Esercizio 7:
Affermazione 1: l'80% delle vittime stradali sono pedoni o motociclisti
Affermazione 2: l'87% degli incidenti mortali è causato da chi ha più di 25 anni

Traduzione della prima affermazione in probabilità:
$P(vittima = pedone \cup vittima = motociclista \mid incidente) = 0.80$
Traduzione della seconda affermazione in probabilità:
$P(età > 25 \mid causa\ di\ incidente\ mortale) = 0.87$ → seconda opzione tra le 3 proposte

Incidente	età $>$ 25	età $\leq$ 25	Totale
Mortale			
Non mortale			
Totale	87%		100%

In questo caso non ha senso secondo me parlare di variabile di risposta, in quanto entrambi i fattori (incidente mortale/non mortale e età maggiore o minore di 25) sono semplicemente dati oggettivi, e quindi variabili esplicative.
Il valore 87% è stato ottenuto tramite il seguente ragionamento:
$P(affermazione\ 2) = \frac{numero\ di\ incidenti\ provocati\ da\ persone\ di\ età > 25}{numero\ di\ incidenti\ totali}$
Il valore 87% è da collocare nella cella "Totale/età > 25", mentre il valore 100% è da collocare nella cella del totale globale.

Esercizio 9:
Specificare cosa succede agli indici D, RR, OR nel caso in cui:
- vengano scambiate le categorie della variabile X
- vengano scambiate X e Y

a. nel caso in cui vengano scambiate le categorie della variabile X si ha:
- $D = D,\ RR = RR,\ OR = OR$

b. nel caso in cui vengano scambiate le due variabili X e Y si ha:
- $D = -D,\ RR = \frac{1}{RR},\ OR = \frac{1}{OR}$

Esercizio 11:
Nel caso posto dall'esercizio 11, abbiamo uno schema di campionamento retrospettivo. Essendo stata casuale la determinazione del trattamento offerto al paziente i-esimo, non sono fissati i totali di riga. Lo scopo dello studio era quello di determinare se l'assunzione costante di aspirina riducesse i problemi cardiovascolari, e questo mi porta a pensare che ci troviamo in una situazione "caso/controllo".

Esercizio 12:
Nel caso posto dall'esercizio 12 abbiamo uno schema di campionamento di tipo casuale semplice, in quanto è stato posto un quesito a n partecipanti totali, e la suddivisione di a, b, c, e d (le 4 celle della tabella in esame) sono semplicemente variabili.

Esercizio 13:
Nel caso proposto dall'esercizio 13 abbiamo uno schema di campionamento di tipo retrospettivo, in quanto si è voluto analizzare la correlazione tra il linfoma di Hodgkin e l'operazione alle tonsille. Non sono fissati i totali di riga, bensì sono fissati quelli di colonna, in quanto il numero di casi di linfoma (assente o presente) preso in esame è noto. Anche qui ci troviamo in una situazione "caso/controllo".

Esercizio 14:

$$RR_{11} = \frac{p_1}{p_2} = \frac{\frac{R_1}{n_1}}{\frac{R_2}{n_2}} = \frac{\frac{189}{11034}}{\frac{104}{11037}} = 1.82$$

$$RR_{12} = \frac{p_1}{p_2} = \frac{\frac{R_1}{n_1}}{\frac{R_2}{n_2}} = \frac{\frac{227}{359}}{\frac{107}{785}} = 4.64$$

$$RR_{13} = \frac{p_1}{p_2} = \frac{\frac{R_1}{n_1}}{\frac{R_2}{n_2}} = \frac{\frac{90}{255}}{\frac{84}{391}} = 1.64$$

$$OR_{11} = \frac{189 \cdot 10933}{104 \cdot 10845} = 1.83$$

$$OR_{12} = \frac{227 \cdot 678}{107 \cdot 132} = 10.89$$

$$OR_{13} = \frac{90 \cdot 307}{84 \cdot 165} = 1.99$$

Nei casi 11 e 13, la misura di RR non è significativa in quanto il campionamento è retrospettivo.
Ora calcolo gli intervalli di confidenza dell'OR:

$$SE(log(OR_{11})) = \sqrt{\frac{1}{189} + \frac{1}{104} + \frac{1}{10845} + \frac{1}{10933}} = 0.123$$

$$IC_{11} = log(OR_{11}) \pm Z_{\frac{\alpha}{2}} * SE(log(OR_{11})) = 0.60 \pm 1.96 * 0.123 = [e^{0.36}, e^{0.84}] = [1.43, 2.31]$$

$$SE(log(OR_{12})) = \sqrt{\frac{1}{227} + \frac{1}{107} + \frac{1}{132} + \frac{1}{678}} = 0.151$$

$$IC_{12} = log(OR_{12}) \pm Z_{\frac{\alpha}{2}} * SE(log(OR_{12})) = 2.39 \pm 1.96 * 0.151 = [e^{2.09}, e^{2.69}] = [8.08, 14.73]$$

$$SE(log(OR_{13})) = \sqrt{\frac{1}{90} + \frac{1}{84} + \frac{1}{165} + \frac{1}{307}} = 0.179$$

$$IC_{13} = log(OR_{13}) \pm Z_{\frac{\alpha}{2}} * SE(log(OR_{13})) = 0.69 \pm 1.96 * 0.179 = [e^{0.39}, e^{1.04}] = [1.48, 2.83]$$

Gli intervalli costruiti possono essere applicati al rischio relativo solo nel caso dell'esempio 11, in quanto le probabilità p_1 e p_2 sono sufficientemente basse da fare in modo che: $RR \approx OR$

Utilizzo la statistica:

$$Z = \frac{\widehat{\psi}}{\sqrt{var(\widehat{\psi})}} = \frac{log(OR)-log(1)}{\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}}$$

$Z_{11} = 4.96$

$Z_{12} = 15.81$

$Z_{13} = 3.83$

Siccome tutti e tre i valori sono molto alti, possiamo ritenere il test significativo.