Esercizio 2.1 (istogramma delle frequenze). Da R, caricare il file *dati_occup*2 col nome *occup*2 e creare $x < -occup2\$occup$ (visualizzare *occup*2 e x, per sicurezza). Visualizzare un istogramma del campione x. Eventualmente, se si hanno i dati su Excel, fare lo stesso con questo programma (sotto "Analisi dati").

Esercizio 2.2 (istogramma delle frequenze *relative*). Aprire nel package "Graphics" la funzione *Histograms*, in modo da scoprire come si può ottenere un istogramma delle frequenze relative, o probabilità empiriche. Rammentare che il rapporto $\frac{n_k}{n}$ tra il numero di occorrenze nella classe k-esima e la numerosità totale n del campione, dovrebbe essere un'approssimazione della probabilità che la v.a. prenda valori nella classe k-esima. Per questo $\frac{n_k}{n}$ porta il nome di *probabilità empirica*. Tracciare quindi l'istogramma delle probabilità empiriche dei dati x. Facoltativamente, cambiare le classi di suddivisione dei dati x a propria scelta e ritracciare l'istogramma, come descritto nella funzione *Histograms*.

Esercizio 2.3 (istogramma cumulativo). Lo scopo di questo esercizio è di raffiguare la funzione di distribuzione cumulativa empirica $\widehat{F}(x)$ del campione in oggetto. Questo comando è automatizzato: *plot.ecdf*. Provare ad usarlo. Però, a titolo di esercizio, si suggerisce di tracciare l'istogramma cumulativo in un modo non automatico, ad esempio il seguente.

Un modo non automatico consiste nel seguire passo per passo la definizone di $\widehat{F}(x)$: si ordina il campione in modo crescente (se non si riesce, vedere il suggerimento), si associa $\frac{1}{n}$ al primo elemento (il più piccolo), $\frac{2}{n}$ al secondo, e così via, fino ad associare 1 all'ultimo (se non si riesce, vedere il suggerimento). Si raffigura quindi questa funzione. Eseguire eventualmente l'esercizio anche con Excel.

[Cercare sort nel package base. Il vettore $y < -1 : n$ è fatto dei numeri da 1 a n. La moltiplicazione di un vettore y per un numero a è semplicemente $a * y$. Dati x, y della stessa lunghezza col significato che $y[i]$ è il valore della funzione y nel punto $x[i]$, plot(x,y) produce un plot per punti di questa funzione. Altre modalità di plot si trovano con l'help.]

Esercizio 2.4 (manipolazione dati regionali della prima lezione). Si invita a caricare da R il file di dati *indicatori_benessere.txt*, visualizzarli, tracciare il grafico delle componenti principali ed una suddivisione in cluster.

[Questo esercizio è risolto nel file *comandi_lezione*1*.txt*]