

OCW 2022

ESTADÍSTIKA INFERENTZIALA R SOFTWARE LIBREA ERABILIZ

Irakaskuntza gida

Eneko Arrospide Zabala

Gorka Bidegain Cancer

Xabier Erdocia Iriarte

Aitziber Unzueta Inchaurre

Bilboko Ingeniaritza Eskola—Matematika Aplikatua saila

AURKIBIDEA

1. SARRERA	1
2. NORI ZUZENDUA ETA AURREBALDINTZAK	1
3. HELBURUAK	1
4. GAITASUNAK	1
5. IKASTAROAREN DESKRIBAPENA	2
6. IRAKASLE TALDEA	2
7. GAITEGIA	3
8. METODOLOGIA	4
9. KRONOGRAMA	5
10. MATERIALAK ETA ERREKURTSOAK	6

1. SARRERA

Metodo eta teknika estatistikoak etengabe erabili behar dira zientzia eta ingeniartzako esparru guztietan. Honen barruan, estatistika inferentziala ezinbesteko tresna da, lagin baten azterketa egiterakoan lortutako emaitzak modu fidagarri batean orokortu ahal izateko.

Gaur egungo datu masiboak kudeatu behar dira, eta datu hauek kudeatzeko tresna informatikoak erabiltzea ezinbestekoa da. Horregatik, R software librearen erabilpena proposatzen da, datuak modu eraginkor batean aztertu eta horietatik ondorioak atera egin daitezkelako. Kurtso hau datuen kudeaketa eta teknika inferentzialak erabiliz lortutako ondorioak interpretatzeko lagungarria izango da.

2. NORI ZUZENDUA ETA AURREBALDINTZAK

Ikastaro hau estatistikako teknika inferentzialei buruz sakondu nahi duen ikasle guztiei zuzendua dago.

Ikastaroa arazorik gabe aurrera eraman ahal izateko, ikasleak estatistika deskribatzaileari buruzko kontzeptuak ezagutu behar ditu bai eta probabilitatearen oinarriak ere. Gainera, oinarritzko matematikako kontzeptuak argi eduki behar ditu.

3. HELBURUAK

Ikastaroaren helburuak honakoak dira:

- Teknika inferentzial ezberdinei buruz sakontzea
- Lagin bateko informazioa modu egokian populazio osoari orokortu ahal izatea
- R software librea erabiltzen ikastea teknika inferentzialak lagin ezberdinei aplikatu ahal izateko

4. GAITASUNAK

Ikastaroa burutzerakoan ikasleak lortu edo/eta indartu beharko lituzkeen gaitasunak honakoak dira:

- G1.** Zorizko laginketako estatistikoak jarraitu dezaketen banaketak identifikatzea eta bereiztea.
- G2.** Lagineko datuetatik abiatuz populazio osoaren parametroetarako estimazioak modu egokian burutzea.

- G3. Parametroei buruzko hipotesi-kontrasteetan ondorio egokiak ateratzea onartu beharreko hipotesia zehaztuz
- G4. R software librea erabiliz estatistika inferentzialaren ariketak ebazteko gai izatea.
- G5. Ingurune edo egoera aldakorretara egokitzeko ahalmena lortzea banakako ikasketa sustatuz.

5. IKASTAROAREN DESKRIBAPENA

Ikastaro honek estatistika inferentzialari buruzko ikuspegi zabala aurkezten du eta honen inguruan agertzen diren ariketak eta problemak ebazteko R software librearen erabilpena deskribatzen du. Kurtsoa bost gaietan banatzen da eta horietako bakoitzean lehenik gai horri buruzko oinarri teorikoak azaltzen dira eta ondoren pausuz pausu ebatzita dauden ariketak erakusten dira.

Gai bakoitzeko teoria eta ariketei zenbatetsitako dedikazioa aldakorra da eta hauen beharrei egokiturik dago, 9. ataleko kronograman agertzen den bezala. Dena den, ikasle bakoitzak denbora egoki iruditzen zaion bezala antolatu dezake ikaskuntzako gaitasunak eskuratzeko.

Ikastaro honetan, lehenik eta behin R software librearen instalazioa eta honen erabilera azaltzen dira. Behin hau ikusirik, estatistika inferentzialaren oinarrizko kontzeptuak erakusten dira, zorizko laginketa eta bertan ager daitezkeen banaketak deskribatuz. Azkenik, parametroen estimazioa eta hipotesi kontraste parametrikokoak zein ez-parametrikokoak azaltzen dira.

Bestalde, ikastaroaren amaieran, autoebaluazio proba bat agertzen da non ikasleak ikastaro honetako kontzeptuak bereganatu dituen konprobatu ahal izango duen.

6. IRAKASLE TALDEA

Ikastaro hau, Bilboko Ingeniaritza Eskolako Matematika Aplikatua saileko ondorengo irakasleek prestatu dute:

Izen-abizenak	Kategoria
Eneko Arrospide Zabala	Irakasle atxikia
Gorka Bidegain Cancer	Irakasle atxikia
Xabier Erdocia Iriarte	Irakasle atxikia
Aitziber Unzueta Inchaurre	Irakasle agregatua

Taula 1. OCW kurtsoa prestatu duten irakasle taldea.

7. GAITEGIA

1. GAIA. R eta Rstudio: ezaugarri nagusiak

Behin R eta RStudio software libreak instalaturik, hauen orokortasunak, oinarrizko kontzeptuak eta estatistika lantzeko eskaintzen dituzten tresnak deskribatzen dira gai honetan. Datuen irakurpena/gordetzea, hauen egitura, funtzioak eta grafikoak landuko dira, software-ak eskaintzan dituen paketeak erabiltzearekin batera.

2. GAIA. Zorizko laginketa. Laginaren estatistikoek banaketak

Laginketan lortutako estatistikoek jarrai dezaketen banaketak adierazten dira gai honetan. Kasu bakoitzean estatistikoek jarraitzen dituzten banaketak identifikatzea ezinbestekoa izango da parametroen estimazioa (3. gaia) eta hipotesi kontrasteak (4. gaia) lantzeko eta hauen inguruan proposaturiko ariketak ebazteko.

3. GAIA. Parametroen estimazioa

2. gaiko estatistikoak erabiliz populazioko parametroen estimazioa burutzen azaltzen da. Horretarako, hasieratik parametroen puntu estimazioa eta tarte estimazioa bereizten dira. Ondoren, parametro ezberdinetarako konfiantza tarteak eraikitzeko azalpenak ematen dira. Parametroen konfiantza tartearen kudeaketa R software librearen bidez azaltzen da.

4. GAIA: Hipotesi-kontraste parametrikoak.

Aztergai den populaziotik hartutako lagin batetik abiatuz, populazioko parametro bati buruzko baieztapen bat onar daitekeen edo ez erabakitzeko hipotesi kontraste parametrikoak azaltzen dira. Lehenenik eta behin, oinarrizko kontzeptuak, hipotesi kontraste motak eta hipotesi kontraste bat egiteko jarraitu beharreko urratsak azaltzen dira. Ondoren, zenbait hipotesi kontraste ikasten dira, bukatzeko, errore motak eta p-balioa ere aztertuz.

5. GAIA: Hipotesi-kontraste ez-parametrikoak.

Aztergai den populazioaren banaketari buruzko edota aztergai diren bi aldagaien menpekotasunari edo independentziari buruzko baieztapenak onartzeko, gai honetan hipotesi kontraste ez-parametrikoak aztertzen dira. Zehazki, Normaltasuna aztertzeko Kolmogórov-Smirnov eta Shapiro-Wilk probak deskribatzen dira. Independentzia proba berriz χ^2 -ren aplikazio bat bezala aztertzen da.

Gaitegi zehaztua hurrengo da:

1. GAIA. R eta Rstudio: ezaugarri nagusiak

- 1.1. Sarrera
- 1.2. Leihoak
- 1.3. Oinarrizko ezagutza
- 1.4. Datu egitura (bektoreak eta data markoak)
- 1.5. Funtzioak
- 1.6. Grafikoak (plot eta curve)
- 1.7. Paketeak
- 1.8. Lan direktorioa
- 1.9. Fitxategi batetik datuen irakurpena
- 1.10. Datu-markoak fitxategi batean gorde

2. GAIA. Zorizko laginketa. Laginaren estatistikoen banaketak

- 2.1. Sarrera
- 2.2. Banaketa normala
 - 2.2.1. Limite zentralaren teorema
 - 2.2.2. Lindeberg-Lévy-ren teorema
- 2.3. Gamma banaketa
- 2.4. Pearson-en χ^2 banaketa
- 2.5. Student-en t banaketa
- 2.6. Fisher-Snedecor-en F banaketa
- 2.7. Laginaren estatistiko garrantzitsuenen banaketak
 - 2.7.1. Laginaren batezbestekoa
 - 2.7.2. Bi laginen batezbestekoen diferentzia
 - 2.7.3. Laginaren bariantza
 - 2.7.4. Bi laginen kuasibariantzen arteko zatiketa
 - 2.7.5. Laginaren proportzioa
 - 2.7.6. Bi laginen proportzioen diferentzia
- 2.8. Banaketen kudeaketa R erabiliz

3. GAIA. Parametroen estimazioa

- 3.1. Estimazioaren kontzeptua
- 3.2. Puntu estimazioa
 - 3.2.1. Estimatzaileen propietateak
 - 3.2.2. Parametroen puntu-estimazioa
- 3.3. Tarte estimazioa
 - 3.3.1. Batezbestekorako konfiantza-tartea
 - 3.3.2. Bariantzarako konfiantza-tartea
 - 3.3.3. Proporziorako konfiantza-tartea
 - 3.3.4. Bi banaketa independenteen batezbestekoen arteko kenduraren konfiantza-tartea

- 3.3.5. Bi banaketa ez-independenteen batezbestekoen arteko kenduraren konfiantza-tartea
- 3.3.6. Bi banaketa independenteen bariantzen arteko zatiduraren konfiantza-tartea
- 3.4. Laginaren tamaina

4. GAIA: Hipotesi-kontraste parametrikoak.

- 4.1. Sarrera
- 4.2. Oinarrizko kontzeptuak
- 4.3. Hipotesi-kontraste motak
- 4.4. Hipotesi-kontrasteen urratsak
- 4.5. Zenbait hipotesi-kontraste
 - 4.5.1. Populazioaren batezbestekorako hipotesi-kontrastea
 - 4.5.2. Bi banaketa independenteren batezbestekoen arteko diferentziarako hipotesi-kontrastea
 - 4.5.3. Populazio normalaren bariantzarako hipotesi-kontrastea
 - 4.5.4. Banaketa normaleko bi populazio independenteren bariantzen arteko zatidurarako hipotesi-kontrastea
 - 4.5.5. Banaketa binomialaren proportziorako hipotesi-kontrastea ($n > 100$)
 - 4.5.6. Bi banaketa binomial independenteren proportzioen arteko diferentziarako hipotesi-kontrastea ($n, m > 100$)
 - 4.5.7. Bi banaketa normal ez independenteren batezbestekoen arteko diferentziarako hipotesi-kontrastea
- 4.6. Errore motak
- 4.7. p-balioa

5. GAIA: Hipotesi-kontraste ez-parametrikoak.

- 5.1. Sarrera
- 5.2. Independentzia probak
- 5.3. Normaltasun probak
 - 5.3.1. Kolmogórov-Smirnov test-a
 - 5.3.2. Shapiro-Wilk test-a

8. METODOLOGIA

Ikastaro hau bost gaietan banatzen da eta bakoitzean teoria eta ariketa zerrenda bat agertzen da.

Ikastaro honi etekinik handiena ateratzeko modu sekuentzian lan egitea gomendatzen da, hasieratik bukaerara arte. Gai bakoitzean jorratutako kontzeptuak hurrengo gaietarako erabilgarriak dira. Beraz, ondorengo metodologia proposatzen da:

1. R eta Rstudio software librea ordenagailu pertsonalean instalatu, ikastaroan argitaraturik dauden jarraibideak jarraituz.
2. Gai bakoitzaren teoria arretaz irakurri kontzeptu garrantzitsuenak ulertuz eta barneratuz. Zalantzarik egotekotan, proposatutako gomendaturiko irakurgaiak eta bestelako materialak erabili hauek argitu ahal izateko.
3. Gai bakoitzeko ebatziak dauden ariketak irakurri, ulertu eta berriro egiten saiatu emaitza berdinak lortzen direla konprobatuz.
4. Autoebaluazioa burutu eta ebatzitako autoebaluazioarekin alderatu irakasgaiko kontzeptu guztiak barneratu direla egiaztatzeko.

9. KRONOGRAMA

Kurtsoaren kronograma sekuentziala da, hau da, ikastaroan aurrera joan ahal izateko aurreko gaietako kontzeptuak barneraturik eduki behar dira. Horrela, ikastaroari ahalik eta etekin handiena ateratzeko ondorengo kronograma proposatzen da:

Astea	Zeregina	Iraupena
1	1. gaiko teoria	7 ordu
2	1. gaiko ariketak	4 ordu
3	2. gaiko teoria	8 ordu
4	2. gaiko ariketak	4 ordu
5	3. gaiko teoria	7 ordu
6	3. gaiko ariketak	4 ordu
7	4. gaiko teoria	7 ordu
8	4. gaiko ariketak	4 ordu
9	5. gaiko teoria	6 ordu
10	5. gaiko ariketak	4 ordu
11	Autoebaluazioa	10 ordu
		GUZTIRA: 65 ordu

Taula 2. Ikastaroko zereginen asteko ordena eta hauei eskaintzea gomendatutako denbora.

		ASTEAK										
		1	2	3	4	5	6	7	8	9	10	11
1. GAIA												
Teoria (7h)	Ariketak (4h)											
2. GAIA												
Teoria (8h)	Ariketak (4h)											
3. GAIA												
Teoria (7h)	Ariketak (4h)											
4. GAIA												
Teoria (7h)	Ariketak (4h)											
5. GAIA												
Teoria (6h)	Ariketak (4h)											
AUTOEBALUAZIOA												
Testa (4h)	Ariketak (6h)											

1. Irudia. Ikastaroaren kronograma.

10. MATERIALAK ETA ERREKURTSOAK

Ikastaro hau burutu ahal izateko, edozein sistema eragile duen ordenagailu bat edukitzea beharrezkoa da R Software librean erabiltzeko. Gainera, ikastaroan eskuragarri dagoen materialaz gain, gomendaturiko irakurgaiak eta bestelako materialak oso lagungarriak dira kurtsua aurrera eraman ahal izateko.