

# Simulation d'examen

**ATTENTION** : Ce sujet n'est qu'une simulation, c'est-à-dire qu'à l'heure où il écrit, il ne couvre pas l'ensemble du programme de l'examen puisqu'il reste encore deux chapitres à traiter.

**Quelques indications pour traiter le sujet :**

- a) Vous devez avoir lu tous les cours qui sont téléchargeables sur le site pour traiter le sujet.
- b) Servez-vous le plus possible des Annexes pour répondre aux questions posées, ce qui signifie que rares sont les calculs.

*Cet exercice est issu du livre d'exercices de François Husson et de Jérôme Pagès intitulé Statistiques générales pour utilisateurs, éditions PUR.*

**Exercice I.1.** Prédiction du maximum d'ozone à Rennes

Air Breizh est un organisme qui travaille sur la qualité de l'air en Bretagne et plus particulièrement sur les prévisions des pics d'ozone (= fortes concentrations en ozone) dans la ville de Rennes. La prévision d'un pic d'ozone incite les autorités locales à prendre des mesures comme la réduction de la vitesse et à prévenir la population des risques liés à la pollution (notamment les asthmatiques et les personnes souffrant de problèmes respiratoires). Pour prévoir des pics d'ozone, Air Breizh utilise 11 données (ou prévisions) météorologiques du jour ainsi que la concentration maximum d'ozone du jour précédent. Un extrait des données est présenté dans le tableau ci-dessous.

- La première colonne correspond à la date de mesure ;
- max03 correspond à la concentration maximum d'ozone atteinte dans la journée (en  $\mu\text{g}/\text{m}^3$ ) ;
- T6, T9, T12, T15, T18 correspondent respectivement à la température prévue à 6h00, 9h00, 12h00, 15h00 et 18h00 ;
- N6, N9, N12, N15 et N18 correspondent à la nébulosité prévue à 6h00, 9h00, 12h00, 15h00 et 18h00 ;
- Vx correspond à la vitesse du vent sur l'axe Est-Ouest (en  $\text{m.s}^{-1}$ ) et
- max03v correspond à la concentration maximum d'ozone mesurée la veille (en  $\mu\text{g}/\text{m}^3$ ).

Date	max03	T6	T9	T12	T15	T18	N6	N9	N12	N15	N18	Vx	max03v
28-07-94	22.2	17.1	17.1	20.2	20.4	20.1	7	8	7	7	4	-0.52	62.8
29-07-04	47.4	15.8	20.9	25.5	27.1	25.4	7	1	0	7	7	0.93	22.2
30-07-94	52	18.7	22.6	26.6	29.2	27.6	6	6	7	4	4	-4.59	47.4
⋮	⋮	⋮	⋮										
29-07-96	65.6	17.7	19.3	24.2	25.5	22.0	8	7	6	6	9	-3.94	46.2
30-07-96	66.8	17.1	19.4	21.0	24.3	23.8	8	7	7	5	6	-1.97	65.6

1. Écrire le modèle permettant de prédire le maximum d'ozone en fonction des données météorologiques.

Les résultats de la régression ont été recopiés dans le tableau ci-dessous.

Analyse de régression : maxO3 en fonction de T6; T9; ...

L'équation de régression est

$$\begin{aligned} \text{maxO3} = & 12,9 - 1,87 \text{ T6} + 1,05 \text{ T9} - 1,46 \text{ T12} - 0,14 \text{ T15} + 3,22 \text{ T18} - 2,63 \text{ Ne6} \\ & + 0,54 \text{ Ne9} + 0,03 \text{ Ne12} - 2,28 \text{ Ne15} + 1,55 \text{ Ne18} + 1,36 \text{ Vx} \\ & + 0,574 \text{ maxO3v} \end{aligned}$$

Régresseur	Coef	Er-T coef	T	P
Constante	12,90	14,48	0,89	0,376
T6	-1,872	1,236	-1,51	0,134
T9	1,051	2,185	0,48	0,632
T12	-1,458	2,375	-0,61	0,541
T15	-0,137	2,135	-0,06	0,949
T18	3,220	1,353	2,38	0,020
Ne6	-2,6254	0,9125	-2,88	0,005
Ne9	0,535	1,376	0,39	0,698
Ne12	0,027	1,560	0,02	0,986
Ne15	-2,284	1,792	-1,27	0,206
Ne18	1,547	1,339	1,16	0,251
Vx	1,3562	0,6839	1,98	0,051
maxO3v	0,57420	0,05634	10,19	0,000

S = 15,77      R-carré = 82,6%    R-carré (ajust) = 79,9%

Analyse de variance

Source	DL	SC	CM	F	P
Régression	12	91998,7	7666,6	30,84	0,000
Erreur résid	78	19389,9	248,6		
Total	90	111388,5			

2. Construire le test permettant de tester la nullité simultanée des coefficients (hypothèses, statistique de test sous  $\mathcal{H}_0$ , décision). Le modèle étudié est-il intéressant ?
3. Détailler le test concernant la significativité du coefficient de la température à 6h00 et du coefficient de la température à 18h00 (hypothèses, statistique de test, loi de la statistique de test sous  $\mathcal{H}_0$ , décision).
4. À partir des informations contenues dans le tableau suivant quel sous-ensemble choisiriez-vous pour construire un modèle de régression ?

La réponse est max03

Vars	R-carré	R-carré(ajust)	C-p	S	T	T	1	1	1	e	e	1	1	1	V	3	x	v
1	58,1	57,7	100,6	22,891														X
1	38,3	37,6	189,4	27,784														X
2	71,4	70,7	43,3	19,036														X
2	69,4	68,7	52,0	19,672														X
3	78,7	78,0	12,4	16,512														X
3	77,7	76,9	16,8	16,891	X													X
4	80,7	79,8	5,4	15,800	X													X
4	80,2	79,3	7,5	15,997														X X
5	81,7	80,6	2,9	15,480	X													X X
5	81,4	80,3	4,5	15,632		X												X X
6	82,0	80,7	3,8	15,466	X	X												X X
6	81,9	80,6	3,9	15,476	X													X X X
7	82,3	80,8	4,4	15,423	X													X X X X
7	82,2	80,7	4,9	15,470	X	X												X X X
8	82,5	80,8	5,5	15,423	X	X												X X X X
8	82,4	80,7	5,7	15,450	X		X											X X X X
9	82,5	80,6	7,3	15,497	X	X	X											X X X X
9	82,5	80,6	7,3	15,500	X	X		X										X X X X
10	82,6	80,4	9,0	15,569	X	X	X		X									X X X X
10	82,6	80,4	9,2	15,584	X	X	X		X	X								X X X X X
11	82,6	80,2	11,0	15,667	X	X	X	X	X	X								X X X X
11	82,6	80,2	11,0	15,667	X	X	X		X	X	X	X	X	X	X	X	X	X X X X
12	82,6	79,9	13,0	15,767	X	X	X	X	X	X	X	X	X	X	X	X	X	X X X X

**Exercice I.2. D'après Frontier, Davout, Gentilhomme, Lagadeuc. *Statistiques pour les sciences de la vie et de l'environnement*. Dunod, 2001.**

Pour étudier la maturation gonadique d'une espèce d'échinoderme, des données de poids sec de gonades, en *mg*, sont récoltées en trois sites différents, pour les mâles et pour les femelles.

• Mâles

Site 1	21,4	20,7	18,3	35,5	23,5	38,7	41,7	35,7	36,2	20,0		
Site 2	75,1	58,9	61,5	58,7	38,4	26,4	38,0	42,5	24,5	32,6	35,5	41,9
Site 3	9,0	18,6	21,7	20,5	22,6	30,4	14,0	17,0	22,1	12,7	27,0	13,9

• Femelles

Site 1	36,7	46,2	45,7	41,2	38,9	44,8	39,7	45,4	29,9	30,7		
Site 2	61,5	53,1	46,3	32,4	23,9	38,0	25,3	51,2	33,7	56,7	48,5	51,4
	23,7	46,0	28,3	32,7								
Site 3	17,5	12,7	20,0	19,1	12,6	16,6	21,8	22,6	22,9	21,0	13,4	14,4

1. Pour chaque sexe, le poids gonadique varie-t-il significativement selon le site ?  
Pour répondre à cette question, vous serez amené à réaliser une analyse de la variance à un facteur fixe, pour chaque sexe.
2. Nous allons maintenant, procéder à une transformation des données brutes.  
Au lieu d'étudier la variable brute « poids sec de gonades », nous allons étudier la variable «  $\log(\text{poids sec de gonades})$  » et ce pour les deux sexes.  
Notons que «  $\log$  » désigne le logarithme népérien.  
Pour chaque sexe, le logarithme du poids gonadique varie-t-il significativement selon le site ? Pour répondre à cette question, vous serez amené à réaliser une analyse de la variance à un facteur fixe, pour chaque sexe.
3. En quoi cette transformation de variable brute était-elle nécessaire ?
4. Pour chaque sexe, entre quels sites existe-t-il une différence ?
5. Tester, si cela est possible, simultanément les significativités des effets du site, du sexe et de leur interaction.

# Annexes

## Exercice 2

```

> SiteM<-rep(1 :3,c(10,12,12))
> SiteM
[1] 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3
> SiteM<-factor(SiteM)
> PoidsM<-c(21.4,20.7,18.3,35.5,23.5,38.7,41.7,35.7,36.2,20,75.1,
+ 58.9,61.5,58.7,38.4,26.4,38,42.5,24.5,32.6,35.5,41.9,9,18.6,21.7,
+ 20.5,22.6,30.4,14,17,22.1,12.7,27,13.9)
> PoidsM
[1] 21.4 20.7 18.3 35.5 23.5 38.7 41.7 35.7 36.2 20.0 75.1 58.9 61.5
+ 58.7 38.4 26.4 38.0 42.5 24.5 32.6 35.5 41.9 9.0 18.6 21.7 20.5
+ 22.6 30.4 14.0 17.0 22.1 12.7 27.0 13.9
> Masculin<-data.frame(SiteM,PoidsM)
> str(Masculin)
'data.frame' : 34 obs. of 2 variables :
SiteM : Factor w/ 3 levels "1","2","3" : 1 1 1 1 1 1 1 1 1 1 ...
PoidsM : num 21.4 20.7 18.3 35.5 23.5 38.7 41.7 35.7 36.2 20 ...
---
> model1<-aov(PoidsM~SiteM,data=Masculin)
> summary(model1)

              Df Sum Sq Mean Sq F value    Pr(>F)
SiteM           2  3912.6   1956.3   15.785 1.873e-05 ***
Residuals     31  3842.0    123.9
---
> residus1<-residuals(model1)
> shapiro.test(residus1)

Shapiro-Wilk normality test

data : residus1
W = 0.9679, p-value = 0.4048
---
> bartlett.test(residus1~SiteM,data=Masculin)

Bartlett test of homogeneity of variances

data : residus1 by SiteM
Bartlett's K-squared = 8.7528, df = 2, p-value = 0.01257
---
> LPoidsM<-log(PoidsM)
> LMasculin<-data.frame(SiteM,LPoidsM)
> str(LMasculin)
'data.frame' : 34 obs. of 2 variables :
SiteM : Factor w/ 3 levels "1","2","3" : 1 1 1 1 1 1 1 1 1 1 ...

```

```

LPoidsM : num 3.06 3.03 2.91 3.57 3.16 ...
---
> model2<-aov(LPoidsM~SiteM,data=LMasculin)
> summary(model2)

              Df Sum Sq Mean Sq F value    Pr(>F)
SiteM          2  4.2464   2.1232  18.389 5.424e-06 ***
Residuals     31  3.5794   0.1155
---
> residus2<-residuals(model2)
> shapiro.test(residus2)

Shapiro-Wilk normality test

data : residus2
W = 0.9688, p-value = 0.4297
---
> bartlett.test(residus2~SiteM,data=LMasculin)

Bartlett test of homogeneity of variances

data : residus2 by Site
Bartlett's K-squared = 0.0666, df = 2, p-value = 0.9672
---
> TukeyHSD(model2)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit : aov(formula = LPoidsM ~ SiteM, data = LMasculin)

SiteM

              diff            lwr            upr            p adj
2-1   0.4128804   0.05479326   0.77096745   0.0210732
3-1  -0.4283498  -0.78643694  -0.07026275   0.0163107
3-2  -0.8412302  -1.18265284  -0.49980755   0.0000030
---
> SiteF<-rep(1 :3,c(10,16,12))
> SiteF
[1] 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3
3 3 3 3
> SiteF<-factor(SiteF)
> PoidsF<-c(36.7,46.2,45.7,41.2,38.9,44.8,39.7,45.4,29.9,30.7,61.5,
+ 53.1,46.3,32.4,23.9,38,25.3,51.2,33.7,56.7,48.5,51.4,23.7,46,28.3,
+ 32.7,17.5,12.7,20,19.1,12.6,16.6,21.8,22.6,22.9,21,13.4,14.4)
> PoidsF
[1] 36.7 46.2 45.7 41.2 38.9 44.8 39.7 45.4 29.9 30.7 61.5 53.1 46.3

```

```

+ 32.4 23.9 38.0 25.3 51.2 33.7 56.7 48.5 51.4 23.7 46.0 28.3 32.7
+ 17.5 12.7 20.0 19.1 12.6 16.6 21.8 22.6 22.9 21.0 13.4 14.4
> Feminin<-data.frame(SiteF,PoidsF)
> str(Feminin)
'data.frame' : 38 obs. of 2 variables :
SiteF : Factor w/ 3 levels "1","2","3" : 1 1 1 1 1 1 1 1 1 1 ...
PoidsF : num 36.7 46.2 45.7 41.2 38.9 44.8 39.7 45.4 29.9 30.7 ...
---
> model3<-aov(PoidsF~SiteF,data=Feminin)
> summary(model3)

              Df Sum Sq Mean Sq F value Pr(>F)
      SiteF    2  4188.8  2094.4   25.893 1.254e-07 ***
Residuals   35  2831.1   80.9
---
> residus3<-residuals(model3)
> shapiro.test(residus3)

Shapiro-Wilk normality test

data : residus3
W = 0.9846, p-value = 0.8691
---
> bartlett.test(residus3~SiteF)

Bartlett test of homogeneity of variances

data : residus3 by SiteF
Bartlett's K-squared = 15.1887, df = 2, p-value = 0.0005033
---
> LPoidsF<-log(PoidsF)
> LPoidsF
[1] 3.602777 3.832980 3.822098 3.718438 3.660994 3.802208 3.681351
+ 3.815512 3.397858 3.424263 4.119037 3.972177 3.835142 3.478158
+ 3.173878 3.637586 3.230804 3.935740 3.517498 4.037774 3.881564
+ 3.939638 3.165475 3.828641 3.342862 3.487375 2.862201 2.541602
+ 2.995732 2.949688 2.533697 2.809403 3.081910 3.117950 3.131137
+ 3.044522 2.595255 2.667228
> LFeminin<-data.frame(SiteF,LPoidsF)
> str(LFeminin)
'data.frame' : 38 obs. of 2 variables :
SiteF : Factor w/ 3 levels "1","2","3" : 1 1 1 1 1 1 1 1 1 1 ...
LPoidsF : num 3.60 3.83 3.82 3.72 3.66 ...
---
> model4<-aov(LPoidsF~SiteF,data=Feminin)
> summary(model4)

```

```

              Df Sum Sq Mean Sq F value Pr(>F)
SiteF        2   5.3369   2.6684  39.581  1.034e-09 ***
Residuals    35   2.3596   0.0674
---
> residus4<-residuals(model4)
> shapiro.test(residus4)

Shapiro-Wilk normality test

data : residus4
W = 0.9596, p-value = 0.1842
---
> bartlett.test(residus4~SiteF)

Bartlett test of homogeneity of variances

data : residus4 by SiteF
Bartlett's K-squared = 4.9664, df = 2, p-value = 0.08347
---
> TukeyHSD(model4)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit : aov(formula = LPoidsF SiteF, data = Feminin)

SiteF
      diff       lwr       upr     p adj
2-1 -0.01438859 -0.2705396  0.2417624 0.9896387
3-1 -0.81498756 -1.0870635 -0.5429116 0.0000000
3-2 -0.80059897 -1.0432588 -0.5579392 0.0000000
---
> Site<-c(SiteM,SiteF)
> Site
[1] 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3
1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3
3 3
> Site<-factor(Site)
> Sexe<-rep(1 :2,c(34,38))
> Sexe
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2
> Sexe<-factor(Sexe)
> LPoids<-c(LPoidsM,LPoidsF)
> LPoids
[1] 3.063391 3.030134 2.906901 3.569533 3.157000 3.655840 3.730501
```



```

3.575151 3.589059 2.995732 4.318821 4.075841 4.119037 4.072440
3.648057 3.273364 3.637586 3.749504 3.198673 3.484312 3.569533
3.735286 2.197225 2.923162 3.077312 3.020425 3.117950 3.414443
2.639057 2.833213 3.095578 2.541602 3.295837 2.631889 3.602777
3.832980 3.822098 3.718438 3.660994 3.802208 3.681351 3.815512
3.397858 3.424263 4.119037 3.972177 3.835142 3.478158 3.173878
3.637586 3.230804 3.935740 3.517498 4.037774 3.881564 3.939638
3.165475 3.828641 3.342862 3.487375 2.862201 2.541602 2.995732
2.949688 2.533697 2.809403 3.081910 3.117950 3.131137 3.044522
2.595255 2.667228

```

```
> Lesdeux<-data.frame(Site,Sexe,LPoids)
```

```
> str(Lesdeux)
```

```
'data.frame' : 72 obs. of 3 variables :
```

```
Site : Factor w/ 3 levels "1","2","3" : 1 1 1 1 1 1 1 1 1 1 ...
```

```
Sexe : Factor w/ 2 levels "1","2" : 1 1 1 1 1 1 1 1 1 1 ...
```

```
LPoids : num 3.06 3.03 2.91 3.57 3.16 ...
```

```
---
```

```
> model5<-aov(LPoids~Site*Sexe,data=Lesdeux)
```

```
> summary(model5)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Site	2	9.0719	4.5360	50.4080	5.142e-14	***
Sexe	1	0.0531	0.0531	0.5904	0.44502	
Site :Sexe	2	0.6055	0.3027	3.3642	0.04062	*
Residuals	66	5.9390	0.0900			

```
---
```

```
> residus5<-residuals(model5)
```

```
> shapiro.test(residus5)
```

Shapiro-Wilk normality test

data : residus5

W = 0.9765, p-value = 0.1953

```
---
```

```
> bartlett.test(residus5~Site*Sexe,data=Lesdeux)
```

Bartlett test of homogeneity of variances

data : residus5 by Site by Sexe

Bartlett's K-squared = 1.6725, df = 2, p-value = 0.4333