

## LABORATORIO – EXCEL XLSTAT 2008 SCHEDE 2 e 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 dell'uomo. 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.

I dati sono raccolti nel dataset: TOPI (in cui si deve prima sostituire al punto decimale la virgola (con Modifica Seleziona tutto Sostituisci) e cancellare gli \*

Numero dei dati raccolti: 157

Nomi delle variabili:

1. Peso Polmone: il peso del polmone del topo espresso in grammi. (la cella vuota 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)

Procediamo con un'analisi descrittiva delle singole variabili. Notiamo che la variabile Peso Polmone è una variabile quantitativa, mentre la variabile Topo è qualitativa e anche la variabile Età si può considerare come qualitativa.

### Tabella di contingenza a una via

Con tabella pivot, come visto nella scheda 1

Conteggio di

Topo	Topo		Totale complessivo
	C57	mdx	
Totale	50,96%	49,04%	100,00%

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

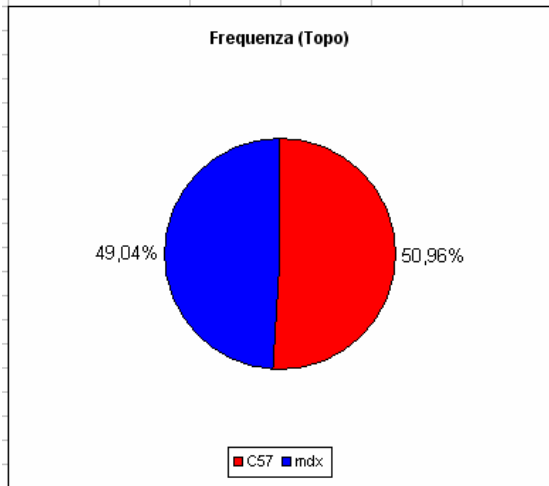
Oppure

Statistiche descrittive (Dati qualitativi):		
Statistica	Topo	
Modalità	C57	mdx
Frequenza	80,000	77,000
Frequenza	50,965	49,045

XL-Stat

- Descrizione dei dati → statistiche descrittive
- Spuntare Dati qualitativi
- Selezionare la Colonna Topo (compresa l'intestazione)
- Scegliere in Risultati le informazioni utili (es. Frequenza per modalità e Frequenza relativa per modalità)

Grafici a torta:



Si può optare per la scelta del grafico (es. grafici a barre o a torta)

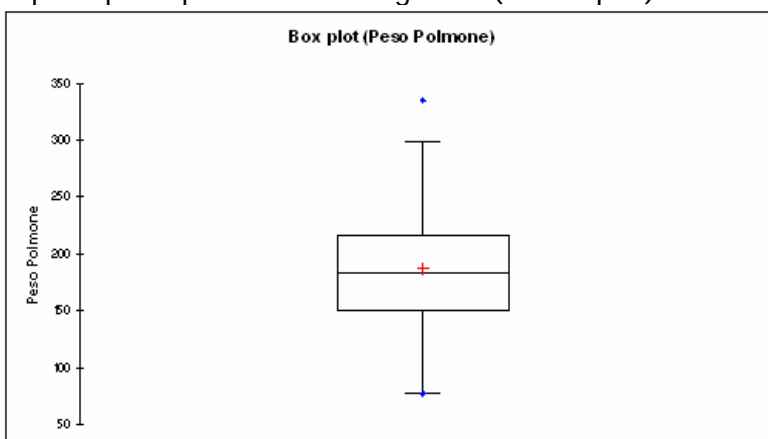
### Principali indici statistici per variabili quantitative

Statistiche descrittive (Dati quantitativi):	
Statistica	Peso Polmone
No. di osservazioni	157
No. di valori mancanti	20
Minimo	76,300
Massimo	335,200
1° Quartile	150,700
Mediana	183,000
3° Quartile	217,200
Media	186,312
Deviazione standard (n)	51,889

XL-Stat

- Descrizione dei dati → statistiche descrittive
- Spuntare Dati quantitativi
- Selezionare la Colonna Peso polmone (compresa l'intestazione)
- Scegliere in Risultati le informazioni utili (es. No. di osservazioni No. di valori mancanti Minimo Massimo 1° Quartile Mediana 3° Quartile Media Deviazione standard (n))

Si può optare per la scelta del grafico (es. boxplot)



**Commento:** ci sono due osservazioni anomale. Avvicinandosi con il mouse agli asterischi, si ottiene il valore del peso del polmone corrispondente.

A questo punto il numero di unità sperimentali è sceso a 137 unità, perché il software elimina i dati mancanti.

## Rappresentazioni grafiche delle variabili considerate singolarmente.

### Istogramma

Selezionare i dati grezzi – Dati Tabella pivot- In Riga mettere Peso polmone, in Colonna Topo, in Dati conteggio peso polmone – Selezionare la tabella – Copia- Modifica incolla speciale  
Dalla nuova tabella estrarre dati per costruire un istogramma.

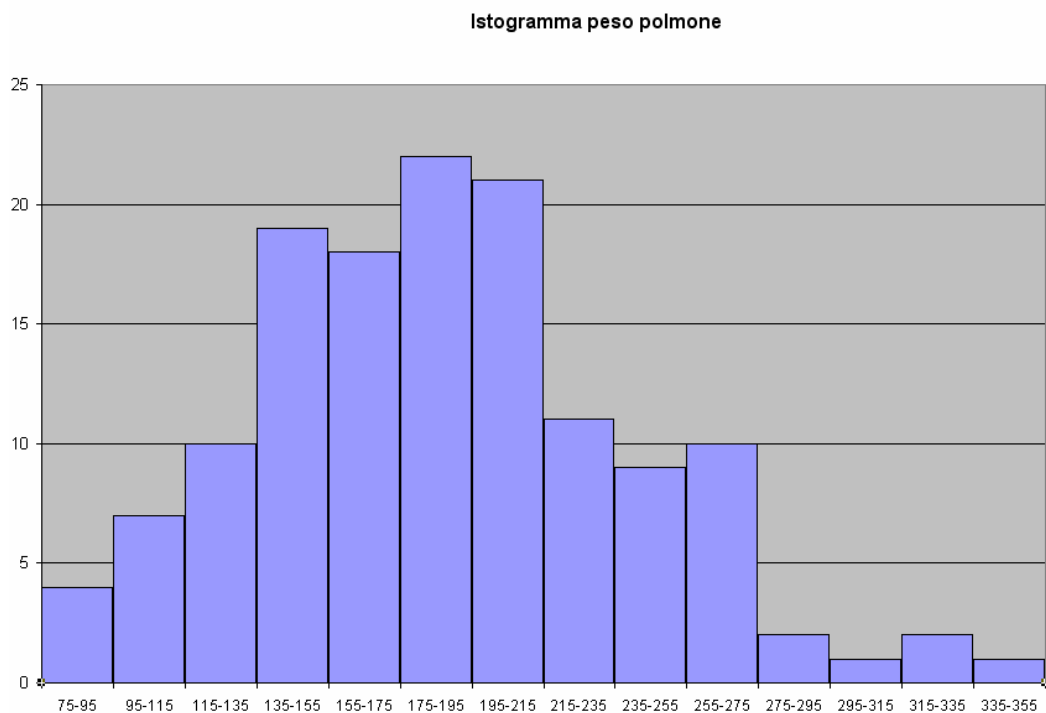
È possibile creare delle classi di uguale ampiezza, ad esempio:

75-95 95-115 115-135 135-155 155-175 175-195 195-215 215-235 235-255 255-275 275-295 295-315 315-335 335-355

Se la tabella è nella zona F8:H38, mediante le istruzioni =SE(F8<=95;SOMMA(G8:H8);" ")  
=SE(E(F8>95;F8<=115);SOMMA(G8:H8);" ") e la funzione Somma, è possibile creare una nuova tabella :

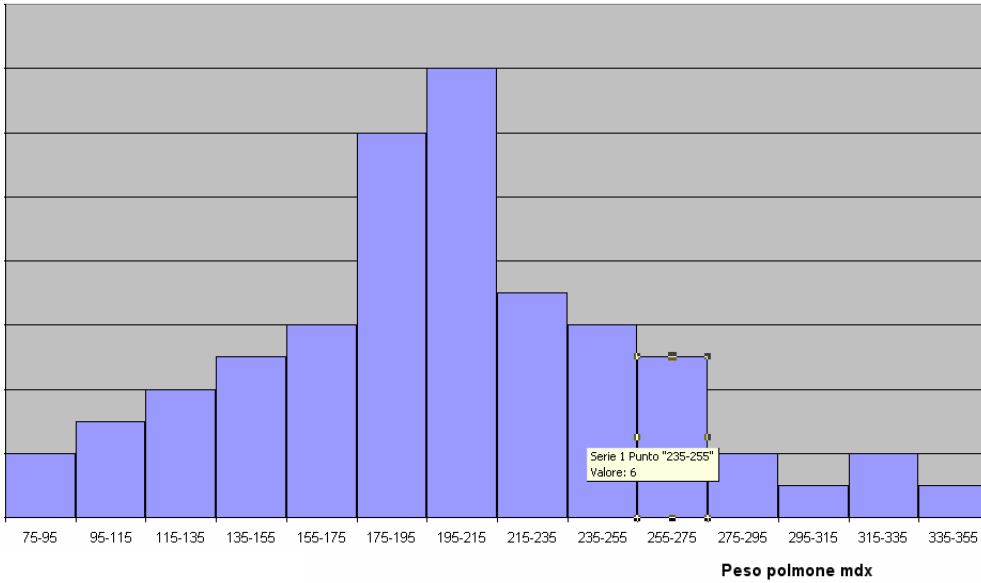
4	7	10	19	18	22	21	11	9	10	2	1	2	1
75-95	95-115	115-135	135-155	155-175	175-195	195-215	215-235	235-255	255-275	275-295	295-315	315-335	335-355

riportata nell'istogramma, in cui i rettangoli sono affiancati mediante l'opzione Distanza tra le barre = 0 in Formato dato (dopo aver selezionato uno dei rettangoli):

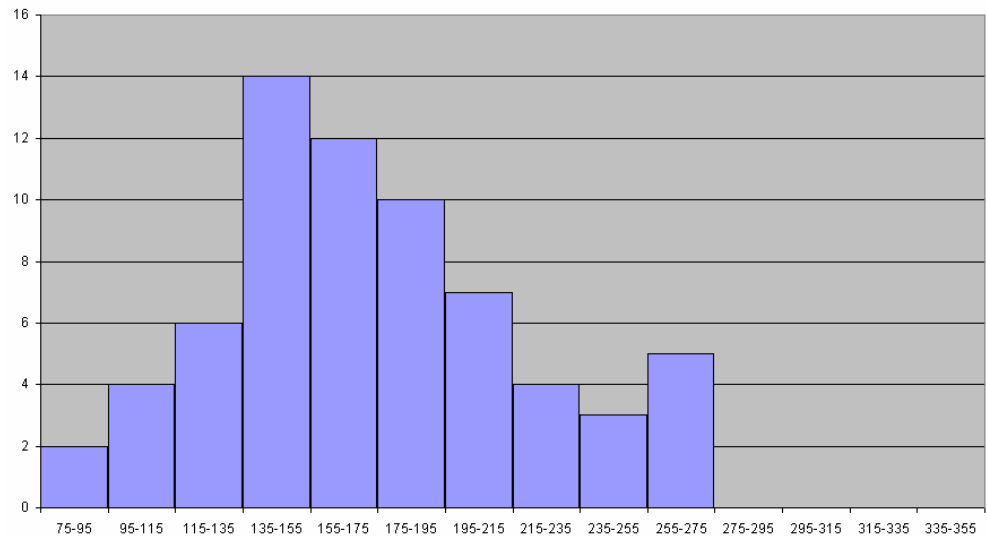


Per riportare sullo stesso piano cartesiano gli istogrammi del peso nel caso C57 e nel caso mdx, si può spuntare (dal menu a tendina) prima la tipologia Topo C57, poi Topo mdx e procedere in modo analogo.

Peso polmone C57



Peso polmone mdx



Riportiamo anche una parte dei fogli di lavoro:

- per l'istogramma peso polmone

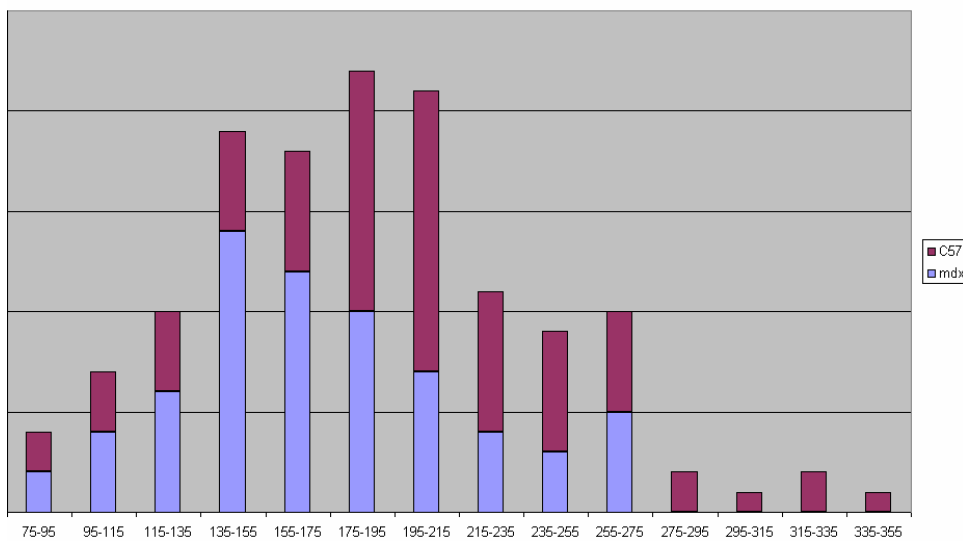
K	L	M
Totali	=SOMMA(L8:L138)	=SOMMA(M8:M138)
Classi C57+mdx	75-95	95-115
	=SE(E(F8>75;F8<=95);SOMMA(G8:H8);" ")	=SE(E(F8>95;F8<=115);SOMMA(G8:H8);" ")
	=SE(E(F9>75;F9<=95);SOMMA(G9:H9);" ")	=SE(E(F9>95;F9<=115);SOMMA(G9:H9);" ")
	=SE(E(F10>75;F10<=95);SOMMA(G10:H10);" ")	=SE(E(F10>95;F10<=115);SOMMA(G10:H10);" ")
	=SE(E(F11>75;F11<=95);SOMMA(G11:H11);" ")	=SE(E(F11>95;F11<=115);SOMMA(G11:H11);" ")
	=SE(E(F12>75;F12<=95);SOMMA(G12:H12);" ")	=SE(E(F12>95;F12<=115);SOMMA(G12:H12);" ")
	=SE(E(F13>75;F13<=95);SOMMA(G13:H13);" ")	=SE(E(F13>95;F13<=115);SOMMA(G13:H13);" ")
	=SE(E(F14>75;F14<=95);SOMMA(G14:H14);" ")	=SE(E(F14>95;F14<=115);SOMMA(G14:H14);" ")
	=SE(E(F15>75;F15<=95);SOMMA(G15:H15);" ")	=SE(E(F15>95;F15<=115);SOMMA(G15:H15);" ")
	=SE(E(F16>75;F16<=95);SOMMA(G16:H16);" ")	=SE(E(F16>95;F16<=115);SOMMA(G16:H16);" ")
	=SE(E(F17>75;F17<=95);SOMMA(G17:H17);" ")	=SE(E(F17>95;F17<=115);SOMMA(G17:H17);" ")
	=SE(E(F18>75;F18<=95);SOMMA(G18:H18);" ")	=SE(E(F18>95;F18<=115);SOMMA(G18:H18);" ")
	=SE(E(F19>75;F19<=95);SOMMA(G19:H19);" ")	=SE(E(F19>95;F19<=115);SOMMA(G19:H19);" ")
	=SE(E(F20>75;F20<=95);SOMMA(G20:H20);" ")	=SE(E(F20>95;F20<=115);SOMMA(G20:H20);" ")
	=SE(E(F21>75;F21<=95);SOMMA(G21:H21);" ")	=SE(E(F21>95;F21<=115);SOMMA(G21:H21);" ")
	=SE(E(F22>75;F22<=95);SOMMA(G22:H22);" ")	=SE(E(F22>95;F22<=115);SOMMA(G22:H22);" ")
	=SE(E(F23>75;F23<=95);SOMMA(G23:H23);" ")	=SE(E(F23>95;F23<=115);SOMMA(G23:H23);" ")
	=SE(E(F24>75;F24<=95);SOMMA(G24:H24);" ")	=SE(E(F24>95;F24<=115);SOMMA(G24:H24);" ")
	=SE(E(F25>75;F25<=95);SOMMA(G25:H25);" ")	=SE(E(F25>95;F25<=115);SOMMA(G25:H25);" ")

- per l'istogramma peso polmone C57:

	=SOMMA(F7:F139)	=SOMMA(G7:G139)
C57	75-95	95-115
	=SE(E(A8>75;A8<=95);B8;" ")	=SE(E(A8>95;A8<=115);B8;" ")
	=SE(E(A9>75;A9<=95);B9;" ")	=SE(E(A9>95;A9<=115);B9;" ")
	=SE(E(A10>75;A10<=95);B10;" ")	=SE(E(A10>95;A10<=115);B10;" ")
	=SE(E(A11>75;A11<=95);B11;" ")	=SE(E(A11>95;A11<=115);B11;" ")
	=SE(E(A12>75;A12<=95);B12;" ")	=SE(E(A12>95;A12<=115);B12;" ")
	=SE(E(A13>75;A13<=95);B13;" ")	=SE(E(A13>95;A13<=115);B13;" ")
	=SE(E(A14>75;A14<=95);B14;" ")	=SE(E(A14>95;A14<=115);B14;" ")
	=SE(E(A15>75;A15<=95);B15;" ")	=SE(E(A15>95;A15<=115);B15;" ")
	=SE(E(A16>75;A16<=95);B16;" ")	=SE(E(A16>95;A16<=115);B16;" ")
	=SE(E(A17>75;A17<=95);B17;" ")	=SE(E(A17>95;A17<=115);B17;" ")
	=SE(E(A18>75;A18<=95);B18;" ")	=SE(E(A18>95;A18<=115);B18;" ")
	=SE(E(A19>75;A19<=95);B19;" ")	=SE(E(A19>95;A19<=115);B19;" ")
	=SE(E(A20>75;A20<=95);B20;" ")	=SE(E(A20>95;A20<=115);B20;" ")
	=SE(E(A21>75;A21<=95);B21;" ")	=SE(E(A21>95;A21<=115);B21;" ")
	=SE(E(A22>75;A22<=95);B22;" ")	=SE(E(A22>95;A22<=115);B22;" ")
	=SE(E(A23>75;A23<=95);B23;" ")	=SE(E(A23>95;A23<=115);B23;" ")
	=SE(E(A24>75;A24<=95);B24;" ")	=SE(E(A24>95;A24<=115);B24;" ")
	=SE(E(A25>75;A25<=95);B25;" ")	=SE(E(A25>95;A25<=115);B25;" ")
	=SE(E(A26>75;A26<=95);B26;" ")	=SE(E(A26>95;A26<=115);B26;" ")

Se si vogliono evidenziare gli individui sani distinguendo da quelli malati, cliccando su uno dei due grafici, scegliere Dati di origine, Aggiungi Serie. Si ottiene il grafico in pila:

Peso polmone mdx e C57

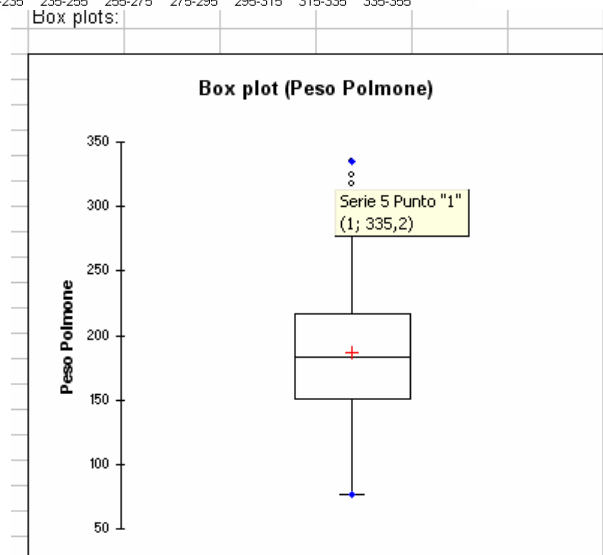


### Box-plot di una singola variabile

Rivediamo il boxplot del peso polmone.

**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 si nota anche dagli istogrammi, 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.

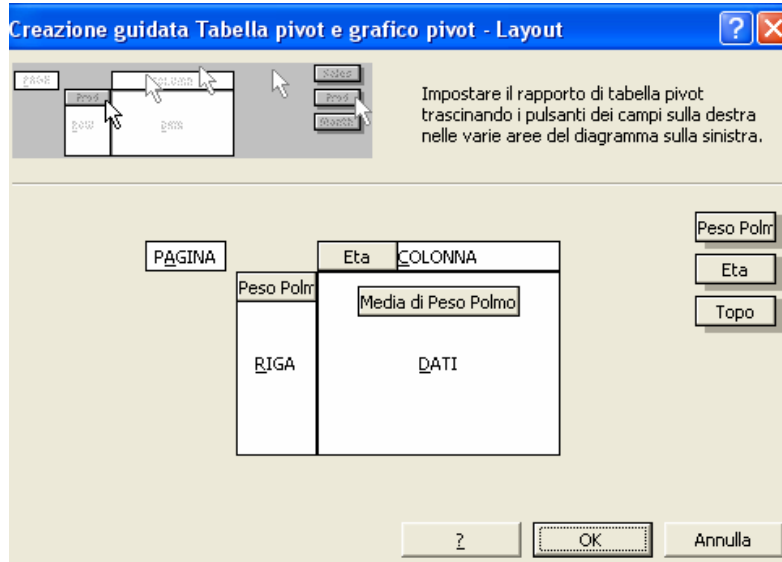


## Indici statistici di una variabile suddivisi per i livelli di un'altra variabile

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

### Primo metodo

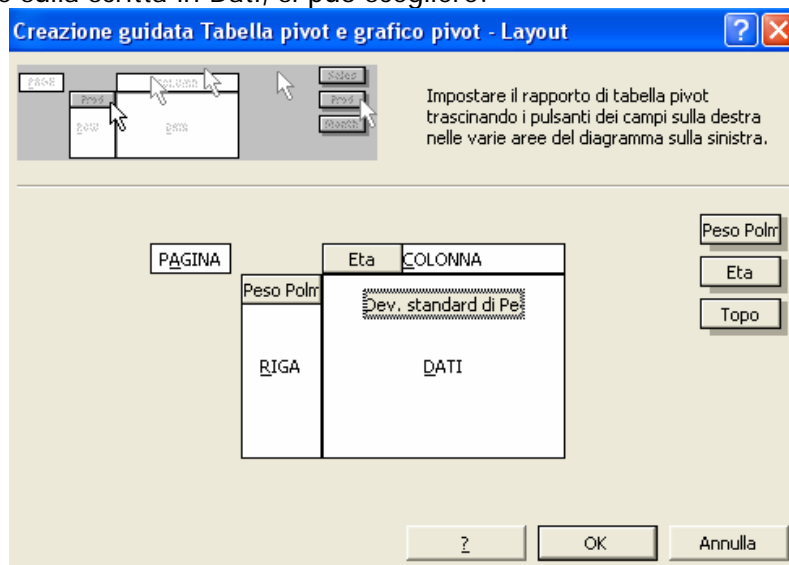
Creare la seguente tabella pivot con "campo pivot table"=media, attivabile cliccando due volte sulla scritta nella zona Dati:



(ricopiata al solito con Incolla speciale Valori) è possibile avere nell'ultima riga le medie del peso per età

Eta	3	6	12	26	39	52	65	78	104
	115,53	172,65	170,97	207,00	201,11	228,47	205,54	177,58	197,27

Cliccando due volte sulla scritta in Dati, si può scegliere:



e si ottiene, nell'ultima riga, la serie delle deviazioni standard:

29,01	35,19	34,34	51,12	52,79	43,62	40,12	29,97	42,05
-------	-------	-------	-------	-------	-------	-------	-------	-------

**Secondo metodo, proseguendo con XL- Stat**

**Ricordiamo che i dati mancanti sono stati esclusi**

Tabella del tipo:

Tabella di contingenza (R1 / C1):											
	104	12	26	3	39	52	6	65	78	Eta	
102	0	0	0	2	0	0	0	0	0	0	0
102,6	0	0	0	1	0	0	0	0	0	0	0
105,8	0	0	0	1	0	0	0	0	0	0	0
111,8	0	0	0	1	0	0	0	0	0	0	0
112	0	0	0	1	0	0	0	0	0	0	0
115	0	0	0	1	0	0	0	0	0	0	0
117,5	0	0	0	1	0	0	0	0	0	0	0
118	0	0	0	1	0	0	0	0	0	0	0
120,2	0	1	0	0	0	0	0	0	0	0	0
121,8	0	0	0	1	0	0	0	0	0	0	0

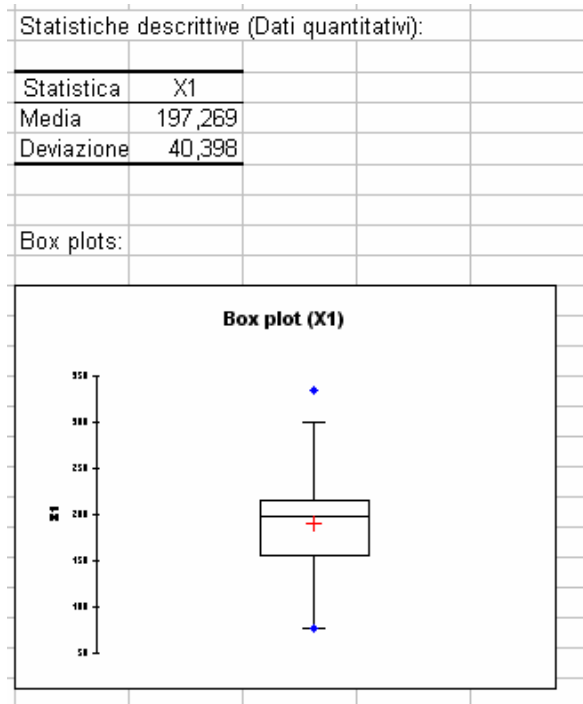
...

Frequenze osservate:											
	104	12	26	3	39	52	6	65	78	Eta	Totale

...

Totale	13	14	22	19	14	17	10	15	13	0	137
--------	----	----	----	----	----	----	----	----	----	---	-----

Si possono avere i profili riga e i profili colonna nel modo seguente:



Si ottiene:

peso	104	12	26	3	39	52	6	65	78
Media	197,269	170,971	206,995	115,526	201,107	228,471	172,650	205,540	177,585
Deviazione	40,398	33,090	49,946	28,237	50,871	42,322	33,387	38,758	28,791

XL Stat

- Preparazione dei dati → Creare una tabella di contingenza
- In Generale selezionare la colonna Peso per Variabile righe e la colonna Età per la Variabile colonne.
- In Dati mancanti selezionare Eliminare le osservazioni
- In Risultati spuntare Tabella di contingenza
- In Grafici nessuna selezione

XL Stat

- Descrizione dei dati → Tabella di contingenza
- In Generale selezionare la tabella di contingenza appena creata e spuntare Foglio
- In Risultati spuntare Frequenze osservate

XL Stat

- Descrizione dei dati → Tabella di contingenza
- In Generale selezionare la tabella di contingenza appena creata e spuntare Foglio
- In Risultati spuntare Proporzioni/Riga Proporzioni/Colonna

Per calcolare le medie di sottogruppo del peso rispetto all'età, ossia per livelli di età, si può completare la tabella appena descritta.

XL Stat

- Descrizione dei dati → Statistiche descrittive
- In Generale selezionare la colonna dei Pesì come Dati quantitativi e spuntare Pesì, selezionando la colonna relativa ad una età (Attenzione: selezionare anche la cella di intestazione)
- In Risultati spuntare Media e Deviazione standard

Fatto una volta, è possibile richiamare lo strumento e sostituire il nome delle celle peso

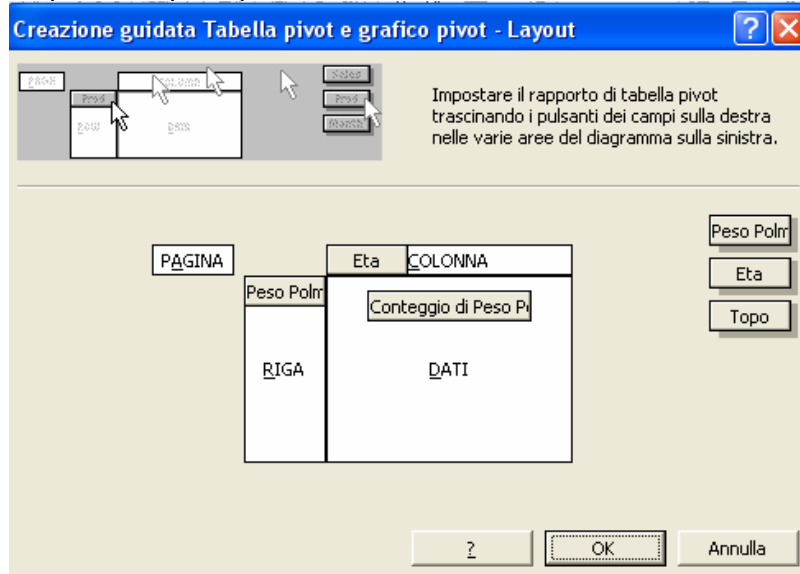
**Commento:** i topi di 3 settimane hanno mediamente il peso del polmone molto più basso degli altri. Questo è dovuto al fatto che i polmoni in quell'età non si sono ancora sviluppati completamente. Potrebbe essere utile procedere all'analisi statistica eliminando anche le osservazioni relative ai topi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

### Diagramma di una variabile suddivisa secondo i livelli di un'altra variabile

Si può procedere così:

- fare la tabella pivot del peso per età



- decidere le classi di età ; ad esempio:

**3      4-12      13-39      40-65      >65**

- copiare la tabella pivot con Incolla speciale Valori e da questa costruire la seguente tabella di contingenza, mantenendo le classi di peso già determinate in precedenza:

	<b>3</b>	<b>4-12</b>	<b>13-39</b>	<b>40-65</b>	<b>&gt;65</b>	
<b>75-95</b>	4	0	0	0	0	4
<b>95-115</b>	7	0	0	0	0	7
<b>115-135</b>	5	2	2	0	1	10
<b>135-155</b>	2	7	4	2	4	19
<b>155-175</b>	0	5	4	4	5	18
<b>175-195</b>	0	5	9	4	4	22
<b>195-215</b>	1	2	6	6	6	21
<b>215-235</b>	0	2	2	4	3	11
<b>235-255</b>	0	1	2	4	2	9
<b>255-275</b>	0	0	3	7	0	10
<b>275-295</b>	0	0	1	0	1	2
<b>295-315</b>	0	0	1	0	0	1
<b>315-335</b>	0	0	2	0	0	2
<b>335-355</b>	0	0	0	1	0	1
	19	24	36	32	26	137

La tabella si può ottenere raggruppando i dati prima per colonne. Se la tabella originaria occupasse la zona da T7 ad AC137, le formule sarebbero

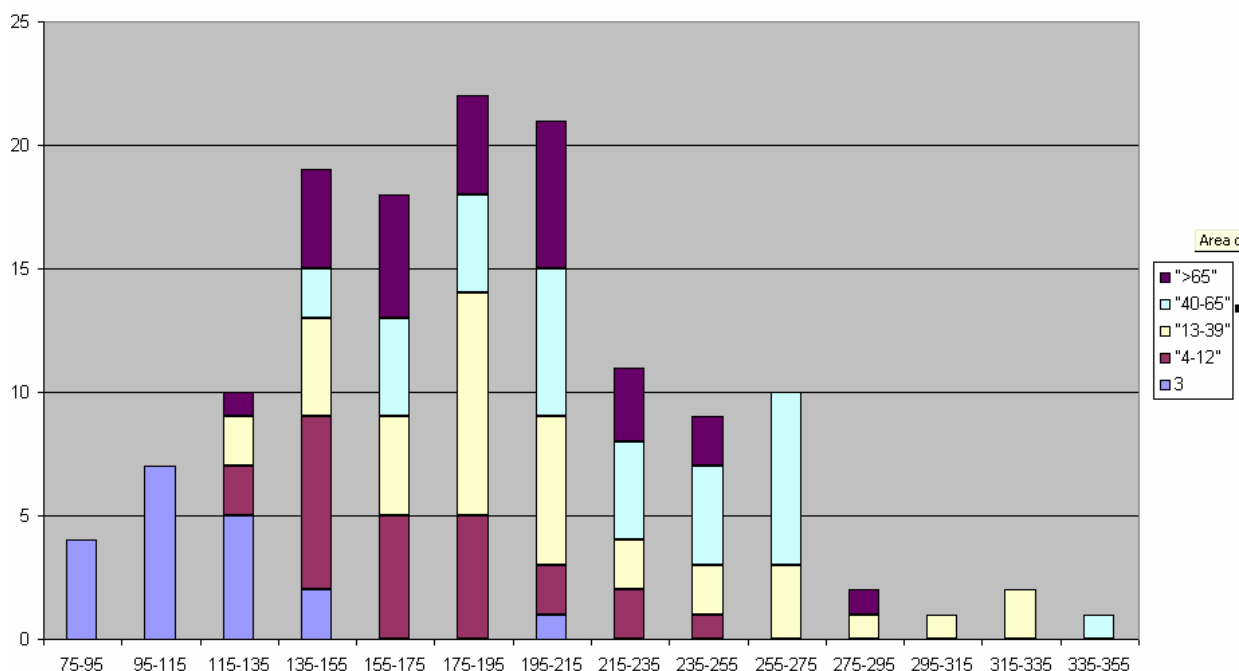


	3	4-12	13-39	40-65	>65	totali
75-95	=SOMMA(U7:U10)	=SOMMA(V7:W10)	=SOMMA(X7:Y10)	=SOMMA(Z7:AA10)	=SOMMA(AB7:AC10)	=SOMMA(M4:Q4)
95-115	=SOMMA(U11:U16)	=SOMMA(V11:W16)	=SOMMA(X11:Y16)	=SOMMA(Z11:AA16)	=SOMMA(AB11:AC16)	=SOMMA(M5:Q5)
115-135	=SOMMA(U17:U26)	=SOMMA(V17:W26)	=SOMMA(X17:Y26)	=SOMMA(Z17:AA26)	=SOMMA(AB17:AC26)	=SOMMA(M6:Q6)
135-155	=SOMMA(U27:U43)	=SOMMA(V27:W43)	=SOMMA(X27:Y43)	=SOMMA(Z27:AA43)	=SOMMA(AB27:AC43)	=SOMMA(M7:Q7)
155-175	=SOMMA(U44:U61)	=SOMMA(V44:W61)	=SOMMA(X44:Y61)	=SOMMA(Z44:AA61)	=SOMMA(AB44:AC61)	=SOMMA(M8:Q8)
175-195	=SOMMA(U62:U81)	=SOMMA(V62:W81)	=SOMMA(X62:Y81)	=SOMMA(Z62:AA81)	=SOMMA(AB62:AC81)	=SOMMA(M9:Q9)
195-215	=SOMMA(U82:U102)	=SOMMA(V82:W102)	=SOMMA(X82:Y102)	=SOMMA(Z82:AA102)	=SOMMA(AB82:AC102)	=SOMMA(M10:Q10)
215-235	=SOMMA(U103:U112)	=SOMMA(V103:W112)	=SOMMA(X103:Y112)	=SOMMA(Z103:AA112)	=SOMMA(AB103:AC112)	=SOMMA(M11:Q11)
235-255	=SOMMA(U113:U121)	=SOMMA(V113:W121)	=SOMMA(X113:Y121)	=SOMMA(Z113:AA121)	=SOMMA(AB113:AC121)	=SOMMA(M12:Q12)
255-275	=SOMMA(U122:U131)	=SOMMA(V122:W131)	=SOMMA(X122:Y131)	=SOMMA(Z122:AA131)	=SOMMA(AB122:AC131)	=SOMMA(M13:Q13)
275-295	=SOMMA(U132:U133)	=SOMMA(V132:W133)	=SOMMA(X132:Y133)	=SOMMA(Z132:AA133)	=SOMMA(AB132:AC133)	=SOMMA(M14:Q14)
295-315	=SOMMA(U134)	=SOMMA(V134:W134)	=SOMMA(X134:Y134)	=SOMMA(Z134:AA134)	=SOMMA(AB134:AC134)	=SOMMA(M15:Q15)
315-335	=SOMMA(U135:U136)	=SOMMA(V135:W136)	=SOMMA(X135:Y136)	=SOMMA(Z135:AA136)	=SOMMA(AB135:AC136)	=SOMMA(M16:Q16)
335-355	=SOMMA(U137)	=SOMMA(V137:W137)	=SOMMA(X137:Y137)	=SOMMA(Z137:AA137)	=SOMMA(AB137:AC137)	=SOMMA(M17:Q17)
totali	=SOMMA(M4:M17)	=SOMMA(N4:N17)	=SOMMA(O4:O17)	=SOMMA(P4:P17)	=SOMMA(Q4:Q17)	=SOMMA(M18:Q18)

Attenzione: avvalersi della copiatura di celle

- rappresentare con un diagramma a pila

Peso per classi di età



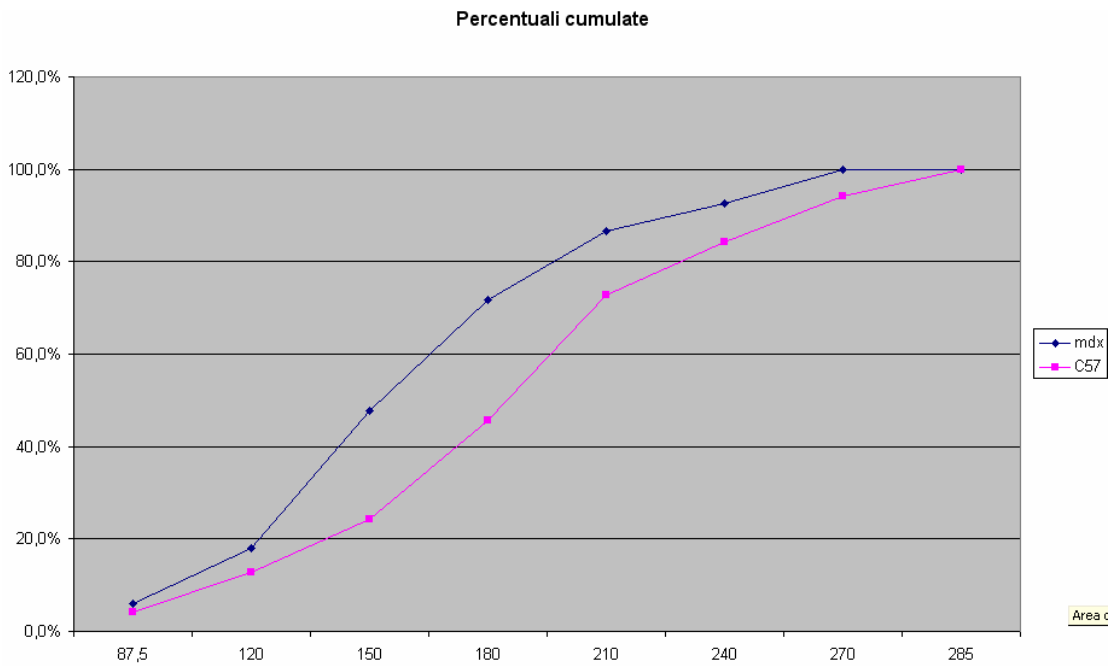
Dal confronto con il grafico di pagina 5 sembra che l'età influenzi il peso del polmoni meno di quanto lo faccia la presenza/assenza della patologia (se si esclude la classe 3 settimane). Per conferma calcoliamo le tabelle a due entrate classi-peso/classi-età e classi-peso/topo.

Mantenendo le classi precedentemente utilizzate, confrontiamo le seguenti tabelle profilo, ottenendo una conferma delle nostre osservazioni.

	3	4-12	13-39	40-65	>65	totali	C57	mdx		
75-95	100%	0%	0%	0%	0%	100%	75-95	50,00%	50,00%	100,00%
95-115	100%	0%	0%	0%	0%	100%	95-115	42,86%	57,14%	100,00%
115-135	50%	20%	20%	0%	10%	100%	115-135	40,00%	60,00%	100,00%
135-155	11%	37%	21%	11%	21%	100%	135-155	26,32%	73,68%	100,00%
155-175	0%	28%	22%	22%	28%	100%	155-175	33,33%	66,67%	100,00%
175-195	0%	23%	41%	18%	18%	100%	175-195	54,55%	45,45%	100,00%
195-215	5%	10%	29%	29%	29%	100%	195-215	66,67%	33,33%	100,00%
215-235	0%	18%	18%	36%	27%	100%	215-235	63,64%	36,36%	100,00%
235-255	0%	11%	22%	44%	22%	100%	235-255	66,67%	33,33%	100,00%
255-275	0%	0%	30%	70%	0%	100%	255-275	50,00%	50,00%	100,00%
275-295	0%	0%	50%	0%	50%	100%	275-295	100,00%	0,00%	100,00%
295-315	0%	0%	100%	0%	0%	100%	295-315	100,00%	0,00%	100,00%
315-335	0%	0%	100%	0%	0%	100%	315-335	100,00%	0,00%	100,00%
335-355	0%	0%	0%	100%	0%	100%	335-355	100,00%	0,00%	100,00%
totali	13,9%	17,5%	26,3%	23,4%	19,0%	100,0%	totali	51,09%	48,91%	100,00%

## Funzione di distribuzione cumulata e percentili

Per realizzare un confronto più evidente possiamo tracciare i grafici delle funzioni di distribuzioni cumulate del peso dei polmoni dei topi sani e dei topi affetti da patologia.



Questa rappresentazione è decisiva: il peso del polmone dei topi affetti da patologia é mediamente piú basso del peso dei polmoni dei topi sani.