

LABORATORIO-MINITAB N. 2-3
VARIABILI QUANTITATIVE

DESCRIZIONE DEI DATI DA ESAMINARE

Sono stati raccolti i dati sul peso del polmone di topi normali e affetti da una patologia simile alla distrofia muscolare degli uomini maschi. Dopo un numero fissato di settimane i topi sono stati sacrificati e i loro polmoni sono stati pesati. È stata anche rilevata l'età. Si vuole capire se il peso del polmone è un indicatore della patologia considerata.

Numero dei dati raccolti: 157

Nomi delle variabili:

1. Peso Polmone: il peso del polmone del topo espresso in grammi. (ND significa non disponibile)
2. Età: l'età del topo espressa in settimane.
3. Topo: il tipo di topo codificato con C57 (normale) o mdx (affetto da distrofia muscolare)

I dati sono raccolti nel dataset: [topi](#)

Procediamo con un'analisi descrittiva delle singole variabili. Notiamo che la variabile topo è qualitativa e anche la variabile Età si può considerare come qualitativa.

Tabella di contingenza a una via

Columns: Topo

"C57"	"mdx"	All
50,96	49,04	100,00

Cell Contents: % of Total

Commenti: il 51% dei topi analizzati è normale e il 49% è affetto da distrofia muscolare.

Ci sono alcune unità sperimentali per cui la misura del peso del polmone non è stata rilevata. Si deve quindi procedere a eliminare queste unità sperimentali dal dataset.

Eliminazione di particolari unità sperimentali dal dataset

Data→Subset Worksheet

Selezionare "Specify which rows to exclude" e scegliere come "Condition" "Peso polmone=ND"

A questo punto il numero di unità sperimentali è sceso a 137 unità; anche in questo sottoinsieme di dati il 51% dei topi normale e il 49% affetto da distrofia muscolare.

Principali indici statistici per variabili quantitative

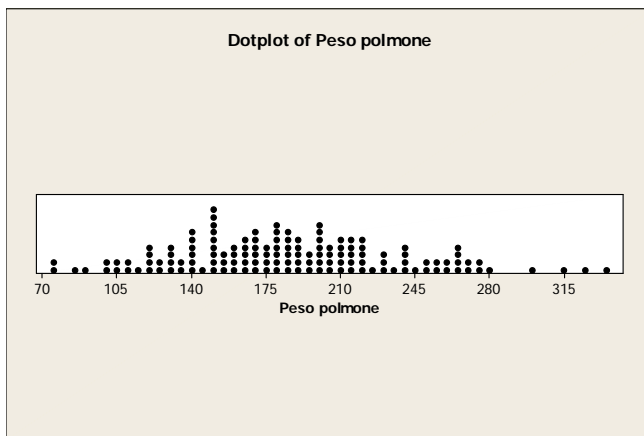
Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Peso polmone	186,31	52,08	76,30	149,90	183,00	217,35	335,20
Eta	41,08	31,69	3,00	12,00	39,00	65,00	104,00

Stat→Basic Statistics→Display Descriptive Statistics.

Scegliere *Peso polmone* e *Eta* come *Variables*. Nel menu *Statistics* selezionare gli indici *Mean*, *Standard Deviation*, *Minimum*, *Maximum*, *First quartile*, *Median*, *Third Quartile*.

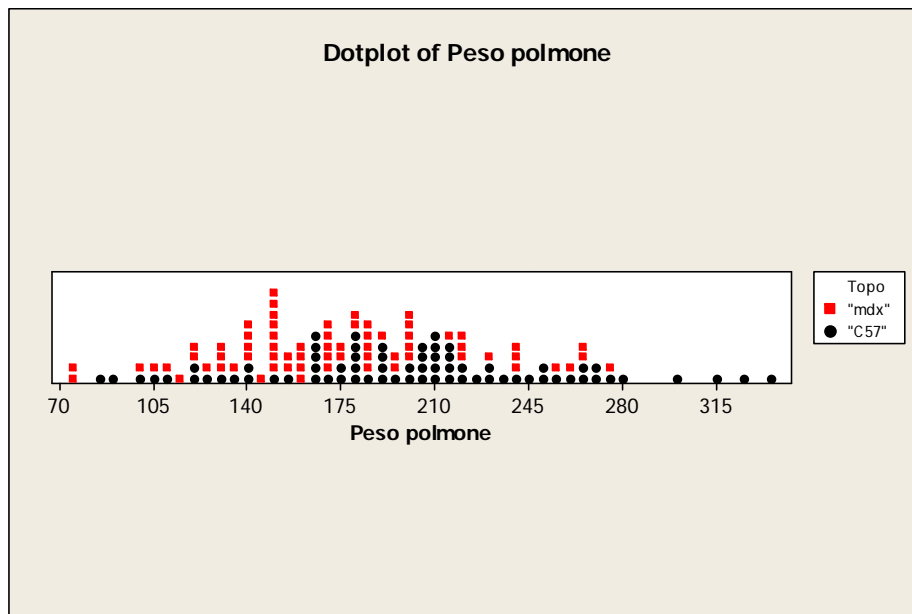
Vediamo anche qualche rappresentazione grafica delle variabili considerate singolarmente.

Dotplot (o diagramma di dispersione)

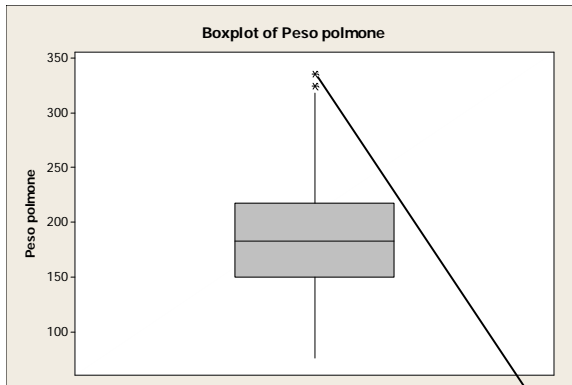


Graph→Dotplots→One way - Simple
Selezionare *Peso polmone* in *Graph variables*.

Se si vuole evidenziare gli individui sani da quelli malati, bisogna scegliere nel menu principale *One way - Stack groups* e selezionare "*Topi*" in *Categorical variables for grouping*.



Box-plot di una singola variabile



Graph→Boxplot→Simple
 Selezionare *Peso polmone* in *Graph variables*.

Commenti: ci sono due osservazioni anomale. "Cliccando" sopra agli asterischi si ottiene il valore del peso del polmone e il numero dell'osservazione corrispondente.

Outlier symbol. Row 83. Peso Polmone= 335.2

Commento: il peso del polmone ha un intervallo di variazione [massimo - minimo] di 258.9 grammi e una distanza interquartile [Q3-Q1] di 67.45; l'intervallo di variazione è quindi circa 4 volte la distanza interquartile. Come di nota anche dai grafici, le "code" della distribuzione sono grandi. Cerchiamo di capire se ciò è dovuto a qualche altra variabile da cui può dipendere il peso: potrebbero essere l'età o la patologia considerata.

Procediamo vedendo il peso medio del polmone rispetto all'età.

Indici statistici di una variabile suddivisi per i livelli di un'altra variabile

Variable	Eta	Count	Mean	StDev
Peso polmone	3	19	115,53	29,01
	6	17	172,6	35,2
	12	15	170,97	34,34
	26	31	207,0	51,1
	39	15	201,1	52,8
	52	19	228,5	43,6
	65	15	205,5	40,1
	78	13	177,58	29,97
	104	13	197,3	42,0

Stat→Basic Statistics→Display
 Descriptive Statistics.
 Scegliere *Peso polmone* come *Variables* e
 Età come "*by variables*".
 Nel menu *Statistics* selezionare gli indici
Mean, Standard Deviation, N Total.

Commento: i topi di 3 settimane hanno mediamente il peso del polmone molto più basso degli altri. Questo è dovuto al fatto che i polmoni in quell'età non si sono ancora sviluppati completamente. Potrebbe essere utile procedere all'analisi statistica eliminando anche le osservazioni relative ai topi di 3 settimane.

Si può anche classificare l'età in classi utilizzando i quartili [min: 3; Q1: 12; mediana: 39; Q3: 65; max: 104] e mantendeno in una classe separata i topi di 3 settimane.

Ricodifica dei dati

Data→Code

Selezionare "Numeric to text". Scegliere la colonna "Età" da ricodificare in "Code data from columns"
Scegliere una colonna in cui mettere i dati codificati nel seguente modo (Original values)

3 New: 3

4:12 New: 4-12

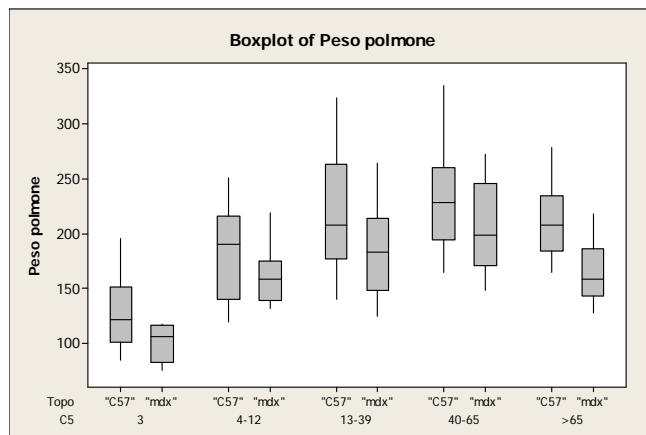
13:39 New: 13-39

40:65 New 40-65

65:200 New >65

Siamo interessati a capire se la patologia influenza il peso dei polmoni; quindi è interessante tracciare i boxplot, divisi per classe di età e per presenza/assenza di patologia.

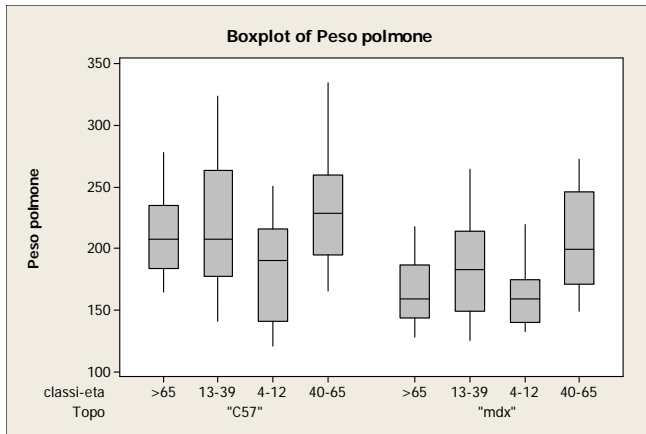
Box-plot di una variabile suddivisa secondo i livelli di un'altra variabile (o di altre variabili)



Graph→Boxplots→With Groups

Selezionare "Peso polmone" in Graph variables e "Classi-età, Topò" in Categorical variables for grouping.

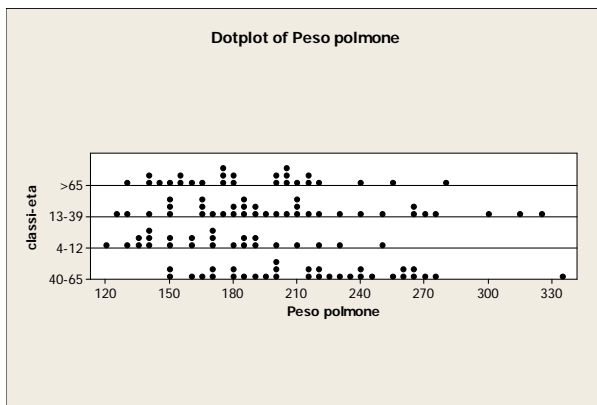
Commento: in ogni classe di età si nota che i topi affetti dalla malattia hanno valori di ogni quartile più bassi rispetto a quelli dei topi sani; inoltre, soprattutto per le classi di età più basse la dispersione del peso è minore per i topi malati rispetto ai topi sani.



Graph→Boxplots→With Groups
 Selezionare "Peso polmone" in Graph variables e "Topo, Classi-età" in Categorical variables for grouping.

Confrontiamoli ancora con il dotplot del peso del polmone suddiviso in classi di età.

Dotplot di una variabile suddivisa secondo i livelli di un'altra variabile



Graph→Dotplot→One Y, with groups
 Selezionare "peso polmone" in Graph variables e "Classi-età" in Categorical variables for grouping

Dal confronto sembra che l'età influenzi il peso del polmoni meno di quanto lo faccia la presenza/assenza della patologia (notiamo che abbiamo escluso la classe 3 settimane). Per maggiore conferma proviamo a calcolare le tabelle a due entrate classi-peso/classi-età e classi-peso/topo. Per far questo dobbiamo suddividere i dati dei pesi del polmone in classi. Siccome il peso massimo é 335.20 e il minimo é 120.20, determiniamo 5 classi di peso

- 1: 120,20:163,20
- 2: 163,20:206,20
- 3: 206,20:249,20
- 4: 249,20:292,20
- 5: 249,20:292,20

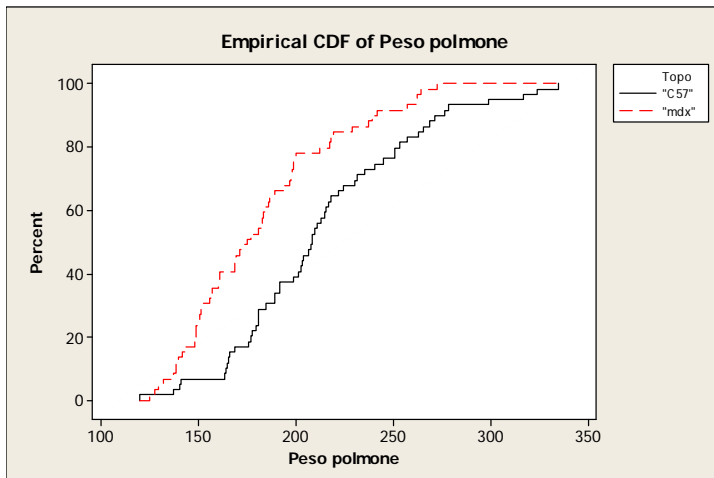
Dopo aver diviso i dati dei pesi in classi, confrontiamo le seguenti tabelle profilo, ottenendo una conferma delle nostre osservazioni.

Rows: classi-peso			Columns: Topo			Rows: classi-peso			Columns: classi-eta		
	"C57"	"mdx"	All		>65	13-39	3	4-12	40-65	All	
1	57,89	42,11	100,00	1	0,0	0,0	100,0	0,0	0,0	100,0	
2	41,67	58,33	100,00	2	0,0	0,0	0,0	100,0	0,0	100,0	
3	63,89	36,11	100,00	3	0,0	100,0	0,0	0,0	0,0	100,0	
4	43,75	56,25	100,00	4	0,0	0,0	0,0	0,0	100,0	100,0	
5	46,15	53,85	100,00	5	100,0	0,0	0,0	0,0	0,0	100,0	
All	51,09	48,91	100,00	All	19,0	26,3	13,9	17,5	23,4	100,0	

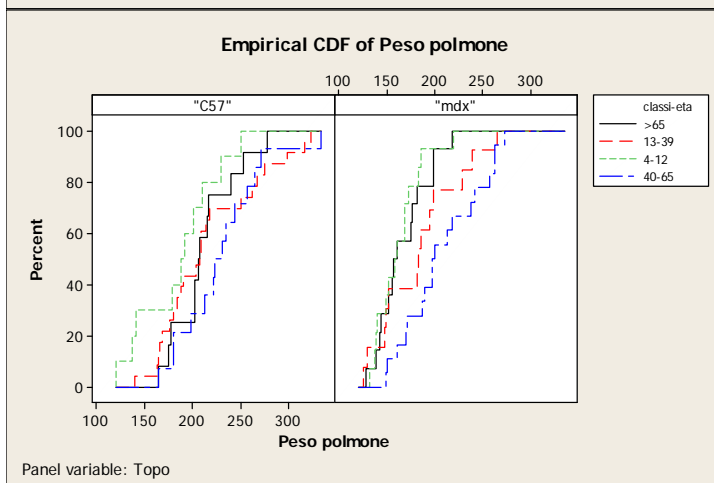
Cell Contents: % of Row

Funzione di distribuzione cumulata e percentili

Per concludere confrontiamo le funzioni di distribuzioni cumulate del peso dei polmoni dei topi sani e dei topi affetti da patologia.



Graph→Empirical CDF-Multiple
 Selezionare "Peso polmone" (esclusi 3 settimane) in *Graph variables* e *Topo* in *Categorical variables for grouping*
 Nel menu *Distribution-Data Display*, scegliere l'opzione *Connect line only*.



Per suddividere ulteriormente in classi d'età, scegliere il menu *Multiple graphs*.

Anche con questa rappresentazione si vede che il peso del polmone dei topi affetti da patologia è mediamente più basso del peso dei polmoni dei topi sani.