

Dibertsitatea

Josemari Sarasola

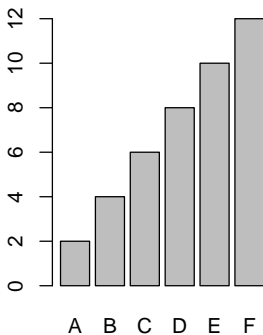
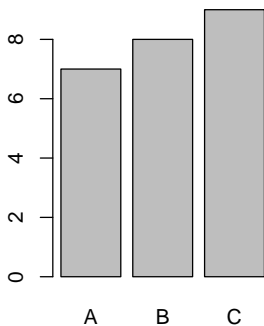
Estatistika eta datuen analisia

Gizapedia



- Maiztasun-banaketa batean (orokorrean, kualitatiboa) dibertsitatea handiagoa izango da, aldi berean, (1) totala kategorien artean parekotasun handiagoarekin (kontzentrazio txikiagoarekin) banatzen denean, edota (2) totala kategoria gehiagoren artean banatzen denean.
- Adibidez (1), har ditzagun bi banaketa hauek: 4-4-4 eta 3-3-6. Bietan kategoria kopurua berdina da, baina dibertsitatea handiagoa da lehenengoan, parekotasun handiagoa dagoelako.
- Adibidez (2), har ditzagun bi banaketa hauek: 4-4-4 eta 6-6-6-6-6. Bietan erabateko berdintasuna dago, kontzentrazioari buruz, eta sakabanatzea zero da, baina dibertsitatea handiagoa da bigarrenean, kategoria gehiagoren artean gertatzen delako.
- Beraz, **dibertsitatea** bi osagaien emaitza da: **aniztasuna** (kategoria edo espezie kopurua) eta **parekotasuna** (ingelesez, *evenness*: elementuak zenbateraino banatzen diren era berdintsuan kategoria horien artean).

Hausnartu: zein banaketatan dago dibertsitate handiena?



- Dibertsitatea neurtzeko **Shannon adierazlea** erabil daiteke (esan behar da beste neurri asko daudela aukeran):

$$H = - \sum p_i \ln p_i$$

p_i izanik banaketako proportzioak batekotan (ez ehunekotan). Zenbat eta handiagoa izan, dibertsitatea orduan eta handiagoa da.

- Baina Shannon adierazlea era horretan adierazita, ezin dira dibertsitatearen bi osagaiak, aniztasuna eta parekotasuna alegia, bereizi. Hobe da **Hill zenbakia** erabiltzea horretarako:

$$e^H$$

- Esan bezala, dibertsitatea bi osagaitan banatzen da: aniztasuna eta parekotasuna. Banaketa zenbat eta anitzagoa eta zenbat eta parekoagoa, orduan eta dibertsitate handiagoa dago.
- Bi osagai horietatik aniztasuna banantzea oso erraza da: aniztasunaren neurri gisa S kategoria-kopurua baino ez dugu hartuko.
- Parekotasuna neurtzea zailagoa da: horretarako jo dezagun berriz ere e^H baliora.

e^H balioak, dibertsitate neurri gisa, maiztasun handia (handia, era lausoan adierazita eta beste maiztasunekin alderatuta) duten kategorien kopurua adierazten du. Aiseago ulertzeko, ondorengo bi muturreko kasuak aztertuko ditugu.

Kategoria	p_i
A	0.97
B	0.01
C	0.01
D	0.01

$$e^H \approx 1$$

Izan ere, beste kategorien aldean, maiztasun handiko kategoria bakarra dago.

Kategoria	p_i
A	0.25
B	0.25
C	0.25
D	0.25

$$e^H = 4$$

Izan ere, beste kategoriekin alderatuta, lau kategoriek dute maiztasun handia.

Beraz **badirudi** e^H parekotasun-neurria dela, aurreko adibideetan ikus daitekeenez, banaketa guztiz parekoa den kasuan (kategoria guztiek %25eko proportzioa dutenean, alegia), e^H neurriak balio handiena hartzen duelako,

baina hau ez da guztiz zuzena e^H balioa S kategoria kopuruaren mendean dagoelako baita ere, aniztasuna ere jasoz ondorioz. Adibidez, $S = 4$ ordez $S = 5$ kategoria edukita, parekotasun osoz $p_i = 0.2$ -ko proportzioekin, $e^H = 5$ izango genuke, parekotasun mailak bietan ($S=4$ eta $S=5$) bietan berdinak izanda ere.

Irtenbidea oso sinplea da: e^H balioa S -rekiko independente egiteko, aski da S -rekin zatitzea, eta orduan horrela geratzen da deskonposaketa:

$$e^H = \frac{e^H}{S} \times S$$

$$dibertsitatea = \text{parekotasuna} \times \text{aniztasuna}$$