

## L'affondamento del Titanic

**Livello scolastico:** 2° biennio

Abilità interessate	Conoscenze	Nuclei coinvolti	Collegamenti esterni
Identificare situazioni che richiedono di rilevare lo stesso carattere su una unità statistica formata da 2 elementi, o 2 caratteri diversi sulla stessa unità statistica. Impostare una tabella a doppia entrata; classificare i dati secondo due caratteri e riconoscere in essa i diversi elementi individuabili. Selezionare, produrre ed usare appropriate rappresentazioni grafiche delle distribuzioni doppie.	Distribuzione doppia di frequenze e tabella a doppia entrata. Distribuzioni condizionate e marginali. Principali rappresentazioni grafiche per le distribuzioni doppie rispetto a caratteri di qualsiasi natura. Concetto e significato di modello: dipendenza e indipendenza stocastica. Probabilità semplice, condizionata e composta.	<u>Risolvere e porsi problemi</u>  Dati e previsioni  Argomentare, congetturare, dimostrare	

### Contesto

Dipendenza stocastica.

Il contesto è di tipo matematico ed extramatematico; per quest'ultimo aspetto si colloca nell'ambito della vita sociale e dell'attualità.

Questo esempio è riferito a dati reali che facilmente possono colpire lo studente nella vita quotidiana attraverso notizie provenienti da televisione e giornali. La parte didatticamente più importante è l'attenzione da porre a dati e percentuali che, a prima vista, potrebbero fuorviare.

### Descrizione dell'attività

Il Titanic trasportava passeggeri in tre classi: 1<sup>a</sup>, 2<sup>a</sup>, 3<sup>a</sup>. Dopo l'affondamento fu stilata la relazione relativa all'incidente ed a tutte le operazioni di salvataggio successive. Alcuni dati sono sinteticamente riportati nelle seguenti tabelle:

Classe dei viaggiatori	Viaggiatori alla partenza			Totale
	Uomini	Donne	Bambini	
1 <sup>a</sup> Classe	175	144	6	325
2 <sup>a</sup> Classe	168	93	24	285
3 <sup>a</sup> Classe	462	165	79	706
Totale	805	402	109	1316

*Tabella 1*

Classe dei viaggiatori	Sopravvissuti			Totale
	Uomini	Donne	Bambini	
1 <sup>a</sup> Classe	57	140	6	203
2 <sup>a</sup> Classe	14	80	24	118
3 <sup>a</sup> Classe	75	76	27	178
Totale	146	296	57	499

*Tabella 2*

Classe dei viaggiatori	Morti o Dispersi			Totale
	Uomini	Donne	Bambini	
1 <sup>a</sup> Classe	118	4	0	122
2 <sup>a</sup> Classe	154	13	0	167
3 <sup>a</sup> Classe	387	89	52	528
Totale	659	106	52	817

Tabella 3

*Osservazione 1:* I passeggeri delle tre classi non viaggiavano tutti nelle stesse condizioni. Ad esempio, i passeggeri di terza classe erano in gran parte emigranti con famiglia e figli, probabilmente non tutti erano perfettamente padroni della lingua inglese (c'erano emigranti francesi, ma anche inglesi usi a parlare in dialetto) e, inoltre, gli alloggi di terza erano collocati in luoghi della nave da cui non era immediato l'accesso al ponte dove si trovavano le lance di salvataggio. Al contrario, i passeggeri di prima erano persone facoltose abituate a parlare in inglese, in generale coppie senza figli in viaggio di divertimento e gli alloggi di prima si affacciavano sul ponte o erano nelle vicinanze. Per dare forza all'osservazione che in terza classe c'erano molti più bambini che nelle altre classi, si può costruire la Tabella 4 con i dati relativi ai bambini alla partenza.

Classe	Viaggiatori alla partenza		
	Bambini	In totale	Percentuale
1 <sup>a</sup> Classe	6	325	1,8 %
2 <sup>a</sup> Classe	24	285	8,4 %
3 <sup>a</sup> Classe	79	706	11,2 %
Totale	109	1316	8,3 %

Tabella 4

Effettivamente, in terza classe i bambini erano molto più numerosi, rappresentavano l'11,2% dei passeggeri di quella classe, in seconda rappresentavano l'8,4% dei passeggeri e in prima rappresentavano solo l'1,8% dei passeggeri. Rivolgiamoci alcune domande atte a capire come stavano effettivamente le cose.

**Domanda:** Osservando i dati riportati nelle tabelle, sembra possibile affermare che, a causa delle maggiori difficoltà incontrate nel salvataggio dei passeggeri di terza classe (con famiglie, con bambini, alloggiati in luoghi poco agevoli, con difficoltà di linguaggio), sono stati salvati più passeggeri di seconda e prima classe che di terza. Questa è solo una impressione o può essere confermata da una analisi puntuale ed approfondita dei dati?

Vediamo alcune (possibili) risposte, suddivise per "categorie".

**Prima risposta (uomini):** Per rispondere alla domanda, si incominci a prendere in considerazione i soli maschi. A partire dalle Tabella 2 e 3 si costruisca la Tabella 5. Salta agli occhi che più della metà dei sopravvissuti maschi (il 51,4%) è di terza classe). Verrebbe da dire, ad una prima affrettata (e, come vedremo, errata) conclusione, che tra gli uomini si sono salvati soprattutto quelli di terza classe (51,4%), seguiti da quelli di prima (39%) e da quelli di seconda (9,6%).

Classe	Sopravvissuti i uomini	Ripartizione percentuale dei sopravvissuti uomini	Uomini alla partenza	Ripartizione percentuale degli uomini alla partenza	Percentuale di sopravvissuti sui parenti
--------	---------------------------	--	-------------------------	--	---

1 <sup>a</sup> Classe	57	39,0	175	21,7	32,6
2 <sup>a</sup> Classe	14	9,6	168	20,9	8,3
3 <sup>a</sup> Classe	75	51,4	462	57,4	16,2
Totale	146	100,0	805	100,0	18,1

Tabella 5

Guardando con più cura la tabella, però, si nota che anche alla partenza ci sono più uomini in terza classe (57,4%), che non in seconda (20,9%) o in prima (21,7%).

A questo punto si può provare a dare una prima risposta (corretta): tra gli uomini, alla partenza i viaggiatori di terza classe sono il 57,4%; tra i sopravvissuti, invece, gli uomini di terza sono solo il 51,4%. In seconda classe, invece: gli uomini sono il 20,9% dei partenti, ma il 9,6% dei sopravvissuti. Diversa ancora la situazione in prima: gli uomini sono il 21,7% dei partenti ed il 39% dei sopravvissuti.

L'ultima colonna della Tabella 5 riporta, per i maschi, la percentuale di sopravvissuti sui partenti, classe per classe. Si vede bene, adesso, che in terza i sopravvissuti sono il 16,2% dei partenti, in seconda sono l'8,3% e in prima il 32,6%.

*Osservazione 2:* Con quanto detto finora, si può fare riferimento alle probabilità condizionate.

Si provi a immaginare di pescare a caso un nome dall'elenco dei viaggiatori: se si pesca tra i nomi di prima classe, la probabilità di trovare il nome di un sopravvissuto è 0,326; se si pesca tra i nomi di seconda classe la probabilità di trovare il nome di un sopravvissuto è 0,083; se si pesca tra i nomi di terza classe la probabilità è 0,162. Appare così evidente, come avevamo sottolineato prima, che è molto più facile trovare il nome di un sopravvissuto in prima classe che non in terza perché le probabilità condizionate di sopravvivenza sono diverse e dipendono dalla classe del viaggiatore.

Seconda risposta (donne): E per le donne?

Si cominci a costruire la Tabella 6, analoga alla 5 ma relativa alle donne. Tra le sopravvissute, il 47,3% è di prima classe, il 27% è di seconda e solamente il 25,7% è di terza. Alla partenza, però, le percentuali sono molto diverse: il 35,8% delle donne viaggia in prima, il 23,1% in seconda e il 41,1% in terza.

Se poi si va a vedere l'ultima colonna della Tabella 5, che riporta, per le donne, la percentuale di sopravvissute sulle partenti, classe per classe, si vede bene che in terza le sopravvissute sono il 46,1% delle partenti, in seconda classe sono l'86,0% e in prima sono state salvate praticamente tutte le passeggere (140 su 144, pari al 97,2%).

Classe	Sopravvissute donne	Ripartizione percentuale delle sopravvissute e donne	Donne alla partenza	Ripartizione percentuale delle donne alla partenza	Percentuale di sopravvissute e sulle partenti
1 <sup>a</sup> Classe	140	47,3	144	35,8	97,2
2 <sup>a</sup> Classe	80	27,0	93	23,1	86,0
3 <sup>a</sup> Classe	76	25,7	165	41,1	46,1
Totale	296	100,0	402	100,0	73,6

Tabella 6

Anche adesso si potrebbe dire che, se si pesca a caso un nome dall'elenco delle viaggiatrici, la probabilità di trovare il nome di una sopravvissuta è 0,972, se si pesca tra i nomi di prima classe; è

0,86, se si pesca tra i nomi di seconda classe; è infine 0,461, se si pesca tra i nomi di terza classe. Anche per le donne, è dunque molto più facile trovare il nome di una sopravvissuta in prima classe che non in terza.

*Osservazione 3:* Anche in questo caso si vede che le probabilità di sopravvivenza sono condizionate alla classe del viaggiatore.

Terza risposta (*tutti = uomini+donne+bambini*): Alla domanda di partenza, può essere risposto anche prendendo in considerazione tutti i passeggeri complessivamente considerati, maschi e femmine, compresi i bambini (*cfr. Tabella 7*).

La domanda di partenza può essere riformulata nei seguenti modi alternativi fra loro equivalenti:

- i passeggeri delle tre classi hanno avuto le stesse opportunità di salvezza?
- la appartenenza a una classe ha modificato in meglio o in peggio le opportunità di salvezza?
- la conclusione favorevole del viaggio è dipesa dalla classe?

Se le opportunità di salvezza sono state (più o meno) eguali per tutti i passeggeri di tutte le classi, in tutte e tre le classi si deve trovare (più o meno) la stessa percentuale di sopravvissuti. Complessivamente, i sopravvissuti sono stati 499 su 1316 passeggeri, pari al 37,9%. La stessa percentuale di sopravvissuti si deve trovare in tutte tre le classi.

Per quanto riguarda la prima classe, il 37,9% di 325 è 123,7: questo è il numero atteso di sopravvissuti nella prima classe, nella ipotesi che questi passeggeri abbiano avuto le stesse opportunità di salvezza degli altri.

Per la seconda classe il 37,9% di 285 è 108,1: questo è il numero atteso di sopravvissuti in seconda, nella stessa ipotesi di indipendenza delle opportunità di salvezza dalla classe.

Analogamente, per quanto riguarda la terza classe, il 37,9% di 706 è 267,7: questo è il numero atteso di sopravvissuti in terza.

Si procede nello stesso modo al calcolo dei morti e dispersi classe per classe. Complessivamente, tra morti e dispersi ci sono stati 817 passeggeri su 1316, pari al 62,1%. La stessa percentuale di dispersi si dovrebbe trovare nelle tre classi.

I numeri attesi, così calcolati, di sopravvissuti e di morti e dispersi sono riportati in ogni casella della tabella 7, in basso a destra, in carattere corsivo.

Si vede che in prima classe i sopravvissuti sono 203, più di quanti (123,2) risultano calcolati nella ipotesi di indipendenza tra esito del viaggio e classe. In terza classe, invece, i sopravvissuti sono 178, meno di quanti (267,7) risultano calcolati nella ipotesi di indipendenza tra esito del viaggio e classe. Per quanto riguarda i dispersi, avviene il contrario. In prima classe i dispersi sono 122, meno di quanti (201,8) risultano calcolati nella ipotesi di indipendenza tra esito del viaggio e classe. In terza classe, invece, i dispersi sono 528, più di quanti (438,3) risultano calcolati nella ipotesi di indipendenza tra esito del viaggio e classe.

Complessivamente, questi risultati permettono di affermare che effettivamente il numero di sopravvissuti, da un lato, e quello di morti e dispersi, dall'altro, sono in qualche modo collegati alla classe dei viaggiatori.

Classe dei viaggiatori	Sopravvissuti	Morti o Dispersi	Totale
1 <sup>a</sup> Classe	203 <i>123,2</i>	122 <i>201,8</i>	325
2 <sup>a</sup> Classe	118 <i>108,1</i>	167 <i>176,9</i>	285
3 <sup>a</sup> Classe	178 <i>267,7</i>	528 <i>438,3</i>	706
Totale	499	817	1316

*Tabella 7*

*Osservazione 4:* Si provi a generalizzare i ragionamenti fatti a riguardo dei dati della tabella 7 riguardo al calcolo delle frequenze attese nell'ipotesi di indipendenza tra esito del viaggio e classe. La tabella 8 è fatta come la 7, l'unica differenza è che si riferisce al caso generale di due variabili X con h modalità (h righe) e Y con k modalità (k colonne). Le frequenze osservate, riportate in ogni casella nella prima riga, sono indicate con  $f_{ij}$ ; le frequenze attese (calcolate nell'ipotesi di indipendenza tra classe e esito del viaggio), riportate in ogni casella nella seconda riga, sono indicate con  $\hat{f}_{ij}$ . Ricordando come sono state calcolate le frequenze attese della tabella 7, nella tabella 8 risulta, in ogni casella,  $\hat{f}_{ij} = T_{i.}T_{.j}/T_{..}$ .

*Osservazione 5:* Verificare che in una tabella a doppia entrata, come la 7 o la 8:

- in ogni riga la somma delle frequenze osservate risulta eguale alla somma delle frequenze attese:  $\sum_{j=1}^k f_{ij} = \sum_{j=1}^k \hat{f}_{ij} = T_{i.} \quad (i=1,2,\dots,h)$ ;
- in ogni colonna la somma delle frequenze osservate risulta eguale alla somma delle frequenze attese:  $\sum_{i=1}^h f_{ij} = \sum_{i=1}^h \hat{f}_{ij} = T_{.j} \quad (j=1,2,\dots,k)$ ;
- il totale generale delle frequenze osservate risulta eguale al totale delle frequenze attese:  $\sum_{i=1}^h \sum_{j=1}^k f_{ij} = \sum_{i=1}^h \sum_{j=1}^k \hat{f}_{ij} = T_{..} \quad (i=1,2,\dots,h; j=1,2,\dots,k)$ .

*Osservazione 6:* Usando la tabella 7 come un'urna e prendendo in considerazione la prima casella si possono facilmente calcolare tre probabilità: quella di estrarre a caso il nome di un viaggiatore di prima classe, quella di estrarre a caso il nome di un viaggiatore sopravvissuto, quella di estrarre a caso il nome di un viaggiatore di prima classe sopravvissuto.

$$P(I^a) = P(\text{di estrarre a caso il nome di un viaggiatore di prima classe}) = 325/1316$$

$$P(S) = P(\text{di estrarre a caso il nome di un viaggiatore sopravvissuto}) = 499/1316$$

$$P(I^a \cap S) = P(\text{di estrarre a caso il nome di un viaggiatore di prima classe sopravvissuto}) = 203/1316$$

Risulta facile verificare la non indipendenza tra i due eventi: i) viaggiatore di prima classe e ii) viaggiatore sopravvissuto:

$$P(I^a) \cdot P(S) = (325/1316) \cdot (499/1316) = 0,09364 \neq P(I^a \cap S) = 203/1316 = 0,1543$$

Variabile X	Variabile Y					Totale
	y <sub>1</sub>	...	y <sub>j</sub>	...	y <sub>k</sub>	
x <sub>1</sub>	$f_{11}$ $\hat{f}_{11}$	...	$f_{1j}$ $\hat{f}_{1j}$	...	$f_{1k}$ $\hat{f}_{1k}$	$T_{1.}$
⋮	⋮		⋮		⋮	⋮
x <sub>i</sub>	$f_{i1}$ $\hat{f}_{i1}$	...	$f_{ij}$ $\hat{f}_{ij}$	...	$f_{ik}$ $\hat{f}_{ik}$	$T_{i.}$
⋮	⋮		⋮		⋮	⋮
x <sub>h</sub>	$f_{h1}$ $\hat{f}_{h1}$	...	$f_{hj}$ $\hat{f}_{hj}$	...	$f_{hk}$ $\hat{f}_{hk}$	$T_{h.}$
Totale	$T_{.1}$	...	$T_{.j}$	...	$T_{.k}$	$T_{..}$

Tabella 8

*Osservazione 7:* Per valutare le distanze tra le frequenze osservate della tabella 7 (in carattere normale) e quelle attese nella ipotesi di indipendenza tra classe e esito del viaggio (in corsivo) è possibile utilizzare anche l'indice sintetico  $\chi^2$ .

Nel caso generale della tabella 8, l'espressione dell'indice è:

$$\chi^2 = \sum_{i=1}^h \sum_{j=1}^k \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

Nel caso particolare della tabella 7, l'indice risulta:

$$\chi^2 = (203-123,2)^2/123,2 + (122-201,8)^2/201,8 + (118-108,1)^2/108,1 + (167-176,9)^2/176,9 + (178-267,7)^2/267,7 + (528-438,3)^2/438,3 = 133,1.$$

Questo valore indica un allontanamento dalla situazione di indipendenza stocastica e, quindi, una dipendenza tra classe ed esito del viaggio.

*Osservazione conclusiva:* Si possono fare due ulteriori considerazioni sul risultato numerico del calcolo dell'indice  $\chi^2$ .

- Il valore di  $\chi^2$  è determinato, oltre che dalla eventuale dipendenza tra i due caratteri X e Y che si vuole misurare, anche dalla numerosità totale T. Infatti, se si moltiplicano tutte le frequenze della tabella 8, e quindi anche il totale T, per una costante  $k > 0$ , le proporzioni interne alla tabella rimangono invariate, ma si verifica facilmente che il nuovo indice risulta  $k\chi^2$

$$\sum_{i=1}^h \sum_{j=1}^k \frac{(kf_{ij} - k\hat{f}_{ij})^2}{k\hat{f}_{ij}} = k\chi^2$$

- Per avere un indice che permetta confronti e giudizi, senza essere influenzato dal numero totale T dei casi osservati e della dimensione  $h \times k$  della tabella, si utilizza un indice normalizzato  $\tilde{\chi}^2$  compreso tra 0 (nel caso di indipendenza) ed 1 (nel caso di massima dipendenza, compatibile con la struttura della tabella: dimensione  $h \times k$  e numerosità T).

$$0 \leq \tilde{\chi}^2 = \frac{\chi^2}{T \min\{(h-1); (k-1)\}} < 1.$$

Nel caso della tabella 7, l'indice normalizzato risulta:

$$\tilde{\chi}^2 = \frac{133,1}{1316 \min\{(3-1); (2-1)\}} = \frac{133,1}{1316 \cdot 2} = 0,05.$$

Questo valore misura l'allontanamento dalla situazione di indipendenza e indica, quindi, una certa dipendenza tra classe ed esito del viaggio.