STT5100 - Automne 2018 - Examen Final

Arthur Charpentier

Examen **A**

Les calculatrices sont autorisées. Les documents sont en revanche interdits, sauf une page d'aide mémoire. L'examen dure 3 heures, toute sortie avant midi est autorisée, et définitive.

Dans les feuilles qui suivent, il y a 30 questions relatives au cours sur les modèles linéaires généralisés (incluant la régression logistique, et Poisson). Pour chaque question, quatre ou cinq réponses sont proposées, une seule est valide, et vous ne devez en retenir qu'une (au maximum),

- vous gagnez 1 point par bonne réponse
- vous ne perdez pas de points pour une mauvaise réponse
- vous ne gagnez pas de point pour plusieurs réponses

Aucune justification n'est demandée. Votre note finale est le total des points (sur 30). Il y a une 31ème question, bonus. Une prédiction parfaite (sur 30) donnera un point bonus qui s'ajoutera à la note.

La page de réponse est à la toute fin du document : merci de décrocher la feuille et de ne rendre que cette dernière, après avoir indiqué votre code permanent en haut. La-dite feuille contient une page avec les réponses à cocher, et au dos, 2 *graphiques* à compléter (pour les questions 5 et 6).

Merci de cocher le carré en bleu ou en noir. En cas d'erreur, vous pouvez cocher une autre case en rouge. Seule cette dernière sera alors retenue.

Formulaire

Pour la loi normale, centrée et réduite ou une loi de Student, on utilisera 1.96 comme valeur du quantile à 97.5%, et 1.64 pour le quantile à 95%.

On notera $x \mapsto \mathbf{1}_A(x)$ la fonction indicatrice vérifiant $\mathbf{1}_A(x) = 0$ si $x \notin A$ et $\mathbf{1}_A(x) = 1$ si $x \in A$. Par extention, si $A = \{a\}$, on notera $x \mapsto \mathbf{1}_a(x)$ la fonction qui vérifie $\mathbf{1}_a(x) = 0$ si $x \neq a$ et $\mathbf{1}_a(x) = 1$ si $x = a$. Dans ce dernier cas, on pourra aussi utiliser la notation $\mathbf{1}(x = a) = 0$.

1 Comment est estimé un modèle logistique

A) par la méthode des moindres carrés sur les y_i

B) par la méthode des moindres carrés sur une transformation logistique de observations y_i

C) par la méthode des moments sur le logarithme des y_i

D) par maximum de vraisemblance car les y_i suivent une loi binomiale conditionnellement aux $\boldsymbol{x}_i$

2 Dans un modèle logistique, si $\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$ vaut 0, quelle sera la prévision pour $\mathbb{P}[Y = 1|x_1, x_2]$?

A) 0.25

B) 0.5

C) 0.75

D) 1

E) aucune des réponses proposées

3 On a obtenu la sortie de régression suivante

```
Call:
glm(formula = cancer ~ genre, family = binomial("logit"), data = df)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.30060    0.08382 -15.517  < 2e-16 ***
genreF                ???     0.11026                 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On sait que les femmes ont deux fois plus de chances d'avoir un cancer que les hommes. Donnez un ordre de grandeur pour la valeur manquante

A) -0.452

B) 0.452

C) 0.693

D) -0.693

4 On a autant d'hommes que de femmes dans l'échantillon

```
Call:
glm(formula = cancer ~ 0 + genre, family = binomial("logit"), data = df)

Coefficients:
         Estimate Std. Error z value Pr(>|z|)
genreH   -1.30060    0.08382  -15.52   <2e-16 ***
genreF   -0.93642    0.07163  -13.07   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Donner un ordre de grandeur du ratio $\mathbb{P}[\text{cancer}|x = \texttt{H}]/\mathbb{P}[\text{cancer}|x = \texttt{F}]$

A) -0.847

B) 0.847

C) 2.601

D) -2.601

5 On considere la sortie de regression suivante, avec une variable binaire $y \in \{0, 1\}$ et deux variables explicatives continues, x_1 et x_2 à valeurs dans l'intervalle $[0, 1]$.

```
Call:
glm(formula = y ~ x1 + x2, family = binomial("logit"), data = df)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.5659     0.1071  -5.283 1.27e-07 ***
x1            0.8571     0.1379   6.215 5.14e-10 ***
x2           -0.2674     0.1390  -1.923   0.0545 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sur la Figure 1, représentez la région des points (x_1, x_2) pour lesquels $\mathbb{P}[Y = 1|x_1, x_2] > \mathbb{P}[Y = 0|x_1, x_2]$.

6 Sur les mêmes données, on décide de couper x_1 et x_2 en classes.

```
Call:
glm(formula = y ~ (x1 > 0.8) + (x2 > 0.65), family = binomial("logit"),
    data = df)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.30533    0.05398  -5.656 1.55e-08 ***
x1 > 0.8TRUE   0.41028    0.09646   4.253 2.11e-05 ***
x2 > 0.65TRUE -0.14849    0.08380  -1.772   0.0764 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sur la Figure 2, représentez la région des points (x_1, x_2) pour lesquels $\mathbb{P}[Y = 1|x_1, x_2] > \mathbb{P}[Y = 0|x_1, x_2]$.

7 Considérons une variable y prenant les valeurs $\{A, B\}$. La régression de $\mathbf{1}_A(y)$ sur x_1 et x_2 donne

```
Call:
glm(formula = (y == "A") ~ x1 + x2, family = binomial, data = df)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.9524     0.2890   3.296 0.000982 ***
x1           -0.4728     0.1454  -3.252 0.001147 **
x2           -0.9676     0.4826  -2.005 0.044957 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La régression de $\mathbf{1}_B(y)$ sur x_1 et x_2 donne

```
Call:
glm(formula = (y == "B") ~ x1 + x2, family = binomial, data = df)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    ???
x1             ???
x2             ???
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Donnez la valeur de $\widehat{\beta}_0$ (estimateur de la constante) dans la seconde régression

A) 0.9524

B) 1.0498

C) -0.9524

D) 0.0476

8 Sur la même sortie que la question précédante, donnez la valeur de $\widehat{\beta}_1$ (estimateur associé a la variable x_1) dans la seconde régression

A) -0.4728

B) 0.4728

C) 2.1151

D) 0.5272

E) on ne peut pas savoir

9 Soit N une variable de comptage (d'accidents de la route), et on considère deux variables explicatives (possibles) $X_1 \in \{\textsf{male}, \textsf{female}\}$ et $X_2 \in \{\textsf{urban}, \textsf{mid-urban}, \textsf{rural}\}$. Soit E le nombre d'individus par classe. On a observé les données suivantes (présentée ici sous forme de tableau de contingence)

E	men	women
urban	100	100
mid-urban	100	100
rural	100	100

N	men	women
urban	10	8
mid-urban	13	10
rural	7	5

On considère une régression de Poisson $(N|X_1, X_2, E) \sim \mathcal{P}(E \cdot \lambda_{x_1,x_2})$ où

$$\lambda_{x_1,x_2} = \exp\left[\gamma_0 + \alpha_{\mathsf{w}}\mathbf{1}(x_1 = \textsf{woman}) + \beta_{\mathsf{r}}\mathbf{1}(x_2 = \textsf{rural}) + \beta_{\mathsf{u}}\,\mathbf{1}(x_2 = \textsf{urban})\right]$$

En utilisant les données individuelles, et en faisant une régression sous R, on obtient la sortie suivante

```
> summary(reg)

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.0388     0.2407  -8.470   <2e-16 ***
woman             ???
rural           -0.6506     0.3561  -1.827   0.0677 .
urban           -0.2451     0.3147  -0.779   0.4360
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Quelle est la valeur de $\widehat{\alpha}_{\mathsf{w}}$ dans la sortie précédante ?

A) -0.3582

B) -0.7814

C) -0.2657

D) 0.1426

10 On suppose que la matrice de variance-covariance pour la sortie précédante a été obtenue

```
> vcov(reg)

             (Intercept)        X1woman        X2rural        X2urban
(Intercept)   0.05794367 -3.333333e-02 -4.347826e-02 -4.347826e-02
X1woman      -0.03333333  7.681159e-02 -7.803914e-18 -2.675628e-18
X2rural      -0.04347826 -7.803914e-18  1.268116e-01  4.347826e-02
X2urban      -0.04347826 -2.675628e-18  4.347826e-02  9.903381e-02
```

Donner un intervalle de confiance à 95% pour $\gamma_0 + \beta_{\mathsf{u}}$

A) $[-4.28; -0.28]$

B) $[-2.33; -2.23]$

C) $[-2.51; -1.55]$

D) $[-2.81; -1.75]$

11 Combien d'accidents pour les hommes vivant en ville (*man* et *urban*) doit-on espérer avoir, avec 95% de chance, quand 100 personnes sont dans le portefeuille de l'assureur

A) $[5.12; 95.29]$

B) $[5.66; 14.74]$

C) $[7.44; 17.40]$

D) $[5.52; 17.52]$

12 On observe les statistiques de décès lors de voyages en avions suivantes (les nombres de passagers sont ici en millions de miles parcourus)

```
   annee deces passagers
1   1976   734    386300
2   1977   516    430000
3   1978   754    502700
4   1979   877    548100
5   1980   814    581400
6   1981   362    603300
7   1982   764    587700
8   1983   809    622300
9   1984   223    743300
10  1985  1066    710700
```

Un premier modèle est considéré, avec une loi binomiale, $D_i \sim \mathcal{B}(E_i, p_i)$, où D est le nombre de décès, et E le nombre de passagers, et p est une transformation logistique d'une fonction linéaire de l'année

```
Call:
glm(formula = cbind(deces, passagers - deces) ~ annee, family = binomial,
    data = df)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) 113.312570   8.425622   13.45   <2e-16 ***
annee        -0.060597   0.004254    ???      ???
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On lit dans un journal que la baisse du nombre de décès (relativement à la hausse du nombre de passagers) n'est pas significative. Qu'en pensez vous ?

A) C'est faux, la baisse est significative

B) C'est vrai, la baisse n'est pas statistiquement significative

C) On n'a pas assez d'observations pour conclure dans un sens ou dans l'autre

13 On tente comme alternative une régression de Poisson $D_i \sim \mathcal{P}(E_i \cdot \lambda_i)$, où λ est une transformation exponentielle d'une fonction linéaire de l'année,

```
Call:
glm(formula = deces ~ annee + offset(log(passagers)), family = poisson,
    data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-22.342   -3.886    3.351    4.778   14.240

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) 113.162703   8.420250   13.44   <2e-16 ***
annee        -0.060522   0.004252  -14.23   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cette sortie donne un modèle très proche du précédant. Parmi les quatre explications suivantes, **laquelle est fausse**

A) Si $Y \sim \mathcal{B}(n, p)$ avec n très grand, et p petit, alors $Y \approx \mathcal{P}(np)$

B) La probabilité de mourir est faible, et $\log(p/(1-p)) \approx \log(p)$

C) La probabilité de mourir est faible, et $e^x/(1+e^x) \approx e^x$

D) Le nombre d'observations est important, et asymptotiquement, la régression de Poisson et la régression logistique sont équivalentes

14 Qu'est-ce que la *surdispersion* dans un modèle ?

A) c'est lorsque la variance de $\widehat{Y}$ est plus grande que la variance de Y

B) c'est lorsque $\text{Var}[Y|X] > \mathbb{E}[Y|X]$

C) c'est lorsque $\text{Var}[Y|X] > \text{Var}[Y]$

D) c'est comme la notion d'homoscédasticité dans les modèles linéaires, mais pour les GLM

15 question 29 (examen CAS)

29. You are given the following information for a fitted GLM:

Response variable	Occurrence of Accidents
Response distribution	Binomial
Link	Logit

Parameter	df	$\hat{\beta}$
Intercept	1	x
Driver's Age	2	
1	1	0.288
2	1	0.064
3	0	0
Area	2	
A	1	-0.036
B	1	0.053
C	0	0
Vehicle Body	2	
Bus	1	1.136
Other	1	-0.371
Sedan	0	0

The probability of a driver in age group 2, from area C and with vehicle body type Other, having an accident is 0.22.

Calculate the odds ratio of the driver in age group 3, from area C and with vehicle body type Sedan having an accident.

A. Less than 0.200
B. At least 0.200, but less than 0.250
C. At least 0.250, but less than 0.300
D. At least 0.300, but less than 0.350
E. At least 0.350

16 question 30 (examen CAS)

30. You are given the following information for a fitted GLM:

Response variable	Occurrence of Accidents
Response distribution	Binomial
Link	Logit

Parameter	df	$\hat{\beta}$	se
Intercept	1	-2.358	0.048
Area	2		
Suburban	0	0.000	
Urban	1	0.905	0.062
Rural	1	-1.129	0.151

Calculate the modeled probability of an Urban driver having an accident.

A. Less than 0.01
B. At least 0.01, but less than 0.05
C. At least 0.05, but less than 0.10
D. At least 0.10, but less than 0.20
E. At least 0.20

17 question 40 (examen CAS)

40. You are given the following results from a fitted GLM on the frequency of accidents:

Parameter	df	$\hat{\beta}$	se
Intercept	1	-11.2141	0.1826
Location	1		
Rural	0	0.0000	
City	1	1.0874	0.3162

Calculate the Wald statistic for testing the null hypothesis of $\beta_{city} = 0$.

A. Less than 10.5

B. At least 10.5, but less than 11.5

C. At least 11.5, but less than 12.5

D. At least 12.5, but less than 13.5

E. At least 13.5

18 question 31 (examen CAS)

31. You are given the following information for a fitted GLM:

Response variable	Claim size
Response distribution	Gamma
Link	Log
Dispersion Parameter	1

Parameter	df	$\hat{\beta}$
Intercept	1	2.100
Zone	4	
1	1	7.678
2	1	4.227
3	1	1.336
4	0	0.000
5	1	1.734
Vehicle Class	6	
Convertible	1	1.200
Coupe	1	1.300
Sedan	0	0.000
Truck	1	1.406
Minivan	1	1.875
Station wagon	1	2.000
Utility	1	2.500
Driver Age	2	
Youth	1	2.000
Middle age	0	0.000
Old	1	1.800

Calculate the predicted claim size for an observation from Zone 3, with Vehicle Class Truck and Driver Age Old.

A. Less than 650
B. At least 650, but less than 700
C. At least 700, but less than 750
D. At least 750, but less than 800
E. At least 800

19 question 34 (examen CAS)

32. You are given the following information for a fitted GLM:

Response variable	Claim size
Response distribution	Gamma
Link	Log
Dispersion parameter	1

Parameter	df	$\hat{\beta}$
Intercept	1	2.100
Zone	4	
1	1	7.678
2	1	4.227
3	1	1.336
4	0	0.000
5	1	1.734
Vehicle Class	6	
Convertible	1	1.200
Coupe.	1	1.300
Sedan	0	0.000
Truck	1	1.406
Minivan	1	1.875
Station wagon	1	2.000
Utility	1	2.500
Driver Age	2	
Youth	1	2.000
Middle age	0	0.000
Old	1	1.800

Calculate the variance of a claim size for an observation from Zone 4, with Vehicle Class Sedan and Driver Age Middle age

A. Less than 55
B. At least 55, but less than 60
C. At least 60, but less than 65
D. At least 65, but less than 70
E. At least 70

20 question 32 (examen CAS)

34. Determine which of the following statements are true.

I. The deviance is useful for testing the significance of explanatory variables in nested models.
II. The deviance for normal distributions is proportional to the residual sum of squares.
III. The deviance is defined as a measure of distance between the saturated and fitted model.

A. I only B. II only C. III only D. All but III E. All

21 question 37 (examen CAS)

37. Determine which of the following GLM selection considerations is true.

A. The model with the largest AIC is always the best model in model selection process.

B. The model with the largest BIC is always the best model in model selection process.

C. The model with the largest deviance is always the best model in model selection process.

D. Other things equal, when the number of observations > 1000, AIC penalizes more for the number of parameters used in the model than BIC.

E. Other things equal, when number of observations > 1000, BIC penalizes more for the number of parameters used in the model than AIC.

22 question 41 (examen CAS)

41. A Poisson regression model with log link is used to estimate the number of diabetes deaths. The parameter estimates for the model are:

Response variable	Number of Diabetes Deaths
Response distribution	Poisson
Link	Log

Parameter	df	$\hat{\beta}$	p-value
Intercept	1	-15.000	<0.0001
Gender: Female	1	-1.200	<0.0001
Gender: Male	1	0.000	
Age	1	0.150	<0.0001
Age^2	1	0.004	<0.0001
Age × Gender: Female	1	0.012	<0.0001
Age × Gender: Male	0	0.000	

Calculate the expected number of deaths for a population of 100,000 females age 25.

A. Less than 3

B. At least 3, but less than 5

C. At least 5, but less than 7

D. At least 7, but less than 9

E. At least 9

23 question 31 (examen CAS)

31.

Given the following information:

- Y is a random variable in the exponential family

$$f(y) = c(y, \phi) * \exp\left[\frac{y\theta - a(\theta)}{\phi}\right]$$

- $a(\theta) = -\sqrt{-2\theta}$
- $\theta = -0.3$
- $\phi = 1.6$

Calculate $E(Y)$.

A. Less than -1
B. At least -1, but less than 0
C. At least 0, but less than 1
D. At least 1, but less than 2
E. At least 2

24 question 36 (examen CAS)

36.

You are given the following information for two potential logistic models used to predict the occurrence of a claim:

- Model 1: (AIC = 262.68)

Parameter	$\hat{\beta}$
(Intercept)	-3.264
Vehicle Value ($000s)	0.212
Gender-Female	0.000
Gender-Male	0.727

- Model 2: (AIC = 263.39)

Parameter	$\hat{\beta}$
(Intercept)	-2.894
Gender-Female	0.000
Gender-Male	0.727

- AIC is used to select the most appropriate model.

Calculate the probability of a claim for a male policyholder with a vehicle valued $12,000 by using the selected model.

A. Less than 0.15
B. At least 0.15, but less than 0.30
C. At least 0.30, but less than 0.45
D. At least 0.45, but less than 0.60
E. At least 0.60

25 question 32 (examen CAS)

32. A GLM is used to model claim size. You are given the following information about the model:

- Claim size follows a Gamma distribution.
- Log is the selected link function.
- Scale parameter is estimated to be 2.
- Model Output:

Variable	$\hat{\beta}$
(Intercept)	2.32
Location - Urban	0.00
Location - Rural	-0.64
Gender - Female	0.00
Gender - Male	0.76

Calculate the variance of the predicted claim size for a rural male.

A. Less than 25
B. At least 25, but less than 100
C. At least 100, but less than 175
D. At least 175, but less than 250
E. At least 250

26 question 31 (examen CAS)

31. Within the context of Generalized Linear Models, suppose that y has an exponential distribution with probability density function expressed as:

$$f(y) = \frac{1}{\mu} \exp(-\frac{y}{\mu}); \text{ for } y > 0$$

Determine the variance of y in terms of μ.

A. $1/\mu$ B. $\sqrt{\mu}$ C. μ D. μ^2
E. Cannot be determined from the given information

27 question 33 (examen CAS)

33.

A distribution belongs to the exponential family if it can be written in the canonical form:

$$f(y;\theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)]$$

Determine the values of $a(y), b(\theta), c(\theta),$ and $d(y)$ for the Poisson distribution.

A. $a(y) = y, b(\theta) = \ln(\theta), c(\theta) = \theta,$ and $d(y) = \ln(y!)$
B. $a(y) = -y, b(\theta) = -\ln(\theta), c(\theta) = \theta,$ and $d(y) = \ln(y!)$
C. $a(y) = y, b(\theta) = \ln(\theta), c(\theta) = \theta,$ and $d(y) = 1/(y!)$
D. $a(y) = y, b(\theta) = \ln(\theta), c(\theta) = -\theta,$ and $d(y) = -\ln(y!)$
E. The answer is not given by (A), (B), (C), or (D)

28 question 37 (examen CAS)

37. You are given the outputs from two GLMs fitted to the same data from a trial of a new drug.

Model 1		Model 2	
Response variable	Number	Response variable	Number
Response distribution	Poisson	Response distribution	Negative binomial
Link	Log	Link	Log
AIC	273.877	AIC	164.880

Parameter	$\hat{\beta}$	s. e. $(\hat{\beta})$	Parameter	$\hat{\beta}$	s. e. $(\hat{\beta})$
Intercept	4.529	0.147	Intercept	4.526	0.595
Treatdrug			Treatdrug		
Placebo	0.000	0.000	Placebo	0.000	0.000
Drug	-1.359	0.118	Drug	-1.368	0.369
Age	-0.039	0.006	Age	-0.039	0.021

Determine which of the following statements is false using the Wald test.

A. Under Model 1, the Treatdrug coefficient has a p-value less than 0.01.
B. Under Model 1, the Age coefficient has a p-value less than 0.01.
C. Under Model 2, the Treatdrug coefficient has a p-value less than 0.01.
D. Under Model 2, the Age coefficient has a p-value less than 0.01.
E. Under both models, the intercept coefficient has a p-value of less than 0.01.

29 question 37 (examen CAS)

37.

Let $Y_1, \dots, Y_n$ be independent Poisson random variables, each with respective mean μ_i for $i = 1, 2, \dots, n$, where:

$$\ln(\mu_i) = \begin{cases} \alpha, & \text{for } i = 1,2,\dots,m \\ \beta, & \text{for } i = m+1, m+2, \dots, n \end{cases}$$

The claims experience for a portfolio of insurance policies with $m = 50$ and $n = 100$ is:

$$\sum_{i=1}^{50} y_i = 563$$

$$\sum_{i=51}^{100} y_i = 1{,}261$$

Denote by $\hat{\alpha}$ and $\hat{\beta}$ the maximum likelihood estimates of α and β, respectively.

Calculate the ratio $\frac{\hat{\alpha}}{\hat{\beta}}$.

A. Less than 0.40
B. At least 0.40, but less than 0.60
C. At least 0.60, but less than 0.80
D. At least 0.80, but less than 1.00
E. At least 1.00

30 question 32 (examen CAS)

32.

Given a family of distributions where the variance is related to the mean through a power function:

$$\mathrm{Var}[Y] = a\mathrm{E}[Y]^p$$

One can characterize members of the exponential family of distributions using this formula.

You are given the following statements on the value of p for a given distribution:

I. Normal (Gaussian) distribution, $p = 0$
II. Compound Poisson–gamma distribution, $1 < p < 2$
III. Inverse Gaussian distribution, $p = -1$

Determine which of the above statements are correct.

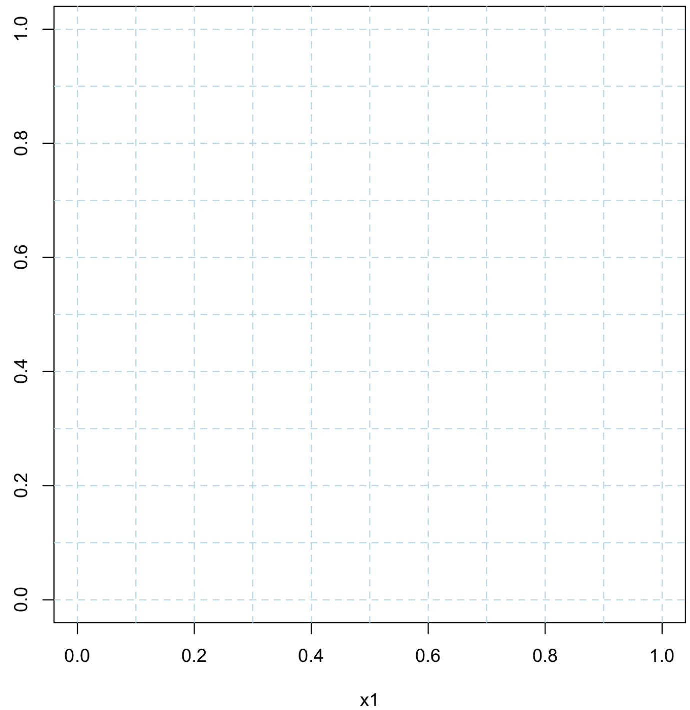
A. I only
B. I and II only
C. I and III only
D. II and III only
E. The answer is not given by (A), (B), (C), or (D)

31 (bonus) Combien de bonne(s) réponse(s) pensez vous avoir sur les 30 premières questions ?

Code permanent : Sujet : A

question 1	☐ A	☐ B	☐ C	☐ D	☐ E
question 2	☐ A	☐ B	☐ C	☐ D	☐ E
question 3	☐ A	☐ B	☐ C	☐ D	☐ E
question 4	☐ A	☐ B	☐ C	☐ D	☐ E
question 5	*au dos*				
question 6	*au dos*				
question 7	☐ A	☐ B	☐ C	☐ D	☐ E
question 8	☐ A	☐ B	☐ C	☐ D	☐ E
question 9	☐ A	☐ B	☐ C	☐ D	☐ E
question 10	☐ A	☐ B	☐ C	☐ D	☐ E
question 11	☐ A	☐ B	☐ C	☐ D	☐ E
question 12	☐ A	☐ B	☐ C	☐ D	☐ E
question 13	☐ A	☐ B	☐ C	☐ D	☐ E
question 14	☐ A	☐ B	☐ C	☐ D	☐ E
question 15	☐ A	☐ B	☐ C	☐ D	☐ E
question 16	☐ A	☐ B	☐ C	☐ D	☐ E
question 17	☐ A	☐ B	☐ C	☐ D	☐ E
question 18	☐ A	☐ B	☐ C	☐ D	☐ E
question 19	☐ A	☐ B	☐ C	☐ D	☐ E
question 20	☐ A	☐ B	☐ C	☐ D	☐ E
question 21	☐ A	☐ B	☐ C	☐ D	☐ E
question 22	☐ A	☐ B	☐ C	☐ D	☐ E
question 23	☐ A	☐ B	☐ C	☐ D	☐ E
question 24	☐ A	☐ B	☐ C	☐ D	☐ E
question 25	☐ A	☐ B	☐ C	☐ D	☐ E
question 26	☐ A	☐ B	☐ C	☐ D	☐ E
question 27	☐ A	☐ B	☐ C	☐ D	☐ E
question 28	☐ A	☐ B	☐ C	☐ D	☐ E
question 29	☐ A	☐ B	☐ C	☐ D	☐ E
question 30	☐ A	☐ B	☐ C	☐ D	☐ E
question 31					

Question 5 :



□ $\mathbb{P}[Y = 1|x_1, x_2] > \mathbb{P}[Y = 0|x_1, x_2]$ ■ $\mathbb{P}[Y = 1|x_1, x_2] < \mathbb{P}[Y = 0|x_1, x_2]$

Question 6 :