Esercitazione 2

CdL in Scienze Politiche

Statistica Sociale

1 Esercizio 1

L'amministrazione regionale ha commissionato un'indagine sulla soddisfazione dei propri utenti dei mezzi di trasporto pubblico. La tabella mostra il numero di utenti rispetto al mezzo di trasporto utilizzato (treno/pullman) e al livello di soddisfazione (Soddisfatto/Non Soddisfatto):

	Soddisfatto	Non Soddisfatto	Totale
Treno	200	100	300
Pullman	50	150	200
Totale	250	250	500

Calcola:

1. Calcola la probabilità di osservare un utente che utilizza il treno.
2. Calcola la probabilità che un utente sia soddisfatto dato che utilizza il treno.
3. Calcola la probabilità che un utente che utilizza il treno sia soddisfatto.
4. Calcola la probabilità che un utente che utilizza il treno sia soddisfatto *o* che un utente che utilizza il pullman sia non soddisfatto.

1.1 Soluzioni

1. 0.6
2. 0.667
3. 0.4
4. 0.7

2 Esercizio 2

I tempi dedicati al tragitto casa-università degli studenti si distribuiscono nella popolazione normalmente con media pari a 81 minuti e deviazione standard pari a 16 minuti. Prendendo in considerazione un'indagine che ha raccolto i dati su un campione di 43 studenti:

1. Calcola la probabilità di osservare un tempo di percorrenza di uno studente che ricade in un intorno di 0.75 deviazioni standard dal valore medio della popolazione.

2.1 Soluzione

Calcolo estremi intervallo:

$$y_1 = \mu - \sigma \times 0.75 = 69 \tag{1}$$

$$y_2 = \mu + \sigma \times 0.75 = 93 \tag{2}$$

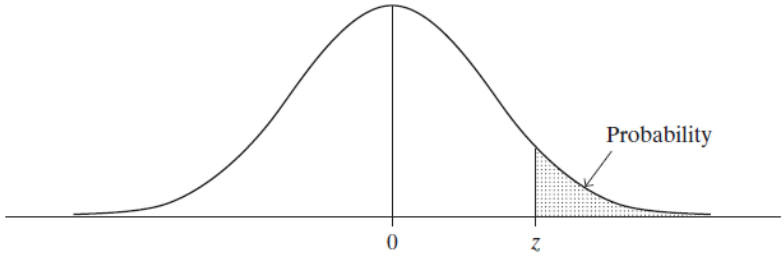
Calcolo gli z-score (esempio di calcolo inverso zscores, non è necessario per l'esercizio):

$$z_2 = \frac{y_2 - \mu}{\sigma} = \frac{93 - 81}{16} = 0.75 \tag{3}$$

$$z_1 = \frac{y_1 - \mu}{\sigma} = \frac{69 - 81}{16} = -0.75 \tag{4}$$

Uso la tavola della distribuzione normale standardizzata in Figura .

TABLE A: Normal curve tail probabilities. Standard normal probability in right-hand tail (for negative values of z, probabilities are found by symmetry).



	Second Decimal Place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Figura 1: Estratto della tavola della distribuzione normale standardizzata

Dalla Figura 1 possiamo vedere che il valore 0.75 è associato a una probabilità di 0.2266. Quindi la probabilità di osservare un valore maggiore a 0.75 *z-scores* o a 93 minuti è pari a 0.2266. Dato che la distribuzione è simmetrica sappiamo che anche la probabilità di osservare un valoreinferiore a -0.75 (quindi a 69 minuti) è 0.2266.

Quindi, se definiamo $P(A)$ come la probabilità di osservare un valore di y **compreso tra** y_1 e y_2 possiamo calcolare:

$$P(A) = P(y_1 < y < y_2) = 1 - (P(y < y_1) + P(y > y_2)) = 1 - (0.2266 + 0.2266) = 0.547 \tag{5}$$

3 Esercizio 3

Una nota catena di negozi ha commissionato un'indagine sulla soddisfazione dei propri clienti nei diversi punti vendita. I punteggi ottenuti dai diversi punti vendita si distribuiscono normalmente con media pari a 205 e deviazione standard pari a 20. Considerando un campione casuale di 168 osservazioni:

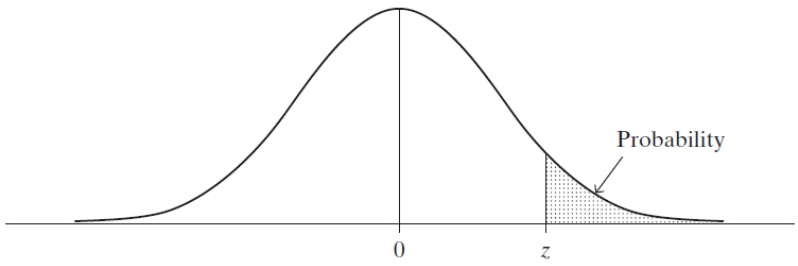
1. Determina l'estremo inferiore di un intervallo centrato intorno al valore medio in cui ti aspetti il 99% dei valori della distribuzione.

2. Determina l'estremo superiore di un intervallo centrato intorno al valore medio in cui ti aspetti il 99% dei valori della distribuzione.

3. Determina la probabilità di osservare un valore superiore a 233.

3.1 Soluzione

Utilizzo l'estratto della tavola in Figura . Dato che l'intervallo deve contenere il 99% della distribuzione allora occorre trovare uno *z-score* che corrisponde a una probabilità sottesa alla coda destra di 0.005.



	Second Decimal Place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									

Figura 2: Estratto della tavola della distribuzione normale standardizzata

Dalla tavola possiamo vedere che il valore che si avvicina di più a 0.005 è 2.58. Quindi, lo *z-score* che corrisponde a un'area sottesa alla coda destra pari a 0.005 è 2.58. Da questo valore possiamo ottenere l'estremo superiore dell'intervallo:

$$y_2 = \mu + z \times \sigma = 205 + 2.58 \times 20 = 256.6 \tag{6}$$

Dato che la distribuzione è simmetrica posso calcolare l'estremo inferiore dell'intervallo:

$$y_1 = \mu - z \times \sigma = 205 - 2.58 \times 20 = 153.4 \tag{7}$$

Per calcolare la **probabilità di osservare un valore superiore a 233** occorre calcolare lo *z-score* corrispondente. In questo caso, 233 corrisponde a uno *z-score* pari a 1.4. Dalla Figura

2 possiamo vedere che 1.4 corrisponde a una probabilità sottesa alla coda destra pari a 0.0808. Quindi la probabilità di osservare un valore maggiore di 233 è pari a 0.0808.

4 Esercizio 4

Il prezzo di uno smartphone di fascia alta venduto nei negozi italiani si distribuisce secondo una normale con media pari a 1770 Euro e deviazione standard pari a 177 Euro. Considerando un'indagine su un campione casuale di 124 osservazioni:

1. Determina la probabilità che la media campionaria sia compresa nell'intervallo tra 1752 Euro e 1790 Euro.

2. Determina l'errore standard della distribuzione della media campionaria se il campione includesse 496 osservazioni.

4.1 Soluzione

Traduco in *z-score* i due estremi dell'intervallo. Dato che stiamo considerando la media campionaria la formula per lo *z-score* è:

$$z_1 = \frac{\bar{y}_1 - \mu}{\sigma_{\bar{y}}} = \frac{1752 - 1770}{177/\sqrt{124}} = -1.13 \tag{8}$$

Applicando questa formula per l'estremo superiore otteniamo 1.26.

Dato che la distribuzione è simmetrica, è possibile calcolare la probabilità corrispondente a -1.13 cercando quella corrispondente a 1.13. Dalla Figura 2 abbiamo che questo valore è 0.1292. Il valore della probabilità nella coda destra corrispondente a $z^2 = 1.26$, invece, è 0.1038.

Quindi la probabilità che la media campionaria $\bar{y}$ sia compresa tra 1760 Euro e 1790 Euro è pari a:

$$P(1760 < \bar{y} < 1790) = 1 - (P(\bar{y} < 1760) + P(\bar{y} > 1790)) = 1 - (0.1292 + 0.1038) = 0.767 \tag{9}$$

Se la dimensione campionaria fosse pari a 496 osservazioni allora l'**errore standard** sarebbe pari a:

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} = \frac{177}{\sqrt{496}} = 7.95 \tag{10}$$