

LABORATORIO-EXCEL N. 2-3
XLSTAT- Pro Versione 7
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. (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)

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 *
Poi copiare in un foglio di Excel dalla cella A1.

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

Con tabella pivot, come visto nella scheda 1

Conteggio di Topo	Topo		Totale complessivo	Commenti: il 51% dei topi analizzati è normale e il 49% è affetto da distrofia muscolare.
	C57	mdx		
Totale	50,96%	49,04%	100,00%	

Nota: Si può anche selezionare la tabella pivot; da Dati Raggruppa e struttura mostra dettaglio si ottiene la suddivisione in C57 e mdx

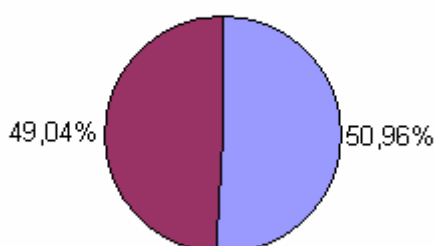
Oppure con XLSTAT

XLSTAT 7.0 - Statistiques descriptives - le 05/11/2008 à 17.26.58								
Description de données qualitatives								
Données : classeur = Cartel1 / feuille = Foglio1 / plage = \$C\$2:\$C\$158 / 157 lignes et 1 colonne								
Pondération uniforme (par défaut)								
Aucune donnée manquante détectée								
Synthèse :								
Variable	données u	données i	opre de mod	Mode	quence mc	% mode	éq. rel. mode	
Topo	157	0	2	C57	80	50,96	0,510	
Topo / 2 Modalités :								
Modalité	Effectif	%	Fréq. rel.					
C57	80	50,96	0,510					
mdx	77	49,04	0,490					

XL-Stat
statistiche
descrittive-
spuntare
Dati
qualitativi
Selezionare
la Colonna
Topo
(compresa
l'intestazione
-

Per ottenere il grafico a torta con Excel, selezionare la tabella con le intestazioni C57 e mdx:

Frequenza (Topo)



Principali indici statistici per variabili quantitative

Con Excel si possono utilizzare le funzioni:

CONTA.SE con il criterio "" per contare i dati mancanti

MIN QUARTILE MAX DEV.ST.POP

Ottenendo :

tot dati	157
dati manca	20
minimo	76,3
massimo	335,2
1° quartile	150,7
Mediana	183
3°quartile	217,2
Media	186,312
dev st	51,889

Oppure con XLSTAT 7.0

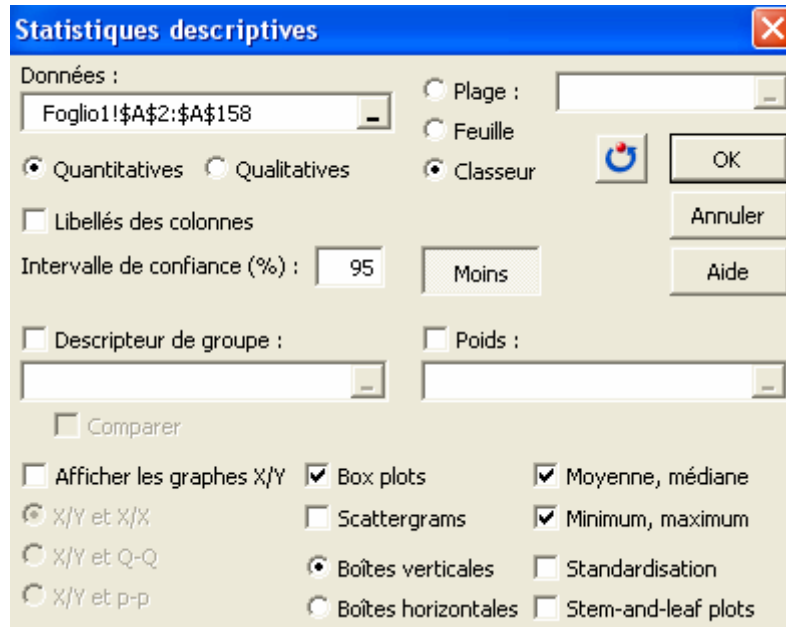
XLSTAT 7.0 - Statistiques descriptives - le 05/11/2008 à 17.30.03	
Description de données quantitatives	
Données : classeur = Cartel1 / feuille = Foglio1 / plage = \$A\$2:\$A\$158 / 157 lignes et 1 colonne	
Les valeurs manquantes dans <Données> ont été ignorées	
Pondération uniforme (par défaut)	
Intervalle de confiance (%): 95,00	
Peso Polmone	
Nbr. de valeurs utilisées	137
Nbr. de valeurs ignorées	20
Nbr. de val. min.	1
% de val. min.	0,730
Minimum	76,300
1er quartile	143,900
Médiane	183,000
3ème quartile	217,350
Maximum	335,200
Etendue	258,900
Somme	25524,700
Moyenne	186,312
Moyenne géométrique	178,858
Moyenne harmonique	171,029
Aplatissement	-0,079
Asymétrie	0,333
Aplatissement	0,008
Asymétrie	0,341
CV (écart-type/moyenne)	0,280
Variance d'échantillon	2692,418
Variance estimée	2712,215
Ecart-type d'échantillon	51,889
Ecart-type estimé	52,079
Ecart absolu moyen	41,144
Ecart absolu médian	33,900
Ecart-type de la moyenne	4,449
Borne inf. IC de la moyenne	177,513
Borne sup. IC de la moyenne	195,111

Nota: abbiamo già evidenziato che la variabile età è considerata qualitativa

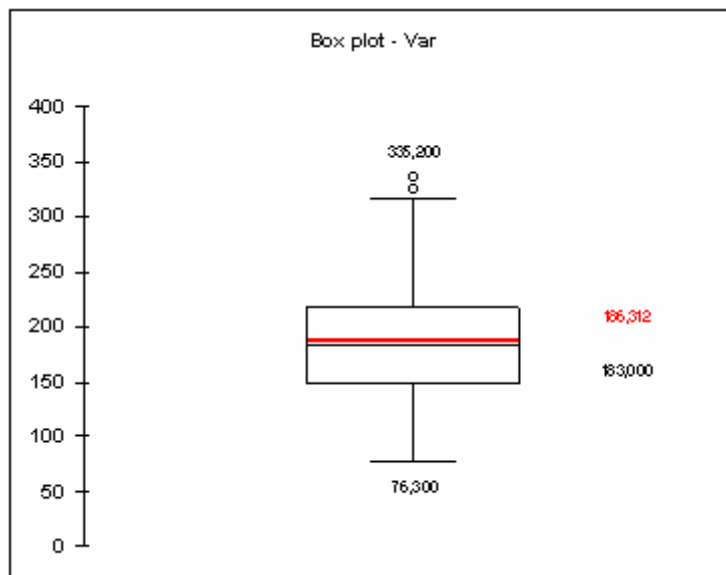
Vediamo anche qualche rappresentazione grafica delle variabili considerate singolarmente.

Box-plot di una singola variabile

Per ottenere il boxplot con XLSTAT si seleziona la serie dei dati; dal menu della statistica descrittiva si spunta il boxplot;



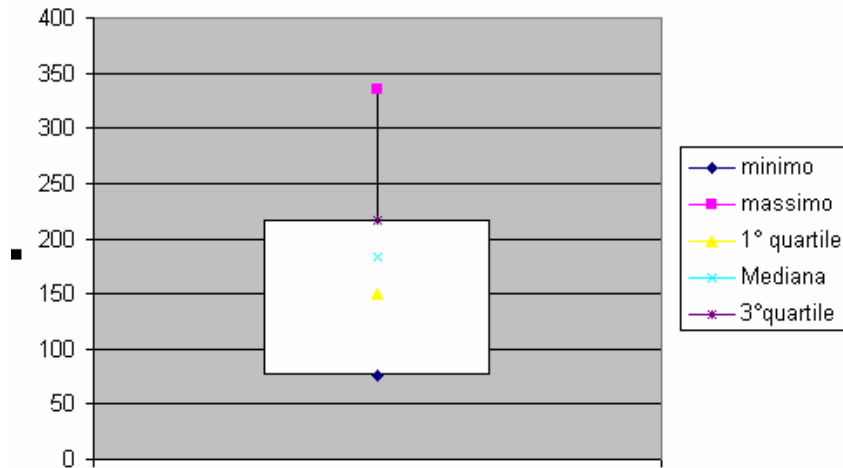
La finestra successiva permette di ignorare i dati mancanti il cui numero viene segnalato. Si ottiene:



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 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.

Per fare un grafico simile al boxplot, con Excel, cliccare su grafico linee, scegliere serie in Righe, selezionare con Ctrl i valori di minimo, massimo, 1° quartile, mediana, 3° quartile con il loro nome, Fine. Fare un doppio clic su uno dei punti ottenuti nel grafico; da Opzioni spuntare linee di min-max e Barre crescenti/decrescenti. Si ha:



Istogramma

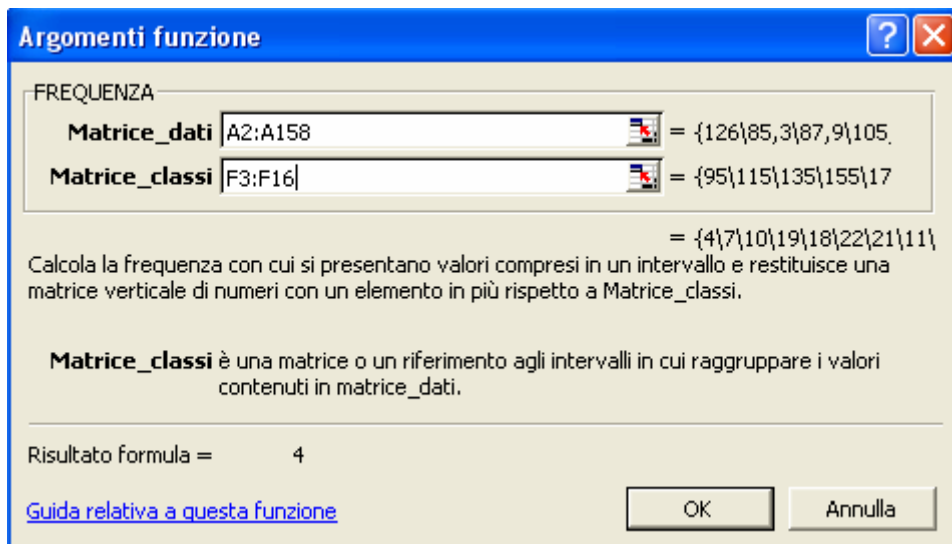
E' possibile creare delle classi di uguale ampiezza, ad esempio:

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

Nel foglio contenente i dati, selezionare una cella (es. E3) e scrivere 75; nella cella sottostante (E4) scrivere 95; selezionare con il mouse queste due celle e copiare (trascinando la crocetta in basso a destra) fino al valore 335. A fianco (in F3) scrivere 95 e poi (in F4) 115. Come prima copiare fino al valore 355. Nella colonna seguente (G3) scrivere: =E3&"-"&F3 e copiare sotto. Si ottiene:

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

Usare la funzione frequenza come segue:



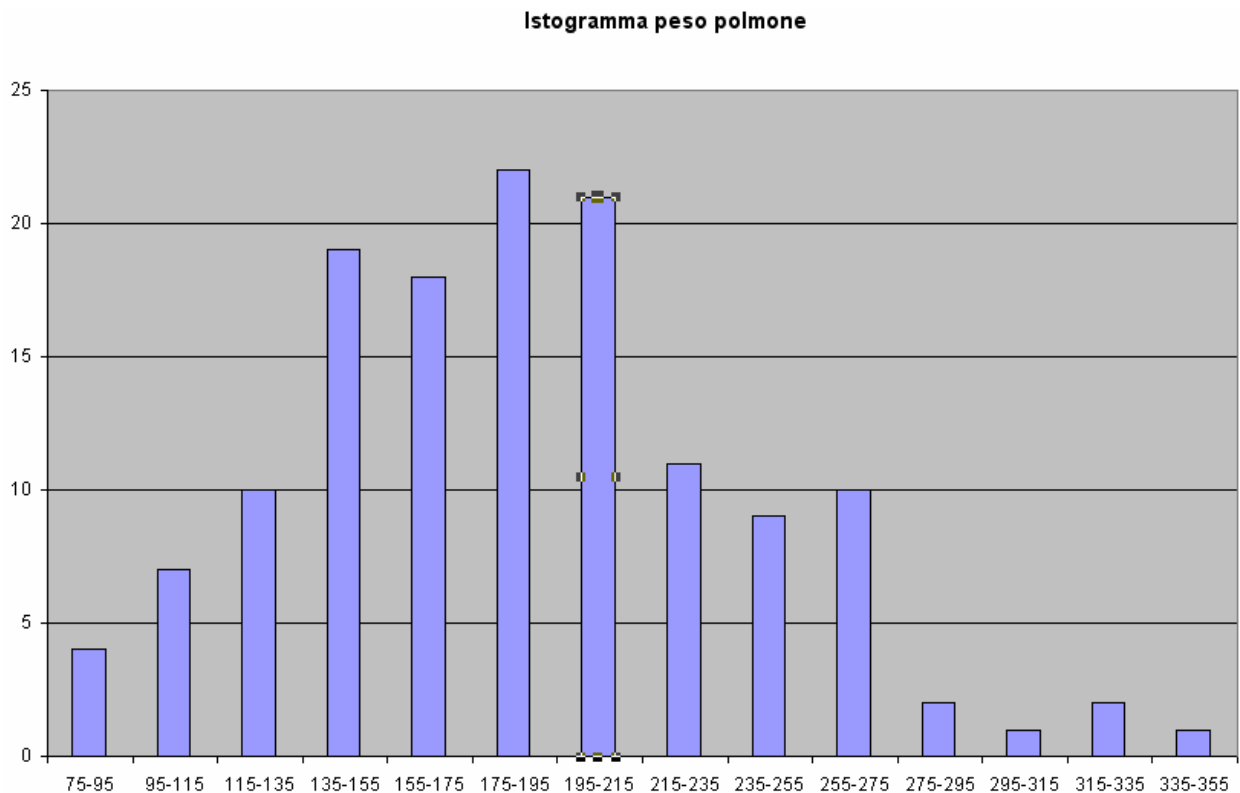
OK

Poiché la funzione frequenza deve restituire una matrice, selezionare con il mouse la zona in cui tale matrice deve apparire (es. da H3 ad H16), cliccare il tasto F2 poi contemporaneamente i tasti Ctrl+Maiusc+ Invio. Si ottiene:

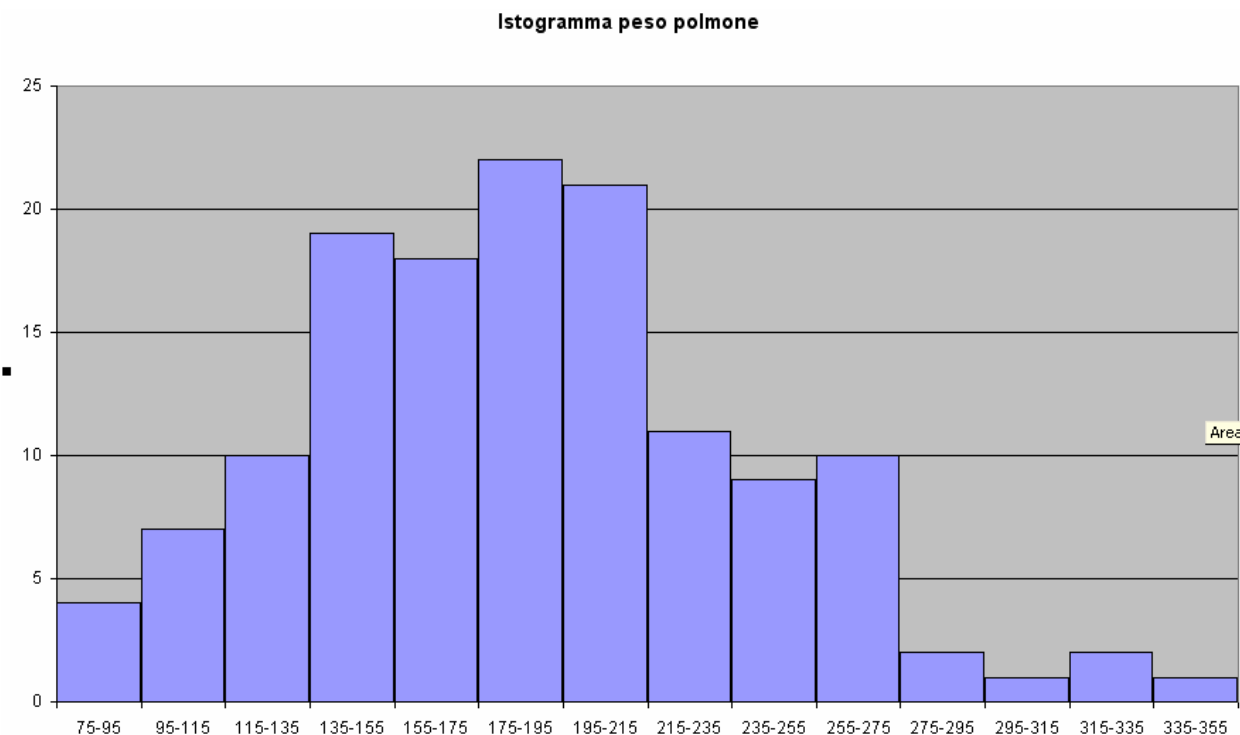
E	F	G	H
75	95	75-95	4
95	115	95-115	7
115	135	115-135	10
135	155	135-155	19
155	175	155-175	18
175	195	175-195	22
195	215	195-215	21
215	235	215-235	11
235	255	235-255	9
255	275	255-275	10
275	295	275-295	2
295	315	295-315	1
315	335	315-335	2
335	355	335-355	1

Riportare in un istogramma, cioè selezionare la zona finale (es. H3:H16) Creazione guidata Grafico Istogramma Avanti ; in Serie Etichette asse categorie X selezionare le classi (da G3 a G16). Si

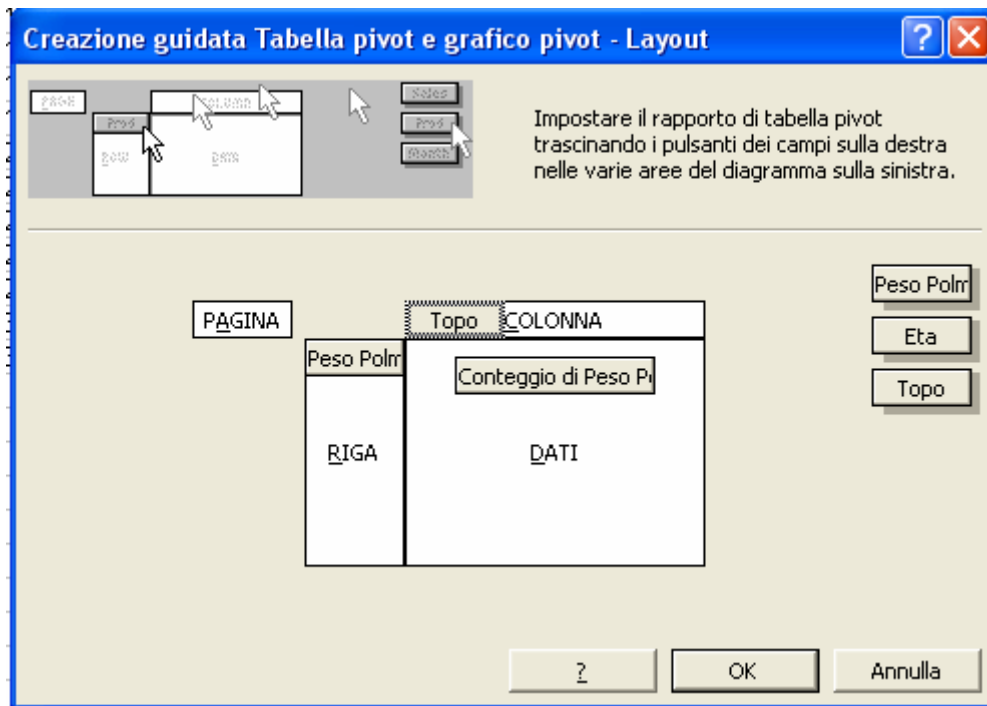
ottiene:



La selezione di un rettangolo come in figura permette di ridurre a zero la distanza tra le barre; in Formato dato Distanza tra le barre =0.



Per riportare sullo stesso piano cartesiano gli istogrammi del peso nel caso C57 e nel caso mdx, si può creare una tabella pivot, come segue:

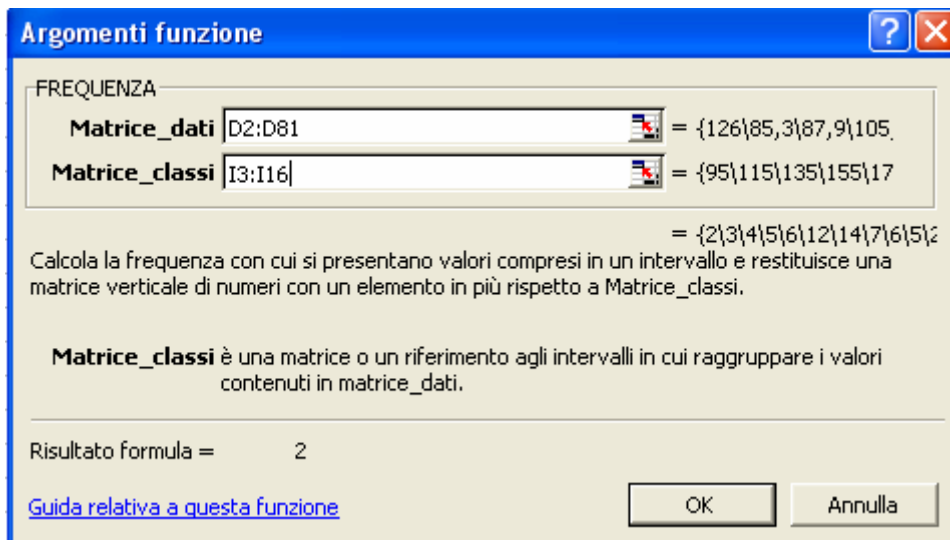


spuntare (dal menu a tendina della tabella pivot) prima la tipologia Topo C57, poi Topo mdx, copiare la colonna del peso polmone Incolla speciale Valori e procedere in modo analogo a quanto realizzato in precedenza.

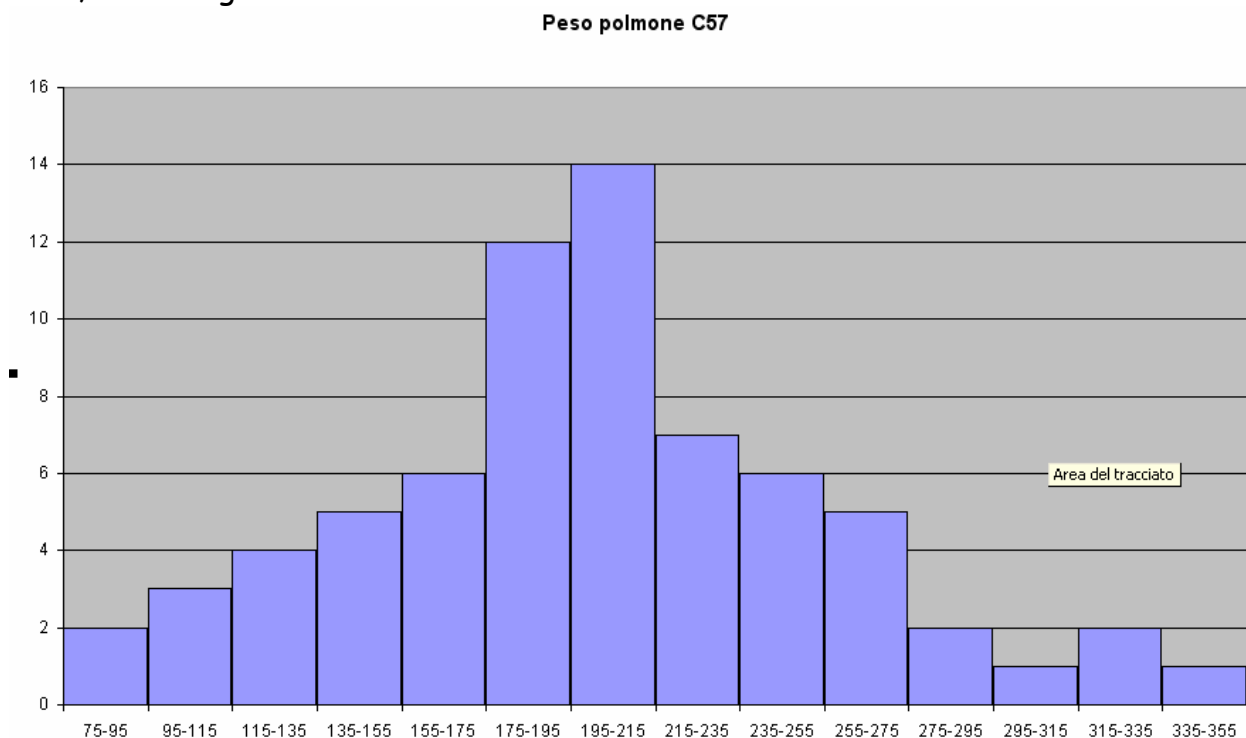
Un modo più rapido è quello di utilizzare il Filtro avanzato nella stessa pagina, cioè: Dal Menu Dati Filtro avanzato, fare questo tipo di opzione:



Si può applicare la funzione frequenza a fianco delle frequenze trovate in precedenza



ecc. , da cui il grafico



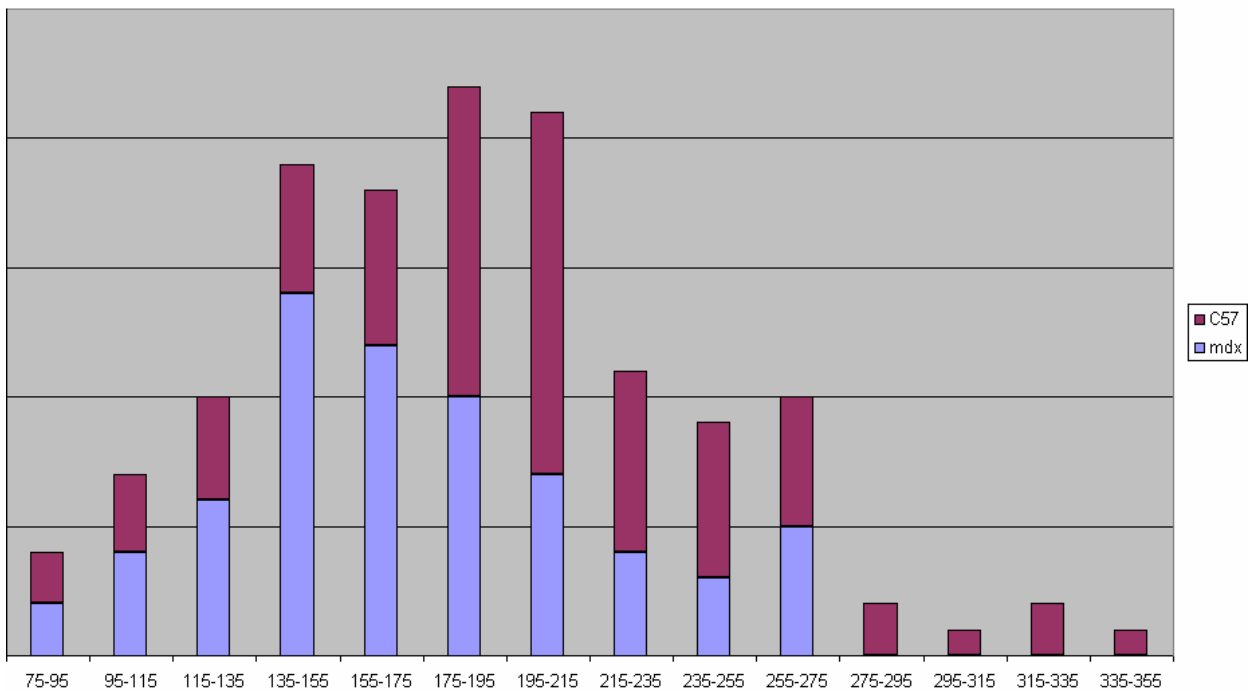
Analogamente per i topi malati.

Se si vogliono evidenziare gli individui sani distinguendo da quelli malati, cliccando su uno dei due grafici, scegliere dal menu Grafico Dati di origine, Aggiungi Serie.

Tipo di grafico...

Si ottiene il grafico in pila:

Peso polmone mdx e C57

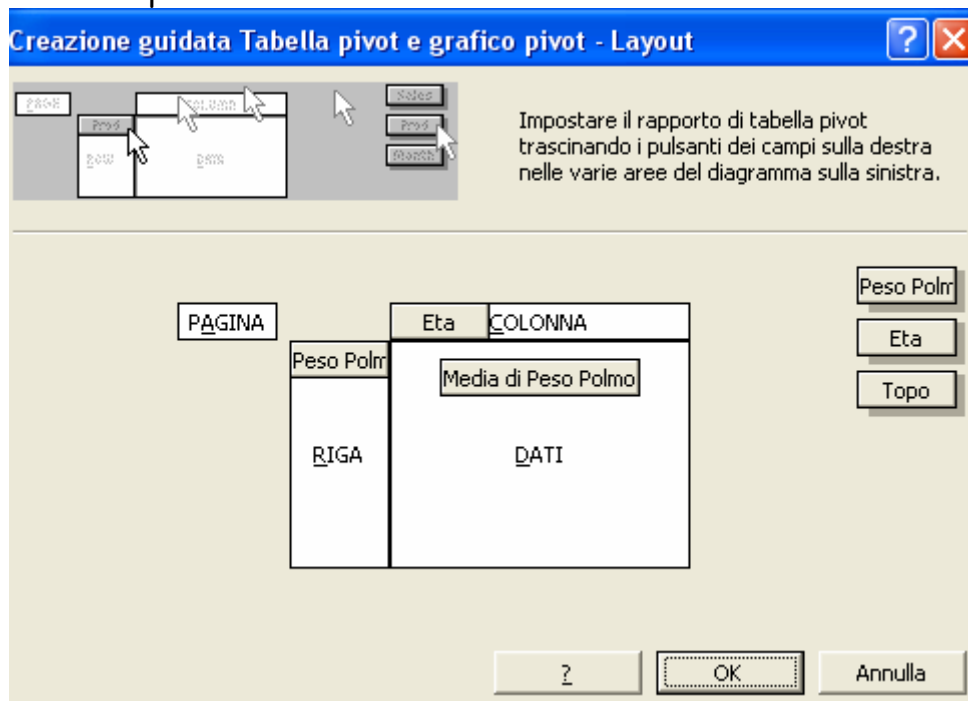


Procediamo vedendo il peso medio del polmone rispetto all'età.

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

Primo metodo

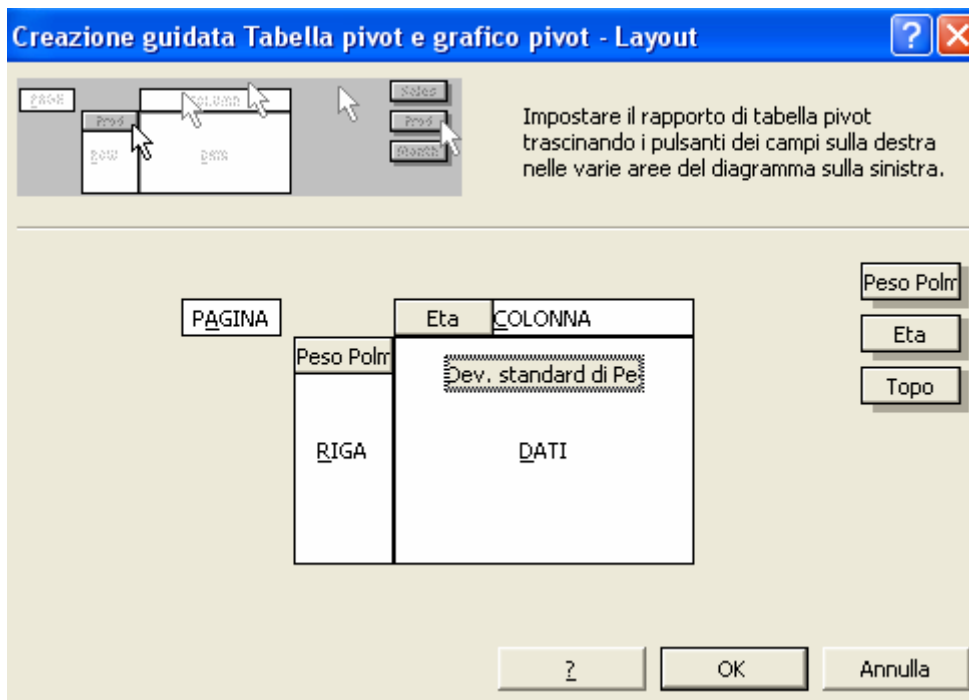
con questa tabella pivot:



(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
-------	-------	-------	-------	-------	-------	-------	-------	-------

Nota: per copiare dati che sono in riga, mettendoli in colonna (o viceversa):

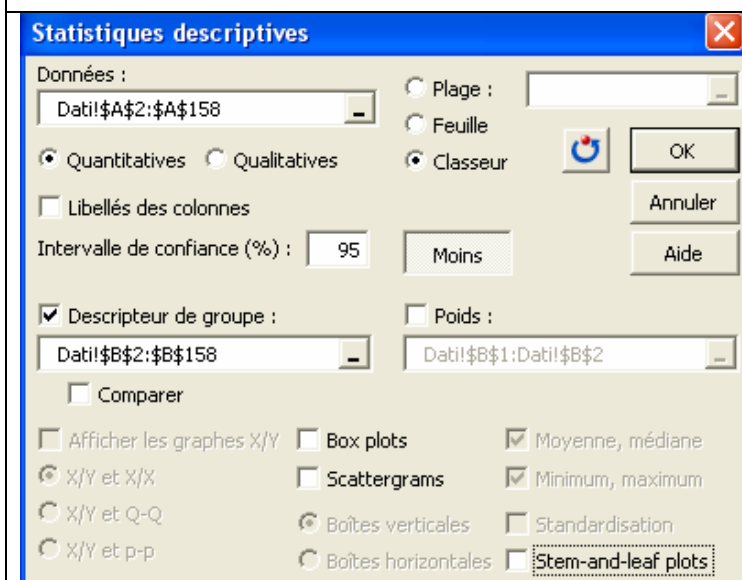
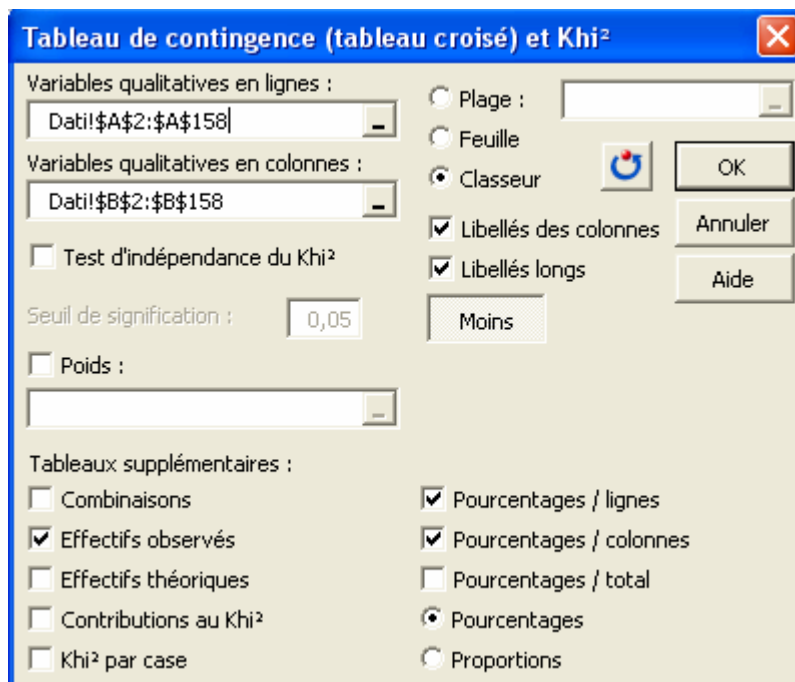
Selezionare le celle in cui si vuole ottenere la colonna (o la riga); digitare la funzione `MATR.TRASPOSTA` con l'indirizzo delle celle da copiare; pigiare simultaneamente `Ctrl + Maiuscolo + Invio`

Secondo metodo, proseguendo con XL- Stat

<p>Tabella del tipo:</p>	<p>XL Stat Creare una tabella di contingenza - Variabile righe (selezionare colonna peso) Variabile colonne (selezionare colonna età). Ignorare i dati mancanti Tra i risultati, si ottiene il conteggio dei dati effettivi osservati.</p>
--------------------------	---

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: XL Stat - Descrizione dei dati- Tabella di contingenza



Per calcolare le medie di sottogruppo del peso rispetto all'età, ossia per livelli di età, si può completare la tabella appena descritta, con XL Stat - Statistiche descrittive - Selezionare la colonna dei Pesi come Dati quantitativi e spuntare Pesi, selezionando la colonna relativa ad una età

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

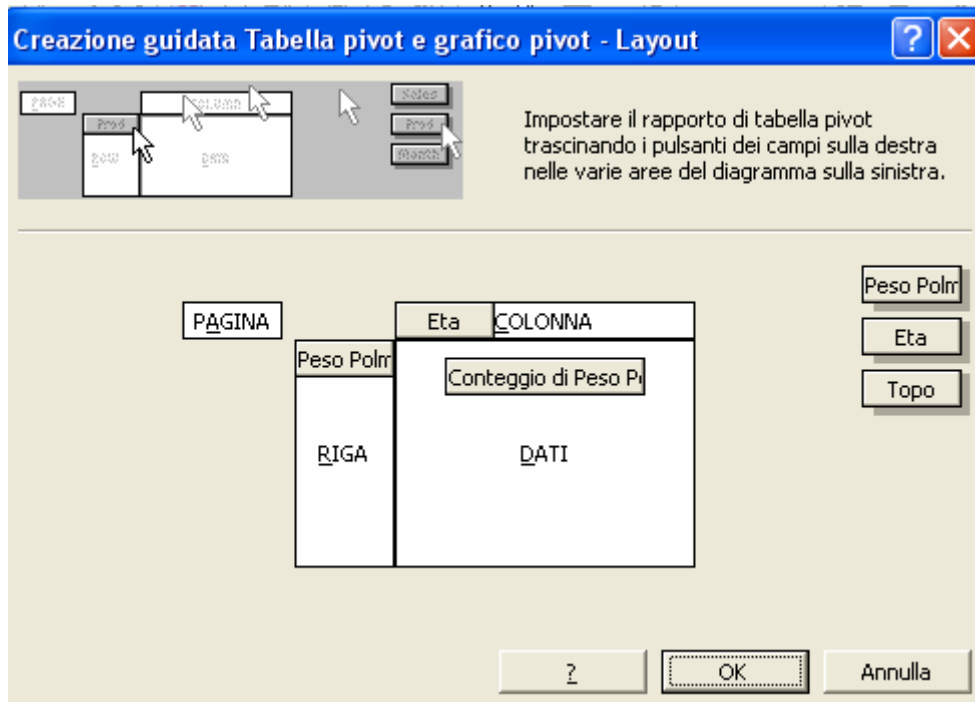
Commento: i topi di 3 settimane hanno mediamente il peso del polmone molto più basso degli altri. Questo è dovuto al fatto che i polmoni in quell'età non si sono ancora sviluppati completamente. Potrebbe essere utile procedere all'analisi statistica eliminando anche le osservazioni relative ai topi di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

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 tabella di contingenza, basata su 137 unità

In questo contesto non è più necessario, però, avere classi piccole, quindi si possono suddividere i dati dei pesi del polmone in 6 classi. Siccome il peso massimo è 335.20 e il minimo è 120.20, determiniamo 6 classi di peso

<120,20

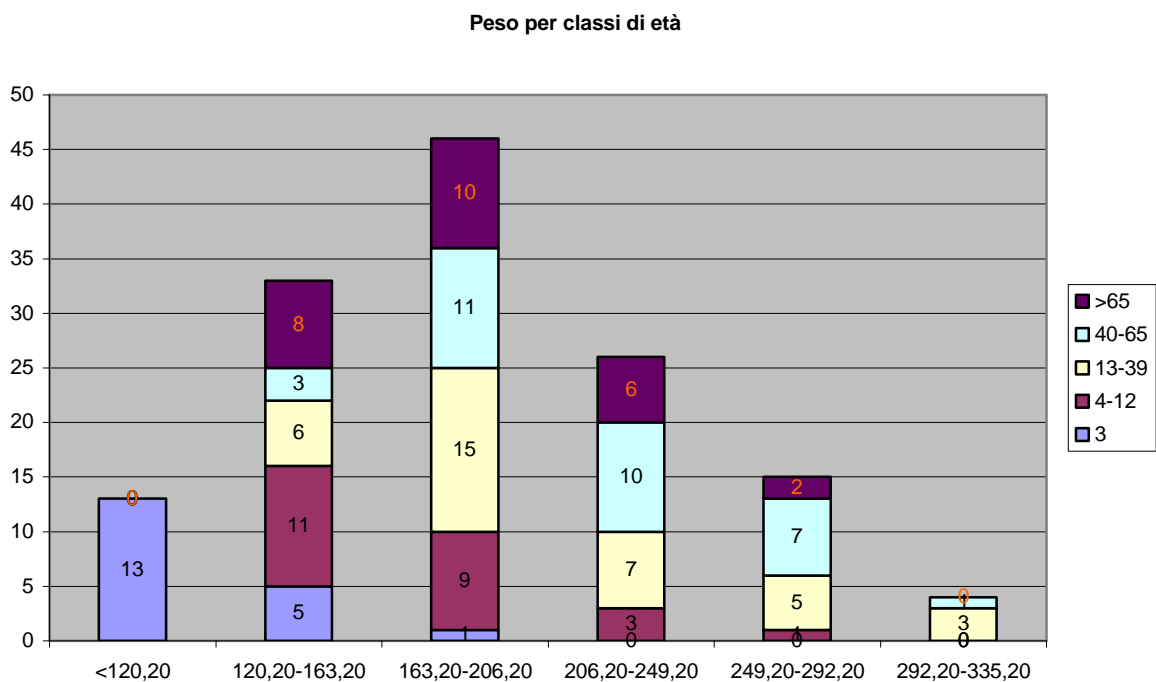
120,20:163,20

163,20:206,20
 206,20:249,20
 249,20:292,20
 249,20:292,20
 292,20:335,20

Usando la formattazione condizionale, è rapida la formulazione della seguente tabella di contingenza:

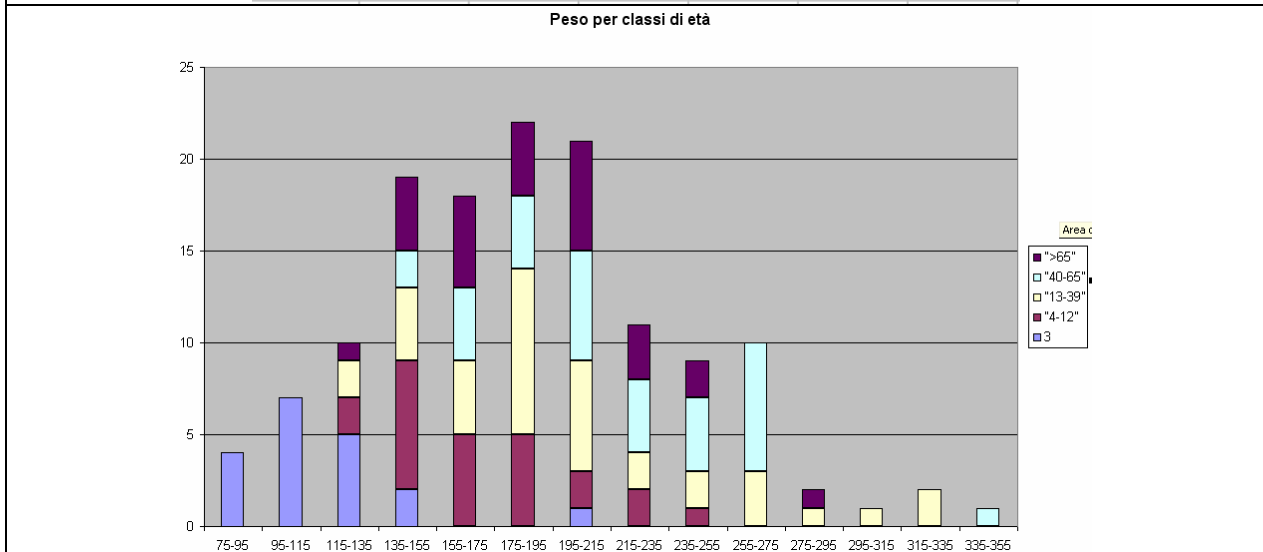
	3	6	12	26	39	52	65	78	104
<120,20	13	0	0	0	0	0	0	0	0
120,20-163,20	5	4	7	3	3	0	3	4	4
163,20-206,20	1	5	4	9	6	6	5	6	4
206,20-249,20	0	0	3	5	2	5	5	3	3
249,20-292,20	0	1	0	3	2	5	2	0	2
292,20-335,20	0	0	0	2	1	1	0	0	0

da cui si ottiene il seguente diagramma in pila:



Con le classi determinate in precedenza, il risultato sarebbe:

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



Dal confronto 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).

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

	3	4-12	13-39	40-65	>65	
<120,20	100,0%	0,0%	0,0%	0,0%	0,0%	100,0%
120,20-163,20	15,2%	33,3%	18,2%	9,1%	24,2%	100,0%
163,20-206,20	2,2%	19,6%	32,6%	23,9%	21,7%	100,0%
206,20-249,20	0,0%	11,5%	26,9%	38,5%	23,1%	100,0%
249,20-292,20	0,0%	6,7%	33,3%	46,7%	13,3%	100,0%
292,20-335,20	0,0%	0,0%	75,0%	25,0%	0,0%	100,0%
	13,9%	17,5%	26,3%	23,4%	19,0%	100,0%

	C57	mdx	totale
<120,20	38,46%	61,54%	100,00%
120,20-163,20	27,27%	72,73%	100,00%
163,20-206,20	52,17%	47,83%	100,00%
206,20-249,20	69,23%	30,77%	100,00%
249,20-292,20	66,67%	33,33%	100,00%
292,20-335,20	100,00%	0,00%	100,00%
	51,09%	48,91%	100,00%

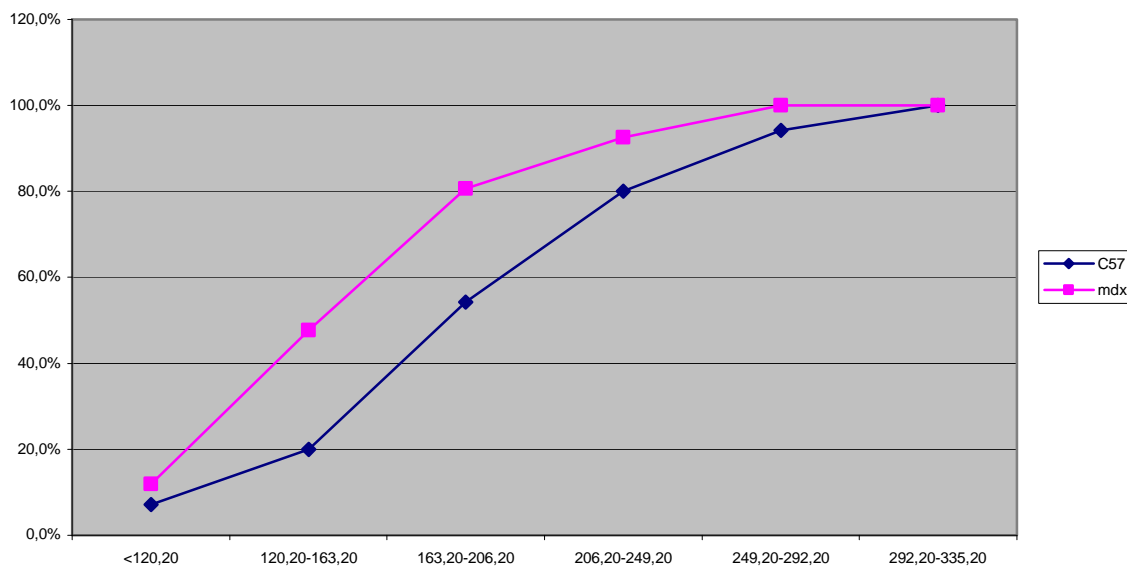
	3	4-12	13-39	40-65	>65
<120,20	68,4%	0,0%	0,0%	0,0%	0,0%
120,20-163,20	26,3%	45,8%	16,7%	9,4%	30,8%
163,20-206,20	5,3%	37,5%	41,7%	34,4%	38,5%
206,20-249,20	0,0%	12,5%	19,4%	31,3%	23,1%
249,20-292,20	0,0%	4,2%	13,9%	21,9%	7,7%
292,20-335,20	0,0%	0,0%	8,3%	3,1%	0,0%
	100,0%	100,0%	100,0%	100,0%	100,0%

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.

	freq cumulate	
	C57	mdx
<120,20	7,1%	11,9%
120,20-163,20	20,0%	47,8%
163,20-206,20	54,3%	80,6%
206,20-249,20	80,0%	92,5%
249,20-292,20	94,3%	100,0%
292,20-335,20	100,0%	100,0%

frequenze cumulate (sei classi peso)



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

