

Compiti a casa assegnati per i corsi di
Analisi di Dati Ecologici
e
Abilità informatiche

Anno accademico 2017 - 2018

Compito per il 9 Ottobre 2017 Svolgete i seguenti esercizi:

1. Trovare la funzione in R che calcola la mediana di una serie di numeri, spiegare cosa è la mediana, trovare la funzione in R per il calcolo della quantità analoga alla deviazione standard (quindi una quantità che calcola la variazione dei numeri), ma riferita alla mediana invece che essere riferita alla media. Spiegare a parole cosa calcola questa funzione. Applicare la funzioni trovate ad un vettore di almeno 7 numeri scelti dallo studente. Spiegare brevemente come siete giunti a trovare le funzioni che cercate e il “percorso” che avete fatto. Se si riesce elencare brevemente alcuni vantaggi e svantaggi nell’uso della mediana rispetto all’uso della media.
2. Cercare i dati di almeno una decina di amici o conoscenti riguardanti il peso (kg) e l’altezza (cm). Immettere i dati in R , calcolare la regressione fra peso e altezza usando il peso come variabile dipendente. Fare un grafico con i punti e la retta di regressione. Interpretare e commentare i risultati dal punto di vista biologico e statistico. Secondo il modello statistico che avete usato (retta di regressione) quanto dovrebbe pesare una persona alta 178 cm? e quanto un bimbo di 65 cm? Sono risultati verosimili? Commentare.

Compito per il 18 Ottobre 2017 Lo scopo di questi due esercizi è quello di imparare a convertire delle formule matematiche in comandi di R. Quando riportate i comandi di R e il suo output, per favore, commentate in modo che io riesca a vedere che voi state effettivamente capendo quello che fate.

1. Cercate la formula della media ponderata e calcolate in R la media dei vostri voti della triennale pesando il voto degli esami per i crediti dell’esame stesso.

2. Cercate e scrivete con l'equation editor di Word o Libreoffice le formule per calcolare:

- la media,
- devianza,
- varianza,
- varianza con la formula computazionale (o alternativa),
- deviazione standard e
- covarianza.

Spiegate brevemente, con parole vostre, il significato statistico. Su un vettore casuale di 20 numeri interi generato da una distribuzione uniforme con minimo=0 e massimo=50, effettuate i calcoli delle statistiche di cui sopra, create le funzioni che effettuano questi calcoli usando come argomento un solo vettore numerico e confrontate i risultati con quelli delle funzioni apposite di R. In R non esiste una funzione apposita per calcolare la devianza. Per la covarianza è necessario generare anche un secondo vettore di dati e quindi creare la funzione anche con un secondo argomento.

Compito per il 26 Ottobre 2017 Il compito numero 1 non è da consegnare.

1. Vi consiglio di esercitarvi ad importare i file in R da Excel o da un altro foglio elettronico. Potete provare ad importare il secondo foglio del file `SLA.xls` che abbiamo usato in classe, ma vi consiglio anche di provare con degli altri dati. Non correggerò questo compito e non voglio che me lo consegniate.
2. Produrre un vettore `weight` (grammi) di 15 dati casuali, estratti da una distribuzione normale con $\mu = 40$ e $\sigma = 10$, create una funzione per standardizzarlo, cioè che tolga la media e divida il risultato per la deviazione standard. Confrontate il risultato con quello della funzione `scale()`.
3. Rispondere alle seguenti domande, riportare i comandi e l'output di R e commentate:
 - Se `x<-c(2,2,3)` e `y<-1:6`. Cosa succede se provo `x * y`?
 - Perché il risultato di `x * y`, dove `x<-c(2,2,3)` e `y<-1:7`, è così ed ho un `warning`? Perché nel caso precedente non ho il `warning`?

- Cosa succede se moltiplico/sommo/divido/sottraggo due vettori di lunghezza diversa? Questa è la “regola del riciclaggio dei vettori” che si applica quando i vettori sono di lunghezza diversa. Provate a spiegare con parole vostre questa regola.
4. Utilizzando la funzione `rnorm` per creare due vettori casuali di almeno 7 elementi (anche correlati) `x` e `y`. Verificate empiricamente che:
- (a) $E(k x) = k E(x)$
 - (b) $var(k x) = k^2 var(x)$
 - (c) $sd(k x) = k sd(x)$
 - (d) $E(x + y) = E(x) + E(y)$
 - (e) $var(x + y) = var(x) + var(y) + 2cov(x, y)$
 - (f) $E(x y) = E(x) E(y) + cov(x, y)$ (Vedi nota:¹)
 - (g) $E\left(\frac{1}{x}\right) \neq \frac{1}{E(x)}$
 - (h) $var(x y) \neq var(x) var(y) + 2cov(x, y)$

dove k è una costante, $E()$ è il valore atteso (media), $var()$ è la varianza, $sd()$ è la deviazione standard, $cov()$ è la covarianza. Usate pure le funzioni apposite di R.

Attenzione che a volte l'operatore `'=='` restituisce **FALSE** anche per piccoli ed inevitabili errori di arrotondamento interni alla macchina. Una delle soluzioni più semplici è arrotondare alla quinta cifra decimale il risultato di destra e sinistra dell'uguaglianza. Oppure potete controllare la differenza tra la parte a destra e a sinistra delle eguaglianze e se la differenza è molto molto piccola, allora si può ritenere verificata l'uguaglianza. Un'ulteriore soluzione è usare la funzione `all.equal()`

Compito per il 8 Novembre 2017 Svolgete i seguenti esercizi:

1. Esercizio sull'uso dei vettori logici come indici:
 - (a) Creare la funzione `is.pari()` che prenda un vettore numerico come argomento e restituisca dei **TRUE** o **FALSE** per ogni elemento del vettore a seconda che il numero sia rispettivamente pari o dispari ².

¹Funziona esattamente solo se la $cov(x, y)$ al denominatore ha n e non $n - 1$

²Suggerimento: usate l'operatore `%/%` per il resto di una divisione

- (b) Generare un vettore `x` casuale di 20 numeri distribuiti normalmente con media 5 e deviazione standard 8 e arrotondarlo a due cifre decimali. Con un unico comando sostituire i numeri negativi di `x` con zero.
- (c) Generare un vettore `a` casuale di 20 numeri interi distribuiti uniformemente da 1 a 50. Utilizzando la vostra funzione `is.pari`, con un solo comando, estrarre i numeri pari di `a` e assegnarli ad un vettore nuovo chiamato `apari`. Fare la stessa cosa con i numeri dispari creando il vettore nuovo `adispari`.
- (d) Estrarre gli elementi di `x` corrispondenti agli elementi di `a` che sono pari e assegnarli ad un vettore nuovo chiamato `xapari`. Quindi fare la stessa cosa con i dispari e creare il nuovo vettore `xadispari`.
- (e) Concatenate `xapari` e `xadispari` in un vettore chiamato `xa`. Quindi, con l'aiuto delle funzioni `rep` e `as.factor`, generate un nuovo vettore `ParDisp` di fattori lungo tanto quanto `xa` e contenente le stringhe "Pari" e "Dispari" a seconda che il corrispondente elemento di `xa` sia originato da `xapari` o `xadispari`.
- (f) Creare un dataframe con i due vettori `xa` e `ParDisp`. Quindi effettuare un test T di Student con la funzione `t.test()` su `xa` per confrontare le medie dei due gruppi `Pari` e `Dispari`. Commentare il risultato e presentare un grafico che lo illustri.
2. Svolgere i seguenti esercizi. Commentate abbondantemente.
- Che differenza c'è fra `rep(1:3,3)` e `rep(1:3,c(3,3,3))` e `rep(1:3,rep(3,3))`?
 - Utilizzando i vettori `ages` e `nomi` usati in classe, descrivere cosa producono le espressioni `sum(ages < 30)`, `sum(ages[ages < 30])` e `sum(nomi[ages < 30])` e perché?
 - Cosa produce la seguente espressione `nomi[nomi=="Stefano" | nomi=="Franco"]`? Perché?
 - Spiegare perché `abs(z)` ha lo stesso effetto di `z[z < 0] <- -z[z < 0]`?
 - Come posso estrarre nella matrice `X` l' i , j -esimo elemento?
 - Come posso estrarre tutte le colonne che vanno da 3 a 6?
 - Come posso estrarre tutte le righe tranne la i -esima?
 - Cosa produce `diag(3)`? e `diag(1:3)`?
3. La Federazione Italiana di Atletica Leggera (FIDAL www.fidal.it) deve convocare, per i Campionati Italiani di Salto in Lungo della

categoria ragazze, le atlete che hanno superato o eguagliato la misura 4.60 metri nel 2013. I dati, tratti dal database della FIDAL, sono nel file per Microsoft Excel su http://elly.scvsa.unipr.it: salto_in_lungo_ragazze.xls. Per poter effettuare la convocazione la FIDAL deve avere la lista dei nomi delle ragazze da convocare e la società sportiva da avvisare. Aiutate la FIDAL nella convocazione producendo i dati richiesti.

Se si escludessero le prestazioni ventose, cioè quelle con vento favorevole (positivo), quali e quante sarebbero le ragazze da convocare?

Non stampate l'intero data set. Stampate solo i dati richiesti! Attenzione che un nome di una società contiene un apostrofo che potrebbe rendere difficoltosa l'importazione dei dati in R.

4. Eseguire manualmente i prodotti di matrici AB e BA dove

$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \\ -3 & 4 \end{pmatrix}$$

e

$$\mathbf{B} = \begin{pmatrix} 1 & -2 & 5 \\ 3 & 4 & 0 \end{pmatrix}$$

e verificare il risultato con l'operatore di R. Mostrare tutti i conti. È possibile presentare questo esercizio scritto a penna, ma si viene penalizzati di un voto.

Compito per il 15 Novembre 2017 Svolgete i seguenti esercizi:

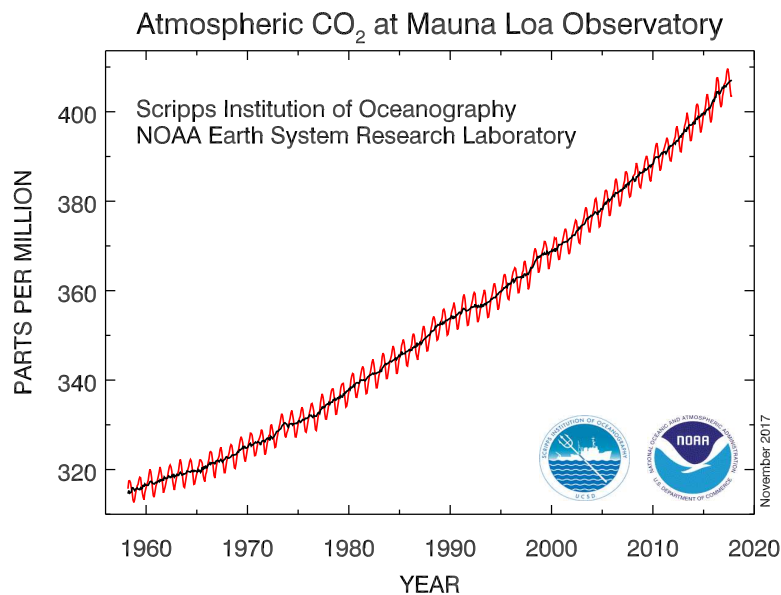
1. Scrivere due funzioni che calcolino la traccia di una matrice quadrata (somma degli elementi sulla diagonale). Nella prima usare la funzione apposita di R per estrarre la diagonale. Nella seconda usare un ciclo `for` per accedere agli elementi della matrice. Applicarle ad una matrice quadrata 12×12 con numeri con due decimali estratti casualmente da una distribuzione normale con $\mu = 10$ e $\sigma = 5$ e verificarne il corretto funzionamento.
2. Utilizzando i cicli `for`, determinare l'insieme intersezione dei due vettori A e B.

```
A <- c(1,4,6,9,15,16,19,22)
```

```
B <- c(1,5,9,11,14,16)
```

A partire da quanto scritto per questo caso particolare, scrivere una funzione generale per calcolare l'insieme intersezione. Verificare se R già possiede una funzione del genere e se esistesse controllare il risultato. **Suggerimento:** può essere utile utilizzare la struttura di controllo `if`.

3. Scrivere, utilizzando un ciclo `for`, una funzione che ritorni un vettore con la media degli elementi di ogni riga di una qualsiasi matrice passata come argomento. Scrivere una seconda funzione che faccia la media delle colonne della matrice. Applicare le funzioni scritte ad una matrice qualsiasi e confrontarne i risultati, dapprima con la funzione `apply` e successivamente anche con le funzioni di sistema `colMeans` e `rowMeans`.
4. Obiettivo del compito è rifare il grafico della famosa curva di Keeling relativa alla concentrazione di CO_2 sull'isola di Mauna Loa.



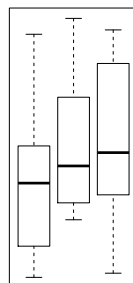
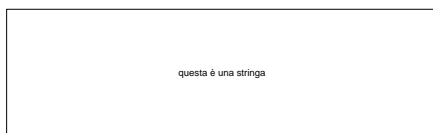
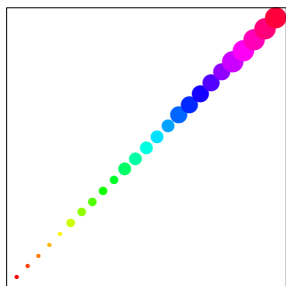
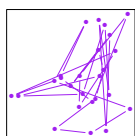
I dati mensili sono disponibili in rete (ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_mm_mlo.txt), importare i dati in R e rifare il grafico come in questa figura. Il grafico deve riportare l'andamento della curva rossa con le relative oscillazioni e deve avere gli assi come nella figura. Riportate i dati delle medie giornaliere (colonna `average`) e non quelli interpolati. La curva in nero non è necessario riportarla. Il grafico però deve inoltre

presentare la retta di regressione lineare fra anni e concentrazione di CO₂. Predire la concentrazione di CO₂ per l'inizio di gennaio 2018 e dire perché non andrà bene. Commentate ampiamente i comandi che date. Ricordo che in un file di dati in formato testo le righe che iniziano con '#' sono ignorate da R, esattamente come i commenti nella riga di comando.

Non stampate l'intero dataset, per favore.

5. Eseguendo in questo preciso ordine la sottostante lista di comandi, si ottiene il seguente grafico:

```
x <- round(rnorm(25,4,3),2)
y <- round(runif(25,1,100),2)
tr <- as.factor(rep(c("Ctrl", "Tr1", "Tr2"), c(8,8,9)))
par(mar=c(1,1,1,1))
plot(x=1:25, col=rainbow(25), cex=rep(1:5, each=5), pch=16, xlab="",
yaxt="n", xaxt="n")
plot(x,y, type="b", col="purple", pch=16, xlab="", ylab="", xaxt="n",
yaxt="n")
boxplot(y~tr, yaxt="n", xaxt="n")
plot(1,1, type="n", xaxt="n", yaxt="n", xlab="", ylab="")
text(1,1, "questa è una stringa")
```



Qual è il comando di layout perché la finestra grafica risulti così impostata? Tenete presente che i parametri widths e heights sono stati usati di default.

6. Spiegate brevemente quali elementi costituiscano un boxplot, con particolare attenzione a descrivere – con parole vostre – il criterio sulla base di cui vengono definiti gli outlier e disegnati i “baffi”. Poi rappresentate con un boxplot il vettore `y`, definito come `y <- round(rnorm(25,1,5),2)`, avendo cura di aggiungere al grafico:
- I valori di mediana, primo quartile, terzo quartile, baffo superiore, baffo inferiore;
 - I punti stessi, al fine di chiarire bene quale sia la distribuzione “sottostante” alla rappresentazione fatta da R . Se riuscite, non lasciate i punti “incolonnati” su un unico valore di `x`, ma rappresentateli come una “nuvola” (SUGGERIMENTO: per fare ciò, consultate l’help della funzione `jitter`).
7. Un gruppo di ricercatori dell’Università della North Carolina hanno seguito la crescita di 27 bambini (16 maschi e 11 femmine) dall’età di 8 anni fino a quella di 14, misurando ogni due anni la distanza (in mm) tra la sella turcica (sede dell’ipofisi) e la fessura pterigopalatina, due punti del cranio che sono facilmente identificabili con l’esposizione ai raggi X. I dati sono riportati in Potthoff & Roy (1964) e in R sono disponibili nel dataset `Orthodont`, in un pacchetto diverso da quello base, il pacchetto `nlme`. In un unico grafico rappresentate il valore medio di distanza per ciascuna età e per entrambi i sessi, commentandone brevemente il significato biologico. Ricordatevi di aggiungere una legenda e le barre di errore.

```
library(nlme) # per caricare il pacchetto appena scaricato
data(Orthodont) # per caricare il dataset
```

8. Nel file `height.xls` sono contenuti i dati di altezza (in m) di 4 diverse specie forestali (abete bianco, frassino, sequoia e nocciolo). Importate il dataset in R per produrre:
- un grafico con i 4 boxplot (uno per ciascuna specie);
 - un barplot con l’altezza media per ciascuna delle 4 specie, con relativa barra d’errore.

Prestate attenzione al fatto che i dati iniziali dovranno essere riorganizzati, di modo da avere una struttura simile a quella del dataset `iris` visto a lezione.

Compito per il 6 dicembre 2017 Svolgere i seguenti esercizi:

1. Ripetere il test di permutazione che ha mostrato John Rauser nel suo video disponibile su youtube o sul sito `elly.scvsa.unipr.it`: “*Statistics Without the Agonizing Pain*”. I dati sono visibili attorno al minuto 3:37 del video. Però non usare, come statistica di riferimento, la differenza fra le medie dei due trattamenti, ma il logaritmo del rapporto fra le medie. Effettuate almeno 50 mila ricombinazioni e abbondate con commenti in modo da far capire che avete afferrato bene il concetto e la parte tecnico-informatica del test.

Spiegare in termini informatici e logici perchè in R posso calcolare il *P-value* della mia statistica osservata con il comando: `2*mean(logratios > truelogratio)` dove `logratios` è il vettore con i logaritmi del rapporto fra le medie dei dati permutati (con le medie “della birra” al numeratore e “dell’acqua” al denominatore) e con `truelogratio` il logaritmo del rapporto fra le medie dei dati veri ³.

2. Usando l’indice di diversità di Simpson dato da

$$H_S = 1 - \sum_i^s p_i^2$$

rifare le curve di rarefazione sui dati di comunità contenuti nel file `alberi.txt`, come sulle dispense, e commentare sia i comandi, sia i risultati. Fare almeno 100 repliche di ogni ricampionamento. Dare un’indicazione su un possibile numero minimo di individui da campionare.

3. Campionare casualmente 200 numeri da un universo `u` di 10000 numeri interi distribuiti normalmente con media 80 e deviazione standard pari a 12. Calcolare quindi la media sui 200 numeri, che chiameremo `media totale`. Permutare casualmente il campione di 200 numeri e suddividerlo in 20 gruppi da 10 numeri. Calcolare la media di ciascuno dei 20 gruppi e dimostrare empiricamente che la media delle medie dei 20 gruppi è uguale alla `media totale`. Cosa deducete da questo piccolo esperimento?
4. Scaricare i file `stormi.zip` dal sito di `elly.biosciienze.unipr.it`. Settando prima il seme uguale al vostro numero di matricola o, se non lo avete, al vostro anno di nascita, estrarre a sorte un numero intero fra 1 e 6, escludendo il 6 che abbiamo analizzato in classe,

³Il ‘>’ diventa ‘<’ se nel rapporto birra e l’acqua vanno rispettivamente al denominatore e al numeratore.

e analizzare il dataset corrispondente `C<numero estratto>.txt` riguardanti gli stormi per testare i 3 modelli spiegati a lezione e riportati a pag. 67 delle dispense. Utilizzando i criteri di cui si è parlato a lezione, scrivere quale ipotesi e modello/i si possono rifiutare e quali si intende non rifiutare e perché. Commentare i comandi e interpretare ecologicamente i risultati.

5. Eseguire i calcoli per SST , SSA e SSE , F e P , effettuare i calcoli per i contrasti di default di R sui dati contenuti nel file `oneway.txt` dentro al dataset di Crawley. I dati riguardano la concentrazione di ozono in due giardini di città. L'obiettivo è testare se i giardini differiscono della loro concentrazione media di ozono. Confrontare i risultati con quelli dell'anova e del `summary` di R. Commentate e interpretate i risultati.

Compito per il 13 dicembre 2017 Svolgete i seguenti esercizi:

1. L'obiettivo del compito è ripassare alcuni concetti importanti che già dovrete conoscere. Il materiale per scrivere le risposte si trova su un qualunque libro di statistica o in rete.
Scrivere in un breve testo (all'incirca 1500 caratteri), con parole vostre, le differenze importanti tra i cosiddetti test statistici *parametrici* e quelli *non parametrici*. Spiegare i vantaggi e gli svantaggi di entrambi gli approcci. Cercate di essere chiari e concisi.
2. In meno di 1000 caratteri spiegare, sempre con parole vostre, le assunzioni dell'analisi della varianza.
3. Utilizzando i dati contenuti nel file `twoway.txt` nel dataset di Crawley, effettuare un ANOVA a due fattori. Il data set contiene dati di crescita di un organismo con sei genotipi diversi cresciuti con quattro fotoperiodi diversi. L'obiettivo è testare se queste due variabili influenzano la crescita. Calcolare SSA , MSA , F e P dovute al Fotoperiodo, SSB , MSB , F e P dovuti al Genotipo e confrontarli con quelli dell'anova di R. Calcolare anche i valori delle `Estimates` del `summary`. Commentate comandi e risultati.
4. Utilizzando i dati contenuti nel file `lmedata3.txt` nel dataset di Crawley, effettuare un ANOVA a due fattori. Il data set contiene dati di crescita in termini di biomassa secca (`weight`) per sei classi di età e sei genotipi. I dati di età vanno considerati in termini non numerici, ma come semplici etichette per identificare una certa classe di età. L'obiettivo è valutare la crescita in funzione di età e genotipo e vedere se genotipi possano dare prestazioni particolari

in certe età. Effettuare l'anova con uno o più modelli appropriati. Commentare i comandi e risultati. Interpretare il significato biologico delle **Estimates** del summary del modello finale. Produrre uno o più grafici che illustrino i risultati.

Infine dare il seguente comando:

```
options(contrasts=c("contr.sum", "contr.poly")), rifare l'analisi e re-interpretare le Estimates del summary.
```

Compito per il 20 dicembre 2017 Svolgete il seguente esercizio:

1. Verificare se la seguente matrice abbia vettori linearmente dipendenti

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

Trovare gli a_i necessari per l'eventuale condizione di dipendenza lineare e dimostrarne l'effettiva dipendenza tramite l'uso dei coefficienti a_i .

2. Calcolare in R la matrice di Devianza-Codevianza e Varianza-Covarianza e Correlazione con i dati nella seguente tabella. Verificare il risultato di quest'ultima con la funzione apposita di R

Tabella:

x	y
35	27
44	22
17	18
45	35
0	3
22	19
46	42
90	73
53	42
30	20
67	43
20	9
7	1
37	27
19	5

Con la x della tabella precedente preparare in R la matrice aumentata \mathbf{X} (cioè la matrice con una colonna di 1 come prima colonna e la x come seconda colonna) e la matrice $\mathbf{X}^T\mathbf{X}$. Dire che tipo di matrice è quest'ultima e se è invertibile. Se lo è, invertirla e dimostrare che la matrice ottenuta è effettivamente l'inversa.

Calcolare, con l'algebra matriciale, le stime di pendenza e intercetta della regressione fra la x e la y della tabella precedente e confrontare i risultati con quelli della funzione apposita di R.

Effettuare l'analisi della varianza sulla regressione precedente e confrontare i risultati con l'`anova()` di R.

3. Utilizzando i dati contenuti nel file `tannin.txt` nel dataset di Crawley, effettuare una regressione con l'algebra matriciale. Il dataset contiene dati di crescita di alcune larve di insetti in funzione del contenuto in tannini della pianta ospite. Calcolare con l'algebra matriciale le stime di pendenza e intercetta della regressione e anche l'ANOVA corrispondente e confrontare i risultati con quelli delle funzioni ufficiali di R `lm()`, `summary()` e `anova()`. Commentate i comandi e i risultati dal punto di vista ecologico e produce un grafico che li illustri.
4. Utilizzando i dati contenuti nel file `ozone.data.txt` nel dataset di Crawley, effettuare una o più regressioni multiple con `lm` per trovare il modello che spieghi adeguatamente la concentrazione di ozono nell'aria. Presentare i risultati del modello più semplice possibile, ma che non abbia significativamente più errore di un modello più complesso. Provate ad utilizzare anche dei termini quadratici per la temperatura, la radiazione e il vento. Non includere i termini quadratici nelle interazioni. Commentate i comandi e i risultati. Verificate anche le assunzioni dell'ANOVA.

Compito per l'8 gennaio 2018 Svolgete i seguenti esercizi. Tenete presente che questi sono un buoni esempi degli esercizi che potrebbero capitare nel esame scritto finale.

1. Eseguite un'ANOVA con le *dummy variables* con l'algebra matriciale per i seguenti dati. Si tratta di dati sulla crescita in grammi di topi allevati con tre diete diverse denominate: "A", "B" e "C". Confrontare i risultati con quelli delle funzioni ufficiali di R `lm()`, `summary()` e `anova()`. Effettuare anche i confronti fra tutti i livelli a-due, a-due con metodo "Tukey's Honest Significant Difference" e confrontate, dove possibile, i risultati con quelli del `summary`,

tenendo presente che il metodo di Tukey è un metodo “a posteriori”, mentre il `summary` consente un test “a priori” ed è quindi più potente. Confrontate i risultati anche con quelli del test non-parametrico Kruskal-Wallis. Commentate i comandi e i risultati e i vari confronti e producite un grafico che illustri i risultati. Verificate anche le assunzioni dell’ANOVA.

A	B	C
8.1	16.5	18.8
10.6	13.2	18.3
7.5	10.1	16.2
-	5.4	10.0
11.0	15.4	17.9
7.5	11.9	15.8
11.5	12.0	-
12.2	14.8	11.6
11.7	14.5	14.6
9.1	13.8	17.3

- Utilizzando i dati contenuti nel file `ipomopsis.txt` nel dataset di Crawley, effettuare un’analisi della covarianza per spiegare se la fruttificazione (*Fruit*) è influenzata dalla grandezza delle radici (*Root*) e dal trattamento: Pascolato/Non pascolato. Commentate i comandi e i risultati, producite un grafico che li illustri e interpretate i coefficienti del `summary`. Verificate anche le assunzioni dell’ANOVA.
- Utilizzando i dati contenuti nel dataset `Orange` che è disponibile in R, effettuare un’analisi della covarianza per spiegare la crescita in circonferenza degli alberi in funzione dell’età e dell’albero. Trasformate il numero dell’albero in un fattore. Cercate informazioni e unità di misura delle variabili con il comando `?Orange`. Commentate i comandi e i risultati, producite un grafico che li illustri e interpretate i coefficienti del `summary`. Verificate e commentate anche le assunzioni dell’ANOVA.