

ESTADÍSTIKA METODOAK INGENIARITZAN

5. ORDENAGAILU PRAKTIKA

R-ko SARRERA, ESTADÍSTIKA DESKRIBATZAILEA ETA ZORIZKO ALDAGAIA

1. Saioa kudeatu

- Lehenik eta behin, **Script** bat ireki behar da: *File->New File-> R Script (Ctrl+Shift+N)*
- Script-ean idatzitakoa gordetzeko: *File->Save as...* Nahi dugun izena jartzen diogu eta gorde nahi den lekua ere.
- Lerroz-lerro exekutatu daiteke (*Ctrl+Intro*) edo Script-ean idatzitako guztia batera exekutatu (*Ctrl+Alt+R*). Bestela *Code->Run Region->...* eta bertan aukeratu zer exekutatu nahi dugun.
- Lan direktorioa aldatzeko: *Session->Set Working Directory->Choose Directory... (Ctrl+Shift+H)*

1

R Studio-rekin LANEAN HASI

2. Datuen bilketa

- Estatistikan datu multzoekin lan egiten da. Hauek gordetzeko hainbat funtzio daude R-n. Garrantzitsuenak: **bektoreak** eta **data frame**.

BEKTOREAK

- Bektoreen datuak zenbakizkoak balioak edo karaktereak izan daitezke.
- Bektoreei hainbat funtzio aplikatu zatzaizkie.
- Bektoreak normalean `c()` komandoa edo funtzioa erabiliz eratu daitezke.
- Bektore bat definitzeko modu ezberdinak daude: `scan()`, `seq()`, `rep()`

[Bektoreen eraketari buruzko UIB-ko MOOC bideoa I](#)

[Bektoreen eraketari buruzko UIB-ko MOOC bideoa II](#)

1

R Studio-rekin LANEAN HASI

2. Datuen bilketa

BEKTOREAK

- Bektore bateko elementu zehatzekin lan egin daiteke.
- Eragile aritmetikoak bektoreko elementuz-elementu lan egiten dute.
- Eragile logikoak erabil daitezke: `<`, `<=`, `>`, `>=`, `==` (berdin), `&` (ebakidura “eta”), `|` (disjuntzio, “edo”)
- Kondizio jakin bat betetzen duten bektore bateko elementuen posizioa jakiteko `which()` funtzioa erabiltzen da.
- Bektore bateko elementu guztiak mota berekoak izan behar dira. Bi salbuespen daude: NA (not available) eta NaN (not a number).

[Bektoreen eragiketei buruzko UIB-ko MOOC bideoa](#)

[Bektoreko elementuen sarbideari buruzko UIB-ko MOOC bideoa](#)

[Bektoreko elementu hutsei buruzko UIB-ko MOOC bideoa](#)

1

R Studio-rekin LANEAN HASI

2. Datuen bilketa

DATA FRAME

- Data frame-ak (Datu markoak) datuak biltzeko gehien erabiltzen den objektua da.
- Data frame bat zutabe ezberdinez eraturik dago, zutabe bakoitza aldagai bat da eta R-n bektore bat bezala definiturik daude. Bektore hauen datuak ez dira zergatik zenbakizkoak izan behar
- Data frame bateko egitura jakiteko `str()` komandoa erabiltzen da.
- Data frame bateko aldagai bat (bektore) eskuragarri jartzeko, `$` komandoa erabiltzen da: (`dataframeizena$zutabearen izena`).
- Aldagai hori eskuragarri jartzeko beste modu bat `attach (dataframeizena)` komandoa erabiliz lortzen da ere. Kontrako komandoa `detach(dataframeizena)` da.

[Data frame-ei sarrerari buruzko UIB-ko MOOC bideoa](#)

[Data frame-en egiturari buruzko UIB-ko MOOC bideoa I](#)

[Data frame-en egiturari buruzko UIB-ko MOOC bideoa II](#)

[Bektoreetatik abiatuz Data frame-en eraketari buruzko UIB-ko MOOC bideoa](#)

4

R Studio-rekin LANEAN HASI

4. Datuen bilketa

DATA FRAME

- Data frame berri bat eratu daiteke aurreko data frame baten datuak erabiliz.
- Data frame bateko aldagai batean hainbat instrukzio exekutatu daitezke `with()` komandoa erabiliz. Era berean, aldagai guztiei funtzio(komando) bat aplikatu zatzaizkie `sapply()` erabiliz.
- Data frame bateko datuak aukeratu daitezke `subset()` komandoa erabiliz.
- Data frame bateko aldagaiak edo eta datuak aldatu daitezke `names()`, `rownames()`, `dinames()` komandoak erabiliz.
- Objektu baten izaera aldatu daiteke adibidez `as.character()`, `as.numeric()`, `as.integer()`...erabiliz.
- Zutabeak (aldagaiak) edo errenkadak (datu berriak) gehitu daitezke Data Frame bati `cbind()` eta `rbind()` komandoak erabiliz hurrenez–hurren, edo `fix(dataframe)` erabiliz.

[Data frame-en bateko datuak aukeratzeari buruzko UIB-ko MOOC bideoa](#)

[Data frame bat aldatzeari buruzko UIB-ko MOOC bideoa](#)

[Data frame bati zutabeak eta errenkadak gehitzeari buruzko UIB-ko MOOC bideoa](#)

[Data frame baten aldagaiei funtzioak aplikatzeari buruzko UIB-ko MOOC bideoa](#)

3. Fitxategi bateko datuen irakurpena

- .txt fitxategi batean dauden datuak, `read.table()` komandoa erabiliz irakurri daitezke.
- Beste modu bat: **Environment** panelean **Import Dataset** sakatu eta inportatu nahi den artxiboa aukeratu.
- Taula moduan dauden fitxategi bateko datuak irakurtzeko eta Data Frame moduan jartzeko `read.delim()` erabili daiteke.
- Irakurri nahi den fitxategiaren direktorio zehatza jarri behar da bai `read.table()` edo `read.delim()` komandoak erabiltzerakoan. Hau ekiditeko, fitxategia gordeta dagoen lan direktoria aldatu behar da saioaren hasieran (7 diapositiba).

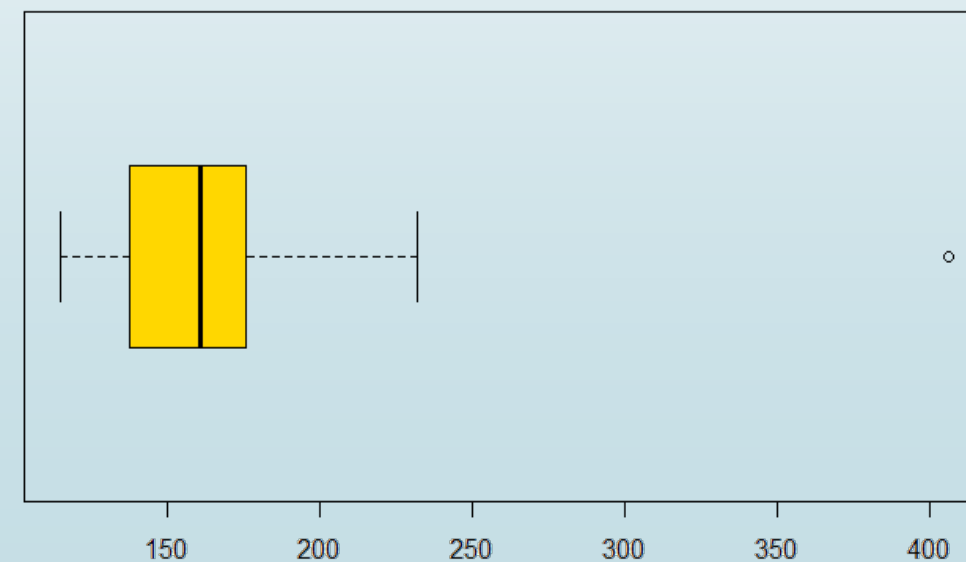
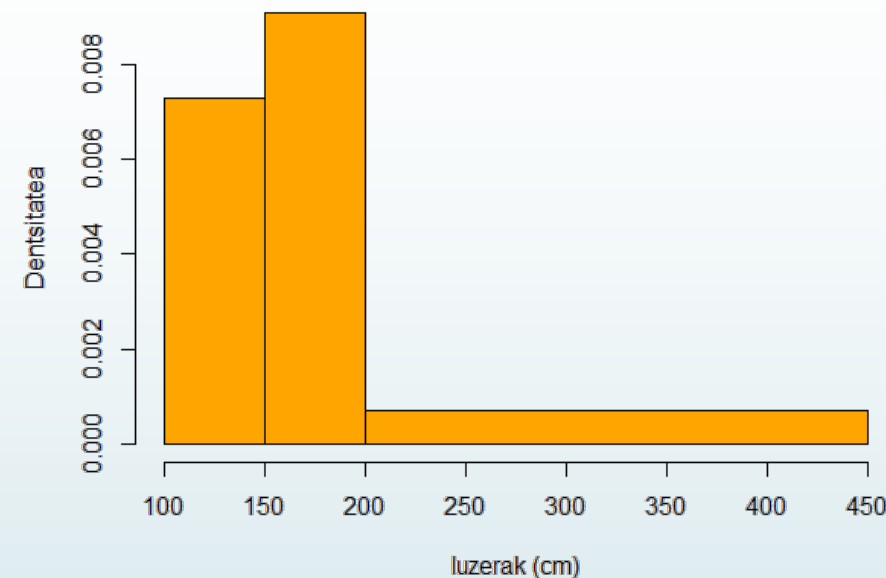
[Fitxategiak irakurtze eta inportatzeari buruzko UIB-ko MOOC bideoa](#)

2

ESTADISTIKA DESKRIBATZAILEA

- **Maiztasun taulak**
- **Adierazpen grafikoak**
 - Barra diagrama/grafikoa
 - Sektore diagrama
 - Histogramak eta maiztasun poligonoak
- **Estatistiko deskribatzaileak**
 - Joera zentraleko neurriak
 - Sakabanaketa neurriak
 - Posiziozko neurriak
 - Formako neurriak
- **Kutxa diagramak**

Luzerak Histograma



Aurrizkiak

Probabilitate funtzioa/Dentsitate funtzioa	d
Banaketa funtzioa	p
Zorizko baloreak sortu	r
Kuantil funtzioa	q

Banaketa diskretuak

Binomiala	binom
Hipergeometrikoa	hyper
Poisson	pois

Banaketa jarraituak

Uniformea	unif
Esponentziala	exp
Normala	norm
χ^2	chisq
Student-en t	t
Snedecor-en F	f

3

ZORIZKO ALDAGIAK

Zorizko aldagai diskretua – BANAKETA GARRANTZITSUAK				
Banaketa	Probabilitate funtzioa: $p(x) = P(X = x) \quad \forall x$	Banaketa funtzioa: $F(x) = P(X \leq x) \quad \forall x$	Kuantilak	Zorizko laginak
Binomiala: $X \sim B(n, p)$	<code>dbinom(x, n, p)</code>	<code>pbinom(x, n, p)</code>	<code>qbinom(v, n, p)</code>	<code>rbinom(o, n, p)</code>
Hipergeometrikoa: $X \sim H(N, n, p = \frac{r}{N})$	<code>dhyper(x, r, N-r, n)</code>	<code>phyper(x, r, N-r, n)</code>	<code>qhyper(v, r, N-r, n)</code>	<code>rhyper(o, r, N-r, n)</code>
Poisson: $X \sim \mathcal{P}(\lambda)$	<code>dpois(x, λ)</code>	<code>ppois(x, λ)</code>	<code>qpois(v, λ)</code>	<code>rpois(o, λ)</code>

1) Notazioa: v: probabilitate-bektorea; o: datu kopurua

2) p eta q funtzioetan `lower.tail=F` argumentua gehi daiteke, defektuz R-k `lower.tail=T` definitua dauka, `lower.tail=F` argumentua gehituz gero $1 - F(x) = P(X > x) \quad \forall x$ probabilitatea kalkulatzen da.

3

ZORIZKO ALDAGIAK

Zorizko aldagai jarraitua – BANAKETA GARRANTZITSUAK				
Banaketa	Dentsitate funtzioa $f(x)$: $P(a < X < b) = \int_a^b f(x) dx$ $\forall a, b \in \mathbb{R} \quad \vee \quad a = -\infty, b = \infty$	Banaketa funtzioa $F(x)$: $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad \forall t \in \mathbb{R}$	Kuantilak	Zorizko laginak
Uniformea: $X \sim UC(a, b)$	<code>dunif(x, a, b)</code>	<code>punif(x, a, b)</code>	<code>qunif(pr, a, b)</code>	<code>runif(o, a, b)</code>
Esponentziala: $X \sim \varepsilon(\beta)$	<code>dexp(x, 1/\beta)</code>	<code>pexp(x, 1/\beta)</code>	<code>qexp(pr, 1/\beta)</code>	<code>rexp(o, 1/\beta)</code>
Normala: $X \sim N(\mu, \sigma)$	<code>dnorm(x, \mu, \sigma)</code>	<code>pnorm(x, \mu, \sigma)</code>	<code>qnorm(pr, \mu, \sigma)</code>	<code>rnorm(o, \mu, \sigma)</code>

1) Notazioa: *pr*: probabilitate-bektorea; *o*: datu kopurua

2) *p* eta *q* funtzioetan `lower.tail=F` argumentua gehi daiteke, defektuz R-k `lower.tail=T` definitua dauka, `lower.tail=F` argumentua gehituz gero $1 - F(x) = P(X > x) \quad \forall x$ probabilitatea kalkulatzen da.

3

ZORIZKO ALDAGAIK**➤ Banaketa diskretuak**

- Zorizko gertaerak sortu. `sample()`
- Probabilitate eta banaketa funtzioak irudikatu. `plot()` *zorizko aldagaiak har ditzakeen baloreak definitu

➤ Banaketa jarraituak

- Dentsitate eta banaketa funtzioak irudikatu. `curve()`