Problema 1

Il testo forniva la seguente tabella:

	A	B	C	D	E
$\bar{x}_g$	5.9	9.8667	$19.\bar{3}$	9.5	$3.\bar{6}$
σ_g	2.0224	3.1170	3.9016	1.9621	0.9428
n_g	10	15	12	10	6

Nota: le "deviazioni standard" sono pari a $\sigma_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ig} - \bar{x}_g)^2}$

a) Il teorema di scomposizione della devianza afferma che

$$D^2 = D^2_{en} + D^2_{tr}$$

avendo definito:

$$D^2 = \sum_{g=1}^{K}\sum_{i=1}^{n_g}(x_{ig} - \bar{x})^2; \quad D^2_{en} = \sum_{g=1}^{K}\sum_{i=1}^{n_g}(x_{ig} - \bar{x}_g)^2; \quad D^2_{tr} = \sum_{g=1}^{K} n_g(\bar{x}_g - \bar{x})^2$$

Nota: le quantità D^2, D^2_{en}, D^2_{tr} andavano opportunamente definite.

b)

$$\bar{x} = \frac{1}{n}\sum_{g=1}^{K} n_g \bar{x}_g = \frac{1}{53}\left(10 \cdot 5.9 + \ldots + 6 \cdot 3.\bar{6}\right) = 10.49$$

$$n = \sum_{g=1}^{K} n_g = 10 + 15 + \ldots + 6 = 53$$

c) Devianza entro i gruppi:

Attenzione a non dimenticarlo

$$D^2_{en} = \sum_{g=1}^{K}\sum_{i=1}^{n_g}(x_{ig} - \bar{x}_g)^2 = \sum_{g=1}^{K} n_g \sigma_g^2 = 10 \cdot 2.0224^2 + \ldots + 6 \cdot 0.948^2 = 413.1377$$

Devianza tra i gruppi

$$D_{tr} = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2 = 10 \cdot (5.9 - 10.49)^2 + \ldots + 6 \cdot (3.\bar{6} - 10.49)^2 =$$

$$= 1444.102$$

$$D^2 = D^2_{en} + D^2_{tr} = 1444.102 + 413.1377 = 1857.239$$

d) Il rapporto di correlazione è definito come

$$\eta^2 = \frac{D^2_{tr}}{D^2} = \frac{1444.102}{1857.239} = 0.7776$$

È quindi presente una forte dipendenza in media. In pratica, questo significa che le linee di produzione sono diverse tra loro in maniera significativa. In particolare, la linea C $(\bar{x}_3 = 19.6)$ è quella che produce quotidianamente il maggior numero di pezzi.

Problema 2

a) $$\bar{x} = \frac{1}{93} \sum_{i=1}^{n} x_i = \frac{340}{93} = 3.6559 \qquad \bar{y} = \frac{1}{93} \sum_{i=1}^{n} y_i = \frac{338}{93} = 3.634$$

$$\bar{z} = \frac{1}{93} \sum_{i=1}^{n} z_i = \frac{402}{93} = 4.3225.$$

Inoltre:

$$\bar{w} = \frac{1}{93} \sum_{i=1}^{n} (x_i + y_i) = \frac{1}{93} \sum_{i=1}^{n} x_i + \frac{1}{93} \sum_{i=1}^{n} y_i = \bar{x} + \bar{y} = 3.65 + 3.634 = 7.2903$$

b) Riporto solamente le formule di alcune varianze, covarianze, correlazioni.

Ad esempio:

Usando alta precisione

$$var(x) = \frac{1}{93}\sum_{i=1}^{m} x_i^2 - \bar{x}^2 = \frac{5780}{93} - 3.6558^2 = \boxed{48.78483}$$

$$var(y) = \frac{1}{93}\sum_{i=1}^{m} y_i^2 - \bar{y}^2 = \frac{5268}{93} - 3.634^2 = \boxed{43.43624}$$

$$cov(x,y) = \frac{1}{93}\sum_{i=1}^{m} x_i y_i - \bar{x}\bar{y} = \frac{5213}{93} - 3.6558 \cdot 3.634 = \boxed{42.7668}$$

$$cor(x,y) = \frac{cov(x,y)}{\sqrt{var(x)\, var(y)}} = \frac{42.7668}{\sqrt{43.43624 \cdot 48.78483}} = \boxed{0.9290456}$$

Proseguendo con i conti, si dovevano quindi ottenere tutte le varianze, covarianze e correlazioni. Nota: il testo chiedeva le matrici, non una lista di varianze/covarianze.

Matrice di covarianza

		Oro	Argento	Bronzo
(x)	Oro	48.78483	"	"
(y)	Argento	42.7668	43.43624	"
(z)	Bronzo	37.2077	35.11793	38.15401

Matrice di correlazione

	Oro	Argento	Bronzo
Oro	1	"	"
Argento	0.9290456	1	"
Bronzo	0.862427	0.862643	1

c) $var(w) = var(x+y) = \underbrace{var(x) + var(y) + 2cov(x,y)}$

Bastava guardare la tabella $= 48.78483 + 43.43624 + 2 \cdot 42.7668 = 177.7547$

d) Si consideri il modello statistico:

$$X_i = \alpha + \beta z_i + \varepsilon_i$$

Allora la stima ai minimi quadrati per $\hat{\beta}$ è:

$$\hat{\beta} = \frac{cov(x,z)}{var(z)} = \frac{37.2077}{38.15401} = \boxed{0.9751995};$$

Nota: ovviamente al denominatore va posto var(z), non var(x)! Le formule non vanno imparate a memoria!

Inoltre:

$$\hat{\alpha} = \bar{x} - \hat{\beta}\bar{z} = 3.6559 - 0.9751995 \cdot 4.3225 = \boxed{-0.5595}$$

e) Alcuni commenti importanti (ulteriori commenti erano ben accetti):

- $\hat{\alpha}$ è negativo, questo significa che il modello prevede un numero negativo di medaglie d'oro se $z = 0$. Tuttavia $\alpha \approx 0$, per cui questo effetto è trascurabile in pratica.

- $\hat{\beta} \approx 1$, pertanto per ogni medaglia di bronzo vinta il modello prevede circa lo stesso numero di medaglie d'oro. Questo è in linea con le aspettative: i paesi tendono ad avere un numero di medaglie abbastanza omogeneo tra le varie tipologie (oro, argento, bronzo).

f) $$R^2 = \rho^2 = 0.862427^2 = 0.7438$$ (Ottimo adattamento ai dati)

$$var(r) = (1 - R^2)\,var(x) = (1 - 0.7438)\,48.78483 = 12.49983$$

g) Il valore previsto dal modello per l'Italia è:

$$\hat{x}_i = \hat{\alpha} + \hat{\beta} \cdot 20 = 18.9449$$

$$\approx 19$$

Nota: il modello non prevede necessariamente un numero intero

Pertanto il residuo corrispondente è:

$$r_i = x_i - \hat{x}_i = 10 - 18.9449 = -8.9449$$

Problema 3

moltiplico e divido per $\tilde{\omega}_m$

$$\bar{x}_{m+1} = \frac{1}{\tilde{\omega}_{m+1}} \sum_{i=1}^{m+1} \omega_i x_i = \frac{1}{\tilde{\omega}_{m+1}} \sum_{i=1}^{m} \omega_i x_i + \frac{1}{\tilde{\omega}_{m+1}} \cdot \omega_{m+1} \cdot x_{m+1}$$

$$= \frac{\tilde{\omega}_m}{\tilde{\omega}_{m+1}} \bar{x}_m + \frac{\omega_{m+1}}{\tilde{\omega}_{m+1}} x_{m+1}$$

Con i dati a disposizione:

$$\bar{x}_{m+1} = \frac{72}{72+12} \cdot 25 + \frac{12}{72+12} \cdot 30 = 25.71429$$