

# DATUEN KUDEAKETA ETA ANALISI ESTATISTIKOA I

**Sara Bidarte Begiristain**

**2018 – 2019 ikasturtea**

## ALDAGAIA, KONSTANTEA ETA ALDAGAI MOTAK:

Estatistikaren oinarria aldagaiak aztertzea izango da, eta Gizarte Zientzietan ez du konstanterik izango, ezaugarriak izango ditugu, zeinak aldatu egingo diren.

Aldagaiak neurgarriak diren eta elementu eta pertsonen artean aldatzen diren elementuak neurtuko dituzte.

Normalean neurketen balio bera izango dute, baina Gizarte Zientzietan ez dago konstanterik; beraz, balioak aldatuko dira.

### ALDAGAI MOTAK:

#### 1) Neurketa eskalaren arabera:

##### a. NOMINALA/ATRIBUTODUNA: (KUALITATIBOA)

Bakarra aukera daiteke eta eksklusiboak dira. Gainera, bakarra dagokigu, ez daude bi aukera.

- Adibidez, sexua dikotomikoa litzateke.

##### b. ORDINALA: (KUALITATIBOA)

Bakarra aukeratu behar da baina mailakatuta daude (txikitik handira)

- Adibidez, unibertsitateko ikasketetatik DBHra, ikasketa maila

##### c. TARTE ESKALA: (KUANTITATIBOA)

Zenbakiak ditugu eskala mota honetan.

- Adibidez, adin tarteak kalkula ditzakegu

##### d. RATIO ESKALA: (KUANTITATIBOA)

Aldagai konposatu bat dago eta bi aldagaien arteko harremana (konparaketa) litzateke.

- Adibidez, jaiotza tasa 1,7 (emakume emankorrek izango dituzten seme-alaben batz bestekoa)

#### 2) Neurtzeko moduaren arabera:

##### a. KUALITATIBOAK EDO KATEGORIKOAK:

Nolakotasuna neurtzen dute eta ez kopurua (zenbakirik ez). Erantzunak kategorietan banatuta daude, mailakatuta egon daitezke edo ez. –

Adibidez, ikasle baten notak jartzerako garaian (gutxi, nahiko, ongi, oso ongi, bikain)

- NOMINALA: aukera bakarra
- ORDINALA: mailaketa

##### b. KUANTITATIBOAK:

Zenbakiekin neur ditzakegunak, kopuru tinkoak edo komadunak izan daitezke.

Adibidez, seme-alaba kopurua...

- DISKRETUAK: zenbaki tinkoak
- JARRAITUAK: kategorien jarraikortasuna

### 3) Haien arteko harremanen arabera

- a. ALDAGAI INDEPENDENTEA
- b. ALDAGAI DEPENDENTEA

### 4) Edukiaren arabera

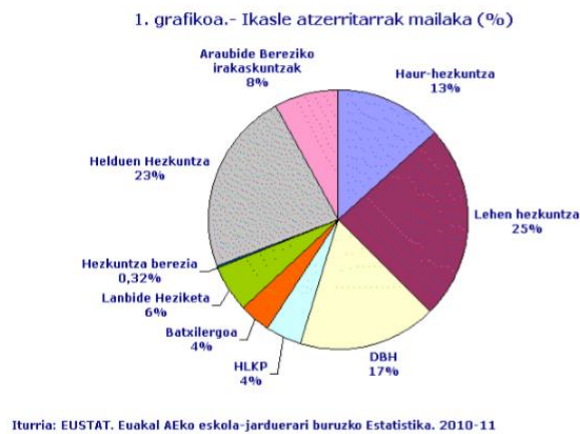
- a. ZENBAKIAK ERABILTZEN DITUENA
- b. EDUKI ALFANUMERIKOAK ERABILTZEN DITUZTENAK

## DATU MULTZOEN ADIERAZPEN GRAFIKOAK ETA DATU DIAGRAMAK:

### 1) SEKTORE DIAGRAMA:

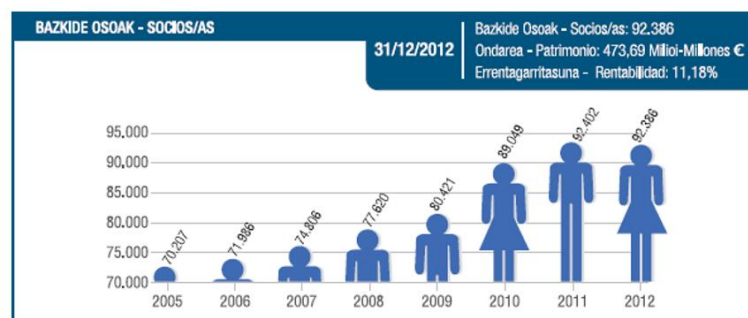
Askotan gazta ere deitzen diogu; hemen, portzentaiak erabiliko ditugu. Beti izan behar du izenburu eta zenbaki bat, eta baita iturria ere.

Banatu egin daitezke. Hots, borobilak ez du zertan elkartuta egon.



### 2) PIKTOGRAMAK:

Bi aldagai nahastuko ditugu, eta aipatu barra diagrama bat dela, baina zutabeak beharrean pertsonak adierazten direla.

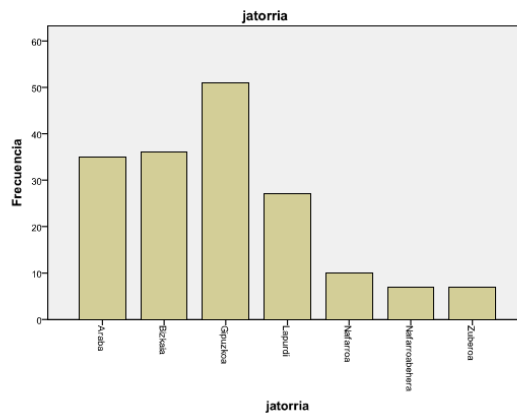


### 3) BARRA DIAGRAMA EDO ZUTABE DIAGRAMA:

Normalean bi aldagai nahastuko ditugun arren, aldagai batekoak zein hirukoak izan daitezke.

Barrak laukizuzen formakoak dira, eta digrama hauek horizontalak zein bertikalak izan daitezke.

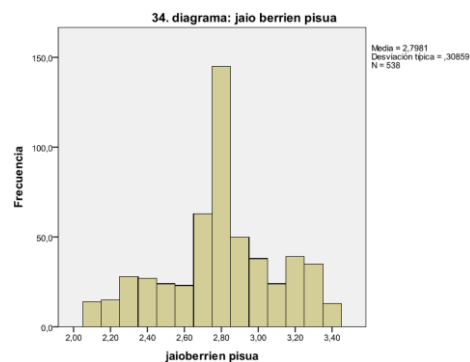
Barra bat edo gehiago izan ditzazke



Iturria: lanketa propioa

#### a. HISTOGRAMA:

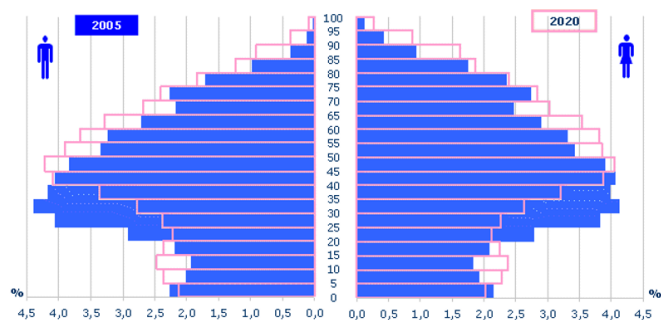
Barra diagrama bat da, baina zutabeak elkar itsatsita daude, eta **aldagai kuantitatibo**entzat dira baliagarriak. Gainera, jarraikortasuna bat dago, eta lotu egin daitezke.



Iturria: nik neuk egina

#### i. BIZTANLERIA PIRAMIDEA:

Histograma bat da, baina hemen hiru aldagai (adina, kopurua eta sexua) izango ditugu. Bi urte ezberdinetakoak aldera ditzakegu.



Iturria: EUSTAT(www.eustat.es)



## ZEIN ALDAGAI MOTATAN ERABIL DEZAKEGU ZEIN DIAGRAMA?

26. taula: diagramen erabilera, aldagai motaren arabera				
Diagrama mota	Aldagai nominala	Aldagai ordinala	Aldagai diskretua	Aldagai jarraitua
Sektore-diagrama	bai	Bai	bai	bai
Piktograma	bai	Bai	bai	bai
Barra-diagrama	bai	Bai	bai	bai
Kartograma	bai	Bai	bai	bai
Histograma	ez	Ez	bai	bai
Maiztasun bakunen poligonoa	ez	Ez	bai	bai
Ojiba edo maiztasun metatuen poligonoa	ez	Ez	bai	bai
Lerro-diagrama	ez	Ez	bai	bai
Kaxa-diagrama	ez	Ez	bai	bai
Iturria: lanketa propioa				

## ESTATISTIKA DESKRIBATZAILEA:

Datu multzo bat dugu, eta lehenengo datuak ordenatu, grafikora pasa eta ordenatuko ditugu.

Datuak ordenatzeko **MAIZTASUN TAULA** erabiliko dugu.

- Taula sortu ahal izateko datuak ordenatuko ditugu (txikitik handira), eta dezimal batera borobilduko ditugu.
- Kategoria bakoitzeko zenbat ditugun adierazi; MAIZTASUN ABSOLUTOAK
- BALIO GALDUA: aurkeztu ez diren ikasle kopurua litzateke
- MAIZTASUN ABSOLUTO METATUAK: maiztasun absolutoen arteko batuketa litzateke
- MAIZTASUN ERLATIBOAK (PROPORTZIOAK):

$$\text{Maiztasun erlatiboa} = \frac{\text{Maiztasun absolutoa}}{\text{Baliiodunak}}$$

- MAIZTASUN ERLATIBOA (PORTZENTAJEAK):

$$\text{Maiztasun erlatiboa} = \frac{\text{Maiztasun absolutoa}}{\text{Baliiodunak}} \cdot 100$$

- MAIZTASUN ERLATIBO METATUAK (PORTZENTAJEAK): maiztasun erlatiboen batuketa.

MAIZTASUNEN INTERPRETAZIOAK (Txostena nola idatziko dugu?)

- Txostena zeren inguruan den, zein gai den, adieraziko diogu irakurleari. Hots, inkestatuen kokapena eta testuingurua, noiz, non eta nola, laginaren tamaina, fidagarritasuna eta baliogarritasuna (Batzuetan ez da dena azalduko)
- Gaiaren aukezpena, galderen formulazioa eta emaniko erantzun kategoriak, diseinuaren eta kontestuaren ikurra.
- Datuak komentatzen hasiko gara:
  - Balio galduaren interpretazioa: normalizat hartu? Baliogarritasun arazorik?
  - Maiztasunik garrantzitsuenak, zer dio jendeak? (Normalean handitik txikira)
  - Beste maiztasun interesgarriak?

- Ausartuko zinateke gehiagorik esaten? Hipotesirik? (Adinak garrantzia du)
- Egin beharreko analisi estatistiko gehiagorik? Beste aldagaiekin gurutzaketak?

## NEURRI ESTATISTIKO DESKRIBATZAILEAK:

### ZENTRO NEURRIAK EDO JOERA ZENTRALEKO ESTATISTIKOAK:

- BATAZ BESTEKO ARITMETIKO SINPLEA:

Bi aukera daude honetarako:

1.

$$\bar{x} = \frac{\sum x}{N}$$

2.

$$\bar{x} = \frac{\sum f \times}{N}$$

*Tarteak* ditugunean erdikoa hartuko dugu. Hots, 2-4 tartean 3 hartuko dugu. Tarteko puntua denez ez dugu informazio guztia lortuko informazioa bidean galduko dugulako.

- BATAZ BESTEKO ARITMETIKO HAZTATUA:

Aldagaia egokitu egingo da biztanleriaren ezaugarrietara. Hots, bere benetako tamainarekin biderkatuko dugu. Adibidez, nota finalak kalkulatzeko egingo dena.

- MOZTUTAKO BATAZ BESTEKO (%20):

Muturreko balioak kenduko ditugu, eta zentrokoak bakarrik hartuko ditugu. Kontuan izan, balioak kendu ditugunez guztien kopurua ere aldatu egiten dela.

$$\bar{x} = \frac{\sum f \times}{N}$$

- MEDIANA:

Zentroa non dagoen adieraziko digu, baina beste modu batera adierazita.

Muturreko neurriak neutralizatuko ditu kalkulu honek; taldea ondo ordezkatzeko du mutur extremoak badaude.

$$Md = \frac{N}{2}$$

- MODA:

Datu multzo batean gehien agertzen den balioa litzateke.

## SAKABANATZE NEURRIAK – DISPERSIOA:

- IBILTARTEA:

Agertutako datuek nondik nora jotzen duten.

$$I = X_{max} - X_{min}$$

- BATAZ BESTEKOAREKIKO BATEZ BESTEKO DESBIDERATZEA:

Zenbat eta handiagoa orduan eta sakabanatuago, eta zenbat eta txikiagoa elkartuago. Beraz, muturreko balioek garrantzia handiagoa dute. Diferentziak kontuan izan behar dira: batz bestekora iritsi gabe negatiboa eta batz bestekoa pasata positiboa.

$$Dx = \frac{\sum |x - \bar{x}| \cdot f}{N}$$

- BARIANTZA:

Batz bestekorik balio bakoitzeko dagoen aldea kalkulatzeko baliagarria da.

Zenbat eta hurrunago balio handiagoak. Honek, analisis egiteko joko handiagoa emango du.

Ez da estatistika deskribatzailean erabiliko batz bestekoarekin zerikusirik ez duelako!

$$S^2 = \frac{\sum (x - \bar{x})^2 \cdot f}{N}$$

$$S^2 = \frac{\sum x^2 \cdot f}{N} - \bar{x}^2$$

- DESBIDERATZE TIPIKOA:

Hau izango da balio duena, eta bariantzaren errokatua litzateke.

$$S = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot f}{N}}$$

$$S = \sqrt{\frac{\sum x^2 \cdot f}{N} - \bar{x}^2}$$

- ALDAKORTASUN KOEFIZIENTEA:

Bi taldeen arteko ezberdintasunak ikusteko erabiliko dugu. Beraz, bi taldeen arteko barne aniztasuna azaltzeko erabiliko dugu.

Emaitza zenbat eta portzentai handiagoa aldakortasun handiagoa eta desberdintasun handiagoa.

$$AK = \frac{S}{\bar{x}} \cdot 100$$



## JOERA ZENTRALEKOAK EZ DIREN POSIZIO NEURRIAK:

Balio metatuetan erabiliko ditugu mediana, koarteilak, dezilak eta pertzentilak. Adibidez, mediana %50 dagoen tokian, eta koartila %25ean.

- KOARTILA (Q): lau zatitan banatuko da
- DEZILA (D): hamar zatitan banatuko da.

Konparazioak egiteko 1 eta 9 dezilak hartuko ditugu. Adibidez, aberastasuna neurtzeko erabiliko da.

- *Dezilen kalkulua balio isolatuetan:*

$$\frac{KN}{10}$$

- PERTZENTILA (P): ehun zatitan banatuko da

- *Pertzentilen kalkulua balio isolatuetan:*

$$\frac{KN}{100}$$

✓ Kasu guztietan kopurua bikoitia bada +1 egingo dugu.

Joera zentralakoak ez diren posizio neurriei buruzko ONDORIO OROKORRAK:

- Kuantilak aldagai kuantitatiboetarako eta kualitatibo ordinaletarako kalkula daitezke.
- Ezin da ahantzi kuantilen artean baliokidetasunak daudela
- Datu multzo bateko koartilak modu grafikoan adierazteko edota aztertzeko maiztasun metatuen poligonoa erabil dezakegu.
- Kuantilak oso egokiak dira maiztasun-banaketa ezberdinak adierazteko.

## FORMA EDO ITXURA NEURRIAK:

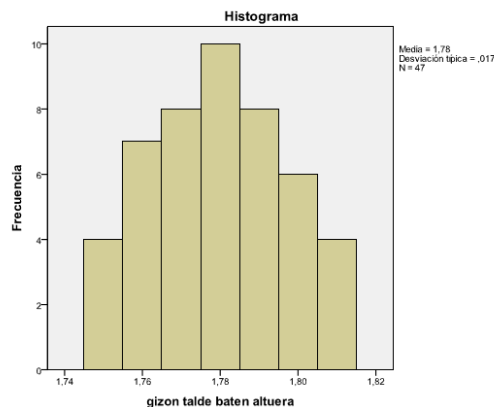
- ALBORAPENA:

Zein motatako asimetria izango dugun adieraziko digu.

Honela interpretatuko ditugu emaitzak:

- $AB = 0$  bada edo Ora hurbiltzen bada, **maiztasun-banaketa simetrikoa edo ia simetrikoa** dela esango dugu.

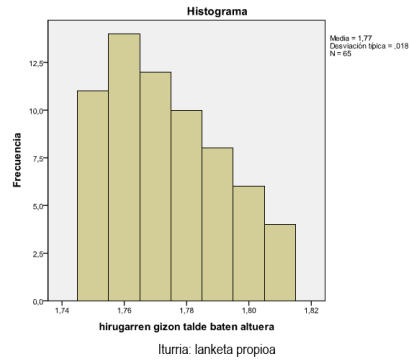
29. diagrama: A gizon taldearen altuerak



Iturria: lanketa propioa

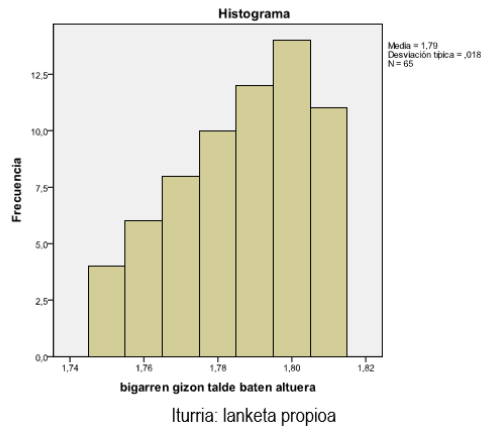
- $AB > 0$  bada, maiztasun-banaketa eskuin aldera alboratuta dagoela edo **alborapen positiboa** duela esango dugu;

31. diagrama: C gizon taldearen altuerak

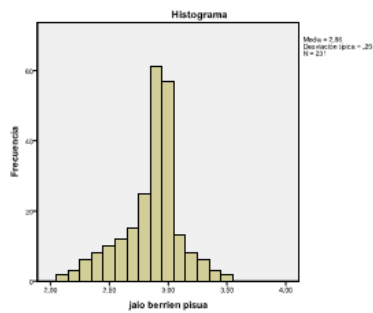


- $AB < 0$  bada, maiztasun-banaketa ezkerrera alboratuta dagoela edo **alborapen negatiboa** duela esango dugu.

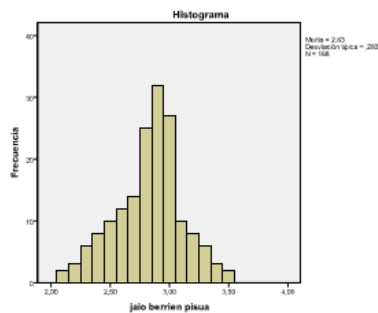
30. diagrama: B gizon taldearen altuerak



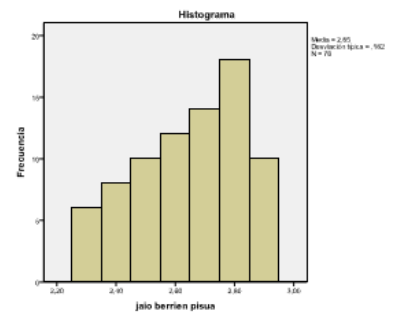
• **KURTOSIA:**



Leptokurtikoa: banaketa normalak baino zorrotasun edo kurtosi handiagoa.



Mesokurtikoa: banaketa normalaren zorrotasun edo kurtosi berdina.  
Iturria: lanketa propioa



Platikurtikoa: banaketa normalak baino zorrotasun edo kurtosi txikiagoa.

Abantailak eta desabantailak:

- Kurtosiak datu multzoaren sakabanatzeri buruzko informazioa ematen digu. Beraz, zorrotasun handiagoa eta eta sakabanatze txikiagoa

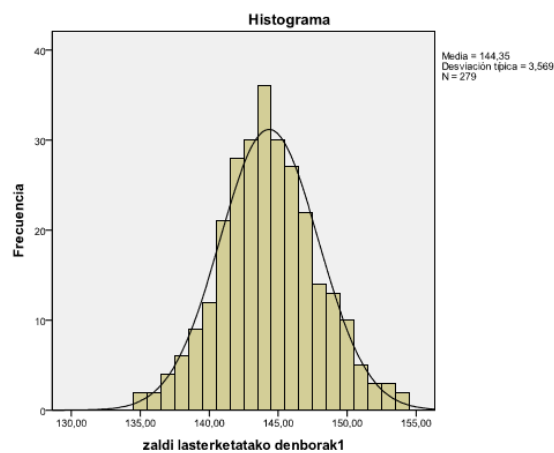
- Abantaila: kalkulatzean dau guztiak erabiltzen dira
- Desabantaila: ez da jasankorra muturreko datuek eragin handia dutelako.

## BANAKETA NORMALA:

### EZAUGARRIAK:

1. Batzbestekoak, medianak eta modak balio bera edo oso antzekoa
2. Aldagai diskretu eta jarraituetan soilik erabiliko dugu
3. Desbiderazi tipikoa txikia izango da balioak elkartuta daudelako
4. Formaren aldetik simetrikoa izango da eta kurtsoiaren aldetik mesokurtikoa

41. diagrama: zaldi-lasterketetako denboren histograma eta kurba normala



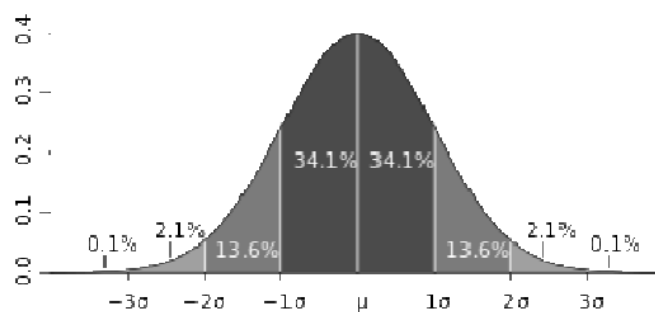
Iturria: lanketa propioa

### BALIOAK TIPIFIKATU:

$$Z = \frac{X - \bar{X}}{S}$$

### GAUSSEN KANPAIA ETA '68 – 99 – 97,5' ARAUA:

44. diagrama: Gauss kanpaia eta "68-95-99,7 araua"



Iturria: [https://commons.wikimedia.org/wiki/File:Standard\\_deviation\\_diagram.svg](https://commons.wikimedia.org/wiki/File:Standard_deviation_diagram.svg)

## ESTADÍSTICA INFERENCIALA:

Kasu honetan lagina ezagutzen dugu, baina ezagutzen ez duguna populazioa da. Hortaz, lagin probabilitikoa erabiliko dugu, eta ez ez-probabilitikoa.

### ESTIMAZIO PUNTUALA:

Lagin bat ezaugarri batzuekin hartuko dugu, eta beste lagin bat hartzean estimazio bat egin. Beraz, lehen estimazioa denentzat.

Kalkulu honetan bi errore izango ditugu. Hots, balio estatistiko bakoitzak bere errorea du.

### TARTE ESTIMAZIOA:

$$PARAMETROA = ESTADISTIKOA \pm KONFIANTZA MAILA \cdot ERRORE TIPIKOA$$

- Batzbestekoarentzako tarte estimazioaren formula:

$$\mu = x \pm z \cdot \sqrt{x}$$

- Portzentaientzako tarte estimazioaren formula:

$$P = p \pm z \cdot \sqrt{p}$$

### BALIO KRITIKO ERABILIENAK:

1 - $\alpha$	$\alpha$	Z
0.90	0.10	1.645
0.95	0.05	1.96
0.99	0.01	2.275

### ERRORE TIPIKOA:

$$1. \sigma_x = \frac{s}{\sqrt{n}}$$

$$2. P = \frac{\sqrt{p \cdot q}}{n}$$

### HIPOTESI FROGA EDO HIPOTESI KONTRASTEAK:

Hemen gure ustea zein den adieraziko da. Beraz, errealitate jakin batetik nire hipotesia ondo dagoen edo ez egiaztatuko dut.

Hipotesia barnean badago ondo legoke (hipotesi nulua). Hala ez bada hipotesi alternatiboa adierazgarritasun maialrekin egiaztatuko dugu.

## HIPOTESI FROGA EGITEKO PAUSOAK:

1. Hipotesi nulua eta alternatiboa zein diren zehaztuko dugu. Hau da, aztergai den populazioaren inguruko baieztapenak egin behar dira. Lagungarria da hipotesi froga modu grafikoan adieraztea.
2. Adierazgarritasun maila finkatu behar da. Hots, aldez aurretik zehaztu behar da.
  - %90  $\rightarrow$  0.1
  - %95  $\rightarrow$  0.05
  - %99  $\rightarrow$  0.01
3. Onarpen eremua zehaztu: erabakitako adierazgarritasun mailaren arabera, onarpen eremua eta eremu kritikoa desberdinduko dituzten eremuak zehaztu.
4. Egiaztatu: lagina aztertuta, hipotesi nulua edo gezurra den erabakitzen lagunduko digun probarako estatistikoa atera.
5. Erabaki estatistikoa hartu: lortutako probarako estatistikoa eremu kritikoa kokatzen bada, hipotesi nulua errefusatuko da, adierazgarritasun mailarekin. Kontrako kasuan, hau da, probarako estatistikoa onarpen eremuan badago, hipotesi nulua onartuko da, adierazgarritasun mailarekin.

## Hipotesi- froga aldebikoak eta aldebakarrekoak:

Egin nahi dugun ebaluazioaren arabera, hiru eratako hipotesi-frogak egin ditzakegu:

### 1. Hipotesi froga aldebikoa

$$H_0 = k \quad H_1 \neq k$$

Hipotesi nulua esaten du parametroa ez dela aldatu, eta hipotesi alternatiboa aldatu dela. Estimazio arrunta egingo dugu. Dena den, kontuan izan behar dugu hipotesiak, betiere, parametroen arabera direla.

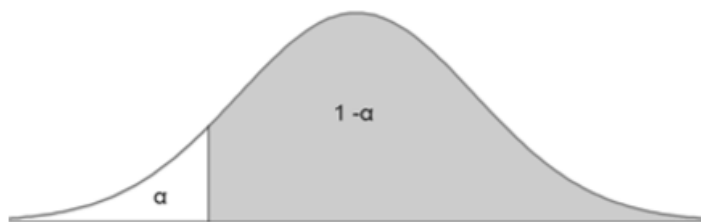


Hemen erabiliko ditugun konfiantza mailak betikoak izango dira!!!

### 2. Hipotesi froga aldebakarrekoa, $\alpha$ ezkerrean dagoenean:

$$H_0 \geq k \quad H_1 < k$$

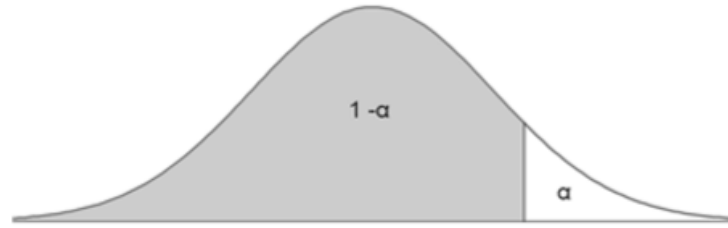
Kasu honetan, hipotesi alternatiboa esaten du parametroa txikitu egin dela.



**3. Hipotesi froga aldebakarrekoa,  $\alpha$  eskuinean dagoenean:**

$$H_0 \leq k \quad H_1 > k$$

Hipotesi alternatiboak esaten du parametroa handitu egin dela.



Aipatu hipotesi froga aldebakarrekoetan muga bakarra dugunez batuketa edo kenketa soilik egingo ditugula, eta konfiantza mailei dagokienez aldaketak izango direla.

$1 - \alpha$	$\alpha$	Z
0.90	0.10	1.28
0.95	0.05	1.645
0.99	0.01	2.33