STATISTICA

Il test di bontà d’adattamento

Esercizio 1 (prosecuzione dalla lezione precedente)

Per il lancio di una nuova pubblicità di una grande azienda d'abbigliamento viene estratto un campione casuale di 1000 unità e viene chiesto di scegliere tra 3 nuovi manifesti stradali (A, B e C) quello preferito. Nella tabella sottostante le risposte sono distinte per fasce d'età:

Età	A	B	C
$18 \dashv 25$	60	10	90
$25 \dashv 35$	115	60	95
$35 \dashv 45$	120	125	75
$45 \dashv 60$	120	90	40

c'è abbastanza evidenza nei dati per sostenere che **la preferenza sui manifesti dipenda dall'età**?

Esercizio 1

Età	A	B	C	$n_{i\cdot}$
$18 \dashv 25$	60	10	90	**160**
$25 \dashv 35$	115	60	95	**270**
$35 \dashv 45$	120	125	75	**320**
$45 \dashv 60$	120	90	40	**250**
$n_{\cdot j}$	**415**	**285**	**300**	**1000**

H_0 : indipend. $(\chi^2 = 0)$

H_1 : associaz. $(\chi^2 > 0)$

$$\chi^2 = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{\left(n_{ij} - n^*{}_{ij}\right)^2}{n^*{}_{ij}}$$

$n^*{}_{ij}$

A	B	C
66.40	45.60	48.00
112.05	76.95	81.00
132.80	91.20	96.00
103.75	71.25	75.00

Esercizio 1

Età	A	B	C	$n_{i\cdot}$
18 ⊣ 25	60	10	90	**160**
25 ⊣ 35	115	60	95	**270**
35 ⊣ 45	120	125	75	**320**
45 ⊣ 60	120	90	40	**250**
$n_{\cdot j}$	**415**	**285**	**300**	**1000**

H_0 : indipendenza

H_1 : associazione

$$\chi^2 = 113.56$$

$$\chi^2(3 \times 2)_{1-\alpha}$$

C'E' UNA FORTE EVIDENZA **CONTRO** L'IPOTESI DI INDIPENDENZA

75.00

$$\chi^2(3 \times 2)_{0.0005} = 24.1028$$

P-valore < 0.0005

Esercizio 1

Età	A	B	C	$n_{i\cdot}$
$18 \dashv 25$	60	**10**	**90**	160
$25 \dashv 35$	115	60	95	270
$35 \dashv 45$	120	**125**	75	320
$45 \dashv 60$	120	90	**40**	250
$n_{\cdot j}$	415	285	300	1000

H_0 : indipendenza

H_1 : **associazione**

A	B	C
66.40	**45.60**	**48.00**
112.05	76.95	81.00
132.80	**91.20**	96.00
103.75	71.25	**75.00**

P-valore < 0.0005

Esercizio 1

Età	A	B	C	$n_{i\cdot}$
$18 \dashv 25$	60	**10**	**90**	160
$25 \dashv 35$	115	60	95	270
$35 \dashv 45$	120	**125**	75	320
$45 \dashv 60$	120	90	**40**	250
$n_{\cdot j}$	415	285	300	1000

H_0 : indipendenza

H_1 : **associazione**

A	B	C
66.40	**45.60**	**48.00**
112.05	76.95	81.00
132.80	**91.20**	96.00
103.75	71.25	**75.00**

P-valore < 0.0005

Esercizio 1

Età	A	B	C	$n_{i\cdot}$
18 ⊣ 25	60	**10**	**90**	160
25 ⊣ 35	115	60	95	270
35 ⊣ 45	120	**125**	75	320
45 ⊣ 60	120	90	**40**	250
$n_{\cdot j}$	415	285	300	1000

H_0 : indipendenza

H_1 : **associazione**

A	B	C
66.40	**45.60**	**48.00**
112.05	76.95	81.00
132.80	**91.20**	96.00
103.75	71.25	**75.00**

P-valore < 0.0005

Esercizio 1

Età	A	B	C	$n_{i\cdot}$
18 ⊣ 25	60	10	90	160
25 ⊣ 35	115	60	95	270
35 ⊣ 45	120	125	75	320
45 ⊣ 60	120	90	40	250
$n_{\cdot j}$	415	285	300	1000

H_0 : indipendenza

H_1 : associazione

$$\chi^2 = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{\left(n_{ij} - n^*{}_{ij}\right)^2}{n^*{}_{ij}}$$

A	B	C
66.40	45.60	48.00
112.05	76.95	81.00
132.80	91.20	96.00
103.75	71.25	75.00

$$\chi^2 = 113.56$$

$$\tilde{\chi}^2 = \frac{\chi^2}{n \times \min(h-1, k-1)} = \frac{113.56}{1000 \times 2}$$

$$\tilde{\chi}^2 = 0.057$$

Esercizio 5

L'apnea ostruttiva del sonno (OSA) è una condizione comune in cui ci sono limitazioni parziali intermittenti (cioè, ipopnee) e complete (cioè apnea) nella respirazione durante il sonno. Poiché la polisonnografia, il test standard per la diagnosi, è costosa sono stati sviluppati questionari per identificare le persone con OSA, tra cui il questionario di Berlino (BQ). Un campione casuale di 733 individui è stato classificato secondo la patologia ed il BQ:

	OSA	Non OSA
BQ=low risk	79	294
BQ=high risk	73	287

Sottoporre a verifica l'ipotesi che la patologia e il BQ siano indipendenti al livello di significatività del 2.5%.

Esercizio 5

L'apnea ostruttiva del sonno (OSA) è una condizione comune in cui ci sono limitazioni parziali intermittenti (cioè, ipopnee) e complete (cioè apnea) nella respirazione durante il sonno. Poiché la polisonnografia, il test standard per la diagnosi, è costosa sono stati sviluppati questionari per identificare le persone con OSA, tra cui il questionario di Berlino (BQ). Un campione casuale di 733 individui è stato classificato secondo la patologia ed il BQ:

	OSA	Non OSA	tot
BQ=low risk	79	294	**373**
BQ=high risk	73	287	**360**
tot	**152**	**581**	**733**

$$n^*_{ij} : \frac{373 \times 152}{733} = \mathbf{77.3}; \quad \frac{373 \times 581}{733} = \mathbf{295.7}$$

$$\frac{360 \times 152}{733} = \mathbf{74.7}; \quad \frac{360 \times 581}{733} = \mathbf{285.3}$$

Esercizio 5

L'apnea ostruttiva del sonno (OSA) è una condizione comune in cui ci sono limitazioni parziali intermittenti (cioè, ipopnee) e complete (cioè apnea) nella respirazione durante il sonno. Poiché la polisonnografia, il test standard per la diagnosi, è costosa sono stati sviluppati questionari per identificare le persone con OSA, tra cui il questionario di Berlino (BQ). Un campione casuale di 733 individui è stato classificato secondo la patologia ed il BQ:

	OSA	Non OSA	tot
BQ=low risk	79 / **77.3**	294 / **295.7**	373
BQ=high risk	73 / **74.7**	287 / **285.3**	360
tot	152	581	733

$$\chi^2 = \frac{(79-77.3)^2}{77.3} + \frac{(294-295.7)^2}{295.7} + \frac{(73-74.7)^2}{74.7} + \frac{(287-285.3)^2}{285.3} = 0.096$$

Esercizio 5

L'apnea ostruttiva del sonno (OSA) è una condizione comune in cui ci sono limitazioni parziali intermittenti (cioè, ipopnee) e complete (cioè apnea) nella respirazione durante il sonno. Poiché la polisonnografia, il test standard per la diagnosi, è costosa sono stati sviluppati questionari per identificare le persone con OSA, tra cui il questionario di Berlino (BQ). Un campione casuale di 733 individui è stato classificato secondo la patologia ed il BQ:

	OSA	Non OSA	tot
BQ=low risk	79 / **77.3**	294 / **295.7**	373
BQ=high risk	73 / **74.7**	287 / **285.3**	360
tot	152	581	733

$$\chi^2 = \frac{(79-77.3)^2}{77.3} + \frac{(294-295.7)^2}{295.7} + \frac{(73-74.7)^2}{74.7} + \frac{(287-285.3)^2}{285.3} = 0.096$$

$\boxed{\alpha = 0.025}$

$\Rightarrow$ $\chi(1)^2{}_{0.975} = 5.02389$ **non** possiamo rifiutare l'ipotesi di indipendenza

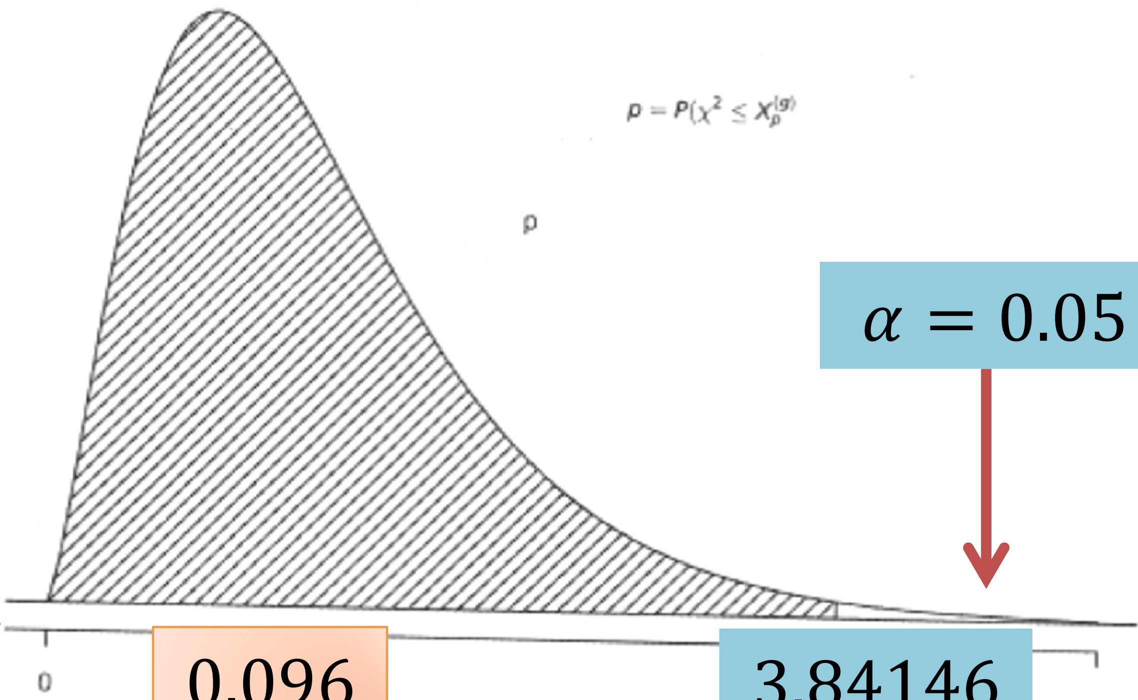
Esercizio 5

L'apnea ostruttiva del sonno (OSA) è una condizione comune in cui ci sono limitazioni parziali intermittenti (cioè, ipopnee) e complete (cioè apnea) nella respirazione durante il sonno. Poiché la polisonnografia, il test standard per la diagnosi, è costosa sono stati sviluppati questionari per identificare le persone con OSA, tra cui il questionario di Berlino (BQ). Un campione casuale di 733 individui è stato classificato secondo la patologia ed il BQ:

	OSA	Non OSA	tot
BQ=low risk	79 / **77.3**	294 / **295.7**	373
BQ=high risk	73 / **74.7**	287 / **285.3**	360
tot	152	581	733

$$\chi^2 = \frac{(79-77.3)^2}{77.3} + \frac{(294-295.7)^2}{295.7} + \frac{(73-74.7)^2}{74.7} + \frac{(287-285.3)^2}{285.3} = 0.096$$

$$\boldsymbol{\alpha = 0.05 \Rightarrow ???}$$

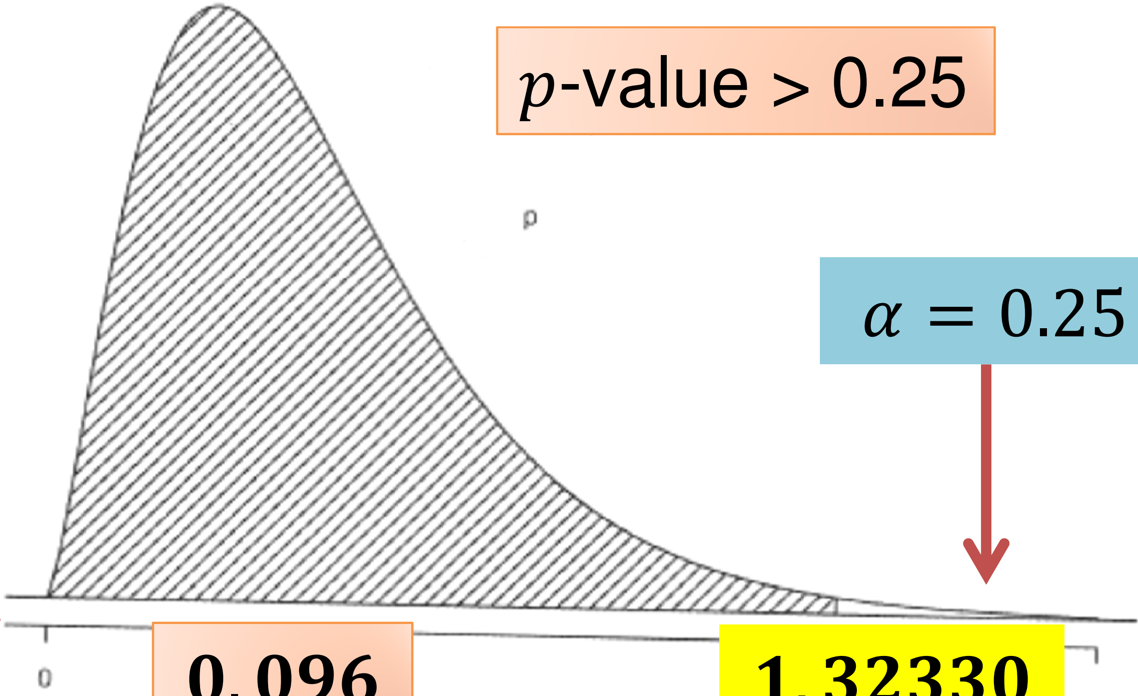


$\nu = 1$, g.d.l.

$1 - \alpha$

p / n	0.750	0.900	0.950	0.975	0.990	0.995	0.9995
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	12.11567
2	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663	15.20180
3	4.10834	6.25139	7.81473	9.34840	11.34487	12.83816	17.73000
4	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026	19.99735
5	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960	22.10533
6	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758	24.10280
7	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774	26.01777
8	10.21885	13.36157	[illegible]	[illegible]	[illegible]	[illegible]	

p-value > 0.25



$\alpha = 0.25$

$\nu = 1$, g.d.l.

$1 - \alpha$

$\mathbf{0.096}$

$\mathbf{1.32330}$

p / n	0.750	0.900	0.950	0.975	0.990	0.995	0.9995
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	12.11567
2	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663	15.20180
3	4.10834	6.25139	7.81473	9.34840	11.34487	12.83816	17.73000
4	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026	19.99735
5	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960	22.10533
6	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758	24.10280
7	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774	26.01777
8	10.21885	13.36157	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Esercizio 5

L'apnea ostruttiva del sonno (OSA) è una condizione comune in cui ci sono limitazioni parziali intermittenti (cioè, ipopnee) e complete (cioè apnea) nella respirazione durante il sonno. Poiché la polisonnografia, il test standard per la diagnosi, è costosa sono stati sviluppati questionari per identificare le persone con OSA, tra cui il questionario di Berlino (BQ). Un campione casuale di 733 individui è stato classificato secondo la patologia ed il BQ:

	OSA	Non OSA	tot
BQ=low risk	79 / **77.3**	294 / **295.7**	373
BQ=high risk	73 / **74.7**	287 / **285.3**	360
tot	152	581	733

$$\chi^2 = \frac{(79-77.3)^2}{77.3} + \frac{(294-295.7)^2}{295.7} + \frac{(73-74.7)^2}{74.7} + \frac{(287-285.3)^2}{285.3} = 0.096$$

$$\boldsymbol{p - \text{value} > 0.25}$$

non possiamo rifiutare l'ipotesi di indipendenza a **nessun livello!**

Eserci

Nel nostro campione il BQ
non è utile per fare previsione
sulla presenza della sindrome

L'apnea os ... o in cui ci sono limitazioni parziali ... e complete (cioè apnea) nella respirazione durante il sonno. Po ... hé la polisonnografia, il test standard per la diagnosi, è costosa **sono stati sviluppati questionari per identificare le persone con OSA, tra cui il questionario di Berlino (BQ).** Un campione casuale di 733 individui è stato classificato secondo la patologia ed il BQ:

	OSA	Non OSA	tot
BQ=low risk	79 / **77.3**	294 / **295.7**	373
BQ=high risk	73 / **74.7**	287 / **285.3**	360
tot	152	581	733

$$\chi^2 = \frac{(79-77.3)^2}{77.3} + \frac{(294-295.7)^2}{295.7} + \frac{(73-74.7)^2}{74.7} + \frac{(287-285.3)^2}{285.3} = 0.096$$

$$\boldsymbol{p - \text{value} > 0.25}$$

non possiamo rifiutare l'ipotesi di indipendenza a **nessun livello!**

Esercizio 6

Un'azienda vuole verificare se l'assistenza fornita via web ai proprio clienti sia omogenea lungo l'intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all'intervento:

	Tempo all'intervento		
Fascia oraria	< 5 min.	5-10 min	> 10 min
8-12	80	78	5
12-16	75	40	85
16-20	0	32	105

a) Calcolare le distribuzioni marginali e il tempo medio di intervento;
b) Guardando alle informazioni contenute nella tabella possiamo trarre un'indicazione riguardo all'omogeneità del servizio?
c) Verificare con un opportuno test statistico quanto indicato al punto b).

Esercizio 6

Un’azienda vuole verificare se l’assistenza fornita via web ai proprio clienti sia omogenea lungo l’intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all’intervento:

	Tempo all’intervento			
Fascia oraria	< 5 min.	5-10 min	> 10 min	**tot**
8-12	80	78	5	**163**
12-16	75	40	85	**200**
16-20	0	32	105	**137**
tot	**155**	**150**	**195**	**500**

a) Calcolare le distribuzioni marginali e il tempo medio di intervento;

Esercizio 6

Un'azienda vuole verificare se l'assistenza fornita via web ai proprio clienti sia omogenea lungo l'intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all'intervento:

	Tempo all'intervento			
Fascia oraria	< 5 min.	5-10 min	> 10 min	
8-12	80	78	5	**163**
12-16	75	40	85	**200**
16-20	0	32	105	**137**
tot	**155**	**150**	**195**	**500**

10-20 min

a) Calcolare le distribuzioni marginali **e il tempo medio di intervento**;

Esercizio 6

Un’azienda vuole verificare se l’assistenza fornita via web ai proprio clienti sia omogenea lungo l’intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all’intervento:

	Tempo all’intervento			
Fascia oraria	< 5 min.	5-10 min	10-20 min	tot
tot	**155**	**150**	**195**	**500**

$$\bar{x}_n = \sum_i x^*{}_i f_i = 2.5 \times \frac{155}{500} + 7.5 \times \frac{150}{500} + 15 \times \frac{195}{500} = 8.875$$

a) Calcolare le distribuzioni marginali **e il tempo medio di intervento**;

Esercizio 6

Un’azienda vuole verificare se l’assistenza fornita via web ai proprio clienti sia omogenea lungo l’intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all’intervento:

	Tempo all’intervento		
Fascia oraria	< 5 min.	5-10 min	> 10 min
8-12	80	78	**5**
12-16	75	40	85
16-20	**0**	32	105

b) Guardando alle informazioni contenute nella tabella possiamo trarre un’indicazione riguardo all’omogeneità del servizio?

Esercizio 6

Un'azienda vuole verificare se l'assistenza fornita via web ai proprio clienti sia omogenea lungo l'intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all'intervento:

	Tempo all'intervento		
Fascia oraria	< 5 min.	5-10 min	> 10 min
8-12	80	78	**5**
12-16	75	40	85
16-20	**0**	32	105

c) Verificare con un opportuno test statistico quanto indicato al punto b).

H_0 : indipendenza (i.e., omogeneità) *vs* H_1 : associazione

Esercizio 6

Un'azienda vuole verificare se l'assistenza fornita via web ai proprio clienti sia omogenea lungo l'intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all'intervento:

	Tempo all'intervento			
Fascia oraria	< 5 min.	5-10 min	> 10 min	**tot**
8-12	80 **50.53**	78 **48.90**	5 **63.57**	**163**
12-16	75 **62**	40 **60**	85 **78**	**200**
16-20	0 **42.47**	32 **41.10**	105 **53.43**	**137**
tot	**155**	**150**	**195**	**500**

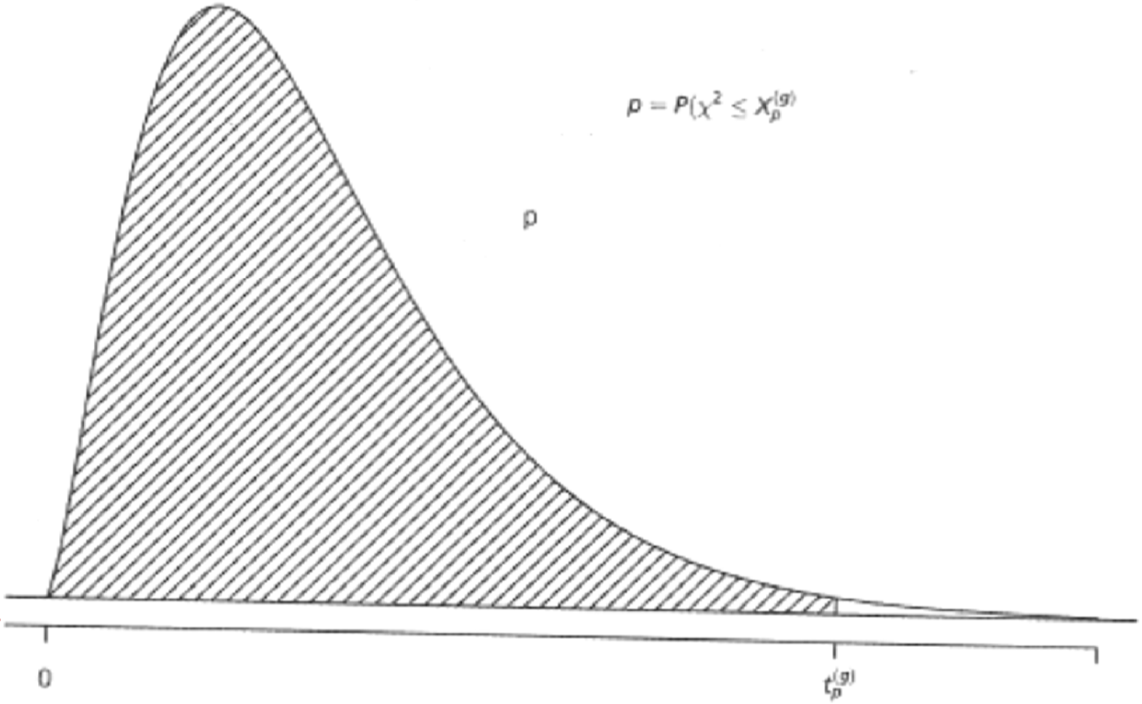
Esercizio 6

Un'azienda vuole verificare se l'assistenza fornita via web ai proprio clienti sia omogenea lungo l'intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all'intervento:

$$\chi^2 = 192.75$$

$$\chi^2(4)_{1-\alpha}$$

	Tempo all'intervento			
Fascia oraria	< 5 min.	5-10 min	> 10 min	**tot**
8-12	80 **50.53**	78 **48.90**	5 **63.57**	**163**
12-16	75 **62**	40 **60**	85 **78**	**200**
16-20	0 **42.47**	32 **41.10**	105 **53.43**	**137**
tot	**155**	**150**	**195**	**500**



$\nu = 1$, g.d.l.

$1 - \alpha$

p / n	0.750	0.900	0.950	0.975	0.990	0.995	0.9995
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	12.11567
2	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663	15.20180
3	4.10834	6.25139	7.81473	9.34840	11.34487	12.83816	17.73000
4	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026	19.99735
5	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960	22.10533
6	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758	24.10280
7	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774	26.01777
8	10.21885	13.36157	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Esercizio 6

Un'azienda vuole verificare se l'assistenza fornita via web ai proprio clienti sia omogenea lungo l'intera giornata (8-20). Per questo seleziona un campione di 500 richieste e le classifica per fascia oraria di ricevimento e tempo fino all'intervento:

$$\chi^2 = 192.75$$

$$\chi^2(4)_{1-\alpha}$$

	Tempo all'intervento			
Fascia oraria	< 5 min.	5-10 min	> 10 min	**tot**
8-12	80 **50.53**	78 **48.90**	5 **63.57**	**163**
12-16	75 **62**	40 **60**	85 **78**	**200**
16-20	0 **42.47**	32 **41.10**	105 **53.43**	**137**
tot	**155**	**150**	**195**	**500**

rifiutiamo l'ipotesi nulla a qualunque livello di significatività => l'assistenza non è omogenea

Un altro uso del test del chi-quadrato

Il **test** del chi-quadrato, II

Lanciamo 120 volte un dado ottenendo la distribuzione di frequenza seguente:

Esito	**1**	**2**	**3**	**4**	**5**	**6**
Fr. $\boldsymbol{n_j}$	20	30	20	25	15	10

Sottoporre a verifica l'ipotesi che il dado sia equilibrato al livello dell'1%

Il **test** del chi-quadrato, II

Esito	1	2	3	4	5	6
Fr. n_j	20	30	20	25	15	10

Se il dado è equilibrato in 120 lanci, ci aspettiamo:

p_j	1/6	1/6	1/6	1/6	1/6	1/6
Fr. att. n^*_j	**20**	**20**	**20**	**20**	**20**	**20**

$$n^*_j = np_j$$

Il **test** del chi-quadrato, II

Esito	1	2	3	4	5	6
Fr. n_j	20	30	20	25	15	10

Se il dado è equilibrato in 120 lanci, ci aspettiamo:

p_j	1/6	1/6	1/6	1/6	1/6	1/6
Fr. att. $n^*{}_j$	20	20	20	20	20	20

$$n^*{}_j = np_j$$

Rifiutiamo al livello α l'ipotesi che il dado sia equilibrato se

$$\sum_{i=1}^{c} \frac{(n^*{}_j - n_j)^2}{n^*{}_j} > \chi^2(c-1)_{1-\alpha}$$

Il **test** del chi-quadrato, II

Esito	**1**	**2**	**3**	**4**	**5**	**6**
Fr. $\boldsymbol{n_j}$	20	30	20	25	15	10

Se il dado è equilibrato in 120 lanci, ci aspettiamo:

$\boldsymbol{p_j}$	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**
Fr. att. $\boldsymbol{n^*{}_j}$	**20**	**20**	**20**	**20**	**20**	**20**

Rifiutiamo al livello α l'ipotesi che il dado sia equilibrato se

H_0 : il dado è equilibrato, H_1 : il dado **non** è equilibrato

Il **test** del chi-quadrato, II

Esito	1	2	3	4	5	6
Fr. n_j	20	30	20	25	15	10

Se il dado è equilibrato in 120 lanci, ci aspettiamo:

p_j	1/6	1/6	1/6	1/6	1/6	1/6
Fr. att. n^*_j	**20**	**20**	**20**	**20**	**20**	**20**

$$\frac{(20-20)^2}{20}+\frac{(20-30)^2}{20}+\frac{(20-20)^2}{20}+$$

$$+\frac{(20-25)^2}{20}+\frac{(20-15)^2}{20}+\frac{(20-10)^2}{20}=12.5$$

$$>\chi^2(\mathbf{5})_{0.99}=15.08627$$

Il **test** del chi-quadrato, II

Esito	**1**	**2**	**3**	**4**	**5**	**6**
Fr. $\boldsymbol{n_j}$	20	30	20	25	15	10

Se il dado è equilibrato in 120 lanci, ci aspettiamo:

$\boldsymbol{p_j}$	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**
Fr. att. $\boldsymbol{n^*{}_j}$	**20**	**20**	**20**	**20**	**20**	**20**

$$\frac{(20-20)^2}{20}+\frac{(20-30)^2}{20}+\frac{(20-20)^2}{20}+$$

$$+\frac{(20-25)^2}{20}+\frac{(20-15)^2}{20}+\frac{(20-10)^2}{20}=12.5$$

$$>\chi^2(\mathbf{5})_{0.99}=15.08627$$

Non possiamo rifiutare l'ip., nulla che il dado sia equilibrato, all'1% di significatività

Il **test** del chi-quadrato, II

Campione *casuale* $X_1, \dots, X_n$ i.i.d

H_0 : $X_i \sim F$, discreta $\qquad$ H_1 : un′altra qualunque, $\neq F$

Si rifiuta l'ipotesi nulla se:

$$\boxed{\sum_{i=1}^{c} \frac{({n^*}_j - n_j)^2}{{n^*}_j} > \chi^2(c-1)_\alpha}$$

Test di buon adattamento (*goodness of fit test*)

Esercizio 1

In una certa regione, l’84% degli automobilisti in un anno non ha nessun incidente, il 14% ne ha uno solo, il 2% ne ha due o più. Su un campione casuale di 400 avvocati, 308 non hanno avuto alcun incidente nell’ultimo anno, 66 ne hanno avuto uno solo e 26 ne hanno avuti due o più. Sulla base di questi dati è possibile affermare che gli avvocati non abbiano lo stesso profilo di incidenti del resto della popolazione?

Esercizio 1

In una certa regione, l'84% degli automobilisti in un anno non ha nessun incidente, il 14% ne ha uno solo, il 2% ne ha due o più. Su un campione casuale di 400 avvocati, **308** non hanno avuto alcun incidente nell'ultimo anno, **66** ne hanno avuto uno solo e **26** ne hanno avuti due o più. Sulla base di questi dati è possibile affermare che gli avvocati non abbiano lo stesso profilo di incidenti del resto della popolazione?

Esito	0	1	≥2
Fr. n_j	308	66	26

p_j	0.84	0.14	0.02
Fr. att.	336	56	8

$$n^*_j = 400 \times 0.84;\ 0.14;\ 0.02$$

Esercizio 1

Esito	**0**	**1**	**≥2**
Fr. $\boldsymbol{n_j}$	308	66	26

$\boldsymbol{p_j}$	**0.84**	**0.14**	**0.02**
Fr. $\boldsymbol{n^*{}_j}$	**336**	**56**	**8**

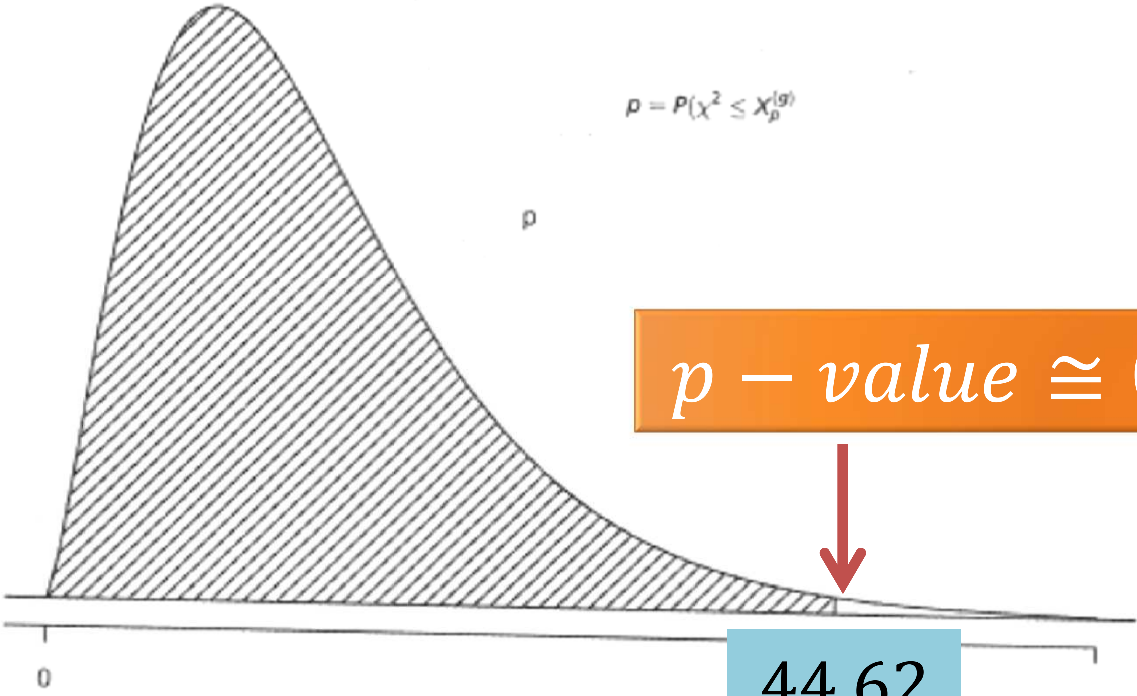
H_0 : $n_j = n^*{}_j$ per ogni j

H_1 : $n_j \neq n^*{}_j$ per almeno un j

$$\sum_{i=1}^{c} \frac{(n_j - n^*{}_j)^2}{n^*{}_j} > \chi^2(c-1)_{1-\alpha}$$

$$\frac{(308-336)^2}{336} + \frac{(66-56)^2}{56} + \frac{(26-8)^2}{8} = 44.62$$

$$\alpha = 1\% \Rightarrow \chi^2(2)_{1-0.01} = 9.21$$



$p - value \cong 0!!$

$\nu = 2$, g.d.l.

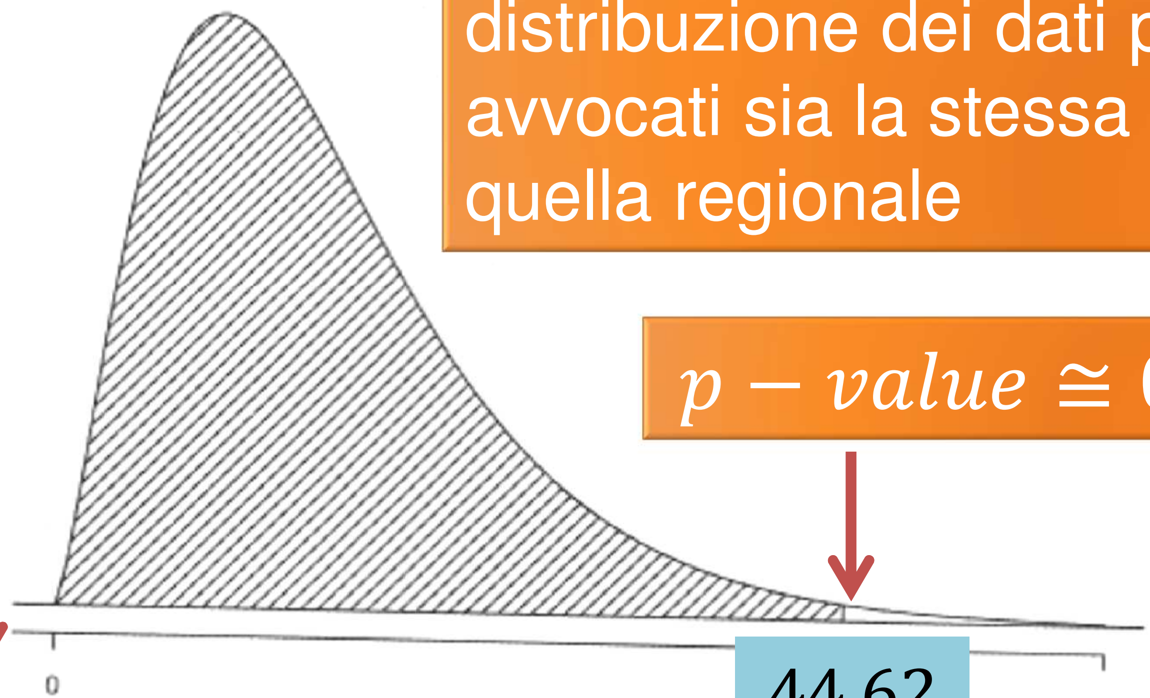
44.62

$1 - \alpha$

p / n	0.750	0.900	0.950	0.975	0.990	0.995	0.9995
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	12.11567
2	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663	15.20180
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026	19.99735
5	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960	22.10533
6	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758	24.10280
7	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774	26.01777
8	10.21885	13.36157	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

C’è un’evidenza molto forte **contro** l’ipotesi che la distribuzione dei dati per gli avvocati sia la stessa di quella regionale

$p - value \cong 0$!!



$\nu = 2$, g.d.l.

44.62

$1 - \alpha$

n \ p	0.750	0.900	0.950	0.975	0.990	0.995	0.9995
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	12.11567
2	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663	15.20180
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026	19.99735
5	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960	22.10533
6	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758	24.10280
7	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774	26.01777
8	10.21885	13.36157	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Esercizio 2

La seguente tabella riporta la ripartizione per fasce d'età delle donne americane non sposate con figli nel 1986 (U.S. National Center for Health Statistics, *Vital Statistics of the United States*):

Fascia d'età	%
≤ 14	1.1
15-|19	32.0
20-|24	36.0
25-|29	18.9
≥ 30	12.0

Da un recente campione di 1000 nati da donne non sposate risulta che 42 delle madri hanno 14 anni o meno, 403 tra i 15 ed i 19 anni, 315 tra 20 e 24 anni, 150 tra 25 e 29 anni e 90 ne hanno almeno 30. Si può sostenere che la distribuzione sia variata?

Esercizio 2

La seguente tabella riporta la ripartizione per fasce d'età delle donne americane non sposate con figli nel 1986 (U.S. National Center for Health Statistics, *Vital Statistics of the United States*):

Fasci a d'età	%	n^*_j	n_i
≤ 14	1.1	11	42
15-|19	32.0	320	403
20-|24	36.0	360	315
25-|29	18.9	189	150
≥ 30	12.0	120	90

Da un recente campione di 1000 nati da donne non sposate risulta …

Esercizio 2

La seguente tabella riporta la ripartizione per fasce d'età delle donne americane non sposate con figli nel 1986 (U.S. National Center for Health Statistics, *Vital Statistics of the United States*):

Fasci a d'età	%	$n^*{}_j$	n_i
≤ 14	1.1	11	42
15-|19	32.0	320	403
20-|24	36.0	360	315
25-|29	18.9	189	150
≥ 30	12.0	120	90

Da un recente campione di 1000 nati da donne non sposate risulta …

$H_0 \ : \ n_j = n^*{}_j$ per ogni j

$H_1 \ : \ n_j \neq n^*{}_j$ per almeno un j

$$\sum_{i=1}^{c} \frac{(n_j - n^*{}_j)^2}{n^*{}_j} > \chi^2(c-1)_{1-\alpha}$$

Esercizio 2

La seguente tabella riporta la ripartizione per fasce d'età delle donne americane non sposate con figli nel 1986 (U.S. National Center for Health Statistics, *Vital Statistics of the United States*):

Fasci a d'età	%	n^*_j	n_i
≤ 14	1.1	**11**	**42**
15-|19	32.0	**320**	**403**
20-|24	36.0	**360**	**315**
25-|29	18.9	**189**	**150**
≥ 30	12.0	**120**	**90**

$$\frac{(42-11)^2}{11} + \frac{(403-320)^2}{320} +$$

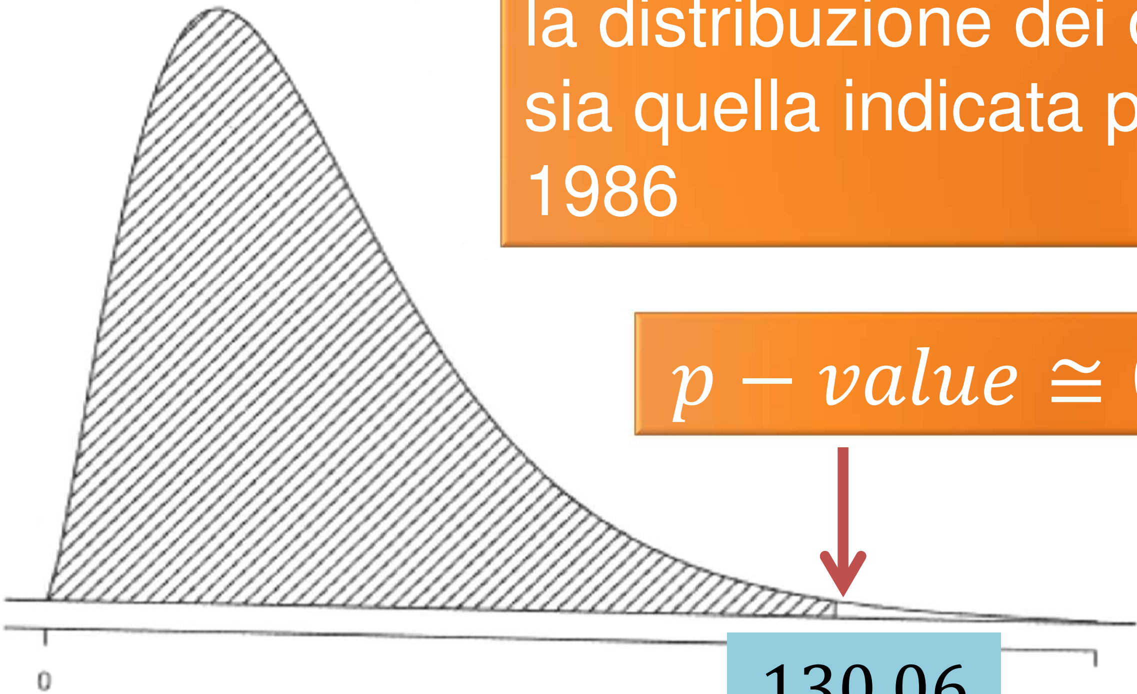
$$\frac{(315-360)^2}{360} + \frac{(150-189)^2}{189}$$

$$+ \frac{(90-120)^2}{120}$$

$$= 130.06$$

C’è un’evidenza molto forte **contro** l’ipotesi che la distribuzione dei dati sia quella indicata per il 1986

$p - value \cong 0$!!



$\nu = 4$, g.d.l.

$1 - \alpha$

130.06

n \ p	0.750	0.900	0.950	0.975	0.990	0.995	0.9995
1	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944	12.11567
2	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663	15.20180
3	4.10834	6.25139	7.81473	9.34840	11.34487	12.83816	17.73000
4	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026	19.99735
5	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960	22.10533
6	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758	24.10280
7	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774	26.01777
8	10.21885	13.36157	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Esercizio 2

La seguente tabella riporta la ripartizione per fasce d'età delle donne americane non sposate con figli nel 1986 (U.S. National Center for Health Statistics, *Vital Statistics of the United States*):

Fasci a d'età	%	n^*_j	n_i
≤ 14	1.1	**11**	**42**
15-|19	32.0	**320**	**403**
20-|24	36.0	**360**	**315**
25-|29	18.9	**189**	**150**
≥ 30	12.0	**120**	**90**

$$\frac{(42-11)^2}{11}+\frac{(403-320)^2}{320}+$$

$$\frac{(315-360)^2}{360}+\frac{(150-189)^2}{189}$$

$$+\frac{(90-120)^2}{120}$$

$$=130.06$$

Esercizio 4, di compito

Da un recente campione di 1000 nati da donne non sposate risulta che 42 delle madri hanno 14 anni o meno, 403 tra i 15 ed i 19 anni, 315 tra 20 e 24 anni, 150 tra 25 e 29 anni e 90 ne hanno almeno 30.

Fascia d'età	n_i
< 15	42
$[15, 20)$	403
$[20, 25)$	315
$[25, 30)$	150
≥ 30	90

a. Rappresentare graficamente la distribuzione di frequenza dell'età al primo parto nel campione.
b. Calcolare l'età media al primo parto, e la varianza nel campione.
c. Calcolare i quartili e disegnare il boxplot