Esercitazioni di probabilità e statistica

Test ipotesi

Esercizio 2.1:

Due ditte producono un certo tipo di macchinario. La ditta A ne ha prodotti 6 e venduto 1, la ditta B ne ha venduti 4 su 9. Si formuli una opportuna ipotesi statistica e sulla base di questa si esegua un opportuno test, determinandone il livello di significatività, per determinare se si può affermare che la ditta B ha maggior probabilità di vendita rispetto alla ditta A.

Soluzione: Per affrontare questi problemi è conveniente mettere in tabella di contingenza: A questo

	A	B	
V	1	4	5
$\overline{V}$	5	5	10
	6	9	15

punto formalizziamo le ipotesi H_0 è l'ipotesi nulla da verificare, mentre H_1 è l'ipotesi alternativa (da considerarsi valida quando H_0 viene rifiutata):

$$H_0 \to A \text{ e } V \text{ sono s-indipendenti}$$
$$H_1 \to A \text{ e } V \text{ sono s-dipendenti, quindi } \Pr\{A|V\} < \Pr\{B|V\}$$

A questo punto si definisce una variabile Y_n con distribuzione nota quando H_0 risulta vera. Si suddivide lo spazio in due insiemi disgiunti uno in cui H_0 viene accettata (C) e uno in cui viene rigettata ($\overline{C}$). La decisione di accettare o rifiutare l'ipotesi nulla in seguito alle osservazioni è sempre affetta da due tipi di errore:
a) l'errore di prima specie α, quello che commettiamo se rigettiamo l'ipotesi nulla H_0 quando essa è vera:

$$\alpha = \Pr\{H_1|H_0\} = \Pr\{Y_n \in \overline{C}|H_0 \; vera\}$$

b) l'errore di seconda specie β, quello che commettiamo se accettiamo H_0 mentre è vera H_1:

$$\beta = \Pr\{H_0|H_1\} = \Pr\{Y_n \in C|H_1 \; vera\}$$

Quindi in questo caso posso scrivere che:

$$Y_5 = \#(A \cap V)$$

e l'errore di prima specie si può scrivere come:

$$\alpha = \Pr\{Y_5 \leq 1|H_0\}$$

Usando il test esatto di Fisher (stiamo considerando numeri piccoli e non si può usare il χ^2) per testare l'ipotesi posso dire che:

$$\begin{aligned}
\alpha &= \Pr\{Y_5 \leq 1|H_0\} = \Pr\{\#(A \cap V) \leq 1|H_0\} \\
&= \Pr\{\#(A \cap V) = 0|H_0\} + \Pr\{\#(A \cap V) = 1|H_0\} \\
&= \frac{\binom{6}{1}\binom{9}{4}}{\binom{15}{5}} + \frac{\binom{6}{0}\binom{9}{5}}{\binom{15}{5}} \\
&= 0.252 + 0.042 \\
&= 29.4\%
\end{aligned}$$

Questo è un risultato a rischio, il livello di significatività si ottiene facendo:

$$1 - \alpha = 70.6\%$$

Il test risulta negativo, non si può dire che si riesce a rigettare H_0 con abbastanza certezza.

Esercizio 2.2:

Un gruppo di medici ha trovato un possibile vaccino contro il comune raffreddore. Ne viene verificata l'efficacia su un gruppo di 24 volontari divisi in 2 gruppi di 12 persone, uno sperimentale e uno di controllo. I membri del gruppo sperimentale vengono vaccinati, mentre quelli del gruppo di controllo no. Dopo sei mesi a tutte le 24 persone si chiede se abbiano avuto raffreddori durante il periodo trascorso. I risultati sono in tabella.

		raffreddati	
		SI	NO
vaccinati	SI	2	10
	NO	7	5

a) Si formuli una ipotesi statistica per la verifica dell'efficacia del vaccino.
b) Sulla base di questa si esegua il test esatto di Fisher, determinandone il livello di significatività.
c) Si può affermare che il vaccino contro il raffreddore è significativamente efficace?

Soluzione: a) Come abbiamo fatto in precedenza mettiamo in tabella i dati, chiamiamo R il caso in sui si sia presentato il raffreddore e $\overline{R}$ il suo complemento, V il caso in cui sia stato somministrato il vaccino $\overline{V}$ il suo complemento. e formuliamo le ipotesi:

	R	$\overline{R}$	
V	2	10	12
$\overline{V}$	7	5	12
	9	15	24

$$H_0 \rightarrow \text{il vaccino è inefficace ossia } V \text{ e } R \text{ sono s-indipendenti}$$
$$H_1 \rightarrow \text{il vaccino è efficace ossia } V \text{ e } R \text{ sono s-dipendenti}$$

In generale i problemi di questo tipo seguono uno schema semplice: b) Il test di Fisher in generale può essere scritto nei seguenti modi:

$$\Pr\{Y_{a+b} \leq a\} = \sum \frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{n}{a+b}} \qquad \textit{notare che usiamo a perchè cerchiamo l'intersezione } R \cap V$$

		Casi		
		R	$\overline{R}$	
v.a.	V	a	b	a+b
	$\overline{V}$	c	d	c+d
		a+c	b+d	n= a+b+c+d

Tabella 1: Tabella generale per eseguire il test esatto di Fisher

oppure in alternativa

$$\Pr\{Y_{a+b} \leq a\} = \sum \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{n!\,a!\,b!\,c!\,d!}$$

Quindi in questo caso:

$$\begin{aligned}
\alpha &= \Pr\{\#(R \cap V) \leq 2\} \\
&= \Pr\{Y_{12} \leq 2 | H_0\} \\
&= \Pr\{Y_{12} = 2 | H_0\} + \Pr\{Y_{12} = 1 | H_0\} + \Pr\{Y_{12} = 0 | H_0\} \\
&= \frac{\binom{9}{2}\binom{15}{10}}{\binom{24}{12}} + \frac{\binom{9}{1}\binom{15}{11}}{\binom{24}{12}} + \frac{\binom{9}{0}\binom{15}{12}}{\binom{24}{12}} \\
&= 0.03998 + 0.0054 + 0.00017 \\
&= 0.045
\end{aligned}$$

c) In questo caso il livello di significatività è sempre $1 - \alpha$

$$1 - \alpha = 95.5\%$$

dal momento che $\alpha \leq 5\%$ il test esatto di Fisher è positivo, ovvero l'assunzione H_0 comporta un evento improbabile. H_0 può essere rigettata.

Esercizio 2.3:

Per sperimentare un fertilizzante, un giardiniere sceglie 14 coppie di piante simili e tratta una pianta di ciascuna coppia con il fertilizzante. Dopo due mesi, 12 delle 14 piante trattare son più rigogliose delle corrispondenti non trattare (per i rimanenti due casi accade l'opposto).
a) Se il fertilizzante fosse inefficace quale sarebbe la probabilità che un puro caso abbia portato a 12 o più successi?
b) I 12 successi forniscono indicazioni riguardo l'efficacia del fertilizzante?

Soluzione: In questo caso possiamo individuare due possibili esiti distinti:

$$H_1 \rightarrow \text{il fertilizzante ha funzionato}$$
$$H_0 \rightarrow \text{il fertilizzante non è efficace}$$

Siamo in un caso in cui l' ipotesi che le due categorie possano occorrere ugualmente è nulla, quindi il test da usare è quello della binomiale. Chiamiamo y la v.a. numero di successi su 14 prove.

$$\Pr\{y\} = \binom{14}{y} \frac{1}{2^y} \frac{1}{2^{14-y}} = \frac{1}{2^{14}} \binom{14}{y}$$

a) Valutiamo quindi il fattore di rischio o errore di prima specie α:

$$\begin{aligned}\alpha &= \Pr\{y \geq 12|H_0\} \\ &= \Pr\{12\} + \Pr\{13\} + \Pr\{14\} \\ &= \binom{14}{12}\frac{1}{2^{14}} + \binom{14}{13}\frac{1}{2^{14}} + \frac{1}{2^{14}} \\ &= 5.55 \cdot 10^{-3} + 8.54 \cdot 10^{-4} + 6.1 \cdot 10^{-5} \\ &= 0.6\%\end{aligned}$$

Questa è la probabilità che il risultato del fertilizzante sia stato un puro caso.

b) Per valutare quanto i 12 successi ci dicano sull'efficacia del fertilizzante basta valutare la significatività del test:

$$1 - \alpha = 99.4\%$$

Esercizion 2.4:

In una certa scuola 420 studenti su un totale di 600 superano un test standardizzato di matematica per il quale il tasso nazionale di promozione è del 60%.
a) Se gli studenti della scuola non hanno particolare attitudine per la matematica (né in positivo né in negativo), quanti studenti dovrebbero superare l'esame mediamente?
b) Qual è la probabilità che alemno 420 studenti superino l'esame?
c) Si può affermare che la scuola impartisce un buona preparazione agli studenti?

Soluzione: In questo caso ho due soli esiti che possono uscire da questa analisi, inoltre uno è mutualmente esclusivo con l'altro quindi il test da usare sarà quello della binomiale. Definiamo la probabilità che il test standardizzato venga passato come $p = 0.6$.

a) La media degli studenti che si si aspetta passino l'esame è:

$$600 \cdot p = 360$$

A questo punto possiamo formulare le ipotesi:

$$\begin{aligned}H_0 &\rightarrow \text{probabilità di superamento uguale alle altre scuole,} \\ H_1 &\rightarrow \text{probabilità di superamento più alta delle altre scuole.}\end{aligned}$$

b) Usando il test della binomiale possiamo scrivere che:

$$\alpha = \Pr\{Y \geq 420\} = \sum_{y=420}^{600} \binom{600}{y} \cdot 0.6^y \cdot 0.4^{600-y}$$

Dobbiamo porre l'attenzione sul fatto che la sommatoria spazia su n grandi, in questo caso sappiamo che la distribuzione binomiale si riduce ad una gaussiana con stessa media e varianza. Quindi:

$$\mu = 360, \qquad \sigma^2 = npq = 600 \cdot 0.6 \cdot 0.4 = 144, \qquad \sigma = 12$$

Possiamo quindi riscrivere:

$$\alpha = \Pr\{Y \geq 420\} \simeq \frac{1}{12 \cdot \sqrt{2\pi}} \int_{420}^{600} e^{-\frac{(Y-360)^2}{2 \cdot 12^2}} dY$$

facciamo un cambio di variabile

$$u = \frac{Y - 360}{12}, \qquad dY = 12du$$

Quindi:

$$\begin{aligned}\alpha &= \frac{1}{\sqrt{2\pi}}\int_{5}^{20} e^{-\frac{u^2}{2}}du \\ &= 3\cdot 10^{\cdot 7} \\ &\sim 0\end{aligned}$$

c) Il livello di rischio è nullo, quindi si può affermare che il livello si istruzione fornito dalla scuola è buono senza alcun dubbio.

Esercizion 2.5:

Su un numero ridotto di cavie si sono sperimentati due pesticidi per verificare se quello di nuova produzione B abbia effetti più letali di quello usato in precedenza A. Il risultato dell'esperimento si può vedere in tabella.

	cavie	
	Sopravvissute	Non sopravvissute
Pesticida A	7	1
Pesticida B	3	6

a) Si formuli una ipotesi statistica per la verifica dell'efficacia del pesticida.
b) Sulla base di questa si esegua il test esatto fi Fisher, determinandone il livello di significatività.
c) Si può affermare che il pesticida di nuova generazione è più efficace del pesticida A?

Soluzione: Costruiamo la tabella di contingenza per quanto riguarda l'esperimento in considerazione:

	A	B	
S	7	3	10
$\overline{S}$	1	6	7
	8	9	17

a) Formuliamo le ipotesi statistiche tenendo a mente che H_0 e H_1 devono essere mutualmente esclusive:

$$H_0 \to \text{A e B hanno la stessa efficacia,}$$
$$H_1 \to \text{A e B hanno efficacia diversa } \Pr\{\overline{S}|B\} > \Pr\{\overline{S}|A\}.$$

b) Eseguiamo il test di Fisher considerando la variabile $\#(B\cap\overline{S})$:

$$\begin{aligned}\alpha &= \Pr\{\#(B\cap\overline{S})\geq 6\} \\ &= \Pr\{\#(B\cap\overline{S}) = 6\} + \Pr\{\#(B\cap\overline{S}) = 7\} \\ &= \frac{\binom{9}{6}\binom{8}{1}}{\binom{17}{7}} + \frac{\binom{9}{7}\binom{8}{0}}{\binom{17}{7}} \\ &= 0.036 \\ &= 3.6\%\end{aligned}$$

c) Calcolando la significatività del test:

$$1-\alpha = 96.4\%$$

si può concludere che l'ipotesi H_0 viene rigettata e si considera attendibile l'ipotesi H_1.

Esercizion 2.6:

Su due campioni indipendenti sono stati registrati le seguenti frequenze di valori eccedenti (+) e non eccedenti (-) la mediana comune (ossia calcolata sull'insieme dei due campioni):

	$+$	$-$	
I	3	6	9
II	5	2	7
	8	8	16

Si verifichi se i due campioni appartengono a popolazioni diverse.

Soluzione: Iniziamo formulando le ipotesi:

$$H_0 \to \text{I e II hanno la stessa mediana,}$$
$$H_1 \to \text{I e II hanno mediane diverse.}$$

poi valutiamo α sull'intersezione tra i casi $(-)$ e il campione II.

$$\begin{aligned}\alpha &= \Pr\{\#(- \cap \text{II}) \leq 2\} \\ &= \Pr\{\#(- \cap \text{II}) = 2\} + \Pr\{\#(- \cap \text{II}) = 1\} + \Pr\{\#(- \cap \text{II}) = 0\} \\ &= \frac{\binom{8}{2}\binom{8}{5}}{\binom{16}{7}} + \frac{\binom{8}{1}\binom{8}{6}}{\binom{16}{7}} + \frac{\binom{8}{0}\binom{8}{7}}{\binom{16}{7}} \\ &= 0.137 + 0.02 + 0.001 \\ &= 0.158 \\ &= 15.7\%\end{aligned}$$

Il test di Fisher risulta negativo. Non riusciamo infatti a scartare l'ipotesi H_0 con l'adeguata significatività:

$$1 - \alpha = 84.3\%$$

Si può procedere anche nel seguente modo, senza nessuna differenza:

	I	II	
$+$	3	5	8
$-$	6	2	8
	9	7	16

e poi procedere con il test di Fisher seguendo la medesima procedura:

$$\begin{aligned}\alpha &= \Pr\{\#(\text{II} \cap -) \leq 2\} \\ &= \Pr\{\#(\text{II} \cap -) = 2\} + \Pr\{\#(\text{II} \cap -) = 1\} + \Pr\{\#(\text{II} \cap -) = 0\} \\ &= \frac{\binom{7}{2}\binom{9}{6}}{\binom{16}{8}} + \frac{\binom{7}{1}\binom{9}{7}}{\binom{16}{8}} + \frac{\binom{7}{0}\binom{9}{8}}{\binom{16}{8}} \\ &= 0.137 + 0.02 + 0.001 \\ &= 0.158 \\ &= 15.7\%\end{aligned}$$