

La distribuzione normale standardizzata

La distribuzione normale standardizzata è importante perché le probabilità per tutte le distribuzioni normali sono calcolate utilizzando la distribuzione normale standard.

Se abbiamo una distribuzione normale le probabilità corrispondenti alle aree sottese alla curva normale tra due estremi x_1 e x_2 possono essere calcolate trasformando la distribuzione nella distribuzione normale standard.

Queste probabilità vengono riportate in un'apposita tavola.

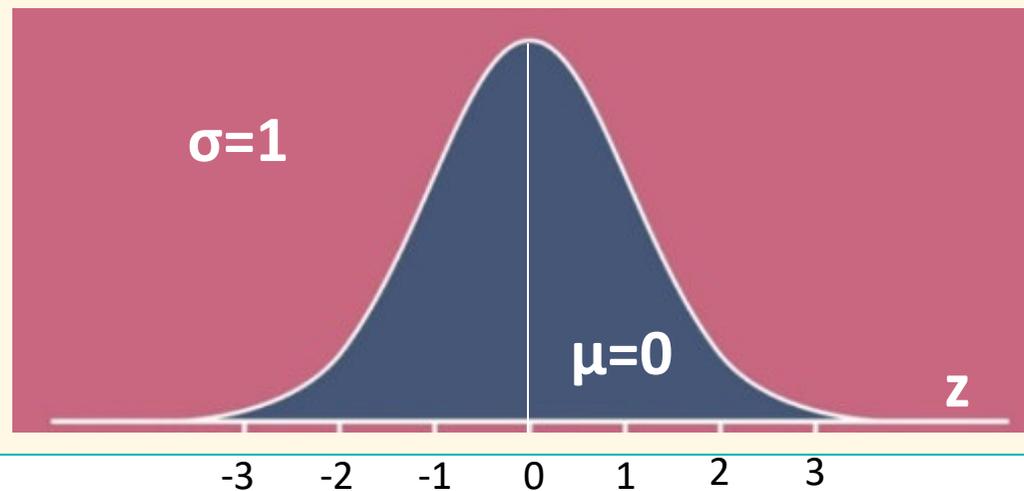
In questo modo è possibile evitare il ricorso a complessi calcoli integrali per trovare le probabilità che una variabile X assuma valori compresi all'interno di determinati intervalli.

Inoltre permette il confronto immediato tra distribuzioni diverse.

La distribuzione normale standardizzata

La distribuzione normale standardizzata, detta z , è quella particolare distribuzione normale con media uguale a 0 e deviazione standard pari a 1.

La conversione ai valori standard modifica solo l'unità di misura, non la posizione relativa o la distanza dei valori della normale.



Trasformazione di una variabile casuale normale standard

Formula per **standardizzare** una qualunque variabile casuale normale:

$$z = \frac{x - \mu}{\sigma}$$

z è il valore standard di x

La standardizzazione è una trasformazione dei dati che consiste nel rendere la media nulla ($\mu = 0$), dato che ad ogni valore della variabile originaria viene sottratta la media della variabile stessa e assumere la deviazione standard σ quale unità di misura ($\sigma = 1$) della nuova variabile, dato che ogni valore viene diviso per σ .

La distribuzione normale standardizzata viene indicata con $N(0,1)$.

Esempio

Qual è il voto più alto? Uno studente ha sostenuto 2 esami presso 2 diverse università straniere.

Nel primo esame ha ottenuto 27/30, mentre nel secondo 75/100.

Sapendo che la media dei voti di tutti gli studenti nei 2 esami è stata rispettivamente 24,5/30 (con σ 2) e di 68/100 (con σ 6), in quale esame lo studente ha avuto la performance migliore?

Con la standardizzazione si possono confrontare 2 fenomeni che hanno scale differenti.

Valore standardizzato di 27/30: $z_1 = (27-24,5)/2 = 1,25$

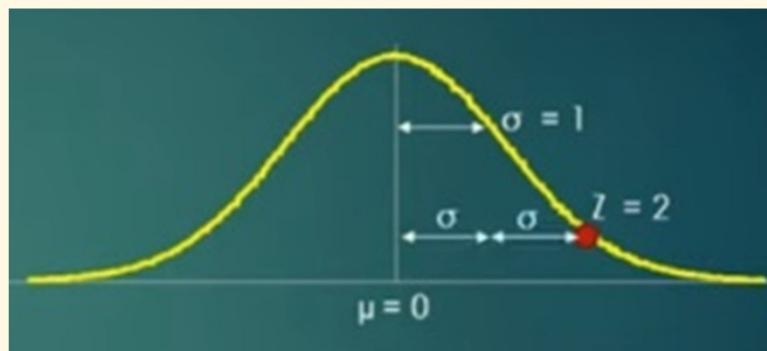
Valore standardizzato di 75/100: $z_2 = (75-68)/6 = 1,17$

Lo studente ha avuto una performance migliore nel primo esame, perché si è allontanato di più dalla media in positivo.

Lo z-score

Lo **z-score** indica di quante deviazioni standard la variabile casuale x si allontana dalla media (sopra o sotto). Fornisce la posizione relativa di x rispetto al centro della distribuzione.

Es. uno z-score di 2 significa che tra la media e la variabile x ci sono 2 deviazioni standard.



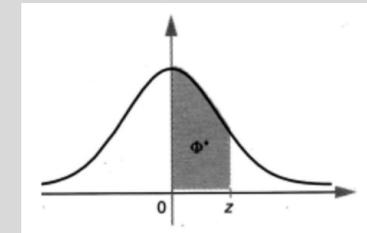
Funzione di densità normale standardizzata

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$\pi = 3.14159$
 $e = 2.71828$ base dei
logaritmi naturali

Funzione di ripartizione della variabile casuale normale standardizzata

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4811
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000



Valori delle aree sotto la curva normale standardizzata da 0 a z

Come si legge la tavola della distribuzione normale standardizzata

Sugli assi sono riportati i valori della variabile z , mentre all'interno si trovano i valori dell'area sotto la curva, calcolati in base al variare di z .

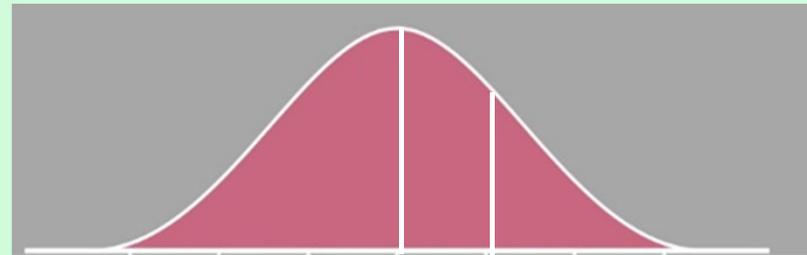
Di solito i valori di z si usano con al massimo 2 cifre decimali, perciò la tabella riporta sull'asse verticale i numeri interi con il primo decimale e su quello orizzontale solo la seconda cifra decimale



Il valore si trova all'incrocio

In virtù della proprietà di simmetria della distribuzione normale, le tavole riportano soltanto i valori dell'area compresa fra lo zero e l'ascissa $+X$, poiché per la simmetria l'area sottesa dall'altra metà della curva è ovviamente uguale.

Esempio

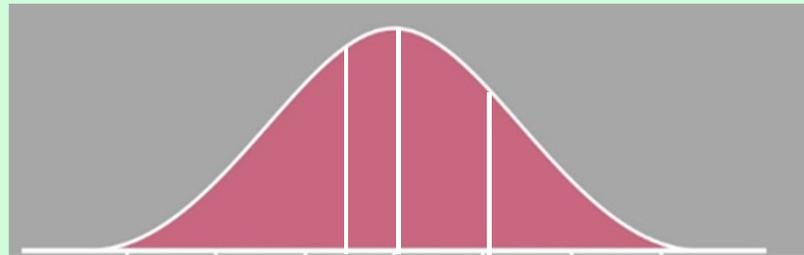


0 1,24

Qual è la probabilità che la variabile normale standardizzata z sia compresa tra 0 e 1,24? 39,25%

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015

Esempio



-0,6 0 1,24

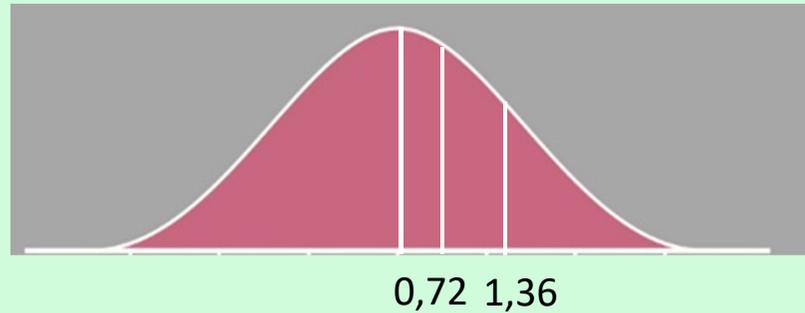
Qual è la probabilità che la variabile normale standardizzata z sia compresa tra -0,6 e 1,24?

$$39,25 + 22,58\% = 61,83\%$$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015

Esempio

Qual è la probabilità che la variabile normale standardizzata z sia compresa tra 0,72 e 1,36? $41,31\% - 26,42\% = 14,89\%$



z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177

Domande di probabilità

I 3 tipi di probabilità che si possono calcolare sono:

- 1- la probabilità che la variabile casuale normale standardizzata z sia \leq a un dato valore
- 2- la probabilità che z sia compreso tra 2 dati valori
- 3- la probabilità che z sia \geq a un dato valore

Domande di probabilità

La tavola della probabilità normale standard ci consente di rispondere alle seguenti domande:

- 1- determinare l'area o la probabilità corrispondente a un valore di z .
- 2- determinare i valori di z a partire da un'area o una probabilità.

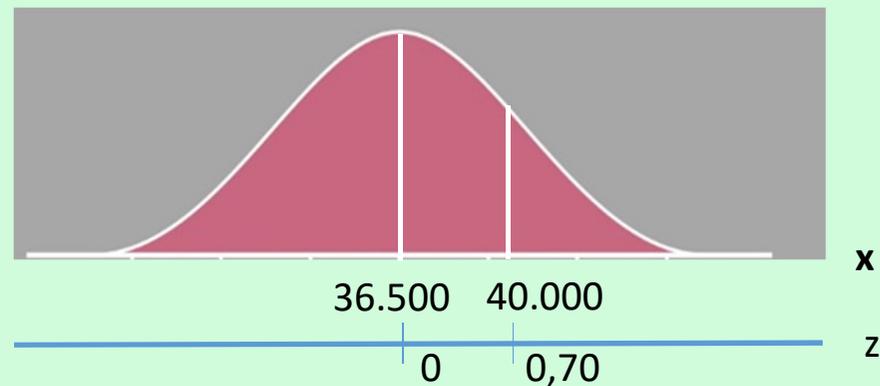
Esempio

La durata media di un pneumatico in termini di distanza percorsa è $\mu = 36.500\text{km}$ e la $\sigma = 5.000$.

I dati raccolti indicano che una distribuzione normale è un'assunzione ragionevole.

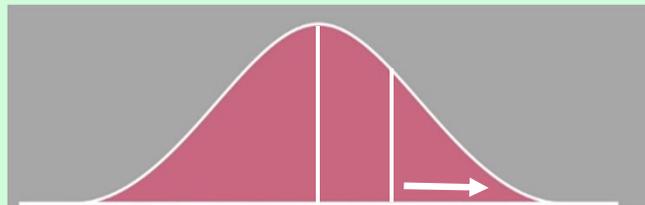
Quale percentuale di pneumatici che durano più di 40.000km ci si può aspettare? Qual è la probabilità che la durata superi i 40.000km ?

$$z = (40.000 - 36.500) / 5.000 = 0,70$$



Esempio

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177

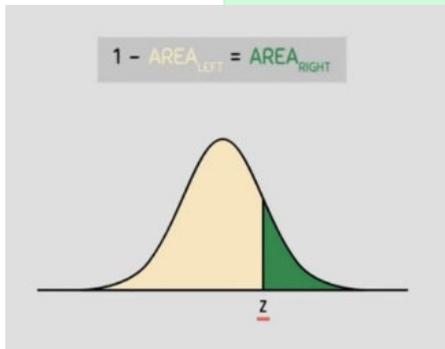
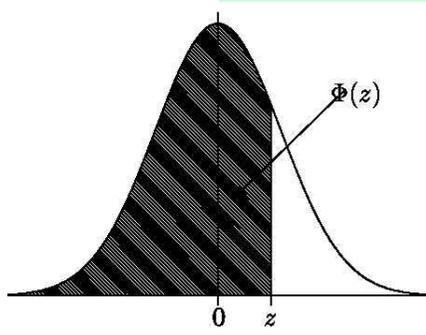


0 0,70

L'area dell'intervallo 0-70 è pari a 0,2580. Per trovare l'area a destra di 0,70, posso fare=
 $1 - (0,5 + 0,2580) = 0,2420$.

La società valuta la possibilità di fornire uno sconto su quegli pneumatici che non forniscono la durata km garantita. Quale dovrebbe essere la durata garantita in km in modo che non più del 10% degli pneumatici portino allo sconto?

Esempio



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Esempio

Sulla tavola si trova il valore di z che corrisponde a un'area a sinistra della media pari a 0,10.

$Z = -1,28$ è il valor che corrisponde alla garanzia desiderata.

Per trovare il valore della x corrispondente:

$$z = \frac{x - \mu}{\sigma} = x = \mu + z\sigma \quad z = -1,28 \quad x - \mu = -1,28\sigma \quad x = \mu - 1,28\sigma$$

$$x = 36.500 - 1,28(5000) = 30.100$$

Una garanzia di 30.100km fa sì che circa il 10% degli pneumatici possa portare all'utilizzo della garanzia da parte degli acquirenti.

Distribuzione di probabilità

Le distribuzioni di probabilità svolgono un ruolo importante nel fornire informazioni utili per le decisioni.

La probabilità non costituisce direttamente un'indicazione sulla decisione da prendere, ma fornisce informazioni che aiutano il soggetto decisore a comprendere meglio i rischi e le incertezze associate al problema.

Esercizio

La paga oraria media per i manager finanziari è 32,62\$ e la deviazione standard è 2,32\$.

Si assume che le paghe siano distribuite secondo una normale.

1- Qual è la probabilità che un manager finanziario guadagni tra i 30\$ e i 35\$ dollari l'ora?

2- Quanto elevata deve essere la paga oraria affinché un manager finanziario faccia parte del 10% dei manager finanziari più pagati?

3- Per un manager finanziario selezionato a caso, qual è la probabilità che guadagni meno di 28\$ l'ora.

Soluzione

$$z_1 = (30-32,62)/2,32 = -1,13$$

$$z_2 = (35-32,62)/2,32 = 1,03$$

$$= 0,3708+0,3485 = 0,7193$$

2- La paga è data da $= 32,62+1,28(2,32) = 35,59\$$

3- $z_3 = (28-32,62)/2,32 = -1,99 = 0,5-0,4767 = 0,0233$

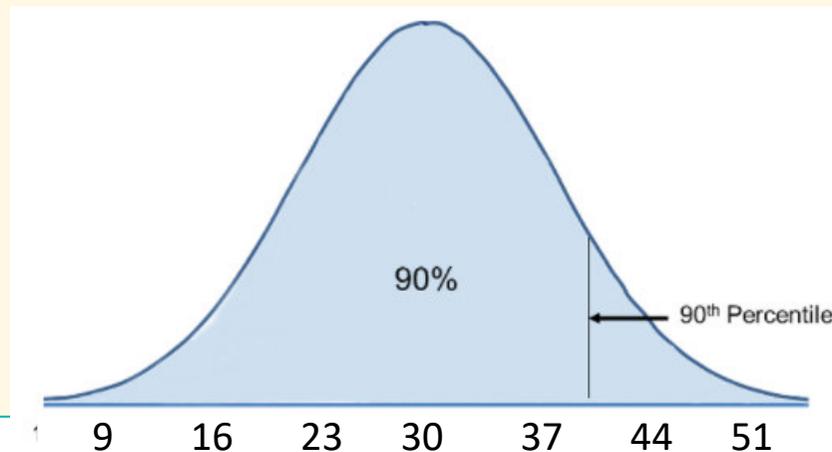
z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1809	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177

La distribuzione normale standardizzata

La distribuzione normale standardizzata può essere usata anche per calcolare i percentili.

$$X = \mu + z\sigma$$

Es. il BMI medio per uomini di 60 anni è 30 e la dev std è 7.
Qual è il 90° percentile del BMI?
In questo caso dobbiamo trovare il valore di X che rappresenti il 90% di probabilità.



Come calcolare i percentili

Con la distribuzione standardizzata possiamo trovare l'area sotto la curva più vicina al 90% e in questo modo determinare il punteggio Z.

Sulla tabella non si trova il valore preciso 0,90, ma il valore 0,8997 che corrisponde al valore Z di 1,28 (l'89,97% dell'area sotto la curva è al di sotto di 1,28).



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177

Come calcolare i percentili

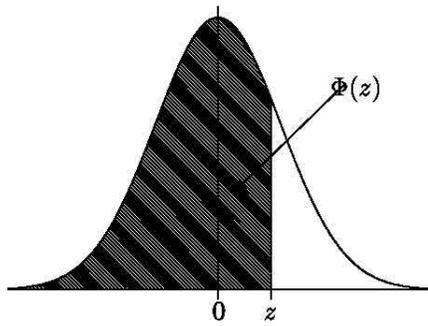
Il valore preciso di Z che contiene fino al 90% dei valori è 1,28.

Usando l'equazione $X = \mu + Z\sigma$

$$X = 30 + 1,28(7) = 38,96.$$

Come interpretare i dati: il 90% del BMI negli uomini di 60 anni è sotto il valore di 38,96, il mentre il 10% è sopra il valore di 38,96.

Il 50esimo percentile?



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Calcolo del 93esimo percentile

La media dei voti di esame è $\mu = 26$ con standard deviation di $\sigma = 2$.

Qual è il punteggio del 93esimo percentile?

Bisogna trovare lo z-score che si avvicina al valore 0,93.

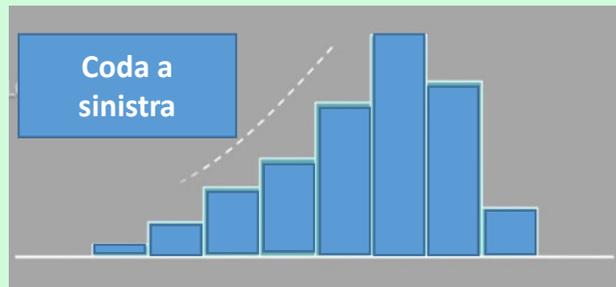
Questo valore sarà 1,48.

$$93\text{rd percentile} = \mu + z\sigma = 26 + (1.48)*2 = 28,96$$

Uno studente al 93esimo percentile dovrebbe ottenere un punteggio di 28,96.

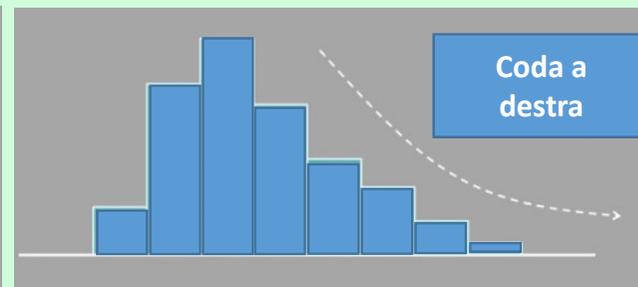
Asimmetria

Asimmetrica a sinistra



Curva asimmetrica negativa

Asimmetrica a destra

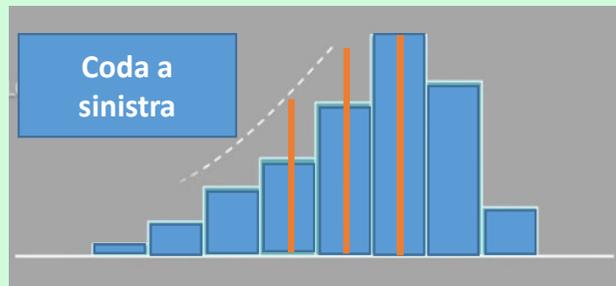


Curva asimmetrica positiva

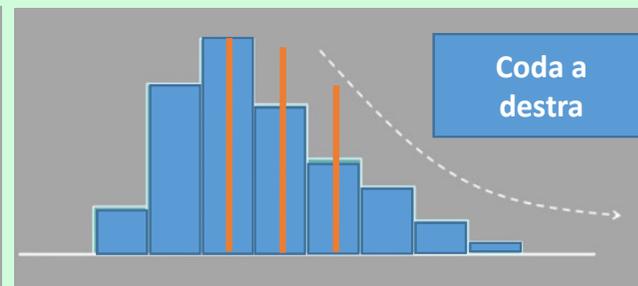
In teoria delle probabilità una distribuzione di probabilità è simmetrica quando la sua funzione di probabilità P (nel caso discreto) o la sua funzione di densità di probabilità (nel caso continuo) siano simmetriche rispetto ad un particolare valore $X = \mu$.

Asimmetria

Asimmetria positiva: $\text{Moda} < \text{Mediana} < \text{Media}$
Asimmetria negativa: $\text{Media} < \text{Mediana} < \text{Moda}$



$\text{Media} < \text{Mediana} < \text{Moda}$



$\text{Moda} < \text{Mediana} < \text{Media}$

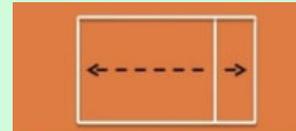
L'elemento fondamentale da ricordare è che la media si trova dalla parte dell'asimmetria perché è influenzata dai punteggi estremi. Così quando la media è maggiore della mediana, la distribuzione può essere denominata con asimmetria positiva; se invece la media aritmetica è più piccola della mediana allora la distribuzione è asimmetrica negativamente.

Boxplot

Il boxplot è molto utile per capire se la distribuzione è simmetrica e per confrontare la forma di più distribuzioni.

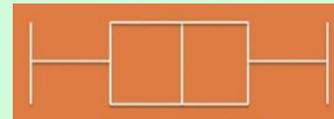
1- Se la mediana non divide la scatola in 2 parti uguali bisogna vedere da che parte è la dispersione maggiore dei valori:

Asimmetrica a sinistra



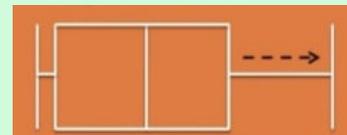
2- Se la scatola è divisa in 2 parti uguali dalla mediana e i baffi hanno la stessa lunghezza:

Curva simmetrica



3- Se la scatola è divisa in 2 parti uguali dalla mediana, bisogna vedere la lunghezza dei baffi

Asimmetrica a destra



Indici di forma

Gli indici di forma di una distribuzione riguardano 2 caratteristiche: la simmetria e la curtosi. Sono indici utilizzati per evidenziare la forma di una distribuzione.

Un **indice di asimmetria** di una distribuzione è un valore che cerca di fornire una misura della sua mancanza di simmetria.

Con il termine **curtosi** si intende la maggiore o minore gibbosità (appiattimento) di una distribuzione in prossimità del suo massimo. Assume rilevanza nell'ambito delle distribuzioni unimodali di forma campanulare.

Misure di asimmetria basate sui quartili

Un indice assoluto di asimmetria basato sulle distanze tra quartili è:

$$aQ = (Q3 - Q2) - (Q2 - Q1)$$

in cui la misura di asimmetria è espressa in termini di differenza tra la distanza che separa il terzo quartile dalla mediana $Q3 - Q2$ e quella che separa la mediana dal primo quartile $Q2 - Q1$.

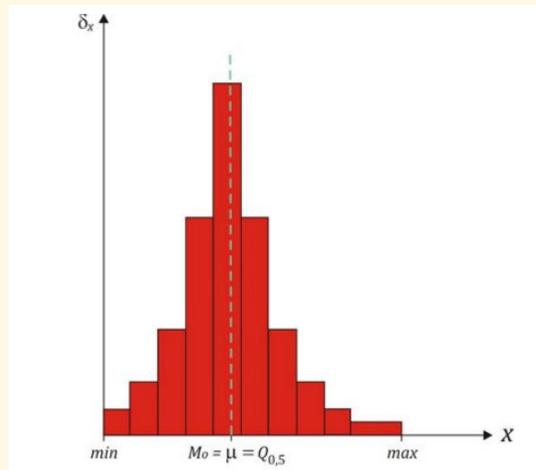
Distinguiamo tre casi:

- 1- Se la distribuzione è simmetrica, la quantità $Q3 - Q2 = Q2 - Q1$, quindi aQ tende ad assumere valori prossimi a zero.
- 2- In caso di asimmetria positiva, la quantità $Q3 - Q2 > Q2 - Q1$, di conseguenza $aQ > 0$
- 3- In caso di asimmetria negativa, la quantità $Q3 - Q2 < Q2 - Q1$ e quindi $aQ < 0$.

$$\text{Indice di Yule e Bowley} = \frac{(Q3 - Q2) - (Q2 - Q1)}{(Q3 - Q1) \text{ scarto interquartile}}$$

Indice di asimmetria

Se in una distribuzione osservata media, moda e mediana coincidono, siamo in presenza di una distribuzione simmetrica?



La non coincidenza dei 3 indici è un indicatore di asimmetria, mentre la coincidenza non garantisce la simmetria.

Momenti

In Statistica esiste un insieme di indici, chiamati momenti, di cui fa parte anche la media aritmetica. Possono essere calcolati solo per variabili quantitative (sia discrete, sia continue).

Servono a descrivere le caratteristiche di una distribuzione. Ogni distribuzione può essere descritta attraverso i momenti.

Il **momento ordinario** di ordine r corrisponde alla media dei valori della variabile elevati alla potenza r -esima.

I momenti centrali sono la media delle erresime potenze degli scarti centrati sulla media aritmetica.

Momenti

Il momento ordinario di ordine zero (ossia calcolato per $r=0$) risulta sempre uguale a 1, mentre per $r = 1$ si ottiene la media delle osservazioni: la media aritmetica, quindi, corrisponde al momento ordinario di ordine 1.

$$\mu = \frac{\sum_{i=1}^N x_i n_i}{\sum_{i=1}^N n_i}$$

Momenti

Momenti di ordine	Momenti	Momenti per distribuzione di frequenza	Indici	Misure
1	$\frac{\sum x_i}{n}$	$\frac{\sum x_i \cdot n_i}{n}$	Media	Tendenza centrale
2 centrale	$\frac{\sum (x_i - \mu)^2}{n}$	$\frac{\sum (x_i - \mu)^2 \cdot n_i}{n}$	Varianza	Dispersione
3 centrale	$\frac{\sum (x_i - \mu)^3}{n}$	$\frac{\sum (x_i - \mu)^3 \cdot n_i}{n}$	Indice di simmetria a M3	Asimmetria
4 centrale	$\frac{\sum (x_i - \mu)^4}{n}$	$\frac{\sum (x_i - \mu)^4 \cdot n_i}{n}$		Curtosi

Momenti centrale di ordine 3

Un indice di asimmetria molto usato è quello basato sul **momento centrato di ordine 3, M3**.

$$\frac{\sum (x_i - \mu)^3 \cdot n_i}{n}$$

L'indicatore dà maggiore rilievo agli scostamenti rispetto alla media, elevando gli scarti ad una potenza dispari: l'elevazione a potenza degli scarti ne aumenta il valore, rendendo quindi più sensibile l'indice.

Indice di asimmetria di Fisher

Se volessimo considerare un indice che non dipende dall'unità di misura del carattere e confrontare caratteri diversi, si può usare la formulazione di Fisher:

L'**Indice di Asimmetria di Fisher** è definito come la Media aritmetica delle terze potenze della variabile standardizzata:

$$\beta = \frac{1}{N} \sum \left(\frac{x_i - \mu}{\sigma} \right)^3$$

Se $\beta > 0$ c'è asimmetria positiva.

Se $\beta = 0$ c'è simmetria.

Se $\beta < 0$ c'è asimmetria negativa.

Indice di asimmetria di Fisher

L'espressione dentro la parentesi è definita scarto standardizzato, per cui l'indice è pari alla media dei cubi degli scarti standardizzati.

Per le distribuzioni di frequenze la formula di Fisher è:

$$\beta = \frac{1}{N} \sum \left(\frac{x_i - \mu}{\sigma} \right)^3 n_i$$

Esercizio

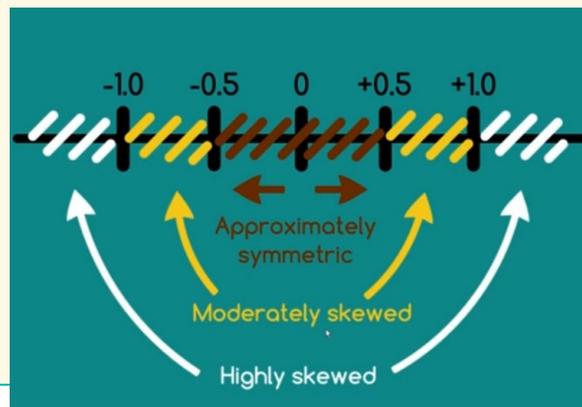
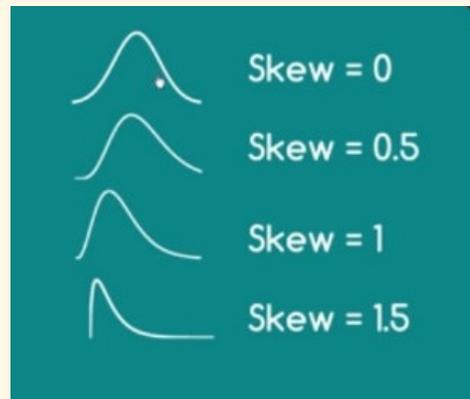
Calcolare l'indice di Fisher per la seguente distribuzione di frequenza:

Numero di stanze	Ospiti
1	115
2	1.563
3	4.879
4	8.055
5	5.321
	19.933

Svolgimento

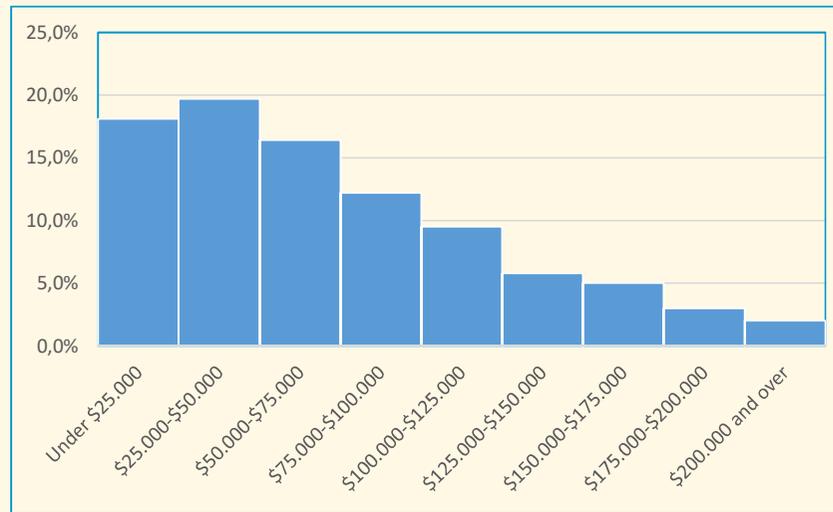
Numero di stanze	Ospiti	$x_i \cdot n_i$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 \cdot n_i$	$(x_i - \mu)^3$	$(x_i - \mu)^3 \cdot n_i$
1	115	115	-2,85	8,12	934	-23	-2.662
2	1.563	3.126	-1,85	3,42	5.349	-6	-9.896
3	4.879	14.637	-0,85	0,72	3.525	-1	-2.996
4	8.055	32.220	0,15	0,02	181	0	27
5	5.321	26.605	1,15	1,32	7.037	2	8.093
	Σ 19.933	Σ 76.703			17.027	-28,57	-7.435,03
		$\mu=3,85$ (76.703/19.933)			$\sigma^2 = 0,85$ ='17027/19.933		$M3=-0,37=$ (-7.435,03/19.933)
					$\sigma = 0,924 =$ $0,85^{(1/2)}$		$\beta = -0,47 (-0,37/(\sigma^3))$

Asimmetria (skewness)



Esempio asimmetria positiva

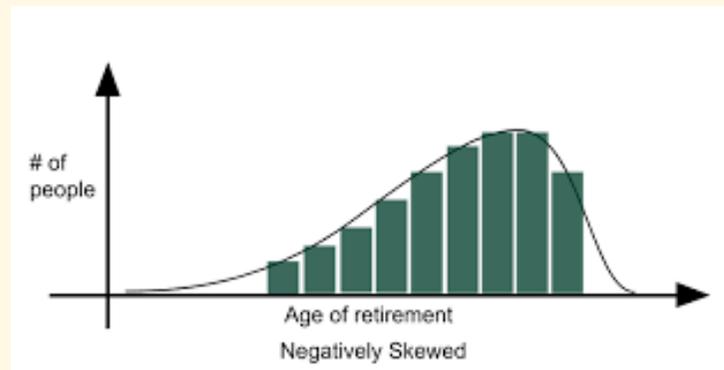
Average Income Distribution



Esempio asimmetria negativa



Esempio asimmetria negativa



Esercizio

Partendo dalla seguente distribuzione 5, 6, 5, 4, 2, 6, 3, 4, 1, 5, 4, 5, 3, 5, 6, 4 calcolare gli indici di simmetria basati sui quartili, M3 e Fisher. Commentare i risultati ottenuti.

Soluzione

xi	ni	Ni
1	1	1
2	1	2
3	2	4
4	4	8
5	5	13
6	3	16
	$\Sigma 16$	

$Q2 = 0,5 * 16 =$ posizione tra la 8

e la 9 = $(4+5)/2 = 4,5$

$Q1 = 0,25 * 16 =$ tra la 4 e 5 =

$(3+4)/2 = 3,5$

$Q3 = 0,75 * 16 =$ tra la 12 e la 13

= $(5+5)/2 = 5$

$aQ = (5-4,5) - (4,5-3,5) = -0,5$

asimmetria negativa

Soluzione

x_i	n_i	$x_i * n_i$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 * n_i$	$(x_i - \mu)^3$	$(x_i - \mu)^3 * n_i$
1	1	1	-3,25	10,56	10,56	-34,33	-34,33
2	1	2	-2,25	5,06	5,06	-11,39	-11,39
3	2	6	-1,25	1,56	3,13	-1,95	-3,91
4	4	16	-0,25	0,06	0,25	-0,02	-0,06
5	5	25	0,75	0,56	2,81	0,42	2,11
6	3	18	1,75	3,06	9,19	5,36	16,08
	$\Sigma 16$	$\Sigma 68$		20,875	$\sigma^2 = 1,9375$		$M3 = -1,97$ $(31,50/16)$
					$\sigma = 1,39$		$\beta = -0,73 (-1,97/1,39^3)$