en fonction des autres.
3. Donner une interprétation de la valeur de R^2 .
4. Que dit la p-valeur de la dernière ligne concernant la validité de ce modèle ? Justifier votre réponse.
5. La ligne concernant la variable TV permet d'effectuer un test sur le coefficient de cette variable dans le modèle.
Préciser les hypothèses de ce test, et sa conclusion en justifiant vos réponses.
6. Quel changement peut-on apporter pour améliorer le modèle ?