Polytechnique Montréal
Département de mathématiques et de génie industriel
MTH2302D - Probabilités et statistique
Examen final - Hiver 2016

Question n° 1 : (8 points)

La masse d'une boule métallique issue d'une production est une variable aléatoire distribuée selon une loi normale de moyenne μ et de variance 0,1 kg. On prélève un échantillon de taille 25 de cette production. Soit $\overline{X}$ et S^2 la moyenne et la variance de cet échantillon.

a) (2 points) Déterminer la valeur de la constante c telle que $P(S^2 \leq c) = 0{,}95$.

X: "Masse d'une boule métallique" $\sim N(\mu\,;\, \overset{\sigma^2}{0{,}1})$

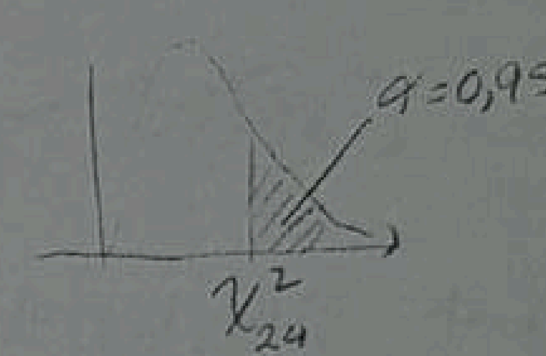
échantillon: $n=25$ $\to \bar{X}$, $\to S^2$

$P(S^2 \leq c) = 0{,}95$

$P\left((n-1)\frac{S^2}{\sigma^2} \leq c\,(n-1)\frac{\cancel{S^2}}{\sigma^2}\right) = P\left(\chi^2_{n-1} \leq c\frac{(n-1)}{S^2}\right)$

$= P(\chi^2_{24}(0{,}95) \leq c)$

$\Rightarrow \underline{c = 13{,}85}$



0,5

b) (2 points) Déterminer la valeur de la constante k telle que $P\left(\frac{\overline{X}-\mu}{S} > k\right) = 0{,}95$.

$P\left(\frac{\bar{X}-\mu}{S} > k\right) = 0{,}95$

$P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} > k\sqrt{n}\right) = 0{,}95$

$1-\alpha$

$P(T > t_{24}(0{,}95)) = P(T > -t_{24}(0{,}05)) = P(T > -1{,}71)$

$k\sqrt{n} = -t_{24}(0{,}05) = -1{,}71$

$\sqrt{25}\,k = -1{,}71$

$k = \frac{-1{,}71}{\sqrt{25}} = \underline{-0{,}342}$

2

Question nº 1 : (suite)

c) **(4 points)** On a observé $\overline{x} = 5{,}9$ et $s^2 = 0{,}16$.

c.1) *(2 points)* Calculer un intervalle de confiance pour la moyenne μ au niveau de confiance $1 - \alpha = 0{,}95$.

Moyenne μ (σ^2 connue)

$n = 25$
$\bar{x} = 5{,}9$
$s^2 = 0{,}16$
$1-\alpha = 0{,}95$
$\alpha = 0{,}05$
$\frac{\alpha}{2} = 0{,}025$

$X \sim N(\mu; 0{,}1)$

① Paramètre d'intérêt : μ

② σ^2 connue et $X \sim N(\mu, \sigma^2)$: (bilateral)

$\sigma = \sqrt{0{,}1}$

③ $IC = \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$z_{0{,}025} = \phi^{-1}(1-0{,}025) = \phi^{-1}(0{,}975) = 1{,}96$

$5{,}9 \pm z_{0{,}025} \frac{\sqrt{0{,}1}}{\sqrt{25}}$

$\downarrow$

$1{,}96$ ✓

$5{,}9 \pm 0{,}0392 = [5{,}\cancel{8608}; \cancel{5{,}9392}]$

1,5

④ ∴ IC pour la moyenne au niveau de confiance $1-\alpha = 0{,}95$
bilateral $[5{,}\cancel{8608}; 5{,}9392]$

c.2) *(2 points)* Calculer un intervalle de confiance unilatéral avec borne inférieure pour σ^2 au niveau de confiance $1 - \alpha = 0{,}95$. Commenter brièvement ce résultat.

$[\ldots, A($

① Paramètre d'intérêt : σ^2

② IC unilatéral borne inférieure ~~(gauche)~~ → Utiliser α

③ $n = 25$
$\bar{x} = 5{,}9$
$s^2 = 0{,}16$
$1-\alpha = 0{,}95$
$\alpha = 0{,}05$

σ^2 connue et $X \sim N(\mu; \sigma^2)$
μ inconnue

1

$\sigma^2 \stackrel{?}{=} 0{,}1$

④ $\left[\frac{(n-1)s^2}{\chi^2_{\alpha; n-1}} ; \frac{(n-1)s^2}{\chi^2_{1-\alpha; n-1}}\right] = \left[\frac{24 \cdot 0{,}16}{\chi^2_{0{,}05; 24}} ; \frac{24 \cdot 0{,}16}{\chi^2_{0{,}95; 24}}\right] = [0{,}105 ; \cancel{0{,}277}]$

$\chi^2_{0{,}05;24} \to 36{,}42$ $\quad \chi^2_{0{,}95;24} \to 13{,}84$

⑤ ∴ IC pour l'écart-type unilateral : $[0{,}105 ; \cancel{0{,}277}]$

Question n° 2 : (6 points)

Dans le but de tester l'effet d'une antibiothérapie systématique sur l'apparition d'une infection post-opératoire, une expérience randomisée est conduite. Un premier groupe de patients reçoit une antibiothérapie ; un deuxième groupe reçoit un placebo. Les résultats observés sont alors les suivants :

	A Antibiothérapie	B Placebo	
Apparition d'infection	10	29	39
Absence d'infection	75	27	102
	85	56	141

Au seuil critique 5%, peut-on dire que l'apparition de l'infection post-opératoire est indépendante du type de thérapie reçue ?

Important *: formuler au préalable les hypothèses H_0 et H_1 du test.*

$\alpha = 0{,}05$ Test d'indépendance H_0 : X_1 indépendante de A et de B vs H_1 : $X_1 \neq$ ind de A et B

1) X_1 : Nombre de personnes où il y a apparition infection

X_2 : " " " absence d'infection

A : Nombre de personne utilisant antibiotique

B : Nombre de personnes utilisant placebo.

Effectif attendus

	A	B	Tot
X_1	$85 \cdot 39/141 = 23{,}5$	$56 \cdot 39/141 = 15{,}49$	$29+10 = 39$
X_2	$85 \cdot 102/141 = 61{,}5$	$102 \cdot 56/141 = 40{,}5$	$75+27 = 102$
Tot	85	56	141

$(10-23{,}5)^2/23{,}5^2 + \ldots$

2) Statistique du test $\chi_0^2 = \frac{(23{,}5-10)^2}{23{,}5} + \frac{(29-15{,}49)^2}{15{,}49} + \frac{(75-61{,}5)^2}{61{,}5} + \frac{(27-40{,}5)^2}{40{,}5}$

$= 27$

3) Regle de décision : Rejet de H_0 si $\chi_0^2 > \chi_v^2(\alpha)$ où $= \chi_1^2(0{,}05) = 3{,}84$

$v = (r-1)\cdot(c-1)$
$= (2-1)(2-1) = 1$

$27 > 3{,}84 \Rightarrow$ On rejette H_0 en faveur de H_1

L'apparition de l'infection post operatoire est pa[s] ...

Polytechnique Montréal
Département de mathématiques et de génie industriel
MTH2302D - Probabilités et statistique
Examen final - Hiver 2016

Question nº 3 : (6 points)

NON a b

Soit $X_1, \ldots, X_n$, un échantillon aléatoire prélevé d'une population (i.e. une variable) X de fonction de densité

$$f(x) = \begin{cases} \dfrac{2}{\theta^2}(\theta - x) & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon,} \end{cases}$$

où θ est un paramètre réel tel que $\theta > 0$.

a) **(4 points)** Trouver l'estimateur $\hat{\theta}$ de θ par la méthode des moments.

$$\mu_1' = E(X) = \int_0^\theta \frac{2x}{\theta^2}(\theta - x)\,dx = \int_0^\theta \left(\frac{2x\cancel{\theta}}{\theta^{\cancel{2}}} - \frac{2x^2}{\theta^2}\right)dx = \int_0^\theta \left(\frac{2x}{\theta} - \frac{2x^2}{\theta^2}\right)dx = \frac{\cancel{2}x^2}{\cancel{2}\theta} - \frac{2x^3}{3\theta^2}\Bigg|_0^\theta$$

$$\frac{\theta^{\cancel{2}}}{\cancel{\theta}} - \frac{2\theta^3}{3\theta^{\cancel{2}}} = \theta - \frac{2}{3}\theta = \frac{1}{3}\theta$$

$$m_1' = \frac{1}{n}\sum_{j=1}^{n} x_j^1 = \bar{X}$$

$$\bar{X} = \frac{\theta}{3} \Rightarrow 3\bar{X} = \theta$$

$$\therefore \hat{\theta} = 3\bar{X}$$

✓

b) **(2 points)** Déterminer le biais de $\hat{\theta}$.

$$\text{Biais}(\hat{\theta}) = E(\hat{\theta}) - \theta = 3\mu - 3\bar{X} = 3\mu - 3\mu = \underline{0}$$

puisque $\bar{X}$ est un bon estima sans biais de μ.

$$E(\hat{\theta}) = E(3\bar{X}) = 3E(\bar{X}) = 3\mu$$

✓

$n = 25$

$N \sim (\mu, 10^2)$

s'assurer que la masse est 500

Question nº 4 : (10 points)

Une machine sert à fabriquer des billes métalliques dont la masse est distribuée selon une loi normale avec un écart type de 10 grammes. L'ingénieur responsable du contrôle de la qualité doit s'assurer que la masse des billes est bien de 500 grammes. Il compte effectuer un test statistique à l'aide d'un échantillon aléatoire de 25 billes. Il conclura que la machine n'est pas bien ajustée si la masse moyenne de l'échantillon est inférieure à 498 grammes ou supérieure à 502 grammes.

$\bar{X}$

a) **(2 points)** Définir la variable ainsi que les paramètres d'intérêt et formuler les hypothèses à tester.

Variable: X = "Masse des billes métalliques" ~~$X \sim N(500, 10^2)$~~ N nc. !!

0

Échantillon: $n = 25$ billes

Paramètre d'intérêt: μ et $\mu_0 = 500$

1.5 ✓ Hypothèses: $\boxed{H_0: \mu = 500 \text{ vs } H_1\ \mu \neq 500}$

Je ne suis pas sûre, alors je pose comme hypothèse ceci pour continuer le numéro.

b) **(2 points)** Calculer le risque de première espèce de ce test.

Risque de première espèce = $\alpha = P(\text{Rejeter } H_0 \mid H_0 \text{ vraie}) = 1 - \beta$

Calcul β σ^2 connue et niveau critique $\alpha = 0{,}05$ (hypothèse)

$\alpha = 1 - \beta = 1 - 0{,}677 = 0{,}323$

(↑ numéro suivant)

$\alpha = P[\text{Rejeter } H_0 \mid H_0 \text{ est vraie}]$

$= 1 - P[498 \leq \bar{X} \leq 502]$

$= 1 - P\left[\frac{498 - 500}{10} \leq Z \leq \frac{502 - 500}{10}\right]$

$=$

Question n° 4 : (suite)

c) **(2 points)** Calculer le risque de deuxième espèce de ce test lorsque la masse moyenne est de 503 grammes.

Pas mentionné dans le texte, alors hypothèse

Calcul de β. Seuil de $\alpha = 0{,}05$ $\frac{\alpha}{2} = 0{,}025$

2

$$\beta = \phi\left(z_{\alpha/2} - \frac{(N - N_0)\sqrt{n}}{\sigma}\right) - \phi\left(-z_{\alpha/2} - \frac{(N-N_0)\sqrt{n}}{\sigma}\right)$$

$$= \phi\left(1{,}96 - \frac{(503-500)\sqrt{25}}{10}\right) - \phi\left(-1{,}96 - \frac{(503-500)\sqrt{25}}{10}\right)$$

$$= \phi(0{,}46) - \phi(-3{,}46)$$

-3,46 3,46

$$= 0{,}67724 - (1 - (\phi(3{,}46))$$

$$= 0{,}67724 - (1 - 0{,}99973) = 0{,}67697 \simeq 0{,}677$$

~~67,7 %~~ = β.

d) **(2 points)** On veut un risque de deuxième espèce de 25%, toujours lorsque la masse moyenne est de 503 grammes, et sans modifier le risque de première espèce. Quelle devrait être la taille d'échantillon nécessaire ?

$\beta = 0{,}25$ $\quad \bar{x} = 503$ $\quad n = ?$ $\quad H_0 : N_0 = 500$ contre $H_1 : N \neq 500$

$\alpha = 1 - 0{,}25 = 0{,}75$ $\quad \alpha/2 = 0{,}375$ $\quad \phi^{-1}(\underbrace{1 - 0{,}375}_{0{,}625}) = 0{,}32$ $\quad \phi^{-1}(\underbrace{1-0{,}25}_{0{,}75}) = 0{,}68$

2

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{(N - N_0)^2} = \frac{(z_{0{,}4875} + z_{0{,}25})^2 \sigma^2}{(503-500)} = \frac{(0{,}32 \cdot 0{,}68)^2 \cdot 10^2}{(503-500)}$$

$$= 1{,}58 \simeq \cancel{2}$$

La taille devrait être de ~~2~~

$\beta = 0{,}25$

α = risque de première espèce

Question nº 4 : (suite)

e) **(2 points)** Calculer la valeur-P du test en a) si la masse moyenne de l'échantillon est de 497 grammes.

$$\alpha_{min} = ? \quad 2\left(1 - \Phi(z_0)\right)$$

$$z_0 = \frac{497 - 500}{10/\sqrt{25}} = -1{,}5$$

$$\alpha_{min} = 2\left[1 - \Phi(-1{,}5)\right] = 2\left[\cancel{1} - \cancel{1} + \Phi(1{,}5)\right]$$

$$= 2\,\Phi(1{,}5) = \ldots$$

Question nº 5 : (15 points)

Afin d'étudier la distance Y (en m) nécessaire pour immobiliser un type de véhicule en fonction de sa vitesse X (en km/h), on a recueilli sept observations de la forme $(x_i, y_i), i = 1, 2, \ldots, 7$. Deux modèles de régression linéaires sont envisagés.

On considère un premier modèle de régression linéaire simple d'équation

$$Y = \beta_0 + \beta_1 X + \epsilon. \tag{1}$$

On dispose des résultats partiels suivants sur l'ajustement du modèle (1) avec les sept observations :

$$\sum_{i=1}^{7} Y_i = 146{,}61; \quad \sum_{i=1}^{7} Y_i^2 = 4836{,}3; \quad SS_R = 1682{,}73.$$

a) **(4 points)** Compléter le tableau d'analyse de la variance ci-dessous, en laissant les cases inutiles vides. Au seuil critique $\alpha = 0{,}05$, tester si le modèle est significatif et conclure.

on rejette pas ⇕ il est pas signif

Source de variation	Somme des carrés	Nombre de degrés d.l.	Moyenne des carrés	F_0
Régression	① $SS_R = 1682{,}73$	③ 1	⑤ 1682,73	⑦ 99.95
Erreur	② 84.18	④ 5	⑥ 16,836	
Total	$S_{YY} = 1766{,}91$	⑧ $n-1=6$		

$n=7$

① SS_R

② $SS_E = S_{YY} - SS_R$

$S_{YY} = \sum_{i=1}^{7} y_i^2 - n(\bar{y})^2 = \sum_{i=1}^{7} y_i^2 - n \sum_{i=1}^{7} (\bar{y})^2$

$\bar{y} = \frac{\sum_{i=1}^{7} y_i}{n} = \frac{146{,}61}{7} = 20{,}94$

$\bar{y}^2 = 438{,}4836$

$S_{YY} = 4836{,}3 - 7 \cdot (438{,}4836) = 1766{,}91$

$SS_E = 1766{,}91 - 1682{,}73 = 84.18$

④ $n-2 = 7-2 = 5$

⑤ $MS_R = \frac{SS_R}{1} = 1682{,}73$

⑥ $MS_E = \frac{SS_E}{n-2} = \frac{84.18}{5} = 16{,}836$

⑦ $F_0 = \frac{MS_R}{MS_E} = \frac{1682{,}73}{16{,}836} = 99{,}95$

4

1) Hypothèse : $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$

2) Statistique : $F_0 = 99{,}95$

3) Règle de décision : On rejette H_0 si $F_0 > F_{1;n-2}(\alpha)$

$F_0 > F_{1;5}(0{,}05)$

99,95

$F_0 > 6{,}61$

4) On rejette H_0 au seuil critique $\alpha = 0{,}05$ en faveur de H_1 puisque $F_0 > 6{,}61$. Le modèle est donc significatif

Question nº 5 : (suite)

$\sum x_i^2$ $Y = \beta_0 + \beta_1 x$

En plus des résultats prédédents, on a

$\sum x_i = 26095$

OK $$\sum_{i=1}^{7} X_i = 401; \quad \sum_{i=1}^{7} X_i^2 = 26095.$$

Moyenne et non pas prédiction (la formule sans 1)

b) **(3 points)** Calculer $\hat{\beta}_0$ et $\hat{\beta}_1$. Donner l'équation de la droite de régression.

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$

$= 401 \cdot 146{,}61 - \frac{401 \times 146{,}61}{7}$

$= 50\,392$

$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 26095 - 7\left(\frac{401}{7}\right)^2 = 3123$

$\hat{\beta}_1 = \frac{50\,392}{3123} = 16{,}14$

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$= \frac{146{,}61}{7} - 16{,}14\left(\frac{401}{7}\right)$

$= -903{,}65$

0,5

$\hat{Y} = 16{,}14x_0 - 903{,}65$

c) **(3 points)** Calculer un intervalle de confiance pour $E(Y|x_0)$ lorsque $x_0 = 10$, au niveau de confiance $1 - \alpha = 0{,}95$. $\alpha = 0{,}05$ $\alpha/2 = 0{,}025$

?

$E(Y|10) \in \hat{Y}_0 \pm t_{n-2}(\alpha/2) \cdot \sqrt{MS_E\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$ $\quad \frac{401}{7} = 57{,}29$

$t_{n-2}(\alpha/2) \rightarrow t_5(0{,}025)$, $MS_E \rightarrow 16{,}836$, $S_{xx} \rightarrow 3123$

$16{,}14(10) - 903{,}65 \pm 2{,}571 \sqrt{16{,}836\left(\frac{1}{7} + \frac{(10-57{,}29)^2}{3123}\right)}$

$-742{,}25 \pm 9{,}777$

IC: $[-752{,}027 ; -732{,}473]$

2,5

$$R^2 = \frac{SC_R}{SC_T} = \frac{\beta_1^2 \cdot SC_x}{SC_Y} = \frac{\beta_1^2 (\Sigma x^2 - n\bar{x}^2)}{SC_Y}$$

$\frac{1}{n} \Sigma$

Question nº 5 : (suite)

Question 5 p: 409 =

$\frac{SC_R}{SC_T} = \frac{\beta_1^2 SC_x}{SC_Y}$

On considère un deuxième modèle de régression linéaire dont l'équation est

$$Y = \beta_0' + \beta_1' X^2 + \epsilon'. \qquad (2)$$

$\beta_1^2 =$

$\frac{SC_{xy}^2}{(SC_x)^2} \cdot SC_x$ / SC_Y

On obtient pour ce deuxième modèle un coefficient de détermination $R^2 = 0{,}985$.

d) (3 points) Au seuil critique $\alpha = 0{,}05$, tester si le modèle (2) est significatif.

$R^2 = 0{,}985 = \frac{SS_R}{SS_R + SS_E}$ $\quad X^2 = \cancel{x}$ $\quad F_0 = \ldots$ $\quad (SC_{xy})^2$

Question précédente

le piège est: $\Sigma \cancel{X} = \Sigma X^2$ donnera SC_Y

$\Sigma \cancel{Y}^2$

$\Sigma X_i = 401$ $\quad \Sigma X_i^2 = 28095$ $\quad \Sigma X^2$? ?

$x = X^2$

$\Sigma x = \Sigma X^2$

Σx^2

On a pas $\Sigma x = \Sigma X^2$

e) (2 points) En se basant sur les coefficients de détermination et les figures 1 et 2 (page suivante), lequel des deux modèles est préférable ? $R^2 \uparrow\uparrow \Rightarrow$ Modèle perfecto

R^2 1

Le modèle de la figure 2 est préférable, car les résidus suivent approximativement plus une loi normale que ~~ceux~~ de la figure 1. En effet, les points de la figure "Normality Probability Plot of Residuals" ont une tendance linéaire. De plus, dans le graphique Residuals en fonction de "Predicted values", les points sont distribués aléatoirement et ne semblent pas démontrer d'autocorrélation. La variance est ainsi constante les résidus. Tandis que dans la figure 1, les résidus semblent avoir une autocorrélation puisque la courbe est de forme parabolique ∴ il y aurait autocorrélation entre les résidus.

∴ Figure 2 meilleur modèle car résidus ~ N et absence d'autocorrélation car variance approximativement

Question n° 5 : (suite)

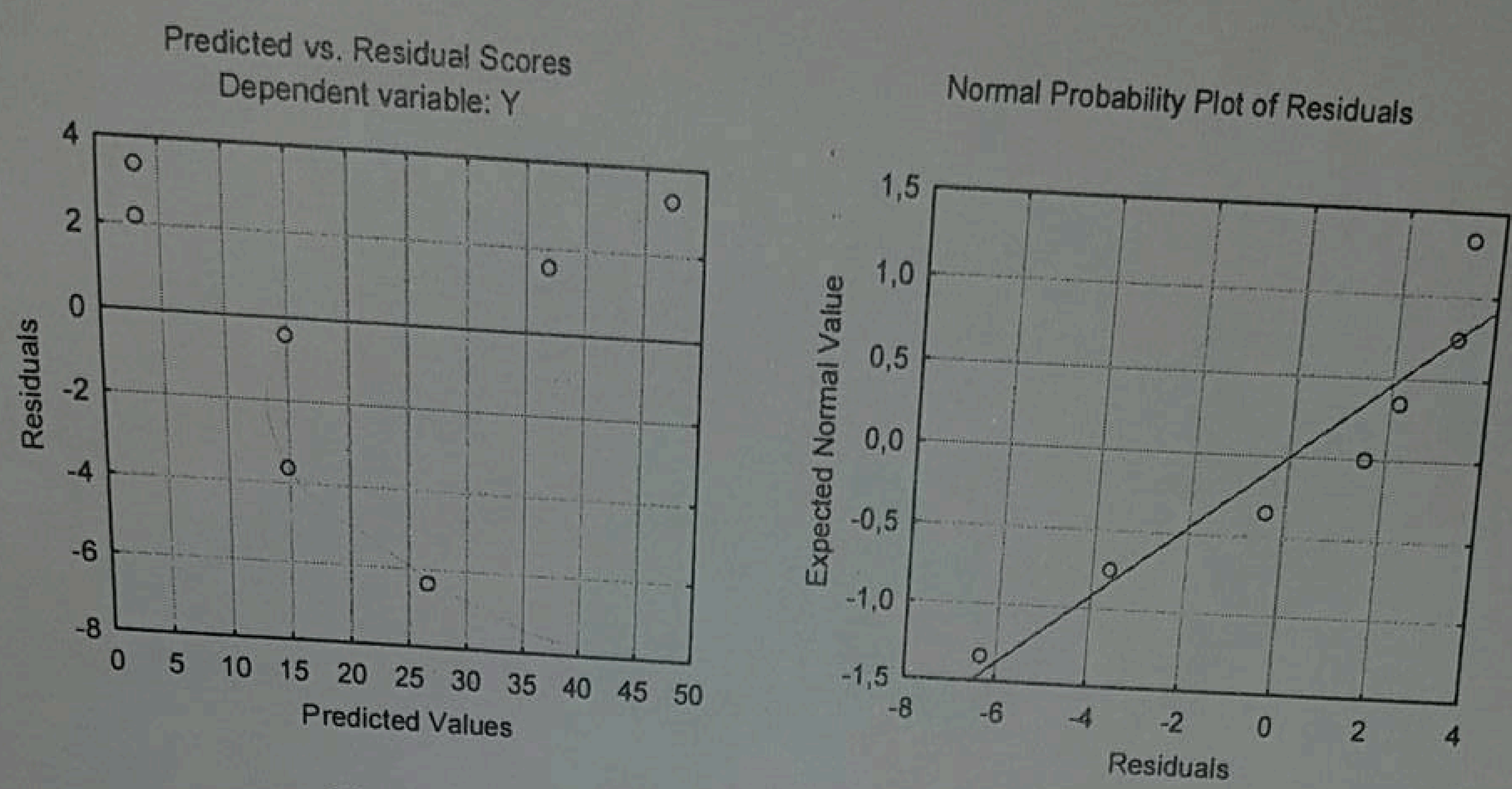


FIGURE 1 – Graphes des résidus du modèle (1).

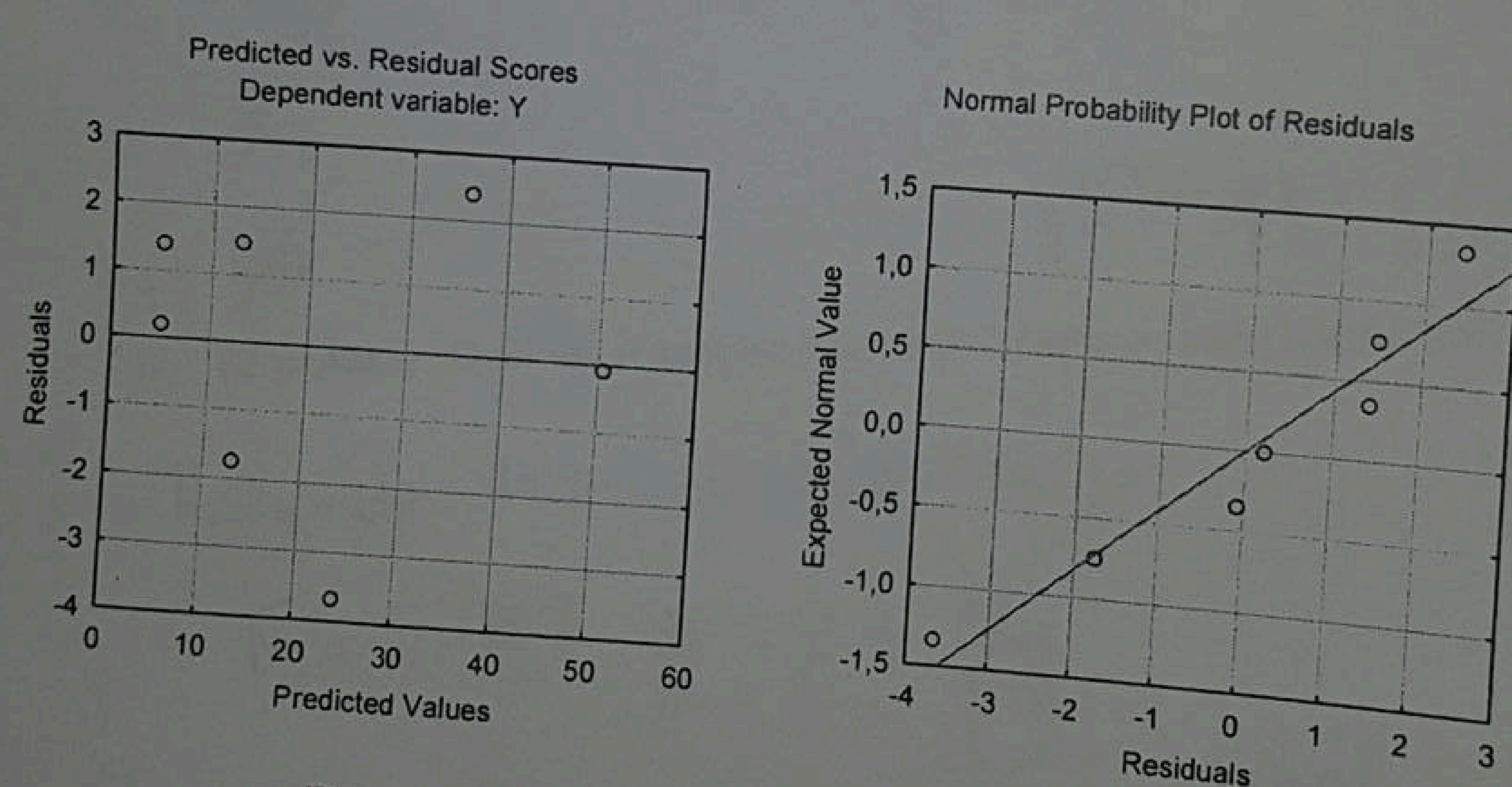


FIGURE 2 – Graphes des résidus du modèle (2).

Vous avez dit en classe que $\rho < 1$...

Question n° 6 : (5 points)

Une boulangerie ne possède qu'une seule caisse et ne peut accueillir plus de 4 clients à la fois. On considère un système $M/M/1/4$ pour modéliser la file de cet établissement. Les clients arrivent en moyenne au rythme de 2 par minute, et sont servis au rythme de 1 par minute.

λ

À l'équilibre : $N = 1 \quad \lambda = 2 \quad k = 4$

$\rho = \frac{\lambda}{\mu} = 2 \neq 1$

$\lambda = 2$ (min)

$\bar{S} = \frac{1}{\mu} = 1$ (min)

$\mu = 1$ (min)

a) (2 points) Quel est le nombre moyen de clients dans la boulangerie ?

Pas la peine de convertir

Nombre moyen de clients $= \bar{N} = E(N)$ N Nombre de clients $\sim G(1-\frac{1}{\rho})$

$\rho' =$ Intensité du trafic $= \frac{\lambda}{\lambda_1} = \frac{2}{1} = 2 = \rho' > 1$ ... inverser $\cancel{\frac{N}{\lambda} = \frac{1}{2} = \rho}$

!!!

$\pi_0 = \frac{P(Y = 1 \cap Y \leq 5)}{P(Y \leq 5)} = \frac{P(Y=1)}{P(Y \leq 5)} = \frac{(\frac{1}{2})^0 (1 - \frac{1}{2})}{1 - (\frac{1}{2})^{4+1}} = \cancel{\frac{16}{31}}$

$\pi_1 = \frac{(\frac{1}{2})^1 (1 - \frac{1}{2})}{1 - (\frac{1}{2})^5} = \cancel{8/31}$

$\bar{N} =$ Calcul

$\pi_2 = \cancel{4/31}$

$\bar{N} = 0 \cdot \frac{16}{31} + 2 \cdot \frac{4}{31} + 3 \cdot \frac{2}{31} + 4 \cdot \frac{1}{31} = \cancel{\frac{18}{31}}$

$\pi_3 = \cancel{2/31}$

$\pi_4 = \cancel{1/31}$

principe (1)

b) (2 points) Quel est le nombre moyen de clients en train d'attendre ?

Nombre moyen de clients en attente $= N_Q$

$\bar{N}_Q = \bar{N} - \bar{N}_S = \frac{18}{31} - \frac{15}{31} = \cancel{3/31}$

$\bar{N}_S = 1 - \pi_0$

$= 1 - \frac{16}{31} = \frac{15}{31}$

princ. (2)

Question n° 6 : (suite)

c) **(1 point)** Quelle est la probabilité de voir des clients attendre en dehors de la boulangerie ?

Attente pleine capacité : π_k

$P(N>4) = \cancel{1 - P(\pi_4)}$

$= \cancel{1 - \pi_4}$

$= \cancel{1 - 1/31} = \underline{0{,}96}$ — pas vraie π_4

$\pi_4 = \ldots$

avec $\rho \neq 1$

0

NON