ESIR 1 2014/2015

Contrôle continu de Statistiques N°2

Jeudi 15 Janvier 2015

Documents et calculatrices autorisés

Durée : 1 heure

Exercice 1

On dispose de deux jeux de données indépendants issus d'un même phénomène aléatoire de moyenne inconnue μ et de variance connue σ^2. Le premier à une taille N_1 et le second à une taille N_2 avec $N_1 < N_2$.

1. Quel jeu de données permet d'avoir la meilleure précision sur l'estimation de la moyenne ?
2. Dans quelle proportion évolue l'intervalle de confiance lorsque l'on passe du premier au second jeu de données (sans les fusionner)
3. Dans quelle proportion évolue l'intevalle de confiance lorsque l'on passe du premier jeu de données à la fusion des deux jeux de données (1 et 2).

Exercice 2

Un échantillon de 15 hommes dont la taille et le poids ont été mesurés approximativement est donné ci dessous.

Index k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_k : Poids (kg)	77	74	92	84	81	83	80	78	75	83	71	87	65	79	111
Y_k : Taille (cm)	191	176	207	186	187	190	187	189	191	205	173	191	172	186	185

Les moyennes empiriques du poids et de la taille sont respectivement:
$\bar{X} = \sum_{k=1}^{15} X_k = 80.33$ et $\bar{Y} = \sum_{k=1}^{15} Y_k = 187.733$
Les variances empiriques du poids et de la taille sont respectivement:
$\hat{\sigma}_X^2 = \sum_{k=1}^{15} X_k^2 - \bar{X}^2 = 102.88$ et $\hat{\sigma}_Y^2 = \sum_{k=1}^{15} Y_k^2 - \bar{Y}^2 = 88.328$
Les variances non biaisées du poids et de la taille sont respectivement:
$s_X^2 = 110.239$ et $s_Y^2 = 94.639$
En appliquant la fonction t.test de R à l'échantillon de poids, on obtient le résultat suivant :

```
One Sample t-test

data:  PH
t = 30,0019, df = 14, p-value = 4,168e-14
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 75,51894 87,14773
sample estimates:
mean of x
 81,33333
```

En appliquant la fonction t.test de R à l'échantillon de taille, on obtient le résultat suivant :

```
One Sample t-test
```

```
data:  TH
t = 74,7402, df = 14, p-value < 2,2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 182,3460 193,1206
sample estimates:
mean of x
 187,7333
```

1. Quelle relation lie la variance non biaisée à la variance empirique?
2. Quelle variance doit être utilisée préférentiellement ici ? Pourquoi ?
3. Préciser l'intervalle de confiance à 95% pour le poids et la taille en détaillant la démarche.
4. Quel aurait été l'intervalle de confiance à 95% en faisant usage de la loi normale centrée réduite et de la variance biaisée ?
5. En considérant les appels à la fonction t.test ci dessus, quelles étaient les hypothèses nulles ? Quelles conclusions peut-on tirer du résulat de ces 2 tests ?

Exercice 3

On dispose d'un jeu de données d . On réalise le test bilatéral suivant :

```
t.test(d,mu=118.6073)

One Sample t-test

data:  d
t = 2,0931, df = 19, p-value = 0,04999
alternative hypothesis: true mean is not equal to 118,6073
95 percent confidence interval:
 118,6074 120,7015
sample estimates:
mean of x
 119,6544
```

1. Quelle est la taille de l'échantillon ?
2. Quelle est l'hypothèse nulle ?
3. Que doit-on conclure de ce test ?

Exercice 4

Lors d'une élection présidentielle au suffrage universel direct à 2 tours, les résultats du premier tour, restreints aux 2 candidats (A et B) arrivés en tête au premier tour, exprimés en nombre de voix est donné dans le tableau 1 ci-dessous pour : la France, la Bretagne, la Loire Atlantique et **l'Ille et Vilaine**. La catégorie "Autres" agrège les voix de tous les autres candidats et les bulletins blancs et nuls.

Zone	Candidat A	Candidat B	Autres
France	10272705	9753629	16558065
Bretagne	628441	508072	869357
Loire Atlantique	245708	201671	340514
Ille et Vilaine	183935	150685	255451

Tableau 1

Le tableau 2 ci dessous désigne les résultats attendus en **Ille et Vilaine** sous l'hypothèse que les résultats se distribuent selon la distribution nationale, Bretonne ou de celle de Loire Atlantique. Les résultats observés en Ille et Vilaine sont rappelés dans la dernière ligne du tableau.

Zone	Candidat A	Candidat B	Autres
Ille et Vilaine ˜ France	165688	157316	267065
Ille et Vilaine˜ Bretagne	184869	149460	255740
Ille et Vilaine ˜ Loire Atlantique	184016	151036	255018
Ille et Vilaine résultats observés	183935	150685	255451

Tableau 2

Le tableau 3 ci-dessous précise des écarts quadratiques normalisés (EQN) utiles au problème posé.

Zone	Candidat A	Candidat B	Autres
EQN France	2009.51794	279.502155	505.064295
EQN Bretagne	4.718779	10.040312	0.326586
EQN Loire Atlantique	0.035655	0.815706	0.735199

Tableau 3

On chercher à situer le vote du département d'Ille et Vilaine vis à vis du vote national, du vote breton et du vote de la Loire-atlantique.

1. Quel type de test allez vous mettre en oeuvre ?
2. Formuler l'hypothèse nulle des tests statistiques à mener pour situer le vote en Ille et Vilaine vis à vis des autres zones géographiques.
3. Indiquez comment est construit le Tableau 2 à partir du tableau 1
4. Indiquez comment est construit le Tableau 3 à partir du tableau 2
5. Formuler des conclusions sur le comportement du corps électoral d'Ille et Vilaine basées sur vos résultats. Détailler la démarche et le raisonnement.

Tables de quantiles

Les tables de quantiles ci-dessous sont fournies à toutes fins utiles.

Loi de Student à n degrés de liberté.

n|p	0.75	0.85	0.9	0.95	0.975	0.98	0.99	0.995	0.9975	0.9995
1	1.	1.96	3.07	6.31	12.7	15.89	31.82	63.65	127.32	636.61
2	0.81	1.38	1.88	2.91	4.3	4.84	6.96	9.92	14.08	31.59
3	0.76	1.24	1.63	2.35	3.18	3.48	4.54	5.84	7.45	12.92
4	0.74	1.18	1.53	2.13	2.77	2.99	3.74	4.6	5.59	8.61
5	0.72	1.15	1.47	2.01	2.57	2.75	3.36	4.03	4.77	6.86
6	0.71	1.13	1.43	1.94	2.44	2.61	3.14	3.7	4.31	5.95
7	0.71	1.11	1.41	1.89	2.36	2.51	2.99	3.49	4.02	5.4
8	0.7	1.1	1.39	1.85	2.3	2.44	2.89	3.35	3.83	5.04
9	0.7	1.09	1.38	1.83	2.26	2.39	2.82	3.24	3.68	4.78
10	0.69	1.09	1.37	1.81	2.22	2.35	2.76	3.16	3.58	4.58
11	0.69	1.08	1.36	1.79	2.2	2.32	2.71	3.1	3.49	4.43
12	0.69	1.08	1.35	1.78	2.17	2.3	2.68	3.05	3.42	4.31
13	0. 69	1.07	1.35	1.77	2.16	2.28	2.65	3.01	3.37	4.22
14	0.69	1.07	1.34	1.76	2.14	2.26	2.62	2.97	3.32	4.14
15	0.69	1.07	1.34	1.75	2.13	2.24	2.6	2.94	3.28	4.07

Loi du χ^2 à n degrés de liberté.

n|p	0.75	0.85	0.9	0.95	0.975	0.98	0.99	0.995	0.9975	0.9995
1	1.32	2.07	2.7	3.84	5.02	5.41	6.63	7.87	9.14	12.11
2	2.77	3.79	4.6	5.99	7.37	7.82	9.21	10.59	11.98	15.2
3	4.1	5.31	6.25	7.81	9.34	9.83	11.34	12.83	14.32	17.72
4	5.38	6.74	7.77	9.48	11.14	11.66	13.27	14.86	16.42	19.99
5	6.62	8.11	9.23	11.07	12.83	13.38	15.08	16.74	18.38	22.1
6	7.84	9.44	10.64	12.59	14.44	15.03	16.81	18.54	20.24	24.1
7	9.03	10.74	12.01	14.06	16.01	16.62	18.47	20.27	22.04	26.01
8	10.21	12.02	13.36	15.5	17.53	18.16	20.09	21.95	23.77	27.86
9	11.38	13.28	14.68	16.91	19.02	19.67	21.66	23.58	25.46	29.66
10	12.54	14.53	15.98	18.3	20.48	21.16	23.2	25.18	27.11	31.41
11	13.7	15.76	17.27	19.67	21.92	22.61	24.72	26.75	28.72	33.13
12	14.84	16.98	18.54	21.02	23.33	24.05	26.21	28.29	30.31	34.82
13	15.98	18.2	19.81	22.36	24.73	25.47	27.68	29.81	31.88	36.47
14	17.11	19.4	21.06	23.68	26.11	26.87	29.14	31.31	33.42	38.1
15	18.24	20.6	22.3	24.99	27.48	28.25	30.57	32.8	34.94	39.71

Loi normale centrée réduite

probabilité	0.75	0.85,	0.9,	0.95	0.975	0.98	0.99
quantile	0.674	1.036	1.281	1.644	1.959	2.053	2