Statistica Matematica A

Esercitatori: Dott.ssa Rosazza - Dott. Zucca

Esercitazione # 7

1.Test χ^2 di adattamento

Esercizio # 1.1

L'azienda di materiale elettronico *Cortocircuito* produce circuiti stampati. Viene avanzata l'ipotesi che la v.a. X che descrive il numero di difetti presentati dai circuiti segua una distribuzione di Poisson. Per verificare tale ipotesi si estrae un campione di 60 circuiti e, per ciascuno di essi, si osserva il numero di difetti che presenta. Il risultato è riassunto nella seguente tabella:

Numero difetti	Frequenza osservata
0	32
1	15
2	9
≥ 3	4

In base a questi dati, si può ritenere al 5% che X segua una legge di Poisson? Com'è definito il p-value del test? Cosa possiamo dire circa il suo valore?

Esercizio # 1.2

Nella sessione di esami 1999-2000 gli studenti hanno ottenuto punteggi finali che sono stati raccolti in quattro categorie **D**, **C**, **B** e **A** rispettivamente con frequenza 40%, 30%, 20% e 10%. In un campione di 2500 studenti che hanno sostenuto sostenuto lo stesso esame nella sessione 2001-2002 si sono osservate le seguenti frequenze:

Punteggio	Frequenza osservata
D	1170
C	585
B	405
A	340

Al livello di significatività dell'1% si può concludere che questo gruppo sia omogeneo all'insieme degli studenti dell'anno precedente?

Esercizio # 1.3

Si suppone che la v.a. X che misura il tempo medio di vita (in mesi) delle lampadine della ditta *Bulbo Incandescente* segua una legge esponenziale. Su un campione di 100 lampadine si sono osservate le seguenti durate

	Frequenza osservata
$X \le 1$	0.39
$1 < X \le 2$	0.24
$2 < X \le 3$	0.12
$3 < X \le 5$	0.16
$5 < X \le 10$	0.09

Si stimi il parametro λ della legge e si appronti un test adatto a valutare l'adattamento della legge trovata.

Esercizio # 1.4

Nel 1856 il biologo Gregor Mendel condusse un esperimento di ibridazione sui piselli. Secondo la teoria sulla trasmissione dei caratteri ereditari da lui stesso proposta, l'esperimento avrebbe dovuto produrre piselli del tipo **RY**, **RG**, **WY** e **WG** rispettivamente con frequenze relative 9/16, 3/16, 3/16 e 1/16. I risultati dell'esperimento furono:

Piselli	RY	RG	WY	WG
Frequenze	315	108	101	32

Questi risultati confermano la teoria di Mendel?

Esercizio # 1.5

Nel paese di *Crazyland* vigono leggi molto peculiari. La premiata ditta di fuochi di artificio *Miccia Corta* afferma che il numero di settimane che trascorrono tra un incidente in fabbrica ed il successivo è a norma di legge (cioè segue una legge esponenziale con media 2). Un ispettore per la sicurezza sul lavoro raccoglie i dati relativi a 100 incidenti e li riassume nella seguente tabella:

Tempo interincidente X	$X < 1$	$1 \le X < 2$	$2 \le X < 3$	$3 \le X < 4$	$4 \le X$
Frequenze	35	19	18	11	17

Può l'ispettore confermare, con un livello di significatività del 5%, l'affermazione di regolarità avanzata dal titolare dell'azienda.

Esercizio # 1.6

Nel bacino del fiume *Fast-and-Furious creek* l'irrigazione comincia il 15 Aprile; un tecnico è interessato alla probabilità di pioggia durante la prima settimana di tale periodo. Dai dati ricavati negli ultimi 100 anni nell'area di interesse, si è ricavata la seguente distribuzione di giorni piovosi per la settimana in questione:

Giorni piovosi	0	1	2	$3,4,5,6,7$	Totale
Frequenza assoluta	57	30	9	4	100

Sulla base dei dati raccolti è plausibile, ad un livello di significatività del 3%, che il numero dei giorni piovosi segua una legge binomiale $\mathcal{B}(7, 0.1)$.

Esercizio # 1.7

Arturo Cecato, altresì noto agli amici come *Killer*, è un giocatore di freccette accanito, ma piuttosto incapace e pericoloso. In una serata tipica (10 partite) il numero di persone colpite dai suoi tiri fuori bersaglio è notevole. La seguente tabella riassume le sue "imprese" relative a 75 serate

Persone colpite	1	2	3	4	5	6	7	8
Frequenza assoluta	1	11	8	13	11	12	10	9

È plausibile assumere che il numero di persone colpite segua una legge di Poisson di parametro 6? In caso negativo, possiamo concludere che la legge non sia una Poisson?

Esercizio # 1.8

Il signor **Ercole Maciste**, titolare della omonima ditta di martinetti idraulici, vuole analizzare la durata del modello "Ercolissimo" in commercio da ben 5 anni. Si analizza un campione di 100 martinetti di cui si è appena rotto l'ultimo ottenendo la seguente tabella che riporta i tempi intercorsi dalla consegna al guasto (calcolati in settimane).

Tempo intercorso X	Frequenza assoluta
$0 \leq X < 300$	55
$300 \leq X < 600$	25
$600 \leq X < 900$	10
$900 \leq X < 1200$	4
$1200 \leq X < 1500$	3
$1500 \leq X < 1800$	2
$1800 \leq X \leq 1825$	1

È possibile affermare che il tempo di vita dell'apparecchio in questione segua una legge esponenziale? Qual è il tempo di vita medio?

Esercizio # 1.9

Un certo protocollo per il controllo qualità classifica i termometri in quattro classi **A**, **B**, **C** e **D**. Sulla base dell'esperienza passata, i termometri della ditta "Fevertime" si distribuiscono tra le categorie secondo la seguente tabella:

Categoria	A	B	C	D
Proporzione	0.87	0.09	0.03	0.01

L'azienda produttrice sottopone un nuovo lotto di 1336 termometri ad un controllo che da il seguente risultato:

Categoria	A	B	C	D
Frequenza assoluta	1188	91	47	10

Si può affermare che il nuovo lotto è conforme agli standard passati?

Esercizio # 1.10

La ditta farmaceutica *Pasticche&Vaccini* sta conducendo uno studio su un ceppo batterico mutante in grado di essere utilizzato per la produzione di nuovi medicinali. La particolarità di questo batterio è che la colonia muta (con il ritmo di una volta ogni 24 ore) le sue caratteristiche. In particolare si suppone che possa assumere 3 "forme" differenti A, B e C con probabilità

Categoria	A	B	C
Probabilità	λ	λ	$1-2\lambda$

dove $\lambda \in (0, 1/2)$. Per verificare tale ipotesi e cercare un valore opportuno per λ si testa la colonia nei primi 24 giorni di vita e si ottiene la seguente tabella

Categoria	A	B	C
Frequenza assoluta	6	8	10

Ci si chiede se esiste un valore di lambda per cui si possa accettare la legge teorica ad un livello di significatività del 5%. Quale valore di λ rende migliore l'adattamento? Cosa si potrebbe concludere nel caso in cui l'ampiezza del campione fosse 2400 così suddivisa

Categoria	A	B	C
Frequenza assoluta	600	800	1000

ad un livello di significatività del 5%?

2.Test di confronto tra due medie e due popolazioni

Esercizio # 2.1

Da un campione di $n_1 := 100$ famiglie di Springfield risulta che il numero medio di figli per famiglia è $\overline{x}_1 := 1.8$, con una deviazione standard pari a $s_1 := 0.6$. La stessa indagine svolta nella vicina Shelbyville rivela un numero medio di figli pari a $\overline{x}_2 := 1.6$ con deviazione standard $s_2 := 0.4$ su un campione di $n_2 := 200$ famiglie. Discutere l'ipotesi nulla

$$H_0 : \mu_1 = \mu_2$$

nei casi

1. varianze supposte note,
2. varianze incognite uguali
3. varianze incognite differenti.

Esercizio # 2.2

In Lombardia negli ultimi 4 inverni sono stati registrati, durante i mesi di Novembre, Dicembre e Gennaio, i seguenti casi di meningite

99/00	13
00/01	15
01/02	20
02/03	18

Supponendo che il numero di abitanti sia rimasto sostanzialmente invariato in questi anni (e pari a 8.940.000) e tenendo come livello di riferimento la media dei primi 3 anni di monitoraggio, si vuole decidere se l'epidemia di quest'anno sia più preoccupante.

1. Si adotti il punto di vista *precauzionale* (non si vuole correre il rischio di sottovalutare l'epidemia): si formuli l'ipotesi nulla adeguata e si discuta la validità dell'ipotesi con un livello di significatività del 5%.

2. Si adotti il punto di vista *non allarmistico* (non si vuole sopravvalutare l'epidemia): si formuli l'ipotesi nulla adeguata, si calcoli il p-value e si discuta la validità dell'ipotesi.

3. Si ripeta il punto (1) utilizzando il confronto tra le medie del 2002 e del 2003 per decidere se quest'anno l' epidemia sia più virulenta.

4. Si studi la possibilità che la media dei primi 3 anni sia uguale a quella dell'ultimo anno supponendo che nell'ultimo anno non siano stati registrati 10 casi (e che quindi in totale siano 28).

3.Test di confronto tra piú medie

Esercizio # 3.1

Per confrontare la resistenza di quattro diverse leghe di metallo (A, B, C e D) sono stati esaminati quattro campioni e su ogni unità è stata misurata la resistenza di rottura. I risultati sono contenuti nella seguente tabella.

A	B	C	D
32	30	24	30
37	31	31	29
34	40	28	28
33	37	32	33
37	38		31
	36		32

Si valuti se le leghe si possono ritenere equivalenti.

Esercizio # 3.2

E' stata studiata la resistenza alla rottura di due tipi di filo di lana. Sappiamo che $\sigma_1 = 5$ e $\sigma_2 = 4$ psi. Per ciascun tipo di filo di lana si è costituito un campione casuale di 20 provini e si è ottenuto, rispettivamente, $\bar{x}_1 = 88$ psi e $\bar{x}_2 = 91$ psi.

1. Usando un intervallo di confidenza al 90% per la differenza delle mediedella resistenza alla rottura, commentare se vi è o no evidenza per affermare che il filo di lana di tipo 2 è più alta.

2. Per fare lo stesso utilizzando un intervallo di confidenza al 98%.

Soluzioni

Soluzione es 1.1:

Stimiamo il parametro λ della legge di Poisson utilizzando la media campionaria che, come è noto è uno stimatore non distorto e consistente; pertanto, essendo l'ampiezza del campione $n = 60$, si ha

$$\lambda = \frac{32 \cdot 0 + 15 \cdot 1 + 9 \cdot 2 + 4 \cdot 3}{60} = 0.75.$$

Per effettuare il test di adattamento si costruiscono i valori teorici di frequenza assoluta della variabile X utilizzando la distribuzione di Poisson $\mathcal{P}(0.75)$ come in tabella (si ricordi che $\mathbb{P}(X = i) = \exp(-\lambda)\frac{\lambda^i}{i!}$):

Numero difetti	Frequenza assoluta osservata	Frequenza assoluta teorica
0	32	28.34
1	15	21.26
≥ 2	13	10.40

Utilizzando la statistica

$$Q = \sum_{i=1}^{N_c} \frac{(n \cdot p_i - F_i)^2}{n \cdot p_i} \approx Y \sim \chi^2(N_c - 1 - r)$$

dove $N_c = 3$ è il numero delle classi, $n \cdot p_i$ sono le frequenze assolute teoriche riassunte nella precedente tabella, F_i sono le frequenze assolute osservate e $r = 1$ è il numero di parametri della distribuzione stimati dai dati e considerando come ipotesi nulla

$$\mathrm{H}_0\text{: le osservazioni provengono dalla distribuzione teorizzata}$$

si ottiene $Q = 2.9659$. Ricordiamo che l'ipotesi nulla viene accettata ad un livello di significatività α se e solo se $Q < \chi^2_\alpha(N_c - 1 - r) = \chi^2_{0.05}(1) = 3.841$. Quindi l'ipotesi nulla viene accettata al 5%. Alternativamente, il p-value del test è $\overline{\alpha} = 1 - F_{\chi^2(1)}(2.9659) = 0.0850$, un valore non molto alto che comporta la forte dipendenza dell'esito del test dalla scelta di α.

Soluzione es 1.2:

Calcoliamo la frequenza attesa per un campione di ampiezza $n = 2500$

Punteggio	Frequenza osservata	Frequenza attesa
D	1170	1000
C	585	750
B	405	500
A	340	250

Utilizzando la statistica

$$Q = \sum_{i=1}^{N_c} \frac{(n \cdot p_i - F_i)^2}{n \cdot p_i} \approx Y \sim \chi^2(N_c - 1)$$

dove $N_c = 4$ è il numero delle classi, $n \cdot p_i$ sono le frequenze assolute attese, F_i sono le frequenze assolute osservate. e considerando come ipotesi nulla

H_0: le osservazioni provengono dalla distribuzione teorizzata

si ottiene $Q = 28.9 + 36.3 + 18.05 + 32.4 = 115.65$. Ricordiamo che l'ipotesi nulla viene accettata ad un livello di significatività α se e solo se $Q < \chi^2_\alpha(N_c - 1) = \chi^2_{0.01}(3) = 11.34$. Quindi l'ipotesi nulla viene rifiutata al livello 1%. Alternativamente, il p-value del test è $\overline{\alpha} = 1 - F_{\chi^2(3)}(115.65) \approx 0$ per cui l'ipotesi nulla è ragionevolmente rifiutabile.

Soluzione es 1.3:

Incominciamo con lo stimare il parametro λ dell'esponenziale. Uno stimatore non distorto e consistente per il parametro λ è $(n-1)/\sum_{i=1}^n x_i$ (per $n > 1$). Utilizzando tale stimatore e approssimando $\sum_{i=1}^n x_i$ con $n \sum_{i=1}^{N_c} \overline{c_j} f_j$ dove f_j è la frequenza relativa della classe j-esima e $\overline{c_j}$ è la media uniforme della classe j-esima si ha

$$\lambda = \frac{0.99}{0.39 \cdot 0.5 + 0.24 \cdot 1.5 + 0.12 \cdot 2.5 + 0.16 \cdot 4 + 0.09 \cdot 7.5} \approx 0.461.$$

Per effettuare il test di adattamento si costruiscono i valori teorici di frequenza utilizzando la distribuzione esponenziale $\exp(0.461)$ come in tabella:

	Frequenza teorica
$X \leq 1$	$1 - \exp(-0.461) \approx 0.3693$
$1 < X \leq 2$	$\exp(-0.461) - \exp(-0.461 \cdot 2) \approx 0.2329$
$2 < X \leq 3$	$\exp(-0.461 \cdot 2) - \exp(-0.461 \cdot 3) \approx 0.1469$
$3 < X \leq 5$	$\exp(-0.461 \cdot 3) - \exp(-0.461 \cdot 5) \approx 0.1511$
$5 < X \leq 10$	$\exp(-0.461 \cdot 5) - \exp(-0.461 \cdot 10) \approx 0.0898$

e si utilizza la seguente statistica

$$Q = n \sum_{i=1}^{N_c} \frac{(p_i - f_i)^2}{p_i} \approx Y \sim \chi^2(N_c - 1 - r)$$

dove $N_c = 5$ è il numero delle classi, p_i sono le frequenze relative teoriche riassunte nella precedente tabella, f_i sono le frequenze relative osservate, r è il numero di parametri della distribuzione stimati dai dati (in questo caso $r = 1$) ed infine $n = 100$ è l'ampiezza del campione. L'ipotesi nulla

$$\mathrm{H}_0\text{: le osservazioni provengono dalla distribuzione teorizzata}$$

viene accettata ad un livello di significatività α se e solo se $Q < \chi^2_\alpha(N_c - 1 - r)$ (dove $\chi^2_\alpha(k)$ è la funzione quantile della distribuzione χ^2 con k gradi di libertà in calcolata in corrispondenza del valore α). Alternativamente, il p-value del test è $\overline{\alpha} = 1 - F_{\chi^2(N_c-1-r)}(Q)$. Utilizzando i dati del problema $Q = 0.6732$ e $\overline{\alpha} = 0.8795$. Il test è quindi accettato al livello di significatività del 10%.

Soluzione es 1.4:

Allo scopo di verificare l'adattamento dei dati reali a quelli teorici calcoliamo le frequenze relative osservate e testiamo l'ipotesi nulla

$$\mathrm{H}_0\text{: i dati provengono dalla legge teorizzata da Mendel.}$$

L'ampiezza del campione è $n = 315 + 108 + 101 + 32 = 556$ quindi

Piselli	RY	RG	WY	WG
Frequenze assolute osservate	315	108	101	32
Frequenze relative osservate	315/556	108/556	101/556	32/556
Frequenze relative teoriche	9/16	3/16	3/16	1/16

da cui

$$Q = n \sum_{i=1}^{N_c} \frac{(p_i - f_i)^2}{p_i} = 0.47.$$

Il p-value del test è pertanto $\overline{\alpha} = 1 - F_{\chi^2(3)}(0.47) \approx 0.9254$, quindi la teoria di Mendel è molto plausibile.

Soluzione es 1.5:

Il lavoro dell'ispettore è quello di svolgere un test di adattamento sugli $n = 100$ dati provenienti dalla variabile X (tempo che intercorre tra un'incidente ed il successivo); l'ipotesi nulla è

$$\mathrm{H}_0\text{: per la legge di } X \text{ vale } \mathcal{L}(X) = \exp(0.5).$$

Ricordando che la funzione di ripartizione della legge $\exp(\lambda)$ è $F(x) = (1 - \exp(-x\lambda))\mathbb{1}_{(0,+\infty)}(x)$ calcoliamo i valori di frequenza assoluta teorici

Tempo interincidente X	$X < 1$	$1 \leq X < 2$	$2 \leq X < 3$	$3 \leq X < 4$	$4 \leq X$
Frequenze assolute osservate	35	19	18	11	17
Frequenze assolute teoriche	39.35	23.87	14.47	8.78	13.53

Utilizziamo la statistica

$$Q = \sum_{i=1}^{N_c} \frac{(n \cdot p_i - F_i)^2}{n \cdot p_i} = 3.7869$$

ed otteniamo il p-value del test è $\overline{\alpha} = 1 - F_{\chi^2(4)}(3.7869) = 0.4356$, pertanto l'ipotesi nulla è accettata ad un livello di significatività del 5% ed la ditta *Miccia Corta* viene considerata in regola.

Alternativamente si confronta $\chi^2_{0.05}(4) = 9.4877$ con Q; essendo $Q < \chi^2_{0.05}(4)$ non possiamo rifiutare l'ipotesi nulla.

Soluzione es 1.6:

Si tratta di testare l'adattamento dei dati ad una legge Binomiale di parametri noti. Osserviamo immediatamente che, se chiamiamo X la variabile aleatoria, allora $\mathbb{P}(X = 0) = 0.4783$, $\mathbb{P}(X = 1) = 0.372$, $\mathbb{P}(X = 2) = 0.124$ e $\mathbb{P}(X \geq 3) = 0.0257$, pertanto

Giorni piovosi	0	1	2	$3, 4, 5, 6, 7$
Frequenza assoluta osservata	57	30	9	4
Frequenza assoluta teorica	47.83	37.2	12.4	2.57

Utilizziamo la statistica

$$Q = \sum_{i=1}^{N_c} \frac{(n \cdot p_i - F_i)^2}{n \cdot p_i} \approx Y \sim \chi^2(N_c - 1 - r)$$

dove $N_c = 4$ è il numero delle classi, $n \cdot p_i$ sono le frequenze assolute teoriche, F_i sono le frequenze assolute osservate, $r = 0$ è il numero di parametri della distribuzione stimati dai dati ed infine $n = 100$ è l'ampiezza del campione considerato. Formuliamo l'ipotesi nulla

H_0: le osservazioni provengono dalla distribuzione teorizzata

ottenendo $Q = 4.8796$ da cui il p-value del test è $\overline{\alpha} = 1 - F_{\chi^2(3)}(4.8796) = 0.1808$; l'ipotesi nulla è accettata ad un livello di significatività del 5%.

Soluzione es 1.7:

Dovendo testare l'adattamento dei dati ad una legge di Poisson di parametro noto procediamo immediatamente con il calcolo dei valori teorici di frequenza assoluta della variabile X utilizzando la distribuzione di Poisson $\mathcal{P}(6)$ (si ricordi che $\mathbb{P}(X = i) = \exp(-\lambda)\frac{\lambda^i}{i!}$); per ottenere una buona stima è necessario accorpare i dati in maniera tale che la frequenza di ogni classe non sia inferiore a 5:

Persone colpite	Frequenza assoluta osservata	Frequenza assoluta teorica
≤ 2	12	4.6477
3	8	6.6926
4	13	10.0389
5	11	12.0467
6	12	12.0467
7	10	10.3258
≥ 8	9	19.2015

Utilizzando la statistica

$$Q = \sum_{i=1}^{N_c} \frac{(n \cdot p_i - F_i)^2}{n \cdot p_i} \approx Y \sim \chi^2(N_c - 1 - r)$$

dove $N_c = 7$ è il numero delle classi, $n \cdot p_i$ sono le frequenze assolute teoriche, F_i sono le frequenze assolute osservate e $r = 0$ è il numero di parametri della distribuzione stimati dai dati ed infine $n = 75$ è l'ampiezza del campione statistico. Formuliamo al solito l'ipotesi nulla

H_0: le osservazioni provengono dalla distribuzione teorizzata.

Con semplici calcoli si ottiene $Q = 18.2809$ pertanto p-value del test è $\overline{\alpha} = 1-F_{\chi^2(6)}(18.2809) = 0.0056$; l'ipotesi nulla può essere ragionevolmente rifiutata.

Naturalmente non possiamo concludere che la legge non sia una Poisson; proviamo a stimare il parametro λ utilizzando la media campionaria

$$\lambda = \frac{1 + 2 \cdot 11 + 3 \cdot 8 + 4 \cdot 13 + 5 \cdot 11 + 6 \cdot 12 + 7 \cdot 10 + 8 \cdot 9}{75} = 4.9067$$

Ripetiamo quindi calcoli analoghi ottenendo

Persone colpite	Frequenza assoluta osservata	Frequenza assoluta teorica
≤ 2	12	9.9550
3	8	10.9226
4	13	13.3985
5	11	13.1485
6	12	10.7526
7	10	7.5371
≥ 8	9	9.2858

da cui $Q = 2.5233$ e $\overline{\alpha} = 1 - F_{\chi^2(5)}(2.5233) \approx 0.773$. Per cui l'adattamento alla legge è buono.

Soluzione es 1.8:

Stimiamo il parametro λ dell'esponenziale utilizzando il campione di $n = 100$ dati a nostra disposizione. Uno stimatore non distorto e consistente per il parametro λ è $(n-1)/\sum_{i=1}^{n} x_i$ (per $n > 1$). Approssimando $\sum_{i=1}^{n} x_i$ con $\sum_{i=1}^{N_c} \overline{c_j} F_j$ dove F_j è la frequenza assoluta della classe j-esima e $\overline{c_j}$ è la media uniforme della classe j-esima si ha

$$\begin{aligned}\lambda &= \frac{99}{55 \cdot 150 + 25 \cdot 450 + 10 \cdot 750 + 4 \cdot 1050 + 3 \cdot 1350 + 2 \cdot 1650 + 1813} \\ &\approx \frac{99}{40363} = 2.453 \cdot 10^{-3}\end{aligned}$$

Per effettuare il test di adattamento si costruiscono i valori teorici di frequenza assoluta utilizzando la distribuzione esponenziale $\exp(2.453 \cdot 10^{-3})$ come in tabella (accorpando opportunamente i dati in modo da avere $n_i \geq 5$):

Tempo intercorso X	Frequenza assoluta dei dati	Frequenza assoluta teorica
$0 \leq X < 300$	55	52.09
$300 \leq X < 600$	25	25
$600 \leq X < 900$	10	11.96
$900 \leq X$	10	11

Utilizziamo ora la seguente statistica

$$Q = \sum_{i=1}^{N_c} \frac{(n \cdot p_i - F_i)^2}{n \cdot p_i} \approx Y \sim \chi^2(N_c - 1 - r)$$

dove $N_c = 4$ è il numero delle classi, $n \cdot p_i$ sono le frequenze assolute teoriche riassunte nella precedente tabella, F_i sono le frequenze assolute osservate e $r = 1$ è il numero di parametri della distribuzione stimati dai dati. L'ipotesi nulla

H_0: le osservazioni provengono dalla distribuzione teorizzata

viene accettata ad un livello di significatività α se e solo se $Q < \chi^2_\alpha(N_c - 1 - r)$ Alternativamente, il p-value del test è $\overline{\alpha} = 1 - F_{\chi^2(N_c-1-r)}(Q)$. Utilizzando i dati del problema $Q = 0.5747$ e $\overline{\alpha} = 0.7502$. Il test è quindi ragionevolmente accettato.

Soluzione es 1.9:

Si consideri la seguente tabella

Categoria	A	B	C	D
Frequenza assoluta na	1188	91	47	10
Frequenza relativa teorica	0.87	0.09	0.03	0.01
Frequenza assoluta teorica nt	1162.32	120.24	40.08	13.36
$(na - nt)^2/nt$	0.5674	7.1106	1.1948	0.845

da cui $Q = 9.718$. Poichè asintoticamente la distribuzione di Q è $\chi^2(3)$ si calcola il p-value $\overline{\alpha} = 1 - F_{\chi^2(3)}(9.718) = 0.0211$, pertanto concludiamo che ragionevolmente il nuovo lotto è difforme dai precedenti.

Soluzione es 1.10:

Per verificare l'adattamento calcoliamo

$$Q = n \sum_{i \in J} \frac{(p_i - f_i)^2}{p_i}$$

e quindi il p-value $\overline{\alpha} = 1 - F_{\chi^2(|J|-1-r)}(Q)$. Per le note proprietà di monotonia delle funzioni di ripartizione, massimizzare il p-value equivale a minimizzare Q. Supponendo che $p_i = p_i(\underline{\lambda})$ (dove $\underline{\lambda} \in \Omega \subset \mathbb{R}^n$ è un vettore di parametri) siano funzioni derivabili e strattamente positive, allora il minimo di Q nell'aperto Ω se esiste deve essere soluzione di $\mathrm{J}Q = 0$ dove, dal teorema di derivazione della funzione composta,

$$\mathrm{J}Q(\underline{\lambda}) = n \sum_{i \in J} \left(1 - \frac{f_i^2}{p_i^2(\underline{\lambda})}\right) \mathrm{J}p_i(\underline{\lambda}).$$

Nel nostro caso, posto $f_1 = 6/24 = 0.25$, $f_2 = 8/24 = 1/3$ e $f_3 = 10/24 = 5/12$, si tratta di risolvere l'equazione

$$1 - \frac{f_1^2}{\lambda^2} + 1 - \frac{f_2^2}{\lambda^2} - 2(1 - \frac{f_3^2}{(1-2\lambda)^2}) = 0$$

con $\lambda \in (0, 1/2)$, la cui soluzione è $\lambda = 1/(2 + \sqrt{(}2f_3^2/(f_1^2 + f_2^2))) \approx 0.2929$ (si vede facilmente che questa corrisponde ad un punto di minimo). Pertanto $Q = 0.2851$ e $\overline{\alpha} = 1 - F_{\chi^2(1)}(Q) = 0.5934$ e l'ipotesi di adattamento è accettata al 5%.

Se ora l'ampiezza del campione fosse $n = 2400$ con le stesse frequenze relative si vedrebbe immediatamente, con calcoli analoghi che $Q = 28.51$ e $\overline{\alpha} = 9.3153 \cdot 10^{-8}$ per cui l'ipotesi sarebbe rifiutata (per ogni $\lambda \in (0, 1/2)$).

Soluzione es 2.1:

1. Il test

$$H_0 : \mu_1 - \mu_2 = 0 =: \Delta_0$$
$$H_1 : \mu_1 - \mu_2 \neq 0 =: \Delta_0$$

ha regione critica e P-value

$$\left|\frac{\overline{x}_1 - \overline{x}_2 - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right| > z_{\alpha/2}$$

$$\overline{\alpha} := 2\left(1 - \phi\left(\left|\frac{\overline{x}_1 - \overline{x}_2 - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right|\right)\right)$$

.

Non essendo suggerito nessun livello di significatività procediamo al calcolo del P-value

$$\left|\frac{\overline{x}_1 - \overline{x}_2 - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right| \approx 3.015$$

$$\overline{\alpha} := 2\left(1 - \phi\left(\left|\frac{\overline{x}_1 - \overline{x}_2 - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right|\right)\right) \approx 0.0026$$

che suggerisce, essendo $\overline{\alpha}$ piuttosto piccolo, un'ipotesi nulla molto poco attendibile.

2. Si utilizza lo stimatore per σ^2

$$S^2 := \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

e la regione di rifiuto e P-value

$$T := \frac{\overline{x}_1 - \overline{x}_2 - \Delta_0}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$|T| > t_{\alpha/2, n_1+n_2-2}$$

$$\overline{\alpha} := 2(1 - F_{T(n_1+n_2-2)}(|T|)).$$

Essendo $T \approx 3.4317$ ed $\overline{\alpha} \approx 6.847 \cdot 10^{-4}$ si ha che l'ipotesi nulla è poco plausibile.

3. In questo caso il test ha regione di rifiuto e P-value

$$T := \frac{\overline{x}_1 - \overline{x}_2 - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\nu := \frac{\left((s_1^2/n_1 + s_2^2/n_2\right)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

$$|T| > t_{\alpha/2, \nu}$$

$$\overline{\alpha} := 2(1 - F_{T(\nu)}(|T|)).$$

Poichè $T \approx 3.0151$, $\nu \approx 144.3428$ e $\overline{\alpha} \approx 0.003$ si ha, anche in questo caso, che l'ipotesi nulla è poco plausibile.

Soluzione es 2.2:

1. La media campionaria dei primi tre anni (assunta come vera) sul campione di ampiezza $n = 8940000$ è $p_0 := (13+15+20)/(3n) \approx 1.7897 \cdot 10^{-6}$, mentre $\overline{p} = 18/n \approx 2.0134 \cdot 10^{-6}$. Il test
$$\begin{aligned} H_0 :& p \geq p_0 \\ H_1 :& p < p_0 \end{aligned}$$
ha come regione critica (o regione di rifiuto) e P-value
$$\begin{aligned} &\frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}} < z_{1-\alpha} \\ &\overline{\alpha} = \phi\left(\frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right). \end{aligned}$$
Essendo $\overline{p} > p_0$ l'ipotesi nulla non può essere rifiutata a livelli inferiori a 0.5. In ogni caso si ha
$$\begin{aligned} &\frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \approx 0.5 \\ &z_{1-0.05} \approx -1.6449 \\ &\overline{\alpha} = \phi\left(\frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right) \approx 0.6915. \end{aligned}$$

2. Questa volta si tratta di studiare il test
$$\begin{aligned} H_0 :& p \leq p_0 \\ H_1 :& p > p_0 \end{aligned}$$
ha come regione critica e P-value
$$\begin{aligned} &\frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}} > z_{\alpha} \\ &\overline{\alpha} = 1 - \phi\left(\frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right). \end{aligned}$$
Eseguendo i calcoli si ha
$$\begin{aligned} &z_{0.05} \approx 1.6449 \\ &\overline{\alpha} = \phi\left(\frac{\overline{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right) \approx 0.3185 \end{aligned}$$

pertanto ancora non si può rifiutare l'ipotesi nulla (equivalentemente utilizzando la regione critica o il P-value).

3. Si considerano $\overline{p}_1 = 18/n$ e $\overline{p}_2 = 20/n$ ed il test

$$\begin{aligned} H_0 :& p_1 \geq p_2 \\ H_1 :& p_1 < p_2 \end{aligned}$$

ha come regione critica (o regione di rifiuto) e P-value

$$\begin{aligned} &\frac{\overline{p}_1 - p_2}{\sqrt{p_1(1-p_1)/n + p_2(1-p_2)/n}} < z_{1-\alpha} \\ &\overline{\alpha} = 1 - \phi\left(\frac{\overline{p}_1 - p_2}{\sqrt{p_1(1-p_1)/n + p_2(1-p_2)/n}}\right). \end{aligned}$$

Eseguendo i calcoli si ha

$$\begin{aligned} &\frac{\overline{p}_1 - p_2}{\sqrt{p_1(1-p_1)/n + p_2(1-p_2)/n}} \approx \frac{-2}{\sqrt{18+20}} \approx -0.3244 \\ &\overline{\alpha} = 1 - \phi\left(\frac{\overline{p}_1 - p_2}{\sqrt{p_1(1-p_1)/n + p_2(1-p_2)/n}}\right) \approx 0.3728 \end{aligned}$$

quindi non possiamo rifiutare l'ipotesi nulla al livello di significatività del 5%.

4. Si considerano $n_1 := n$, $n_2 := 3n$, $\overline{p}_1 = 28/n_1 \approx 3.132 \cdot 10^{-6}$ e $\overline{p}_2 = (20+13+15)/n_2 \approx 1.7897 \cdot 10^{-6}$ ed il test

$$\begin{aligned} H_0 :& p_1 = p_2 \\ H_1 :& p_1 \neq p_2 \end{aligned}$$

ha come regione di accettazione e P-value

$$\begin{aligned} &z_{1-\alpha/2} < \frac{\overline{p}_1 - \overline{p}_2}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} < z_{\alpha/2} \\ &\overline{\alpha} = 2\left(1 - \phi\left(\frac{|\overline{p}_1 - \overline{p}_2|}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}}\right)\right). \end{aligned}$$

Eseguendo i calcoli si ha

$$\begin{aligned} &\frac{\overline{p}_1 - \overline{p}_2}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \approx 2.0785 \\ &z_{0.025} = -z_{0.975} \approx 1.9600 \\ &\overline{\alpha} \approx 0.0377 \end{aligned}$$

pertanto rifiutiamo l'ipotesi nulla al livello di significatività del 5%.

Soluzione es 3.1:

Consideriamo l'ipotesi nulla

$$\mathrm{H}_0\colon\ \mu_A = \mu_B = \mu_C = \mu_D$$

e l'ipotesi alternativa

$$\mathrm{H}_1\colon \text{ almeno due medie sono diverse}$$

Calcoliamo i valori

	A	B	C	D		
$Y_{i.}$	173	212	115	183	$Y_{..} =$	683
n_i	5	6	4	6	$N =$	21
$\overline{Y}_{i.}$	34.6	35.33	28.75	30.5	$\overline{Y}_{..} =$	32.52

Inoltre abbiamo $\sum_{i=1}^{4}\sum_{j=1}^{n_i} Y_{ij}^2 = 22521$ e $\sum_{i=1}^{4}\frac{Y_{i.}^2}{n_i} = 22364.217$.
Possiamo ora calcolare:

$$SS_E = \sum_{i=1}^{4}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2 = \sum_i\sum_j Y_{ij}^2 - \sum_i \frac{Y_{i.}}{n_i} = 22521 - 22364.217 = 156.783$$

$$SS_F = \sum_i n_i(\overline{Y}_{i.}^2 - \overline{Y}_{..}^2)^2 = \sum_i \frac{Y_{i.}}{n_i} - \frac{Y_{..}}{N} = 22364.217 - \frac{683^2}{21} = 150.455$$

Ricordando che

$$MS_F := \frac{SS_F}{n-1}, \qquad MS_E := \frac{SS_E}{\sum_{i=1}^{n} n_i - n}$$

(in questo caso $n = 4$), possiamo ora costruire la tabella ANOVA:

Fonte di variabilità	SS	$g.l.$	MS	F
Fattore	150.455	3	50.15	5.44
Errore	156.783	17	9.22	
Tot	307.238	20		

poichè $\alpha = 0.05$ e $F_{0.05;3;17} = 3.2$ l'ipotesi H_0 di uguaglianza fra le resistenze delle leghe va rifiutata al livello del 5% dato che $F = 5.44 > F_{0.05;3;17} = 3.2$.

Soluzione es 3.2:

Ricordiamo che la statistica test utilizzata per ipotesi sul confronto di due medie con varianze note è

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)\,.$$

L'intervallo di confidenza quindi è del tipo

$$\bar{X}_1 - \bar{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

In questo caso conosciamo $n_1 = n_2 = 20$; $\sigma_1 = 5$; $\sigma_2 = 4$; $\bar{x}_1 = 88$ e $\bar{x}_2 = 91$.

1. per l'intervallo di confidenza al 90%: $\alpha = 0.1$ abbiamo $z_{\alpha/2} = 1.64$, quindi l'intervallo cercato è il seguente:

   $$-5.348 \leq \mu_1 - \mu_2 \leq -0.652.$$

   Quindi, siccome l'intervallo contiene solo valori strettamente negativi, possiamo affermare che il filo di lana di tipo 2 ha più alta resistenza media.

2. per l'intervallo di confidenza al 98%: $\alpha = 0.02$ abbiamo $z_{\alpha/2} = 2.32$, quindi l'intervallo cercato è il seguente:

   $$-6.322 \leq \mu_1 - \mu_2 \leq 0.322$$