

# Statistica I

## Esercitazione 7: tabelle di contingenza

**Tommaso Rigon**

**Università Milano-Bicocca**



# Descrizione del problema



- Per verificare l'effetto della vitamina C sull'accrescimento dei maiali, a  $n = 30$  **maiali** sono state somministrate dalla nascita dosi diverse di acido ascorbico.
- Ad una età prefissata è stata poi misurata la **lunghezza media dei denti**, usata come una misura della crescita.

# Dati grezzi

## Lunghezza media dei denti in mm

```
[1] 15.2 21.5 17.6  9.7 14.5 10.0  8.2  9.4 16.5  9.7 19.7 23.3 23.6
[14] 26.4 20.0 25.2 25.8 21.2 14.5 27.3 25.5 26.4 22.4 24.5 24.8 30.9
[27] 26.4 27.3 29.4 23.0
```

## Dose di acido ascorbico in mg

```
[1] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 1.0 1.0 1.0 1.0 1.0 1.0
[17] 1.0 1.0 1.0 1.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0
```

## Domande

- Si costruisca la **tabella di contingenza** congiunta e si dica se esiste o non esiste **indipendenza** in distribuzione utilizzando l'indice  $\chi^2$ .
- Si calcolino le medie della lunghezza condizionate alla dose. Si dica se esiste **dipendenza in media**. Si ottenga il rapporto di correlazione  $\eta^2$ .
- Si determini la **retta di regressione** e la si disegni tramite grafico opportuno assieme ai dati. Si ottenga l'indice di bontà d'adattamento  $R^2$  e il coefficiente di correlazione  $\rho$ .

# Tabella di contingenza

	0.5	1	2	Totale
8.2	1	0	0	1
9.4	1	0	0	1
9.7	2	0	0	2
10	1	0	0	1
14.5	1	1	0	2
15.2	1	0	0	1
16.5	1	0	0	1
17.6	1	0	0	1
19.7	0	1	0	1
20	0	1	0	1
21.2	0	1	0	1
21.5	1	0	0	1
22.4	0	0	1	1
23	0	0	1	1
23.3	0	1	0	1
23.6	0	1	0	1
24.5	0	0	1	1
24.8	0	0	1	1
25.2	0	1	0	1
25.5	0	0	1	1
25.8	0	1	0	1
26.4	0	1	2	3
27.3	0	1	1	2
29.4	0	0	1	1
30.9	0	0	1	1

# Tabella frequenze attese

	0.5	1	2	Totale
8.2	0.33	0.33	0.33	1.00
9.4	0.33	0.33	0.33	1.00
9.7	0.67	0.67	0.67	2.00
10	0.33	0.33	0.33	1.00
14.5	0.67	0.67	0.67	2.00
15.2	0.33	0.33	0.33	1.00
16.5	0.33	0.33	0.33	1.00
17.6	0.33	0.33	0.33	1.00
19.7	0.33	0.33	0.33	1.00
20	0.33	0.33	0.33	1.00
21.2	0.33	0.33	0.33	1.00
21.5	0.33	0.33	0.33	1.00
22.4	0.33	0.33	0.33	1.00
23	0.33	0.33	0.33	1.00
23.3	0.33	0.33	0.33	1.00
23.6	0.33	0.33	0.33	1.00
24.5	0.33	0.33	0.33	1.00
24.8	0.33	0.33	0.33	1.00
25.2	0.33	0.33	0.33	1.00
25.5	0.33	0.33	0.33	1.00
25.8	0.33	0.33	0.33	1.00
26.4	1.00	1.00	1.00	3.00
27.3	0.67	0.67	0.67	2.00
29.4	0.33	0.33	0.33	1.00
30.9	0.33	0.33	0.33	1.00

- Pertanto si ottiene che

$$\chi^2 = \frac{(1 - 0.3333)^2}{0.3333} + \frac{(0 - 0.3333)^2}{0.3333} + \dots + \frac{(1 - 0.3333)^2}{0.3333} = 50.$$

- Di conseguenza, l'indice di connessione normalizzato è

$$\chi_{\text{norm}}^2 = \frac{50}{30 \min\{25 - 1, 3 - 1\}} = \frac{50}{60} = 0.83333.$$

- Tale risultato indica una forte **dipendenza in distribuzione**.

# Analisi della varianza

- Otteniamo ora la tabella delle **medie** e **varianze** condizionate.

	Numerosità	Media	Varianza	Devianza
0.5	10	13.23	17.90	179.00
1	10	22.70	13.77	137.66
2	10	26.06	6.34	63.44

- La media complessiva è pari a  $\bar{y} = 20.66333$ . Da questa tabella è quindi agevole calcolare la **devianza tra i gruppi**, ottenendo:

$$\mathcal{D}_{\text{tr}}^2 = 10(13.23 - 20.66)^2 + 10(22.70 - 20.66)^2 + 10(26.06 - 20.66)^2 = 885.265.$$

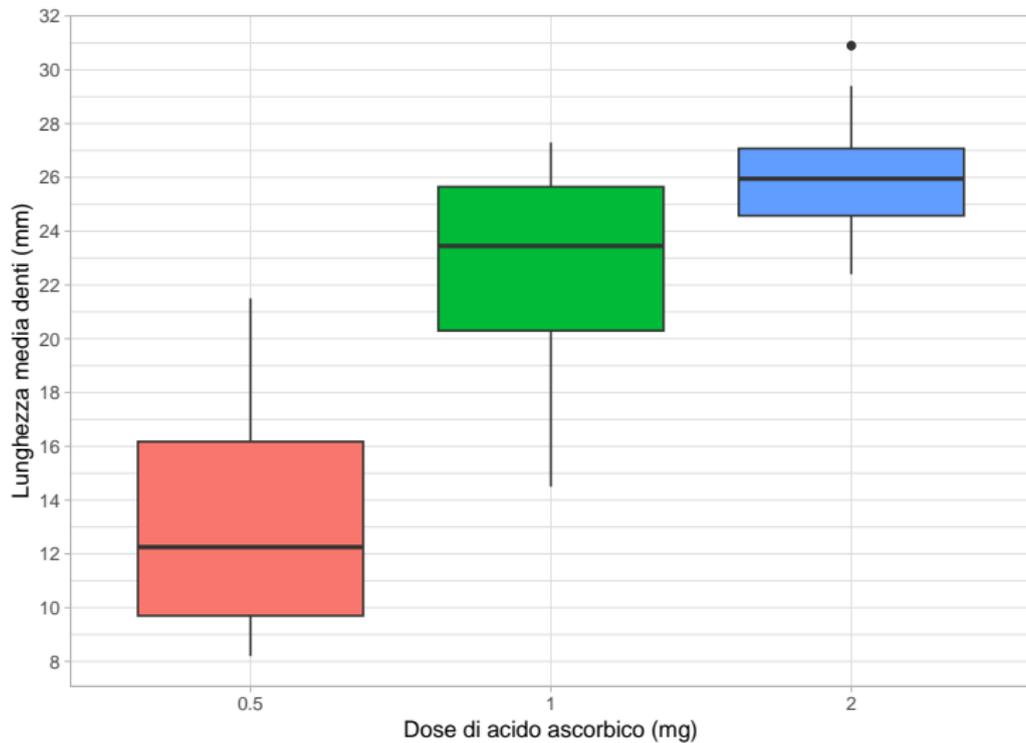
- Inoltre, la **devianza entro i gruppi** è pari a

$$\mathcal{D}_{\text{en}}^2 = 179.001 + 137.660 + 63.444 = 380.105,$$

da cui si ottiene che  $\mathcal{D}^2 = 885.265 + 380.105 = 1265.37$ . Il **rapporto di correlazione** è pertanto pari a

$$\eta^2 = \frac{\mathcal{D}_{\text{tr}}^2}{\mathcal{D}^2} = \frac{885.265}{1265.37} = 0.699.$$

# Boxplot



# Retta di regressione

- Otteniamo in primo luogo alcune quantità di interesse:

$$\sum_{i=1}^n x_i y_i = 814.35, \quad \sum_{i=1}^n x_i^2 = 52.5, \quad \sum_{i=1}^n y_i^2 = 14074.57,$$

- Inoltre, è possibile calcolare le **medie aritmetiche** delle due variabili precedenti:

$$\bar{x} = 1.166667, \quad \bar{y} = 20.66333.$$

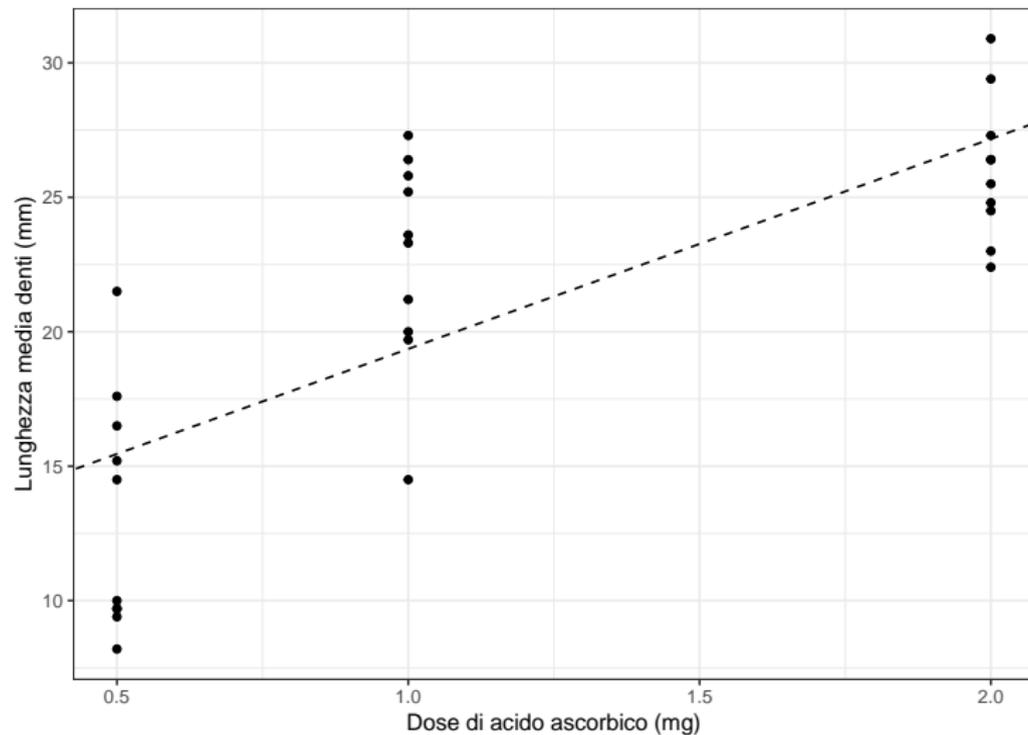
- Dalle precedenti quantità, è possibile ottenere che:

$$\text{cov}(x, y) = 3.037778, \quad \text{var}(x) = 0.3888889, \quad \text{var}(y) = 42.17899.$$

- Di conseguenza, otteniamo le seguenti **stime ai minimi quadrati**:  $\hat{\beta} = 7.811$  e  $\hat{\alpha} = \bar{y} - 7.811\bar{x} = 11.550$ .

- Inoltre, semplici calcoli portano a  $\rho = 0.75$  e quindi  $R^2 = \rho^2 = 0.5626$ .

# Diagramma a dispersione



# Commento ai risultati

- Tutti gli indicatori di dipendenza (connessione, rapporto di correlazione e correlazione di Pearson) evidenziano una **marcata dipendenza** tra dose e lunghezza dei denti.
- Tuttavia le conclusioni sono via via più dettagliate:
  - Indice  $\chi^2$  denota una differenza tra le tre distribuzioni “di qualsiasi tipo”.
  - Indice  $\eta^2$  denota una differenza tra le **medie** delle distribuzioni.
  - Indice  $\rho$  suggerisce che all'aumentare della dose la lunghezza cresce a sua volta di un determinato ammontare.
- Inoltre, il **modello lineare** si adatta abbastanza bene ai dati, però sembra **sottostimare** la lunghezza media dei denti per la dose pari a 1mg.
- Il modello lineare prevede che in assenza di acido ascorbico, la lunghezza media dei denti sarebbe di  $\hat{\alpha} = 11.550mm$ .