Nota 1. Come accorgimento pratico, conviene creare un semplice file testo in cui scrivere i comandi usati, es. quelli di caricamento dati o di calcolo di una media. Questi comandi verranno usati per tutto il corso ed è quindi molto più semplice ricuperarli da questo file piuttosto che tutte le volte con l'help o dalle dispense.

Nota 2. Il programma R si scarica dal sito

$$http: //www.r-project.org/$$

Se si vuole trovare una funzione - es. quella che calcola un valor medio - si può usare "Cerca nella guida" immettendo il nome inglese, es. mean, deviation, histogram, o una parte di esso. Escono le voci interessate, con l'indicazione del Package in cui si trovano. I package si aprono da "Guida Html". Aprendoli si trovano i dettagli delle varie funzioni. Però spesso i dettagli non servono: l'uso istintivo dei comandi è in genere quello giusto, es. mean(x).

Esercizio 1.1 (caricamento dati).

 $Enunciato \ in forma \ concisa:$ caricare in Ri dati Istat regionali sull'occupazione del 2002.

Enunciato in forma estesa. Trovare sul sito Istat

sotto "mercato del lavoro" i dati regionali 2002 del "tasso di occupazione". Copiarli come testo in uno dei due modi seguenti:

- il più diretto è un file tipo testo (es. con "Blocco Note" di Windows). Conviene accorciare il nome delle regioni (es. Piem, VaAo, ecc.), bisogna eliminare gli spazi nei nomi delle regioni, ed infine sostituire la virgola dei decimali col punto (notazione internazionale), ad es. col sostituisci automatico. Salvate il file col nome "dati_occup" in una cartella di cui conoscete il percorso.
- Il meno diretto ma più utile per eseguire altri esercizi più elaborati, è salvare i dati su un foglio Excel (o editore analogo). Anche in tal caso vanno accorciati i nomi, eliminati gli spazi nei nomi, trasformata la virgola in punto (si cambia sotto "Opzioni", "Internazionale"). Poi, oltre a salvare il foglio Excel per successive elaborazioni, va salvato come "testo delimitato da tabulazione". Aprirlo con un editore appropriato (es. "Blocco Note" di Windows) per verificare il contenuto.

Nota: dai dati, eliminate Bolzano, Trento ed i riassunti nazionali. Tenete cioè solo le regioni.

Ottenuto così un file di testo con due colonne, la prima coi nomi delle regioni, la seconda con le percentuali di occupazione, caricarlo da R col comando

Visualizzare occup. Devono comparire tre colonne: la prima di numerazione, poi la colonna V1 coi nomi delle regioni e la V1 con le percentuali di occupazione.

Esercizio 1.2 (calcolo indicatori elementari).

Enunciato in forma concisa: calcolare media, deviazione standard e standardizzazione, per i dati dell'esercizio precedente..

Enunciato in forma estesa. La media empirica (o campionaria) di un campione $x_1, ..., x_n$ è

 $\overline{x} = \frac{x_1 + \dots + x_n}{n}$

che indichiamo anche con $\widehat{\mu}$. La deviazione standard empirica s, che indichiamo anche con $\widehat{\sigma}$, è la radice quadrata della varianza campionaria

$$s^{2} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{k} - \overline{x})^{2} = \frac{1}{n-1} \sum_{k=1}^{n} x_{k}^{2} - \frac{n}{n-1} \overline{x}^{2}.$$

Calcolare con R e facoltativamente con Excel, media e deviazione e varianza empiriche del campione regionale caricato all'esercizio precedente. Infine, calcolare il campione standardizzato (cioè quello di elementi $\frac{x_k - \overline{x}}{s}$). Eventualmente, calcolarne media e dev.st empiriche.

[Note: su Excel media e dev.st si trovano tra le funzioni statistiche. Su R il primo problema è scorporare la colonna V2 dalla tabella occup. Si deve allora introdurre il vettore di dati

$$x < -occup\$V2$$

Lo si visualizzi, scrivendo x. Ora si possnono calcolare mean e deviation di x, cercando il comando da dare con l'help, "cerca nella guida".] [Soluzioni: $\overline{x}=44.107,\,s=6.583622$.]

Esercizio 1.3 (altro caricamento dati).

Enunciato in forma concisa: fare in modo che i dati dell'esercizio 1.1, quando vengono caricati da R, appaiano disposti solo su due colonne: la prima, senza titolo, con i nomi delle regioni; la seconda, col titolo occup, con i dati numerici.

Enunciato in forma estesa. Modificare la procedura dell'esercizio 1.1 in modo da avere un file testo la cui prima riga sia composta da: uno spazio (es. di tabulazione) sopra i nomi delle regioni, un nome, es. occup, sopra le percentuali di occupazione. Dare il nome $dati_occup2$ a questo file. Caricarlo con R, dando alla tabella il nome occup2 (cioè col comando occup2 < -read.table ecc.). Leggere occup2 e riconoscere che questo è un modo molto più naturale di aver caricato i dati. Creare ora

$$x2 < -occup2\$occup$$

e calcolare media e dev.st empiriche di x2.