UNITAT 8: PROVA DE KHI QUADRAT.

Situació 1

1. Determini si existeix relació estadísticament significativa entre el sexe i el fet de ser fumador.

1er. Plantejament de la hipòtesi nul·la:

H_0: $fo_{ij} = fe_{ij}$ → No hi ha relació estadísticament significativa entre el sexe i el fet de ser fumador.

2n. Càlcul de l'estadístic de contrast: Khi-quadrat de Pearson

		Fumador		Total
		Sí	No	
Sexe	Home	11 (9,07)	5 (6,93)	16
	Dona	6 (7,93)	8 (6,07)	14
Total		17	13	30

$$fe_{11} = \frac{16 \cdot 17}{30} = 9{,}07 \qquad fe_{12} = \frac{16 \cdot 13}{30} = 6{,}93$$
$$fe_{21} = \frac{14 \cdot 17}{30} = 7{,}93 \qquad fe_{22} = \frac{14 \cdot 13}{30} = 6{,}07$$

$$\chi^2 = \sum\sum \frac{\left(fo_{ij} - fe_{ij}\right)^2}{fe_{ij}}$$
$$= \frac{(11-9{,}07)^2}{9{,}07} + \frac{(5-6{,}93)^2}{6{,}93} + \frac{(6-7{,}93)^2}{7{,}93} + \frac{(8-6{,}07)^2}{6{,}07}$$
$$= 2{,}0386$$

3er. Prendre decisió

```
> pchisq(c(2.0386), df=1, lower.tail=FALSE)
[1] 0.1533507
```

p ($\chi^2 \geq 2{,}0386$ | H_0) = 0,1533507 → No es rebutja la hipòtesi nul·la

4t. Conclusions

No es pot afirmar que existeixi relació estadísticament significativa entre el sexe i el fet de ser fumador, donat que el valor de χ^2 obtingut per la nostra mostra és de 2,0386 que amb un grau de llibertat té un grau de significació associat de 0,1533507, valor lo suficientment gran com per a no rebutjar la hipòtesi nul·la.

Llistat R:

```
> local({
+   .Table <- xtabs(~Sexe+Estatus.Fumador, data=Situació1)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   .Test <- chisq.test(.Table, correct=FALSE)
+   print(.Test)
+   cat("\nExpected counts:\n")
+ })

Frequency table:
    Estatus.Fumador
Sexe No Sí
   D  8  6
   H  5 11

    Pearson's Chi-squared test
data:  .Table
X-squared = 2.0386, df = 1, p-value = 0.1533

Expected counts:
    Estatus.Fumador
Sexe        No        Sí
   D 6.066667 7.933333
   H 6.933333 9.066667
```

2. Determini si existeix relació estadísticament significativa entre el fet de patir dependència a la nicotina i la intenció de deixar de fumar.

1er. Plantejament de la hipòtesi nul·la:

H_0: $fo_{ij} = fe_{ij}$ → No hi ha relació estadísticament significativa entre el fet de patir dependència a la nicotina i la intenció de deixar de fumar.

2n. Càlcul de l'estadístic de contrast: Khi-quadrat de Pearson

		Intenció deixar fumar		Total
		Sí	No	
Dependència nicotina	Sí	7 (3,29)	1 (4,71)	8
	No	0 (3,71)	9 (5,29)	9
Total		7	8	17

$$fe_{11} = \frac{8 \cdot 7}{17} = 3{,}29 \qquad fe_{12} = \frac{8 \cdot 8}{17} = 4{,}71$$

$$fe_{21} = \frac{9 \cdot 7}{17} = 3{,}71 \, fe_{22} = \frac{9 \cdot 8}{17} = 5{,}29$$

No es compleix la condició d'aplicació perquè tres de les quatre freqüències esperades són inferiors a cinc, la freqüència esperada més petita és de 3,29, i com que estem en una taula de contingència d'ordre 2 x 2, per tant s'ha d'obtenir la khi-quadrat amb la correcció de continuïtat de Yates:

$$\chi^2_{Yates} = \sum\sum \frac{\left(\left|fo_{ij} - fe_{ij}\right| - 0{,}5\right)^2}{fe_{ij}}$$
$$= \frac{(|7 - 3{,}29| - 0{,}5)^2}{3{,}29} + \frac{(|1 - 4{,}71| - 0{,}5)^2}{4{,}71} + \frac{(|0 - 3{,}71| - 0{,}5)^2}{3{,}71} + \frac{(|9 - 5{,}29| - 0{,}5)^2}{5{,}29} = 10{,}0187$$

3er. Prendre decisió

```
> pchisq(c(10.0187), df=1, lower.tail=FALSE)
[1] 0.001549588
```

p ($\chi^2 \geq$ 10,0187 | H_0) = 0,0015496 → Es rebutja la hipòtesi nul·la

4t Grandària de l'efecte

$$V = \sqrt{\frac{\chi^2}{n \cdot (q-1)}} = \sqrt{\frac{10{,}0187}{17 \cdot (2-1)}} = 0{,}7677$$

5è. Conclusions

Es pot afirmar que existeixi relació estadísticament significativa entre el fet de patir dependència a la nicotina i la intenció de deixar de fumar, donat que el valor de χ^2 amb la correcció de continuïtat de *Yates* obtingut per la nostra mostra és de 10,0187 que amb un grau de llibertat té un grau de significació associat de 0,001549588, valor lo suficientment petit com per a rebutjar la hipòtesi nul·la. A més a més, el valor de la V de *Cramér*, 0,7677, indica que la intensitat de la relació és elevada. Si observem la taula de contingència podem veure com un 87,9% de les persones que sí tenen dependència a la nicotina sí tenen intenció de deixar de fumar en tant que entre les persones que no tenen dependència a la nicotina aquest percentatge és del 0%.

Llistat R:

```
> local({
+   .Table <- xtabs(~Dep.Nicotina+Intenció.Deixar, data=Situació1)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   .Test <- chisq.test(.Table, correct=TRUE)
+   print(.Test)
+   cat("\nExpected counts:\n")
+   print(.Test$expected)
+ })

Frequency table:
            Intenció.Deixar
Dep.Nicotina No Sí
          No  9  0
          Sí  1  7
```

```
    Pearson's Chi-squared test with Yates' continuity correction
data:  .Table
X-squared = 10.019, df = 1, p-value = 0.00155

Expected counts:
            Intenció.Deixar
Dep.Nicotina       No       Sí
          No 5.294118 3.705882
          Sí 4.705882 3.294118
```

3. Determini si existeix relació estadísticament significativa entre el sexe i el fet de patir dependència a la nicotina.

1er. Plantejament de la hipòtesi nul·la:

H_0: $fo_{ij} = fe_{ij}$ → No hi ha relació estadísticament significativa entre el sexe i el fet de patir dependència a la nicotina.

2n. Càlcul de l'estadístic de contrast: Khi-quadrat de Pearson

		Dependència nicotina		Total
		Sí	No	
Sexe	Home	7 (5,18)	4 (5,82)	11
	Dona	1 (2,82)	5 (3,18)	6
Total		8	9	17

$$fe_{11} = \frac{11 \cdot 8}{17} = 5{,}18 \qquad fe_{12} = \frac{11 \cdot 9}{17} = 5{,}82$$
$$fe_{21} = \frac{6 \cdot 8}{17} = 2{,}82 \qquad fe_{22} = \frac{6 \cdot 9}{17} = 3{,}18$$

No es compleix la condició d'aplicació perquè dos de les quatre freqüències esperades són inferiors a cinc, la freqüència esperada més petita és de 2,82, i com que estem en una taula de contingència d'ordre 2 x 2, s'ha d'obtenir la prova exacta de Fisher, el grau de significació associat a aquesta prova el podem obtenir utilitzant el software R-commander:

Llistat R:

```
> local({
+   .Table <- xtabs(~Sexe+Dep.Nicotina, data=Situació1)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   cat("\nExpected counts:\n")
+   print(.Test$expected)
+   print(fisher.test(.Table))
+ })
```

```
Frequency table:
    Dep.Nicotina
Sexe No Sí
   D  5  1
   H  4  7

Expected counts:
    Dep.Nicotina
Sexe       No       Sí
   D 3.176471 2.823529
   H 5.823529 5.176471

    Fisher's Exact Test for Count Data
data:  .Table
p-value = 0.1312
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
   0.5607447 469.8386363
sample estimates:
odds ratio
  7.632907
```

3er. Prendre decisió

p (Prova exacta de Fisher) = 0,1312 → No es rebutja la hipòtesi nul·la

4t. Conclusions

No es pot afirmar que existeixi relació estadísticament significativa entre el sexe i el fet de partir dependència a la nicotina. Com que no es compleix la condició d'aplicació i partim d'una taula d'ordre 2 x 2, on la freqüència esperada més petita és de 2,82 interpretem el grau de significació associat a la prova exacta de Fisher, que per les nostres dades és de 0,1312, valor lo suficientment gran com per a no rebutjar la hipòtesi nul·la.

4. Determini si existeix relació estadísticament significativa entre el sexe i la intenció de deixar de fumar.

1er. Plantejament de la hipòtesi nul·la:

H_0: $fo_{ij} = fe_{ij}$ → No hi ha relació estadísticament significativa entre el sexe i la intenció de deixar de fumar.

2n. Càlcul de l'estadístic de contrast: Khi-quadrat de Pearson

		Intenció deixar fumar		Total
		Sí	No	
Sexe	Home	6 (4,53)	5 (6,47)	11
	Dona	1 (2,47)	5 (3,53)	6
Total		7	10	17

$$fe_{11} = \frac{11 \cdot 7}{17} = 4{,}53 \qquad fe_{12} = \frac{11 \cdot 10}{17} = 6{,}47$$
$$fe_{21} = \frac{6 \cdot 7}{17} = 2{,}47 \qquad fe_{22} = \frac{6 \cdot 10}{17} = 3{,}53$$

No es compleix la condició d'aplicació perquè tres de les quatre freqüències esperades són inferiors a cinc, la freqüència esperada més petita és de 2,47, i com que estem en una taula de contingència d'ordre 2 x 2, s'ha d'obtenir la prova exacta de Fisher, el grau de significació associat a aquesta prova el podem obtenir utilitzant el software R-commander:

Llistat R:

```
> local({
+   .Table <- xtabs(~Sexe+Intenció.Deixar, data=Situació1)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   .Test <- chisq.test(.Table, correct=FALSE)
+   print(.Test)
+   cat("\nExpected counts:\n")
+   print(.Test$expected)
+   print(fisher.test(.Table))
+ })

Frequency table:
    Intenció.Deixar
Sexe No Sí
   D  5  1
   H  5  6

    Pearson's Chi-squared test
data:  .Table
X-squared = 2.2998, df = 1, p-value = 0.1294

Expected counts:
    Intenció.Deixar
Sexe       No       Sí
   D 3.529412 2.470588
   H 6.470588 4.529412

    Fisher's Exact Test for Count Data
data:  .Table
p-value = 0.3043
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
   0.3970979 329.3206419
sample estimates:
odds ratio
  5.403483
```

3er. Prendre decisió

p (Prova exacta de Fisher) = 0,3043 → No es rebutja la hipòtesi nul·la

4t. Conclusions

No es pot afirmar que existeixi relació estadísticament significativa entre el sexe i la intenció de deixar de fumar. Com que no es compleix la condició d'aplicació i partim d'una taula d'ordre 2 x 2, on la freqüència esperada més petita és de 2,47 interpretem el grau de significació associat a la prova exacta de Fisher, que per les nostres dades és de 0,3043, valor lo suficientment gran com per a no rebutjar la hipòtesi nul·la.

Situació 2

5. Determini si existeix relació estadísticament significativa entre el sexe i la classe social.

1er. Plantejament de la hipòtesi nul·la:

H_0: $fo_{ij} = fe_{ij}$ → No hi ha relació estadísticament significativa entre sexe i classe social.

2n. Càlcul de l'estadístic de contrast: Khi-quadrat de Pearson

		Nivell social				Total
		Baix	Mig	Alt	Molt alt	
Sexe	Home	2 (2,67)	1 (2,67)	3 (2,00)	4 (2,67)	10
	Dona	6 (5,33)	7 (5,33)	3 (4,00)	4 (5,33)	20
Total		8	8	6	8	30

$$fe_{11} = \frac{10 \cdot 8}{30} = 2{,}67 \qquad fe_{12} = \frac{10 \cdot 8}{30} = 2{,}67 \quad fe_{13} = \frac{10 \cdot 6}{30} = 2$$

$$fe_{14} = \frac{10 \cdot 8}{30} = 2{,}67 \qquad fe_{21} = \frac{20 \cdot 8}{30} = 5{,}33 \quad fe_{22} = \frac{20 \cdot 8}{30} = 5{,}33$$

$$fe_{23} = \frac{20 \cdot 6}{30} = 4 \quad fe_{24} = \frac{20 \cdot 8}{30} = 5{,}33$$

No es compleix la condició d'aplicació perquè cinc de les vuit freqüències esperades són inferiors a cinc, la freqüència esperada més petita és de 2, i com que estem en una taula de contingència d'ordre 2 x 4, recodifiquem valors, en aquest cas de la variable classe social, unim les categories baix amb mitjan i alt amb molt alt, així doncs la nova taula es mostra a continuació:

		Nivell social		Total
		Baix-Mig	Alt-Molt alt	
Sexe	Home	3 (5,33)	7 (4,67)	10
	Dona	13 (10,67)	7 (9,33)	20
Total		16	14	30

Com es pot veure a la taula anterior, continua havent una freqüència esperada inferior a 5, la freqüència esperada més petita és de 4,67, i com que ara treballem amb una taula 2 x 2, hem de fer l'estadístic khi-quadrat aplicant la correcció de continuïtat de Yates, i que a continuació es mostra:

$$\chi^2_{Yates} = \sum\sum \frac{\left(\left|fo_{ij} - fe_{ij}\right| - 0{,}5\right)^2}{fe_{ij}}$$
$$= \frac{(|3 - 5{,}33| - 0{,}5)^2}{5{,}33} + \frac{(|7 - 4{,}67| - 0{,}5)^2}{4{,}67} + \frac{(|13 - 10{,}67| - 0{,}5)^2}{10{,}67} + \frac{(|7 - 9{,}33| - 0{,}5)^2}{9{,}33} = 2{,}0257$$

3er. Prendre decisió

```
> pchisq(c(2.0257), df=1, lower.tail=FALSE)
[1] 0.1546576
```

p ($\chi^2 \geq 2{,}0257$ | H_0) = 0,1546576 → No es rebutja la hipòtesi nul·la

4t. Conclusions

No es pot afirmar que existeixi relació estadísticament significativa entre el sexe i la classe social, donat que el valor de χ^2 amb la correcció de continuïtat de *Yates* obtingut per la nostra mostra és de 2,0257 que amb un grau de llibertat té un grau de significació associat de 0,1546576, valor lo suficientment gran com per a no rebutjar la hipòtesi nul·la.

Llistat R:

```
> local({
+   .Table <- xtabs(~SEXE+NSocial_r, data=Situació2)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   .Test <- chisq.test(.Table, correct=TRUE)
+   print(.Test)
+   cat("\nExpected counts:\n")
+   print(.Test$expected)
+ })

Frequency table:
    NSocial_r
SEXE Alt-Molt Alt Baix-Mig
   D             7       13
   H             7        3

    Pearson's Chi-squared test with Yates' continuity correction
data:  .Table
X-squared = 2.0257, df = 1, p-value = 0.1547

Expected counts:
    NSocial_r
SEXE Alt-Molt Alt  Baix-Mig
   D     9.333333 10.666667
   H     4.666667  5.333333
```

6. Determini si existeix relació estadísticament significativa entre sexe i la severitat de l'estat del malalt (8 o més molt sever respecte a la resta de puntuacions).

1er. Plantejament de la hipòtesi nul·la:
H_0: $fo_{ij} = fe_{ij}$ → No hi ha relació estadísticament significativa entre sexe i la severitat de l'estat del malalt recodificada en molt sever (8 o més punts en GDS) respecte a la resta de puntuacions (no molt sever).

2n. Càlcul de l'estadístic de contrast: Khi-quadrat de Pearson

		Severitat de l'estat del malalt		Total
		Molt sever	No molt sever	
Sexe	Home	3 (3,33)	7 (6,67)	10
	Dona	7 (6,67)	13 (13,33)	20
Total		10	20	30

$$fe_{11} = \frac{10\cdot 10}{30} = 3{,}33 \qquad fe_{12} = \frac{10\cdot 20}{30} = 6{,}67$$
$$fe_{21} = \frac{10\cdot 20}{30} = 6{,}67 \qquad fe_{22} = \frac{20\cdot 20}{30} = 13{,}33$$

No es compleix la condició d'aplicació perquè una de les quatre freqüències esperades és inferiors a cinc, la freqüència esperada més petita és de 3,33, i com que estem en una taula de contingència d'ordre 2 x 2, hem d'obtenir el valor de l'estadístic khi-quadrat amb la correcció de continuïtat de Yates, i que a continuació es mostra:

$$\chi^2_{Yates} = \sum\sum \frac{\left(\left|fo_{ij} - fe_{ij}\right| - 0{,}5\right)^2}{fe_{ij}}$$
$$= \frac{(|3 - 3{,}33| - 0{,}5)^2}{3{,}33} + \frac{(|7 - 6{,}67| - 0{,}5)^2}{6{,}67} + \frac{(|7 - 6{,}67| - 0{,}5)^2}{6{,}67} + \frac{(|13 - 13{,}33| - 0{,}5)^2}{13{,}33} \cong 0$$

3er. Prendre decisió

```
> pchisq(c(0), df=1, lower.tail=FALSE)
[1] 1
```

p ($\chi^2 \geq 0$ | H_0) = 1 → No es rebutja la hipòtesi nul·la

4t. Conclusions
No es pot afirmar que existeixi relació estadísticament significativa entre el sexe i i la severitat de l'estat del malalt recodificada en molt sever (8 o més punts en GDS) respecte a la resta de puntuacions (no molt sever), donat que el valor de χ^2 amb la correcció de continuïtat de *Yates* obtingut per la nostra mostra és pràcticament igual

a cero, que amb un grau de llibertat té un grau de significació associat de 1, valor lo suficientment gran com per a no rebutjar la hipòtesi nul·la.

Llistat R:

```
> local({
+   .Table <- xtabs(~SEXE+GDS_r, data=Situació2)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   .Test <- chisq.test(.Table, correct=TRUE)
+   print(.Test)
+   cat("\nExpected counts:\n")
+   print(.Test$expected)
+ })

Frequency table:
    GDS_r
SEXE Molt sever No molt sever
   D          7            13
   H          3             7

    Pearson's Chi-squared test with Yates' continuity correction
data:  .Table
X-squared = 1.1833e-31, df = 1, p-value = 1

Expected counts:
    GDS_r
SEXE Molt sever No molt sever
   D   6.666667     13.333333
   H   3.333333      6.666667
```

7. Determini si existeix relació estadísticament significativa entre la classe social i la severitat de l'estat del malalt (8 o més molt sever respecte a la resta de puntuacions).

1er. Plantejament de la hipòtesi nul·la:

H_0: $fo_{ij} = fe_{ij}$ → No hi ha relació estadísticament significativa entre classe social i la severitat de l'estat del malalt recodificada en molt sever (8 o més punts en GDS) respecte a la resta de puntuacions(no molt sever).

2n. Càlcul de l'estadístic de contrast: Khi-quadrat de Pearson

		Nivell social				Total
		Baix	Mig	Alt	Molt alt	
Severitat de l'estat del malalt	Molt sever	3 (2,67)	3 (2,67)	2 (2,00)	2 (2,67)	10
	No molt sever	5 (5,33)	5 (5,33)	4 (4,00)	6 (5,33)	20
Total		8	8	6	8	30

$$fe_{11} = \frac{10 \cdot 8}{30} = 2{,}67 \qquad fe_{12} = \frac{10 \cdot 8}{30} = 2{,}67 \quad fe_{13} = \frac{10 \cdot 6}{30} = 2$$

$$fe_{14} = \frac{10\cdot 8}{30} = 2{,}67 \qquad fe_{21} = \frac{20\cdot 8}{30} = 5{,}33 \quad fe_{22} = \frac{20\cdot 8}{30} = 5{,}33$$
$$fe_{23} = \frac{20\cdot 6}{30} = 4 \quad fe_{24} = \frac{20\cdot 8}{30} = 5{,}33$$

No es compleix la condició d'aplicació perquè cinc de les vuit freqüències esperades són inferiors a cinc, la freqüència esperada més petita és de 2, i com que estem en una taula de contingència d'ordre 2 x 4, recodifiquem valors, en aquest cas de la variable classe social, unim les categories baix amb mitjan i alt amb molt alt, així doncs la nova taula es mostra a continuació:

		Nivell social		Total
		Baix-Mig	Alt-Molt alt	
Sexe	Home	6 (5,33)	4 (4,67)	10
	Dona	10 (10,67)	10 (9,33)	20
Total		16	14	30

Com es pot veure a la taula anterior, continua havent una freqüència esperada inferior a 5, la freqüència esperada més petita és de 4,67, i com que ara treballem amb una taula 2 x 2, hem de fer l'estadístic khi-quadrat aplicant la correcció de continuïtat de Yates, i que a continuació es mostra:

$$\chi^2_{Yates} = \sum\sum \frac{\left(\left|fo_{ij} - fe_{ij}\right| - 0{,}5\right)^2}{fe_{ij}}$$
$$= \frac{(|6 - 5{,}33| - 0{,}5)^2}{5{,}33} + \frac{(|4 - 4{,}67| - 0{,}5)^2}{4{,}67}$$
$$+ \frac{(|10 - 10{,}67| - 0{,}5)^2}{10{,}67} + \frac{(|10 - 9{,}33| - 0{,}5)^2}{9{,}33} = 0{,}01674$$

3er. Prendre decisió

```
> pchisq(c(0.01674), df=1, lower.tail=FALSE)
[1] 0.8970545
```

p ($\chi^2 \geq 0{,}01674 \mid H_0$) = 0,8970545 → No es rebutja la hipòtesi nul·la

4t. Conclusions

No es pot afirmar que existeixi relació estadísticament significativa entre la classe social i la severitat de l'estat del malalt recodificada en molt sever (8 o més punts en GDS) respecte a la resta de puntuacions(no molt sever), donat que el valor de χ^2 amb la correcció de continuïtat de *Yates* obtingut per la nostra mostra és de 0,01674 que amb un grau de llibertat té un grau de significació associat de 0,8970545, valor lo suficientment gran com per a no rebutjar la hipòtesi nul·la.

Llistat R:

```
> local({
+   .Table <- xtabs(~GDS_r+NSocial_r, data=Situació2)
+   cat("\nFrequency table:\n")
+   print(.Table)
+   .Test <- chisq.test(.Table, correct=TRUE)
+   print(.Test)
+   cat("\nExpected counts:\n")
+   print(.Test$expected)
+ })

Frequency table:
               NSocial_r
GDS_r            Alt-Molt Alt Baix-Mig
  Molt sever                4        6
  No molt sever            10       10

    Pearson's Chi-squared test with Yates' continuity correction
data:  .Table
X-squared = 0.016741, df = 1, p-value = 0.8971

Expected counts:
               NSocial_r
GDS_r            Alt-Molt Alt  Baix-Mig
  Molt sever         4.666667  5.333333
  No molt sever      9.333333 10.666667
```