5. Hipotesi-kontraste ez-parametrikoak

1. ARIKETA

Azken hamarkadan Bidasoa ibaiara isurtzen diren kutsatzaileen kantitatea murrizteko, fabrikek euren instalazioetan iragazkiak instalatu dituzte. Iragazki hauek, ultra-iragazki edo nano-iragazki motakoak izan daitezke. Dena den, nahiz eta filtroak instalatu, zenbait fabrikek isunak ordaindu behar izan dituzte kutsatzaileen isurketagatik. Ondorengo taulan, zoriz hartutako 300 fabrikek instalaturik dituzten iragazkien eta ordaindutako isun moten datuak azaltzen dira.

Iragazkia / Isuna	Nano-iragazkia	Ultra-iragazkia	**Guztira**
Isun arina	34	92	**126**
Isuna ertaina	22	102	**124**
Isun larria	9	41	**50**
Guztira	**65**	**235**	**300**

%2ko adierazgarritasun mailaz, independentziarik al dago isunaren eta instalatutako iragazki motaren artean?

Ebazpena :

Lehenik bi faktoreak edo aldagaiak definituko dira.

X: “Fabrikak jasotako isun mota”.

Y: “Fabrikak instalatutako iragazki mota”.

Ondoren, ariketa honi dagokion hipotesi-kontrastea zehaztuko da:

H_0: X eta Y aldagaiak independenteak dira.

H_a: X eta Y aldagaiak ez dira independenteak.

Behin aldagaiak definiturik eta hipotesi-kontrastea zehazturik, R-n kontingentzia-taula eraikiko da, datuak zutabeka sartuz:

```
>Nanoiragazkia<-c(34,22,9)
>Ultrairagazkia<-c(92,102,41)
>Kontingentzia.taula<-
   data.frame(Nanoiragazkia,Ultrairagazkia)
```

Kontingentzia-taula eratu ondoren, aldagaien independentzia onartu edo baztertu ahal izateko χ^2-ren testa aplika daiteke

```
>chisq.test(Kontingentzia.taula)

[1] Pearson's Chi-squared test

data:  Kontingentzia.taula
X-squared = 3.6206, df = 2,
p-value = 0.1636
```

χ^2 testaren emaitzaren interpretazioa honakoa litzateke:

Kontrasterako estatistikoaren balioa $\chi^2 = 3.6206$ da,

$\mathrm{df} = 2$ askatasun graduak dira $(n-1)\cdot(m-1) = (2-1)\cdot(3-1)$

p-balioa 0.1636 da.

$0.1636 > 0.02$, hau da, p-balioa adierazgarritasun maila baino handiagoa denez hipotesi nulua onartuko litzateke. %2ko adierazgarritasun mailaz, fabrikak instalatutako iragazki motaren eta honek jasotako isunaren arteko independentzia onartuko litzateke.

2. ARIKETA

Ohiko ikuskapen batean, hiru marka komertzialeko hainbat gailuren seinalearen intentsitatea neurtu da. Neurketa horietan, gailuen seinalearen intentsitatea taula honetan ikusten den bezala kalifikatu da:

Marka / Intentsitatea	A marka	B marka	C marka	**Guztira**
Txarra	22	24	20	**66**
Onargarria	50	42	60	**152**
Ona	18	22	16	**56**
Bikaina	6	12	8	**26**
Guztira	**96**	**100**	**104**	**300**

%5eko adierazgarritasun-mailaz, gailuen intentsitatea marka komertzialaren araberakoak diren edo ez konprobatu.

Ebazpena :

Lehenik bi faktoreak edo aldagaiak definituko dira.

X: “Gailuaren seinalearen intentsitatea”.

Y: “Gailuaren marka”.

Ondoren, ariketa honi dagokion hipotesi-kontrastea zehaztuko da:

H_0: X eta Y aldagaiak independenteak dira.

H_a: X eta Y aldagaiak ez dira independenteak.

Behin aldagaiak definiturik eta hipotesi-kontrastea zehazturik, R-n kontingentzia-taula eraikiko da, datuak zutabeka sartuz:

```
>A.marka<-c(22,50,18,6)
>B.marka<-c(24,42,22,12)
>C.marka<-c(20,60,16,8)
>Kontingentzia.taula<-data.frame(A.marka,B.marka,C.marka)
```

Kontingentzia-taula eratu ondoren, aldagaien independentzia onartu edo baztertu ahal izateko χ^2-ren testa aplikatu daiteke

```
>chisq.test(Kontingentzia.taula)

     Pearson's Chi-squared test

data:  Kontingentzia.taula
X-squared = 6.3533, df = 6, p-value = 0.3848
```

χ^2-test-aren emaitzaren interpretazioa honakoa litzateke:

Kontrasterako estatistikoaren balioa $\chi^2 = 6.3533$ da,

$\text{df} = 6$ askatasun graduak dira $(n-1)\cdot(m-1) = (4-1)\cdot(3-1)$

p-balioa 0.3848 da.

$0.3848 > 0.05$, hau da, p-balioa adierazgarritasun maila baino handiagoa denez hipotesi nulua onartuko litzateke. %5eko adierazgarritasun mailaz, gailuaren seinalearen intentsitatearen eta gailuaren markaren arteko independentzia onartuko litzateke.

3. ARIKETA

Bi industria farmazeutikok Payer eta Bizer, kaleratu dituzten txertoek ikterizia izaneko efektu sekundarioan eragina duten egiaztatu nahi da. Horretarako txertoa jaso duten 500 pertsonari segimendua egin zaie ikterizia garatu duten edo ez azterturik. Lortutako datuak ondorengo taulan agertzen dira:

Marka / Ikterizia	Payer	Bizer	**Guztira**
Bai	11	21	**32**
Ez	227	241	**468**
Guztira	**238**	**262**	**500**

%10eko adierazgarritasun-mailaz, ikterizia garatzea txertoaren markarekiko independentea den onartu al daiteke?

Ebazpena :

Lehenik bi faktoreak edo aldagaiak definituko dira.

X: “Ikteriziaren garapena”.

Y: “Txertoa ekoiztu duten industria farmazeutikoaren marka”.

Ondoren, ariketa honi dagokion hipotesi-kontrastea zehaztuko da:

H_0: X eta Y aldagaiak independenteak dira.

H_a: X eta Y aldagaiak ez dira independenteak.

Behin aldagaiak definiturik eta hipotesi-kontrastea zehazturik, R-n kontingentzia-taula eraikiko da, datuak zutabeka sartuz:

```
>Payer<-c(11,227)
>Bizer<-c(21,241)
>Kontingentzia.taula<-data.frame(Payer, Bizer)
```

Kontingentzia-taula eratu ondoren, aldagaien independentzia onartu edo baztertu ahal izateko χ^2-ren testa aplikatu daiteke. Hala ere, kasu honetan kontingentzia-taula 2x2-koa denez Yates-en zuzenketa aplikatu behar den edo ez konprobatu behar da.

Horretarako, itxarotako maiztasunak $\left(e_{ij}\right)$ kalkulatuko dira eta guztiak 5 baino handiagoak badira ez litzateke Yates-en zuzenketa aplikatu behar, beste kasuan ordea bai.

```
>e11<-c(sum(Kontingentzia.taula[1,])*
sum(Kontingentzia.taula[,1])/sum(Kontingentzia.taula))
>e11
[1] 15.232

>e12<-c(sum(Kontingentzia.taula[1,])*
sum(Kontingentzia.taula[,2])/sum(Kontingentzia.taula))
>e12
[1] 16.768

>e21<-c(sum(Kontingentzia.taula[2,])*
sum(Kontingentzia.taula[,1])/sum(Kontingentzia.taula))
>e21
[1] 222.768

>e22<-c(sum(Kontingentzia.taula[2,])*
sum(Kontingentzia.taula[,2])/sum(Kontingentzia.taula))
>e22
[1] 245.232
```

Ikusi daitekeen moduan $e_{ij} > 5 \; \forall (i, j)$, beraz ez da Yates-en zuzenketa aplikatu behar. Hau jakinik, χ^2-ren testa aplikatuko da.

```
>chisq.test(Kontingentzia.taula, correct=F)

     Pearson's Chi-squared test

data:  Kontingentzia.taula
X-squared = 2.3973, df = 1, p-value = 0.1215
```

χ^2-test-aren emaitzaren interpretazioa honakoa litzateke:

Kontrasterako estatistikoaren balioa $\chi^2 = 2.3973$ da,

df $=1$ askatasun graduak dira $(n-1)\cdot(m-1) = (2-1)\cdot(2-1)$

p-balioa 0.1215 da.

$0.1215 > 0.10$, hau da, p-balioa adierazgarritasun maila baino handiagoa denez hipotesi nulua onartuko litzateke. %10eko adierazgarritasun mailaz, ikteriziaren garapena eta txertoa ekoiztu duen marka farmazeutikoaren independentzia onartuko litzateke.

4. ARIKETA

Normaltasun_datuak.txt fitxategian 4 lagin ditugu: *data1*, *data2*, *data3*, *data4*. Lehenengo bien tamaina 100 datukoa da eta azken bien tamaina 30 datukoa da.

%10eko adierazgarritasun-mailaz, *data1* eta *data2* laginen jatorria populazio normal bat den onartu al daiteke?

Ebazpena :

Ondoren, ariketa honi dagokion hipotesi-kontrasteak zehaztuko da:

data1 laginaren kasuan:

H_0: *data1* laginaren jatorria populazio normala da

H_a: data1 laginaren jatorria ez da populazio normala

data2 laginaren kasuan:

H_0: data2 laginaren jatorria populazio normala da

H_a: data2 laginaren jatorria ez da populazio normala

Bi lagin hauen tamaina 100 datukoa da, beraz Kolmogórov-Smirnov (Lilliefors zuzenketarekin) normaltasun proba aplikatuko dugu R erabiliz.

data1 laginarentzat:

```
> lillie.test(data1)

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  data1
D = 0.047933, p-value = 0.8279
```

K-S testaren emaitzaren interpretazioa honakoa litzateke:

Kontrasterako estatistikoaren balioa $D = 0.48$ da eta p-balioa 0.83 da.

$0.83 > 0.10$, hau da, p-balioa adierazgarritasun maila baino handiagoa denez hipotesi nulua onartuko litzateke. %10eko adierazgarritasun mailaz, *data1* laginaren jatorria populazio normala bat da

data2 laginarentzat:

```
> lillie.test(data2)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  data2
D = 0.16502, p-value = 4.46e-07
```

K-S testaren emaitzaren interpretazioa honakoa litzateke:

Kontrasterako estatistikoaren balioa $D = 0.17$ da eta p-balioa 0.000000446 da.

$0.000000446 > 0.10$, hau da, p-balioa adierazgarritasun maila baino txikiagoa denez hipotesi nulua baztertuko litzateke. %10eko adierazgarritasun mailaz, ezin da onartu *data2* laginaren jatorria populazio normala bat denik.

5. ARIKETA

Normaltasun_datuak.txt fitxategian 4 lagin ditugu: *data1*, *data2*, *data3*, *data4*. Lehenengo bien tamaina 100 datukoa da eta azken bien tamaina 30 datukoa da.

%5eko adierazgarritasun-mailaz, *data3* eta *data4* laginen jatorria populazio normal bat den onartu al daiteke?

Ebazpena :

Ondoren, ariketa honi dagokion hipotesi-kontrasteak zehaztuko da:

data3 laginaren kasuan:

H_0: *data3* laginaren jatorria populazio normala da

H_a: data3 laginaren jatorria ez da populazio normala

data4 laginaren kasuan:

H_0: data4 laginaren jatorria populazio normala da

H_a: data4 laginaren jatorria ez da populazio normala

Bi lagin hauen tamaina 30 datukoa da, beraz Shapiro-Wilk normaltasun proba aplikatuko dugu R erabiliz.

data3 laginarentzat:

```
> shapiro.test(data3)

        Shapiro-Wilk normality test

data:  data3
W = 0.97695, p-value = 0.7398
```

S-W testaren emaitzaren interpretazioa honakoa litzateke:

Kontrasterako estatistikoaren balioa $W=0.98$ da eta p-balioa 0.74 da.

$0.74>0.05$, hau da, p-balioa adierazgarritasun maila baino handiagoa denez hipotesi nulua onartuko litzateke. %5eko adierazgarritasun mailaz, *data3* laginaren jatorria populazio normala bat dela onartu daiteke.

data4 laginarentzat:

```
> shapiro.test(data4)

        Shapiro-Wilk normality test

data:  data4
W = 0.9383, p-value = 0.08186
```

S-W testaren emaitzaren interpretazioa honakoa litzateke:

Kontrasterako estatistikoaren balioa $W=0.94$ da eta p-balioa 0.08 da.

$0.08>0.05$, hau da, p-balioa adierazgarritasun maila baino haundiagoa denez hipotesi nulua onartuko litzateke. %5eko adierazgarritasun mailaz, *data4* laginaren jatorria populazio normala bat dela onartu daiteke.