

Guida Completa ai Comandi R per l'Analisi Esplorativa dei Dati

Basato sulla dispensa di Nicola Torelli

17 giugno 2025

Indice

1	Introduzione	3
2	Caricamento di Dati e Pacchetti	3
2.1	Caricamento di Pacchetti	3
2.2	Caricamento di Dataset	3
3	Esplorazione Iniziale dei Dati	3
3.1	Struttura e Dimensioni	3
3.2	Riassunti Descrittivi	4
4	Tabelle di Frequenza	4
4.1	Tabelle Semplici	4
4.2	Tabelle a Doppia Entrata	4
5	Grafici per Variabili Categoriali	4
5.1	Diagramma a Torta	4
5.2	Diagramma a Barre	5
6	Grafici per Variabili Quantitative	5
6.1	Diagramma Ramo e Foglie	5
6.2	Diagramma a Punti	5
6.3	Istogramma	6
6.4	Box Plot	6
6.5	Funzione di Ripartizione Empirica	6
7	Indici Numerici	7
7.1	Indici di Tendenza Centrale	7
7.2	Indici di Dispersione	7
7.3	Indici di Forma	7
8	Gestione dei Dati Mancanti	8
8.1	Identificazione e Conteggio	8
8.2	Gestione Parametrica	8

9	Trasformazioni delle Variabili	8
9.1	Trasformazioni Quantitative	8
9.2	Trasformazioni Categoricali	9
10	Analisi Bivariata	9
10.1	Due Variabili Categoricali	9
10.2	Variabile Quantitativa per Gruppi	9
10.3	Due Variabili Quantitative	9
11	Regressione Lineare	10
11.1	Regressione Semplice	10
11.2	Regressione Multipla	10
11.3	Diagnostiche del Modello	11
12	Analisi Multivariata	11
12.1	Matrici di Correlazione	11
12.2	Grafici Multipli	11
13	Analisi di Raggruppamento (Clustering)	11
13.1	Calcolo delle Distanze	11
13.2	K-means	12
13.3	Clustering Gerarchico	12
13.4	PAM (Partitioning Around Medoids)	12
14	Grafici con ggplot2	13
14.1	Grafici Base	13
14.2	Grafici Avanzati	13
15	Test Statistici	14
15.1	Test per Variabili Categoricali	14
15.2	Test per Variabili Quantitative	14
16	Q-Q Plot e Test di Normalità	14
17	Lisciamento e Densità	14
18	Parametri Grafici	15
18.1	Controllo della Finestra Grafica	15
18.2	Personalizzazione Grafica	15
19	Funzioni Utili	15
19.1	Manipolazione Dati	15
19.2	Controllo e Debug	16
20	Esempio Completo: Analisi Esplorativa	17
21	Risoluzione Problemi Comuni	18
21.1	Errori Frequenti	18
22	Risorse Aggiuntive	18

1 Introduzione

Questa guida raccoglie tutti i comandi R presenti nel corso di "Analisi dei dati - Introduzione all'analisi esplorativa dei dati con R", organizzati per argomento con spiegazioni dettagliate ed esempi pratici.

2 Caricamento di Dati e Pacchetti

2.1 Caricamento di Pacchetti

```
1 # Caricamento di pacchetti essenziali
2 library(MASS)           # Contiene dataset e funzioni statistiche
3 library(insuranceData)  # Dataset assicurativi
4 library(lattice)        # Grafici a pannelli multipli
5 library(ggplot2)        # Grafici avanzati
6 library(cluster)        # Analisi di raggruppamento
7 library(moments)        # Calcolo di asimmetria e curtosi
```

2.2 Caricamento di Dataset

```
1 # Caricamento di dataset incorporati
2 data(Cars93)            # Dataset automobili
3 data(AutoBi)            # Dataset sinistri assicurativi
4 data(iris)              # Dataset fiori di iris
5 data(eurodist)          # Distanze tra città europee
6
7 # Lettura da file CSV
8 dati <- read.csv("nomefile.csv", header=TRUE, sep=",")
```

Nota Importante

Il parametro `header=TRUE` indica che la prima riga contiene i nomi delle variabili.
Il parametro `sep=","` specifica il separatore dei campi.

3 Esplorazione Iniziale dei Dati

3.1 Struttura e Dimensioni

```
1 # Struttura del dataset
2 str(Cars93)             # Mostra struttura e tipi di variabili
3 dim(Cars93)             # Dimensioni: righe x colonne
4 nrow(Cars93)            # Numero di righe
5 ncol(Cars93)            # Numero di colonne
6 names(Cars93)           # Nomi delle variabili
7 head(Cars93)            # Prime 6 righe
8 tail(Cars93)            # Ultime 6 righe
9 head(Cars93, n=10)      # Prime 10 righe
```

3.2 Riassunti Descrittivi

```
1 # Riassunto generale
2 summary(Cars93)          # Riassunto per tutte le variabili
3
4 # Per variabili quantitative
5 summary(Cars93$Price)    # Riassunto per singola variabile
6 mean(Cars93$Price)       # Media
7 median(Cars93$Price)     # Mediana
8 var(Cars93$Price)        # Varianza
9 sd(Cars93$Price)         # Deviazione standard
```

4 Tabelle di Frequenza

4.1 Tabelle Semplici

```
1 # Tabella di frequenze assolute
2 table(AutoBi$ATTORNEY)
3
4 # Tabella di frequenze relative
5 prop.table(table(AutoBi$ATTORNEY))
6
7 # Tabella con valori mancanti
8 table(AutoBi$MARITAL, useNA="ifany")
```

4.2 Tabelle a Doppia Entrata

```
1 # Tabella di contingenza
2 tab1 <- table(AutoBi$ATTORNEY, AutoBi$CLMSEX)
3 tab1
4
5 # Frequenze relative per riga
6 prop.table(tab1, 1)
7
8 # Frequenze relative per colonna
9 prop.table(tab1, 2)
10
11 # Frequenze relative sul totale
12 prop.table(tab1)
```

5 Grafici per Variabili Categoricali

5.1 Diagramma a Torta

```
1 # Diagramma a torta semplice
2 tabtipo <- table(Cars93$Type)
```

```

3 pie(tabtipo)
4
5 # Con personalizzazioni
6 pie(tabtipo, main="Distribuzione per Tipo di Auto")

```

Nota Importante

Il diagramma a torta è sconsigliato per l'analisi dei dati. È preferibile usare il diagramma a barre.

5.2 Diagramma a Barre

```

1 # Diagramma a barre semplice
2 maritab <- table(AutoBi$MARITAL)
3 barplot(maritab)
4
5 # Con personalizzazioni
6 barplot(maritab,
7         main="Stato Civile",
8         names.arg=c("sposato", "single", "vedovo", "divorziato"),
9         col="blue")
10
11 # Barre affiancate per confronti
12 barplot(table(AutoBi$ATTORNEY, AutoBi$CLMSEX),
13         beside=TRUE,
14         legend=TRUE)
15
16 # Barre sovrapposte
17 barplot(prop.table(table(AutoBi$ATTORNEY, AutoBi$LOSSclass), 2),
18         legend=TRUE)

```

6 Grafici per Variabili Quantitative

6.1 Diagramma Ramo e Foglie

```

1 # Stem-and-leaf plot
2 stem(Cars93$Length)

```

6.2 Diagramma a Punti

```

1 # Stripchart
2 stripchart(Cars93$Length, pch=19, method="stack", cex=1.2)
3
4 # Dotchart
5 dotchart(Cars93$Length)
6
7 # Plot semplice (mostra indice vs valore)

```

```
8 plot(Cars93$Length)
```

6.3 Istogramma

```
1 # Istogramma base
2 hist(Cars93$Length)
3
4 # Istogramma con densit 
5 hist(Cars93$Length, prob=TRUE)
6
7 # Personalizzazioni
8 hist(Cars93$Length,
9       prob=TRUE,
10      breaks=12,
11      main="Distribuzione Lunghezza Auto",
12      xlab="Lunghezza (pollici)",
13      col="lightblue")
14
15 # Con classi personalizzate
16 hist(Cars93$Length,
17       breaks=c(140,160,170,180,190,200,220),
18       prob=TRUE)
```

6.4 Box Plot

```
1 # Box plot semplice
2 boxplot(Cars93$Length)
3
4 # Box plot orizzontale
5 boxplot(Cars93$Length, horizontal=TRUE)
6
7 # Box plot multipli
8 boxplot(Sepal.Length ~ Species, data=iris)
9
10 # Con personalizzazioni
11 boxplot(Sepal.Length ~ Species,
12         data=iris,
13         main="Lunghezza Sepali per Specie",
14         xlab="Specie",
15         ylab="Lunghezza Sepalo",
16         col=c("red", "green", "blue"))
```

6.5 Funzione di Ripartizione Empirica

```
1 # Funzione di ripartizione empirica
2 plot(ecdf(Cars93$Length))
3
4 # Con personalizzazioni
```

```

5 plot(ecdf(Cars93$Length),
6      main="Funzione di Ripartizione Empirica",
7      xlab="Lunghezza",
8      ylab="Proporzione")

```

7 Indici Numerici

7.1 Indici di Tendenza Centrale

```

1 # Media aritmetica
2 mean(AutoBi$LOSS, na.rm=TRUE)
3
4 # Media sfrondata (trimmed)
5 mean(AutoBi$LOSS, na.rm=TRUE, trim=0.05)
6
7 # Mediana
8 median(AutoBi$LOSS, na.rm=TRUE)
9
10 # Quantili
11 quantile(AutoBi$LOSS, na.rm=TRUE) # Quartili
12 quantile(AutoBi$LOSS, probs=seq(0,1,0.1), na.rm=TRUE) # Decili
13 quantile(AutoBi$LOSS, probs=0.5, na.rm=TRUE) # Mediana
14
15 # Riassunto dei 5 numeri
16 fivenum(AutoBi$LOSS)

```

7.2 Indici di Dispersione

```

1 # Varianza
2 var(AutoBi$LOSS, na.rm=TRUE)
3
4 # Deviazione standard
5 sd(AutoBi$LOSS, na.rm=TRUE)
6
7 # Scarto interquartile
8 Q1 <- quantile(AutoBi$LOSS, 0.25, na.rm=TRUE)
9 Q3 <- quantile(AutoBi$LOSS, 0.75, na.rm=TRUE)
10 SI <- Q3 - Q1
11
12 # MAD (Median Absolute Deviation)
13 mad(AutoBi$LOSS, constant=1, na.rm=TRUE)

```

7.3 Indici di Forma

```

1 # Asimmetria (skewness) - richiede library(moments)
2 library(moments)
3 skewness(AutoBi$CLMAGE, na.rm=TRUE)

```

```

4
5 # Curtosi - richiede library(moments)
6 kurtosis(AutoBi$CLMAGE, na.rm=TRUE)
7
8 # Indice di Galton (basato sui quartili)
9 Qloss <- fivenum(AutoBi$LOSS)
10 galton <- (Qloss[4] + Qloss[2] - 2*Qloss[3]) / (Qloss[4] - Qloss
    [2])

```

8 Gestione dei Dati Mancanti

8.1 Identificazione e Conteggio

```

1 # Verifica presenza di NA
2 is.na(AutoBi$CLMAGE)           # Vettore logico
3 anyNA(AutoBi$CLMAGE)           # TRUE se ci sono NA
4 sum(is.na(AutoBi$CLMAGE))      # Conteggio NA
5
6 # Rimozione casi con NA
7 AutoBi_complete <- na.omit(AutoBi) # Rimuove righe con NA
8 anyNA(AutoBi_complete)         # Verifica

```

8.2 Gestione Parametrica

```

1 # Molte funzioni hanno il parametro na.rm
2 mean(AutoBi$CLMAGE, na.rm=TRUE)
3 var(AutoBi$CLMAGE, na.rm=TRUE)
4 sd(AutoBi$CLMAGE, na.rm=TRUE)

```

9 Trasformazioni delle Variabili

9.1 Trasformazioni Quantitative

```

1 # Standardizzazione
2 AutoBi$LOSS_std <- scale(AutoBi$LOSS)
3
4 # Trasformazione logaritmica
5 AutoBi$LOSS_log <- log(AutoBi$LOSS)
6
7 # Radice quadrata
8 AutoBi$LOSS_sqrt <- sqrt(AutoBi$LOSS)
9
10 # Normalizzazione min-max
11 AutoBi$LOSS_norm <- (AutoBi$LOSS - min(AutoBi$LOSS, na.rm=TRUE))
    /
12                               (max(AutoBi$LOSS, na.rm=TRUE) - min(AutoBi$
    LOSS, na.rm=TRUE))

```


9.2 Trasformazioni Categoriali

```
1 # Creazione di classi da variabile quantitativa
2 AutoBi$LOSSclass <- cut(AutoBi$LOSS,
3                          breaks=c(0, 0.5, 2, 4, 8, 1100),
4                          labels=c("Molto Basso", "Basso", "Medio",
5                                  "Alto", "Molto Alto"))
6
7 # Ricodifica di fattori
8 AutoBi$ATTORNEY <- factor(AutoBi$ATTORNEY)
9 levels(AutoBi$ATTORNEY) <- c("yes", "no")
10
11 # Cambio del livello di riferimento
12 AutoBi$MARITAL <- relevel(AutoBi$MARITAL, ref="single")
```

10 Analisi Bivariata

10.1 Due Variabili Categoricali

```
1 # Tabella di contingenza
2 tab <- table(AutoBi$ATTORNEY, AutoBi$CLMSEX)
3
4 # Test Chi-quadrato
5 chisq.test(tab)
6 chisq.test(tab, correct=FALSE) # Senza correzione di continuit 
7
8 # Odds ratio (per tabelle 2x2)
9 OR <- (tab[1,1] * tab[2,2]) / (tab[1,2] * tab[2,1])
10 LOR <- log(OR) # Log odds ratio
```

10.2 Variabile Quantitativa per Gruppi

```
1 # Box plot multipli
2 boxplot(Sepal.Length ~ Species, data=iris)
3
4 # Confronto funzioni di ripartizione
5 plot(ecdf(iris$Sepal.Length[iris$Species=="setosa"]),
6      main="Confronto Funzioni di Ripartizione")
7 plot(ecdf(iris$Sepal.Length[iris$Species=="versicolor"]),
8      col="red", add=TRUE)
9
10 # ANOVA
11 testanova <- aov(Sepal.Length ~ Species, data=iris)
12 summary(testanova)
```

10.3 Due Variabili Quantitative

```

1 # Diagramma di dispersione
2 plot(Cars93$Weight, Cars93$MPG.city,
3       xlab="Peso", ylab="Consumi Urbani",
4       main="Relazione Peso-Consumi")
5
6 # Correlazione
7 cor(Cars93$Weight, Cars93$MPG.city, use="complete.obs")
8
9 # Covarianza
10 cov(Cars93$Weight, Cars93$MPG.city, use="complete.obs")

```

11 Regressione Lineare

11.1 Regressione Semplice

```

1 # Modello di regressione semplice
2 modello <- lm(Gas ~ Temp, data=whiteside)
3 summary(modello)
4
5 # Coefficienti
6 coef(modello)                # Tutti i coefficienti
7 modello$coefficients[1]      # Intercetta
8 modello$coefficients[2]      # Coefficiente angolare
9
10 # Valori predetti e residui
11 yhat <- fitted(modello)      # Valori predetti
12 residui <- residuals(modello) # Residui
13
14 # Aggiungere la retta al grafico
15 plot(whiteside$Temp, whiteside$Gas)
16 abline(modello, col="red")

```

11.2 Regressione Multipla

```

1 # Modello con pi variabili
2 modello_mult <- lm(Gas ~ Temp + Insul, data=whiteside)
3 summary(modello_mult)
4
5 # Modello con interazione
6 modello_int <- lm(Gas ~ Temp + Insul + Temp:Insul, data=whiteside)
7
8 # Equivalente a:
9 modello_int <- lm(Gas ~ Temp * Insul, data=whiteside)
10
11 # Selezione automatica (stepwise)
12 modello_step <- step(modello_mult, direction="both")

```

11.3 Diagnostiche del Modello

```
1 # Grafici diagnostici
2 par(mfrow=c(2,2))
3 plot(modello) # 4 grafici standard
4 par(mfrow=c(1,1))
5
6 # Grafici specifici
7 plot(modello, which=1) # Residui vs Fitted
8 plot(modello, which=2) # Q-Q plot
9 plot(modello, which=3) # Scale-Location
10 plot(modello, which=4) # Cook's distance
```

12 Analisi Multivariata

12.1 Matrici di Correlazione

```
1 # Matrice di correlazione
2 matrice_cor <- cor(iris[,-5]) # Esclude variabile categoriale
3 matrice_cor
4
5 # Matrice di covarianza
6 matrice_cov <- cov(iris[,-5])
7
8 # Visualizzazione matrice correlazioni (richiede ggcorrplot)
9 library(ggcorrplot)
10 ggcorrplot(matrice_cor, type="lower", lab=TRUE)
```

12.2 Grafici Multipli

```
1 # Matrice di scatterplot
2 pairs(iris[,-5])
3
4 # Con ggplot2 esteso (richiede GGally)
5 library(GGally)
6 ggpairs(iris)
7
8 # Grafici con lattice
9 library(lattice)
10 histogram(~ Sepal.Length | Species, data=iris, layout=c(1,3))
11 densityplot(~ Sepal.Length | Species, data=iris, layout=c(1,3))
```

13 Analisi di Raggruppamento (Clustering)

13.1 Calcolo delle Distanze

```

1 # Distanze euclidee
2 dist_euclidea <- dist(iris[, -5])
3 dist_euclidea <- daisy(iris[, -5])
4
5 # Distanze con standardizzazione
6 dist_std <- daisy(iris[, -5], stand=TRUE)
7
8 # Distanza di Gower (per dati misti)
9 dist_gower <- daisy(iris, metric="gower")

```

13.2 K-means

```

1 # Clustering K-means
2 set.seed(123)
3 km_result <- kmeans(iris[, -5], centers=3, nstart=20)
4
5 # Risultati
6 km_result$cluster      # Assegnazione ai cluster
7 km_result$centers      # Centroidi
8 km_result$size         # Dimensione dei cluster
9
10 # Visualizzazione
11 plot(iris$Sepal.Length, iris$Petal.Length,
12       col=km_result$cluster, pch=19)
13
14 # Confronto con classificazione vera
15 table(km_result$cluster, iris$Species)

```

13.3 Clustering Gerarchico

```

1 # Clustering gerarchico agglomerativo
2 library(cluster)
3 hc_result <- agnes(iris[, -5], method="ward")
4
5 # Dendrogramma
6 pltree(hc_result, hang=-1, cex=0.8)
7
8 # Taglio del dendrogramma
9 cluster_labels <- cutree(hc_result, k=3)
10
11 # Aggiungere rettangoli al dendrogramma
12 rect.hclust(hc_result, k=3, border="red")

```

13.4 PAM (Partitioning Around Medoids)

```

1 # Clustering PAM
2 pam_result <- pam(iris[, -5], k=3)

```

```

3
4 # Medoidi
5 pam_result$medoids
6
7 # Grafico silhouette
8 plot(pam_result, which=2) # Silhouette plot

```

14 Grafici con ggplot2

14.1 Grafici Base

```

1 library(ggplot2)
2
3 # Scatterplot
4 ggplot(iris, aes(x=Sepal.Length, y=Petal.Length)) +
5   geom_point()
6
7 # Con colori per gruppi
8 ggplot(iris, aes(x=Sepal.Length, y=Petal.Length, color=Species))
9   +
10  geom_point()
11
12 # Istogramma
13 ggplot(iris, aes(x=Sepal.Length)) +
14   geom_histogram(bins=20)
15
16 # Box plot
17 ggplot(iris, aes(x=Species, y=Sepal.Length)) +
18   geom_boxplot()

```

14.2 Grafici Avanzati

```

1 # Con linea di tendenza
2 ggplot(iris, aes(x=Sepal.Length, y=Petal.Length)) +
3   geom_point() +
4   geom_smooth(method="lm")
5
6 # Curve loess per gruppi
7 ggplot(iris, aes(x=Sepal.Length, y=Petal.Length, color=Species))
8   +
9   geom_point() +
10  geom_smooth(se=FALSE)
11
12 # Faceting (pannelli multipli)
13 ggplot(iris, aes(x=Sepal.Length, y=Petal.Length)) +
14   geom_point() +
15   facet_wrap(~Species)

```

15 Test Statistici

15.1 Test per Variabili Categoricali

```
1 # Test Chi-quadrato
2 chisq.test(table(AutoBi$ATTORNEY, AutoBi$CLMSEX))
3
4 # Test esatto di Fisher (per tabelle 2x2)
5 fisher.test(table(AutoBi$ATTORNEY, AutoBi$CLMSEX))
```

15.2 Test per Variabili Quantitative

```
1 # Test t per confronto di medie
2 t.test(Sepal.Length ~ Species,
3        data=iris[iris$Species %in% c("setosa", "versicolor"),])
4
5 # ANOVA
6 summary(aov(Sepal.Length ~ Species, data=iris))
7
8 # Test di Kolmogorov-Smirnov per confronto distribuzioni
9 ks.test(iris$Sepal.Length[iris$Species=="setosa"],
10         iris$Sepal.Length[iris$Species=="versicolor"])
```

16 Q-Q Plot e Test di Normalità

```
1 # Q-Q plot contro normale
2 qqnorm(iris$Sepal.Length)
3 qqline(iris$Sepal.Length)
4
5 # Q-Q plot tra due campioni
6 qqplot(iris$Sepal.Length[iris$Species=="setosa"],
7        iris$Sepal.Length[iris$Species=="versicolor"])
8
9 # Test di normalit 
10 shapiro.test(iris$Sepal.Length) # Test di Shapiro-Wilk
```

17 Lisciamiento e Densit 

```
1 # Stima della densit  con metodo del nucleo
2 densita <- density(iris$Sepal.Length)
3 plot(densita)
4
5 # Sovrapposizione a istogramma
6 hist(iris$Sepal.Length, prob=TRUE)
7 lines(density(iris$Sepal.Length), col="red", lwd=2)
8
```

```

9 # Regressione liscia (lowess/loess)
10 plot(Cars93$Weight, Cars93$MPG.city)
11 lines(lowess(Cars93$Weight, Cars93$MPG.city), col="red")

```

18 Parametri Grafici

18.1 Controllo della Finestra Grafica

```

1 # Divisione della finestra grafica
2 par(mfrow=c(2,2)) # 2x2 pannelli
3 par(mfrow=c(1,2)) # 1x2 pannelli
4 par(mfrow=c(1,1)) # Torna a singolo pannello
5
6 # Margini
7 par(mar=c(4,4,2,1)) # bottom, left, top, right

```

18.2 Personalizzazione Grafica

```

1 # Simboli e colori
2 plot(x, y, pch=19) # Punto pieno
3 plot(x, y, pch=1) # Cerchio vuoto
4 plot(x, y, col="red") # Colore rosso
5 plot(x, y, col=2) # Colore rosso (numero)
6 plot(x, y, cex=1.5) # Dimensione simboli
7
8 # Linee
9 plot(x, y, type="l") # Solo linee
10 plot(x, y, type="b") # Punti e linee
11 plot(x, y, type="o") # Punti sovrapposti alle linee
12 lines(x, y, lwd=2) # Spessore linea
13 lines(x, y, lty=2) # Tipo linea (tratteggiata)
14
15 # Etichette e titoli
16 plot(x, y, main="Titolo", xlab="Asse X", ylab="Asse Y")
17
18 # Aggiungere elementi
19 abline(h=0) # Linea orizzontale
20 abline(v=0) # Linea verticale
21 abline(a, b) # Retta con intercetta a e pendenza b
22 text(x, y, "etichetta") # Testo
23 legend("topright", legend=c("A", "B"), col=c(1,2), pch=c(1,2))

```

19 Funzioni Utili

19.1 Manipolazione Dati

```

1 # Selezione di subset
2 subset_data <- subset(iris, Species == "setosa")
3 subset_data <- iris[iris$Species == "setosa", ]
4
5 # Ordinamento
6 iris_ordinato <- iris[order(iris$Sepal.Length), ]
7
8 # Funzioni apply
9 apply(iris[,1:4], 2, mean)      # Media per colonna
10 sapply(iris[,1:4], mean)      # Stesso risultato, output
    semplificato

```

19.2 Controllo e Debug

```

1 # Controllo struttura
2 str(iris)                      # Struttura
3 class(iris$Species)           # Classe di una variabile
4 levels(iris$Species)          # Livelli di un fattore
5
6 # Informazioni sulla sessione
7 sessionInfo()                  # Versione R e pacchetti caricati

```


20 Esempio Completo: Analisi Esplorativa

Esempio

```
1 # Esempio completo di analisi esplorativa
2 # 1. Caricamento dati
3 data(iris)
4
5 # 2. Esplorazione iniziale
6 str(iris)
7 summary(iris)
8
9 # 3. Analisi univariata
10 # Variabile quantitativa
11 hist(iris$Sepal.Length, prob=TRUE, main="Distribuzione
    Lunghezza Sepalo")
12 lines(density(iris$Sepal.Length), col="red", lwd=2)
13
14 # Statistiche descrittive
15 mean(iris$Sepal.Length)
16 sd(iris$Sepal.Length)
17 quantile(iris$Sepal.Length)
18
19 # Variabile categoriale
20 table(iris$Species)
21 barplot(table(iris$Species), main="Distribuzione Specie")
22
23 # 4. Analisi bivariata
24 # Quantitativa vs Categoriale
25 boxplot(Sepal.Length ~ Species, data=iris)
26 summary(aov(Sepal.Length ~ Species, data=iris))
27
28 # Quantitativa vs Quantitativa
29 plot(iris$Sepal.Length, iris$Petal.Length)
30 cor(iris$Sepal.Length, iris$Petal.Length)
31
32 # Regressione
33 modello <- lm(Petal.Length ~ Sepal.Length, data=iris)
34 summary(modello)
35 abline(modello, col="red")
36
37 # 5. Analisi multivariata
38 pairs(iris[, -5])
39 cor(iris[, -5])
40
41 # Clustering
42 km <- kmeans(iris[, -5], 3)
43 table(km$cluster, iris$Species)
```

21 Risoluzione Problemi Comuni

21.1 Errori Frequenti

Nota Importante

Errore: "object not found"

Soluzione: Verificare che il dataset sia caricato e i nomi delle variabili siano corretti.

Errore: "missing values"

Soluzione: Aggiungere `na.rm=TRUE` alle funzioni o rimuovere i dati mancanti.

Errore: Grafici non visualizzati

Soluzione: Controllare le impostazioni grafiche con `par()` e assicurarsi che i dispositivi grafici siano aperti.

22 Risorse Aggiuntive

- **Help in R:** `?nome_funzione` o `help(nome_funzione)`
- **Esempi:** `example(nome_funzione)`
- **Vignette:** `vignette(package="nome_pacchetto")`
- **Ricerca:** `??termine_ricerca`