

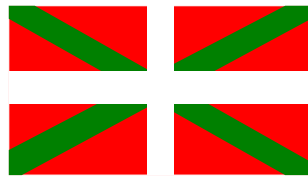
Estatistika enpresara aplikatua

EAZ Gradua: bigarren ikasturtea

Irakaslea: Lorea Mendiola

Ebaluazio jarraitua eta azterketak

Donostiako Ekonomia eta Enpresa Fakultatea
EHU



Egilea: Beñat Zunzunegi

Ebazpen estraofiziala:
egileak eta Gizapediak
ez dute ebazpenaren zuzentasunari buruz
inongo ardura bere gain hartzen.



Gizapedia

gizapedia.hirusta.io

Ebaluazio jarraituko ariketak

1. problema

Bi katalizatzaile aztertzen dira, banaketa normal bati jarraitzen dion prozesu kimiko baten batezbesteko errendimenduari nola eragiten dioten zehazteko. 1.katalizatzailea izan da orain arte erabili dutena. 2.katalizatzailea merkaturatu berria da eta, garestiagoa bada ere, fabrikatzaileek diote prozesuaren batezbesteko errendimendua handitu egiten duela. Hau horrela balitz, 2.katalizatzailea erabiliko litzateke. Proba pilotua egin da eta hauek dira lortutako errendimenduak ehunekotan:

1. katalizatzailea	2. katalizatzailea
91,50	89,19
94,18	90,95
92,18	90,46
95,39	93,21
91,79	97,19
89,07	97,04
94,72	91,07
89,21	92,75

Esan dezakegu %10eko esangura-mailarekin 2.katalizatzaileak prozesuaren batezbesteko errendimendua handitu egiten duela eta erabiltzea komeni dela?

Populazio normalen batezbestekoen berdintasuna kontrastatu behar da, populazioak independenteak izanik eta bariantzak ezezagunak:

- μ_1 : 1. katalizatzailearen populazio-batezbestekoa
- μ_2 : 2. katalizatzailearen populazio-batezbestekoa

Galdetzen da ea 2. katalizatzaileak batezbestekoa handitzen duen ($\mu_1 < \mu_2$), beraz aurkakoa hartzen dugu hipotesi nulu gisa:

$$H_0 : \mu_1 \geq \mu_2 \rightarrow \mu_1 - \mu_2 \geq 0$$

$$H_a : \mu_1 < \mu_2 \rightarrow \mu_1 - \mu_2 < 0$$

Kontrasterako estatistikoa eta dagokion lagin-banaketa hauek izango dira orduan:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2} \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \sim t_{n_1 + n_2 - 2}$$

Joan gaitezen estatistikoa kalkulatzeko:

- $n_1 = n_2 = 8$
- $\bar{x}_1 = (91,50 + \dots + 89,21)/8 = 92,25$
- $s_1^2 = (91,50 - 92,25)^2 + \dots + (89,21 - 92,25)^2 / (8 - 1) = 5,68$
- $\bar{x}_2 = (89,19 + \dots + 92,75)/8 = 92,73$
- $s_2^2 = (89,19 - 92,73)^2 + \dots + (92,75 - 92,73)^2 / (8 - 1) = 8,90$

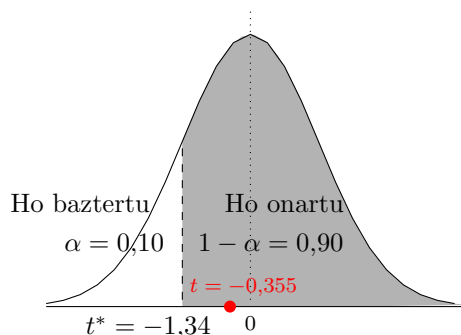
t estatistikoaren balioa hau da orduan:

$$t = \frac{(92,25 - 92,73) - 0}{\sqrt{\frac{1}{8} + \frac{1}{8} \sqrt{\frac{(8 - 1) \times 5,68 + (8 - 1) \times 8,90}{8 + 8 - 2}}} = -0,355$$

Hipotesi nulua baztertzen dugu t estatistikoa oso txikia denean, edo bestela esanda, eremu kritikoa ezker aldean dago. Eman dezagun balio kritikoa, 0.1eko esangura-mailarekin, $t_{8+8-2} = t_{14}$ banaketa batean:

$$t_{0,1,14}^* = -1,34$$

t estatistikoak (-0.355) ez du gainditzen ezkerretik balio kritikoa (-1.34). Beraz, hipotesi nulua onartu egiten dugu: ez dago ebidentzia aski baieztatzeko 2. katalizatzailea hobea dela.



1. problema

Valentziako Unibertsitateak prebentzio-jarduerak diseinatu nahi ditu lantegietan COVID-19ak sortutako pandemia egoerari aurre egiteko. Horretarako, langileen zaugarritasun mailarekin (vulnerabilidad) erlazionatuta dauden faktoreak identifikatu nahi ditu eta 419 langileko lagina erabiliz, ondorengo maiztasun-taula lortu ditu (datu errealak):

	Zaugarritasun-maila:		
	Txikia	Ertaina	Altua
Adina (↓)			
< 60	151	109	7
60 -65	78	68	4
> 65	1	1	0
Generoa (↓)			
Gizonezkoa	123	92	7
Emakume ez haurduna	95	70	4
Emakume haurduna	12	16	0
Erretzailea (↓)			
Bai	211	135	6
Ez	19	43	5
Gizentasuna (↓)			
GMI < 30	210	120	4
GMI 30-40	20	53	5
GMI > 40	0	5	2

(GMI: Gorputz masaren indizea)

Taularen arabera, zeintzuk dira COVID-19aren aurrean langileek duten zaugarritasun mailarekin lotutako faktoreak? %5eko esangura-maila erabili ezazu.

Independentzia-proba bana garatu behar da emandako lau faktoreetarako. Horretarako, E_i maiztasun teorikoak (O_i enpirikoak edo behatuak dira) kalkulatzeko aurrean, bazter maiztasunak bidertuz kontingentzia-taula bakoitzean eta 419ko lagin tamainarekin zatituz (maiztasun batzuk 5 baino txikiagoak izan arren, Yatesen zuzenketa ez dugu garatuko, ikasi egin ez dugulako):

	Zaugarritasun-maila:		
	Txikia	Ertaina	Altua
Adina (↓)			
< 60	146.5	113.4	7.01
60 -65	82.3	63.7	3.94
> 65	1.1	0.85	0.05
Generoa (↓)			
Gizonezkoa	121.8	94.3	5.8
Emakume ez haurduna	92.7	71.8	4.4
Emakume haurduna	15.4	11.9	0.7
Erretzailea (↓)			
Bai	193.2	149.5	9.2
Ez	36.7	28.4	1.75
Gizentasuna (↓)			
GMI < 30	183.3	141.9	8.7
GMI 30-40	42.8	33.1	2.05
GMI > 40	3.8	2.97	0.18

Eta haietatik, khi-karratu estatistikoak kalkulatzeko ditugu lau faktoreetarako, gelaska guztietarako $\frac{(O_i - E_i)^2}{E_i^2}$ balioak gehituz:

$$\mathbf{X}_{adina}^2 = \frac{(151 - 146,5)^2}{146,5^2} + \dots + \frac{(0 - 0,05)^2}{0,05^2} = 0,9$$

$$\mathbf{X}_{generoa}^2 = \frac{(123 - 121,8)^2}{121,8^2} + \dots + \frac{(0 - 0,7)^2}{0,7^2} = 3,33$$

$$\mathbf{X}_{erretzailea}^2 = \frac{(211 - 193,2)^2}{193,2^2} + \dots + \frac{(5 - 1,75)^2}{1,75^2} = 26,17$$

$$\mathbf{X}_{gizentasuna}^2 = \frac{(210 - 183,3)^2}{183,3^2} + \dots + \frac{(2 - 0,184)^2}{0,184^2} = 61,34$$

Eta orain balio kritikoak kalkulatzeko ditugu faktore bakoitzeko:

- Adinerako, generorako eta gizentasunerako (errenkada eta zutabe kopuruak berdinak dira hirurotan):

$$\chi_{0,05;(3-1)(3-1)}^2 = 9,49$$

- Erretzeko joerari buruz:

$$\chi_{0,05;(3-1)(2-1)}^2 = 5,99$$

Khi-karratu estatistikoak balio kritikoa gainditzen du erretzeko joeran eta gizentasunean. Beraz, lau faktoreetatik gaixotasunarekiko zaugarritasunean eragiten dutenak bi horiek dira: erretzeko joera eta gizentasuna.

Maiatzaren 24ko azterketa

1. problema

Enpresa publikoetako langileen arteko genero-berdintasuna aztertzeko, 120 langile hautatu dira ausaz eta 24 emakumeak dira.

- (0.5 puntu) Zein da aztertu nahi diren populazioa eta ausazko aldagaia?
- (1 puntu) %90eko konfiantza-mailarekin, zein da enpresa publikoetako langileen arteko emakumeen proportzioa?
- (0.5 puntu) Zein da laginaren proportzioaren balioa eta zurenb estimazioaren errore-marjina (zabalera erdia)?
- (0.5 puntu) Errore-marjina 0.03ra murriztu nahi bada, zenbatean handitu behar da lagina?

a.

Populazioa enpresa publikoetako langile guztiek osatzen dute, eta ausazko aldagaia laginean, 120 langileen artean alegia, suertatzen diren emakumeen kopurua.

b.

Proportzioari buruzko tartearen formula hau da:

$$I_{1-\alpha}(p) : f \pm z_{\alpha/2} \sqrt{\frac{f(1-f)}{n}}$$

Gure kasuan, $1 - \alpha$ konfiantza-maila %90 da, eta beraz $z_{0,05}$ balioa bilatu behar dugu, gaineratik %5eko probabilitatea uzten duen banaketa normal estandarren balioa: 1.6448.

f estatistikoa $24/120=0.2$ da (emakumeak %20 dira laginean).

Beraz, tartea honela geratzen da:

$$I_{0,90}(p) : 0,2 \pm 1,6448 \sqrt{\frac{0,2 \times 0,8}{120}} : 0,2 \pm 0,06 : 0,14 - 0,26$$

Beraz, emakumeen proportzioa enpresa publikoetan %14- %26 bitartekoa da.

c.

Laginaren proportzioaren balioa $0.2 = \%20$ da, eta errore marjina $0.06 = \%0.06$.

d.

Lagin tamainaren formula hau da:

$$\epsilon = z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \rightarrow n = \frac{z_{\alpha/2}^2 p(1-p)}{\epsilon^2}$$

Kasu okerreanean (lagin handiena) jartzen gara $p = 0,5$ dugunean:

$$n = \frac{1,6448^2 \times 0,5 \times 0,5}{0,03^2} = 751,5 \rightarrow 752$$

752 langile hautatu behar dira, errore marjina 0-03 izateko %90eko konfiantzarekin.

2. problema

KNOLIVE markako olio-botilen etiketek produktu horren 250ml dutela adierazten dute. Halere, Gipuzkoako Kontsumitzaileen Batasunak hainbat keza jaso ditu edukia txikiagoa dela esanez. Enpresaren aurkako nuerriak hartu aurretik, egoera ikertu nahi dute estafa bat ote dagoen jakiteko. Botilen edukia 10ml-ko desbideratze tipikoko banaketa normala jarraitzen duela dakite. 100 taminako lagin bat hartu dute botilek diotena egia den egiaztatzeko eta 248ml-ko lagin-batezbestekoa lortu dute.

- (1 puntu) Zer erabaki behar du Gipuzkoako Kontsumitzaileen Batasunak %95eko konfiantza-mailarekin? Estafa bat izan daiteke?
- (1.5 puntu) Gipuzkoako Kontsumitzaileen Batasuna gehien kezkatzen duena erabaki okerra hartzea da. Hau da, KNOLIVE enpresaren aurkako ekintzak martxan jartzea benetan estafarik egon gabe. Errore mota hori egiteko probabilitatea murrizteko hipotesi-kontraste berri bat egin nahi du, baina ez du argi %90eko edo %99ko konfiantza-maila aukeratu behar duen. Zein aukeratu behar du? Zergatik? Egin hipotesi-kontraste berria konfiantza-maila egokienarekin. Erabakia aldatuko du?

a.

Hipotesi nuluzat uste edo susmatzen denaren aurkako hartzen da, estafarik ez dagoela alegia, eta beraz batez besteko edukia 250 edo handiagoa dela:

$$H_0 : \mu \geq 250$$

$$H_a : \mu < 250$$

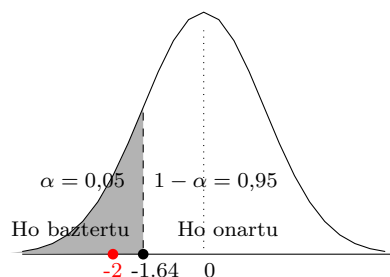
Batezbestekoari buruzko kontrastea da beraz, populazio normala izanda eta bariantza (eta desbideratze) ezagunarekin ($\sigma = 10$). Estatistikoa lagin-batezbestekoa da, kasu honetan honela banatzen dena, hipotesi nulua egiaztat harturik:

$$\bar{x} \sim N\left(\mu = 250, \sigma_{\bar{x}} = \frac{10}{\sqrt{100}} = 1\right) \rightarrow \frac{\bar{x} - 250}{1} \sim N(0, 1)$$

Gure kasuan estatistikoaren balioa $(248-250)/1 = -2$ da.

Eman dezagun orain balio kritikoa. Horretarako, estatistikoa $Z \sim N(0, 1)$ banaketa normal estandarren arabera banatzen dela baliatu behar dugu. *Hipotesi nulua, batezbestekoa 250 edo handiagoa dela alegia, baztertuko dugu, estatistikoa oso txikia denean.* Konfiantza %95 da eta gainera utziko dugu beraz, eta %5eko esangura-maila azpitik. Hartara, banaketa normal estandarrean azpitik %5eko probabilitatea uzten duen balioa bilatu behar dugu: -1.64. Horixe da balio kritikoa.

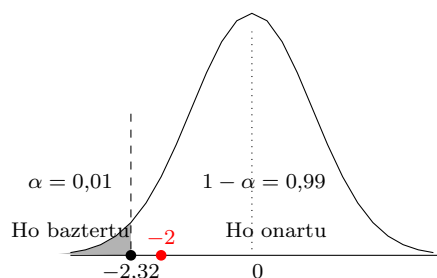
Kontrastearen eskema hau da:



Estatistikoaren balioa baztertze-aldean dago beraz hipotesi nulua baztertu, eta olio-enpresak iruzurra egiten ari dela baieztatzeko ebidentzia edo arrazoi aski dagoela erabaki behar da.

b. Konfiantza-mailak %90 eta %99 izanik, esangura-mailak, α alegia, %10 eta %1 dira, hurrenik hurren. Esangura-maila bat dator I motako errorea deitzen denarekin, hipotesi nulua (enpresak ez du iruzurrik egiten) egia izanda, enpresak iruzurra egiten duela erabakitze arriskua da. Arrisku hori txikiena izatea nahi dugu, eta beraz %1ekoa aukeratu dugu, %99ko konfiantza-maila alegia.

Gara dezagun proba berria. Balio kritiko berria, aurrekoan bezala, taula normalean begiratu, -2.32 da. Beraz, hau da probaren eskema:



Oraingo honetan, erabakia da enpresak ez duela iruzurrik egiten, ez dago horretarako arrazoi edo ebidentzia askirik. Aldaketaren arrazoia hau da: hipotesi nuluari konfiantza handiagoa eman diogu oraino honetan.