

BIOMETRIA EVOLUZIONISTICA ED ECOLOGICA

Software R:

- Analisi dati;
- è un software per fare conti; x analisi statistica di dati sperimentali; manipolazione dati (manipolare e alterare i dati e si fanno anche i plot)
- Anche linguaggio di programmazione;
- Usa la riga di comando con propria sintassi e **linguaggio** (scrive dei veri e propri programmi);
- Gratuito;
- Multi-piattaforma;
- Flessibile: si possono fare tante cose al suo interno;
- www.r-project.org: scaricalo;
- download → CRAN; cran mirrors (significa che è sotto forma di rete; i mirrors sono dei ripetitori a partire da un cloud di partenza). Posso aprirne uno qualunque perché sono tutti uguali → scelgo Mac → base → download e poi lo installo; Lingua: scegli inglese;
- Nel sito iniziale trovo anche i codici sorgente che generano il software;
- Puoi trovarci anche versioni di anni precedenti;
- Contributed extension packages (più sotto nella pagina di R) → per installare pacchetti diversi (noi abbiamo il base). 21414 pacchetti. Ogni pacchetto ha il suo nome. Noi useremo alcuni pacchetti;
- Consideralo come una grande calcolatrice sovradimensionata;
- Finestra bianca: console di R;
- Le prime sono righe informative;
- Stai sempre aggiornata perché ne escono 2 all'anno circa;
- Condizioni: riconoscimento lavoro altrui;
- Se farai pubblicazioni o nella tesi cita le references degli studenti che usi: semmai cita R se lo utilizzerai;
- Cosa scrivere: es. 2+4 e invio;
- -: meno; *: moltiplicazione; /: divisione. Il punto è la nostra virgola. Uso il punto come separatore in numeri decimali; ^: elevamento a potenza.
- Posso fare anche più operazioni in un colpo solo: uso le parentesi. Nella parentesi è la prima operazione. Uso solo parentesi tonde, non altre parentesi. Uso sempre la tonda per incorporare altre tonde. Il numero di tonde che apriamo deve essere uguale al numero di tonde che chiudiamo. L'ordine è determinato dalle tonde. Le altre parentesi esistono in R ma hanno un altro significato;
- Per fare un **elevazione a potenza** si usa una funzione. Il comando che fa radice quadrata è: **sqrt**. → sqrt(25) → radice quadrata di 5.
- Quindi: comando (argomento/i). I vari argomenti vengono separati da **virgole** (separatore di argomenti).
- Ogni funzione di R ha una help page (ci insegna come usarla). Per chiamarla si fa: ?comando —> es: ?sqrt e si apre la help page. In fondo a qualunque page ci sono esempi.
- Usando le frecce su e giù si scorrono tutti i comandi che si hanno già fatto.
- Comando: lines() → si fanno le linee.

VARIABILI

- R oltre a questo fa di più;
- Definisce le variabili;
- Variabili: oggetti di R che hanno un certo valore. Si dà un valore alla variabile, che può essere numerico, ecc... es: x=5. Lo imposto io scrivendo: x=5. Da questo momento in poi se io scrivo x e lo invio ci ridà il valore = 5. Può essere anche un'operazione es: Y=(2+5)*3. O z=x+y.
- Ls: sta per lista e scrivo: ls(). Vengono fuori gli oggetti che abbiamo creato adesso, cioè le variabili con i valori.
- Per cancellare oggetti: uso rm(valore). Es: rm(xx).
- Se voglio cambiare una variabile la sovrascrivo, cioè scrivo la variabile=a ciò che voglio. Es: x=7. E si cambia da 5 a 7.
- Non esistono solo i numeri per le variabili, ma anche altre che non sono i numeri, es. i **CARATTERI**. Es: y="C". NB. Se voglio mettere un carattere lo metto tra "". Se dimentico le virgolette mi dà ERRORE. R capisce che quello è il nome di qualcosa che ho creato io, ma io non ho creato nulla con scritto C.
- Altro variabile: variabile booleana (buleana) che dà il valore di vero e falso. Es: z=TRUE (senza virgolette). Le usiamo in determinate condizioni. Sono risposte a domande. Non la useremo

BIOMETRIA EVOLUZIONISTICA ED ECOLOGICA

molto. Le uso per verificare una certa condizione. Es: $2 > 7$. Se io faccio $z = 2 > 7$, lo riciclo e diventa FALSE.

- Ci sono anche variabili fattoriali e qualitative;

VETTORI

- le variabili sono MONODIMENSIONALI sempre. I vettori hanno tanti valori contemporaneamente. È un insieme di dati ORDINATI (l'ordine è importante), OMOGENEO (deve contenere tutti numeri, o tutti caratteri o tutti valori booleani. Niente mischioni).
- Per creare vettori serve una funzione. Ci sono tante funzioni per creare vettori: c, seq, rep.
- C: "concatena"; seq: "sequenza"; rep: "ripetizione/repetition."
- C: funziona con tutti. Es: $v = c(1,3,7)$ → così concateno i numeri e creo un vettore. Virgola = separatore. Ora v contiene 1, 3 e 7.
- Esistono anche scorciatoie: se voglio creare un vettore di numeri che vanno da 1 a 10, scrivo $v = c(1:10)$ invece di $v = c(1,2,3,4,5,6,7,8,9,10)$. Oppure fare $v = 1:10$ e funziona lo stesso.
- Funziona anche al rovescio. $v = 10:1$. Si inverte l'ordine.
- Si possono anche unire vari insiemi. Es: $v = c(1:10, 12:15, 16:19)$.
- Si possono concatenare anche vettori di caratteri. Es: $s = c("a", "j", "o")$.
- Seq: è NUMERICA SPECIFICA. Si usa solo con vettori numerici. Genera una sequenza di numeri. Es: $w = seq(7, 12)$. Si scrivono i due estremi della sequenza. Conterrà i numeri interi da 7 a 12. Vantaggio: posso aggiungere un altro argomento che è l'INCREMENTO. Es: incrementare di due di due: $s = seq(7, 12, 2)$ → 7 9 12. Posso mettere anche numeri decimali. Es: $s = seq(7, 12, 0.2)$.
- Rep: fa ripetizioni. $z = rep(5, 100)$ → ripeterà 5 x 100 volte.
- Se faccio $z = rep(c(7, 9, 13), 3)$ → si ripeterà per 3 volte l'insieme. Prima viene eseguito il comando c e poi rep. Oppure $z = rep(c(7, 9, 13), 1:3)$ → 1:3 è un vettore che contiene 1, 2, 3. I due vettori hanno lunghezza 3. Risultato: 7 9 9 13 13 13. 7 corrisponde a 1, 9 a 2, 13 a 3.

A volte si trovano vettori già fatti, tipo LETTERS. Se io scrivo letters in minuscolo compaiono tutti i vettori.

- Oppure LETTERS che scrive la stessa cosa in maiuscolo. Oppure month.name. Sono tutti vettori di carattere.
- Paste: è una funzione per creare etichette o dare nomi ad altre cose. Serve per incollare cose → > incolla vettori tra di loro elemento per elemento (il 1 con il 1 dell'altro, il 2 con il 2 dell'altro e così via).
- Es. $labels = paste(rep("A", 10), 1:10, sep = "")$ → si incolleranno 10 A con la seq di 1 a 10. Labels ora si chiamerà in questo modo → NB c'è un !=! Sep = "/" → mi separa gli elementi con /.
- Paste restituisce **sempre** un vettori di caratteri. I numeri sono considerati come caratteri. Concatena caratteri.

Variabili qualitative o fattoriali: sottoforma non numerica e descrivono i dati sottoforma di categorie. Esprimibili attraverso stati diversi (es. Mendel → o liscio o rugoso). Divide i dati per caratteristiche. Per far capire a R che non è una lista di caratteri si usa un comando:

$f = factor(c(rep("treatmentA", 5), rep("treatmentB", 3), "treatmentC"))$ → un gruppo di persone è stata sottoposta al trattamento A, un gruppo di persone al B e alcune in C.

In fondo mi da i levels: possibili stati di una variabile qualitativa.

Quindi 4 tipologie di fattori: fattoriali, caratteriali, numeriche e bouleriane.

- Con il comando $class(\text{cosa voglio sapere})$ → mi dice se è un factor (qualitativa), carattere (caratteriale), integer (numeri interi), numeric (numeri decimali, es. 0.1, 0.2, ecc...).
- LETTERS lo uso per avere la seq delle lettere.
- Se voglio la sequenza da 1 a 26 scrivo: 1:26.
- Se voglio solo lettere dispari scriverò: $seq(1, 15, 2)$.
- NB con paste metti sempre il separatore! Anche se non voglio separare devo scriverlo.
- $class(es1) = numeric$; $class(es2) = character$.

VETTORI NUMERICI - STATISTICHE DI BASE

Statistiche di base sono:

- $length(z)$: dice quanto è lungo il vettore. Ovvero numero degli elementi del vettore. Si applica a tutti i tipi di vettore.
- $min(z)$: da il valore più piccolo del vettore;

BIOMETRIA EVOLUZIONISTICA ED ECOLOGICA

- $\max(z)$: più alto;
- $\text{sum}(z)$: da la somma di tutti i dati del vettore;
- $\text{mean}(z)$: media del vettore;
- $\text{median}(z)$: da la mediana \rightarrow se ho in ordine i dati è il dato che sta in mezzo. Numero che divide i dati in un 50% più piccolo ed in un 50% più grande. Es. 1 3 5 7 9 \rightarrow media: 5; mediana: 5 (numero che spezza a metà i dati più piccoli e i dati più grandi).
- $\text{var}(z)$: varianza dei valori del vettore.
- $\text{sd}(z)$: standard dei valori del vettori.
- Varianza \rightarrow misura della dispersione dei dati. Più i dati sono dispersi e più la varianza è elevata. Formula varianza: cerca su internet. Sempre positivo.
- Deviazione standard: radice quadrata della varianza.
- Differenza tra i due: unità di misura. Es. se ho due vettori $v_1(28\ 29\ 30\ 31\ 32)$ e $v_2(10\ 20\ 30\ 40\ 50)$. Hanno tutti e due la stessa media e mediana, però se uso la varianza cambiano i numeri. $\text{var}(v_1) = 2.5$ e $\text{var}(v_2)=250$. Il primo è molto più compatto e meno disperso. Il secondo invece è molto dispersivo.

QUANTILI

- la mediana è un quantile empirico. Tutti i quantili fanno lo stesso ragionamento della mediana ma in modo diverso.
 - Il primo quartile è il numero che divide i dati in un 25% più piccolo e un 75% più grande (come mediana ma in proporzione diverse).
 - Il terzo quartile: 75% e 25% più grande.
 - La mediana è il secondo quartile.
- \Rightarrow questi dati insieme dividono i dati in quarti.
- i decili sono numeri che funzionano come i quartili solo che dividono di 10 in 10. Es. 20% e un 80%.
 - In R uso la funzione: $\text{quantile}(z,0.25)$ \rightarrow trovo il QUANTILE al 25%. Posso calcolare tutte le % che voglio, anche 95% es.
 - Posso calcolare tanti quantili in un colpo solo: $\text{quantile}(es1,c(0.25,0.50,0.75))$ e trovo tutte le % che voglio.
 - Funzione summary \rightarrow $\text{summary}(es1)$ = ti da tante info es. min mediana max 1 quartile e 3 quartile media.
 - Se usiamo summary con $f(\text{variabile qualitativa})$ \rightarrow non ci darà tutta questa cosa qua xk sono fattori, ma ci da una tabellina delle frequenze di quante volte compare treatmentA,B,C \rightarrow tabellina con le frequenze assolute.

VETTORI NUMERICI - CAMPIONAMENTO CASUALE E SELEZIONE DI ELEMENTI

È possibile estrarre/selezionare elementi/parti da un vettore più grande? Sì.

Due modi:

- 1) campionamento casuale: uso la funzione sample \rightarrow $r=\text{sample}(1:10,6)$ \rightarrow da un campione da 1 a 10, mi estrae 6 elementi casuali. Se voglio la re-immissione uso $r=\text{sample}(1:10,6,\text{replace}=T)$. Replace si presenta sotto forma bouleriana. T sta per TRUE. Prende 6 numeri casuali e posso comparire anche gli stessi numeri es. 8 e 8. Nel primo non succedeva.

Può essere anche non-numerico: es. lancio la moneta 10 volte \rightarrow $t=\text{sample}(c("T", "C"),10,\text{replace}=T)$.

- 2) Selezione di elementi:

LEZIONE 3

Come poter estrarre alcuni elementi da un vettore in modo specifica:

Se ho un vettore $v=\text{rep}(\text{seq}(3,7,2),4)$, per estrarre tipo il 7 uso le parentesi $[]$. Quindi, il loro scopo è di estrarre elementi specifici.

$\rightarrow v[3]$ \rightarrow estraggo il valore 7 (è il terzo elemento del vettore).

Se ne voglio estrarre di più di uno:

$v[c(1,3,5)]$ \rightarrow è un info sotto forma di vettore.

Posso scrivere anche $v[1:3]$ \rightarrow posizione.

Oppure, posso usare le quadre per creare dei filtri. Es. $v[v>4]$ \rightarrow estraggo numeri maggiori di 4.

NB. Se scrivo $v>4$ si genera un vettore bouleriano \rightarrow false true true false ...

Quindi in queste parentesi c'è un vettore bouleriano.

BIOMETRIA EVOLUZIONISTICA ED ECOLOGICA

Ergo, estraiamo elementi sulla base della loro posizione o del loro filtro. Posso salvare il filtro in un altro vettore:

$v1=v[v>4]$ → filtro.

NB. Ciò che viene mostrato non viene salvato, ciò che non compare viene mostrato.

Ciò che sta nelle quadre è visto da R come un vettore sempre.

Il vettore può essere numerico o bouleriano.

Esercizio:

NB. Per i numeri decimali non funziona:

Scrivo **$s=seq(0.001,1,0.001)$**

Chiamo il nuovo vettore "g" e scrivo: **$g=sample(s,100)$** → NB. Si crea un nuovo vettore.

Calcolare le statistiche di base:

Uso summary per fare prima. Scrivo **summary(g)**

var(g)

sd(g)

Per calcolare i decili: **quantile(g,seq(0.1,0.9,0.1))**

Ultimo stap: fare un filtro → $w=g[g<0.3]$ oppure senza creare il vettore → $g[g<0.3]$

Per vedere la lunghezza: **length(g)** oppure **length(w)**.

MATRICI

Sono insieme di elementi ordinati, però con un ordinamento di 2D (il vettore ne ha solo una). Avrà una forma come una tabella, infatti si chiama righe e colonne. Per il resto sono simili ai vettori.

Devono essere omogenee e non miste. Come i vettori. Fatta di righe e colonne.

Come creare una matrice:

Creare una matrice di 12 elementi, avrà 4 colonne.

$M=matrix(1:12,3,T)$

La funzione Matrix prende 3 argomenti:

- vettore: elenco il contenuto della matrice → 1:12.

- Gli altri 2 x specificare come farla → nrow (numero delle righe) e byrow (in forma bouleriana).

Devo riempire la matrice secondo riga [T] oppure no [F, quindi x colonne]? Specifica l'ordine di riempimento della matrice.

Es. $M=matrix(1:12,nrow=3,byrow=T)$

T=riempimento secondo riga, F=riempimento secondo colonna.

Per dare nomi specifici alle righe e alle colonne faccio così:

• **rownames(M)=c("A","B","C")**

→ a,b e c sono le righe

• **colnames(M)=c("I","II","III","IV")**

Posso creare una seconda matrice e faccio **$M2=t(M)$** e M2 sarà uguale a M ma con il nome delle colonne e delle righe invertite.

Se voglio sapere le dimensioni della mia matrice mi restituisce il numero delle righe e delle colonne.

dim(M)

Altro modo per costruire le matrici è usano la funzione:

• **abc=rbind(a,b)** → creano matrici appiccicando i vari vettori. I vettori diventano righe di una matrice;

• **abr=cbind(a,b)** → i vettori diventano colonne di una matrice.

Es. Creo due vettori: $a=1:7$ e $b=8:12$ → NB. Devono avere la stessa lunghezza e devono essere omogenei.

$ab=rbind(a,b)$

Esercizio: 2 modi per farlo.

Modo 1:

$A=seq(1,10,2)$

$B=seq(2,10,2)$

$C=rep(0,5)$

Ho creato i 3 vettori.

Creo la matrice: **$ABC=cbind(A,B,C)$**

Modo 2:

$ABC=matrix(c(A,B,C),nrow=5,byrow=F)$ → uso i vettori che ho creato prima.

Per cambiare i nomi:

rownames(ABC)=letters[1:5] → NB. Per fare prima ho usato letters.

BIOMETRIA EVOLUZIONISTICA ED ECOLOGICA

colnames(ABC)=LETTERS[1:3]

DATA FRAMES

Sono molto importanti in R. Tipica forma in cui in R si organizzano i dati. A primo d'occhio è uguale alla matrice. Nel data frame essere riga o colonna ha un significato specifico:

- Riga: le nostre osservazioni,
- Colonne: sono le variabili che abbiamo misurate.

Altra differenza dalle matrici è che la matrice funziona come i vettori, deve essere omogenea, il data frame invece no, può avere colonne di natura diversa (all'interno della colonna deve però esserci omogeneità).

Come costruire un data frame:

Funzione principale: `data.frame(var3,var2,var1)` → appiccica insieme dei vettori facendolo diventare delle colonne.

Quindi, creo 3 vettori con i dati delle variabili.

- `var1=1:10`
- `var2=16:25`
- `var3=factor(c(rep("a",2),rep("b",5),rep("c",3)))` → variabile qualitativa!

DF=data.frame(var1,var2,var3)

Per essere sicuri che è un data frame scrivo: `class(DF)` e mi dice subito se è una matrice o un data frame.

Poi, sui suoi appunti ci sono le diverse cose per vedere le diverse cose.

`head(DF)`: mostra di default le prime 6 righe, però se voglio vedere solo 3 righe scrivo: `head(DF,3)`.

Uguale per `tail` (mostra le ultime righe).

Esercizio:

`var1=1:50`

`var2=c(rep(TRUE,30),rep(FALSE,20))` → NB. True e false senza apicetti.

Oppure

`var2=var1<31`

`var3=rep(LETTERS[17:26],5)`

Oppure

`tail(LETTERS,10)`

`var3=rep(tail(LETTERS,10),5)`

`DF=data.frame(var1,var2,var3)`

Poi guardo appunti prof e vedo tutti i parametri.

DATA IMPORT/EXPORT

Uso `read.table`. Importa dati dall'esterno e mi genera un data.frame. I dati però devono già avere la forma di tabella con righe e colonne. NB. Noi dobbiamo sapere che la prima riga contiene i nomi delle variabili. A volte compaiono le sigle NA (not available, significa che un dato è mancante).

[con Mac vado su `misc` in alto ecc...]

Come importare: prima di tutto devo informare R dove è il file. Si trova nella cartella Basic sul desktop → file → cambia directory → Basic

Scrivo: `getwd()`

`data=read.table("alkfos.txt",header=T)`

NB. Specificare l'header (sapere se contiene o meno l'header, cioè se c'è l'intestazione nella prima riga o no; cioè se nella prima riga si contengono delle variabili o no, se le contiene `header=T`!)

Metti in R ?`read.table` e leggi cosa c'è scritto.

LEZIONE 4

Conto NA's → conto dei dati mancanti (NA significa not available).

Grp: variabile qualitativa o fattoriale. Nel data frame è stata messo come carattere, dobbiamo cambiarlo e mettere come variabile qualitativa.

Varianti di `read.table`: funzionano ugualmente, però si usano con il formato csv dei file di testo → `read.csv` e `read.csv2`. Qui i separatori sono o virgole o punto e virgole.

Se ho dei file in formato excel la cosa più facile da fare è importare il file in csv.

Come specificare grp come variabile qualitativa: `data$grp` e invio. Si estrai il vettore corrispondente della colonna. Lo posso fare per tutte le variabili.

`data$grp` non cambio. Per cambiare: `data$grp=factor(data$grp)` → lo sovrascrivo come fattore.

BIOMETRIA EVOLUZIONISTICA ED ECOLOGICA

Se faccio summary, mi compare una tabellina e non mi dice più che è un carattere.

NB sempre specificare se sono variabili qualitative quando uno i data frame.

Per esportare da R asco write.table

- 1) `alkfos.csv` → do il nome che voglio dare al file con csv (lo voglio esporre con csv);
- 2) `Row.name` → x definire i file di output deve avere i nomi delle righe, voglio i nomi delle righe?
F=no, T=si;
- 3) `col.name` → "" delle colonne; vogliono i nomi delle colonne? Si=T (sono i nomi delle variabili);
- 4) `Quote` → F, perché non voglio mettere gli apicetti;
- 5) `Sep` → il separatore che voglio è il ";"; quello internazionale ",";
- 6) `Dec` → il decimale metto la "."; quello internazionale ".";

Ora dovrei trovare in Basic il file.

`data[2,3]` → per estrarre un numero da una tabella, 2 è la riga e 3 la colonna.

Per estrarre un pezzettino di tabella → `data[1:3,2:5]` è quindi possibile estrarre più elementi in un colpo solo.

Se scrivo `data[2,]` → è estratta tutta la riga dall'inizio alla fine e la stessa cosa vale x le colonne (si basa sulla posizione).

Un altro modo per estrarre le colonne posso usare il \$. Devo xro sapere il nome della colonna → `data$c3` → viene estratta tutta la colonna (si basa sul nome).

==> 2 modi basati o sulla posizione o sul nome.

`data[,c(1,3,6)]` → per estrarre 3 colonne.

Salvarlo su un oggetto: `sub7 = data[,c(1,3,6)]`.

Gli indici nelle quadre possono funzionare anche al contrario basta metterci un - davanti! Estraggo tutto tranne quelle colonne con il -.

`data[,-c(1,3,6)]`.

Usando lo stesso concetto si può anche modificare la tabella! Se faccio `data[3,2]=100` → mi cambia il numero che c'era in questa posizione con 100!!

Posso fare la stessa cosa con le colonne → `data[,4]=data[,6]` ho due colonne uguale alla 6.

Si può fare la stessa cosa con il dollaro → `data$c18=data$24`.

`data[,8:1]` → mi riscrive tutta la tabella solo al contrario.

LEZIONE 5

Grafici e funzioni grafiche: lavoreremo con variabili di tipo quantitativo (variabili espresse con numeri) e qualitative (non espresse con numeri). Rappresentazioni grafiche delle variabili (anche in modo misto).

Grafici per le variabili qualitative o fattoriali: iniziamo costruendo un vettore che corrisponde alle classi dei passeggeri del Titanic. Costruiamo un vettore numerico con 4 numeri.

`titanclass=c(325,285,706,885)`

`Names` → serve da etichetta per dare il nome alle righe e alle colonne. Abbiamo dato etichette al vettore che esiste.

`names(titanclass)=c("1st","2nd","3rd","Crew")`

Nella prima classe c'erano 325 individui ecc... → SONO VARIABILE QUALITATIVA CON 4 POSSIBILI VALORI.

Lo rappresentino con il grafico a torta.

`pie(titanclass)`

I colori sono dati di default. Se vogliamo dare un titolo e selezionare colori: `main` = specifica il titolo principale del grafico e `col` = sta per Colors e serve x specificare un vettore di colori che vogliamo utilizzare (tanti colori quanti sono le classi).

NB. I colori vanno messi in apicetti.

`colors()` → da l'elenco dei nomi di tutti i colori codificati nella versione Basic.

Ci sono dei vettori con colori standardizzati con alcuni colori, come `pal=rainbow()` e nelle parentesi metto quanti colori voglio (qui si usano i codici alfanumerici dei colori).

Per esportare: la finestra del grafico DEVE ESSERE APERTA SE SI VUOLE SALVARE. Uso `dev.copy2pdf(file="pieplot_col.pdf")` → salva un file in pdf nella nostra directory. I file li troverò in documents del computer.

Con `grey` → 0 corrisponde al nero e 1 al bianco → voglio 4 colori allora la nostra spaziatura sarà di 0.33.

Per aprire i pacchetti: `library(nomedelpacchettosenzaapicetti)`

Grafico a barre: lungo l'asse orizzontali ci sono le 4 classi e in verticale ci sono le frequenze.

NB. Guarda appunti.

Se vogliamo il grafico in orizzontale scriverò:

BIOMETRIA EVOLUZIONISTICA ED ECOLOGICA

`barplot(titanclass,main="Passegens of Titanic",col=pall,hor=T)`

Grafico con 2 variabili qualitative → usiamo il grafico a mosaico.

Uso `library(MASS)` → già presente su R, quindi basta che scrivo questo e poi scrivo `data(caith)` e poi solo `caith`.

`Mosaicplot(...,shade=T)`

Bianco: differenze non significative, blu: eccesso dei rappresentazione rispetto all'atteso, rossi: carenza di individui nel nostro dataset rispetto all'atteso.

Linee tratteggiate rosse: meno individui di quelli che ci aspettavamo;

Linee piene blu: più individui.

Variabile quantitativa: tutte le variabili che possono essere misurate.

`[crabs$sp=="O"]` → è un filtro. Vogliamo solo le dimensioni del carapace di granchi arancioni. Se il filtro è TRUE ci verranno fuori le info.

`==` → controllo di una uguaglianza. Chiediamo: è vera questa cosa?

Uso un istogramma per rappresentare una variabile continua.

NB. Non è un barplot perché una variabile quantitativa e non qualitativa.

Con `freq=F` avremo le densità di frequenza (frequenza relativa [frequenza assoluta di ogni classe/ totale delle osservazioni] di ogni classe e si divide per l'ampiezza della nostra classe).

La somma dell'area di tutti i rettangoli fa 1 (per la probabilità) e la densità di probabilità funziona bene anche per classi di grandezza diverse tra di loro.

Il modo migliore per fare un istogramma è usare la densità perché si può generalizzare. Quindi se posso scegliere uso la densità e non la frequenza.

Area di ogni rettangolo è proporzionale alla probabilità che un dato cada là.

`lines(d)` → aggiunge una stima generale (linea di tendenza) SOPRA il nostro grafico.

`Br` → numeri da 10 a 50 spazati di ampiezza 10.

Boxplot: margine inferiore → 1 quartile, sup → 4 quartile, in mezzo → mediana o 2 quartile, i due batti (linee sopra e sotto) includono tutti gli altri valori dal valore minimo al max fino a 1.5. gli altri sono outliers.

NB. L'istogramma divide in classi di grandezza. Il boxplot no, ma li divide in quartili.

`las` → gira i numeri della scala (prima erano allineati con l'asse, ora li gira così da avere una lettura fluida).

`rug(caleor,side=2)` → rappresentazione grafica in tacche dei nostri dati sul lato 2.

Side 1 in basso, side 3 in alto e 4 a dx.

Notch: aggiunge una tacca in corrispondenza della mediana dei nostri dati. Disegnata in maniera proporzionale rispetto alla varianza dei nostri dati (tacca grande: il nostro dataset è molto grande)

→ permette di confrontare a livello grafico 2 distribuzioni. Se ho due boxplot e voglio confrontarli (se ho la stessa varianza e distribuzione allora sono simili) e li sovrappongo e le tacche sono uguali allora sono simili, se non si sovrappongono non sono simili.

Grafico per due variabili quantitative contemporaneamente: uso Scatterplot e rappresentazioni multi-frame.

Una variabile sull'asse delle x e l'altra variabile quantitativa sulla asse delle y.

È un grafico a dispersione. Uso `plot(caleor,cawior)`

`Caleor` va sull'asse delle x, `cawior` sull'asse delle y.

Ogni dato è un granchio.

Qui le due variabili sono proporzionali tra di loro. Cresce una e cresce anche l'altra.

Sono legate tra di loro con relazioni lineare, direttamente proporzionale. Al crescere cresce anche l'altra.

Si vede che non è casuale.

`Par`: serve anche per creare un plot multiframe → serie di grafici rappresentati nello stesso pannello.

`mfrow=c(1,3)` → si chiede di suddividere il pannello grafico in un area di una riga e tre colonne.

Dividiamo il pannello grafico in 3 parti uguali. `mfrow = "multiframe"`.

`plot(caleor,cawior)` → il primo grafico si inserisce nel primo pannello.

`plot(cawior~caleor,main="Length and Width of carapaces")`. Va nella seconda parte.

(Ondina window: alt 1,2 e 6; Mac option e 5).

E poi faccio il terzo plot che andrà nella terza parte.

`pch=20` → puntino più piccolo ma pieno.

Per tornare ad un pannello normale (rimane diviso in tre finché non si chiude. Se aggiungiamo un altro plot, si metterà nella prima parte della finestra).

BIOMETRIA EVOLUZIONISTICA ED ECOLOGICA

Scrivo: `par(mfrow=c(1,1))` → si torna ad un pannello intero x ogni immagine.

`Xlab` e `ylab` → decidiamo noi cosa mettere cosa dove.

Per aggiungere linee → uso `abline` (fa linee rette, non curve). H: riga orizzontale, V: linee verticale.
lty: x decidere che linea usare e modificare l'estetica delle linee rette. Col: scelgo il colore delle linee.

Trovo il punto dove sin incrociano le 2 medie.

Per evidenziare il punto uso `points`. `cex=2` → punto grande il doppio rispetto al default (che sarebbe 1).