

Document de Treball 6

Objectiu

L'objectiu d'aquesta sessió és:

1. Taules de contingència: Anàlisi d'una variable categòrica en dues o més poblacions o dues variables categòriques.
2. Exemple.
3. Activitat.

1. Taules de contingència: Una variable categòrica en dues o més poblacions o dues variables categòriques

Material de l'aulavirtual:

Teoria	Activitats amb R	Activitats sense R	Indestap
Secció: Unitat 3. Anàlisi estadística de dues o més variables. Pdf: Tema 8. Comparació de variables categòriques	Document: Exercicis unitat 3. Dades: Dades unitat 3. Exercicis: 7,8,9,14, 15 i 25.	Secció: Problemes Carpeta: Categòriques.	Document: Indestap Secció del document: 2.6 Subsecció:2.6c
Unitat 3. Anàlisi estadística de dues o més variables  Tema 5 Comparació de dues mitjanes poblacionals Pujat 27/01/2020 10:18  Tema 6 Comparació de k mitjanes poblacionals Pujat 27/01/2020 10:18  Tema 8 Comparació de variables categòriques Pujat 27/01/2020 10:19  Exercicis unitat 3 Pujat 27/01/2020 10:19  Dades unitat 3 Pujat 27/01/2020 10:20  Exercici Pujat 13/03/2020 13:20		Problemes  Probabilitat  Una mostra  Dues mostres  k mostres  Categòriques	Indestap  Instal·lació Rcmdr Pujat 3/02/2020 09:40  Instal·lació Rcmdr en MacOs Pujat 3/02/2020 09:40  Resolució Problemes Rcmdr MacOs Pujat 3/02/2020 09:41  Indestap Pujat 3/02/2020 09:42  DietasDieabetes Pujat 3/02/2020 09:43

En aquest últim tema tenim informació de dues variables categòriques o d'una variable categòrica en dues o més poblacions. L'únic estudi que realitzarem en aquest cas és resoldre un contrast d'hipòtesi del tipus:

$$\begin{cases} H_0: \text{homogeneïtat} \\ H_A: \text{No homogeneïtat} \end{cases} \quad \begin{cases} H_0: \text{independència} \\ H_A: \text{No independència} \end{cases}$$

La resolució és igual en ambdós contrastos però caldrà saber, a partir de l'enunciat del problema, diferenciar quin dels dos contrastos hem de definir.

Destresa:

- a) Diferenciar entre un contrast d'homogeneïtat i un contrast d'independència.

Escenari 1: Volem saber si una determinada característica (variable categòrica) té la mateixa distribució en diverses poblacions (Test d'**Homogeneïtat**)

Escenari 2: Volem saber si dues característiques diferents (dues variables categòriques) estan relacionades entre sí o son independents (Test d'**Independència**)

El primer pas és saber diferenciar el tipus de contrast: homogeneïtat o independència?

- i) Si “nosaltres” (l'investigador) controlem una de les dues variables categòriques, el que estem fent es crear poblacions o grups i parlarem d'homogeneïtat de les poblacions o grups.

Exemple 1: Dieta i Salut {AHA, mediterrània}{no sa, sa } (pàgina 4 document de teoria)

Molts especialistes coincideixen a senyalar que la dieta mediterrània, rica en vegetals, fruites i llegums, és més saludable que una dieta hipercalòrica per a disminuir el risc cardiovascular. De fet s'ha demostrat que pot disminuir el risc de patir malalties coronaries. Es realitzà un estudi per a comprovar si la dieta mediterrània era també recomanable front a dietes hipocalòriques, recomanades per l'AHA (American Heart Association).

A un grup de 605 malalts que havien sobreviscut a un atac al cor se'ls assignà aleatòriament a dos subgrups. Als malalts del grup 1 se'ls va donar una dieta hipocalòrica basada en les normes de la AHA , mentre que als malalts del grup 2 se'ls va indicar que havien de seguir una dieta tipus-mediterrània. Als quatre anys es comptabilitzaren els malalts que seguien sans, els que havien patit alguna malaltia i els que havien faltat. S'obtingueren els resultats següents que guardem en una taula de contingència.

En aquest exemple, el que una persona seguisca la dieta AHA o mediterrània el controla l'investigador, per tant, el que estem fent és crear poblacions o grups i parlarem d'homogeneïtat de les poblacions o grups {AHA, mediterrània} i volem veure si es comporta igual respecte la variable categòrica {no sa, sa } (que no controlem, ja que el que una persona estiga sana o no ho controla l'investigador)

Població 1: Malalts que segueixen la dieta AHA

Població 2: Malalts que segueixen la dieta mediterrània

Una única variable categòrica: {Estar sa, No estar sa}

El contrast a triar en aquest cas és

H_0 : Homogeneïtat: estar sans o no estar-ho es distribueix per igual en les poblacions que han seguit una de les dues dietes

H_A : No homogeneïtat: L'estat de salut no es distribueix de la mateixa forma en els malalts que han seguit una o l'altra dieta

- ii) Si nosaltres no controlem ninguna de les dues variables categòriques, parlarem d'independència de les dues variables categòriques.

Exemple 2: Daltonisme i sexe {daltònic, no daltònic}{xiquet, xiqueta } (pàgina 5)

En un estudi sobre daltonisme, un grup d'investigadors examinà a un gran nombre d'estudiants noruecs i s'obtingueren els següents resultats per a les dues variables categòriques {daltonisme, sexe}.

Una única població: escolars noruecs, i dues variables qualitatives:

Variable 1: {xiquet, xiqueta} Variable 2: {daltònic, no daltònic}

En aquest exemple, el sexe i el que siga daltònic o no no ho controla l'investigador. El contrast a triar en aquest cas és:

H_0 : Independència: El daltonisme és independent del sexe

H_A : No independència: El daltonisme està relacionat amb el sexe

b) Resolució d'aquests contrast d'hipòtesi:

La forma de resoldre estos dos contrastos és la mateixa, mitjançant els test de la Khi-quadrat per a taules de contingència (Pearson's Chi-squared test). Si $p\text{-valor} < \alpha$, aleshores es rebutja H_0 .

Resolem el contrast d'homogeneïtat Exemple 1 (anàlogament es faria l'exemple d'independència)

H_0 : estar sans o no estar-ho es distribueix per igual en les poblacions que han seguit una de les dues dietes

H_A : L'estat de salut no es distribueix de la mateixa forma en els malalts que han seguit una o l'altra dieta

```
> .Table # Counts
      1  2
1 64 239
2 29 273

> colPercents(.Table) # Column Percentages
      1      2
1 68.8 46.7
2 31.2 53.3
Total 100.0 100.0
Count 93.0 512.0

> .Test <- chisq.test(.Table, correct=FALSE)

> .Test

      Pearson's Chi-squared test

data:  .Table
X-squared = 15.4282, df = 1, p-value = 8.57e-05
```

Percentatges observats de cada cel·la (per columnes)

p-valor < $\alpha = 0.05$. Rebutjar H_0

Es rebutja H_0 , podem afirmar que: "l'estat de salut no es distribueix de la mateixa manera en els malalts que han seguit aquestes dues dietes diferents".

La proporció de malalts sans seguint la dieta Mediterrània fou del 90%,

$$\Pr(\text{estar sa} \mid \text{dMed}) = \frac{273}{302} = 0.9040$$

mentre que en el grup de la dieta AHA fou del 79%.

$$\Pr(\text{estar sa} \mid \text{AHA}) = \frac{239}{303} = 0.7888$$

Podem concloure que: La dieta mediterrània és millor que la suggerida per la AHA.

c) Condicions d'aplicabilitat (pàgina 19)

Condicions sobre el disseny de l'experiment.

Les dades han de ser d'un d'aquests dos tipus:

Test d'homogeneïtat:

Dues ó mes mostres aleatòries, observades respecte d'una variable categòrica

Test d'independència:

Una mostra aleatòria, observada respecte de dues variables categòriques

Condicions sobre la grandària de la mostra.

Mostra o mostres suficientment grans.

Cada freqüència esperada ha de ser major o igual a 5

Anem a veure com es calculen les freqüències esperades (pàgina 7 i 8)

($E(i, j) \geq 5$, per a tota cel·la (i,j) de la taula de contingència)

$$E(\text{Fila } i, \text{Columna } j) = \frac{(\text{Total fila } i) \times (\text{Total columna } j)}{\text{Total general}}$$

Per tant en l'exemple1 Dieta i Salut:

Freqüències observades (taula de contingència)

	No sans	Sans	TOTAL
AHA	64	239	303
dMed	29	273	302
TOTAL	93	512	605

Freqüències esperades
(Si la hipòtesi nul·la és certa)

	No sans	Sans
AHA	46.58	256.42
dMed	46.42	255.58

Càlcul de les freq. esperades
= (total de la fila x total de la columna) / total

$$E_{11} = \frac{303 \times 93}{605} = 46.58$$

$$E_{21} = \frac{302 \times 93}{605} = 46.42$$

$$E_{12} = \frac{303 \times 512}{605} = 256.42$$

$$E_{22} = \frac{302 \times 512}{605} = 255.58$$

d) Càlcul de proporcions

La proporció de malalts sans seguint la dieta Mediterrània fou del 90%,

$$\Pr(\text{estar sa} | \text{dMed}) = \frac{273}{302} = 0.9040$$

mentre que en el grup de la dieta AHA fou del 79%.

$$\Pr(\text{estar sa} | \text{AHA}) = \frac{239}{303} = 0.7888$$

Podem concloure que: La dieta mediterrània és millor que la suggerida per la AHA.

Destresa	Pàgina del document on es troba la resposta	Exemples en el document
a)	3	4 i 5
b)	6, 9, 10 al 18	11 al 18
c)	6	6, 14 i 16
d)	19, 7,8	7 i 8

Activitat

Recordeu que podeu realitzar totes les activitats esmenades en l'apartat de material de l'aula virtual.

2. Exemple

Exemple: Es denomina epistaxis a les hemorràgies amb origen en les fosses nasals. Normalment es tracta d'un procés que remiteix espontàniament. La primera mesura és baixar-li el cap al malalt i indicar-li que es pressioni fortament el nas amb dos dits durant cinc minuts cronometrats. Si després de cinc minuts de compressió persisteix el sagnat, l'epistaxi ha de ser avaluada per un metge. Quan els sagnats són recurrents poden causar anèmia a causa d'una deficiència de ferro.

En un estudi retrospectiu es van analitzar les dades de 200 malalts adults amb diagnòstic d'epistaxis ingressats en un hospital de referència entre els anys 1997 i 2004 tots dos inclosos (8 anys)

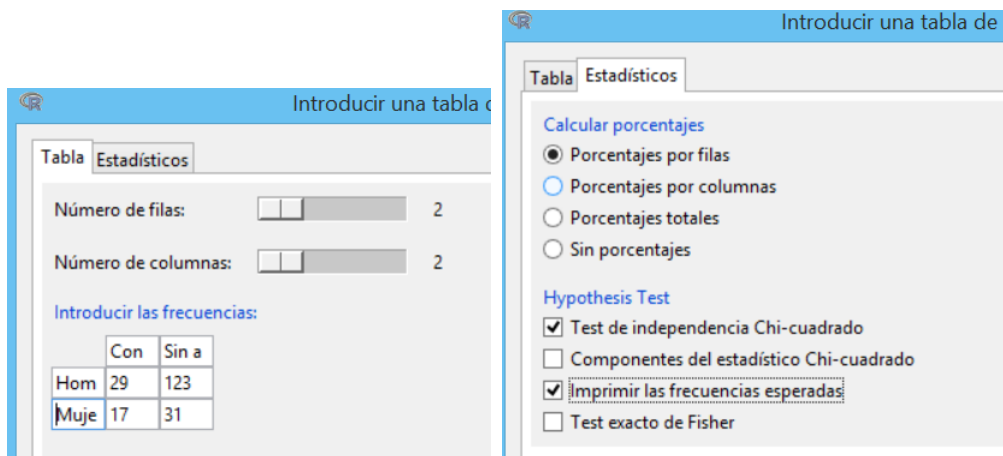
Si dels malalts, 152 eren homes i 48 dones i, dels 46 que van desenvolupar anèmia durant la seua estada hospitalària, 29 eren homes i 17 eren dones, realitzeu el contrast adient per verificar si els homes amb epistaxis tenen una menor incidència d'anèmia que les dones.

Pel que fa al contrast de la incidència de l'anèmia entre homes i dones, serà necessari recórrer a les taules de contingència, utilitzant directament les dades que ens han proporcionat:

	Amb anèmia	Sense anèmia	Total
Homes	29	123	152
Dones	17	31	48
Total	46	154	200

Tenim dades observades de dues variables categòriques. Anem al menú:

Estadístics -> Taules de contingència -> Introduir i analitzar una taula de doble entrada



NOTA: El contrast que realitza en aquest cas el R Commander (taules de contingència) és sempre no direccional. Si volem el p-valor del contrast direccional (només en taules 2 x 2) hem de comprovar que l'adreça de les dades és la mateixa que la de la hipòtesi alternativa i, en aquest cas, dividir per

dos el p-valor que ens proporciona el R; si l'adreça de les dades és la contrària a la hipòtesi alternativa, directament ens quedem amb la hipòtesi nul·la.

El contrast d'hipòtesi que volem fer és el següent:

$$H_0: P(\text{anèmia} \mid \text{home}) = P(\text{anèmia} \mid \text{dona})$$
$$H_A: P(\text{anèmia} \mid \text{home}) < P(\text{anèmia} \mid \text{dona})$$

Per aquesta raó hem demanat els percentatges per files que ens estimaran les proporcions d'interès en la mostra (dividint per 100, clar).

```
> .Table # Counts
      Con anemia Sin anemia
Hombres      29      123
Mujeres      17      31
```

Taula de valors observats

```
> rowPercents(.Table) # Row Percentages
      Con anemia Sin anemia Total Count
Hombres      19.1      80.9     100    152
Mujeres      35.4      64.6     100     48
```

Taula de percentatges per files

```
> .Test <- chisq.test(.Table, correct=FALSE)
```

```
> .Test
```

Pearson's Chi-squared test

Test chi-quadrat para la taula de contingència

```
data: .Table
```

```
X-squared = 5.4982, df = 1, p-value = 0.01904
```

```
> .Test$expected # Expected Counts
```

```
      Con anemia Sin anemia
Hombres      34.96      117.04
Mujeres      11.04       36.96
```

Taula de valors esperats

En la taula de percentatges per fila veiem que la proporció d'incidència d'anèmia observada en homes és 0.191 i en dones 0.354, per la qual cosa estan en l'adreça de la hipòtesi alternativa.

Pel que fa al test de Chi-quadrat ens dona un p-valor de 0.01904, però com aquest resultat correspon a un contrast no-direccional i estem fent un contrast direccional amb unes dades que recolzen la hipòtesi alternativa, hem de dividir per dos el p-valor que ens ha retornat. Per tant, el **p-valor** és **0.00952** i rebutgem la hipòtesi nul·la, per la qual cosa concloem que **hi ha evidència estadística que els homes amb epistaxis tenen una menor incidència d'anèmia que les dones**.

Finalment, com **tots els valors esperats són majors o iguals que 5**, el mètode és vàlid.

3. Activitat d'entrega

Trobareu en l'aula virtual un exercici que s'anomena Activitat3. Crearé una tasca perquè entregueu tot i els pugeu a l'aula virtual. S'ha d'entregar en word, i heu d'assenyalar la població, mostra, variables i els p-valors així com les taules utilitzades.