# Matematika Estatistikoa

Oihana Aristondo Echeberria
José David Núñez González
Lucía Porlán Ferrando

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

# Matematika Estatistikoa

# Matematika Estatistikoa

Oihana Aristondo Echeberria
José David Núñez González
Lucía Porlán Ferrando

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

*CIP. Unibertsitateko Biblioteka*

**Aristondo Echeberria, Oihana**

Matematika Estatistikoa [Recurso electrónico] / Oihana Aristondo Echeberria, José David Nuñez González, Lucía Porlán Ferrando. – Datos. – [Leioa] : Universidad del País Vasco / Euskal Herriko Unibertsitatea, Argitalpen Zerbitzua = Servicio Editorial, [2025]. – 1 recurso en línea: PDF (92 p.). – (Unibertsitateko Eskuliburuak = Manuales Universitarios)

Modo de acceso: World Wide Web

ISBN: 978-84-9082-965-3

1. Estadística matemática. I. Núñez González, José David, coaut. II. Porlán Ferrando, Lucía, coaut.

(0.034)519.2

Gipuzkoako Ingeniaritza Eskola, Eibarko atala
Matematika Aplikatua Saila

© Servicio Editorial de la Universidad del País Vasco
Euskal Herriko Unibertsitateko Argitalpen Zerbitzua

ISBN: 978-84-9082-965-3

# Gaien Aurkibidea


**1 Estatistika deskribatzailea** **1**
- 1.1 Populazioa eta lagina . . . 1
- 1.2 Ezaugarri motak . . . 1
- 1.3 Ezaugarri Kuantitatiboak . . . 2
  - 1.3.1 Maiztasun-banaketa . . . 2
  - 1.3.2 Deskribapen grafikoak . . . 4
- 1.4 Ezaugarri Kualitatiboak . . . 6
  - 1.4.1 Maiztasun-banaketa . . . 6
  - 1.4.2 Deskribapen grafikoak . . . 6
- 1.5 Lagin kuantitatibo sinpleen estatigrafo nagusiak . . . 8
  - 1.5.1 Zentralizazio neurriak . . . 8
  - 1.5.2 Sakabanatze-neurriak . . . 11
  - 1.5.3 Asimetria-neurriak . . . 12
  - 1.5.4 Kutxa-Diagramak . . . 14

**2 Erregresioa eta korrelazioa** **17**
- 2.1 Sakabanatze eta erregresio-diagramak . . . 17
- 2.2 Erregresioa: Zer da? . . . 18
  - 2.2.1 Erregresio motak . . . 18
  - 2.2.2 Parametroak kalkulatzea . . . 18
- 2.3 Korrelazioa: Zer da? . . . 19
- 2.4 Erregresio-parametroak aurkitu . . . 19
  - 2.4.1 Erregresio lineala . . . 19
  - 2.4.2 Koefiziente horien eta korrelazioaren arteko erlazioa . . . 19
  - 2.4.3 Erregresio esponentziala . . . 20
  - 2.4.4 Erregresio potentziala . . . 20
  - 2.4.5 Erregresio parabolikoa . . . 21

**3 Konbinatoria** **23**
- 3.1 Aldakuntzak . . . 23
  - 3.1.1 Errepikapen gabe . . . 23
  - 3.1.2 Errepikapenekin . . . 24
- 3.2 Konbinazioak . . . 25
  - 3.2.1 Errepikapen gabe . . . 25

3.2.2 Errepikapenekin . . . 25
3.3 Permutazioak . . . 26
3.3.1 Errepikapen gabe . . . 26
3.3.2 Errepikapenekin . . . 26
3.4 Eskema . . . 27
3.5 Zenbaki konbinatorioak . . . 28

**4 Probabilitatea** **29**
4.1 Definizioak . . . 29
4.2 Probabilitatea . . . 31
4.3 Baldintzazko probabilitate . . . 32
4.4 Biderketa-arauak . . . 33
4.5 Probabilitate totala . . . 34
4.6 Bayes-en teorema . . . 35

**5 Ausazko Aldagai Diskretuen Banaketa** **37**
5.1 Ausazko aldagaia . . . 37
5.2 Ausazko aldagai diskretua . . . 37
5.3 Aldagai diskretuen banaketen batez bestekoa eta bariantza . . . 39
5.4 Ausazko aldagai diskretuen banaketa . . . 40
5.4.1 Banaketa hipergeometrikoa . . . 40
5.4.2 Banaketa binomiala . . . 41
5.4.3 Poisson-en banaketa . . . 42
5.4.4 Banaketa multinazionala/polinomiala . . . 43
5.4.5 Banaketa binomial negatiboa . . . 43
5.4.6 Banaketa geometrikoa . . . 43
5.5 Laburpen-Eskema . . . 44

**6 Ausazko aldagai jarraituen banaketa** **45**
6.1 Banaketa-funtzioa . . . 45
6.2 Dentsitate-funtzioa . . . 45
6.3 Banaketak . . . 46
6.3.1 Banaketa normala . . . 46
6.3.2 Pearsonen Chi-karratua banaketa . . . 48
6.3.3 Student-en T banaketa . . . 49
6.3.4 Snedecor-en F banaketa . . . 50

**7 Laginketa eta Zenbatespena** **51**
7.1 Parametroen estimazioa . . . 51
7.1.1 Estimazio puntuala . . . 51
7.1.2 Konfiantza-tarteak . . . 51
7.1.3 Normaltasun kontrasteak . . . 53
7.2 Laginketa . . . 54
7.2.1 Laginketa motak . . . 54

7.2.2 Aldagaiak . . . . . 55

**8 Hipotesien kontrasteak** **57**
8.1 Errore motak . . . . . 57
8.2 Hipotesi motak . . . . . 57
8.3 Hipotesiak onartzeko tarteak . . . . . 57

**9 Bariantzaren Analisia (ANOVA)** **61**
9.1 ANOVA taula eraikitzea . . . . . 61
9.2 ANOVA bi faktorerekin . . . . . 62
9.3 Aldakuntza-analisia . . . . . 63

**10 Ariketak** **65**
10.1 Erregresioa eta Korrelazioa . . . . . 65
10.2 Konbinatoria . . . . . 77
10.3 Probabilitatea . . . . . 79
10.4 Ausazko aldagai diskretuen banaketa . . . . . 83
10.5 Ausazko aldagai jarraituen banaketa . . . . . 87
10.6 Laginketa eta Estimazioa . . . . . 91
10.7 Hipotesi-kontrasteak . . . . . 99
10.8 ANOVA . . . . . 103

# 1. Estatistika deskribatzailea

## 1.1 Populazioa eta lagina

Estatistika-azterketa bakoitzeko objektu nagusiek osatutako taldeari **populazioa** edo **unibertsoa** deitzen zaio. Populazio bat ondo definituta egon dadin, elementu zehatzen bat biztanleriarena den ala ez jakin behar da. Gainera, populazioak bi motatakoak izan daitezke: finituak edo infinituak.

$$\Omega = \text{populazioa}$$

**Lagin** bat biztanleriaren azpimultzo adierazgarria da. Laginaren elementu kopuruari laginaren tamaina esaten zaio.

$$\Omega' = \text{lagina}, \text{non } \Omega' \subset \Omega$$

Populazio baten emaitzak fidagarriak izan daitezen, azpipopulazio adierazgarriena aukeratuko dugu lagin gisa.

## 1.2 Ezaugarri motak

Ikerketa batean, interesgarria da laginaren ezaugarriak batera aztertzea. Ezaugarrien balioak zenbakiak edo atributuak izan daitezke; beraz, sailkapen hau izango da:

- **Ezaugarri kualitatiboak:** balioak atributuak direnean.
- **Ezaugarri kuantitatiboak:** balioak zenbakiak direnean. Kategoria hau honela banatzen da:
  - **Ezaugarri diskretuak:** baloreek zenbaki-kopuru finituak hartzen dituztenean.
  - **Ezaugarri jarraituak:** tarte baten barruan edozein balio hartzen denean.

Adibideak:

1. Tailer baten lanbide-kategoria → Ezaugarri kualitatiboa.
2. Erregai motak → Ezaugarri kualitatiboa.
3. Platerak egiteko tailer batean 30 plater hartzen baditugu, bakoitzak zenbat akats dituen jakin nahi dugu. Ezaugarriaren balioak multzo honetan hartuko dira:{0, 1, 2, 3,...}. → Ezaugarri kuantitatibo diskretua.

4. Torloju-tailer batean lagin bat hartzen bada torlojuen diametroa neurtzeko. (Ezaugarriaren balioa tarte baten barruan har daiteke) → Ezaugarri kuantitatibo jarraitua.

## 1.3 Ezaugarri Kuantitatiboak

### 1.3.1 Maiztasun-banaketa

#### Taldekatu gabeko datuak edo datu diskretuak

Demagun $n$ tamainako lagin bat dugula, $n$ dimentsioak ditugula, eta, lan egin aurretik, balio horiek txikienetik handienera ordenatu behar ditugula, eta balio desberdinak baino ez ditugula hartuko $\{x_1, x_2,..., x_m\}$, non $x_1 < x_2 < ... < x_m$.

Balio horiek eta horien maiztasunak, **maiztasun taula**n jarriko ditugu:

| Balioak | Maiztasun absolutua | Maiztasun erlatiboa | Maiztasun absolutu metatua | Maiztasun erlatibo metatua |
|---|---|---|---|---|
| $x_1$ | $n_1$ | $f_1$ | $N_1$ | $F_1$ |
| $x_2$ | $n_2$ | $f_2$ | $N_2$ | $F_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_j$ | $n_j$ | $f_j$ | $N_j$ | $F_j$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_m$ | $n_m$ | $f_m$ | $N_m = n$ | $F_m = 1$ |
| Guztira | $n$ | 1 | | |

Taula 1.1: Taldekatu gabeko datuen maiztasun taula

- **1go zutabea:** Maiztasunen balioak.
- **2. zutabea:** $x_j$ balio bakoitzaren maiztasun absolutua, eta $n_j$ bitartez adieraziko dugu. Hau da, $x_j$ balioa $n_j$ aldiz errepikatzen da. Hauen batura $\sum_{j=1}^{m} n_j = n$.
- **3. zutabea:** Maiztasunen erlatiboak. $f_j = \dfrac{n_j}{n}$ ikurra erabiliko dugu, non batura $\sum_{j=1}^{m} f_j = 1$. Datu hauek lagin desberdinak konparatzeko erabiltzen dira.
- **4. zutabea:** Maiztasun absolutu metatuak, $N_j$. Bere definizioa hurrengoa da: $N_j = \sum_{k=1}^{j} n_k$. Orduan $N_m = n$.
- **5. zutabea:** Maiztasun erlatibo metatuak $F_j$ ikurrarekin adierazten dira, non $F_j = \dfrac{N_j}{n}$, eta gainera, $F_j = \sum_{k=1}^{j} f_k$, orduan $F_m = 1$.

Oharra: Maiztasun erlatiboak portzentai bezala adieraz daiteke.

Adibidea: $f_i$ maiztasun erlatiboa 0,15 bada, $f_i$ bider 100 biderkatzean 15 lortuko dugu. Horrek esan nahi du $x_i$balioa laginaren % 15 dela.

Bestalde, $F_i$ maiztasun metatua 0,46 bada, $F_i$ bider 100, 46 lortuko dugu. Horrek esan nahi du $x_i$ balioa laginaren balio guztien % 46 edo txikiagoa dela.

Laburbilduz:

- Maiztasun erlatiboak $x_i$ balioari dagokion laginaren ehunekoa erakusten du.
- Metatutako maiztasunak laginaren ehunekoa erakusten du, $x_i$ balioa baino txikiagoa edo berdina dena.

### Datu taldekatuak edo jarraituak

Lagineko balio kopurua 30 edo handiagoa denean, datuak klase-tarteetan taldekatzen dira lana errazteko. Orduan, $x_j$-ren ordez, normalean tartearen batez besteko puntua kokatzen da, $k_j$. Balio hori klase-ordezkari izendatzen da.

Taldekatze on bat izateko, zenbait puntu hartu behar dira kontuan:

- Gutxieneko errorea ematen duen klase-tartearen zabalera hautatu behar da.
- Laginaren ibilbidea klase-tarteen zabaleraren multiplo batera biribiltzen da. Laginaren ibilbidea baliorik altuenaren eta baliorik txikienaren arteko aldea izango da.

Adibideak:

1. Ibilbidea $x_m - x_1 = 0,33$ da eta tartearen zabalera $0,05$,beraz $0,02$ batuko diogu ibilbideari: $x_1$ balioaren ordez $x_1 - 0,01$ jarriko dugu eta $x_m$ balioaren ordez $x_m = x_m + 0,01$. Horrela, ibilbide biribildua $0,35$-koa izango da, eta 7 espazioek $0,05$-eko zabalera izango dute. Neurketa bat mutur baten berdina ez izateko, oso komenigarria da tarteen muturrek laginaren tamainak baino hamartar bat gehiago izatea.
2. Aurreko adibidean, ibilbidea $x_1 = 12,20$ eta $x_m = 12,53$ izanik, $x_m - x_1 = 12,54 - 12,19 = 0,35$ hurrengo tarteak izango ditu:

   (12,19 - 12,24); (12,24 - 12,29); (12,29 - 12,34); (12,34 - 12,39);
   (12,39 - 12,44); (12,44 - 12,49); (12,49 - 12,54)

   hauek klase-markak:

$$k_1 = 12,215;\ k_2 = 12,265;\ k_3 = 12,315;\ k_4 = 12,365;\ k_5 = 12,415;$$
$$k_6 = 12,465;\ k_7 = 12,515$$

Hori egin ondoren, maiztasun taula egiten da, aurreko kasuan bezala, baina zutabe bat gehiago izango du.

| Tarteak | Klase-marka | Maiztasun absolutua | Maiztasun erlatiboa | Maiztasun absolutua metatua | Maiztasun erlatiboa metatua |
|---|---|---|---|---|---|
| $(x_1,\ x_2]$ | $k_1$ | $n_1$ | $f_1$ | $N_1$ | $F_1$ |
| $(x_2,\ x_3]$ | $k_2$ | $n_2$ | $f_2$ | $N_2$ | $F_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $(x_j,\ x_{j+1}]$ | $k_j$ | $n_j$ | $f_j$ | $N_j$ | $F_j$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $(x_m,\ x_{m+1}]$ | $k_m$ | $n_m$ | $f_m$ | $N_m = n$ | $F_m = 1$ |
| Guztira | | $n$ | 1 | | |

Taula 1.2: Datu-maiztasun taldekatuen banaketa-taula

- **1go zutabea:** Maiztasun-tarteak $x_i,\ x_{i+1}$.
- **2. zutabea:** Klase-markak $k_j = \dfrac{x_i + x_{i+1}}{2}$ dira eta $n_i,\ (x_i,\ x_{i+1}]$ lagineko balore kopurua tartean. Oro har, balio bakoitza non dagoen argi uzteko, tarteak erdi irekiak izango dira.
- Gainerako zutabeek maiztasun erlatiboak, absolutuak eta metatuak adierazten dituzte, taula 1.1 modu berean .

Gainerako zutabeek maiztasun erlatiboak, absolutuak eta metatuak adierazten dituzte, taulako modu berean.

### 1.3.2 Deskribapen grafikoak

Aztertzen ari garen ezaugarriak nola jokatzen duen tauletan ikus dezakegun arren, datuak errazago ulertuko ditugu deskribapen grafikoen bidez.

#### Maiztasun absolutuko Barra-Diagrama eta Maiztasun-Poligonoak eta Lerro-Diagramak

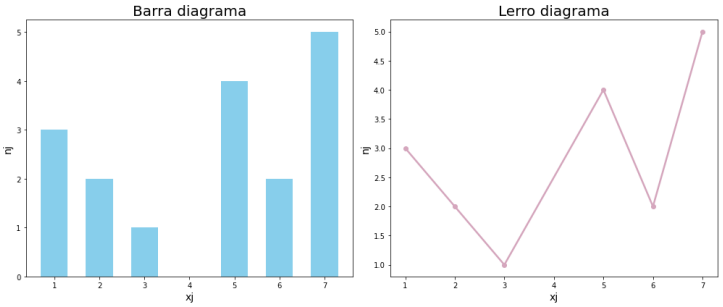
Lagin batean taldekatu gabeko datuak daudenean, bi grafiko kartesiar egin ditzakegu taulako lehen bi zutabeak erabiliz. Ezaugarriaren balioak $OX$ ardatzean kokatzen dira, eta balio bakoitzaren maiztasuna $OY$ ardatzean, $(x_j, n_j)$ puntuak osatuz.

**Barra-diagramak** egiteko $(x_j, 0)$ y $(x_j, n_j)$ puntuak barra baten bitartez lotuko ditugu.

**Maiztasun-Poligonoak** edo **Lerro-Diagramak** egiteko, $(x_1, n_1); ...; (x_m, n_m)$ puntuak zuzen-segmentuen bidez konektatuko ditugu.

Adibidea: Taulako datuetatik abiatuta, egin Barra-Diagrama bat eta Lerro-Diagrama bat.

| $x_j$ | $n_j$ |
|---|---|
| 1 | 3 |
| 2 | 2 |
| 3 | 1 |
| 5 | 4 |
| 6 | 2 |
| 7 | 5 |




Oharra: Grafiko berak egin daitezke maiztasun erlatiboak zein metatuak irudikatzeko.

### Histogramak eta Maiztasun-Poligonoak

Irudikapen grafiko hori datuak taldekatuta daudenean erabiltzen da. **Histograma** delakoen eraikuntzaren interpretazioa tarteen zabaleraren arabera sailkatuko dugu.

1. Tarte guztiek zabalera bera dutenean:

   Ardatz kartesiarretan, oinarri bereko eta altuera desberdinetako laukizuzenak marraztuko ditugu, non oinarria tartearen zabalera izango den eta maiztasuna altuera.

   Tarteak $OX$ ardatzean jarriz eta tartearen maiztasuna $OY$ ardatzean.

2. Tarteek zabalera desberdina dutenean:

   Kasu honetan, aurreko kasuaren antzera egingo da, baina laukizuzen bakoitzaren altuera haren maiztasuna zati zabalera izango da, hau da, $\frac{n_i}{z_i} = h_i$ non $z_i$ tartearen zabalera den.

Oharra: Histogramak maiztasun erlatiboak eta metatuak irudikatzeko ere erabil daitezke.

**Maiztasun-Poligonoak** egiteko, tarte bakoitzean $(k_i, n_i)$ puntuak hartuko ditugu kontuan eta segmentu zuzenekin konektatuko ditugu.

## 1.4 Ezaugarri Kualitatiboak

### 1.4.1 Maiztasun-banaketa

Lagina kualitatiboa denean, datu diskretu gisa funtzionatuko du. Kasu honetan, atributuak edo kualitateak lehen zutabean jarriko ditugu.

| Atributuak | Maiztasun absolutua | Maiztasun erlatiboa | Maiztasun absolutua metatua | Maiztasun erlatiboa metatua |
|---|---|---|---|---|
| 1. Atributua | $n_1$ | $f_1$ | $N_1$ | $F_1$ |
| 2. Atributua | $n_2$ | $f_2$ | $N_2$ | $F_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| j. Atributua | $n_j$ | $f_j$ | $N_j$ | $F_j$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| m. Atributua | $n_m$ | $f_m$ | $N_m = n$ | $F_m = 1$ |
| Total | $n$ | 1 | | |

Taula 1.3: Karateristika kualitatiboetarako maiztasunen banaketa-taula.

- **1go zutabea:** Atributoak.
- Gainerako zutabeek maiztasun erlatiboak, absolutuak eta metatuak adierazten dituzte, 1.1 taulako modu berean.

Gainerako zutabeek maiztasun erlatiboak, absolutuak eta metatuak adierazten dituzte, 1.1 taulako modu berean.

### 1.4.2 Deskribapen grafikoak

Ezaugarri kualitatiboak irudikatzeko, bi grafiko mota ikusiko ditugu:

- Sektore-Diagrama
- Piktograma

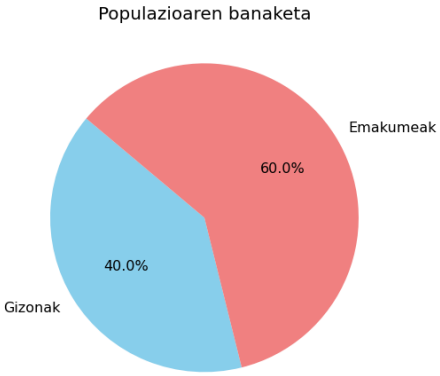
#### Sektore-Diagramak

Adierazpen horretan, kategoria bakoitzaren maiztasuna zirkunferentziaren eremuaren zati batek adierazten du.

Horretarako, angeluak kalkulatuko ditugu hiruko erregela hau erabiliz:

$$\text{Guztira} = \sum_{i=1}^{n} x_i, \;\; \text{Angelua}_i = \frac{x_i}{\text{Guztira}} \; 360^\circ \tag{1.1}$$

Adibidea: 60.000 gizon eta 90.000 emakume diren biztanleriaren Sektore-Diagrama egin.

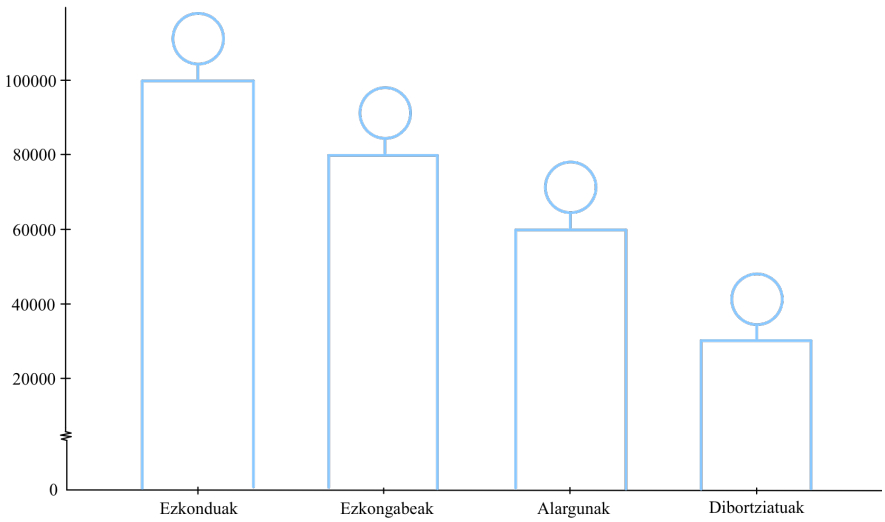



### Piktograma

**Piktogramak** marrazkien bidez egiten dira, marrazkien tamainak kategoriaren maiztasuna adierazten du.

Adibidea: Herri batean zenbat ezkondu, ezkongabe, dibortziatu eta alargun dauden aztertzen ari gara, eta emaitza hauek lortu ditugu:

- Ezkondu: 100.000
- Ezkongabe: 80.000
- Alargun: 60.000
- Dibortziatu 30.000



## 1.5 Lagin kuantitatibo sinpleen estatigrafo nagusiak

Datu-multzo baten ezaugarriak deskribatzeko, hiru estatigrafo mota erabiltzen dira:

- Zentralizazio neurriak.
- Sakabanatze-neurriak.
- Asimetria-neurriak.

### 1.5.1 Zentralizazio neurriak

Estatigrafo horiek laginaren erdiko balioa modan, medianan eta batez bestekoan kokatzeko erabiltzen dira.

#### Moda

Ezaugarririk ohikoena **moda** da. Maiztasun-banaketaren taula hartu eta $M_o$ sinboloarekin adierazten da. Maiztasun bera duten bi balio desberdin badaude, banaketa bimodala izango da. Kasu honetan, bi balioak har daitezke, edo hautaketa-irizpide bat ezar daiteke (ertaina, txikiagoa, handiagoa, ausazkoa...).

Baina datu taldekatuak daudenean, formula honen bidez kalkulatzen da:

$$M_o = L_{i-1} + \frac{n_{i+1}}{n_{i+1} + n_{i-1}} \; z_i \tag{1.2}$$

$L_{i-1}$ beheko muturra da, $n_{i+1}$ ondorengo tartearen maiztasun absolutua da, $n_{i-1}$ aurreko tartearen maiztasun absolutua da eta $z_i$ klasearen zabalera da.

#### Mediana

$Me$ adierazten da, eta ordenatutako datu-multzo batean erdiko posizio-aldagaiaren balioa adierazten du. Balio horretatik beherako balioek lagin osoaren 50% osatzen dute. Hau da, $F_{Me}$ $100 = 50$ denean. Laginak elementu kopuru bikoitia badu, erdiko bi elementuen arteko batez bestekoa da mediana.

Datuak taldekatuta daudenean, aurreko azalpenak mediana zer tartetan dagoen jakiteko balio du, baina balioa kalkulatuko dugu formula hau erabiliz:

$$Me = L_{i-1} + \frac{n/2 - N_{i-1}}{n_i} \; z_i \tag{1.3}$$

$L_{i-1}$ mediana dagoen tartearen mutur txikiena da, $N_{i-1}$ mediana dagoen aurreko tartearen maiztasun absolutu metatua da, $n_i$ mediana dagoen tartearen maiztasun absolutua da eta $z_i$ klasearen anplitudea da.

**batez besteko Aritmetikoa**

**Media aritmetikoa** lagin baten estatigrafo nagusia da, $\overline{x}$ sinboloarekin adierazten da eta formula honen bidez definitzen da:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1.4}$$

Baina laginean behin eta berriz errepikatzen diren balioak daudenean, formula hau erabiltzen da:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i \; n_i}{n} \tag{1.5}$$

Notazioan laginaren eta biztanleriaren batez besteko aritmetikoa bereiziko dugu, hurrenez hurren, $\overline{x}$ eta $\mu$.

Oharrak:

1. Laginak balio desberdin gutxi eta tartekatu gabeak dituenean, taula hauek erabiliko ditugu:

   Adibidea:

| $x_i$ | $n_i$ | $x_i \cdot n_i$ |
|---|---|---|
| 2,3 | 10 | 23 |
| 4,5 | 6 | 27 |
| 1,5 | 4 | 10 |
| Guztira | 20 | 60 |

   Kasu hauetan: $\overline{x} = \frac{60}{20} = 3$

2. Laginaren balioak taldekatuta ez badaude eta oso handiak badira, aldagaia aldatuko da. Lagin-balio bakoitzari "$a$" balio finko bat kenduko zaio emaitzak txikiak izan daitezen, hau da, $x'_i = x_i - a$.

   Orduan, $\sum_{i=1}^{m} x'_i = \sum_{i=1}^{m} x_i - a$ denez, n-rekin zatituz: $\overline{x}' = \overline{x} - a \rightarrow \overline{x} = \overline{x}' + a$.

   Adibidea: Balio finkoa $a = 126$ izanik, $\overline{x} = \frac{-17}{30} = -0.566$ orduan $\overline{x} = -0.566 + 126 = 125.434$.

| $x_i$ | $n_i$ | $f_i$ | $N_i$ | $x'_i = x_i - a$ | $x'_i \cdot n_i$ |
|---|---|---|---|---|---|
| 121 | 5 | | | -5 | -25 |
| 122 | 4 | | | -4 | -16 |
| 123 | 3 | | | -3 | -9 |
| 124 | 2 | | | -2 | -4 |
| 126 | 1 | | | 0 | 0 |
| 127 | 6 | | | 1 | 6 |
| 128 | 4 | | | 2 | 8 |
| 130 | 3 | | | 4 | 12 |
| 131 | 1 | | | 5 | 5 |
| 132 | 1 | | | 6 | 6 |
| Total | 30 | | | | -17 |

3. Laginaren datuak $b$ zabalerako tarte taldekatuak direnean, batez bestekoa kalkulatzeko erabiliko diren balioak klasearen ordezkariak izango dira.

### batez besteko Geometrikoa

**batez besteko geometrikoa** oso gutxitan erabiltzen da; erabileretako bat banku-interesak kalkulatzeko da. $G$ letrarekin idazten da, eta formula hau du:

$$G = \sqrt[n]{x_1^{n_1} \; x_2^{n_2} \; ... \; x_m^{n_m}} \tag{1.6}$$

### batez besteko Harmonikoa

**batez besteko harmonikoa**, aurrekoa bezala, oso arraroa eta gutxi erabiltzen da. $H$ letrarekin idazten da, eta honela adierazten da:

$$H = \frac{n}{\sum \frac{1}{x_i} \; n_i} \tag{1.7}$$

### batez besteko Koadratikoa

**batez besteko Koadratikoa**, desbiderapen estandarra edo RMS (Root Mean Square) ere esaten zaio, eta maiz erabiltzen da estatistikan eta datuen analisian, balio multzo baten aldakortasuna neurtzeko. $Q$ letrarekin idazten da, eta formula hau du:

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2} \tag{1.8}$$

### 1.5.2 Sakabanatze-neurriak

#### Ibilbidea

Baliorik handienaren eta txikienaren artean dago.

$$\text{Ibilbidea} = x_m - x_i \tag{1.9}$$

#### Pertzentilak

Medianaren antzeko kontzeptu bat da, baina datuak erdibitu beharrean, ehunekotan banatzen dituzte. Zehazki, datuak txikienetik handienera ordenatzen badira, k ordenako pertzentila da zure ezkerreko datuen $k\%$ uzten duen balioa. Kasu honetan, $k = 1, 2, 3, ..., 99$ pertzentil izango ditugu.

- Kasu diskretuan, definizioa aplikatzen da.
- Kasu jarraituan, formula hau aplikatu beharko da:

$$P_k = l_i + \frac{\frac{k\ n}{100} - N_{i-1}}{n_i}\ z_i \tag{1.10}$$

Pertzentilak definitzeko moduaren antzera, dezilak eta kuartilak ere definitzen dira.

- Dezilak: Bereziki, $k$ ordenako dezilak distribuzioko datuen %10 bere ezkerrera uzten ditu, non $k = 1, 2, ..., 9$. Hau da: $D_1 = P_{10}$, $D_2 = P_{20}$, ..., $D_9 = P_{90}$, $D_{10} = P_{100}$.
- Kuartilak: Definizioz, $k$ ordenako kuartilak distribuzioko datuen %25 bere ezkerrera uzten ditu, non $k = 1, 2, 3$. Hau da: $Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$, $Q_4 = P_{100}$.

Jakina, bigarren kuartila mediana da: $Q_2 = Me$. Kuartilen kasuan, kuartil arteko tartea definitzen da, hau da, hirugarren eta lehen ordenako kuartilen arteko distantzia: $R_1 = Q_3 - Q_1$.

#### Bariantza eta Desbiderapen estandarra

Laginketa-balioen batez besteko aritmetikoarekiko sakabanatze-maila neurtzen duen estatigrafoari **bariantza** deitzen zaio, eta $s_x^2$ sinboloarekin izendatzen da, lagina bada, edo $\sigma_x^2$ sinboloarekin, populazioarena bada. Biztanleriaren bariantzarako adierazpena hau da:

$$\sigma_x^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} \tag{1.11}$$

Formula garatuz:

$$\sum_{i=1}^{n}(x_i-\overline{x})^2=\sum(x_i^2+\overline{x}^2-2\ x_i\ \overline{x})=\sum x_i^2+n\ \overline{x}^2-2\ \overline{x}\ \sum x_i=$$

$$=\sum x_i^2+n\ \overline{x}^2-2\ \overline{x}\ n\ \overline{x}=\sum x_i^2-n\ \overline{x}^2=\sum x_i^2-\frac{(\sum x_i)^2}{n}$$

Onduan $s_x^2=\frac{1}{n}\ [\sum x_i^2-\frac{1}{n}\ (\sum x_i)^2]=\frac{1}{n}\ \sum x_i^2-(\overline{x})^2$

**Desbiderapen estandarra** bariantzaren erro karratua baino ez da:

$$\sigma_x=\sqrt{\frac{\sum(x_i-\overline{x})^2}{n}}=\sqrt{\frac{1}{n}\ \sum\left(x_i^2-\overline{x}\right)^2} \tag{1.12}$$

Laginak datu errepikatuak dituenean, maiztasun absolutuak formula hauetan sartzen dira:

$$\sigma_x^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2\ n_i}{n}=\frac{1}{n}\ \left[\sum x_i^2\ n_i-\frac{1}{n}\ \left(\sum x_i\ n_i\right)^2\right] \tag{1.13}$$

Balio txikiak dispertsio gutxi dagoela adierazten du, baina balio altuak dispertsio handia dagoela.

Aurretik aipatutako definizio eta formula guztietan biztanleriaren datuak aipatzen dira. **Bessel-en zuzenketa** k *(n - 1)*-ren ordez *n*-ren artean zatitzea dakar laginketa-bariantzaren eta laginaren desbiderapen estandarraren formuletan, lagin bat denez gutxienez elementu bat faltako baita. Horrela, gerta daitezkeen alborapenak zuzentzen dira.

Ikus ditzagun laginaren bariantza eta desbiderapen estandarrari dagozkion adierazpenak:

$$s_x^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1} \tag{1.14}$$

$$s_x=\sqrt{\frac{\sum(x_i-\overline{x})^2}{n-1}}=\sqrt{\frac{1}{n-1}\ \sum x_i^2-(\overline{x})^2} \tag{1.15}$$

### 1.5.3 Asimetria-neurriak

Datuak aldagaiaren ibilbidearen zati batean kontzentratuta dauden zehazteko erabiltzen dira. Hau da, batez besteko aritmetikoarekiko simetrikoak diren ala ez.

### Asimetria

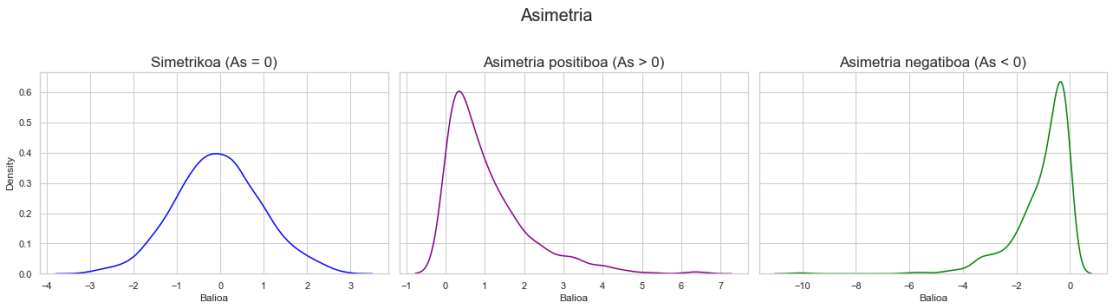
**Asimetria** da aldagai batek batez besteko aritmetikoarekiko duen banaketaren simetria adierazten duen neurria, irudikapen grafiko bat egin beharrik gabe.

batez bestekoaren eta modaren zuzeneko konparazioa egitea, $\overline{x} - M_o$, ez da oso fidagarria. Beraz, Pearson-en koefizientea erabiltzen da, non:

$$A_s = \frac{\overline{x} - M_o}{s_x}$$

$A_s$ balioa kontuan hartuta, hau dugu:

- $A_s = 0$ bada, lagina simetrikoa da.
- $A_s < 0$ bada, laginak asimetria du ezkerretik edo asimetria negatiboa.
- $A_s > 0$ bada, laginak asimetria du eskuinetik edo asimetria positiboa.




Irudia 1.1: Asimetria-neurriak. Asimetria.

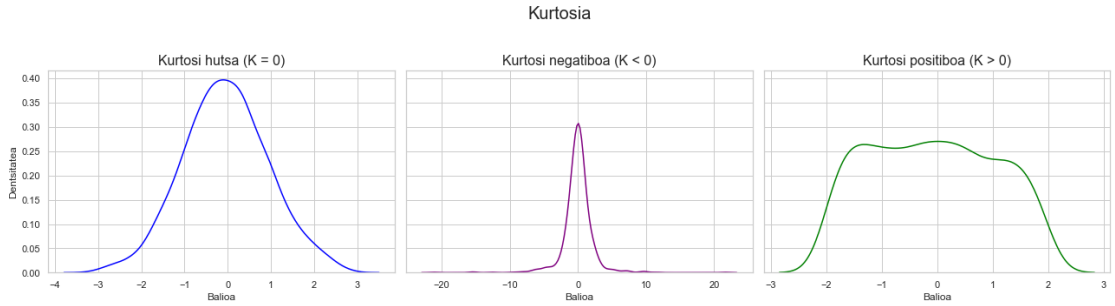
### Kurtosia

**Kurtosia** banaketa bat banaketa normalarekiko txikitzearen edo zorroztearen neurria da. Honela adierazten du:

$$K = \frac{\sum \frac{(x_i - \overline{x})^4 \, n_i}{n}}{s_x^4} - 3 \tag{1.16}$$

$K$ balioaren arabera, sailkapen hau izango dugu:

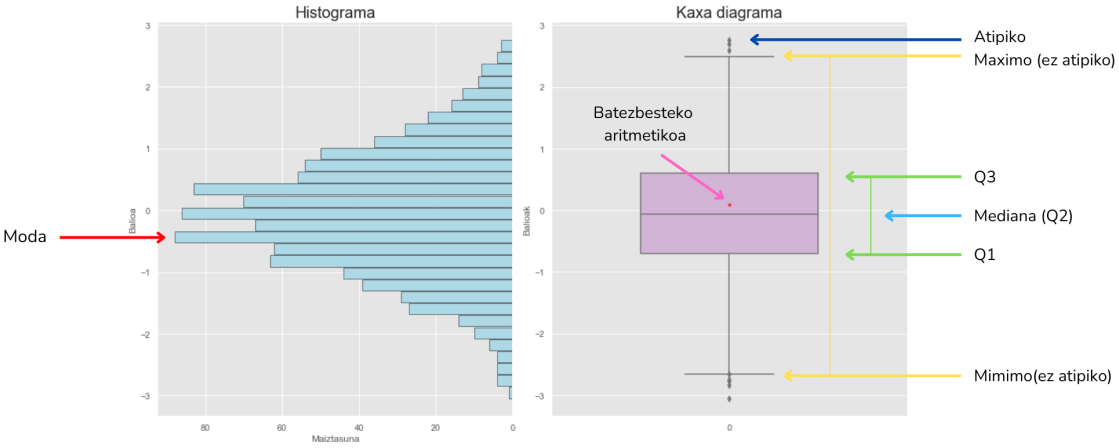
- $K = 0$ bada, kurtosi nulua du eta lagina mesokurtikoa dela esaten da.
- $K < 0$ bada, kurtosi negatiboa du eta lagina platikurikoa da..
- $K > 0$ bada, kurtosi positiboa du eta lagina leptokurikoa da. .




Irudia 1.2: Asimetria-neurriak. Kurtosia.

### 1.5.4 Kutxa-Diagramak

**Kutxa-Diagramek** zenbakizko aldagai baterako datuen banaketa erakusten dute. Datu horiek zentroa eta datuen hedapena ikusten laguntzen dute. Ikusizko tresna gisa ere erabil daitezke, normaltasuna egiaztatzeko edo ezohiko balioak izan daitezkeen puntuak identifikatzeko.




Irudia 1.3: Kutxa-Diagrama.

Jarraian, Kutxa-Diagrama baten oinarrizko zatiak azalduko ditugu:

- Kutxaren erdiko lerroak datuen mediana adierazten du. Datuak simetrikoak badira, kaxaren erdian egongo da mediana.
- Kutxaren goiko eta beheko muturrek 25 eta 75 pertzentilak adierazten dituzte. Kaxaren luzera bi pertzentil horien arteko aldea da, eta kuartilen arteko tartea (IQR) esaten zaio.

- Kutxatik zabaltzen diren lerroei biboteak deitzen zaie. Bibote horiek datuek espero duten bariantza adierazten dute.
- Biboteen muturren azpitik edo gainetik geratzen diren datuak puntuz adierazten dira, eta puntu horiek balio atipikoak dira.
- Izartxoak datu multzoaren batez besteko aritmetikoaren balioa adierazten du.

# 2. Erregresioa eta korrelazioa

**Erregresioa** eta **korrelazioa** populazio berean definitutako ausazko bi aldagai erlazionatzeko erabiltzen diren teknikak dira. $(X, Y)$ gisa idatziko ditugu bi dimentsioko ausazko aldagai horiek.

Abidideak:

1. Eibarko pertsonen altuerari eta pisuari buruzko datuak hartuko bagenitu, nolabait lotuta leudeke.
2. Haurren altuera eta adina ere lotuta egongo lirateke.
3. Finantzetan, adibidez, Telefonicaren eta Euskaltelen eguneko prezioak ere nolabait lotuta egongo dira.

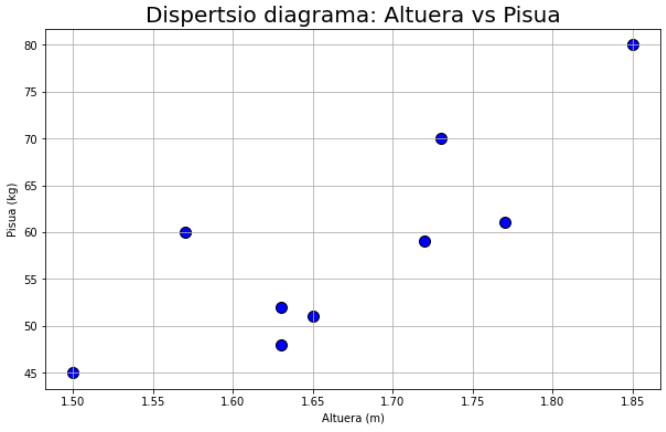
## 2.1 Sakabanatze eta erregresio-diagramak

$(X, Y)$ bi dimentsioko ausazko aldagai bat grafikoki irudikatzeko, sakabanatze-diagramak erabiltzen dira. Horietan, aldagaietako bat x ardatzean kokatuko da, eta bestea, berriz, y ardatzean, eta X bakoitzari Y bat egokituko zaio.

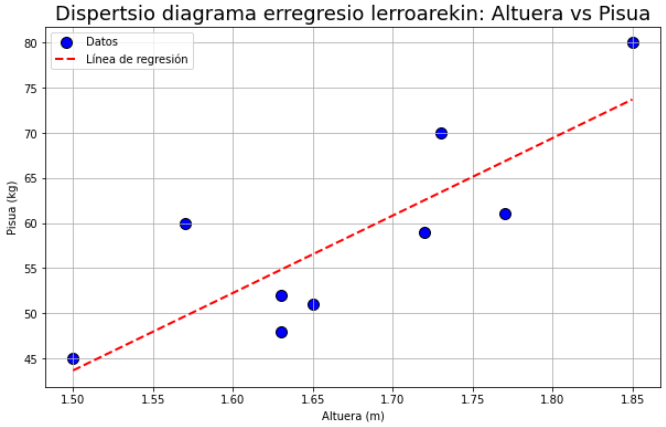
Adibidea: Demagun Eibarko 10 emakumeri altuera eta pisua galdetzen dizkiegula eta datu hauek ematen dizkigutela::

| Pertsona | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| X = altuera | 1,65 | 1,85 | 1,72 | 1,57 | 1,73 | 1,77 | 1,63 | 1,63 | 1,50 |
| Y = pisua | 51 | 80 | 59 | 60 | 70 | 61 | 48 | 52 | 45 |

Orduan, datu horien sakabanatze-diagrama bat honela ikusiko litzateke:



Eta erregresio zuzena, puntu hauek hobekien laburbilduko lituzkeena, honako hau litzateke:




## 2.2 Erregresioa: Zer da?

Kasu honetan, bi aldagaiak ekuazio zehatz baten bidez erlazionatzen dira. Horrela, bietako bat, gutxi gorabehera, bestearentzako hurbilketa bat da.

Erregresioarekin hasi aurretik, definizio batzuk ikusiko ditugu:

$$S(x) = \sum x; \quad S(xy) = \sum xy; \quad S(y^2) = \sum y^2; \quad cov(x,y) = \frac{1}{n} \cdot \sum x \cdot y - \overline{x} \cdot \overline{y}$$

### 2.2.1 Erregresio motak

Erregresioak forma asko izan ditzake:

- $y = a + b \cdot x$
- $y = a + b \cdot x + c \cdot x^2$
- $y = a \cdot e^{b \cdot x}$
- $y = a \cdot x^b$

Horrela, x eta y egiazkoetatik abiatuta, y ($\hat{y}$)-ren balio zenbatetsiak aurki daitezke.

### 2.2.2 Parametroak kalkulatzea

Parametro guztiak kalkulatzeko metodo on bat dagoela esan genezake, karratu minimoen metodoa deritzona. Metodo honek $\sum(y_i - \hat{y}_i)^2$ ahalik eta txikiena izateko parametroak aurkitzen ditu.

## 2.3 Korrelazioa: Zer da?

Horrela, koefiziente bakarra erabiltzen da bi aldagai erlazionatzeko: $r$ korrelazio-koefizientea. $r$ balioa [-1, 1] tartean egongo da, eta haren balio absolutua 1-etik zenbat eta gertuago egon, orduan eta handiagoa izango da bi aldagaien arteko erlazioa:

$$r = \frac{conv(x,y)}{s_x \cdot s_y} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y})}{s_x \cdot s_y} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \overline{x}) \cdot (y - \overline{y}) \cdot f(x,y) \cdot dx \cdot dy}{s_x \cdot s_y} \tag{2.1}$$

Oharra: Kasu honetan, badakigu lotura dagoen ala ez bi aldagaien artean, baina ez dakigu bata bestearen arabera aurkitzen.

**Kobariantza** honela kalkula daiteke:

$$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i - \overline{x} \cdot \overline{y} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f(x,y) \cdot dx \cdot dy - \overline{x} \cdot \overline{y} \tag{2.2}$$

## 2.4 Erregresio-parametroak aurkitu

Lehen esan bezala, hurbilketa eta benetako balioa ahalik eta antzekoenak izatea eragingo duten balioak bilatuko ditugu.

### 2.4.1 Erregresio lineala

$y = a + b \cdot x$ adierazpena izanik, $\sum(y - a - b \cdot x)^2$ minimoa egiten duten $a$ eta $b$ parametroak aurkituko ditugu.

Deribatuak eginez:

$$\begin{cases} \sum 2 \cdot (y - a - b \cdot x) \cdot (-1) = 0 \\ \sum 2 \cdot (y - a - b \cdot x) \cdot (-x) = 0 \end{cases} \rightarrow \begin{cases} \sum (y - a - b \cdot x) = 0 \\ \sum (y \cdot x - a \cdot x - b \cdot x^2) = 0 \end{cases}$$

Eta koefizienteak hurrengoak izango dira:

$$b = \frac{cov(x,y)}{s_x^2}; \quad a = \overline{y} - b \cdot \overline{x} \tag{2.3}$$

### 2.4.2 Koefiziente horien eta korrelazioaren arteko erlazioa

$$r_z = \frac{s_x}{s_y} \cdot b \tag{2.4}$$

**Estandarizazioa**

Zer gertatzen da $x' = \dfrac{x + x_o}{u}$ y $y' = \dfrac{y + y_o}{v}$ badugu?

Orduan,

- $r(x, y) = r(x', y')$
- $b(x, y) = \dfrac{u}{v} \cdot b(x', y')$

**Estimazio errorea**

Zenbatespen-errorea, $y$ balio errealaren eta $\widehat{y}$ estimatuaren arteko bariantza gisa kalkulatzen da.

$$s_{ey} = \sqrt{\frac{\sum y_i^2 - a \cdot \sum y_i - b \cdot \sum x_i \cdot y_i}{n - 2}} \tag{2.5}$$

### 2.4.3 Erregresio esponentziala

$y = a \cdot e^{b \cdot x}$ adierazpena izanik, logaritmoak aplikatuko ditugu:

$$ln\, y = ln(a \cdot e^{b \cdot x}) = ln\, a + ln(e^{b \cdot x}) = ln\, a + b \cdot x$$

Horrela, aurreko kasuan genuen ekuazioaren antzekoa izango dugu, eta, beraz,

$$b = \frac{cov(x, ln\, y)}{s_x^2};\ \ a = e^{\overline{s} - b \cdot \overline{x}}\ \text{ non }\ s = ln\, y$$

Korrelazio-koefizienteak adierazpen hau du:

$$r_e = \frac{s_x}{s_{ln\, y}} \cdot b$$

### 2.4.4 Erregresio potentziala

$y = a \cdot x^b$ adierazpena izanik, logaritmoak aplikatuko ditugu:

$$ln\, y = ln(a \cdot x^b) = ln\, a + ln(x^b) = ln\, a + b \cdot (ln\, x)$$

Horrela, aurreko kasuan genuen ekuazioaren antzekoa dugu, eta, beraz,

$$b = \frac{cov(ln\, x, ln\, y)}{s_{ln\, x}^2};\ \ a = e^{\overline{s} - b \cdot \overline{t}}\ \text{ non }\ s = ln\, y\ \text{ eta }\ t = ln\, x$$

Korrelazio-koefizienteak adierazpen hau du:

$$r_p = \frac{s_{ln\, x}}{s_{ln\, y}} \cdot b$$

### 2.4.5 Erregresio parabolikoa

$y = a + b \cdot x + c \cdot x^2$ adierazpena izanik, hiru parametro daudenez, ezin da parametroen balio zehatzik eman eman. Kasu bakoitzean ekuazio-sistema hau ebatzi beharko da:

$$\begin{cases} a \cdot n + b \cdot \sum_i x_i + c \cdot \sum_i x_i^2 = \sum_i y_i \\ a \cdot \sum_i x_i + b \cdot \sum_i x_i^2 + c \cdot \sum_i x_i^3 = \sum_i x_i \cdot y_i \\ a \cdot \sum_i x_i^2 + b \cdot \sum_i x_i^3 + c \cdot \sum_i x_i^4 = \sum_i x_i^2 \cdot y_i \end{cases}$$

Kasu horretan, errorea honela zenbatetsiko da:

$$s_{ey} = \sqrt{\frac{\sum y^2 - a \cdot \sum y - b \cdot \sum x \cdot y - c \cdot \sum x^2 \cdot y}{n}}$$

# 3. Konbinatoria

Konbinatoriak zehaztutako irizpide jakin batzuk betetzen dituzten multzo bateko elementuen konfigurazioak zenbatzeko metodoak aztertzen ditu.

Konbinazio-arazo guztietan, funtsezko kontzeptu batzuk bereizi behar dira:

- **Populazioa** Aztertzen ari diren elementuen multzoa. Multzo honetako elementu kopurua $\boldsymbol{n}$ izango da. Horrela, A multzoak $\{a_1, ..., a_n\}$ elementuak dituela esango dugu, eta hurrengo eran idatziko dugu $A = \{a_1, a_2, ..., a_n\}$
- **Lagina** Biztanleriaren azpimultzoa da eta $\boldsymbol{m}$ da lagina osatzen duten elementuen kopurua. Bi azpimultzo desberdinak izango dira, gutxienez elementu desberdin bat badute. Adibidez: $\{a_1, a_2\} \neq \{a_1, a_3\}$. Gainera, laginaren ezaugarriak hauek dira:
  - **Ordena** Lagineko elementuak ordenatuta agertzea garrantzitsua den ala ez. Orduan:
    - Ordena garrantzitsua bada, orduan $\{a_1, a_2\} \neq \{a_2, a_1\}$ eta, beraz, desberdinak dira.
    - Ordena EZ bada garrantzitsua, orduan $\{a_1, a_2\} = \{a_2, a_1\}$ eta, beraz, berdintzat jotzen dira.
  - **errepikapenak** Elementuak errepikatzeko edo ez errepikatzeko aukera.

Azken batean, bi azpimultzo desberdinak dira gutxienez elementu desberdin bat badute edo ordena garrantzitsua denean elementuen ordena desberdina bada.

## 3.1 Aldakuntzak

Aldaketak multzo bateko elementuak ordena kontuan hartuta taldekatzeko moduak dira.

### 3.1.1 Errepikapen gabe

$m$-naka hartutako $n$ elementuren **errepikatuz gabeko aldakuntzak**, $m$ elementu desberdinekin osatutako multzoak dira, eta ditugun $n$ elementuen artean aukeratzen dira. Beste aldakuntza dela esango dugu, baldin eta:

- $m \leq n$, nahiz eta $m = n$ kasua permutazioetarako baztertu.
- bi azpimultzo desberdinak dira, baldin eta gutxienez elementu desberdin bat badute edo, elementuen ordena desberdina bada.

Eraiki daitezkeen aldakuntzen kopurua formula honen bidez kalkula daiteke:

$$\mathrm{V}_{\mathrm{n}}^{\mathrm{m}} = \frac{n!}{(n-m)!} \tag{3.1}$$

Adibidea: $A = \{\mathrm{a}_1, \mathrm{a}_2, \mathrm{a}_3\}$ multzo batetatik, elementuak 2-naka aukeratzen dira. Zenbat talde egin daitezke?

Guztira 3 elementu ditugu, beraz $n$ balioa 3 da. Eta elementuak 2-ko taldetan aukeratzen dira; beraz,$m$ balioa 2-koa da. Bi parametro horien balioak kontuan hartuta, konbinazio hauek egin daitezke:

$$\mathrm{V}_3^2 = \frac{3!}{(3-2)!} = \frac{3!}{1!} = 6$$

Konbinazio horiek eskuz egiten baditugu, kalkulua ondo dagoela ikusiko dugu:

$$\begin{array}{ll} \{\mathrm{a}_1, \mathrm{a}_2\} & \{\mathrm{a}_2, \mathrm{a}_1\} \\ \{\mathrm{a}_1, \mathrm{a}_3\} & \{\mathrm{a}_3, \mathrm{a}_1\} \\ \{\mathrm{a}_2, \mathrm{a}_3\} & \{\mathrm{a}_3, \mathrm{a}_2\} \end{array}$$

### 3.1.2 Errepikapenekin

$m$-naka hartutako $n$ elementuren **errepikatuzko aldakuntzak**, $m$ errepikatu daitezken $m$ elementuk osatutako multzoak dira, eta ditugun $n$ elementuen artean aukeratzen dira. Bi aldakuntza desberdinak izango dira, elementuren batean desberdinak badira edo ordena desberdinean kokatuta badaude. Eraiki daitezkeen aldakuntzen kopurua formula honen bidez kalkula daiteke:

$$\mathrm{VR}_{\mathrm{n}}^{\mathrm{m}} = n^m \tag{3.2}$$

Adibidea: $A = \{1,\ 2,\ 3,\ 4,\ 5\}$ multzo bat izanik, 5 zifrako zenbakiak sortu nahi dira. Errepikatu ahal direlarik.

Guztira 5 elementu ditugu, beraz $n$ balioa 5 da. Eta elementuak 5-eko taldetan aukeratzen dira; beraz,$m$-ren balioa 5-ekoa da. Bi parametro horien balioak eta bilatzen diren zenbakien zifrak kontuan hartuta, egin daitezkeen errepikatuzko aldakuntzak hurrengoak dira:

$$\mathrm{VR}_5^5 = 5^5 = 3125$$

## 3.2 Konbinazioak

Konbinazioak multzo bateko elementuak taldekatzeko moduak dira, honako hauek kontuan hartuta:

- EZ du eraginik jartzen diren ordenan.

### 3.2.1 Errepikapen gabe

$m$-naka hartutako $n$ elementuren **errepikatuz gabeko konbinazioak**, $m$ elementu desberdinekin osatutako multzoak dira, eta ditugun $n$ elementuen artean aukeratzen dira. Bi multzo desberdinak izango dira, baldin eta elementuren batean desberdinak badira. (Ez du eraginik elementuen kokapen-ordenan) Eraiki daitezkeen konbinazioen kopurua formula honen bidez kalkula daiteke:

$$\mathrm{C}_{\mathrm{n}}^{\mathrm{m}} = \binom{n}{m} = \frac{n!}{(n-m)!\; m!} \tag{3.3}$$

Adibidea: $A = \{\mathrm{a}_1, \mathrm{a}_2, \mathrm{a}_3\}$ multzotik elementuak 2-naka aukeratzen dira. Zenbat konbinazio izango ditugu?

Guztira 3 elementu ditugu, beraz $n$ balioa 3 da. Elementuak 2-ko taldetan aukeratzen dira; beraz, $m$ -ren balioa 2-koa da. Bi parametro horien balioak kontuan hartuta, konbinazio hauek egin daitezke:

$$\mathrm{C}_3^2 = \frac{3!}{(3-2)!\; 2!} = \frac{3!}{2!} = 3$$

Konbinazio horiek eskuz egiten baditugu, kalkulua ondo dagoela ikusiko dugu:

$$\{\mathrm{a}_1, \mathrm{a}_2\}$$
$$\{\mathrm{a}_1, \mathrm{a}_3\}$$
$$\{\mathrm{a}_2, \mathrm{a}_3\}$$

### 3.2.2 Errepikapenekin

$m$-naka hartutako $n$ elementuren **errepikatuzko konbinazioak**, $m$ errepikatu daitezken $m$ elementuk osatutako multzoak dira, eta ditugun $n$ elementuen artean aukeratzen dira. Bi konbinazio desberdinak direla esango dugu, baldin eta elementuren batean diferenteak badira. (Ez du eraginik elementuak jartzeko ordenan). Eraiki daitezkeen konbinazioen kopurua formula honen bidez kalkula daiteke:

$$\mathrm{CR}_{\mathrm{n}}^{\mathrm{m}} = \binom{n+m-1}{m} = \frac{(n+m-1)!}{(n-1)!\; m!} \tag{3.4}$$

Adibidea: Gozotegi batean 5 pastel mota daude, eta horietako 4 aukeratzen uzten digute. Zenbat pastel konbinazio hauta daitezke?

Guztira 5 elementu ditugu, beraz $n$ balioa 5 da. Eta elementuak 4-ko taldetan aukeratzen dira; beraz,$m$ balioa 4-koa da. Bi parametro horien balioak kontuan hartuta, konbinazio hauek egin daitezke:

$$\mathrm{CR}_5^4 = \binom{5+4-1}{4} = \frac{(5+4-1)!}{(5-1)!\ 4!} = 70$$

## 3.3 Permutazioak

Permutazioak, edo ordenazioak ere esaten zaie, multzo baten elementuak taldekatzeko moduak dira, honako hauek kontuan hartuta:

- Zer ordenatan jartzen diren, horrek eragina du.
- Dauden elementu guztiak hartuko ditugu.

### 3.3.1 Errepikapen gabe

$n$ elementuko **Errepikapen gabeko permutazioak** horiek guztiak ordenatzeko modu guztiak dira. Hauen arteko desberdintasun bakarra elementuen kokapen-ordena da. Permutazio horien kopurua honako hau izango da:

$$\mathrm{P_n} = n! \tag{3.5}$$

Adibidea: Zenbat modutan eser daitezke 3 pertsona 3 besaulkitan?

$$\mathrm{P}_3 = 3! = 6 \tag{3.6}$$

### 3.3.2 Errepikapenekin

$n$ elementuko **Errepikapenezko permutazioak** horiek guztiak ordenatzeko modu guztiak dira. Kasu honetan zenbaki errepikatuak izango ditugu, hau da, $A = \{a, a, a, ..., a, b, b, b, \ ..., b, c, c, c, ..., c\}$. Permutazio horien kopurua honako hau izango da:

$$\mathrm{PR}_{\mathrm{n}}^{\alpha,\beta,\gamma} = \frac{n!}{\alpha!\ \beta!\ \gamma!} \tag{3.7}$$

Non:

- $\alpha$: a agertzen den kopurua.

- $\beta$: b agertzen den kopurua.
- $\gamma$: c agertzen den kopurua.

Adibidea: Zenbat zenbaki eraiki daitezke hurrengo multzoko zifra guztiak erabiliz? $A = \{4, 4, 4, 3, 3, 5\}$?,

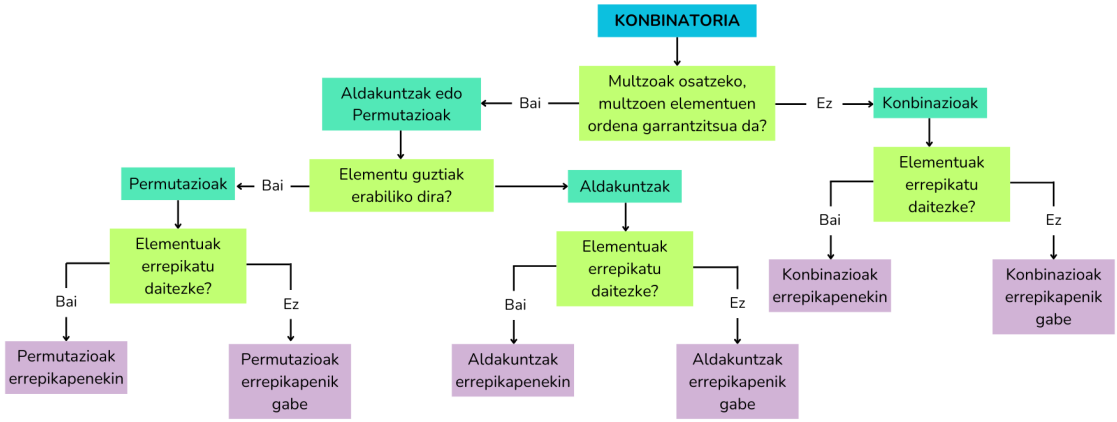
Guztira 6 elementu ditugu, beraz $n$ balioa 6 da. Eta parametroen balioetarako $\alpha$,$\beta$ eta $\gamma$ elementuak zenbat aldiz errepikatzen diren ikusi behar da:

- $\alpha = 3$, 4 zenbakia 3 aldiz errepikatzen da.
- $\beta = 2$, 3 zenbakia 2 aldiz errepikatzen da.
- $\gamma = 1$, 5 zenbakia 1 aldiz errepikatzen da.

$$\mathrm{PR}_6^{3,2,1} = \frac{6!}{3!\ 2!\ 1!} = 60$$

## 3.4 Eskema

Eskema honek adierazten du zer urrats egin behar diren zer konbinatoria mota dagoen ikusi ahal izateko.




Irudia 3.1: Eskema konbinatoria.

## 3.5 Zenbaki konbinatorioak

Zenbaki konbinatorioak edo koefiziente binomialak $\binom{n}{m}$ formako zenbakiak dira, $m$ eta $n$ osoko ez-negatiboak $m \leq n$ izanik. Lehen esan bezala, formula honen bidez kalkulatzen dira:

$$\binom{n}{m} = \frac{n\ (n-1)\ (n-2)...(n-m+1)}{m!} = \frac{n!}{m!\ (n-m)!} \tag{3.8}$$

Adierazten dute:

- $m$-naka hartutako $n$ elementuan errepikapen gabeko konbinazioak.
- $n$ elementuko multzo batek dituen $m$ elementuko multzo kopurua.

Oinarrizko propietateak:

1. $\binom{n}{m} = \binom{n}{n-m}$
2. $\binom{n}{m} = \binom{n-1}{m-1} + \binom{n-1}{m}$
3. $\binom{n}{0} = \binom{n}{n} = 1$
4. $\binom{n}{1} = n$

# 4. Probabilitatea

**Probabilitatea** zorizkoak edo ausazkoak diren gertakizunak azaltzeko erabiltzen da. Probabilitatearen diziplinak hainbat gertakarirekin lotutako aukerak eta probabilitateak kuantifikatzeko metodoak ematen ditu.

## 4.1 Definizioak

- **Ausazko saiakuntza:**
  - Era infinituan errepika daitezke antzeko esperimentuak.
  - Esperimentu bakoitzean lor daitezkeen emaitzek unibertsoa edo lagin-espazioa osatzen dute.
  - Esperimentu bat egin aurretik ezin da jakin zer emaitza lortuko den (ausazkotasun-baldintza).
  - Esperimentuen kopurua handitzen denean, emaitza bakoitzaren maiztasun erlatiboa (edo lagin-espazioaren puntu bakoitza) egonkortzeko joera dago.
- **Ausazko gertaera:** lagin-espazioaren azpimultzoa.

Adibideak:

1. Dado batean zenbaki bikoitiak ateratzeko probabilitatea.

$$\text{Ausazko saiakuntza} \rightarrow E = \{1, 2, 3, 4, 5, 6\}$$
$$\text{Ausazko azpimultzoa} \rightarrow S = \{2, 4, 6\}$$

2. Ausazko esperimentua: bi dado botatzea. Ausazko gertaera: "dadoen batura 7 da".

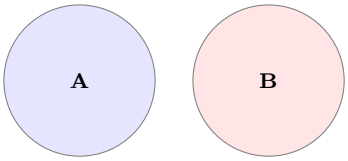
| | | | | | |
|---|---|---|---|---|---|
| 1, 1 | 1, 2 | 1, 3 | 1, 4 | 1, 5 | 1, 6 |
| 2, 1 | 2, 2 | 2, 3 | 2, 4 | 2, 5 | 2, 6 |
| 3, 1 | 3, 2 | 3, 3 | 3, 4 | 3, 5 | 3, 6 |
| 4, 1 | 4, 2 | 4, 3 | 4, 4 | 4, 5 | 4, 6 |
| 5, 1 | 5, 2 | 5, 3 | 5, 4 | 4, 5 | 5, 6 |
| 6, 1 | 6, 2 | 6, 3 | 6, 4 | 6, 5 | 6, 6 |

$$\text{N} = 36 = 6^2$$

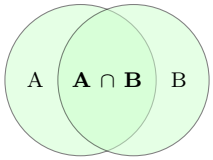
$$\{(1,6), (2,5), (3,4), (5,2), (6,1)\}$$

- **Eragiketak:**

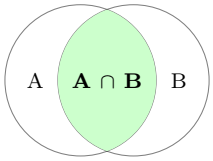
- Gertaerak letra larriz adierazten dira (A, B, C,...)



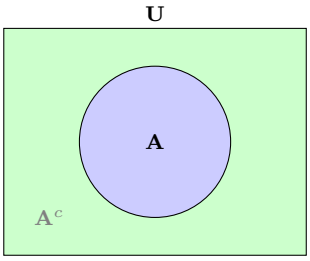

- Elementuak gertaerei dagozkie (a ∈ A, b ∈ B, ...)
- Gertaeren batura: $A \cup B = \{x \in S \mid x \in A \vee x \in B\}$




- Gertaeren ebakidua: $A \cap B = \{x \in S \mid x \in A \wedge x \in B\}$




- Osagarria: $\overline{A} = A^c = \{x \in S \mid x \notin A\}$




- **Oharrak:**
  - Gertaera bat ezinezkoa denean adierazten duen multzoa "hutsa"da. P $(\emptyset) = 0$
  - Gertaera bat gertatzen denean, beti deitzen da "gertaera segurua". P $= 100\%$
  - Gertaera bat ezinezkoa edo segurua ez denean, "gertaera probablea".
- **Oharra:** Batura eta ebakidura trukakorrak dira. Hau da, $A \cap B = B \cap A$, baita, $A \cup B = B \cup A$

Adibideak:

1. Bi dadoren jaurtiketa eta bien produktuak $\geq$40 izan behar du.
   Ezinezko gertaera. Ezinezkoa da $\geq$40 ateratzea, bi datuen arteko produktuaren gehienezko balioa 36-koa baita.

2. Bi dadoren jaurtiketa eta bien batura $\leq$12 izatea.
   Gertaera segurua. Dadoen balioen arteko batura honako hauen artean egongo baita: $1+1 \leq x \leq 6+6 = 2 \leq x \leq 12$.

3. Dadoak botatzea eta batura 8 izatea:

$$\mathrm{A} = \{(2,6),(4,4),(3,5),(5,3),(6,2)\}$$

   Dadoak ateratzea eta produktua 12 izatea:

$$\mathrm{B} = \{(2,6),(3,4),(4,3),(6,2)\}$$

   - $A \cup B$: $\{(2,6),(4,4),(3,5),(5,3),(6,2),(3,4),(4,3)\}$
   - $A \cap B$: $\{(2,6),(6,2)\}$
   - A gertakarian guztira dauden 36 aukeretatik 5 ditugu. Beraz, $\overline{A}$-n 36 - 5 = 31 aukera izango ditugu.

## 4.2 Probabilitatea

**Laplaceren erregela:** ausazko esperimentu baten emaitza guztiak ekiprobableak badira, Laplacek A gertaera baten probabilitatea honela definitzen du: esperimentuan A gertaera gertatzearen aldeko emaitzen kopuruaren eta esperimentuaren emaitza posibleen kopuruaren arteko zatidura. Horrela, formula honen bidez adieraz daiteke:

$$P(A) = \frac{\text{A-ren aldeko kasuak}}{\text{Kasu posible guztiak}} \tag{4.1}$$

**Propietateak**:

- $0 \leq P(A) \leq 1$
- $\mathrm{P(E)} = 1$
- $\mathrm{P}(\emptyset) = 0$
- Gertaera bateragarriak: Bi gertaerak batera gerta daitezkeenean, hau da, oinarrizko gertaera komunen bat dutenean.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Gertaera bateraezinak: aldi berean bi gertaera edo gehiago gerta ezin daitezkeenean, hau da, gertaera komunik ez dutenean.

$$P(A \cup B) = P(A) + P(B)$$

- $\bigcup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} P(A)$
- $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) + P(A \cap C) - P(A \cap B \cap C)$

Adibideak:

1. Ikasle batek matematika gainditzeko probabilitatea $2/3$ da, ingelesa gainditzeko probabilitatea $4/9$ da eta biak gainditzeko probabilitatea $1/2$ da. Zein da gutxienez bat gainditzeko probabilitatea?

$$P(M \cup I) = P(M) + P(I) - P(M \cap I) = 2/3 + 4/9 - 1/2 = 31/36$$

2. Bi dado jaurtitzean, 7 edo 11 lortzeko probabilitatea.

   - "7 *lortu*" $\rightarrow A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$
   - "11 *lortu*" $\rightarrow B = \{(5,6), (6,5)\}$

$$P(A) = \frac{6}{36} = \frac{1}{6}\,; \quad P(B) = \frac{2}{36} = \frac{1}{18}$$

$A \cup B = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1), (5,6), (6,5)\}$
$A \cap B = \emptyset$
$P(A \cup B) = P(A) + P(B) = 1/6 + 1/18 = 2/9$
$P(A \cap B) = 0$

## 4.3 Baldintzazko probabilitate

**Probabilitate baldintzatua** gertaera bat gertatzeko aukera da, A deitzen dioguna, beste gertaera bat gertatu delako (B deitzen diogu). Formula honen bidez adieraz daiteke:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{4.2}$$

Formula hau irakur daiteke: A gertatzeko probabilitatea, B gertatu denez, A eta B gertatzearen probabilitatearen berdina da, aldi berean, B-ren probabilitatearen artean.

Adibidea:

1. Dado batean gertaera hauek daude:

$$\text{“zenbaki karratu bat lortzea”} \rightarrow A = \{1, 4\}$$
$$\text{“zenbaki bat 3 baino handiagoa lortzea”} \rightarrow B = \{4, 5, 6\}$$
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}; \quad P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

- $A \cap B : \{4\}$
- $P(A) = \frac{2}{6} = \frac{1}{3}; \quad P(B) = \frac{3}{6} = \frac{1}{2}; \quad P(A \cap B) = \frac{1}{6}$

$$P(A \mid B) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{2}{6} = \frac{1}{3}; \quad P(B \mid A) = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{3}{6} = \frac{1}{2}$$

2. Hegazkin bat orduka ateratzeko probabilitatea 0,83 da, orduan iristeko probabilitatea 0,82 da eta hegazkina irten eta orduan iristeko probabilitatea 0,78 da.

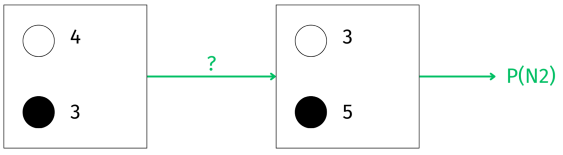
$$P(S) = 0.83; \quad P(Ll) = 0.82; \quad P(S \cap Ll) = 0.78$$
$$P(Ll \mid S) = \frac{P(S \cap Ll)}{P(S)}; \quad P(S \mid Ll) = \frac{P(S \cap Ll)}{P(Ll)}$$
$$P(Ll \mid S) = \frac{0.78)}{0.83} = 0.94; \quad P(S \mid Ll) = \frac{0.78}{0.82} = 0.95$$

## 4.4 Biderketa-arauak

Izan bitez A eta B gertaerak:

$$\begin{cases} P(A \cap B) = P(A) \cdot P(B \mid A) \\ P(B \cap A) = P(B) \cdot P(A \mid B) \end{cases} \rightarrow P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$$

Adibidea: Bi kaxa ditugu, lehenengoan 4 bola zuri eta 3 bola beltz. Lehenengo kaxatik bola bat atera eta bigarren kutxan sartuko dugu. Bertan, hasiera batean 3 bola zuri eta 5 bola beltz zeuden. Zein da bigarren kaxatik bola bat ateratzean beltza izateko probabilitatea?




$$P(N_1 \cap N_2) = P(N_1) \cdot P(N_2 \mid N_1) = \frac{3}{7} \cdot \frac{6}{9} = \frac{18}{63}$$
$$P(B_1 \cap N_2) = P(B_1) \cdot P(N_2 \mid B_1) = \frac{4}{7} \cdot \frac{5}{9} = \frac{20}{63}$$
$$P(N_2) = P(N_1 \cap N_2) + P(B_1 \cap N_2) = \frac{18}{63} + \frac{20}{63} = \frac{38}{63}$$

Bi gertaera aske A y B $\iff$ $P(A \cap B) = P(A) \cdot P(B)$ *o* $P(A \mid B) = P(A)$ *y* $P(B \mid A) = P(B)$

Adibidea: Txanpon bat jaurti, aurpegia atera, dadoa jaurti eta zenbaki bikoitia atera.

Gertaera horiek ez dute inolako loturarik, eta, beraz, gertaera askeak dira.

## 4.5 Probabilitate totala

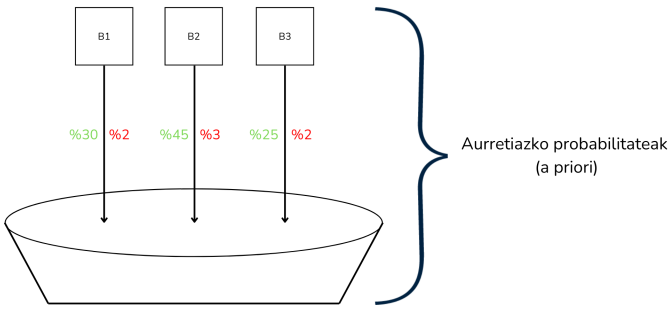
Egoera batzuetan, hainbat gertaerek parte hartzen dute lagin-espazio bereko beste gertaeraren bat egiten. Izan bitez $B_1$, $B_2$,..., $B_n$gertaerak:

Demagun S espazioaren banaketa bat dela, hau da, gertaerak askeak direla eta $S = \bigcup_{i=1}^{n} B_i$ *o* $S = B_1 \cup ... \cup B_i \cup ... \cup B_n$.

Formula honen bidez, S (A∈S) -ko A gertaeraren probabilitatea kalkula daiteke, $B_1$, $B_2$,..., $B_n$ gertaeren probabilitate-balioak ezagututa.

$$P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A \mid B_i) \tag{4.3}$$

Adibidea: Tailer batean piezen 3 makina ditugu $B_1$, $B_2$ eta $B_3$, hurrenez hurren % 30 % 45 eta % 25 ekoizten dutenak. Makina horiek pieza akastunen % 2 % 3 eta % 2 ekoizten dute, hurrenez hurren. Pieza bat ausaz hartzen badugu, zein da akastuna izateko probabilitatea?




Gertaera hauek identifikatuko dira:

- A: "piezak akatsa du"
- $B_1$: "1go makinako pieza"

- $B_2$: "2. makinako pieza"
- $B_3$: "3. makinako pieza"

$$P(A) = P(B_1) \cdot P(A \mid B_1) + P(B_2) \cdot P(A \mid B_2) + P(B_3) \cdot P(A \mid B_3) = \\ 0.3 \cdot 0.02 + 0.45 \cdot 0.03 + 0.25 \cdot 0.02 = 0.0245 \rightarrow 2.45\%$$

## 4.6 Bayes-en teorema

Izan bitez $B_1, B_2, ..., B_i$ elkarrekiko gertakari baztertzaileak eta E lagin-espazioa dutenak, hau da, $B_1 \cup B_2 \cup ... \cup B_n = E$. A beste gertaera bat bada, orduan:

$$P(B_i \mid A) = \frac{P(B_i) \cdot (P(A \mid B_i))}{\sum_{k=1}^{n} P(B_k) \cdot P(A \mid B_k)} = \frac{P(B_i \cap A)}{P(A)} \tag{4.4}$$

Adibidea: Aurreko adibidean (3 makinena). Zein da akastun pieza $B_2$ makinatik ateratzeko probabilitatea ?

$P(A) = 0.0245$

$P(B_2 \mid A) = \frac{P(B_2) \cdot P(A|B_i)}{\sum_{k=1}^{n} P(B_k) \cdot P(A|B_k)} = \frac{0.45 \cdot 0.03}{0.0245} = 0.538 \rightarrow 53.8\%$

# 5. Ausazko Aldagai Diskretuen Banaketa

Aurretik gaietan, S lagin-espazio baten gertaeren probabilitatea kalkulatzen ikasi dugu. Unitate honetan, S elementuen eta zenbaki errealen arteko korrespondentziak ezartzeko arauak aztertuko ditugu, ondoren probabilitate-balio bat esleitzeko.

## 5.1 Ausazko aldagaia

Lagin-espazioko puntu bakoitzari zenbaki erreal bat esleitzen dion funtzioari*aldagai aleatorio* deitzen zaio.

Ausazko aldagaia letra larriz idazten da, eta haren balioak letra xehez. Orduan$X$ aldagai aleatorio bat da eta $x$ aldagaiaren balio bat da.

$$\text{X} = \text{x; x zenbaki erreal bat izanik}$$

## 5.2 Ausazko aldagai diskretua

- Ausazko aldagai bat diskretua da, hartu ditzakeen balioak zenbakigarri mugatuak edo infinituak badira.
- Zer EZ litzateke aldagai diskretu bat? Donostiara iristeko denboraren probabilitatea kalkulatzea, denbora, balio jarraitua denez, ez bailitzateke balio mugatua edo infinitu zenbakigarria izango.

### Probabilitate-funtzio

Laginaren puntuen probabilitateak ezagutzen direnean, ausazko aldagai diskretuaren balio bakoitzari probabilitatea esleitzen dion funtzioari **probabilitate-funtzioa** deitzen zaio.

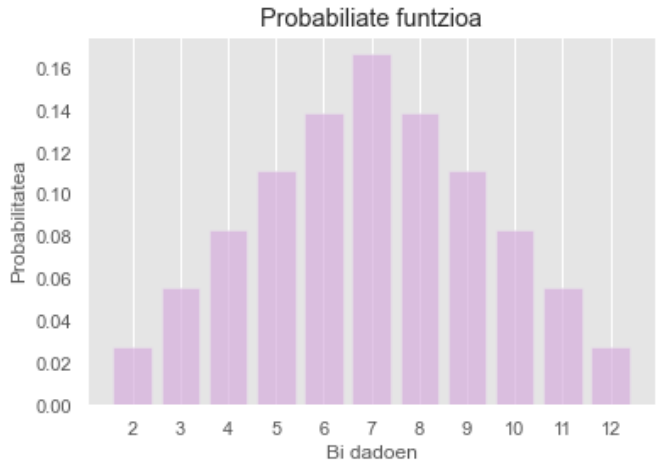
$$P(X = x) = P_i \tag{5.1}$$

Adibidea: Bi dado jaurtitzeko probabilitate-funtzioa:

- $x = 2 \rightarrow A_2 = \{(1,1)\} \rightarrow P(A_2) = \frac{1}{36}$
- $x = 3 \rightarrow A_3 = \{(1,2),(2,1)\} \rightarrow P(A_3) = \frac{2}{36} = \frac{1}{18}$
- $x = 4 \rightarrow A_4 = \{(1,3),(2,2),(3,1)\} \rightarrow P(A_4) = \frac{3}{36} = \frac{1}{12}$

- $x = 5 \rightarrow A_5 = \{(1,4),(2,3),(3,2),(4,1)\} \rightarrow P(A_5) = \frac{4}{36} = \frac{1}{9}$
- $x = 6 \rightarrow A_6 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\} \rightarrow P(A_6) = \frac{5}{36}$
- $x = 7 \rightarrow A_7 = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \rightarrow P(A_7) = \frac{6}{36} = \frac{1}{6}$
- $x = 8 \rightarrow A_8 = \{(2,6),(3,5),(4,4),(5,3),(6,2)\} \rightarrow P(A_8) = \frac{5}{36}$
- $x = 9 \rightarrow A_9 = \{(3,6),(4,5),(5,4),(6,3)\} \rightarrow P(A_9) = \frac{4}{36} = \frac{1}{9}$
- $x = 10 \rightarrow A_{10} = \{(4,6),(5,5),(6,4)\} \rightarrow P(A_10) = \frac{3}{36} = \frac{1}{12}$
- $x = 11 \rightarrow A_{11} = \{(5,6),(6,5)\} \rightarrow P(A_{11}) = \frac{2}{36} = \frac{1}{18}$
- $x = 12 \rightarrow A_{12} = \{(6,6)\} \rightarrow P(A_{12}) = \frac{1}{36}$

| $x_i$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $P(X = x_i)$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |




### Banaketa-funtzioa

Ausazko aldagai baten banaketa-funtzioa probabilitate metatuaren funtzioa da.

$$F(x) = P(X \leq x) = \sum_{i=1}^{j} P(x = i) \tag{5.2}$$

Adibidea: Kalkulatu aurreko adibideko F(4).

$$F(4) = P(x \leq 4) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4) =$$
$$= 0 + 0 + \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36} = \frac{1}{6}$$

Propietateak:

- $F(x_i) = P(x_1) + P(x_2) + P(x_3) + ... + P(x_i) = F(X_{i-1} + P(x_i)$
- $P(x_i \leq x \leq x_j) = P(x \leq x_j) - P(x < x_i) = F(x_j) - F(x_{i-1})$
- $P(x_i < x < x_j) = P(x < x_j) - P(x \leq x_i) = F(x_{j-1}) - F(x_i)$
- $P(x_i < x \leq x_j) = P(x \leq x_j) - P(x \leq x_i) = F(x_j) - F(x_i)$
- $P(x_i \leq x < x_j) = P(x < x_j) - P(x < x_i) = F(x_{j-1}) - F(x_{i-1})$

## 5.3 Aldagai diskretuen banaketen batez bestekoa eta bariantza

### batez bestekoa

Dakigunez, banaketa diskretu baten batez besteko aritmetikoa $\overline{X} = \sum_{i=1}^{k} x_i \cdot \frac{n_i}{n}$ da, eta $x_i$ eta $n_i$ aldagaiaren balioak maiztasun absolutua dira.

Baina lagina oso handia denean, adibidez, $n \to \infty$, $\frac{n_i}{n} \to P(x_i)$ orduan $\lim_{x\to\infty} \overline{x} = \lim_{n\to\infty} \sum x_i \cdot \frac{n_i}{n} = \sum x \cdot P(x_i)$ betetzen da, eta adierazpen horri ausazko aldagaiaren batez bestekoa edo itxaropen matematikoa deitzen zaio, eta $E(X)$ edo $mu_x$ zeinuaren bidez adierazten da. Laburbilduz:

$$E(x) = \sum_{i=1}^{k} x_i \cdot P(x_i) \tag{5.3}$$

Adibidea: Aurreko dadoen adibidean:

$E(x) = 2 \cdot P(2) + 3 \cdot P(3) + 4 \cdot P(4) + 5 \cdot P(5) + 6 \cdot P(6) + 7 \cdot P(7) + 8 \cdot P(8) + 9 \cdot P(9) + 10 \cdot P(10) + 11 \cdot P(11) + 12 \cdot P(12) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = \frac{6}{36} = \frac{1}{6}$

### Bariantza eta desbiderapen tipikoa

Lehen egin dugun bezala, banaketa diskretuaren bariantza $s_x^2 = \frac{\sum (x_i - \overline{x})^2 \cdot n_i}{n}$ da. Orduan, $n \to \infty$ $\overline{X} \to \mu_x$ eta $\frac{n_i}{n} \to P(x_1)$ betetzen direnean, aurreko adierazpenaren mugari ausazko aldagaiaren bariantza deitzen zaio eta $\sigma_x^2$ edo $Var(x)$.

$Var(x) = \lim_{x\to\infty} s_x^2 = \sum_{i=1}^{k} (x_i - \overline{x})^2 \cdot P(x_i)$ eta karratua garatuz, hurrengoa dugu: $Var(x) = \sum_{i=1}^{k} x_i^2 \cdot P(x_i) - \overline{x}^2$.

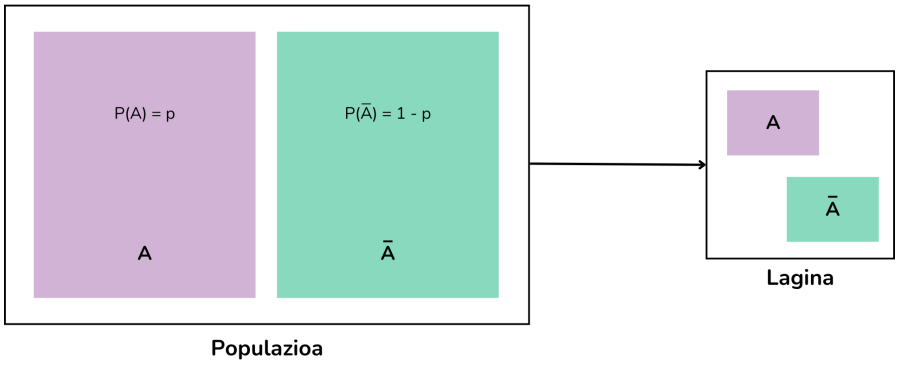
## 5.4 Ausazko aldagai diskretuen banaketa

### 5.4.1 Banaketa hipergeometrikoa

Banaketa hipergeometrikoaren ezaugarriak:

- N elementuko populazio mugatua dugu.
- Elementu guztiak bi mota baztertzailetan sailkatuta daude: A y $\overline{A}$.
- $P(A) = p; \quad P(\overline{A}) = 1 - p$
- n tamainako laginak konfiguratzen ditugu biztanleria finitua oinarri hartuta. **ORDEZKAPENIK GABE**

  Oharra: Lagin bat ordezkapenik gabe hartzeak esan nahi du elementuak banan-banan hartzen direla, itzulketarik gabe. Beraz, ondorioztatu dezakegu saiakuntzak jada ezin direla independentetzat jo, arrakastaren probabilitatea elementu berri bakoitza hartzean aurreko saiakuntzen emaitzak eragiten duelako, populazioaren elementu kopurua saiakuntza bakoitzean aldatzen delako.




Irudia 5.1: Populazio batetik ateratako lagina.

X (aldagai aleatorioa): laginean dauden A motako elementuen kopurua.

$$X = \{x \mid max[0, n - N(1-p)] \leq x \leq min[n, Np]\} \tag{5.4}$$

$$P(X=x)=\frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}};\quad E(x)=np;\quad Var(x)=\frac{N-n}{N-1}np(1-p)\quad (5.5)$$

Adibidea: Kutxa batek 100 alkandora ditu, 60 onak eta 40 txarrak. 70 alkandorako lagin bat konfiguratzen dut.

$$A:\text{"}camisa\ buena\text{"};\quad \overline{A}:\text{"}camisa\ mala\text{"};\quad N=100;\quad A=60;\quad \overline{A}=40;\quad n=70$$

a. Zenbat alkandora on eduki ditzake laginak?

$$X=\{x\mid max[0,70-100\cdot(1-0.6)]\leq x\leq min[70,100\cdot 0.6]\}\rightarrow 30\leq x\leq 60$$

b. Laginean 40 alkandora on izateko probabilitatea.

$$P(x=40)=\frac{\binom{100\cdot 0.6}{40}\binom{100(1-0.6)}{70-40}}{\binom{100}{70}};$$

$$P(x=40\ buenas)=P(x=30\ malas)$$

### 5.4.2 Banaketa binomiala

Banaketa hori oso garrantzitsua da, eta maiz erabiltzen da. Ausazko aldagaiaren interesa saiakuntza ezberdinetan lortzen diren arrakasta kopuruarekin lotuta dago.

Propietateak:

- n saiakera berdin-berdinak egiten dira..
- Saiakera bakoitzean bi emaitza posible baino ez daude: arrakasta eta porrota.
- Arrakastaren probabilitatea $p$ bada, porrotaren probabilitatea*1 - $p$* da. Hau da, $P=p$ eta $P(porrota)=1-p$.
- Probabilitateak beti dira konstanteak, ez dira saiakuntza batetik bestera aldatzen.
- Entsegu guztiak libreak dira (lehenengoan porrota atera bada, ez du baldintzatzen bigarrenean arrakasta edo porrota ateratzea).

X ausazko gertaera definitzen da: "n saiakeretako arrakasta-kopurua";
Izate eremua: $\{0, ..., n\}$

$$X = B(n, p) \tag{5.6}$$

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad \sum P(X = x) = 1 \tag{5.7}$$

$$E(x) = np; \quad Var(x) = np(1-p) \tag{5.8}$$

Banaketa binomial motak:

- P(X = $x_i$) → Binomial puntuala
- P(X ≤ $x_i$) → Binomial metatua → $P(X \leq x_i) = 1 - P(X \geq x_i)$,horrela definitze da $X = B(n, p, x)$

Adibidea: "X: Lau jaurtiketatan hiru alde lortu"

$$P(X = 3) = \binom{4}{3} 0,5^3 (1-0,5)^{4-3}; \tag{5.9}$$

### 5.4.3 Poisson-en banaketa

Poissonen banaketa zehaztutako eskualde batean edo denbora-tartean batean, lortuko liratekeen arrakasten kopuruari dagokion probabilitatea kalkulatzeko erabil daiteke, gertatzen diren arrakasten batez besteko kopurua ezagutzen bada.

X gertakariren probabilitatea denbora jakin batean, gertaeraren probabilitatea m denean.

$$\text{m} \to \infty \qquad \text{p} \to 0$$

$$P(X = x) = \frac{m^x e^{-m}}{x!}; \quad m = np \tag{5.10}$$

$$E(x) = Var(x) = m = np \tag{5.11}$$

### 5.4.4 Banaketa multinazionala/polinomiala

Banaketa binomialaren orokortze bat da, proba bakoitzean n gertaera baztertzaile ditugu $A_1, A_2, ..., A_n$, eta hauen probabilitateak $p_1, p_2, ..., p_n$, eta guztien batura unitatearen berdina da.

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = \frac{n!}{x_1!\ x_2!\ ...\ x_n!}\ p_1^{x_1} p_2^{x_2} ... p_n^{x_n} = \frac{n!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} p_i^{x_i} \tag{5.12}$$

### 5.4.5 Banaketa binomial negatiboa

Probabilitate-eredu horrek eredu binomialaren antzeko ezaugarriak ditu:

- Saiakuntza independenteak, saiakuntza bakoitzak bi emaitza posible baino ez ditu.
- P(arrakasta) = p; P(porrota) = 1 - p
- X: "arrakasta lortu arteko saiakuntza-kopurua"

$$P(X = x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r;\ \ E(x) = \frac{r(1-p)}{p};\ \ Var(x) = \frac{r(1-p)}{p^2} \tag{5.13}$$

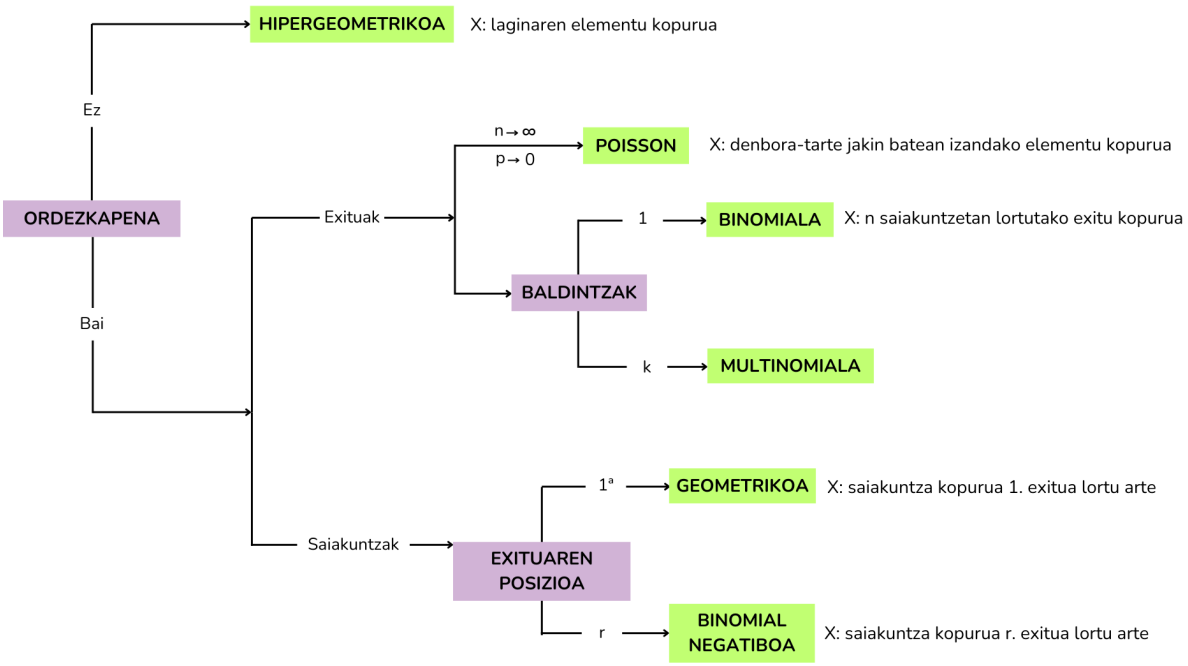
### 5.4.6 Banaketa geometrikoa

Banaketa binomial negatiboaren kasu berezi bat da, $r = 1$ denean. Hau da, lehenengo arrakasta lortu arte egiten diren saiakuntza kopuruarekiko probabilitatea ezagutu nahi da.

- Saiakera independenteak edo askeak.
- p(arrakasta) = p; p(porrota) = 1 - p
- X: "saiakuntza-kopurua, lehen arrakasta lortu arte"

$$P(X = x) = (1-p)^{x-1} \cdot p;\ \ E(x) = \frac{1}{p};\ \ Var(x) = \frac{1-p}{p^2} \tag{5.14}$$

## 5.5 Laburpen-Eskema




Irudia 5.2: Eskema. Ausazko aldagai diskretuaren banaketa.

# 6. Ausazko aldagai jarraituen banaketa

Gogoratu zer zen aldagai jarraitu bat: zuzen errealean edozein balio har dezakeen aldagai bat. Beraz, banaketa jarraitua aldagai jarraitu baten probabilitateak kalkulatzeko erabiltzen den banaketa izango da.

## 6.1 Banaketa-funtzioa

X ausazko aldagai jarraitu bat da, eta haren balioak txikienetik handienera ordenatuta daudela uste dugu. X aldagaiaren banaketa-funtzioa deituko diogu, eta *F (x)* idatziko diogu funtzio honi:

- $F(x) = P(X \leq x)$
- $P(x_1 \leq x \leq x_2) = F(x_2) - F(x_1)$
- $P(x > x_2) = 1 - P(x \leq x_2)$

## 6.2 Dentsitate-funtzioa

Gogoan izan probabilitateak interesatzen zaizkigula, baina oraindik ez dugu probabilitate-funtziorik definitu. Aldagai jarraituetan ezin dugu probabilitate-funtzioa erabili, aldeko kasua bakarra baita eta kasu posibleak infinituak baitira. $P(X = a) = \frac{1}{\infty} = 0$.

**Dentsitate-funtzioak** ausazko aldagai horrek balio jakin bat hartzeko probabilitate erlatiboa deskribatzen du. $f(x)$ gisa definitzen da, banaketa-funtzioaren deribatu gisa:

$$\text{f(x)} = \frac{dF(x)}{dx} \rightarrow F(x) = \int_{-\infty}^{x} f(x)dx \tag{6.1}$$

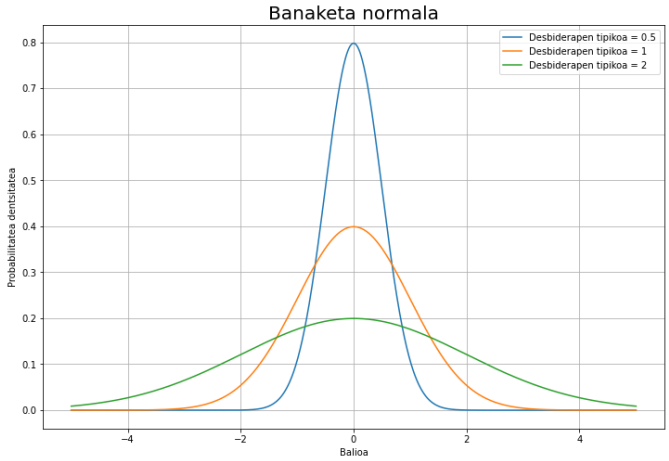
Jarraian, dentsitate-funtzioaren propietateren bat erakutsiko da:

- $P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x)dx$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $E(x) = \mu = \int_{-\infty}^{\infty} xf(x)dx$
- $Var(x) = \sigma_x^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$

## 6.3 Banaketak

### 6.3.1 Banaketa normala

X ausazko aldagai jarraitu batek batez bestekoa $\mu$ eta desbiderapen tipikoko $\sigma$ banaketa normal bati jarraitzen dio, eta $\mathrm{N}(\mu, \sigma)$ bidez izendatzen da. Banaketa mota horien ezaugarria da batez bestekoak, modak eta medianak balio bera dutela, eta $\mu$ izena duela.




Irudia 6.1: Banaketa normala.

#### Funciones

La función de densidad de esta distribucuón y, en consecuencia, la función de distribución son las siguientes:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \rightarrow F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \tag{6.2}$$

Banaketa hori simetrikoa denez,

$$P(-z < Z < z) = F(z) - F(-z) = F(z) - [1 - F(z)] = 2F(z) - 1$$

#### Aldagaiaren tipifikazioa

Hala ere, banaketa arrunt guztien artean ezagunena $N(0, 1)$ da. Distribuzio normala da, batez bestekoa 0 eta desbiderapen tipikoa 1 dituena, eta normal estandarra deituko diogu. Gainera, erabiliko diren taulak banaketa estandarrekoak izango dira.

Banaketa normal estandarrik ez dagoenean, hau da, batez bestekoa ez denean 0 eta desbideratzea ez denean 1, aldagaia tipifikatu behar da banaketa hori edukitzeko.

$$N(\mu \neq 0,\ \sigma \neq 1) \rightarrow N(\mu = 0,\ \sigma = 1)$$
$$P(X < x) \rightarrow P(Z < z)$$

Horretarako, N($\mu, \sigma$) banaketa jarraitzen duen X aldagaia N(0, 1) banaketa jarraitzen duen Z aldagai bihurtu behar dugu.

$$F_{\mu,\,\sigma}(x) = P(X \leq x_i) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x_i-\mu}{\sigma}\right) = P\left(Z \leq \frac{x_i-\mu}{\sigma}\right) = F_{0,\,1}\left(\frac{x-\mu}{\sigma}\right)$$

$$P(x_1 \leq X \leq x_2) = P\left(\frac{x_1-\mu}{\sigma} \leq Z \leq \frac{x_2-\mu}{\sigma}\right)$$

**Moivre-ren teorema**

Izan bedin X aldagai binomial bat non $n$ handia den. Beraz, X aldagai binomiala aldagai normal baten bidez hurbil daiteke. Hau da:

$$B \sim N(\mu,\ \sigma);\ X = N(np,\ \sqrt{np(1-p)}) \tag{6.3}$$

Aldagaia tipifikatzeko orduan, honela geratuko zaigu:

$$z = \frac{x-np}{\sqrt{np(1-p)}} \tag{6.4}$$

Puntu baten probabilitatea kalkulatu nahi badugu, $P(X = x_i)$, Z aldagai normala denez, puntu baten probabilitatea zero izango da. Problema hori gainditzeko, $P(X = x_i) = P(x_i - 0.5 \leq x \leq x_i + 0.5)$ kalkulatuko dugu.

$$B \sim N(\mu,\ \sigma) \rightarrow P(X = x_i) = P(x_i - 0.5 \leq Y \leq x_i + 0.5) \tag{6.5}$$

Hona hemen:

$$P(X = x_i) = P(x_i{-}0.5 \leq Y \leq x_i{+}0.5) = P\left(\frac{x_i - 0.5 - np}{\sqrt{np(1-p)}} < Z < \frac{x_i + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

- $P(x_1 \leq X \leq x_2) = P(x_1 - 0.5 < Y < x_2 + 0.5)$
- $P(x_1 < X < x_2) = P(x_1 + 0.5 \leq Y \leq x_2 - 0.5)$

Adibideak: Izan bedi $X$ ausazko aldagai binomial bat, non $X = B(16, 0.5)$. Orain, taula binomialak eta Moivreren teorema erabiliz, probabilitate hauek kalkulatuko ditugu:

1. $P(X = 8);\ X = B(16;\ 0.5)$

$$P(X = 8) = P(8 - 0.5 < Y < 8 + 0.5) = P(7.5 < Y < 8.5)$$

Aldagaia tipifikatuko dugu:

$$P\left(\frac{7.5 - 160.5}{\sqrt{160.5(1 - 0.5)}} \leq Z \leq \frac{8.5 - 160.5}{\sqrt{160.5(1 - 0.5)}}\right) = P\left(-0.25 \leq Z \leq 0.25\right) =$$

$$= 1 - 2 \cdot P(z \geq 0.25) = 1 - 2 \cdot 0.4013 = 0.1974$$

2. $P(X < 6);\ X = B(16,\ 0.5)$

$$P(X < 6{+}0.5) = P(Y \leq 6.5) \sim P\left(Z < \frac{6.5 - 16 \cdot 0.5}{\sqrt{16 \cdot 0.5 \cdot (1 - 0.5)}}\right) = P(Z < -0.75) =$$

$$= P(Z > 0.75) = 0.2266$$

3. $P(6 < X < 10);\ X = B(16,\ 0.5)$

$$P(6 < X < 10) = P(6 - 0.5 \leq Y \leq 10 + 0.5) = P(5.5 \leq Y \leq 10.5) \sim$$

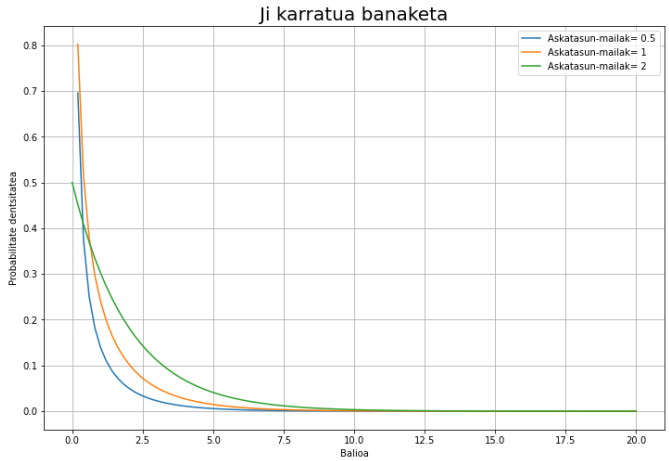
$$\sim P\left(\frac{5.5 - 16 \cdot 0.5}{\sqrt{16 \cdot 0.5 \cdot (1 - 0.5)}} < Z < \frac{10.5 - 16 \cdot 0.5}{\sqrt{16 \cdot 0.5 \cdot (1 - 0.5)}}\right) = P(-1.25 < Z < 1.25) =$$

$$= 1 - 2 \cdot P(Z > 1.25) = 1 - 2 \cdot 0.1056 = 0.7888$$

### 6.3.2 Pearsonen Chi-karratua banaketa

Pearsonen Chi-karratua banaketa funtzioa, $\chi^2$, aldagai normalen $n$ karratuen batura gisa sortzen da. Karratuen batura denez, beti izango da positiboa. Izan bitez $z_1,\ z_2,\ ...,\ z_n$ ausazko aldagai independenteak, guztiak banaketa normal estandarrarekin, orduan $z_1^2 + z_2^2 + ... + z_n^2$ ausazko aldagaiak Pearsonen Chi-karratu izeneko banaketa jarraitzen du $n$ askatasun-mailarekin, eta $\chi_n^2$ bidez adierazten da.

$$\chi_n^2 = z_1^2 + z_2^2 + ... + z_n^2 \quad (z_i \sim N(0,\ 1)) \tag{6.6}$$



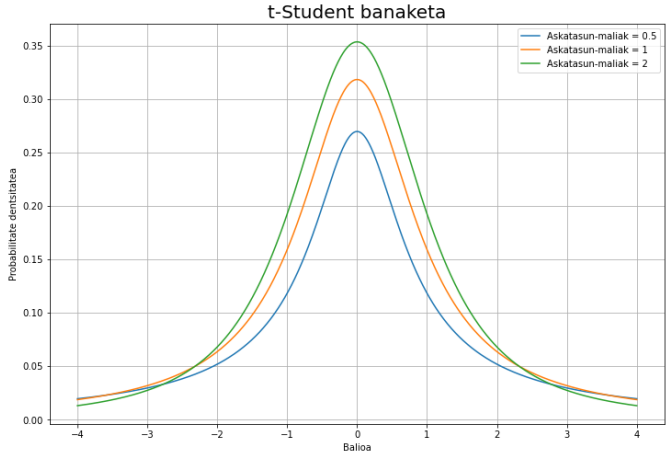

Irudia 6.2: Chi-karratua banaketa.

Kasu honetan parametro bakar baten mende dago, $n$: $\chi^2_n$ izenekoa. Hori ere taulen bidez kalkulatuko da.

### 6.3.3 Student-en T banaketa

$z$ -k banaketa normal estandarra jarraitzen badu eta $x^2_n$ z-rekiko independentea bada, hurrengo ausazko aldagaiak **t de Student** izeneko banaketa jarraitzen du $n$ askatasun-mailarekin, eta honela definitzen da $t_n$.

$$t_n = \frac{z \cdot \sqrt{n}}{\sqrt{x^2_n}} \tag{6.7}$$




Irudia 6.3: Student-en T banaketa.

Banaketa normalaren antzeko forma du, baina punta gehiago du. Banaketa

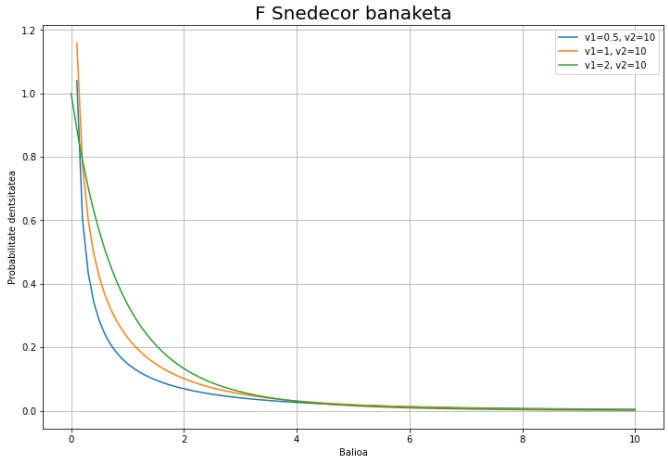
hori ere simetrikoa da, eta $(-\infty, \infty)$ tartean dago definituta. Tamaina txikiko laginetarako erabilgarria izaten da.

### 6.3.4 Snedecor-en F banaketa

Izan bitez $\chi^2_{v_1}$ eta $\chi^2_{v_2}$ bi Chi-karratu banaketa $v_1$ eta $v_2$ askatasun-gradurekin. Fisher-Snedecor banaketa honela sortzen da:

$$X^2_{v_1},\ X^2_{v_2};\quad F_{v_1;\,v_2} = \frac{X^2_{v_1}/v_1}{X^2_{v_2}/v_2} \tag{6.8}$$

Banaketa-funtzioa $(0, \infty)$ tartean definitzen da.




Irudia 6.4: Snedecor-en F banaketa.

Oharra: Taulek $\alpha < 0.5$ datuak baino ez dituzte. $\alpha > 0.5$ denean, honako propietate hau hartu behar da kontuan:

$$\frac{1}{F_{1-\alpha;\,v_2;\,v_1}} = F_{\alpha;\,v_1;\,v_2} \tag{6.9}$$

# 7. Laginketa eta Zenbatespena

## 7.1 Parametroen estimazioa

**Populazio baten parametroa** populazio horren ezaugarri garrantzitsu bati buruzko informazioa adierazten duen zenbaki bat da.

Adibidea: batez bestekoa eta bariantza populazio baten parametroak dira. *T-student* batean, aldiz, parametroa askatasun graduak dira. Populazio horren lagin bat daukagunez, parametro horiek zenbatetsiko ditugu.

### 7.1.1 Estimazio puntuala

**Estimatzaile puntuala** da laginaren parametroaren estimazioa emango digun formula.

batez bestekoa eta lagin-bariantza populaziokoak izango dira, baina balio bakar batekin zenbatesteak arazo asko sortzen dituenez (lagin bakoitzean balio hori alda daiteke), konfiantza-tarteak erabiliko ditugu.

#### batez besteko eta bariantzaren zenbateslea

Lagin-batez bestekoa ($\overline{x}$) eta lagin-bariantza ($s^2$) biztanleriaren batez bestekoaren ($\mu$) eta bariantzaren ($\sigma^2$) zenbatesleak dira, hurrenez hurren. Hau da:

$$\hat{y} = \overline{x} = \frac{1}{n} \cdot \sum_i x_i; \quad \hat{\sigma}^2 = s^2 = \frac{1}{n} \cdot \sum_i x_i^2 \tag{7.1}$$

Zenbatespen puntualak eragozpen handi bat du: populazio-ezaugarri edo parametro baten zenbatesle onena erabilita ere, zenbatespenean ez asmatzeaz gain (asmatzeko aukera urruna da), ez dugu haren zehaztasun- eta fidagarritasun-maila ezagutuko. Horregatik, estimazioa konfiantza-tarteetan oinarrituta egingo dugu.

### 7.1.2 Konfiantza-tarteak

**Konfiantza-tarteetan** egongo da zenbatespena, eta, horretarako, konfiantza- edo garrantzi-maila egokitu behar da (1 -$\alpha$).

Oharra: Adibidez, %95eko konfiantza-maila definitzen badugu, esan nahi du konfiantza-tarteak gure interesaren biztanleria-parametroaren (ertaina edo bariantza) egia-balioa izango duela, %95eko probabilitatearekin. Horrela, emandako populazioan $n$ tamainako lagin posible guztiak eraikitzen baditugu, kalkulatutako konfiantza-tarteen %95ek gure intereseko populazio-parametroaren egia-balioa izango du, eta gainerako %5etan ez da parametro hori sartuko.

**KONFIANTZA-TARTEA POPULAZIO NORMAL BATEN batez besteKORAKO**

1. Populazio-bariantza ezaguna eta lagin handia (n > 30)

$$T_\mu = \left[\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] \tag{7.2}$$

2. Populazio-bariantza ezezaguna eta lagin handia (n > 30)

$$T_\mu = \left[\overline{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}},\ \overline{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right] \tag{7.3}$$

3. Populazio-bariantza ezezaguna eta lagin txikia (n < 30)

$$T_\mu = \left[\overline{x} - t_{\alpha/2;\, n-1} \cdot \frac{s}{\sqrt{n}},\ \overline{x} + t_{\alpha/2;\, n-1} \cdot \frac{s}{\sqrt{n}}\right] \tag{7.4}$$

**BI POPULAZIO NORMALEN batez besteKOEN DIFERENTZIARAKO KONFIANTZA-TARTEA**

1. Populazio-bariantza ezagunak eta lagin handiak ($n_x > 30$, $n_y > 30$)

$$T_{\mu_1-\mu_2} = \left[(\overline{x_1} - \overline{x_2}) - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ (\overline{x_1} - \overline{x_2}) + z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] \tag{7.5}$$

2. Populazio-bariantza ezezaguna eta lagin handiak ($n_x > 30$, $n_y > 30$)

$$T_{\mu_1-\mu_2} = \left[(\overline{x_1} - \overline{x_2}) - z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ (\overline{x_1} - \overline{x_2}) + z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right] \tag{7.6}$$

3. Bariantzak, populazio ezezagunak baina berdinak eta lagin txikiak ($n_x < 30$, $n_y < 30$)

$$T_{\mu_1-\mu_2} = \left[(\overline{x_1} - \overline{x_2}) \mp t_{\alpha/2;\, n_1+n_2-2} \cdot \sqrt{\frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] \tag{7.7}$$

4. Populazio-bariantza ezezagunak eta lagin txikiak ($n_x < 30$, $n_y < 30$)

$$T_{\mu_1-\mu_2} = \left[(\overline{x_1} - \overline{x_2}) \mp t_{\alpha/2;\, m} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right] \tag{7.8}$$

$$c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}};\quad \frac{1}{m} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1} \tag{7.9}$$

**BARIANTZA EDO DESBIDERAPENA**

1. POPULAZIO NORMAL BATEN BARIANTZA EDO DESBIDERATZEA

$$\text{Bariantza: } T_{\sigma^2} = \left[ \frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2;\, n-1}},\ \frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2;\, n-1}} \right] \tag{7.10}$$

$$\text{Desbiderapen tipikoa: } T_{\sigma} = \left[ \frac{\sqrt{n-1}\cdot s}{\sqrt{\chi^2_{\alpha/2;\, n-1}}},\ \frac{\sqrt{n-1}\cdot s}{\sqrt{\chi^2_{1-\alpha/2;\, n-1}}} \right] \tag{7.11}$$

2. BI POPULAZIO NORMALEN BARIANTZEN EDO DESBIDERATZEEN ARRAZOIA

$$\text{Bariantza: } I = \left[ \frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\alpha/2;\, n_1-1;\, n_2-1}},\ \frac{s_1^2}{s_2^2} \cdot F_{\alpha/2;\, n_1-1;\, n_2-1} \right] \tag{7.12}$$

$$\text{Desbiderapen tipikoa: } I = \left[ \frac{s_1}{s_2} \cdot \frac{1}{\sqrt{F_{\alpha/2;\, n_1-1;\, n_2-1}}},\ \frac{s_1}{s_2} \cdot \sqrt{F_{\alpha/2;\, n_1-1;\, n_2-1}} \right] \tag{7.13}$$

Oharra:

1. Bi lagin daudenean, populazioak nola eraikitzen diren ikusi behar da:
   - Lagin ez-parekatua: bi populazio desberdin
   - Lagin parekatua: populazio bat
2. Bi populaziorekin, [A, B], lan egiten denean, aukera hauek izan ditzakegu:
   - $[-,-] \rightarrow$ B populazioa A populazioa baino hobea
   - $[-,+] \rightarrow$ zin da definitu zein den onena
   - $[+,+] \rightarrow$ A populazioa B populazioa baino hobea
3. Bariantzen arrazoian:
   - $[,] \in 1 \rightarrow$ ezin dugu esan bariantzak desberdinak direnik
   - $[,] \notin 1 \rightarrow$ badakigu bariantzak desberdinak direla

### 7.1.3 Normaltasun kontrasteak

Populazio batek laginaren bidez banaketa normal bat jarraitzen duen egiaztatzeko, proba hauek egiten dira:

$$\text{lagina} \begin{cases} n \geq 50 \rightarrow \text{Kolmogórov-Smirnov} \\ n < 50 \rightarrow \text{Shapiro-Wilk} \end{cases} \rightarrow \begin{cases} H_0 : \text{banaketa normala jarraitzen du} \\ H_1 : \text{ez du banaketa normala jarraitzen} \end{cases} \rightarrow$$

| | 2 populazio | 3 populazio |
|---|---|---|
| $H_0$ | t-Student | ANOVA |
| $H_1$ | Wilcoxon | Kruskal Wallis |

## 7.2 Laginketa

Laginketa esaten zaio biztanleria estatistiko batetik lagin bat hautatzeko teknikari.

### 7.2.1 Laginketa motak

- **Probabilitate-laginketa:** Populazio-unitate bakoitzak laginean hautatua izateko probabilitate bat du, eta probabilitate hori zehaztasunez zehaztu daiteke. Elementu guztiek hautatuak izateko probabilitate bera dute.
  - Ausazko laginketa sinplea: Metodo honen bidez, edozein lagin ateratzeko probabilitatea kalkula daiteke.
    - Elementuak ordeztu gabe.
    - Elementuak ordezkatuz.
  - Laginketa sistematikoa: Biztanleria handia denean edo denboran zehar lagindu behar denean erabiltzen da.
  - Laginketa estratifikatua: Aldez aurretik biztanleria talde homogeneotan zatitzean datza, azterketarako ezaugarri garrantzitsuren batekin alderatuta.
    - Esleipen proportzionala: laginaren tamaina estratu bakoitzaren tamainarekiko proportzionala da.
    - Esleipen optimoa: Laginak aldakortasun handiagoko estratuak lehenesten ditu, eta populazioa aldez aurretik ezagutzea eskatzen du.
- **Laginketa ez-probabilistikoa:** Laginketa mota honetan ezin da kalkulatu lagin espezifiko bat ateratzeko probabilitatea, subjektu guztiek ez baitute hautatuak izateko probabilitate bera.
  - Kuoten araberako laginketa: Ikertzaileak modu askean aukeratzen ditu laginaren subjektuak estratu bakoitzaren barruan.
  - Elur-bolaren laginketa: Irizpide jakin batzuk betetzen dituzten gizabanakoen hasierako talde batetik abiatuta, irizpide horiek antzeko ezaugarriak dituzten beste batzuen lokalizatzaile gisa balio dute.

– Laginketa subjektiboa erabaki arrazoitu bidez: Laginaren unitateak ezaugarri espezifikoen arabera hautatzen dira, modu arrazionalean, eta ez ausaz.

### 7.2.2 Aldagaiak

batez bestekoa, bariantza eta desbiderapenaz gain, meta-aldagaiak kalkula daitezke.

- **Populazio infinituak:**

$$\text{Baturenak: } \begin{cases} \mu_{\overline{x}} = \mu \\ \sigma^2_{\overline{x}} = \dfrac{\sigma^2}{n} \rightarrow \sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}} \end{cases} \rightarrow \overline{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$$

$$\text{Baturenak: } \begin{cases} \mu_T = n \cdot \mu \\ \sigma^2_T = n \cdot \sigma^2 \rightarrow \sigma_T = \sqrt{n} \cdot \sigma \end{cases} \rightarrow x_1 + x_2 + x_3 + ... + x_n$$

- **Populazio finituak:**

$$\text{Baturenak: } \begin{cases} \mu_{\overline{x}} = \mu \\ \sigma^2_{\overline{x}} = \dfrac{\sigma^2}{n} \cdot \dfrac{N-n}{N-1} \end{cases}$$

$$\text{Baturenak: } \begin{cases} \mu_T = n \cdot \mu \\ \sigma^2_T = n \cdot \sigma^2 \cdot \dfrac{N-n}{N-1} \end{cases}$$

Adibidea: $\mu = 5{,}2$; $\sigma = 0{,}5$

a. Baturen eta batez bestekoen batez bestekoa eta desbideratzea (n = 4).

$$\text{Baturenak: } \begin{cases} \mu_{\overline{x}} = \mu = 5.2 \\ \sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.5}{\sqrt{4}} = 0.25 \end{cases}$$

$$\text{Baturenak: } \begin{cases} \mu_T = n \cdot \mu = 4 \cdot 5.2 = 20.8 \\ \sigma_T = \sqrt{n} \cdot \sigma = \sqrt{4} \cdot 0.5 = 1 \end{cases}$$

b. Lau ikaslek lortutako kalifikazioen batura 22 edo handiagoa izateko probabilitatea.

$$P(x \geq 22) = P\left(z \geq \frac{22 - 20.8}{1}\right) = P(z \geq 1.2) = 1 - P(z \leq 1.2) = 1 - 0.8849 = 0.1151$$

c. Kalifikazioen batez bestekoa 4,5 baino txikiagoa izateko probabilitatea.

$$P(x < 4.5) = P\left(z < \frac{4.5 - 5.2}{0.25}\right) = P(z < -2.8) = 1 - P(z < 2.8) = 1 - 0.9974 = 0.0026$$

# 8. Hipotesien kontrasteak

Askotan interesatzen zaigu jakitea parametro batek balio jakin bat hartuko duen edo maila bat gainditukо duen.

Adibidea: Ingeniari batek esaten badu bere fabrikan enkargatutako piezen batez besteko luzera 5 cm-koa dela, eta beste batek ez. Hipotesi hori onartu edo baztertu beharko dugu.

## 8.1 Errore motak

Bi errore mota hartzen ditugu kontuan:

- Mota 1: hipotesia onartzea egia ez denean.
- Mota 2: hipotesiari uko egitea egia denean.

## 8.2 Hipotesi motak

- **Aldebakarreko hipotesia:**

$$\begin{cases} H_o : \theta \leq \theta_o \\ H_i : \theta > \theta_o \end{cases} \qquad \begin{cases} H_o : \theta \geq \theta_o \\ H_i : \theta < \theta_o \end{cases}$$

- **Aldebiko hipotesia:**

$$\begin{cases} H_o : \theta = \theta_o \\ H_i : \theta \neq \theta_o \end{cases}$$

## 8.3 Hipotesiak onartzeko tarteak

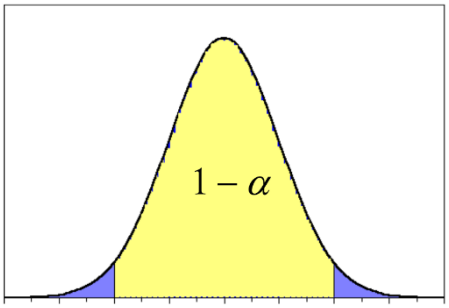
Zenbatetsitako tarteak dira, non zenbateslea egon daiteke. Hipotesia onartzeko tarteekin zuzenean lotuta daude. Formulak ez bikoizteko, aldi berean ikusiko dugu. Bi kasuetan konfiantza-maila bat finkatu behar da (n.c. = 1 - $\alpha$).

### 1 POPULAZIOA

1. **Populazio-bariantza ezaguna eta lagin handia (n > 30)**

   - Aldebiko hipotesia

$$\begin{cases} H_o : \mu = \mu_o \\ H_i : \mu \neq \mu_o \end{cases} \quad ; \quad z_o = \frac{(\overline{x} - \mu_o)}{\frac{\sigma}{\sqrt{n}}}; \quad z_o \in [-z_{\alpha/2};\ z_{\alpha/2}]$$




Oharra: Ikus daitekeenez, onarpen-tartea errefus-tartea baino askoz handiagoa da. Beraz, lehentasunez onartuko dugu, baina onartu beharrean ez dugula baztertzen esango dugu.

- Aldebakarreko hipotesia

$$\begin{cases} H_o : \mu \leq \mu_o \\ H_i : \mu > \mu_o \end{cases} ; \quad z_o = \frac{(\overline{x} - \mu_o)}{\frac{\sigma}{\sqrt{n}}}; \quad z_o \in [-\infty;\ z_\alpha]$$

$$\begin{cases} H_o : \mu \geq \mu_o \\ H_i : \mu < \mu_o \end{cases} ; \quad z_o = \frac{(\overline{x} - \mu_o)}{\frac{\sigma}{\sqrt{n}}}; \quad z_o \in [-z_\alpha;\ +\infty]$$

**2. Populazio-bariantza ezezaguna eta lagin handia (n > 30)**

- Aldebiko hipotesia

$$\begin{cases} H_o : \mu = \mu_o \\ H_i : \mu \neq \mu_o \end{cases} ; \quad z_o = \frac{(\overline{x} - \mu_o)}{\frac{s}{\sqrt{n}}}; \quad z_o \in [-z_{\alpha/2};\ z_{\alpha/2}]$$

- Aldebakarreko hipotesia

$$\begin{cases} H_o : \mu \leq \mu_o \\ H_i : \mu > \mu_o \end{cases} ; \quad z_o = \frac{(\overline{x} - \mu_o)}{\frac{s}{\sqrt{n}}}; \quad z_o \in [-\infty;\ z_\alpha]$$

$$\begin{cases} H_o : \mu \geq \mu_o \\ H_i : \mu < \mu_o \end{cases} ; \quad z_o = \frac{(\overline{x} - \mu_o)}{\frac{s}{\sqrt{n}}}; \quad z_o \in [-z_\alpha;\ +\infty]$$

**3. Populazio-bariantza ezezaguna eta lagin txikia (n < 30)**

- Aldebiko hipotesia

$$\begin{cases} H_o : \mu = \mu_o \\ H_i : \mu \neq \mu_o \end{cases} ; \quad t_o = \frac{(\overline{x} - \mu_o)}{\frac{s}{\sqrt{n}}}; \quad t_o \in [-t_{\alpha/2;\, n-1};\ z_{\alpha/2;\, n-1}]$$

- Aldebakarreko hipotesia

$$\left\{ \begin{array}{l} H_o : \mu \leq \mu_o \\ H_i : \mu > \mu_o \end{array} \right. ; \quad t_o = \frac{(\overline{x} - \mu_o)}{\frac{s}{\sqrt{n}}}; \quad t_o \in [-\infty;\ t_{\alpha;\, n-1}]$$

$$\left\{ \begin{array}{l} H_o : \mu \geq \mu_o \\ H_i : \mu < \mu_o \end{array} \right. ; \quad t_o = \frac{(\overline{x} - \mu_o)}{\frac{s}{\sqrt{n}}}; \quad t_o \in [-t_{\alpha;\, n-1};\ +\infty]$$

## 2 POPULAZIOAK

### 1. Populazio-bariantza ezagunak eta lagin handiak ($n_x > 30$; $n_y > 30$)

- Aldebiko hipotesia

$$\left\{ \begin{array}{l} H_o : \mu_x = \mu_y \\ H_i : \mu_x \neq \mu_y \end{array} \right. ; \quad z_o = \frac{(\overline{x} - \overline{y})}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}; \quad z_o \in [-z_{\alpha/2};\ z_{\alpha/2}]$$

- Aldebakarreko hipotesia

$$\left\{ \begin{array}{l} H_o : \mu_x \leq \mu_y \\ H_i : \mu_x > \mu_y \end{array} \right. ; \quad z_o = \frac{(\overline{x} - \overline{y})}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}; \quad z_o \in [-\infty;\ z_\alpha]$$

$$\left\{ \begin{array}{l} H_o : \mu_x \geq \mu_y \\ H_i : \mu_x < \mu_y \end{array} \right. ; \quad z_o = \frac{(\overline{x} - \overline{y})}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}; \quad z_o \in [-z_\alpha;\ +\infty]$$

### 2. Populazio-bariantza ezezagunak eta lagin handiak ($n_x > 30$, $n_y > 30$)

- Aldebiko hipotesia

$$\left\{ \begin{array}{l} H_o : \mu_x = \mu_y \\ H_i : \mu_x \neq \mu_y \end{array} \right. ; \quad z_o = \frac{(\overline{x} - \overline{y})}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}; \quad z_o \in [-z_{\alpha/2};\ z_{\alpha/2}]$$

- Aldebakarreko hipotesia

$$\left\{ \begin{array}{l} H_o : \mu_x \leq \mu_y \\ H_i : \mu_x > \mu_y \end{array} \right. ; \quad z_o = \frac{(\overline{x} - \overline{y})}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}; \quad z_o \in [-\infty;\ z_\alpha]$$

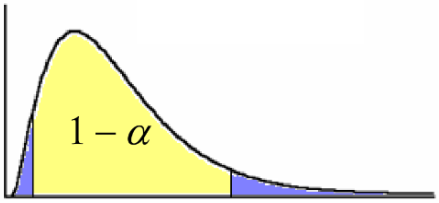
$$\left\{ \begin{array}{l} H_o : \mu_x \geq \mu_y \\ H_i : \mu_x < \mu_y \end{array} \right. ; \quad z_o = \frac{(\overline{x} - \overline{y})}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}; \quad z_o \in [-z_\alpha;\ +\infty]$$

## BARIANTZA ETA DESBIDERATZEA

### 1. Populazio bat

- Aldebiko hipotesia

$$\left\{ \begin{array}{l} H_o : \sigma = \sigma_o \\ H_i : \sigma \neq \sigma_o \end{array} \right. ; \quad \chi_o^2 = \frac{(n-1) \cdot s^2}{\sigma^2}; \quad \chi_o^2 \in [\chi^2_{1-\alpha/2;\, n-1};\ \chi^2_{1-\alpha/2;\, n-1}]$$




- Aldebakarreko hipotesia

$$\left\{ \begin{array}{l} H_o : \sigma \leq \sigma_o \\ H_i : \sigma > \sigma_o \end{array} \right. ; \quad \chi_o^2 = \frac{(n-1) \cdot s^2}{\sigma^2}; \quad \chi_o^2 \in [0,\ \chi^2_{\alpha;\, n-1}]$$

$$\left\{ \begin{array}{l} H_o : \sigma \geq \sigma_o \\ H_i : \sigma < \sigma \end{array} \right. ; \quad \chi_o^2 = \frac{(n-1) \cdot s^2}{\sigma^2}; \quad \chi_o^2 \in [\chi^2_{1-\alpha;\, n-1};\ +\infty]$$

Oharra: Banaketa Chi-karratuarekin ziurgabetasun-eremu bat aurkituko dugu.

$\chi_o$ ziurgabetasun-eskualdean erortzen bada $[\chi^2_{1-\alpha;\, n-1}, \chi^2_{\alpha;\, n-1}]$ ezin dugu bariantzetatik inolako ondoriorik atera. Arazo hori konpontzeko modu bat lagina handitzea da.

# 9. Bariantzaren Analisia (ANOVA)

Proba hau laginen batez bestekoak batez besteko berdinak dituzten populazioetatik datozen zehazteko erabiltzen da, aztertutako bi populazio baino gehiago daudenean.

Demagun k populazio desberdin alderatu nahi ditugula. Beraz, hauxe izango da gure hipotesi-kontrastea:

$$\begin{cases} H_o : \mu_1 = \mu_2 = \mu_3 = ... = \mu_n \\ H_i : \text{kontrako kasuan} \end{cases}$$

Kontraste horretarako, honako suposizio hauek egin behar ditugu:

- Populazio guztiek banaketa normala dute. Banaketa normal bat jarraituko ez balute, Kruscal-Wallisen analisia aplikatu beharko litzateke.
- Populazio guztiak askeak dira.
- Populazio guztiek bariantza bera dute.

Hiru baldintzak betetzen badira. Hipotesiak ANOVA taulen bidez kontrastatuko dira.

## 9.1 ANOVA taula eraikitzea

| | $A_1$ | $A_2$ | ... | $A_n$ |
|---|---|---|---|---|
| | $x_{11}$ | $x_{12}$ | ... | $x_{ik}$ |
| | $x_{21}$ | $x_{22}$ | ... | $x_{2k}$ |
| | ... | ... | ... | ... |
| | | $x_{n_2 2}$ | ... | $x_{n_k k}$ |
| | $x_{n_1 1}$ | | ... | |
| B | $B_1$ | $B_2$ | ... | $B_k$ |
| N | $n_1$ | $n_2$ | ... | $n_k$ |

$$B = \sum_{i=1}^{k} B_i; \quad \text{Laginaren tamaina guztira: } N = \sum_{i=1}^{k} n_i \tag{9.1}$$

$\sigma^2$ bariantza bera populazio guztietarako onartu dugunez, balio hori bi modutara kalkula daiteke: klase arteko bariantza erabiliz edo klase barruko bariantza erabiliz.

- **klase arteko bariantza:**

$$S_A^2 = \frac{1}{k-1} \cdot \left[ \sum_{j=1}^{k} (\frac{B_j^2}{n_j}) - \frac{B^2}{N} \right] \tag{9.2}$$

non $k$ populazio kopurua den eta *(k - 1)* askatasun maila.

- **Klase barneko bariantza:**

$$S_B^2 = \frac{1}{N-k} \cdot \sum_{j=1}^{n} \sum_{i=1}^{n} x_{ij}^2 - \sum_{j=1}^{n} \frac{B_j^2}{n_j} \tag{9.3}$$

non *(N - k)* askatasun maila.

Beraz, klase arteko eta klase barruko bariantzak zatituta, Fisher Snedecorreren banaketa lortuko dugu, $(k-1)$ eta $(N-k)$ askatasun mailarekin.

$$F = \frac{S_A^2}{S_B^2} \qquad \begin{cases} \mathrm{H}_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_n \quad \rightarrow \quad F \leq F_{\alpha;\, k-1;\, N-k} \\ \mathrm{H}_i : \text{kontrako kasuan} \quad \rightarrow \quad F > F_{\alpha;\, k-1;\, N-k} \end{cases} \tag{9.4}$$

## 9.2 ANOVA bi faktorerekin

Bi kontraste mota izango ditugu, zutabeka eta lerroka.

$$\begin{cases} H_o : \mu_{Z_1} = \mu_{Z_2} = \mu_{Z_3} = ... = \mu_{Z_k} \\ H_i : \text{kontrako kasuan} \end{cases} \quad \text{y} \quad \begin{cases} H_o : \mu_{E_1} = \mu_{E_2} = \mu_{E_3} = ... = \mu_{E_n} \\ H_i : \text{kontrako kasuan} \end{cases}$$

$\sigma^2$ bariantza bera populazio guztietarako onartu dugunez, balio hori bi modutara kalkula daiteke: klase arteko bariantza erabiliz edo maila barneko bariantza erabiliz.

- **Klase arteko bariantza: Klase arteko bariantza** aurreko kasuan bezala kalkulatuko da, baina kasu honetan zutabe edo errenkada bidez.

$$KB_{AZ} = \sum_{j=1}^{k} \frac{B_j^2}{n} - \frac{B^2}{N}; \;\; k-1 \text{ askatasun maila}; \;\; S_Z^2 = \frac{KB_{AZ}}{k-1} \tag{9.5}$$

$$KB_{AE} = \sum_{j=1}^{k} \frac{D_j^2}{n} - \frac{D^2}{N}; \;\; n-1 \text{ askatasun maila}; \;\; S_E^2 = \frac{KB_{AE}}{n-1} \tag{9.6}$$

- **Klase barneko bariantza:**

$$KB_{B,H} = \sum_{j=1}^{k} \sum_{i=1}^{n} x_{ij}^2 - \sum_{j=1}^{k} \frac{B_j^2}{n} - \sum_{i=1}^{n} \frac{B_i^2}{n} - \frac{B^2}{N}; \;\; (n-1)(k-1) \text{ askatasun maila} \tag{9.7}$$

$$S_B^2 = \frac{KB_{B,H}}{(n-1)(k-1)}$$

Azken bariantza hori oso handia denez, bariantza totala kalkula dezakegu,

$$KB_T = \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}^2 - \frac{B^2}{N} \text{ non } KB_T = KB_{B,H} + KB_{AZ} + KB_{AE} \qquad (9.8)$$

## 9.3 Aldakuntza-analisia

| Aldakuntza | SK | Askatasun maila | $S^2$ | F |
|---|---|---|---|---|
| Interklasea ($KB_A$) | $\sum_{j=1}^{k} \frac{B_j^2}{n_j} - \frac{B^2}{N}$ | k - 1 | $\frac{KB_A}{k-1} = S_A^2$ | $\frac{S_A^2}{S_B^2}$ |
| Interklasea ($KB_B$) | $\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}^2 - \sum_{j=1}^{2} \frac{B_j^2}{n_j}$ | N - k | $\frac{KB_B}{N-k} = S_B^2$ | |
| Guztira | $\sum_{j=1}^{k}\sum_{i=i}^{n_j} x_{ij}^2 - \frac{B^2}{N}$ | N - 1 | | |

| Aldakuntza | SK | Askatasun maila | $S^2$ | F |
|---|---|---|---|---|
| Errenkaden artean | $KB_{AE}$ | n - 1 | $\frac{KB_{AE}}{n-1} = S_E^2$ | $\frac{S_E^2}{S_B^2}$ |
| Zutabeen artean | $KB_{AZ}$ | k - 1 | $\frac{KB_{AZ}}{k-1} = S_Z^2$ | $\frac{S_Z^2}{S_B^2}$ |
| Hondar aldakuntza | $\sum_{j=1}^{k}\sum_{i=i}^{n} x_{ij}^2 - \frac{B_{Z_j}^2}{n} - \sum_{i=1}^{n} \frac{B_{E_i}^2}{k} + \frac{B^2}{n \cdot k}$ | (n - 1)(k - 1) | $\frac{KB_{AH}}{(n-1)(k-1)} = S_B^2$ | |

# 10. Ariketak

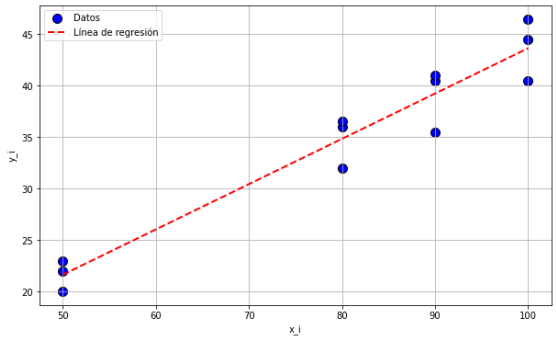
## 10.1 Erregresioa eta Korrelazioa

**1. 12 pertsonen artean odol analisia egiten da. Odol-bolumena eta bakoitzaren globulu gorrien zenbaketa ehunekotan neurtzen dira. Emaitzak taula honetan biltegiratzen dira:**

| Pertsona | $x_i$ | $y_i$ |
|---|---|---|
| 1 | 100 | 40,5 |
| 2 | 100 | 44,5 |
| 3 | 100 | 46,5 |
| 4 | 90 | 35,5 |
| 5 | 90 | 40,5 |
| 6 | 90 | 41,0 |
| 7 | 80 | 32,0 |
| 8 | 80 | 32,0 |
| 9 | 80 | 36,5 |
| 10 | 50 | 20,0 |
| 11 | 50 | 22,0 |
| 12 | 50 | 23,0 |

**Kalkulatu zuzena eta erregresio errorea, eta korrelazio-koefizientea.**

Hasi aurretik, ikus dezagun grafikoki nola ikus daitekeen erregresioa.




Erregresio linealaren antza du. Beraz, taula osagarri bat sortuko dugu:

| Pertsona | $x_i$ | $y_i$ | $x_i \cdot y_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|---|
| 1 | 100 | 40,5 | 4050 | 10000 | 1640,25 |
| 2 | 100 | 44,5 | 4450 | 10000 | 1980,25 |
| 3 | 100 | 46,5 | 4650 | 10000 | 2162,25 |
| 4 | 90 | 35,5 | 3195 | 8100 | 1260,25 |
| 5 | 90 | 40,5 | 3645 | 8100 | 1640,25 |
| 6 | 90 | 41,0 | 3690 | 8100 | 1681,00 |
| 7 | 80 | 32,0 | 2560 | 6400 | 1024,00 |
| 8 | 80 | 36,0 | 2880 | 6400 | 1296,00 |
| 9 | 80 | 36,5 | 2920 | 6400 | 1332,25 |
| 10 | 50 | 20,0 | 1000 | 2500 | 400,00 |
| 11 | 50 | 22,0 | 1100 | 2500 | 484,00 |
| 12 | 50 | 23,0 | 1150 | 2500 | 529,00 |
| Totala | 960 | 418,0 | 35290 | 81000 | 15429,50 |

Orduan,

$$b = \frac{cov(x,\ y)}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i - \overline{x} \cdot \overline{y}}{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} = \frac{35290/12 - 960/12 \cdot 418/12}{81000/12 - (960/12)^2} = 0.44047$$

$$a = \overline{y} - b \cdot \overline{x} = 418/12 - 0.44047 \cdot 960/12 = -0.40476$$

Beraz,

$$\hat{y} = -0.40476 + 0.44047 \cdot x$$

Orduan, 100-eko presio arteriala duen pertsona batentzat kalkulatutako globulu gorrien kopurua hurrengoa izango da:

$$\hat{y} = -0.40476 + 0.44047 \cdot 100 = 43.642$$

Beraz, zein da errore koadratikoa?

$$s_{ey} = \sqrt{\frac{\sum y_i^2 - a \cdot \sum y_i - b \cdot \sum x_i \cdot y_i}{n-2}} =$$

$$= \sqrt{\frac{15429.5 - (-0.40476) \cdot 418 - 0.44047 \cdot 35290}{12-2}} = 2.359$$

Korrelazio koefizientea kalkulatuko dugu:

$$r = \frac{s_x}{s_y} \cdot b = \frac{\sqrt{81000/12 - (960/12)^2}}{\sqrt{15429,5/12 - (418/12)^2}} \cdot 0,44047 = 0,968$$

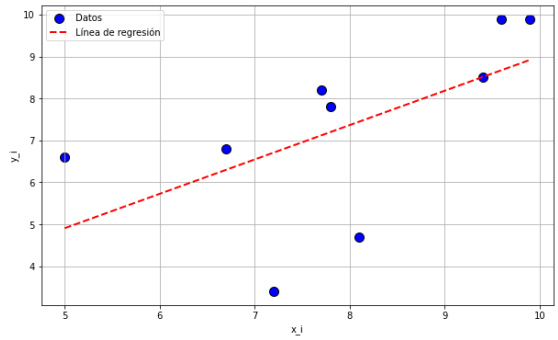
Lortutako korrelazio koefizientearen balioa unitatetik oso hurbil dago; beraz, oso erlazio handia dago.

**2. 9 ikasleren kalifikazioak taula honetan ageri dira. X lehen lauhileko eta Y bigarreneko kalifikazioak dira:**

| Pertsona | $x_i$ | $y_i$ |
|---|---|---|
| 1 | 7,7 | 8,2 |
| 2 | 5,0 | 6,6 |
| 3 | 7,8 | 7,8 |
| 4 | 7,2 | 3,4 |
| 5 | 8,1 | 4,7 |
| 6 | 9,4 | 8,5 |
| 7 | 9,6 | 9,9 |
| 8 | 9,9 | 9,9 |
| 9 | 6,7 | 6,8 |
| Totala | 71,4 | 6,8 |

**Lehen lauhilekoan lortutako notak bigarren lauhilekoan duen eragina ikusteko, kalkulatu erregresio zuzena eta erregresio-errorea.**

Hasi aurretik, ikus dezagun grafikoki nola ikusiko litzatekeen erregresioa:




Erregresio linealaren antza du. Beraz, taula osagarri bat sortuko dugu:

| Pertsona | $x_i$ | $y_i$ | $x_i \cdot y_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|---|
| 1 | 7,7 | 8,2 | 63,14 | 59,29 | 67,24 |
| 2 | 5,0 | 6,6 | 33,00 | 25,00 | 43,56 |
| 3 | 7,8 | 7,8 | 60,84 | 60,84 | 60,84 |
| 4 | 7,2 | 3,4 | 24,48 | 51,84 | 11,56 |
| 5 | 8,1 | 4,7 | 38,07 | 65,61 | 22,09 |
| 6 | 9,4 | 8,5 | 79,90 | 88,36 | 72,25 |
| 7 | 9,6 | 9,9 | 95,04 | 92,16 | 98,01 |
| 8 | 9,9 | 9,9 | 98,01 | 98,01 | 98,01 |
| 9 | 6,7 | 6,8 | 45,56 | 44,89 | 46,24 |
| Totala | 71,4 | 65,8 | 538,04 | 586,00 | 519,80 |

Orduan,

$$b = \frac{cov(x,\ y)}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i - \overline{x} \cdot \overline{y}}{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} = \frac{538,04/9 - 71,4/9 \cdot 65,8/9}{586/9 - (71,4/9)^2} = 0,819$$

$$a = \overline{y} - b \cdot \overline{x} = 65,8/9 - 0,819 \cdot 71,4/9 = 0,814$$

Beraz,

$$\hat{y} = 0,814 + 0,819 \cdot x$$

Orduan, lehenengo lauhilekoan 8,5 lortu zuen pertsonak, bigarren lauhilekoan lortuko duen gutxi gorabeherako kalifikazioa:

$$\hat{y} = 0,814 + 0,819 \cdot 8,5 = 7,776$$

Beraz, zein da akats koadratikoa?

$$s_{ey} = \sqrt{\frac{\sum y_i^2 - a \cdot \sum y_i - b \cdot \sum x_i \cdot y_i}{n-2}} =$$

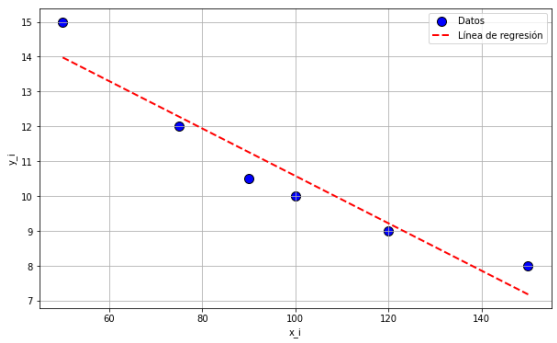
$$= \sqrt{\frac{519,8 - 0,814 \cdot 65,8 - 0,819 \cdot 71,4}{9-2}} = 7,632$$

**3. Hainbat auto motaren potentzia (X), zaldietan neurtuta, eta azelerazio ahalmena (Y), segundoetan neurtua 0tik 100era iristeko, taula honetan laburbiltzen da:**

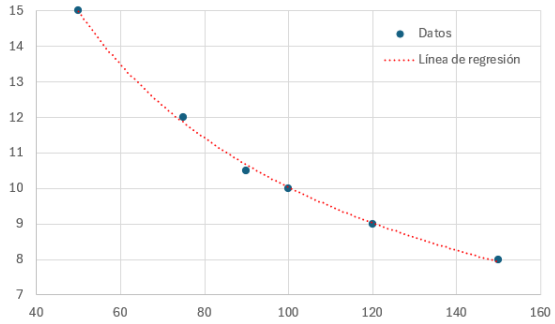
| Autoa | $x_i$ | $y_i$ |
|---|---|---|
| 1 | 50 | 15,0 |
| 2 | 75 | 12,0 |
| 3 | 90 | 10,5 |
| 4 | 100 | 10,0 |
| 5 | 120 | 9,0 |
| 6 | 150 | 8,0 |

**Kalkulatu kurba eta erregresio-errorea.**

Hasi aurretik, ikus dezagun erregresioa grafikoki.



Ez du erregresio linealaren antz handirik. Ikus dezagun erregresio potentzialetik gertuago dagoen:




Argi dago kurba hori gure datuetara askoz gehiago hurbiltzen dela. Orduan, nola kalkulatzen ditugu koefizienteak?

$$y = \alpha \cdot x^{\beta} \rightarrow ln\, y = ln(\alpha \cdot x^{\beta}) = ln\, \alpha + ln(x^{\beta}) = ln\, \alpha + \beta \cdot (ln\, x)$$

Erregresio potentzial bat dugu. Orduan, hurrengo taula sortuko dugu:

| Autoa | $x_i$ | $y_i$ | $ln\, x_i$ | $ln\, y_i$ | $x_i \cdot y_i$ | $x_i \cdot ln\, y_i$ | $ln\, x_i \cdot ln\, y_i$ | $x_i^2$ | $(ln\, x_i)^2$ | $y_i^2$ | $(ln\, y_i)^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 50 | 15,0 | 3,91 | 2,71 | 750 | 135,40 | 10,59 | 2500 | 15,30 | 225,00 | 7,33 |
| 2 | 75 | 12,0 | 4,32 | 2,48 | 900 | 186,37 | 10,73 | 5625 | 18,64 | 144,00 | 6,17 |
| 3 | 90 | 10,5 | 4,49 | 2,35 | 945 | 211,62 | 10,58 | 8100 | 20,25 | 110,25 | 5,53 |
| 4 | 100 | 10,0 | 4,61 | 2,30 | 1000 | 230,26 | 10,60 | 10000 | 21,21 | 100,00 | 5,30 |
| 5 | 120 | 9,0 | 4,79 | 2,19 | 1080 | 263,67 | 10,52 | 14400 | 22,92 | 81,00 | 4,83 |
| 6 | 150 | 8,0 | 5,01 | 2,08 | 1200 | 311,92 | 10,42 | 22500 | 25,11 | 64,00 | 4,32 |
| Totala | 585 | 64,5 | 27,13 | 14,12 | 5875 | 1339,24 | 63,45 | 63125 | 123,43 | 724,25 | 33,49 |

Orduan,

$$b = \frac{cov(ln\, x, ln\, y)}{s^2_{ln(x)}} = \frac{\frac{1}{n}\sum_{i=1}^{n} ln\, x_i \cdot ln\, y_i - \overline{ln\, x} \cdot \overline{ln\, y}}{\frac{\sum_{i=1}^{n}(ln\, x_i - \overline{ln\, x})^2}{n}} =$$

$$= \frac{63,45/6 - 27,13/6 \cdot 14,12/6}{123,43/6 - (27,13/6)^2} = -0,578646$$

$ln\, a = \overline{ln\, y} - b \cdot \overline{ln\, x} = (14,12/6) - (-0,578646) \cdot (27,13/6) = 4,969777663 \rightarrow a = 143,994868$

Beraz,

$$\hat{y} = 143,994868 \cdot x^{-0,578646}$$

Orduan, 76 zaldiko autoa 0tik 100era pasatzeko aurreikusitako denbora hau izango da:

$$\hat{y} = 143,994868 \cdot 76^{-0,578646} = 11,7495^{***} \text{ segundu}$$

Beraz, zein da errore koadratikoa?

$$s_{ey} = \sqrt{\frac{\sum (ln\ y_i)^2 - ln\ a \cdot \sum ln\ y_i - b \cdot \sum ln\ x_i \cdot ln\ y_i}{n-2}} =$$

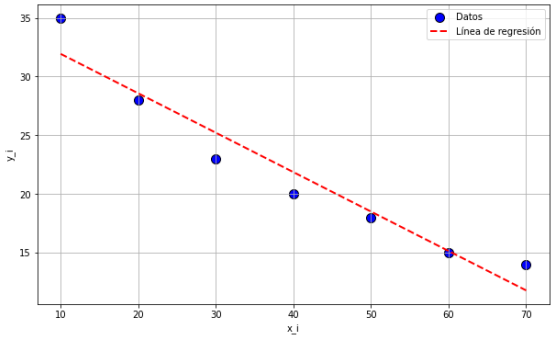
$$= \sqrt{\frac{33,49 - 4,97 \cdot 14,12 - (-0,578646) \cdot 63,45}{6-2}} = 0.048339251$$

4. **Fabrika batean pieza berri bat egiten hasi dira. Langileen eraginkortasuna denboran zehar aztertzeko, datu hauek gehitu dira: pieza egiteko minutu kopurua (X) eta pieza fabrikan fabrikatzen daramaten egun kopurua (Y).**

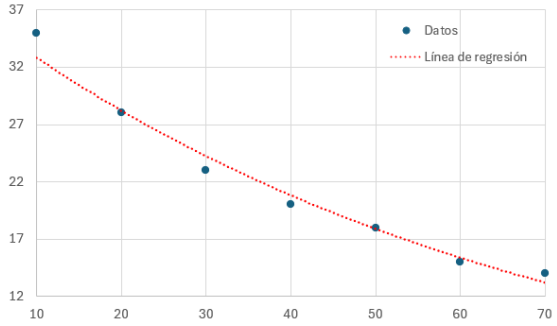
| Pertsona | $x_i$ | $y_i$ |
|---|---|---|
| 1 | 10 | 35 |
| 2 | 20 | 28 |
| 3 | 30 | 23 |
| 4 | 40 | 20 |
| 5 | 50 | 18 |
| 6 | 60 | 15 |
| 7 | 70 | 14 |

**Kalkulatu kurba eta erregresio-errorea.**

Hasi aurretik, ikus dezagun grafikoki nola ikus daitekeen erregresioa.




Ez du erregresio lineal baten antz handirik. Ikus dezagun erregresio esponentzialera gehiago hurbiltzen den:




Argi dago lerro hau gure datuetara gehiago hurbiltzen dela. Orduan, nola kalkulatzen ditugu koefizienteak?

$$y = \alpha \cdot e^{\beta \cdot x} \rightarrow ln\ y = ln\ (\alpha \cdot e^{\beta \cdot x}) = ln\ \alpha + ln\ (e^{\beta \cdot x}) = ln\ \alpha + \beta \cdot x$$

Erregresio esponentziala dugu. Beraz, taula osagarri hau sortuko dugu::

| Pertsona | $x_i$ | $y_i$ | $ln\ x_i$ | $ln\ y_i$ | $x_i \cdot y_i$ | $x_i \cdot ln\ y_i$ | $ln\ x_i \cdot ln\ y_i$ | $x_i^2$ | $(ln\ x_i)^2$ | $y_i^2$ | $(ln\ y_i)^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10 | 35 | 2,30 | 3,56 | 350 | 35,55 | 8,19 | 100 | 5,30 | 1225 | 12,64 |
| 2 | 20 | 28 | 2,99 | 3,33 | 560 | 66,64 | 9,98 | 400 | 8,97 | 784 | 11,10 |
| 3 | 30 | 23 | 3,40 | 3,14 | 690 | 94,06 | 10,66 | 900 | 11,57 | 529 | 9,83 |
| 4 | 40 | 20 | 3,69 | 2,99 | 800 | 119,83 | 11,05 | 1600 | 13,61 | 400 | 8,97 |
| 5 | 50 | 18 | 3,91 | 2,89 | 900 | 144,52 | 11,31 | 2500 | 15,30 | 324 | 8,35 |
| 6 | 60 | 15 | 4,09 | 2,71 | 900 | 162,48 | 11,09 | 3600 | 16,76 | 225 | 7,33 |
| 7 | 70 | 14 | 4,25 | 6,34 | 980 | 184,73 | 11,21 | 4900 | 18,05 | 196 | 6,96 |
| Totala | 280 | 153 | 24,64 | 21,25 | 5180 | 807,82 | 73,49 | 14000 | 89,56 | 3683 | 65,20 |

Orduan,

$$b = \frac{cov(x,\ ln\ y)}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot ln\ y_i - \overline{x} \cdot \overline{ln\ y}}{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} =$$

$$= \frac{807,82/7 - 280/7 \cdot 21,25/7}{14000/7 - (280/7)^2} = -0,0151$$

$ln\ a = \overline{ln\ y} - b \cdot \overline{x} = (21,25/7) - (-0,0151) \cdot (280/7) = 3,639 \rightarrow a = e^{ln\ a} = e^{3,639} = 38.05376393$

Beraz,

$$\hat{y} = 38.05 \cdot e^{-0,0151 \cdot x}$$

Beraz, zein da errore koadratikoa?

$$s_{ey} = \sqrt{\frac{\sum (ln\ y_i)^2 - ln\ a \cdot \sum ln\ y_i - b \cdot \sum x_i \cdot ln\ y_i}{n-2}} =$$

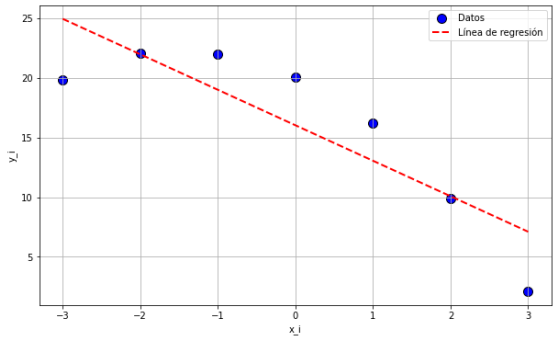
$$= \sqrt{\frac{65,20 - 3,639 \cdot 21,25 - (-0,0151) \cdot 807,82}{7-2}} = 0.117755679$$

5. **Korazatu batetik jaurtitako jaurtigai baten altuera jaurtiketaren ondorengo lehen 7 segundoetan neurtu zen. Kalkuluak sinplifikatzeko, denboren hasiera nahi bezala hartu da. Hona hemen lortutako datuak:**

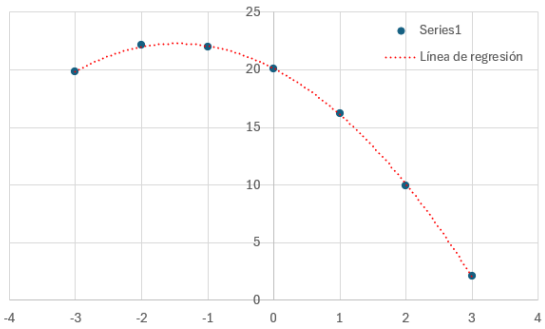
| $x_i$: Denbora (s) | $y_i$: Altuera (m) |
|---|---|
| -3 | 19,8 |
| -2 | 22,1 |
| -1 | 22,0 |
| 0 | 20,1 |
| 1 | 16,2 |
| 2 | 9,9 |
| 3 | 2,1 |

**Kalkulatu erregresio-kurba edo parabola kasu honetan eta erregresio-errorea.**

Hasi aurretik, ikus dezagun grafikoki nola ikus daitekeen erregresioa.




Ez du erregresio lineal baten antz handirik. Ikus dezagun erregresio parabolikora gehiago hurbiltzen den:



Argi dago kurba hau gure datuetara askoz gehiago hurbiltzen dela. Orduan, nola kalkulatzen ditugu koefizienteak? Ebatz dezagun ekuazio-sistema hau:

$$\begin{cases} a \cdot n + b \cdot \sum_i x_i + c \cdot \sum_i x_i^2 = \sum_i y_i \\ a \cdot \sum_i x_i + b \cdot \sum_i x_i^2 + c \cdot \sum_i x_i^3 = \sum_i x_i \cdot y_i \\ a \cdot \sum_i x_i^2 + b \cdot \sum_i x_i^3 + c \cdot \sum_i x_i^4 = \sum_i x_i^2 \cdot y_i \end{cases}$$

Horretarako, taula osagarri bat sortuko dugu:

| Zenba | $x_i$ | $y_i$ | $x_i \cdot y_i$ | $x_i^2$ | $x_i^3$ | $x_i^4$ | $x_i^2 \cdot y_i$ | $y_i^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | -3 | 19,8 | -59,4 | 9 | -27 | 81 | 178,2 | 329,04 |
| 2 | -2 | 22,1 | -44,2 | 4 | -8 | 16 | 88,4 | 488,41 |
| 3 | -1 | 22,0 | -22 | 1 | -1 | 1 | 22 | 484 |
| 4 | 0 | 20,1 | 0 | 0 | 0 | 0 | 0 | 404,01 |
| 5 | 1 | 16,2 | 16,2 | 1 | 1 | 1 | 16,2 | 262,44 |
| 6 | 2 | 9,9 | 19,8 | 4 | 2 | 16 | 39,6 | 98,01 |
| 7 | 3 | 2,1 | 6,3 | 9 | 27 | 81 | 18,9 | 4,41 |
| Totala | 0 | 112,2 | -83,3 | 28 | 0 | 196 | 363,3 | 2133,32 |

Orduan,

$$\begin{cases} a \cdot 7 + b \cdot 0 + c \cdot 28 = 112,2 \\ a \cdot 0 + b \cdot 28 + c \cdot 0 = -83,3 \\ a \cdot 28 + b \cdot 0 + c \cdot 196 = 363,3 \end{cases} \rightarrow \begin{cases} a = 20,1 \\ b = -2,975 \\ c = -1,0178 \end{cases}$$

Beraz,

$$\hat{y} = 20,1 - 2,975 \cdot x - 1,0178 \cdot x^2$$

Bukatzeko, errore koadratikoa kalkulatzen dugu:

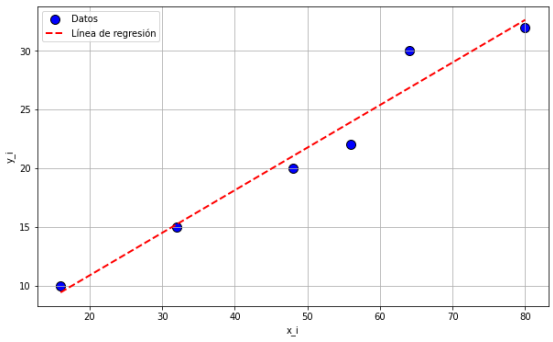
$$s_{ey} = \sqrt{\frac{\sum y^2 - a \cdot \sum y - b \cdot \sum x \cdot y - c \cdot \sum x^2 \cdot y}{n-2}} =$$

$$= \sqrt{\frac{2133,32 - 20,1 \cdot 112,2 - (-2,975) \cdot (-83,3) - (-1,0178) \cdot 363,3}{7-2}} = 0,09923$$

**6. Enpresa jakin bati buruzko datu hauek ezagutzen dira, azken 6 urteetako salmenta-bolumenari (milaka eurotan) eta publizitate-gastuari (eurotan) buruzkoak:**

| $x_i$: Publizitate-gastuak (eurotan) | $y_i$: Salmenta-bolumena (milaka euro) |
|---|---|
| 16 | 10 |
| 32 | 15 |
| 48 | 20 |
| 56 | 22 |
| 64 | 30 |
| 80 | 32 |

**a. Lotura lineala al dago enpresaren salmenten eta publizitate-gastuen artean?**




Hori kontuan hartuta, bi aldagaien artean erlazio lineala dagoela esan dezakegu. Orain, determinazio linealeko koefizientea kalkulatuko dugu, aldagaien arteko asoziazio-maila linealaren neurri deskribatzailea lortzeko. Kalkulu hori egiteko, taula hau sortu behar da:

| | $x_i$ | $y_i$ | $x_i \cdot y_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|---|
| | 16 | 10 | 160 | 256 | 100 |
| | 32 | 15 | 480 | 1024 | 225 |
| | 48 | 20 | 960 | 2304 | 400 |
| | 56 | 22 | 1232 | 3136 | 484 |
| | 64 | 30 | 1920 | 4096 | 900 |
| | 80 | 32 | 2560 | 6400 | 1024 |
| Totala | 296 | 129 | 7312 | 17216 | 3133 |

$$b = \frac{cov(x,\ y)}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i - \overline{x} \cdot \overline{y}}{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} = \frac{7312/6 - 296/6 \cdot 129/6}{17216/6 - (296/6)^2} = 0.363$$

$$r = \frac{s_x}{s_y} \cdot b = \frac{\sqrt{17216/6 - (296/6)^2}}{\sqrt{3133/6 - (129/6)^2}} \cdot 0.363 = 0,9787 \rightarrow r^2 = 0.957$$

**b. Lortu gutxieneko erregresio koadratikoko zuzenak.**

$\hat{y} = a + b \cdot x$ gisa adierazten badugu, koefizienteak honela kalkulatuko dira:

$$b = \frac{cov(x,y)}{s_x^2} = 0,363; \quad a = \overline{y} - b \cdot \overline{x}$$

$$a = \frac{129}{6} - 0.363 \cdot \frac{296}{6} = 3.592 \rightarrow \hat{y} = 3.592 + 0.363 \cdot x$$

c. **Enpresaren zer salmenta-bolumen espero liteke publizitatetik 60 € gastatzea urtebetean? Eta 200 €-ko publizitate-gasturako?**

Salmenta-bolumena iragartzeko, salmentek publizitateko gastuen arabera duten erregresio-zuzena erabiliko dugu. 60 €-ko publizitate-gasturako, honako salmenta-bolumen hau lortuko dugu:

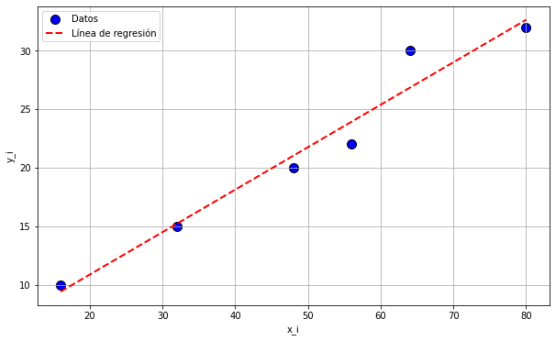
$$\hat{y} = 3.592 + 0.363 \cdot x = 3.592 + 0.363 \cdot 60 = 25,372$$

Gastua 200 € bada, ezin dugu erregresio-zuzena erabili, 200 balioa publizitateko gastuaren ibilbidetik kanpo baitago.

**Enpresa jakin bati buruzko datu hauek ezagutzen dira, azken 6 urteetako salmenta-bolumenari (milaka eurotan) eta publizitate-gastuari (eurotan) buruzkoak:**

| $x_i$: Publizitate-gastuak (eurotan) | $y_i$: Salmenta-bolumena (milaka euro) |
|---|---|
| 16 | 10 |
| 32 | 15 |
| 48 | 20 |
| 56 | 22 |
| 64 | 30 |
| 80 | 32 |

a. **Lotura lineala al dago enpresaren salmenten eta publizitate-gastuen artean?**




Hori kontuan hartuta, bi aldagaien artean erlazio lineala dagoela esan dezakegu. Orain, determinazio linealeko koefizientea kalkulatuko dugu, aldagaien arteko asoziazio-maila linealaren neurri deskribatzailea lortzeko. Kalkulu hori egiteko, taula hau sortu behar da:

| | $x_i$ | $y_i$ | $x_i \cdot y_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|---|
| | 16 | 10 | 160 | 256 | 100 |
| | 32 | 15 | 480 | 1024 | 225 |
| | 48 | 20 | 960 | 2304 | 400 |
| | 56 | 22 | 1232 | 3136 | 484 |
| | 64 | 30 | 1920 | 4096 | 900 |
| | 80 | 32 | 2560 | 6400 | 1024 |
| Totala | 296 | 129 | 7312 | 17216 | 3133 |

$$b = \frac{cov(x,\ y)}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i - \overline{x} \cdot \overline{y}}{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} = \frac{7312/6 - 296/6 \cdot 129/6}{17216/6 - (296/6)^2} = 0.363$$

$$r = \frac{s_x}{s_y} \cdot b = \frac{\sqrt{17216/6 - (296/6)^2}}{\sqrt{3133/6 - (129/6)^2}} \cdot 0.363 = 0,9787 \rightarrow r^2 = 0.957$$

**b. Lortu gutxieneko erregresio koadratikoko zuzenak.**

$\hat{y} = a + b \cdot x$ gisa adierazten badugu, koefizienteak honela kalkulatuko dira:

$$b = \frac{cov(x, y)}{s_x^2} = 0,363; \quad a = \overline{y} - b \cdot \overline{x}$$

$$a = \frac{129}{6} - 0.363 \cdot \frac{296}{6} = 3.592 \rightarrow \hat{y} = 3.592 + 0.363 \cdot x$$

**c. Enpresaren zer salmenta-bolumen espero liteke publizitatetik 60 € gastatzea urtebetean? Eta 200 €-ko publizitate-gasturako?**

Salmenta-bolumena iragartzeko, salmentek publizitateko gastuen arabera duten erregresio-zuzena erabiliko dugu. 60 €-ko publizitate-gasturako, honako salmenta-bolumen hau lortuko dugu:

$$\hat{y} = 3.592 + 0.363 \cdot x = 3.592 + 0.363 \cdot 60 = 25,372$$

Gastua 200 €koa bada, ezin dugu erregresio-zuzena erabili, 200 balioa publizitateko gastuaren ibilbidetik kanpo baitago.

## 10.2 Konbinatoria

1. **Hiru opari desberdin banatzen dira bost mutilen artean. Zenbat modutan egin dezakete, baldin eta:**

   Guztira 5 mutil eta 3 opari ditugu banatzeko, beraz$n$ balioa 5 da eta $m$ balioa 3 da.

   a. **Mutil bakoitzak opari bakarra jaso dezake.**

      Zer konbinatoria mota den identifikatzeko, honako alderdi hauek hartu behar dira kontuan::

      - Ordenak garrantzia du.
      - Ez dute aipatzen elementu guztiak erabiltzen direnik.
      - Mutil bakoitzak opari bakarra jaso dezake; beraz, ez dira errepikatzen.

      $$V_n^m = \frac{n!}{(n-m)!} \rightarrow V_5^3 = \frac{5!}{(5-3)!} = 60$$

   b. **Mutil bakoitzari opari bat baino gehiago toka dakioke.**

      - Ordenak garrantzia du.
      - Ez dute aipatzen elementu guztiak erabiltzen direnik..
      - Mutil bakoitzari opari bat baino gehiago toka dakioke; beraz, elementuak errepikatu egin daitezke.

      $$VR_n^m = n^m \rightarrow VR_5^3 = 5^3 = 125$$

   c. **Mutil bakoitzak opari bakarra jaso dezake, baina hirurak berdinak dira.**

      - Ordenak ez du axola, elementuak berdin-berdinak baitira.
      - Ez dute aipatzen elementu guztiak erabiltzen direnik.
      - Opari bakarra jaso dezake mutil bakoitzak, beraz, ez dira elementuak errepikatzen.

      $$C_n^m = \frac{n!}{(n-m)!\,m!} \rightarrow C_5^3 = \frac{5!}{(5-3)!\,3!} = 10$$

2. **Zifra bakoitiekin, hiru zifrako zenbat zenbaki osa daitezke zifrak errepikatu ahal izanez gero?**

Zenbaki bakoitiak 1, 3, 5, 7 eta 9 dira, eta hirunaka bildu nahi ditugu. Ordenak eragina duenez eta zifrak errepika daitezkeenez, errepikatu egiten dira. Beraz, emaitza honako hau da:

$$VR_5^3 = 5^3 = 125$$

Hau da, 125 aukera daude.

**3. Belenek 4 pertsona aukeratu behar ditu, dituen 20 hautagaien artean, bere lantaldea osatzeko. Zenbat modutan egin dezake hautaketa?**

Aukeraketa egiteko ordenak ez du eraginik:

$$C_{20}^4 = \frac{20!}{(20-4)!\ 4!} = 4845$$

4845 modu ditu hautaketa egiteko.

**4. Zortzi txirrindulari bidegorritik doaz ilaran. Zenbat modutan ordenatu daitezke?**

Ilarako ordenak garrantzia du eta elementu guztiak erabiltzen dira. Beraz, permutazio bat da:

$$P_8 = 8! = 40320$$

40320 modu desberdinetan jar daitezke.

**5. Itsasontzi baten seinale-makilan hiru bandera gorri, bi urdin eta lau berde igo daitezke. Zenbat seinale desberdin adieraz daitezke bederatzi banderak jarrita?**

Hemen elementu guztiak erabiltzen ditugu, ordenak garrantzia du eta hainbat aldiz errepikatzen diren elementuak daude. Hau da, errepikapena duen permutazioa da:

$$PR_9^{3,\,2,\,4} = \frac{9!}{3!\ 2!\ 4!} = 1260 \text{ aukera}$$

**6. Telefono-zenbaki batek zazpi zifra oso ditu. Demagun:**

- **Lehen zifrak 2 eta 9 arteko zenbakia izan behar du, biak barne.**
- **Bigarren eta hirugarren zifrak 1etik 9ra bitartekoak izan behar dira, biak barne.**

- **Gainerako zifra bakoitza 0 eta 9 arteko zenbaki bat da, biak barne.**

Beraz, jarritako baldintza bakoitzerako errepikapenezko aldakuntzak dira, ordenak eragina duelako, elementu guztiak erabiltzen ez direlako eta errepikatu egin daitezkeelako. Errepikapenezko aldakuntzen formula honako hau da: $V_n^m = n^m$.

- Lehenengo zifrarako 8 kasu posible ditugu, izan ere:

$$V_8^1 = 8^1 = 8$$

- Bigarren eta hirugarrenerako biak batera, hauek ditugu:

$$V_9^2 = 9^2 = 81$$

- Hurrengo zenbakietarako:

$$V_{10}^4 = 10^4 = 10000$$

Beraz, hurrengo aukerak ditugu $8 \cdot 81 \cdot 10000 = 6480000$.

7. **Fabrika batean, 5 pieza mota daude, produktu bat mihiztatzeko erabil daitezkeenak: A, B, C, D eta E. Ekoizpen-ingeniari batek 8 piezako multzo bat hautatu behar du prototipo bat eraikitzeko, baina ez da beharrezkoa pieza guztiak desberdinak izatea. Gainera, pieza mota bakoitzaren erabilgarritasun mugagabea dago.**

Aukeraketa egiteko ordenak ez du eraginik, eta piezak errepika daitezke; beraz, errepikapena duten konbinazioak dira.

$$\mathrm{CR}_5^8 = \binom{5+8-1}{8} = \frac{(5+8-1)!}{(5-1)!\,8!} = 495$$

495 modu daude 5 motako 8 piezako multzo bat hautatzeko, errepikapena baimentzen denean.

## 10.3 Probabilitatea

1. **Arte eszenikoen akademia batean dantza- eta antzerki-eskolak ematen dira. Dantzakoa, dantza klasikoko modalitatea eta kabareta daude. Akademian, %17k dantza klasikoa egiten du, %45ek kabareta eta %5ek bi modalitateak. Akademia horretara doan pertsona**

**bat aukeratzen badugu:**

Gertaeren arabera: A = "Dantza klasikoa praktikatzea", B = "Kabareta praktikatzea"eta T = "Antzerkia praktikatzea".

a. **Kalkulatu dantza motaren bat (edo biak) egiteko probabilitatea.**

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = 0,17 + 0,45 - 0,05 = 0,57$$

b. **Kalkulatu antzerkia bakarrik praktikatzeko probabilitatea.**

$$p(T) = p(\overline{A \cup B}) = 1 - 0,57 = 0,43$$

2. **Kalitateko sail batean ibilgailu elektriko eta hibridoen motorraren softwarearen funtzionamendua aztertzen da. 85 auto elektriko eta 145 auto hibrido aztertu ziren. Guztira, 43 autok akatsak zituzten beren motorren softwarean. Gainera, software akastuneko motorretatik 12 auto elektrikoei zegozkien.**

Datuak sarrera bikoitzeko taula batean antolatuko ditugu.

| | Elektrikoak (E) | Hibridoak (H) | GUZTIRA |
|---|---|---|---|
| Akasduna (D) | 12 | 31 | 43 |
| Ez akasduna ($\overline{D}$) | 73 | 114 | 187 |
| GUZTIRA | 85 | 145 | 230 |

a. **Kalkulatu ausaz hautatutako auto bat hibridoa izateko eta zure motorraren softwarea egokia izateko probabilitatea.**

Aztertutako 230 autoetatik 114 hibridoak eta motor softwarea egokia dutenak dira:

$$p(H \cap \overline{D}) = \frac{114}{230} = 0,4957$$

b. **Kalkulatu ausaz hautatutako auto hibrido batek motorraren softwarea akastuna izateko probabilitatea.**

145 auto hibridoen artean, 31ak motorraren softwarea akasduna dute:

$$p(D/H) = \frac{31}{145} = 0,2138$$

3. **Ikasle batek bi proba egiten ditu egun berean. Lehenengo proba gainditzeko probabilitatea 0.6koa da; bigarrena gainditzekoa**

**0.8koa, eta biak gainditzekoa 0.5ekoa. Aurkitu aukera hauek:**

Izan bitez hurrengo gertaerak:

- A = "lehenengoa gainditu' $\rightarrow$ P(A) = 0.6
- B = "bigarrengoa gainditu" $\rightarrow$ P(B) = 0.8
- A $\cap$ B = "A eta B frogak gainditu" $\rightarrow$ P(A $\cap$ B) = 0.5

**a. Gutxienez froga bat gainditzea.**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.8 - 0.5 = 0.9$$

**b. Froga bat ere ez gainditzea.**

$$P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.9 = 0.1$$

**c. Frogak, gertaera independenteak dira?**

A eta B gertaera independenteak izateko, hurrengoa bete behar da: $P(A \cap B) = P(A) \cdot P(B)$.

$$P(A \cap B) = 0.5 \neq 0.6 \cdot 0.8 = 0.48 = P(A) \cdot P(B)$$

Beraz, A eta B gertaerak, ez dira askeak.

**d. Lehenengoa gainditu ezean, bigarren proba pasatzea.**

$$P(B/\overline{A}) = \frac{P(B \cap \overline{A})}{P(\overline{A})} = \frac{0.3}{0.4} = 0.75$$

$B = (A \cap B) \cup (\overline{A} \cap B) \rightarrow P(B) = P(A \cap B) + P(\overline{A} \cap B) \rightarrow$ izanik, $0.8 = 0.5 + P(\overline{A} \cap B) \rightarrow P(\overline{A} \cap B) = 0.3$

**4. Aurpegiak 1etik 6ra zenbakituta dituen dado bat trukatuta dago, eta zenbaki bat lortzeko probabilitatea zenbaki horrekiko zuzenki proportzionala da.**

$$P(1) = x \quad P(2) = 2x \quad P(3) = 3x \quad P(4) = 4x \quad P(5) = 5x \quad P(6) = 6x$$

$$P(\Omega) = 1 = x + 2x + 3x + 4x + 5x + 6x = 21x \rightarrow x = \frac{1}{21}$$

**a. Aurkitu 3 ateratzeko probabilitatea, bakoitia atera dela jakinda.**

$$P(3/\text{bakoitia}) = \frac{P(3)}{P(1,3,5)} = \frac{\frac{3}{21}}{\frac{1}{21} + \frac{3}{21} + \frac{5}{21}} = \frac{3}{9} = \frac{1}{3}$$

b. **Kalkulatu bikoitia ateratzeko probabilitatea, 3 baino gehiago atera zela jakinda.**

$$P(\text{bikoitia/3 baino handiagoa}) = \frac{P(4,6)}{P(4,5,6)} = \frac{\frac{4}{21}+\frac{6}{21}}{\frac{4}{21}+\frac{5}{21}+\frac{6}{21}} = \frac{10}{15} = \frac{2}{3}$$

**5. Mikrotxipak ekoizteko instalazio batean, hiru txip mota sortzen dira: A mota, B mota eta C mota. Ekoitzitako txip bat mota bakoitzekoa izateko probabilitateak hauek dira:**

- **A mota:** $P(A) = 0,5$
- **B mota:** $P(B) = 0,3$
- **C mota:** $P(C) = 0,2$

**Gainera, txip mota bakoitzak akatsak izan ditzake:**

- **Txipa A motakoa bada, akastuna izateko probabilitatea** $P(D|A) = 0,02$**da**.
- **Txipa B motakoa bada, akastuna izateko probabilitatea** $P(D|B) = 0,05$ da.
- **Txipa C motakoa bada, akastuna izateko probabilitatea** $P(D|C) = 0,1$ da.

**Erantzun galdera hauei:**

a. **Zein da ausaz hautatutako txip bat akastuna izateko probabilitatea?**

Probabilitate Osoaren Legea erabiltzen dugu $P(D)$kalkulatzeko, txip bat akastuna izateko probabilitatea:

$$P(D) = P(D|A) \cdot P(A) + P(D|B) \cdot P(B) + P(D|C) \cdot P(C) =$$

$$= 0,02 \cdot 0,5 + 0,05 \cdot 0,3 + 0,1 \cdot 0,2 = 0,045$$

Beraz, txip bat akastuna izateko probabilitatea 0,045 edo % 4,5 da.

b. **Txip akastun bat hautatzen bada, zein da txip mota bakoitzarena izateko probabilitatea?**

Baldintzapeko probabilitatea erabiltzen dugu txip akastun bat mota bakoitzekoa izateko probabilitatea kalkulatzeko, hots, $P(A|D)$, $P(B|D)$, y $P(C|D)$.

$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)} = \frac{0,02 \cdot 0,5}{0,045} = 0,222$$

$$P(B|D) = \frac{P(D|B) \cdot P(B)}{P(D)} = \frac{0,05 \cdot 0,3}{0,045} = 0,333$$

$$P(C|D) = \frac{P(D|C) \cdot P(C)}{P(D)} = \frac{0,1 \cdot 0,2}{0,045} = 0,444$$

Beraz, txip akastun bat mota bakoitzekoa izateko probabilitateak gutxi gorabehera hauek dira:

- Tipo A: 22,2%.
- Tipo B: 33,3%.
- Tipo C: 44,4%.

## 10.4 Ausazko aldagai diskretuen banaketa

**1. Etxalde batean egunero ekoizten diren arrautzetatik, %20 bota behar dira kontsumitzeko egokiak ez direlako. Ausaz eta modu independentean 5 arrautza hautatzen dira:**

Banaketa binomiala da. Hauek dira gertaerak: A = "Arrautza ez da kontsumorako egokia"eta B = "Arrautza kontsumorako egokia da". Arrautza bat kontsumitzeko egokia ez izateko probabilitatea p(A) = p = 0,20 da, eta aurkakoaren probabilitatea (kontsumorako egokia da) p(B) = q = 0,80 da. Ausaz 5 arrautza aukeratzen direnez, banaketa binomial baten aurrean gaude B(5; 0,20).

**a. Kalkula ezazu hautatutako arrautzetako bat (gutxienez 1) botatzeko probabilitatea.**

Arrautzetako bat botatzea "denak kontsumitzeko egokiak diren gertakarien aurkakoa da".

$$\text{B(5; 0,80): } p[x = 5] = \binom{5}{5} \cdot 0,80^5 \cdot 0,20^0 = 0,32768$$

$$\text{gero: } p = 1 - 0,32768 = 0,67232$$

b. • **Zer da gertagarriagoa, gai ez diren 2 arrautza egotea, edo zehazki gai ez diren 3 arrautza egotea? Lortu probabilitate hauek.**

$$p[x=2] = \binom{5}{2} \cdot 0,20^2 \cdot 0,80^3 = \frac{5!}{2! \cdot 3!} \cdot 0,20^2 \cdot 0,80^3 = 0,2048$$

$$p[x=3] = \binom{5}{3} \cdot 0,20^3 \cdot 0,80^2 = \frac{5!}{3! \cdot 2!} \cdot 0,20^3 \cdot 0,80^2 = 0,0512$$

Gero, litekeena da kontsumorako egokiak ez diren 2 arrautza egotea.

• **Nola arrazoituko zenuke aurreko galderaren erantzuna kalkulagailua erabili gabe?**

Arrautza bat ausaz ateratzean, txikiagoa da kontsumitzeko egokia ez izateko probabilitatea, egokia izatearena baino. Orduan, litekeena da kontsumorako egokiak ez diren arrautzak ateratzea hiru ateratzea baino. Orduan, probabilitate handiago du kontsumitzeko egokiak ez diren bi arrautza ateratzea, hiru ateratzea baino.

**2. A gertatzeko probabilitatea 1/5 bada.**

**a. Zein da esperimentua errepikatu behar den gutxieneko aldia, gutxienez A gertakaria behin gertatzeko probabilitatea 1/2 baino handiagoa izan dadin?**

X = "n probetako arrakasta kopurua", X $\sim$ B(n, 0.2), p = 0.2, q = 0.8

$$\mathrm{P}(\mathrm{X} \geq 1) > \frac{1}{2} \rightarrow \mathrm{P}(\mathrm{X} < 1) \leq \frac{1}{2} \rightarrow \mathrm{P}(\mathrm{X} = 0) \leq \frac{1}{2}$$

$$P(X=0) = \binom{n}{0} \cdot 0.2^0 \cdot 0.8^n = 0.8^n \leq \frac{1}{2} \rightarrow (0.8^4 = 0.4096) \rightarrow n = 4$$

A gertaerak gutxienez 1/2 baino probabilitate handiagoa izateko, prozesua gutxienez 4 aldiz errepikatu behar da.

**b. Zein da esperimentua 5 aldiz egitean gutxienez A bi aldiz gertatzeko probabilitatea?**

B(5; 0.2) banaketa binomiala da.

$$P(X \geq 2) = 1 - [P(X=0) + P(X=1)] = 1 - \left[\binom{5}{0} \cdot 0.2^0 \cdot 0.8^5 + \right.$$

$$\left. + \binom{5}{1} \cdot 0.2^1 \cdot 0.8^4 \right] = 1 - (0.3277 + 0.4096) = 0.2627$$

**3. A gertaera bat gertatzeko probabilitatea 1/5 bada.**

**a. Zein da esperimentua errepikatu behar den aldi minimoa, gutxienez A gertaera behin gertatzeko probabilitatea 1/2 baino handiagoa izan dadin?**

$$\text{X} = \text{“exitu kopurua } n \text{ saiakeretan”, X} \sim \text{B(n, 0.2), p} = 0.2, \text{ q} = 0.8$$

$$\text{P(X} \geq 1) > \frac{1}{2} \rightarrow \text{P(X} < 1) \leq \frac{1}{2} \rightarrow \text{P(X} = 0) \leq \frac{1}{2}$$

$$P(X = 0) = \binom{n}{0} \cdot 0.2^0 \cdot 0.8^n = 0.8^n \leq \frac{1}{2} \rightarrow (0.8^4 = 0.4096) \rightarrow n = 4$$

A gertaerak gutxienez 1/2 baino probabilitate handiagoa izateko, gutxienez 4 aldiz errepikatu behar da prozesua.

**b. Zein da esperimentua 5 aldiz egitean gutxienez A bi aldiz gertatzeko probabilitatea?**

B banaketa binomiala da (5; 0.2)

$$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - \left[\binom{5}{0} \cdot 0.2^0 \cdot 0.8^5 + \right.$$

$$\left. + \binom{5}{1} \cdot 0.2^1 \cdot 0.8^4\right] = 1 - (0.3277 + 0.4096) = 0.2627$$

**4. Garraioen erabilerari buruzko inkesta batean, 50 pertsonak aukeratzen dute garraiobide gogokoena 4 aukeren artean: autoa, autobusa, bizikleta eta trena. Pertsona batek modu bakoitza aukeratzeko probabilitatea 0.4, 0.3, 0.2 eta 0.1 da, hurrenez hurren. Inkestatutako 50 pertsonetatik 20k nahiago dute autoa, 15ek autobusa, 10ek bizikleta eta 5ek trena?**

Hau da banaketa multinazionalaren kasua.

$$P(X_1 = 20, X_2 = 15, X_3 = 10, X_4 = 5) =$$

$$= \frac{50!}{20! \cdot 15! \cdot 10! \cdot 5!} \cdot 0.4^{20} \cdot 0.3^{15} \cdot 0.2^{10} \cdot 0.1^5 = 0.00355$$

**5. Ingeniari batek saiakera bakoitzean 0.2 huts egiteko probabilitatea duen makina bat probatzen du. $X$ ausazko aldagaia honela definitzen du: lehen hutsegitea behatu arteko saiakera-kopurua.**

$p = 0.2$ parametroa duen banaketa geometrikoa da.

a. **Zein da hirugarren saiakeran lehenengoak huts egiteko probabilitatea?**

$$P(X = 3) = (1 - 0.2)^{3-1} \cdot 0.2 = 0.8^2 \cdot 0.2 = 0.128$$

Beraz, lehenengoak hirugarren saiakeran huts egiteko probabilitatea %12.8 da.

b. **Zein da lehenengoak huts egiteko probabilitatea bosgarren saiakeran edo saiakera baino lehen?**

$$P(X \leq 5) = P(X = 1)+P(X = 2)+P(X = 3)+P(X = 4)+P(X = 5) =$$
$$= (1-0.2)^{1-1}\cdot 0.2+(1-0.2)^{2-1}\cdot 0.2+(1-0.2)^{3-1}\cdot 0.2+(1-0.2)^{4-1}\cdot 0.2+$$
$$+(1 - 0.2)^{5-1} \cdot 0.2 = 0.67232$$

**6. Gai elektrikoen denda batek 20 xafla ditu, eta horietatik 5 horiak dira. Ausaz eta ordezkapenik gabe ateratzen badira 10 xafla, zein da horietako bi horiak izateko probabilitatea?**

Plantxak ausaz ateratzen direnez eta gero ordezkapenik (ordezkapenik) ez dutenez, banaketa hipergeometrikoa da. Xafla bat horia izateko probabilitatea hau da: $p = \frac{5}{20} = 0.25$.

$$P(X = 2) = \frac{\binom{20 \cdot 0.25}{2}\binom{20 \cdot (1 - 0.25)}{10 - 2}}{\binom{20}{10}} = 0.3483$$

**7. Neurketa-gailu jakin batek gehiegizko desbiderapena erakusteko probabilitatea 0.05ekoa bada, zein da probabilitatea?**

a. **Probatu diren neurgailu horietako seigarrena gehiegizko desbiderapena erakusten duen hirugarrena da.**

Izan bedi:

- x = 6 neurgailu.
- r = 3 gailu, gehiegizko desbideratzea erakusten dutenak.
- p = p (gehiegizko desbiderapena erakusten duen gailua) = 0.05
- q = p (gehiegizko desbideratzerik erakusten ez duen gailua) = 0.95

$$P(X = 6) = \binom{6 - 1}{3 - 1}(1 - 0.05)^{6-3}0.05^3 = 0.001072$$

b. **Probatu diren neurgailu horietako zazpigarrena izan dadila gehiegizko desbideratzerik erakusten ez duen laugarrena.**

Sea:

- x = 7 neurgailu.
- r = Desbiderapen handiegia duten 4 gailu.
- p = p (gehiegizko desbideratzerik erakusten ez duen gailua) = 0.95
- q = p (gehiegizko desbiderapena erakusten duen gailua) = 0.05

$$P(X = 7) = \binom{7-1}{4-1}(1 - 0.95)^{7-4}0.95^4 = 0.002036$$

8. **Failetan mascletà batean (zarata handiz lehertzen diren su artifizialak) pertsona bat konortea galtzeko probabilitatea 0.001 da. San Jose egunean 5.000 pertsona inguru joaten direla mascletà ikustera kontuan hartuta, zein da 25 pertsona zorabiatzeko probabilitatea?**

Poissonen banaketa bat da probabilitatea 0 ingurukoa delako eta laginaren tamaina $\infty$-ra hurbil daitekeelako. Horregatik, $m = 5000 \cdot 0.001 = 50$.

$$P(X = 25) = \frac{50^{25} \cdot e^{-50}}{25!} = 3.70 \cdot 10^{-5}$$

## 10.5 Ausazko aldagai jarraituen banaketa

1. **Herri bateko jaioberrien pisuak 3300 gramoko batez besteko banaketa normala eta 465 gramoko desbiderapen tipikoa jarraitzen du. Jaioberri batek pisu txikia du 2.500 gramotik beherako pisua badu.**

   a. **Zein da herri honetan jaioberri batek pisu txikia izateko probabilitatea?**

   Eskura ditugun datuak $\mu$= 3300 g-ko batez bestekoa eta $\sigma$= 465 g-ko desbiderapen tipiko duen banaketa normal bat jarraitzen dute, hau da, N (3300, 465).

   N(0, 1) banaketaren taula erabiltzeko aldagaia tipifikatu behar dugu. $x \in N(3300,\ 465)$ orduan $z \in N(0,\ 1)$:

$$P[x \leq 2500] = P\left[z \leq \frac{2500 - 3300}{465}\right] = P[z \leq -1,72] = P[z \geq -1,72] =$$

$$= 1 - P[z \leq 1,72] = 1 - 0,9573 = 0,0427$$

**b. Zein da herri honetan jaioberri batek 3500 eta 4.000 gramo arteko pisua izateko probabilitatea?**

Aldagaiak N(3300, 465) banaketa normala jarraitzen dute. Tipifikatu behar dira:

$$P[3500 \leq x \leq 400] = P\left[\frac{3500 - 3300}{465} \leq z \leq \frac{4000 - 3300}{465}\right] =$$

$$= P[0,43 \leq z \leq 1,51] = P[z \leq 1,51] - P[z \leq 0,43] = 0,9345 - 0,6664 = 0,2681$$

**2. Emakume heldu baten serum-burdinaren kantitateak batez bestekoa 120 $\mu g/dl$ eta 30 $\mu g/dl$ desbiderapen tipikoaren banaketa normala jarraitzen du. Burdina faltagatik emakume batek anemia mota bat duela ulertuko da bere burdin kantitatea 75 $\mu g/dl$-ra iristen ez bada.**

**a. Zein da emakume heldu batek burdin faltagatik anemia izateko probabilitatea?**

Ditugun datuan banaketa normal bat jarraitzen dute. Bere batez bestekoa $\mu = 120$ $\mu g/dl$ eta desbiderapena $\sigma = 30$ $\mu g/dl$, hau da N(120, 30). N(0, 1) normalari dagokion taula erabiltzeko, aldagaia tipifikatu behar dugu, $x \in N(120,\ 30)$ normalera bihurtuz $z \in N(0,\ 1)$.

$$P[x \leq 75] = P\left[z \leq \frac{75 - 120}{30}\right] = P[z \leq -1,5] = P[z \geq 1,5] =$$

$$= 1 - P[z \leq 1,5] = 1 - 0,9332 = 0,0668$$

**b. Emakume helduen %45ek, $k$ baino burdin gehiago dute serumean. Bila ezazu k-ren balioa.**

Aldagaiak N(120, 30) banaketa normalari dagozkio; beraz, tipifikatu egin behar ditugu:

$$P[x > k] = P\left[\frac{x - 120}{30} > \frac{k - 120}{30}\right] = P\left[z > \frac{k - 120}{30}\right] = 0,45 \rightarrow$$

$$\rightarrow 1 - P\left[z < \frac{k - 120}{30}\right] = 0,45 \rightarrow P\left[z < \frac{k - 120}{30}\right] = 0,55 \rightarrow$$

$$\rightarrow \frac{k - 120}{30} = 0,125 \rightarrow k = 123,75\ \mu g/dl$$

3. **Emakume heldu baten serum-burdinaren kantitateak 120 $\mu g/dl$ eta 30 $\mu g/dl$ desbiderapen tipikoaren batez besteko banaketa normala jarraitzen du. Burdina faltagatik emakume batek anemia mota bat duela joko da bere burdin kantitatea 75 $\mu g/dl$-ra iristen ez bada.**

   a. **Zein da emakume heldu batek burdin faltagatik anemia izateko probabilitatea?**

   Ditugun datuak $\mu = 120\ \mu g/dl$ batez besteko banaketa normal batekoak dira, eta desbiderapen tipikoa $\sigma = 30\ \mu g/dl$, hau da, normal N (120, 30). N (0, 1) normalari dagokion taula erabiltzeko, $x \in N(120,\ 30)$ aldagaiak tipifikatu behar ditugu $z \in N(0,\ 1)$ aldagai bihurtuz.

$$P[x \leq 75] = P\left[z \leq \frac{75-120}{30}\right] = P[z \leq -1,5] = P[z \geq 1,5] =$$
$$= 1 - P[z \leq 1,5] = 1 - 0,9332 = 0,0668$$

   b. **Emakume helduen % 45ek k baino burdina gehiago dute serumean. Ikertu k-ren balioa**

   Aldagaiak N banaketa normalari dagozkio (120, 30); beraz, honela tipifikatu behar ditugu:

$$P[x > k] = P\left[\frac{x-120}{30} > \frac{k-120}{30}\right] = P\left[z > \frac{k-120}{30}\right] = 0,45 \rightarrow$$
$$\rightarrow 1 - P\left[z < \frac{k-120}{30}\right] = 0,45 \rightarrow P\left[z < \frac{k-120}{30}\right] = 0,55 \rightarrow$$
$$\rightarrow \frac{k-120}{30} = 0,125 \rightarrow k = 123,75\ \mu g/dl$$

4. **Unibertsitate bateko ikasleen garaierak banaketa normal bati jarraitzen dio: $\mu$= 170 zm eta desbiderapen estandarra $\sigma$= 10 zm.**

   a. **Kalkulatu ikasle batek 160 zm-tik beherako garaiera izateko probabilitatea.**

$$P[X < 160] = P\left[Z < \frac{160-170}{10}\right] = P[Z < -1] = P[Z > 1] =$$
$$= 1 - P[Z < 1] = 1 - 0.8413 = 0.1587$$

   Ikasle batek 160 zm-tik beherako garaiera izateko probabilitatea % 15.87 da.

**b. Zein da ikasle batek 165 zm eta 185 zm arteko garaiera izateko probabilitatea?**

$$P[165 < X < 185] = P\left[\frac{165-170}{10} < Z < \frac{185-170}{10}\right] = P[-0.5 < Z < 1.5] =$$

$$= P[Z < 1.5] - P[Z < -0.5] = P[Z < 1.5] - P[Z > 0.5] = P[Z < 1.5] -$$

$$-(1 - P[Z < 0.5]) = 0.9332 - (1 - 0.6915) = 0.6247$$

Ikasle batek 165 zm eta 185 zm arteko garaiera izateko probabilitatea % 62,47koa da.

**c. Aurki ezazu 90. perzentilari dagokion garaiera.**

90. pertzentilerako, $X$ balioa bilatuko dugu, $P[X < x_{90}] = 0.90$.

Lehenik eta behin, banaketa normal estandarrean 90eko pertzentilari dagokion $Z$ balioa aurkituko dugu:

$$Z_{0.90} \approx 1.28$$

Ondoren, $Z$balio hori $N(170, 10)$banaketa-eskalara aldatuko dugu:

$$Z = \frac{X - \mu}{\sigma} \rightarrow x_{90} = \mu + Z_{0.90} \cdot \sigma = 170 + 1.28 \cdot 10 = 182.8$$

90eko pertzentilari dagokion garaiera 182,8 zm-koa da, gutxi gorabehera.

**5. Chi-karratu banaketan, 20 graduko askatasunarekin, zenbat da 10. perzentila?**

$$\chi^2_{0.10;\,20} = P(\chi^5_{20} \geq 0.10) = P(\chi^5_{20} \leq (1 - 0.10)) = P(\chi^5_{20} \leq 0.90) = 12.443$$

**6. Snedecor-en F banaketan, zenbakitzailean 10 askatasun-gradu eta izendatzailean 10 gradu dituelarik, zer perzentil dagokio 2.3226 balioari?**

*Snedecor-en F banaketan* delakoaren VII. taulan, $n_1$= 10 eta $n_2$= 10 kasuetarako, ikusten da 3.3226 balioa $\alpha$= 0.10 dela. Horrek esan nahi du aldagaiak behaketen % 90 azpitik uzten duela, eta, beraz, 90. pertzentila dagokiola ($P_{90}$).

## 10.6 Laginketa eta Estimazioa

**1. Enpresa batek bere langileek hilero prestakuntza-zereginetan ematen duten batez besteko denbora (ordutan) aztertu nahi du. Enpresak 200 langile ditu guztira. 30 langileko ausazko lagin sinple bat hartzen da, eta prestakuntza-denborari buruzko datu hauek lortzen dira ordutan:**

4.5, 6.3, 3.2, 5.1, 7.5, 4.8, 6.0, 5.5, 3.9, 4.2, 5.6, 4.3, 6.1, 5.8, 4.0, ,7.0, 4.7, 3.6, 5.2, 4.4, 6.5, 4.1, 5.3, 3.7, 6.4, 5.7, 4.9, 3.8, 5.4, 6.2

**a. Kalkulatu prestakuntza-denboraren batez besteko lagina.**

$$\overline{x} = \frac{\sum x}{n} = 4.81$$

**b. Kalkulatu lagin-bariantza eta lagin-desbideratzea.**

$$s^2 = \frac{\sum (x - \overline{x})^2}{n-1} \approx 1.394 \rightarrow s \approx 1.18$$

**c. Lortutako datuak erabiliz, kalkulatu prestakuntza-denboraren batez besteko populazioa.**

$$\mu_{\overline{x}} = \overline{x} = 4.81$$

**d. Zein izango litzateke laginen batez bestekoen bariantza 30 langileren laginak behin eta berriz hartuz gero?**

$$\sigma^2_{\overline{x}} = \frac{s^2}{n} \cdot \frac{N-n}{N-1} = \frac{1.394}{30} \cdot \frac{200-30}{200-1} \approx 0.0429$$

**2. Landaketa batean, 1000 fruta-arbola daude, eta ekoiztutako fruituen guztizko kopurua kalkulatu nahi da. 50 zuhaitzeko ausazko lagin sinple bat aukeratzen da, eta horietako bakoitzean dauden fruituak kontabilizatzen dira; batez beste, 120 fruitu lortzen dira zuhaitz bakoitzeko, eta 15 frutuko desbideratze estandarra. Kalkulatu landaketa osoan sortutako fruituen zenbatetsitako batura.**

$$\mu = \overline{x} \cdot n = 120 \cdot 1000 = 120000$$

**3. Gozoki fabrika batek 100 gramo pisatu behar duten txokolate barrak ekoizten ditu. 50 barreko ausazko lagina hartzen da, eta batez besteko pisua 101 gramokoa da. $\sigma^2$ 4 gramokoa da. Aurkitu konfiantza-tartea:$\%95$ txokolate-barren batez besteko pisurako.**

- Laginaren tamaina ($n$) = 50
- Laginaren batez bestekoa ($\overline{x}$) = 101 gramo
- Populazio-bariantza ($\sigma^2$) = 4 gramo $\rightarrow$ $\sigma$ =2 gramo
- Konfiantza-maila = % 95

Emandako datuekin, konfiantza-tarte bat da biztanleria normal baten batez bestekoarentzat, eta ezagutzen den biztanleria-bariantza eta lagin handi bat ditu. Konfiantzaren % 95erako, $\alpha$= 1 - 0,95 = 0,05.

$$T_\mu = \left[\overline{x} - z_{0.05/2} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{0.05/2} \cdot \frac{\sigma}{\sqrt{n}}\right] =$$

$$= \left[101 - 1.96 \cdot \frac{2}{\sqrt{50}},\ 101 + 1.96 \cdot \frac{2}{\sqrt{50}}\right] = [100.446, 101.554]$$

% 95eko konfiantzarekin, txokolate-barren batez besteko pisua 100,45 gramo eta 101,55 gramo artean dagoela esan dezakegu.

**4. Irakurtzeko ohituren azterketa batek neurtzen du zenbat liburu irakurtzen dituzten hilean unibertsitate bateko ikasleek. 100 ikasleko lagin bat hartzen da, eta, batez beste, irakurritako 5 liburu eta 1,5 liburuko laginketa-desbideratze estandarra lortzen dira. Kalkulatu % 90eko konfiantza-tartea ikasleek hilean irakurritako liburuen batez besteko kopururako.**

- Laginaren tamaina ($n$) = 100
- Laginaren batez bestekoa ($\overline{x}$) = 5 liburu
- Laginaren desbiderapena ($s$) = 1.5 liburu
- Konfidantza maila = 90%

Emandako datuekin, konfiantza-tarte bat da biztanleria normal baten batez bestekoarentzat, eta biztanleriaren bariantza ezezaguna eta lagin handia ditu. Konfiantzaren % 90erako, $\alpha$= 1 - 0,90 = 0,10.

$$T_\mu = \left[\overline{x} - z_{0.1/2} \cdot \frac{s}{\sqrt{n}},\ \overline{x} + z_{0.1/2} \cdot \frac{s}{\sqrt{n}}\right] =$$

$$= \left[5 - 1.645 \cdot \frac{1.5}{\sqrt{100}},\ 5 + 1.645 \cdot \frac{1.5}{\sqrt{100}}\right] = [4.753, 5.247]$$

% 90eko konfiantzarekin, esan dezakegu unibertsitate honetako ikasleek irakurri dituzten liburuen batez besteko kopurua hileko 4.753 eta 5.247 liburu artean dagoela.

5. **Ikertzaile batek bateria-mota jakin batek erresistentzia proba batean irauten duen batez besteko denbora kalkulatu nahi du. 9 bateria probatzen dira, 18 orduko batez besteko denbora eta 2 orduko desbiderapena aurkituz. Aurkitu konfiantza-tartea: % 95 baterien batez besteko iraupenerako.**

   - Laginaren tamaina ($n$) = 9
   - Laginaren batez bestekoa ($\overline{x}$) = 18 ordu
   - Laginaren desbiderapena ($s$) = 2 ordu
   - Konfidantza maila = 95%

   Emandako datuekin, konfiantza-tarte bat da biztanleria normal baten batez bestekoarentzat, eta populazio-bariantza ezezaguna eta lagin txikia ditu. Konfiantzaren % 95erako, $\alpha$= 1 - 0,95 = 0,05.

$$T_\mu = \left[\overline{x} - t_{0.05/2;\,9-1} \cdot \frac{s}{\sqrt{n}},\ \overline{x} + t_{0.05/2;\,9-1} \cdot \frac{s}{\sqrt{n}}\right] =$$

$$= \left[18 - 2.306 \cdot \frac{2}{\sqrt{18}},\ 18 + 2.306 \cdot \frac{2}{\sqrt{18}}\right] = [16.913, 19.087]$$

   % 95eko konfiantzarekin, baterien batez besteko iraupena populazioan 16.913 eta 19.087 ordu artean dago.

6. **Bi hiritako langileen soldatak aztertu dira. A hiriko 200 langileko lagin batean, batez besteko soldata 1200 €-koa da, eta ezagutzen den desbideratze estandarra 100 €-koa da. B hiriko 150 langileko lagin batean, batez besteko soldata 1150 €da, 120 €-ko desbideratze estandarrarekin. Kalkula ezazu % 95eko konfiantza-tartea bi hirien arteko batez besteko soldaten diferentziarako.**

   - A hiriaren datuak:
     - Laginaren tamaina ($n_A$) = 200
     - Laginaren batez bestekoa ($\overline{x}_A$) = 1200
     - ($\sigma_A$) = 100
   - B hiriaren datuak:
     - Laginaren tamaina ($n_B$) = 150
     - Laginaren batez bestekoa ($\overline{x}_B$) = 1150
     - ($\sigma_B$) = 120

   Emandako datuekin, konfiantza-tarte bat da bi populazio normalen batez bestekoen diferentziarako, eta populazio-bariantza ezagunak eta lagin handiak ditu. Konfiantzaren % 95erako, $\alpha$= 1 - 0,95 = 0,05.

$$T_{\mu_A-\mu_B} = \left[(\overline{x_A} - \overline{x_B}) \mp z_{0.05/2} \cdot \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_2^B}{n_B}}\right] =$$

$$= \left[(1200 - 1150) \mp 1.96 \cdot \sqrt{\frac{100^2}{200} + \frac{120^2}{150}}\right] = [26.33, 73.67]$$

% 95eko konfiantzarekin, bi hirien arteko batez besteko soldaten arteko aldea 26,33 € eta 73,67 € artean dago.

**7. Ikasleek ariketan ematen duten denborari buruzko ikerketa batean, bi unibertsitatetako ikasleen laginak hautatzen dira. X unibertsitatean, 100 ikaslek astean 6 orduko ariketa dute batez beste, 1,8 orduko laginketa-desbideratze estandarrarekin. Y unibertsitatean, 120 ikaslek 5 orduko batez bestekoa ematen dute, 2 orduko lagin-desbideratze estandarrarekin. Kalkula ezazu % 90eko konfiantza-tartea bi unibertsitateek ariketan ematen duten denboraren arteko diferentziarako.**

- X unibertsitatearen datuak:
  - Laginaren tamaina ($n_X$) = 100
  - ($\overline{x}_X$) = 6 ordu
  - ($s_X$) = 1,8 ordu
- Y unibertsitatearen datuak:
  - Laginaren tamaina ($n_Y$) = 120
  - Laginaren batez bestekoa ($\overline{x}_Y$) = 5 ordu
  - Laginaren desbiderapena ($s_Y$) = 2 ordu

Emandako datuekin, bi populazio normalen batez bestekoen diferentziarako konfiantza-tarte bat da, eta populazio-bariantza ezezagunak eta lagin handiak ditu. Konfiantzaren % 90erako, $\alpha$= 1 - 0,90 = 0,10.

$$T_{\mu_X-\mu_Y} = \left[(\overline{x_X} - \overline{x_Y}) \mp z_{0.10/2} \cdot \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}\right] =$$

$$= \left[(6 - 5) \mp 1.645 \cdot \sqrt{\frac{1.8^2}{100} + \frac{2^2}{120}}\right] = [0.578, 1.422]$$

% 90eko konfiantzarekin, X Unibertsitateko eta Y Unibertsitateko ikasleen artean ariketari eskainitako batez besteko denboraren aldea 0.578 eta 1.422 ordu artean dagoela baiezta dezakegu. Horrek iradokitzen du, batez beste,

X Unibertsitateko ikasleek astean 0.578 eta 1.422 ordu bitartean ematen dituztela Y Unibertsitateko ikasleen ariketan.

**8. Bezeroari arreta emateko bi zerbitzuren erantzun-denborak alderatzen dira. Zerbitzu bakoitzeko 12 erantzunen laginak hautatzen dira. A zerbitzuan, batez besteko denbora 2,5 minutukoa da, 0,8 minutuko desbideratze estandarrarekin. B zerbitzuan, batez besteko denbora 3 minutukoa da, 0,7 minutuko desbideratze estandarrarekin. Bariantzak bi populazioetan berdinak direla suposatuz. Aurkitu konfiantza-tartea, % 95, bi zerbitzuen arteko batez besteko erantzun-denboraren diferentziarako.**

- A zerbitzuaren datuak:
  - Laginaren tamaina ($n_A$) = 12
  - Laginaren batez bestekoa ($\overline{x}_A$) = 2,5 minutu
  - Laginaren desbiderapena ($s_A$) = 0,8 minutu
- B zerbitzuaren datuak:
  - Laginaren tamaina ($n_B$) = 12
  - Laginaren tamaina ($\overline{x}_B$) = 3 minutu
  - Laginaren desbiderapena ($s_B$) = 0,7 minutu

Emandako datuekin, konfiantza-tarte bat da bi populazio normalen batez bestekoen diferentziarako, eta populazio-bariantza ezezagunak baina berdinak eta lagin handiak ditu. Konfiantzaren % 95erako, $\alpha$= 1 - 0,95 = 0,05.

$$T_{\mu_A-\mu_B} = \left[(\overline{x_A} - \overline{x_B}) \mp t_{0.05/2;\ 12+12-2} \cdot \sqrt{\frac{(n_A-1)\cdot s_A^2 + (n_B-1)\cdot s_B^2}{n_A+n_B-2}} \cdot \sqrt{\frac{1}{n_A}+\frac{1}{n_B}}\right] =$$

$$= \left[(2.5-3) \mp 2.074 \cdot \sqrt{\frac{(12-1)\cdot 0.8^2 + (12-1)\cdot 0.7^2}{12+12-2}} \cdot \sqrt{\frac{1}{12}+\frac{1}{12}}\right] = [-1.137, 0.137]$$

%95eko konfiantzarekin, A zerbitzuaren eta B zerbitzuaren arteko batez besteko erantzun-denboren aldea -1.137 eta 0.137 minutu artean dago. Horrek adierazten du, batez beste, A zerbitzua B zerbitzua baino 1.137 minutu azkarragoa izan daitekeela, edo B zerbitzua baino 0.137 minutu motelagoa izan litekeela. Tarteak 0 balioa hartzen duenez, horrek iradokitzen du ez dagoela alde estatistikoki esanguratsurik bi zerbitzuen erantzun-denboretan.

**9. Enpresa batek bi trebakuntza-metodo desberdinen eraginkortasuna ebaluatzen du. A metodoa erabiltzen duten 10 langileko lagin bat hartzen da, ondorengo azterketa batean 85eko batez besteko puntuazioarekin, eta 5eko desbiderapen estandarrarekin. B metodoa erabiltzen duten 8 langileko beste lagin bat hartu da, 80ko batez besteko ziztadarekin eta 6ko desbiderapen estandarrarekin. Kalkula ezazu %95eko konfiantza-tartea bi gaikuntza-metodoen arteko batez besteko puntajeen diferentziarako.**

- A metodoaren datuak:
  - Laginaren tamaina ($n_A$) = 10
  - Laginaren batez bestekoa ($\overline{x}_A$) = 85
  - Laginaren desbiderapena ($s_A$) = 5
- B metodoaren datuak:
  - Laginaren tamaina ($n_B$) = 8
  - Laginaren batez bestekoa ($\overline{x}_B$) = 80
  - Laginaren desbiderapena ($s_B$) = 6

Emandako datuekin, bi populazio normalen batez bestekoen diferentziarako konfiantza-tarte bat da, eta populazio-bariantza ezezagunak eta lagin txikiak ditu. Konfiantzaren % 95erako, $\alpha$= 1 - 0,95 = 0,05.

$$T_{\mu_A-\mu_B} = \left[(\overline{x_A} - \overline{x_B}) \mp t_{0.05/2;\, m} \cdot \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}\right]$$

$$c = \frac{\frac{s_A^2}{n_A}}{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = \frac{\frac{5^2}{10}}{\frac{5^2}{10} + \frac{6^2}{8}} = \frac{5}{14} \approx 0.3571$$

$$\frac{1}{m} = \frac{c^2}{n_A - 1} + \frac{(1-c)^2}{n_B - 1} \rightarrow \frac{1}{m} = \frac{0.3571^2}{10-1} + \frac{(1-0.3571)^2}{8-1} \rightarrow m = 13.65 \approx 13$$

$$T_{\mu_A-\mu_B} = \left[(85-80) \mp 2.160 \cdot \sqrt{\frac{5^2}{10} + \frac{6^2}{8}}\right] = [-0.715, 10.715]$$

% 95eko konfiantzarekin, esan dezakegu bi trebakuntza-metodoen arteko batez besteko puntuen aldea -0.715 eta 10.715 puntu artean dagoela. Horrek adierazten du, A metodorako batez besteko puntuazioa altuagoa dela dirudien arren, konfiantza-tarteak balio negatiboak dituela, eta horrek iradokitzen du ez dagoela estatistikoki esanguratsua den alderik bi metodoen artean.

**10. Makina batek 5 zm-ko diametro izendatua duten piezak ekoizten ditu. Doitasuna egiaztatzeko, 15 piezako lagin bat hartu eta diametroa neurtzen da. Laginaren bariantza 0.04 zm $^2$-koa da. Kalkula ezazu % 95eko konfiantza-tartea makinak sortutako piezen diametroaren bariantzarako.**

- Laginaren tamaina ($n$) = 15
- Laginaren bariantza ($s^2$) = 0.04 zm$^2$
- Konfidantza maila = 95%

Emandako datuekin, konfiantza-tarte bat da biztanleria normal baten bariantzarako. Konfiantzaren % 95erako, $\alpha$= 1 - 0,95 = 0,05.

$$T_{\sigma^2} = \left[ \frac{(n-1)\cdot s^2}{\chi^2_{0.05/2;\ 15-1}},\ \frac{(n-1)\cdot s^2}{\chi^2_{1-0.05/2;\ 15-1}} \right] =$$

$$= \left[ \frac{(15-1)\cdot 0.04}{26.119},\ \frac{(15-1)\cdot 0.04}{5.629} \right] = [0.0214, 0.0995]$$

% 95eko konfiantzarekin, makinak sortutako piezen diametroaren bariantza 0.0214 cm $^2$eta 0.0995 cm $^2$artean dagoela esan dezakegu. Horrek esan nahi du, 15 piezako lagin asko hartuko bagenitu, modu berean kalkulatutako tarteen % 95ek makina honen ekoizpeneko diametroaren benetako bariantza izango lukeela.

**11. Bi hornitzaile entregatzeko denboraren aldakortasunak alderatzen dira. A hornitzailearen 12 entregako lagin batean, bariantza 16 minutukoa da $^2$; B hornitzailearen 15 entregako lagin batean, berriz, bariantza 25 minutukoa da $^2$. Kalkula ezazu % 90eko konfiantza-tartea bi hornitzaileen arteko entrega-denboraren bariantzen arrazoirako.**

- A hornitzailearen datuak:
  - Laginaren tamaina ($n_A$) = 12
  - Laginaren bariantza ($s^2_A$) = 16 minutu
- B hornitzailearen datuak:
  - Laginaren tamaina ($n_B$) = 15
  - Laginaren bariantza ($s^2_B$) = 25 minutu

Emandako datuekin, konfiantza-tarte bat da bi populazio normalen bariantzen arrazoirako. Konfiantzaren % 90erako, $\alpha$= 1 - 0,90 = 0,10.

$$I = \left[\frac{s_A^2}{s_B^2} \cdot \frac{1}{F_{0.1/2;\ 12-1;\ 15-1}},\ \frac{s_A^2}{s_B^2} \cdot F_{0.1/2;\ 12-1;\ 15-1}\right] =$$
$$= \left[\frac{16}{25} \cdot \frac{1}{2.484},\ \frac{16}{25} \cdot 2.484\right] = [0.258, 1.589]$$

% 90eko konfiantzarekin, esan dezakegu bi hornitzaileen arteko entrega-denboraren bariantzen arrazoia 0.2577 eta 1.5898 artekoa dela. Horrek iradokitzen du bi hornitzaileen entrega-denboraren bariantzak desberdinak izan daitezkeen arren, hornitzaile baten entrega-denboren aldakortasuna ez dela konfiantza-tarte horren barruan bestearena baino nabarmen handiagoa.

## 10.7 Hipotesi-kontrasteak

1. **Badirudi oilasko okelaren kontsumoak gora egin duela duela hilabete batzuk beste haragi mota bati alarma piztu zitzaionetik. Harategi batean, oilasko-okelaren eguneroko salmentak 19 kilo erregistratu zituen batez beste ordura arte, eta 3 kiloko desbideratze tipikoa. Aipatutako alarmaren ondorengo 35 eguneko lagin batean, egunean saldutako 21 kilo oilasko haragi lortu ziren batez beste. Salmentek ohiko desbideratze bera duen banaketa normal bati jarraitzen diotela suposatuz:**

   a. **Test bat planteatzen da oilaskoaren salmenta handitu ez dela egiaztatzeko, datuek adierazten duten bezala. Zer ondoriotara iristen da %5eko garrantziarekin?**

   Bariantza ezaguna duen populazioaren batez bestekorako aldebakarreko kontrastea da. Hurrengo hipotesiak ezartzen dira:

   $$\text{Hipotesi nulua } H_o\text{: } \mu \leq 19 \text{ kg}$$
   $$\text{Kontrako hipotesia } H_i\text{: } \mu > 19 \text{ kg}$$

   $H_o$ onartzen da kontraste-estatistikoa, $z_o$, (-$\infty$; $z_\alpha$) onarpen-eremuaren barruan sartzen bada.

   Kontuan izanda: $\overline{x} = 21$ kg, n = 35 egun, $\sigma = 3$ kg eta $\alpha = 0.05$. $z_\alpha = z_{0.05} = 1.645$ izango dugu.

   Izan bedi $z_o = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{21 - 19}{3/\sqrt{35}} = 3.94$. Hipotesi nulua baztertu da, eta, ondorioz, hipotesi alternatiboa onartu da, esanez oilasko-haragiaren batez besteko kontsumoa handitu egin dela alarmarekin, % 95eko fidagarritasunarekin.

   b. **Kalkulatu %95eko konfiantza-tartea oilasko-okelaren eguneko batez besteko salmentarako.**

   Ezagutzen den bariantza-banaketaren batez besteko populazioarentzako konfiantza-tartea:

   $$T_\mu = \left[\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

   non $z_{\alpha/2} = z_{0.025} = 1.96$ beraz, $T_\mu = \left[21 \mp 1.96 \cdot \frac{3}{\sqrt{35}}\right] = [20.01;\ 21.99]$. Ikus dezakegu $19 \notin [20.01;\ 21.99]$, konfiantza tartetik kanpo dagoela.

**2. Fabrikatzaile batek ziurtatzen du baloien diametroak 22 zm-ko batez bestekoa duela, 2 zm-ko desbiderapen estandarrarekin. 100 baloiko ausazko lagin bat hautatzen da, eta 21,6 zm-ko batez bestekoa ikusten da. Baloien batez besteko diametroa zehaztutako balioaren desberdina dela ondoriozta dezakegu? % 5eko garrantzia erabiltzen du**

Batez besteko diametroa zehaztutako balioarekiko desberdina den frogatu nahi dugu; beraz, aldebiko proba bat dugu.

$$\text{Hipotesi nulua } H_o\text{: } \mu = 22$$
$$\text{Kontratko Hipotesia } H_1\text{: } \mu \neq 22$$

Ariketaren datuak:

- Populazioak duen batez besteko hipotetikoa ($\mu_0$) = 22 zm
- Populazioak duen desbiderapena ($\mu$) = 2 zm
- Laginaren tamaina ($n$) = 100
- Laginaren batez bestekoa ($\overline{x}$) = 21.6 zm
- Esangura maila $\alpha$ = 0,05 da

Populazioaren desbiderapen estandarra ezagutzen dugunez ($\sigma = 2$) eta laginaren tamaina handia denez ($n = 100$), Z proba erabiltzen dugu. Probako estatistikoa (Z) honela kalkulatzen da:

$$Z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{21.6 - 22}{2/\sqrt{100}} = -2$$

Alde biko proba denez, zati 2 egingo dugu, eta horrek $\alpha/2 = 0.025$ emango du banaketa normalaren kola bakoitzerako. Banaketa normal estandarraren taulan bilatuta, $\alpha/2$= 0.025 kantitaterako balio kritikoak 1,96 direla aurkituko dugu. Beraz, $z_0 \notin [-1.96, 1.96]$.

Z = -2 < -1,96 denez, balioa ez dago onarpen tartean. $H_0$ hipotesi nulua baztertuko dugu, % 5eko esangura-mailan. Horrek adierazten du nahikoa ebidentzia estatistiko dagoela baloien batez besteko diametroa 22 zm-koa ez dela ondorioztatzeko.

**3. Industria-instalazio batean, produktu bat fabrikatzeko batez besteko denbora 15 minutukoa da. 81 behaketako lagin batek 14,5 minutuko batez bestekoa eta 2,8 minutuko laginaren desbideratze estandarra erakusten ditu. Fabrikazio-denbora nabarmen txikiagoa al da? Erabili % 1eko garrantzia.**

Fabrikazio-denbora zehaztutako balioa (15 minutu) baino txikiagoa den jakin nahi dugunez, alde bakarreko proba bat dugu ezkerrean.

$$\begin{array}{c}\text{Hipotesi nulua } H_o\text{: } \mu \geq 15 \\ \text{Kontratko Hipotesia } H_1\text{: } \mu < \ 15\end{array}$$

Ariketaren datuak:

- Populazioak duen batezbesteko hipotetikoa ($\mu_0$) = 15 minutu
- Laginaren desbiderapena ($s$) = 2,8 minutu
- Laginaren tamaina ($n$) = 81
- Laginaren batez bestekoa ($\overline{x}$) = 14,5 minutu
- Esangura maila $\alpha = 0.01$ da

Populazioaren desbiderapen estandarra ezagutzen ez dugunez eta laginaren tamaina handia denez ($n = 81$), gutxi gorabeherako Z proba bat erabil dezakegu, laginaren desbideratze estandarra erabiliz. Z probako estatistikoa honela kalkulatzen da:

$$Z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{14.5 - 15}{2.8/\sqrt{81}} = -1.607$$

Ezkerreko aldebakarreko proba denez, banaketa normalaren kola batean $\alpha = 0.01$-ko garrantzi-maila baterako balio kritikoa bilatzen dugu. Banaketa normal estandarraren taulan, $\alpha = 0.01$-rako balioa 2,33 da, gutxi gorabehera. Beraz, $z_0 \in [-2.33, +\infty]$.

Z-ren balioa onarpen tartean dago; horregatik, $H_0$hipotesi nulua ez dugu baztertzen % 1eko adierazpen-mailara. Horrek adierazten du ez dagoela nahikoa ebidentzia estatistiko fabrikazio-denbora batez bestekoa 15 minutu baino txikiagoa dela ondorioztatzeko.

4. **Hegazti espezie bati buruzko ikerketa batean, batez besteko pasabidea 300 gramokoa dela baieztatzen da. 16 hegaztiren lagina hartzen da, eta 290 gramoko batez bestekoa eta 12 gramoko desbideratze estandarra lortzen da. Arrazoizkoa al da batez besteko pisua 300 gramotik beherakoa dela ondorioztatzea? % 5eko garrantzia erabiltzen du.**

Hipotesi nulua $H_o$: $\mu \geq 300$
Kontrako hipotesia $H_1$: $\mu < 300$

Ariketaren datuak:

- Populazioaren batez besteko hipotetikoa ($\mu_0$) = 300 gramo
- Laginaren desbideratze estandarra ($s$) = 12 gramo
- Laginaren tamaina ($n$) = 16
- Laginaren batez bestekoa ($\overline{x}$) = 290 gramo
- Emandako esangura-maila $\alpha$= 0.05 da

Populazioaren desbideratze estandarra ezagutzen ez dugunez eta laginaren tamaina txikia denez ($n = 16$), Studenten $t$proba erabiliko dugu. $t$probaren estatistikoa honela kalkulatzen da:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{290 - 300}{12/\sqrt{16}} = -3.33$$

Student-en $t$taulan, 15 askatasun-gradu dituen $\alpha = 0.05$-rako balioa -1,753 da. Beraz, $t_0 \notin [-1,753, +\infty]$.

$H_0$ hipotesi nulua errefusatuko dugu, % 5eko esangura-mailan. Horrek adierazten du nahikoa ebidentzia estatistiko dagoela hegaztien batez besteko pisua 300 gramotik beherakoa dela ondorioztatzeko.

5. **Azterlan batek bi konpainietako (A eta B.) enplegatuen artean zeregin bat betetzeko batez besteko denbora konparatzen du. A konpainiako 100 langileko lagin batean 30 minutuko batez bestekoa eta 5 minutuko desbideratze estandarra lortu ziren. B motako 80 langileko lagin batean, 32 minutuko batez bestekoa eta 4 minutuko desbideratze estandarra lortu ziren. Desberdina al da bi konpainien arteko batez besteko denbora? % 5eko garrantzia erabiltzen du.**

Ataza betetzeko batez besteko denbora bi konpainien artean desberdina den frogatu nahi dugu; beraz, aldebiko proba bat da.

Hipotesi nulua $H_o$: $\mu_A = \mu_B$
Kontrako hipotesia $H_1$: $\mu_A \neq \mu_B$

Ariketaren datuak:

- Garrantzi maila $\alpha$= 0,05 da.
- A konpainiaren datuak:
  - Laginaren tamaina $(n_A) = 100$
  - Laginaren batez bestekoa $(\overline{x}_A) = 30$ minutu
  - Laginaren desbideratze estandarra $(s_A) = 5$ minutu
- B konpainiaren datuak:
  - Laginaren tamaina $(n_B) = 80$
  - Laginaren batez bestekoa $(\overline{x}_B) = 32$ minutu
  - Laginaren desbideratze estandarra $(s_B) = 4$ minutu

Lagin handiak ditugunez, bi populazioen batez bestekoak alderatu nahi ditugu, eta Z proba bat erabil dezakegu bi lagin independentetarako. Z probako estatistikoa honela kalkulatzen da:

$$Z = \frac{\overline{x}_A - \overline{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{30 - 32}{\sqrt{\frac{5^2}{100} + \frac{4^2}{80}}} \approx -2.99$$

$\alpha = 0.05$adierazpen-maila duen aldebiko proba denez, $\alpha$zati 2 egingo dugu, bi banaketa-kolei esleitzeko; horrela, $\alpha/2 = 0.025$. Banaketa normal estandarraren taulan bilatuta, $\alpha/2 = 0.025$-rako balioa 1.96 dela aurkituko dugu. Beraz, $z_0 \notin [-1.96, 1.96]$.

$H_0$hipotesi nulua ez dugu onartuko, % 5eko garrantzi-mailan. Horrek adierazten du nahikoa ebidentzia estatistiko dagoela lana betetzeko batez besteko denbora bi konpainien artean nabarmen desberdina dela ondorioztatzeko.

## 10.8 ANOVA

1. **Txirrindulari talde bateko kideak hiru taldetan banatzen dira ausaz, eta metodo desberdinekin entrenatzen dute. Lehen taldeak ibilbide luzeak egiten ditu erritmo lasaian, bigarren taldeak intentsitate handiko serie laburrak egiten ditu eta hirugarrenak gimnasioan pisuekin lan egiten du eta goi-maiztasuneko pedalei eragiten trebatzen da. Hilabeteko entrenamenduaren ondoren, errendimendutest bat egiten da, 9 km-ko ibilbide kronometratua. Honako denbora hauek erabili ziren:**

| | 1 metodoa | 2 metodoa | 3 metodoa |
|---|---|---|---|
| | 15 | 14 | 13 |
| | 16 | 13 | 12 |
| | 14 | 15 | 11 |
| | 15 | 16 | 14 |
| | 17 | 14 | 11 |
| B | 77 | 72 | 61 |
| N | 5 | 5 | 5 |

**%95eko konfiantza-mailarekin hiru metodoek emaitza baliokideak sortzen dituztela esan al daiteke? Edo, aitzitik, ba al dago besteak baino metodo handiagorik?**

$$k = 3; \;\; B = \sum_{i=1}^{k} B_i = 77 + 72 + 61 = 210; \;\; N = \sum_{i=1}^{n} n_i = 5 + 5 + 5 = 15$$

$$S_A^2 = \frac{1}{k-1} \cdot \left[ \sum_{j=1}^{k} \left( \frac{B_i^2}{n_i} \right) - \frac{B^2}{N} \right] = \frac{1}{3-1} \cdot \left[ \left( \frac{77^2}{5} + \frac{72^2}{5} + \frac{61^2}{5} \right) - \frac{210^2}{15} \right] = 13.4$$

$$S_B^2 = \frac{1}{N-k} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}^2 - \sum_{i=1}^{n} \frac{B_i^2}{n_i} = \frac{1}{15-3} \cdot \Big[ (15^2+16^2+14^2+15^2+17^2+14^2+13^2+$$

$$+15^2+16^2+14^2+13^2+12^2+11^2+14^2+11^2) - \left( \frac{77^2}{5} + \frac{72^2}{5} + \frac{61^2}{5} \right) \Big] = 1.43$$

$$F = \frac{S_A^2}{S_B^2} = \frac{13.4}{1.43} = 9.37$$

$$F_{\alpha;\, k-1;\, N-k} = F_{0.05;\, 3-1;\, 15-3} = F_{0.05;\, 2;\, 12} = 3.8853$$

$F > F_{\alpha;\, k-1;\, N-k}$ betetzen denez, hipotesi nulua baztertzen dugu eta hiru entrenamendu-metodoek desberdintasun esanguratsuak sortzen dituztela ondorioztatzen da.

**2. Zentzurik gabeko hitzen zerrenda bat agertzen da ordenagailuaren pantailan, ausaz talde bati esleitutako lau prozedura desberdinekin. Ondoren, hitz horiek gogoratzeko proba bat egiten zaie, eta honako emaitza hauek ateratzen dira:**

| | 1 prozedura | 2 prozedura | 3 prozedura | 4 prozedura |
|---|---|---|---|---|
| | 5 | 9 | 8 | 1 |
| | 7 | 11 | 6 | 3 |
| | 6 | 8 | 9 | 4 |
| | 3 | 7 | 5 | 5 |
| | 9 | 7 | 7 | 1 |
| | 7 | | 4 | 4 |
| | 4 | | 4 | |
| | 2 | | | |
| B | 43 | 42 | 43 | 18 |
| N | 8 | 5 | 7 | 6 |

**Zer ondorio atera daitezke aurkezteko lau moduei buruz, %5eko garrantzi-mailarekin?**

$$k = 4; \quad B = \sum_{i=1}^{k} B_i = 43{+}42{+}43{+}18 = 146; \quad N = \sum_{i=1}^{n} n_i = 8{+}5{+}7{+}6 = 26$$

$$S_A^2 = \frac{1}{k-1} \cdot \left[ \sum_{j=1}^{k} \left( \frac{B_i^2}{n_i} \right) - \frac{B^2}{N} \right] =$$

$$= \frac{1}{4-1} \cdot \left[ \left( \frac{43^2}{8} + \frac{42^2}{5} + \frac{43^2}{7} + \frac{18^2}{6} \right) - \frac{146^2}{26} \right] = 27.41$$

$$S_B^2 = \frac{1}{N-k} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}^2 - \sum_{i=1}^{n} \frac{B_i^2}{n_i} = \frac{1}{26-4} \cdot \Big[ (5^2{+}7^2{+}6^2{+}3^2{+}9^2{+}7^2{+}4^2{+}2^2{+}$$

$$+9^2{+}11^2{+}8^2{+}7^2{+}7^2{+}8^2{+}6^2{+}9^2{+}5^2{+}7^2{+}4^2{+}4^2{+}1^2{+}3^2{+}4^2{+}5^2{+}1^2{+}4^2) -$$

$$- \left( \frac{43^2}{8} + \frac{42^2}{5} + \frac{43^2}{7} + \frac{18^2}{6} \right) \Big] = 3.91$$

$$F = \frac{S_A^2}{S_B^2} = \frac{27.41}{3.91} = 7.01$$

$$F_{\alpha;\, k-1;\, N-k} = F_{0.05;\, 4-1;\, 26-4} = F_{0.05;\, 3;\, 22} = 3.0491$$

$F > F_{\alpha;\, k-1;\, N-k}$ betetzen denez, hipotesi nulua baztertzen dugu eta lau aurkezpen-prozedurek alde nabarmenak sortzen dituztela ondorioztatzen da.

**Unibertsitateko eskuliburuak**
**Manuales universitarios**

ISBN: 978-84-9082-965-3

**UPV/EHUko Argitalpen Zerbitzua**
argitaletxea@ehu.eus • 94 601 2227
Biblioteka eraikuntza, 1. solairua
Sarriena auzoa z/g. Bizkaiko campusa
www.ehu.eus/argitalpenak

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

**Servicio Editorial de la UPV/EHU**
editorial@ehu.eus • 94 601 2227
Edificio Biblioteca, 1ª planta
Bº Sarriena s/n. Campus de Bizkaia
www.ehu.eus/argitalpenak