BIOESTATISTIKAREN AZTERKETA

2005eko ekainaren 13a

1.- ACB (Asociación de Clubs de Baloncesto) ligako finalerdietan, Malagako Unicaja eta Tau Baskonia taldeen arteko norgehiagokan, oso partida orekatuak jokatzen ari dira. Hirugarren partidan, esate baterako, 10 minutuko azken laurdenean Tauk zeuzkan abantailak hauexek ziren:

$X \equiv$	*azken laurdeneko beteriko minutuak*	1	2	3	4	5	6	7	$\cdots$	10
$Y \equiv$	*minutuka, Tauren abantailak*	7	5	6	3	4	1	2	$\cdots$	?

(a) Azter ezazu, X eta Y aldagaien arteko korrelazioa (1 puntu)

(b) Erregresio lineala erabiliz, estima ezazu nola bukatu zen partida, hau da, azken laurdeneko 10. minutua bete zeneko Tauren abantaila edo desabantaila eta berdinketa zein minututan lortu zen. (2 puntu)

2.- Demagun $\mathcal{A}$ ontzian 5 bola gorri eta 3 beltz daudela eta $\mathcal{B}$ ontzian 4 gorri eta 5 beltz. Zoriz aukeraturiko ontzi batetik ateratako bi bola erakutsiko balizkizuke eta zein ontzitakoak diren asmatzeko eskatuko balizute, zein izango litzateke, kasu bakoitzean, emango zenukeen erantzuna? (2 puntu)

3.- Azter dezagun, lau zifrako zenbaki esangarri bat zoriz aukeratzean datzan esperimentua; ($1000 \leq x \leq 9999$) betetzen duena, hain zuzen ere.

a) Esperimentu hau burutzean, konproba ezazu, zenbaki kapikua bat lortzeko probabilitatea 0.01 dela. Esperimentua 400 aldiz burutuko bagenu, zein izango litzateke, gutxienez, 5 zenbaki kapikua lortzeko probabilitatea? (1.75 puntu)

b) Esperimentu bera burutzean, konproba ezazu, lorturiko zenbakiaren zifra guztiak desberdinak izateko probabiliatea $\frac{63}{125}$ dela. Esperimentua 2500 aldiz burutuko balitz, zein izango litzateke, gutxienez, horrelako 1245 zenbaki (zifra desberdinetakoak) eta, gehienez, 1255 zenbaki lortzeko probabilitatea? (1.75 puntu)

c) Demagun programaturiko makina batek, geuk eskatuta, lau zifrako zenbaki esangarriak ematen dituela eta 100 zenbaki eskatu ostean, zifra desberdinetako 58 zenbaki lortu direla. Datu honekin, estima ezazu, makina honek lau zifra desberdinetako zenbaki esangarri bat sortzeko daukan probabilitatea. Aurreko atalean egindakoa hontuan hartuta, esango al zenuke, zenbakiak emateko, makinak erabilitako prozedura onargarria dela? (1.5 puntu)

Lagungarria izan daitekeena

Demagun, n tamainuko populazio edo lagin batean X eta Y aldagaiek, hurrenez-hurren, $x_1, x_2, \ldots, x_n$ eta $y_1, y_2, \ldots, y_n$ balio desberdinak hartzen dituztela.

Orduan, X aldagaiaren batezbestekoa:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

X aldagaiaren bariantza:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2$$

X eta Y aldagaien arteko kobariantza:

$$s_{XY} = \frac{\sum_{i=1}^{n} x_i \, y_i}{n} - \bar{x} \, \bar{y}$$

X eta Y aldagaien arteko Pearson-en korrelazio koefizientea:

$$\rho_{XY} = \frac{s_{XY}}{s_X \, s_Y}$$

Emaitzak:

1.- (a) Emandako datuekin, ondoko taula eraiki daiteke:

	x_i	y_i	x_i^2	y_i^2	$x_i\,y_i$
	1	7	1	49	7
	2	5	4	25	10
	3	6	9	36	18
	4	3	16	9	12
	5	4	25	16	20
	6	1	36	1	6
	7	2	49	4	14
Guztira	28	28	140	140	87

Beraz,

$$\bar{x}=4,\;\; \bar{y}=4,\;\; s_x^2=4,\;\; s_y^2=4 \text{ eta } s_{xy}=\frac{\sum x_i\,y_i}{n}-\bar{y}\,\bar{x}=-\frac{25}{7}\simeq -3{,}5714.$$

Horrez gain, Pearson-en koefizienteak,

$$\rho_{XY}=\frac{s_{XY}}{s_X\,s_Y}=-\frac{25}{28}\simeq -0{,}8929$$

esaten digu, X eta Y aldagaien arteko korrelazioa, nahiz eta negatiboa izan, handi samarra dela.

1.- (b) Orduan, X gaineko, Y-ren erregresio zuzena,

$$\tilde{Y}-\bar{y}=\frac{s_{xy}}{s_x^2}(X-\bar{x}) \Longleftrightarrow \tilde{Y}-4=-\frac{25}{28}(X-4) \Longleftrightarrow$$

$$\tilde{Y}=\frac{53}{7}-\frac{25}{28}\,X\simeq 7{,}5714-0{,}8929\,X$$

$X=10$ denean, $\tilde{Y}=-\frac{19}{14}\simeq -1{,}35$ Tauren desabantaila izango da.

Eta, Y gaineko, X-en erregresio zuzena,

$$\tilde{X}-\bar{x}=\frac{s_{xy}}{s_y^2}(Y-\bar{y}) \Longleftrightarrow \tilde{X}-4=-\frac{25}{28}(Y-\frac{103}{3}) \Longleftrightarrow$$

$$\tilde{X}=\frac{53}{7}-\frac{25}{28}\,Y\simeq 7{,}5714-0{,}8929\,Y$$

berdinketa suertatzen denean, hau da, $Y=0$ denean, $\tilde{X}=\frac{53}{7}\simeq 7{,}57$ (7 minutu eta 34 segundu)

2.- Lortzen diren bi bola horietarako hiru aukera dago: biak gorriak izatea (2g), biak beltzak (2b) eta bata gorria eta bestea beltza izatea (gb)
Biak gorriak izan badira, egindako galderari erantzuteko, alderatu beharko genuke $p(A|_{2g})$ eta $p(B|_{2g})$, hau da, bi bola gorri horiek A ontzitakoak izateko probabilitatea eta B ontzitakoak izatekoa. Hala ere, kontuan har daiteke, $p(A|_{2g})=1-p(B|_{2g})$ dela.

Izan ere,

$$p(A|_{2g}) = \frac{p(A)\,p(2g|_A)}{p(2g)} \;\text{ non,}$$

$$p(A) = p(B) = \frac{1}{2} \;\text{ eta }\; p(2g) = p(A)\,p(2g|_A) + p(B)\,p(2g|_B) = \frac{1}{2}\,(\frac{C_5^2}{C_8^2} + \frac{C_4^2}{C_9^2}) = \frac{11}{42}$$

$$\text{Beraz, } p(A|_{2g}) = \frac{\frac{1}{2}\frac{5}{14}}{\frac{11}{42}} = \frac{15}{22} \simeq 0{,}6818 \;\text{ eta }\; p(B|_{2g}) = \frac{7}{22} \simeq 0{,}3182$$

Bi bolak beltzak direnean, ondokoa lortzen da,

$$p(2b) = p(A)\,p(2b|_A) + p(B)\,p(2b|_B) = \frac{1}{2}\,(\frac{C_3^2}{C_8^2} + \frac{C_5^2}{C_9^2}) = \frac{97}{504}$$

$$\text{eta, } p(A|_{2b}) = \frac{p(A)\,p(2b|_A)}{p(2b)} = \frac{\frac{1}{2}\frac{3}{28}}{\frac{97}{504}} = \frac{27}{97} \simeq 0{,}27835 \;\text{ eta }\; p(B|_{2b}) = \frac{70}{97} \simeq 0{,}72165$$

Bukatzeko, bata gorria eta bestea beltza direnean,

$$p(gb) = p(A)\,p(gb|_A) + p(B)\,p(gb|_B) = \frac{1}{2}\,(\frac{15}{C_8^2} + \frac{20}{C_9^2}) = \frac{275}{504}$$

$$\text{eta, } p(A|_{gb}) = \frac{p(A)\,p(gb|_A)}{p(gb)} = \frac{\frac{1}{2}\frac{15}{28}}{\frac{275}{504}} = \frac{27}{55} \simeq 0{,}4909 \;\text{ eta }\; p(B|_{gb}) = \frac{28}{55} \simeq 0{,}5091$$

Beraz, bi bolak gorriak direnean, probableena da A ontzitakoak izatea eta biak beltzak edo bata gorria eta bestea beltza direnean, probableena B ontzitakoak izatea da.

Problema hau, beste era honetara ere, pentsa daiteke; ateratako bolak, esate baterako, beltzak izan direla suposaturik, egindako galderari erantzuteko, konparatu behar ditugu bi probabilitate hauek: $p(A|_{2b})$ eta $p(B|_{2b})$ Baina,

$$p(A|_{2b}) = \frac{p(A)\,p(2b|_A)}{p(2b)} \;\text{ eta }\; p(B|_{2b}) = \frac{p(B)\,p(2b|_B)}{p(2b)}$$

Orduan,

$$p(A|_{2b}) \;\begin{matrix} < \\ = \\ > \end{matrix}\; p(B|_{2b}) \Longleftrightarrow \frac{p(A)\,p(2b|_A)}{p(2b)} \;\begin{matrix} < \\ = \\ > \end{matrix}\; \frac{p(B)\,p(2b|_B)}{p(2b)} \Longleftrightarrow p(2b|_A) \;\begin{matrix} < \\ = \\ > \end{matrix}\; p(2b|_B)$$

Kasu honetan, $p(2b|_A) = \frac{3}{28}$ eta $p(2b|_B) = \frac{5}{18}$, hau da, $p(2b|_B) > p(2b|_A)$ denez, $p(B|_{2b}) > p(A|_{2b})$ izango da eta esango genuke, bi bola beltz lortzen denean, probableena B ontzitakoak izatea dela.

3.-(a)
Kasu posibleak: lehen zifrarako ($\neq 0$) 9 aukera eta besteentzako 10; guztira 9 $\times$ 10 $\times$ 10 $\times$ 10 = 9000 edo ($9\,VR^3_{10}$)
Aldeko kasuak: lehen zifrarako 9 aukera, bigarrenerako 10 eta hirugarren eta laugarrenerako aukera bakarra; guztira 90 aldeko kasu.
Beraz, p(kapikua)= $\dfrac{90}{9000} = 0{,}01$

Esperimentua 400 aldiz burutzen bada eta X aldagaiak lorturiko zenbaki kapikuak zenbatzen baditu, X aldagaiaren banaketa $Bin\,(400; 0{,}01)$ banaketa binomiala izango da edo, hurbilketa gisa, $P\,(4)$, 4 parametroko Poissonen banaketa. Orduan, azken banaketa hau erabiliz,

$$p(X \geq 5) = 1 - p(X \leq 4) = 1 - \sum_{j=1}^{4} \frac{4^i\, e^{-4}}{i!} = 1 - \frac{103}{3}\, e^{-4} \simeq 0{,}3712$$

3.-(b)
Aldeko kasuak: lehen eta bigarren zifretarako 9 aukera (lehenak $\neq 0$ izan behar duelako eta bigarrenak lehenaren desberdina), hirugarrenerako 8 eta laugarrenerako 7; guztira 9 $\times$ 9 $\times$ 8 $\times$ 7 = 4536 aldeko kasu.
Beraz, p (zifra desberdinak)= $\dfrac{4536}{9000} = \dfrac{63}{125} = 0{,}504$

Esperimentua 2500 aldiz burutzen bada eta X aldagaiak lorturiko zifra desberdinetako zenbakiak zenbatzen baditu, X aldagaiaren banaketa $Bin\,(2500; 0{,}504)$ banaketa binomiala izango da edo, hurbilketa gisa, $N(1260; 24{,}9992)$, banaketa normala. Orduan, azken banaketa hau erabiliz,

$$\begin{aligned} p(1245 \leq X \leq 1255) &\simeq p(\frac{1244{,}5 - 1260}{24{,}9992} < Z = \frac{X - 1260}{24{,}9992} < \frac{1255{,}5 - 1260}{24{,}9992}) \\ &\simeq p(-0{,}62 < Z < -0{,}18) = p(Z > 0{,}18) - p(Z > 0{,}62) \simeq 0{,}1610 \end{aligned}$$

3.-(c) Laginetik lortzen ditugu estatistikoak, hauexek: arrakastaren proportzioa, kasu honetan, zifra desberdinetako zenbakien proportzioa, $\bar{p} = 0{,}58$, porrotarena, $\bar{q} = 0{,}42$ eta laginaren tamainua, $n = 100$.

$\bar{p}$ motako estatistikoak neurtzen dituen aldagaiak, $\bar{P}$, badakigu, $N(p; \sqrt{\dfrac{\bar{p}\,\bar{q}}{n}})$, banaketa normalari jarraitzen diola, non p zifra desberdinetako zenbaki bat lortzeko probabilitatea den.

Orduan, p -rako % Ako konfiantzazko tarteak, horrelakoak izango dira:

$$(\bar{p} - k\sqrt{\frac{\bar{p}\,\bar{q}}{n}}\,;\, \bar{p} + k\sqrt{\frac{\bar{p}\,\bar{q}}{n}}) = (0{,}58 - k \times 0{,}04996\,;\, 0{,}58 + k \times 0{,}04996),$$

non k % Ako konfiantzari dagokion konstantea den.

Izan ere,

%90eko konfiantzari dagokion tartea, $(0{,}58-1{,}645\times 0{,}04996\,;\,0{,}58+1{,}645\times 0{,}04996)=(0{,}4978\,;\,0{,}6621)$ izango da;

%95eko konfiantzari dagokiona, $(0{,}58-1{,}96\times 0{,}04996\,;\,0{,}58+1{,}96\times 0{,}04996)=(0{,}4821\,;\,0{,}6779)$ eta

%99ko konfiantzari dagokiona, $(0{,}58-2{,}58\times 0{,}04996\,;\,0{,}58+2{,}58\times 0{,}04996)=(0{,}4511\,;\,0{,}7089)$

Kasualitatez, aurreko atalean kalkulatu dugulako, kasu honetan, badakigu, zifra desberdinetako zenbaki bat lortzeko probabilitatea $p=0{,}504$ dela eta ikus daiteke, balio hau, konfiantzazko tarte guztietan dagoela. Hau dela eta, esan daiteke, nonbait, makinak erabilitako prozedura egokia izan dela.