

ESTADÍSTIKA ENPRESARA APLIKATUA

Azterketa ebatziak

2023-2024 ikasturtea

Donostiako Ekonomia eta Enpresa Fakultatea. EHU

Egilea eta irakasgaiaren irakaslea: Josemari Sarasola



Gizapedia

gizapedia.org

ESTADÍSTIKA ENPRESARA APLIKATUA

Irakaslea: Josemari Sarasola

Data: 2024ko maiatzaren 15a, 10:00

Iraupena: 90 min

I. ebazkizuna (2.25 puntu)

Soziologo batek ikerketa batean planteatu du *gay* edo gizon homosexualek lanbide fisiko neketsuak saihestu eta, aldi berean, jendearekin harremanak izatea dakarten lanbide erosoak bilatzen dituztela. Ikerketa garatzeko, EUSTAK estatistika-erakunde ofizialaren datua hartu du abiapuntutzat, zeinaren arabera 18-65 urte bitarteko eta beraz lan egiteko adinean dauden gizon homosexualak esparru bereko gizon guztien populaziotik %5 direla Gipuzkoan.

Soziologoak bi inkesta garatu ditu bere ikerketan: lehenengoan, KARA moda-enpresako Gipuzkoako dendetan bezeroekiko harremanetarako enplegatutik dauden 60 gizonetatik 9 *gay* direla suertatu da; bigarrenean, GOGORKI eraikuntza-enpresan behargin edo obrero gisa enplegatutik dauden 20 gizonetatik inor ez da *gay* suertatu.

Egin beharreko atazak:

Aurreko inkestetako datuetatik:

- KARA enpresako inkesta harturik, **De Moivre-Laplace teorema baliatuz eta jarraitutasun-zuzenketa aplikatuz**, erabaki ezazu %1eko esangura-mailarekin, gizon homosexualek publikoarekin harremanak izatea eskatzen duten lanbideetan jarduteko joera handiagoa duten.
- GOGORKI enpresako inkesta harturik, **Larson nomograma baliatuz eta baita probabilitatea behar den formularekin zehatz kalkulatz ere**, erabaki ezazu %1eko esangura-mailarekin, gizon homosexualek lan neketsuak hartzeko joera txikiagoa duten.
- Azkenik, enpresa horietan egindako inkestak aparte utzita, zenbatekoa populazio orokorretik gizonak zoriz aukeratzen hasita, lehen homosexuala aurkitu arte 9 gizon heterosexual suertatzeko probabilitatea?

Oharra: (a) eta (b) ataletan, **hipotesi nulua zehazki eta modu arrazoituan planteatu, eta baliatzen den X edo dena delakoaren aldagaiaren banaketa modu egokian adierazi, parametroak zehaztuz. Azken erabakia ez adierazi onartu/baztertu dikotomia gisa, interpretazio osoa eginez baizik.**

a

Galderan planteatzen den eta KARA enpresako datuek erakusten dutenaren aurkakoa hartzen dugu hipotesi nulutzat (KARA enpresan gizonen $9/60 = \%15$ dira *gay*, populazio orokorrean %5 direlarik; beraz, badirudi KARA enpresara *gay* diren gizonak gehiago sartzen direla, %5eko proportzioan baino maizago alegia):

$$H_0 : p \leq 0.05$$

p izanik, enpresa horretan, eta hedaduraz, mota horretako enpresetan gizon langilea *gay* izateko probabilitatea.

Hipotesi nulu horren pean, enpresa horretan lan egiten duten *gay* gizonen kopurua banaketa binomialaren arabera da, eta n handia denez, De Moivre-Laplace teorema erabiltzen dugu:

$$X \sim B(n = 60, p = 0.05) \longrightarrow N(\mu = np = 3, \sigma = \sqrt{npq} = 1.688)$$

Harritzekoa goitik dago: %10 asko da %5 aldean. Beraz, probaren norabidea goranzkoa da. Eman dezagun horrela p -balioa edo ebidentziaren probabilitatea, jarraitutasun-zuzenketa lehen pausoan aplikatuz:

$$P[X_{disk} \geq 9] = P[X_{jar} > 8.5] = P\left[Z > \frac{8.5 - 3}{1.688}\right] = P[Z > 3.25] = 1 - P[Z < 3.25] = 1 - 0.9997674 = 0.0002 < \alpha$$

Beraz, hipotesi nulua baztertu eta *gay* gizonen harreman publikoetako lanak nahiago izateko joera handiagoa dutela ondorioztatu behar da, horretarako ebidentzia aski sendoa delako.

b

Galderan planteatzen den eta GOGORKI enpresako datuek erakusten dutenaren aurkakoa hartzen dugu hipotesi nulutzat (GOGORKI enpresan ez dago gay den gizonik, populazio orokorrean %5 direlarik; beraz, badirudi GOGORKI enpresara gay diren gizonak gutxiago sartzen direla, %5eko proportzioan baino gutxiagotan alegia):

$$H_0 : p \geq 0.05$$

p izanik, enpresa horretan, eta hedaduraz, mota horretako enpresetan gizon langilea gay izateko probabilitatea.

Hipotesi nulu horren pean, enpresa horretan lan egiten duten gay gizonen kopurua banaketa binomialaren arabera da, eta n handia denez, De Moivre-Laplace teorema erabiltzen dugu:

$$X \sim B(n = 20, p = 0.05)$$

Harritzekoa behetik dago: %0 (gayrik ez) gutxi da %5 aldean. Beraz, probaren norabidea beheranzkoa da. Eman dezagun horrela p -balioa edo ebidentziaren probabilitatea:

$$P[X \leq 0] = P[X = 0] = 0.05^0 \times 0.920 \times \frac{20!}{0!20!} = 0.35 > \alpha$$

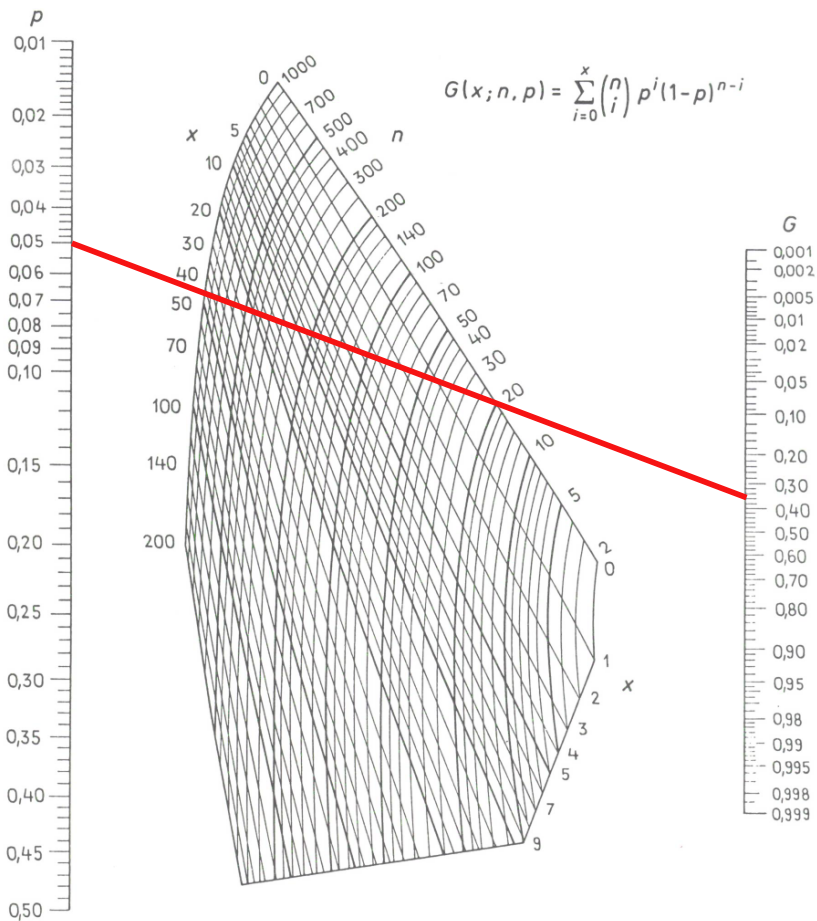
Beraz, hipotesi nulua onartu eta beraz ebidentzia (0 gay 20 gizonetik) ez da aski sendoa baieztatzeko gay gizonen lan neketsuak saihesten dituztela.

c

Banaketa geometrikoaren formula erabili behar da.

$$P[\text{lehen gay arte 9 hetero}] = 0.95^9 \times 0.05$$

LARSON NOMOGRAMA - (B) ATALA -



II. ebazkizuna (0.75 puntu)

Sexologian frogaturik dago emakumeak multiorgasmikoagoak direla gizonak baino, hau da, emakumeak sexu harreman bakarrean orgasmo gehiago eta maizago izateko joera izaten duela. Emakume gazte heterosexual eta bikotekide bakarrekoen artean egindako ikerketa baten arabera, sexu saio batean zehar emakumeak sentitzen dituen orgasmoen kopuruak Poisson prozesu bati jarraitzen dio, aldi errefraktarioa, hau da, orgasmo batetik besterako batez besteko denbora (eta memoria-ezaren efektuagatik, baita orgasmoa izan arteko batez besteko denbora ere) 12 minutukoa izanik.

Egin beharreko ataza:

30 minutuko sexu harreman baterako kalkulatu talde horretako emakume batek gutxienez 2 orgasmo sentitzeko probabilitatea, hots, sexu-harreman multiorgasmikoa izateko probabilitatea.

Oharra: Ez erabili Thorndike nomograma kalkulua egiteko. Soluzioa eman ezazu behar den formula garatuz eta adierazita utzi, azken emaitza kalkulatu gabe.

Ematen den datua batez besteko denborari buruzkoa da. Beraz, banaketa esponentzialaren batezbestekoa adierazten du:

$$\frac{1}{\lambda} = 12min \longrightarrow \lambda_{1min} = 1/12 \text{ orgasmo} \longrightarrow \lambda_{1h} = 60/12 = 2.5 \text{ orgasmo}$$

30 minutuko sexu harreman bati buruz galdetzen digutenez, epe horretara moldatzen dugu lambda parametroa:

$$\lambda_{1h} = 60/12 = 5 \longrightarrow \lambda_{30min} = 2.5 \text{ orgasmo}$$

Eskatutako probabilitatea eman dezagun (X: 30 minutuko orgasmo kopurua):

$$P[X \geq 2] = 1 - P[X \leq 1] = 1 - P[X = 0] - P[X = 1] = 1 - \frac{e^{-2.5} \times 2.5^0}{0!} - \frac{e^{-2.5} \times 2.5^1}{1!}$$

III. ebazkizuna (2 puntu)

Enpresa batean egindako kontu-ikuskaritza batean fakturen proportzio bat B kontabilitatean edo *beltzean* kobratu dela jakin da. Proportzio hori faktura kopuru totalarekin bidertuz, faktura irregular horiek 66 direla zenbatetsi da. Faktura horiek guztiak aurkitzeko oztopoak direla eta, haietako bakoitzaren zenbatekoa zorizkotzat hartu **100€-500€ tartean uniformeki** banatzen dela ezarri da.

Egin beharreko atazak:

- (a) 66 faktura horiek guztiek %95eko probabilitatearekin osatzen duten gutxienezko zenbatekoa eman ezazu, LTZ baliatuz.
- (b) Azkenean 66 fakturako kopuru hori ez da ziurtzat jotzen eta gehiago izan litezkeela uste da. (b1) 30.000€-ko zenbateko totalera gutxienez heltzeko, zenbat faktura beharko lirateke batez beste? (b2) Berriz ere LTZ baliatuz, zenbat faktura *beltzean* izango lirateke guztira faktura horien guztien bitartez ezkutatutako diru-zenbatekoa **30.000€ baino handiagoa izateko probabilitatea 0.99 dela** baieztatu ahal izateko?

a

66 fakturen zenbatekoa honela adierazi eta kalkulatzeko dugu, batugai bakoitza faktura bakoitzaren zenbatekoa izanik:

$$\mathbf{X} = X_1 + X_2 + \dots + X_{66}$$

Faktura kopurua aski handia denez, hots, 30 edo handiagoa, eta haiek elkarrekiko independenteak direnez, LTZ aplikatu daiteke. X_i batugai bakoitza uniformeki banajitzen da 100-500 tartean. LTZ aplikatzeko, bakoitzaren itxaropena eta bariantza behar ditugu:

$$\mu = \frac{100 + 500}{2} = 300 ; \sigma^2 = \frac{(500 - 100)^2}{12} = 13333$$

Horrela, LTZ-aren arabera $\mathbf{X}$ fakturen zenbateko totala honela banatzen da:

$$\mathbf{X} \sim N(\mu = 300 \times 66 = 19800; \sigma = \sqrt{13333 \times 66} = 938)$$

Eskatzen dena %95eko probabilitatea duen kopuru minimoa da:

$$P[\mathbf{X} > x] = 0.95$$

Estandartu dezagun aurrekoa hasieratik:

$$P[\mathbf{X} > x] = P\left[Z > \frac{x - 19800}{938}\right] = 0.95$$

Banaketa normalaren pertzentilen taulan bilatuta:

$$\frac{x - 19800}{938} = -1.64 \longrightarrow x = 18261.68$$

Azken emaitza hau da erantzuna: 66 fakturen gutxienezko zenbateko totala 18261.68€ da.

b1

Faktura bakoitzeko batez besteko zenbatekoa 300 denez, 30.000-ra heltzeko, $30.000/300=100$ faktura beharko lirateke gutxienez batez beste.

b1

Orain, faktura kopurua (n) da ezezaguna:

$$\mathbf{X} = X_1 + X_2 + \cdots + X_n$$

LTZ aplikatuz:

$$\mathbf{X} \sim N(\mu = 300n; \sigma = \sqrt{13333 \times n} = 115.46\sqrt{n})$$

Ondoko hau betetzea nahi dugu:

$$P[\mathbf{X} > 30.000] = 0.99$$

Estandartuz:

$$P[\mathbf{X} > 30.000] = P\left[Z > \frac{30.000 - 300n}{115.46\sqrt{n}}\right] = 0.99$$

Banaketa normalaren pertzentilen taulan bilatuta:

$$\frac{30.000 - 300n}{115.46\sqrt{n}} = -2.32 \longrightarrow 300n - 267.9\sqrt{n} - 30.000 = 0$$

$n = x^2$ aldagai aldaketa eginez eta soluzio negatiboa baztertuz (sigma negatiboa ematen duelako):

$$300x^2 - 267.9x - 30.000 = 0 \longrightarrow x = 10.45$$

Aldagai-aldaketa deseginez eta, 30.000 gainditu nahi dugunez, faktura kopurua gehiegiz borobilduz:

$$n = 10.45^2 = 109.2 \rightarrow 110 \text{ faktura}$$

Beraz, 110 faktura behar dira gutxienez 30.000€-ko zenbateko totala gainditzeko 0.99ko probabilitateaz.

ESTADISTIKA ENPRESARA APLIKATUA

Irakaslea: Josemari Sarasola

Data: 2024ko uztailaren 5a, 16:00

Iraupena: 60 min (orientagarria)

I. ebazkizuna (2 puntu 10etik)

Josemari Zarazola irakaslea hil zen duela denbora bat. Bere bizitzan zehar erakutsitako dohai jainkotiarrek direla eta, Eliza Katolikoa santu izendatzeko izapideak garatzen hasi da. Jakina denez, pertsona bat santu izendatzeko (kasu honetan San Joxemari Estatistiko izenarekin santutegian sartzeko) mirariak egin dituela frogatu behar da. Kasu honetan, Josemari Zarazolak ustez egindako gertaera miragarriak hau da: bere klaseetara bertaratu eta berarekin estatistika ikasi duten neska asko (erlatiboki) lekaime edo moja sartu dira, hain zuzen ere populazio orokorrean lekaime sartzen diren emakumeak oso gutxi diren garaian. Zehatzago, populazio orokorrean batez beste 100.000 emakumeetatik 2 lekaime sartzen dela zenbatetsi da, erbata zoriz. Josemari Zarazolaren klaseetara bertaratu diren gutxi gorabehera 5.000 emakumeetatik, berriz, 5 dira lekaime sartu direnak.

Egin beharreko atazak:

- Adierazi zein den ebidentzia.
- Adierazi eta justifikatu hipotesi nulua, horretarako λ parametroari buruzko tarte egokia eman.
- p balioa zein probabilitateari dagokion adierazi, eta probabilitate hori nola kalkulatu den adierazi, azken kalkulua egin gabe, banaketa binomialaren Poisson-en hurbilketa baliatuz (edo zuzenean Poisson-en banaketa baliatuz).
- Ariketa honetan ematen den eranskinetik aukeratu zein den p balio zehatz hori.
- Azkenik, p balio hori harturik, erabaki ezazu Josemari Zarazolaren klaseetan emakume ikasleekin gertatutakoa benetan miraria dela baieztatzeko arrazoi nahikoa duen Eliza Katolikoak, esangura-maila %1 izanik.

Eranskina:

```
> ppois(5,0.1,lower.tail=F)      [1] 0.124652      > 1-ppois(4,0.1)
[1] 1.274899e-09                  > ppois(2,4)      [1] 7.667802e-08
> ppois(5,2,lower.tail=F)        [1] 0.2381033     > 1-ppois(5,2)
[1] 0.01656361                   > ppois(2,6)      [1] 0.01656361
> ppois(4,2,lower.tail=F)        [1] 0.0619688     > 1-ppois(4,2)
[1] 0.05265302                   > 1-ppois(5,0.1)  [1] 0.05265302
> ppois(6,2,lower.tail=F)        [1] 1.274899e-09   > 1-ppois(6,2)
[1] 0.004533806                  > 1-ppois(6,0.1)  [1] 0.004533806
> ppois(2,5)                     [1] 1.818012e-11
```

(a)

Ebidentzia da 5.000 emakume ikasleetatik 5 lekaime sartu direla.

(b)

Orokorrean, 100.000 emakumeetatik 2 sartzen dira lekaime; Josemari Zarazolaren klaseetan, berriz, 5.000 emakumeetatik 5. Bere klaseetan proportzio orokorra beteko balitz, $\frac{5000}{100000} \times 2 = 0.1$ emakume ikasle batezbeste sartu beharko lirateke lekaime. Beraz, argi dago badirudiela bere klaseetara joan ezker, lekaime sartzeko aukera handiagoak daudela. Beraz, datuek erakusten dutenaren aurkakoa harturik, hau izango da hipotesi nulua (X : 5.000 emakumeetatik lekaime sartzen den emakume kopurua):

$$X \sim B(n = 5000, p = 2/100000 = 0.00002) \rightarrow P(\lambda = np = 0.1)$$

$$H_0 : p \leq 0.00002 \equiv \lambda \leq 0.1$$

(c)

Probaren norabidea zehazteko, hipotesi nulua (emakume gutxi sartzen da lekaime) goitik (emakume asko sartzen denean lekaime) baztertzen dela hartzen da kontuan.

$$p = P[X \geq 5/\lambda = 0.1] = 1 - P[X \leq 4/\lambda = 0.1] = 1 - \left(\frac{e^{-0.1} \times 0.1^0}{0!} + \frac{e^{-0.1} \times 0.1^1}{1!} + \frac{e^{-0.1} \times 0.1^2}{2!} + \frac{e^{-0.1} \times 0.1^3}{3!} + \frac{e^{-0.1} \times 0.1^4}{4!} \right)$$

(d)

Hau da emaitza zehatza ematen duen agindua:

```
> 1-ppois(4,0.1)
[1] 7.667802e-08
```

(e)

p balioa 0.0000000766 da. Adierazgarritasun- edo esangura-maila (0.01) baino txikiagoa denez, hipotesi nulua baztertu behar da, hots, batez besteko kopurua 5.000 ikasleen artean 0.1 baino handiagoa dela baieztatu behar da, horretarako ebidentzia (5.000 emakumeetatik 5 lekaima) aski sendoa delako. Beraz, Josemari Zarazola klaseetan gertatzen dena aski harrigarria da eta miraritzat hartu behar du Eliza Katolikoak.

II. ebazkizuna (1 puntu 10etik)

Josemari Zarazolaren ekimenez lekaiame sartutako neska gazteen gehiengo bat Bergarako Maria Stella Matutina Ahizpen komentura joan da. Komentu horretara bizitza kontenplatiboa garatzeko asmoarekin nobizia edo lekaimagaiak zoriz eta independentziaz sartzen dira, sarrera batetik bestera batezbeste 6 hilabete daudelarik.

Egin beharreko atazak:

1. (a) Zenbat da 4 lekaiame sartu bitartean 2 urte baino gutxiago pasatzeko probabilitatea? Kalkulua adierazita utzi, azken emaitza kalkulatu gabe.
2. (b) Nola banatzen da 4 lekaiame sartu bitarteko denbora (notazio laburrarekin behar bezala adierazi, parametroak izendatuz eta zehaztuz)?

Lekaimagaiak zoriz eta independentziaz sartzen direnez, Poisson-en prozesu bati jarraitzen diote. Eman dezagun lehenbizi, beraz, Poisson-en prozesua definitzen duen lambda parametroa. Lekaimagaien arteko batezbesteko denbora ematen da; beraz, banaketa esponentzialaren batezbestekoarekin berdindu behar dugu:

$$\frac{1}{\lambda} = 6 \text{ hilabete} \rightarrow \lambda_{hil} = \frac{1}{6} \rightarrow \lambda_{urte} = 12 \times \frac{1}{6} = 2 \text{ lekaimagai}$$

(a)

2 urteko epe bati buruz galdetzen digutenez, eman dezagun 2 urteko epeari dagokion lambda parametroa:

$$\lambda_{2 \text{ urte}} = 2 \times 2 = 4 \text{ lekaimagai}$$

4 lekaiame sartu bitartean 2 urte baino gutxiago igaroko dira, 2 urtetan $X=4$ lekaimagai edo gehiago sartzen direnean (X : 2 urtetan sartzen diren lekaimagaiak):

$$P[X \geq 4] = 1 - P[X \leq 3] = 1 - \left(\frac{e^{-2} \times 2^0}{0!} + \frac{e^{-2} \times 2^1}{1!} + \frac{e^{-2} \times 2^2}{2!} + \frac{e^{-2} \times 2^3}{3!} \right)$$

(b)

Y izendatzen badugu 4 lekaimagai sartu bitarteko denbora, Y aldagaia Gamma banaketaren arabera da:

$$Y \sim \Gamma(k = 4, \lambda = 4)$$

III. ebazkizuna (2 puntu 10etik)

Bergarako komentura sartzen den lekaimagai bakoitza Agustin Hiponakoa kristau filosofoaren lan guztiak latinez irakurri behar ditu, egunero pixka bat, poliki eta behar bezala ulertuz. Filosofoaren lan guztiek 12.345.678 hitzeko testua osatzen dute. Lekaimagai bakoitzak egunero irakurtzen duen testuaren bolumena probabilitate-banaketa ezezagun baten arabera banatzen da, 40.000 hitzeko eta 10.000 hitzeko batezbesteko eta desbideratzearekin, hurrenez hurren.

Egin beharreko atazak:

- (a) Zenbat egun behar ditu gutxienez lekaime batek filosofoaren lan guztiak irakurtzeko %99ko probabilitatearekin?
- (b) Zenbat da dena irakurtzeko 305 egun edo gutxiago behar izateko probabilitatea?
- (c) Egunero irakurtzen den testu-bolumenaren banaketa ezezaguna izan ordez, normala baliz, batezbesteko eta desbideratze berdinekin betiere, azken emaitza berdina izango al litzateke (ez duzu azken kalkulua egin behar)?
- (d) Eta uniforme baliz?

(a)

n egunetan irakurtzen X testu-bolumena honela banatzen da LTZ erabiliz (suposatuz egun batetik bestera independentzia dagoela eta aintzat hartuz egun kopuru hori handia izan behar dela derrigorrez kopuruak ikusita):

$$X \sim N(\mu = 40.000n; \sigma = \sqrt{10.000^2 n} = 10.000\sqrt{n})$$

Lan guztiak irakurriko ditu n egun horietan irakurtzen duen testu-bolumena 12.345.678 hitz baino handiagoa denean:

$$P[X > 12.345.678] = P\left[Z > \frac{12.345.678 - 40.000n}{10.000\sqrt{n}}\right] = 0.99$$

Pertzentilen-taulan bilatuta:

$$\begin{aligned} \frac{12.345.678 - 40.000n}{10.000\sqrt{n}} &= -2.32 \\ 40.000n - 23.200\sqrt{n} - 12.345.678 &= 0 \end{aligned}$$

$n = x^2$ aldagai-aldaketa eginez eta soluzio negatiboa baztertuz (desbideratze negatiboa ematen duelako):

$$40.000x^2 - 23.200x - 12.345.678 = 0 \rightarrow x = 17.86 \rightarrow n = 318.9$$

Beraz, motz ez geratzeko, 319 egun aurreikusi behar dira lan guztiak irakurtzeko.

(b)

305 egunetan irakurritako testu-bolumena nola banatzen den eman dezagun:

$$X \sim N(\mu = 40.000 \times 305 = 12.200.000; \sigma = 10.000\sqrt{305} = 174.642, 5)$$

Dena irakurtzeko 305 egun edo gutxiago beharko dira, egun horietan irakurtzen dena 12.345.678 hitz baino gehiago direnean:

$$P[X > 12.345.678] = P\left[Z > \frac{12.345.678 - 12.200.000}{174.642, 5}\right] = P[Z > 0.83] = \{taulak\} = 1 - 0.7967 = 0.2033$$

(c-d)

LTZ-k batuketan txertatzen diren banaketen banaketaren batezbestekoa eta desbideratzea baino kontuan hartzen ez dituenenez, horren aplikazioa independentea da batugai den banaketa bakoitzaren forma edo nolakotasun konkretuarekiko. Beraz, normala nahiz uniforme izanda emaitza berbera izango genuke. Normalarekin ugalkortasunaren propietatea erabil genezake baino orduan ere emaitza berdina da.