Come interpretare e confrontare le probabilità e gli odds di "avere successo negli studi" che vengono forniti dalle tabelle e dai modelli statistici.

Nella prima riga della tabella 5 della sezione 3.2 si legge che, considerando la popolazione degli studenti della classe L02,

la percentuale degli studenti che hanno ottenuto almeno 40 CFU entro il primo anno, tra quelli che si trovano ***nel quarto più basso*** **della graduatoria dell'intero test**, è il **18,9%**. Diremo che questa è la **percentuale di successo** nel primo quarto.

Scendendo lungo la colonna, vediamo inoltre che le percentuali di successo nei diversi quarti della graduatoria sono

1° quarto	18,9%
2° quarto	40,9%
3° quarto	51,8%
4° quarto	70,6%

Vediamo così che tali percentuali crescono al variare della fascia di punteggio nel test. Questo fatto giustifica l'idea che

- è opportuno consigliare (eventualmente obbligare) gli studenti che sono in una fascia di punteggio basso di svolgere attività di consolidamento della propria preparazione iniziale, *fornendo nel contempo adeguate opportunità formative a tale scopo* (OFA);
- se è necessario selezionare un numero programmato di studenti, si possa usare la graduatoria di punteggio nel test e prendere gli studenti che hanno i punteggi più alti.

Vogliamo ora discutere alcuni modi per interpretare questi dati in termini di *probabilità* e di *odds*. Cominciamo considerando una situazione semplificata, in cui abbiamo soltanto due fasce:

fascia A: studenti che sono nel primo quarto, ossia che sono sotto il primo quartile,
fascia B: tutti gli altri studenti.

Nella tabella non si legge immediatamente la percentuale di studenti che hanno successo nella fascia B, ma, supponendo che il numero degli studenti in ogni quarto di graduatoria sia lo stesso[1], si vede che tale percentuale è circa del 54%.

[1] Tale approssimazione è del tutto sensata ai nostri fini.

Possiamo rappresentare la situazione mediante la seguente tabella di percentuali

	Percentuale di successo	Percentuale di insuccesso
Fascia A	**18,9%**	**81,1%**
Fascia B	**54%**	**46%**

Questa tabella mostra un'evidente differenza tra le percentuali di successo e di insuccesso nelle due fasce e può essere utilizzata per convincere uno studente della fascia A, se a questa fascia abbiamo assegnato un OFA, che per decidere abbiamo adottato un criterio valido.
Un altro modo per interpretare i dati nella tabella delle percentuali qui sopra è considerare il rapporto tra percentuale di successo e insuccesso in ciascuna fascia. Nella prima fascia si ha

successo/insuccesso $= \frac{18,9}{81,1} \cong 0,23 \cong 1:4$

Questo rapporto si può interpretare dicendo che *per uno studente che ha successo, ce ne sono 4 che non ci riescono*. Un altro modo di interpretare il rapporto $1:4$ è dire che i numeri 1 e 4 sono le quote di una scommessa equa sul successo di uno studente che si trova nel primo quarto della graduatoria. Il termine inglese per questo rapporto è "***odd***" e precisamente possiamo scrivere

odd di successo nella fascia A $= \frac{18,9}{81,1} \cong 0,23 \cong 1:4$

In modo analogo possiamo scrivere

odd di successo nella fascia B $= \frac{18,9}{81,1} \cong 1,17 \cong 7:6$

e questo vuol dire che, nella fascia B, per 7 studenti che hanno successo ce ne sono 6 che non riescono.

Possiamo interpretare la tabella delle percentuali in termini di probabilità, pensando ad esempio che se si prende a caso il nome di uno studente che si trova nel quarto più basso della graduatoria, e si va poi a vedere se lo studente ha ottenuto almeno 40 CFU alla fine del primo anno, la probabilità di questo evento è del 18,9%. In questo senso usiamo il termine *probabilità*. Ma bisogna sempre avere bene in mente che le carriere degli studenti non sono esperimenti casuali! e anzi l'obiettivo dei sistemi di istruzione è proprio quello di fornire agli studenti l'opportunità di "battere le probabilità" (in inglese "beat the odds").

Scriviamo allora la tabella di percentuali usando le probabilità di successo e insuccesso, espresse come numeri tra zero e 1, invece che come percentuali.

	Probabilità di successo	Probabilità di insuccesso
Fascia A	**0,19**	**0,81**
Fascia B	**0,54**	**0,46**

Vogliamo ora trovare un modo per misurare di quanto il successo è più probabile nella seconda fascia rispetto alla prima. Ci possono essere diversi modi per fare questo, che hanno ciascuno vantaggi e svantaggi. Due modi che vengono in mente subito, poiché li usiamo comunemente per confrontare due dati, sono i seguenti.

Fare la **differenza tra le probabilità** di successo nelle due fasce:

0,54 – 0,19 = **0,35**

Questo calcolo si può esprimere a parole dicendo che nella fascia B la probabilità aumenta di 0,35 (o 35 punti percentuali) rispetto alla fascia A.

Fare il **rapporto tra le probabilità** di successo nelle due fasce:

0,54/0,19 ≅ **2,9**

Questo calcolo si può esprimere a parole dicendo che nella fascia B la probabilità è circa tre volte maggiore che nella fascia B.

C'è però un altro modo che può avere dei vantaggi ed è **fare il rapporto tra gli odds**:

$$^{1{,}17}/_{0{,}23} \cong 5$$

Il termine inglese per il rapporto suddetto è ***Odds Ratio***, che si abbrevia OR.

Possiamo così scrivere la tabella di contingenza per le quattro fasce che si trovano tra i quartili nella tabella 5 della sezione 3.2, per la classe L02 e relative all'intero test.

	% successo	% insuccesso	odds	Odds Ratio OR
min – Q1	**18,9**	**81,1**	**0,2**	**1**
Q1 – Q2	**40,9**	**59,1**	**0,7**	**2,97**
Q2 – Q3	**51,8**	**48,2**	**1,1**	**4,6**
Q3 – Q3	**70,6**	**29,4**	**2,4**	**10,3**

Dove, nell'ultima colonna, per ogni riga, abbiamo calcolato i rapporti tra gli odds della fascia stessa e gli odds nella fascia 1, che abbiamo preso come riferimento.

Odds Ratio fascia 1 = 0,2/0,2 = 1

Odds Ratio fascia 2 = 0,7/0,2 = 2,97

Odds Ratio fascia 3 = 1,1/0,2 = 4,6

Odds Ratio fascia 4 = 2,4/0,2 = 10,3

In generale gli odds di un evento sono definiti come il rapporto tra la probabilità che l'evento si verifichi e la probabilità che l'evento non si verifichi. In altre parole, se p indica la probabilità dell'evento, si ha

Odds = p/(1-p)

Pertanto, mentre la probabilità è un numero compreso fra zero e 1, gli odds sono un numero tra zero e $+\infty$.

Conoscendo gli odds di un evento, la probabilità si ricava invertendo la relazione precedente e precisamente si ha

$$p = \frac{odds}{1+ odds} \qquad (1)$$

Pertanto, conoscere la probabilità di un evento è esattamente equivalente a conoscere gli odds; si tratta soltanto di due scale diverse per indicare la stessa cosa.

L'utilità di usare il rapporto tra gli odds, invece che la differenza delle probabilità, per confrontare due situazioni si ha in particolare quando i possibili valori delle probabilità da confrontare sono vicini a zero oppure vicini a 1. Vediamo un esempio. Supponiamo che un vaccino cambi dal 5 per mille all'1 per mille, ossia da 0,005 a 0,001, la probabilità di un esito infausto di una malattia. In altre parole, il vaccino riduce a un quinto i casi infausti. Questo può essere un buon risultato, se il vaccino ha bassi effetti collaterali. In questo caso il rapporto tra gli odds è appunto circa un quinto e cattura questa informazione, ma la differenza tra le probabilità è 0,004, che sembra quasi non apprezzabile.

Inoltre è importante comprendere l'uso che si fa del concetto di odds nei modelli logit. Ad esempio supponiamo di voler studiare l'andamento della probabilità ***p*** che uno studente abbia successo nella carriera universitaria a seconda del punteggio **t** in un test.
Per approssimare una funzione è spesso efficace usare un modello lineare. Però le funzioni lineari variano tra $-\infty$ e $+\infty$ e quindi non possono approssimare in modo soddisfacente la funzione $t \to p(t)$, dove p varia tra zero e 1, o come la funzione $t \to odds(t)$, dove gli odds variano tra zero e $+\infty$. Ma **il logaritmo degli odds**, che denoteremo con la lettera ***x***, varia tra $-\infty$ e $+\infty$ e ci si può chiedere se è possibile trovare dei parametri a, b tali che la funzione lineare

$$t \to x = at + b$$

approssima ragionevolmente[2] i dati sperimentali. D'altra parte, se chiamiamo y il logaritmo degli odds, ossia

$$y = \log(\text{odds}) \qquad (2)$$

abbiamo ovviamente

$$\text{odds} = \exp(y)$$

e, ricordando che $p = \frac{odds}{1+ odds}$, otteniamo[3]

$$p = \frac{\exp(x)}{1+\exp(x)}$$

ossia

$$p(t) = = \frac{\exp(at+b)}{1+\exp(at+b)} ,$$

E questo è un modello logit che esprime la probabilità p (t) di avere successo per uno studente che ha punteggio t nel test.

[2] Questo si può trattare come un problema di regressione lineare, col metodo dei minimi quadrati.
[3] La funzione che manda x in exp(x)/(1+exp(x)) si chiama "logistica" e ha numerose applicazioni. In particolare si incontra come soluzione delle equazioni differenziali che sono state utilizzate all'inizio dell'800 da Pierre François Verhulst per modellizzare la dinamica delle popolazioni.