

Corrigés des exercices de Statistique Inférentielle

1. On va faire un test du chi-deux d'indépendance. La population est alors l'ensemble des personnes vivant en Europe. L'échantillon étudié est constitué de $N = 60$ d'entre elles. Les variables considérées sont X : réponse à la question "pensez-vous être superstitieux ?" variable nominale à deux modalités (oui et non) et Y : influence de la superstition dans le comportement de la vie quotidienne, variable nominale (voire ordinale) à 3 modalités (très (influencé), un peu (influencé), pas du tout (influencé)).

On écrit alors toutes les étapes nécessaires pour effectuer le test :

- Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle

H_0 : "Sur la population, il n'y a pas de lien entre la réponse à la question et le comportement" (l'hypothèse alternative étant H_1 : "Sur la population, il existe un lien entre la réponse à la question et le comportement").

- Détermination de la valeur critique χ_α^2 :

→ le seuil imposé par l'énoncé est $\alpha = 0.05$ (= 5%) ;

→ X ayant $p = 2$ modalités et Y ayant $q = 3$ modalités, on a $d.d.l. = (p - 1)(q - 1) = 2$;

→ on lit dans la table du chi-deux $\chi_\alpha^2 = 5.991$; (ainsi, si H_0 est vraie, pour 95% des échantillons, la valeur du χ^2 correspondant est inférieure à 5.991).

- Calcul du χ^2 correspondant à notre échantillon :

Pour cela, on commence par établir le tableau des effectifs observés pour chacune des modalités du couple de variables (X, Y) :

L'énoncé nous donne :

→ les marges de X , soit $L_1 = 28$ et $L_2 = 32$ (superflu, car on sait que $N = 60$)

→ une partie des marges de Y : $C_2 = 30$ (une personne sur deux), $C_3 = 14$ et on en déduit $C_1 = N - C_2 - C_3 = 16$.

→ $O_{21} = 0$: aucune personne de l'échantillon ayant répondu "non" n'est très influencée par la superstition ;

→ Sur les 30 personnes un peu influencées, deux-tiers, c'est-à-dire 20 ont répondu "non" (et donc 10 ont répondu "oui", ce qui permet d'avoir entièrement la deuxième colonne).

On a, pour l'instant :

$X Y$	très	un peu	du tout	total
oui		10		28
non	0	20		32
total	16	30	14	60

On peut alors facilement compléter le tableau et on a le tableau des effectifs observés O_{ij}

$X Y$	très	un peu	du tout	total
oui	16	10	2	28
non	0	20	12	32
total	16	30	14	60

puis le tableau des effectifs théoriques T_{ij} :

$X Y$	très	un peu	du tout	total
oui	7,5	14	6,5	28
non	8,5	16	7,5	32
total	16	30	14	60

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test et, sans même calculer χ^2 , on a $\chi^2 \geq \frac{(O_{21} - T_{21})^2}{T_{21}} = T_{21} \approx 8.5 > 5.991$.

• Conclusion du test : $\chi^2 > 5.991 = \chi_{\alpha}^2$ ce qui nous conduit à rejeter H_0 , c'est-à-dire que les observations faites sur l'échantillon ne sont pas compatibles avec l'indépendance de la réponse à la question et du comportement au seuil de 5%. Le test est donc significatif au seuil de 5%. C'est le risque d'erreur en ce sens que, si X et Y ne sont pas liées, notre échantillon figure parmi les 5% dont le χ^2 est supérieur à la valeur critique.

Remarque : Le résultat est rassurant quant à la connaissance qu'ont les personnes d'elles-mêmes...

2. On va faire un test du chi-deux d'indépendance. L'échantillon étudié ici est constitué de $N = 106$ patients américains, Ω est la population des américains. Sur cette population, les variables considérées sont X "fumeurs?" qui associe à chaque individu la modalité "est fumeur" ou "n'est pas fumeur" suivant son comportement, et Y "cancer du poumon" qui associe à chaque individu la modalité "a un cancer du poumon" ou "n'a pas de cancer du poumon" suivant son état. On écrit alors toutes les étapes nécessaires pour effectuer le test :

• Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle H_0 : "Sur la population des américains, il n'y a pas de lien entre le fait de fumer et avoir un cancer du poumon" l'hypothèse alternative étant H_1 : "Sur la population, il existe un lien entre le fait de fumer et avoir un cancer du poumon".

• Détermination de la valeur critique χ_{α}^2 :
 → le seuil imposé par l'énoncé est $\alpha = 0.05$ (= 5%) ;
 → X ayant $p = 2$ modalités et Y ayant $q = 2$ modalités, on a $d.d.l. = (p - 1)(q - 1) = 1$;
 → on lit dans la table du chi-deux : $\chi_{\alpha}^2 = 3.841$; (ainsi, si H_0 est vraie, pour 95% des échantillons, la valeur du χ^2 correspondante est inférieure à 3.841).

• Calcul du χ^2 correspondant à notre échantillon :

tableau des O_{ij}	cancer	pas de cancer	Total
fumeur	60	32	92
pas fumeur	3	11	14
Total	63	43	106

tableau des T_{ij}	cancer	pas de cancer	Total
fumeur	55	37	92
pas fumeur	8	6	14
Total	63	43	106

tableau des $\frac{(O_{ij} - T_{ij})^2}{T_{ij}}$	cancer	pas de cancer
fumeur	0.52	0.76
pas fumeur	3.4	3.98

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test et, sans même calculer χ^2 , on a $\chi^2 \geq \frac{(O_{22} - T_{22})^2}{T_{22}} \approx 3.98 > 3.841$.

• Conclusion du test : $\chi^2 > 3.841 = \chi_\alpha^2$ ce qui nous conduit à rejeter H_0 , c'est-à-dire que les observations faites sur l'échantillon ne sont pas compatibles avec l'absence de lien entre fumer et avoir un cancer du poumon. Le test est donc significatif au seuil de 5%. C'est le risque d'erreur en ce sens que, si X et Y ne sont pas liées, notre échantillon figure parmi les 5% dont le χ^2 est supérieur à la valeur critique.

3. Ω est ici la population des malades en général, soignés par un certain médicament. Sur cette population, les variables considérées sont X "moment de prise du médicament" avec pour modalités : matin, après-midi, soir, et Y "guérison" avec pour modalités : oui, non. On effectue un test du χ^2 d'indépendance et on écrit alors toutes les étapes nécessaires pour effectuer le test :

• Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle H_0 : "Sur la population des malades Ω , il n'y a pas de lien entre le moment de prise de médicament et la guérison" ; l'hypothèse alternative étant H_1 : "Sur la population Ω , il existe un lien entre le moment de prise de médicament et la guérison".

• Détermination de la valeur critique χ_α^2 :
 → le seuil imposé par l'énoncé est $\alpha = 0.05$ (= 5%) ;
 → X ayant $p = 3$ modalités et Y ayant $q = 2$ modalités, on a *d.d.l.* = $(p - 1)(q - 1) = 2$;
 → on lit dans la table du chi-deux : $\chi_\alpha^2 = 5.991$; (ainsi, si H_0 est vraie, pour 95% des échantillons, la valeur du χ^2 correspondant est inférieure à 5.991).

• Calcul du χ^2 correspondant à notre échantillon :

tableau des O_{ij}	oui	non	Total
matin	141	59	200
après-midi	125	75	200
soir	154	46	200
Total	420	180	600

tableau des T_{ij}	oui	non	Total
matin	140	60	200
après-midi	140	60	200
soir	140	60	200
Total	420	180	600

tableau des $\frac{(O_{ij} - T_{ij})^2}{T_{ij}}$	oui	non
matin	0.01	0.02
après-midi	1.61	3.75
soir	1.40	3.27

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test et, en faisant la somme des coefficients dans le tableau précédent, on obtient $\chi^2 = 10.05$.

• Conclusion du test : $\chi^2 = 10.05 > 5.991 = \chi_\alpha^2$ ce qui nous conduit à rejeter H_0 , c'est-à-dire que les observations faites sur l'échantillon ne sont pas compatibles avec l'absence de lien entre le moment de prendre le médicament et la guérison. Le test est donc significatif au seuil de 5%. C'est le risque d'erreur en ce sens que, si X et Y ne sont pas liées, notre échantillon figure parmi les 5% dont le χ^2 est supérieur à la valeur critique.

4. Ω est ici la population des habitants des États-Unis appartenant à l'une des quatre catégories socio-professionnelles citées. Sur cette population, les variables considérées sont X "Catégorie Socio-professionnelle" avec pour modalités : clergé, éducateurs, dirigeant d'entreprise, commerçant et Y "alcoolique ?" avec pour modalités : oui, non. On effectue un test du χ^2 d'indépendance et on écrit alors toutes les étapes nécessaires pour effectuer le test :

• Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle H_0 : "Sur la population des américains, il n'y a pas de lien entre le fait d'être alcoolique et la catégorie socio-professionnelle" ; l'hypothèse alternative H_1 étant : "Sur la population des américains, il existe un lien entre le fait d'être alcoolique et la catégorie socio-professionnelle".

• Détermination de la valeur critique χ_α^2 :
 → le seuil imposé par l'énoncé est $\alpha = 0.10$ (= 10%) ;
 → X ayant $p = 4$ modalités et Y ayant $q = 2$ modalités, on a $d.d.l. = (4 - 1)(2 - 1) = 3$;
 → on lit dans la table du chi-deux : $\chi_\alpha^2 = 6.251$; (ainsi, si H_0 est vraie, pour 90% des échantillons, la valeur du χ^2 correspondante est inférieure à 6, 251).

• Calcul du χ^2 correspondant à notre échantillon :

Tableau des O_{ij}	Alcoolique	Non alcoolique	Total
Clergé	32	268	300
Éducateurs	51	199	250
Dirigeants d'entreprise	67	233	300
Commerçants	83	267	350
Total	233	967	1200

Tableau des T_{ij}	Alcoolique	Non alcoolique
Clergé	58.25	241.75
Éducateurs	48.54	201.46
Dirigeants d'entreprise	58.25	241.75
Commerçants	67.96	282.04

Tableau des $\frac{(O_{ij} - T_{ij})^2}{T_{ij}}$	Alcoolique	Non alcoolique
Clergé	11.83	2.85
Éducateurs	0.12	0.03
Dirigeants d'entreprise	1.31	0.32
Commerçants	3.33	0.80

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test et, en faisant la somme des coefficients dans le tableau précédent, on obtient $\chi^2 = 20.6$.

• Conclusion du test : $\chi^2 = 20.6 > 6.251 = \chi_\alpha^2$ ce qui nous conduit à rejeter H_0 , c'est-à-dire que les observations faites sur l'échantillon sont incompatibles avec l'absence de lien entre la

catégorie socio-professionnelle et l'alcoolisme. Le test est donc significatif au seuil de 5%. C'est le risque d'erreur en ce sens que, si X et Y ne sont pas liées, notre échantillon figure parmi les 5% dont le χ^2 est supérieur à la valeur critique.

5. a) Un individu est ici un adulte européen pratiquant l'apprentissage d'une langue. La variable X associée à chaque individu sa durée d'apprentissage de la langue. X est une variable quantitative dont les modalités sont réparties en 4 classes.

b) Tableau qui permet de comparer les observations au tableau "théorique"

durée d'apprentissage	n_i	f_i	Nf_i	$\frac{(n_i - Nf_i)^2}{Nf_i}$
moins de 6 mois	3	0.30	9	9
de 6 mois à moins de 12	12	0.32	9.6	0.6
de 12 mois à moins de 24	12	0.18	5.4	8.07
24 mois et plus	3	0.20	6	1.5
Total	30	1	30	19.17

c) On va faire ici un test du chi-deux d'ajustement. On écrit alors toutes les étapes nécessaires pour effectuer le test :

- Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle

H_0 : "Sur la population Ω des adultes européens apprenant une langue, la durée d'apprentissage est inférieure à 6 mois pour 30% d'entre eux, de 6 à 12 mois pour 32%, de 12 à 24 mois pour 18% et supérieure à 24 mois pour 20%" ; l'hypothèse alternative étant H_1 : "Sur la population Ω , l'une au moins des fréquences ci-dessus est incorrecte".

- Détermination de la valeur critique χ_α^2 :

→ le seuil n'étant pas imposé par l'énoncé, on prendra $\alpha = 0.05$ (= 5%) ;

→ X ayant $k = 4$ modalités, on a $d.d.l. = k - 1 = 3$;

→ on lit dans la table du chi-deux : $\chi_\alpha^2 = 7.814$; (ainsi, si H_0 est vraie, pour 95% des échantillons, la valeur du χ^2 correspondante est inférieure à 7.814).

- Calcul du χ^2 correspondant à notre échantillon :

Voit b). Les effectifs théoriques étant tous supérieurs strictement à 5, on peut effectuer le test.

- Conclusion du test : $\chi^2 = 19.17 > 7.814 = \chi_\alpha^2$ ce qui nous conduit à rejeter H_0 , c'est-à-dire

que les observations faites sur l'échantillon sont incompatibles avec les fréquences annoncées. Le test est donc significatif au seuil de 5%. C'est le risque d'erreur en ce sens que, si les fréquences annoncées sont correctes, notre échantillon figure parmi les 5% dont le χ^2 est supérieur à la valeur critique.

6. On travaille sur la population Ω des ménages d'un quartier qui est partagée en deux classes : celle des ménages ayant un micro-ondes, et celle des ménages n'en ayant pas. Les données sont observées sur un échantillon de 500 ménages. Sur cet échantillon, on observe que 250 ménages (150 + 100) ont un four à micro-ondes.

Il s'agit d'effectuer un test de distribution de fréquences (nombres de classes $k = 2$).

- Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle

H_0 : "Sur la population Ω des ménages du quartier, la fréquence des ménages ayant un micro-ondes est 0.4" ; l'hypothèse alternative étant H_1 : "Sur la population Ω des ménages du quartier, la fréquence des ménages ayant un micro-ondes n'est pas 0.4".

- Détermination de la valeur critique χ_α^2 :

→ le seuil imposé par l'énoncé est $\alpha = 0.05$ (= 5%) ;

→ La variable considérée ayant $k = 2$ modalités, on a $d.d.l. = k - 1 = 1$;

→ on lit dans la table du chi-deux : $\chi_\alpha^2 = 3.841$; (ainsi, si H_0 est vraie, pour 95% des échantillons, la valeur du χ^2 correspondante est inférieure à 3.841).

- Calcul du χ^2 correspondant à notre échantillon :

micro-ondes	n_i	f_i	Nf_i	$\frac{(n_i - Nf_i)^2}{Nf_i}$
oui	250	0.40	200	12.5
non	250	0.60	300	8.33
Total	500	1	500	20.83

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test.

• Conclusion du test : $\chi^2 = 20.83 > 3.841 = \chi_\alpha^2$ ce qui nous conduit à rejeter H_0 , c'est-à-dire que les observations faites sur l'échantillon sont incompatibles avec la fréquence annoncée.

Le test est donc significatif au seuil de 5%. C'est le risque d'erreur en ce sens que, si les fréquences annoncées sont correctes, notre échantillon figure parmi les 5% dont le χ^2 est supérieur à la valeur critique.

7. On travaille sur la population Ω des français regardant la télévision plus de 4 heures par jour, partagée en 4 classes suivant l'âge (variable X). Les données sont observées sur un échantillon de taille $N = 30$. Il s'agit d'effectuer un test de distribution de fréquences (nombres de classes $k = 4$).

- Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle

H_0 : "Sur la population Ω , la fréquence dans chaque tranche d'âge est 0.25" ; l'hypothèse alternative étant H_1 : "Sur la population Ω , la fréquence n'est pas de 0.25 pour chaque tranche d'âge".

- Détermination de la valeur critique χ_α^2 :

→ le seuil imposé par l'énoncé est $\alpha = 0.05$ (= 5%) ;

→ La variable considérée ayant $k = 4$ modalités, on a $d.d.l. = k - 1 = 3$;

→ on lit dans la table du chi-deux : $\chi_\alpha^2 = 7.814$; (ainsi, si H_0 est vraie, pour 95% des échantillons, la valeur du χ^2 correspondante est inférieure à 7.814).

- Calcul du χ^2 correspondant à notre échantillon :

X	n_i	f_i	Nf_i	$\frac{(n_i - Nf_i)^2}{Nf_i}$
moins de 20 ans	11	0.25	7.5	1.63
de 20 à 35 ans	4	0.25	7.5	1.63
de 35 à 50 ans	5	0.25	7.5	0.83
plus de 50 ans	10	0.25	7.5	0.83
Total	30	1	30	4.93

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test.

- Conclusion du test : $\chi^2 = 4.93 < 7.814 = \chi_\alpha^2$ ce qui nous conduit à accepter H_0 , c'est-à-dire que les observations faites sur l'échantillon sont compatibles avec les proportions annoncées par la presse. Le test est donc non significatif au seuil de 5%.

c) On refait le même exercice avec la même population mais seulement 2 classes ($k = 2$).

- Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle H_0 : "Sur la population Ω des français regardant la télévision plus de 4 heures par jour, la fréquence des moins de 20 ans est 0.25" ; l'hypothèse alternative étant H_1 : "Sur la population Ω , la fréquence des moins de 20 ans n'est pas 0.25".

- Détermination de la valeur critique χ_α^2 :
 → le seuil imposé par l'énoncé est $\alpha = 0.05$ (= 5%) ;
 → La variable considérée ayant $k = 2$ modalités, on a $d.d.l. = k - 1 = 1$;
 → on lit dans la table du chi-deux : $\chi_\alpha^2 = 3.841$; (ainsi, si H_0 est vraie, pour 95% des échantillons, la valeur du χ^2 correspondante est inférieure à 3.841).

- Calcul du χ^2 correspondant à notre échantillon :

X	n_i	f_i	Nf_i	$\frac{(n_i - Nf_i)^2}{Nf_i}$
moins de 20 ans	11	0.25	7.5	1.63
plus de 20 ans	19	0.75	22.5	0.54
Total	30	1	30	2.17

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test.

- Conclusion du test : $\chi^2 = 2.17 < 3.841 = \chi_\alpha^2$ ce qui nous conduit à accepter H_0 , c'est-à-dire que les observations faites sur l'échantillon sont compatibles avec les proportions annoncées par la presse. Le test est donc non significatif au seuil de 5%.

8. On travaille sur la population Ω des enfants de 10 ans. La variable X est le nombre de mots mémorisés sur une liste de 15 mots, variable quantitative dont les modalités sont regroupées en 3 classes. Il s'agit d'effectuer un test de distribution de fréquences (nombres de classes $k = 3$).

- Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle H_0 : "Sur la population Ω des enfants de 10 ans, 25% des enfants de 10 ans en France trouvent au plus 5 mots, 45% de 6 à 10 mots, 30% de 11 à 15 mots" ; l'hypothèse alternative étant H_1 : "Sur la population Ω , une fréquence au moins est différente de celles citées ci-dessus".

- Détermination de la valeur critique χ_α^2 :
 → le seuil imposé par l'énoncé est $\alpha = 0.10$ (= 10%) ;
 → La variable considérée ayant $k = 3$ modalités, on a $d.d.l. = k - 1 = 2$;
 → on lit dans la table du chi-deux : $\chi_\alpha^2 = 4.605$; (ainsi, si H_0 est vraie, pour 90% des échantillons, la valeur du χ^2 correspondante est inférieure à 4.605).

- Calcul du χ^2 correspondant à notre échantillon :

X	n_i	f_i	Nf_i	$\frac{(n_i - Nf_i)^2}{Nf_i}$
au plus 5 mots	7	0.25	7.5	0.04
de 6 à 10 mots	10	0.45	13.5	0.91
de 11 à 15 mots	13	0.30	9	1.78
Total	30	1	30	2.73

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test.

- Conclusion du test : $\chi^2 = 2.73 < 4.605 = \chi_\alpha^2$ ce qui nous conduit à accepter H_0 , c'est-à-dire que les observations faites sur l'échantillon sont compatibles avec les proportions annoncées par l'étude. Le test est donc non significatif au seuil de 10%.

9. On travaille sur la population Ω des nourissons français qui est partagée en deux classes suivant le sexe. Il s'agit d'effectuer un test de distribution de fréquences (nombres de classes $k = 2$).

- Formulation de l'hypothèse : l'hypothèse à tester est l'hypothèse nulle

H_0 : "Sur la population Ω des nourissons français, la proportion des filles est de 48.78% (et celle des garçons de 51.22%)" ; l'hypothèse alternative étant H_1 : "Sur la population Ω , la proportion des filles n'est pas de 48.78%".

- Détermination de la valeur critique χ_α^2 :

→ le seuil imposé par l'énoncé est $\alpha = 0.05$ (= 5%) ;

→ La variable considérée ayant $k = 2$ modalités, on a $d.d.l. = k - 1 = 1$;

→ on lit dans la table du chi-deux : $\chi_\alpha^2 = 3.841$;

(ainsi, si H_0 est vraie, pour 95% des échantillons, la valeur du χ^2 correspondante est inférieure à 3.841).

- Calcul du χ^2 correspondant à notre échantillon :

sexe	n_i	f_i	Nf_i	$\frac{(n_i - Nf_i)^2}{Nf_i}$
filles	40	0.4878	36.59	0.32
garçon	35	0.5122	38.42	0.30
Total	75	1	75	0.62

Les effectifs théoriques étant tous supérieurs strictement à 5, on peut bien effectuer le test et $\chi^2 = 0.62$.

- Conclusion du test : $\chi^2 = 0.62 < 3.841 = \chi_\alpha^2$ ce qui nous conduit à accepter H_0 , c'est-à-dire que les observations faites sur l'échantillon sont compatibles avec 48.78% de filles sur Ω . Le test est donc non significatif au seuil de 5%.