



UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA TRIENNALE IN
STATISTICA PER LE TECNOLOGIE E LE SCIENZE

STATISTICA MEDICA

EFFICIENZA DEL LAVORO MUSCOLARE

PROFF.SSA

LAURA VENTURA

AUTORE

ALESSANDRO FABRIS

MATRICOLA

1169093

1 Introduzione

In uno studio per valutare l'efficienza del lavoro muscolare vengono misurate le calorie consumate in seguito all'attività fisica svolta su una bicicletta stazionaria. Per ogni soggetto è inoltre nota la misura della massa corporea. L'obiettivo dello studio è modellare il consumo calorico in funzione della massa corporea e dell'intensità del lavoro compiuto. I dati utilizzati fanno riferimento allo studio “On the Efficiency of Muscular Work” di M. Greenwood (1918) pubblicato nella Royal Society [1].

Le analisi sono state eseguite con il *software* **R** nella versione 4.2.2. Il livello di significatività è fissato al 5%. Per tutti gli approfondimenti sui test utilizzati in questa analisi si rimanda a “Biostatistica” di Ventura e Racugno [2].

Lo studio è stato svolto su un campione di 24 soggetti, per ciascuno dei quali sono state rilevate:

- **Body Mass** (*massa corporea*) misurata in *kg*;
- **Work Level** (*intensità di lavoro*) misurata in *cal/ora*;
- **Heat Output** (*consumo calorico*) misurata in *cal*.

2 Analisi esplorative

Nelle analisi esplorative viene svolta una preliminare descrizione delle informazioni date dalle variabili osservate. Si individuano inoltre le relazioni tra coppie di tali variabili mediante opportuni test statistici.

2.1 Analisi univariate

2.1.1 Body Mass

La **Body Mass** *media* è di 57.54 *kg* mentre la *mediana* è di 58.80 *kg*, il *valore minimo* e il *valore massimo* sono, rispettivamente, pari a 43.70 *kg* e 66.70 *kg*. La *deviazione standard* è pari a 6.59 *kg* mentre lo *scarto interquartile* è pari a 7.30 *kg*.

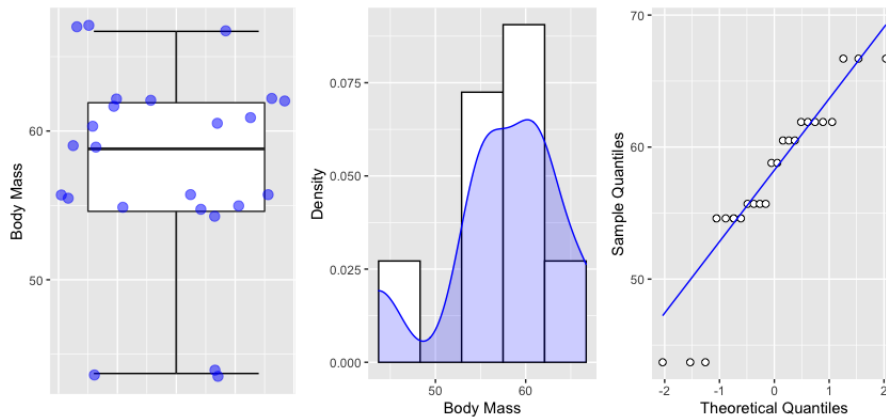


Figura 1: Boxplot a sinistra, istogramma al centro con densità stimata e diagramma quantile-quantile a destra per la variabile **Body Mass**.

Dai 3 grafici in Figura 1 si nota una leggera asimmetria negativa (confermata dal fatto che la media è leggermente inferiore alla mediana) e dalla presenza di una coda pesante a sinistra. L'ipotesi di normalità, valutata attraverso la statistica test di Shapiro-Wilk e il diagramma quantile-quantile (Figura 1 a destra), viene pertanto rifiutata con un alpha fissato a 0.05 ($W = 0.88005$, $p\text{-value} = 0.00833$).

2.1.2 Work Level

Il **Work Level** *media* è di 34.04 *cal/ora* mentre la *mediana* è di 38.75 *cal/ora*, il *valore minimo* e il *valore massimo* sono, rispettivamente, pari a 13.00 *cal/ora* e 56.00 *cal/ora*.

La *deviazione standard* è pari a 16.36 *cal/ora* mentre lo *scarto interquartile* è pari a 24.00 *cal/ora*.

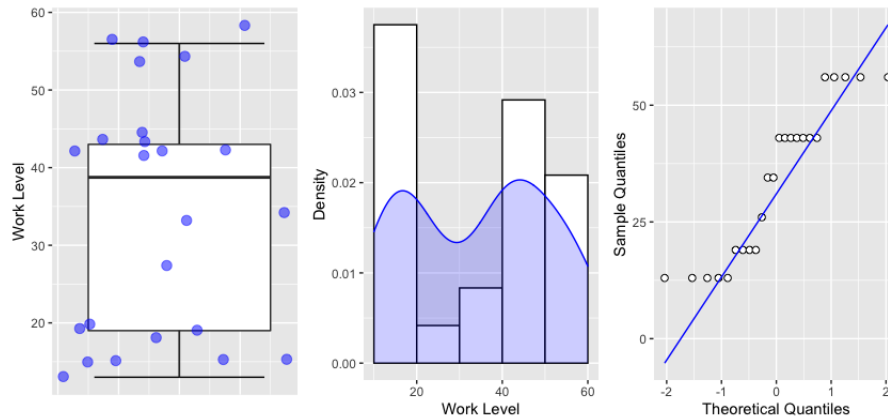


Figura 2: Boxplot a sinistra, istogramma al centro con densità stimata e diagramma quantile-quantile a destra per la variabile **Work Level**.

Dai 3 grafici in Figura 2 si nota che la distribuzione non è simmetrica. L'ipotesi di normalità, valutata attraverso la statistica test di Shapiro-Wilk e il diagramma quantile-quantile (Figura 2 a destra), viene pertanto rifiutata con un alpha fissato a 0.05 ($W = 0.86019$, $p\text{-value} = 0.003399$).

2.1.3 Heat Output

L' **Heat Output** *media* è di 260.0 *cal* mentre la *mediana* è di 272.0 *cal*, il *valore minimo* e il *valore massimo* sono, rispettivamente, pari a 160.0 *cal* e 352.0 *cal*. La *deviazione standard* è pari a 65.9 *cal* mentre lo *scarto interquartile* è pari a 113.8 *cal*.

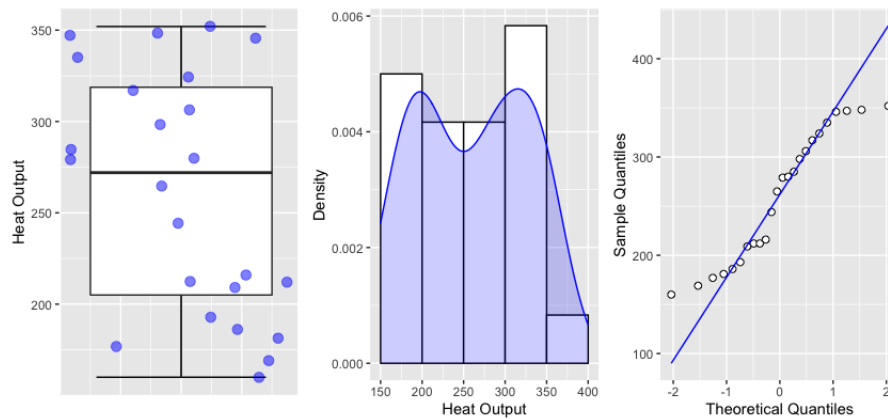


Figura 3: Boxplot a sinistra, istogramma al centro con densità stimata e diagramma quantile-quantile a destra per la variabile **Heat Output**.

Dai 3 grafici in Figura 3 si nota che la distribuzione è simmetrica e nel diagramma quantile-quantile (Figura 3 a destra) sembra ci siano delle code pesanti.

L'ipotesi di normalità, valutata attraverso la statistica test di Shapiro-Wilk e il diagramma quantile-quantile (Figura 3 a destra), viene pertanto rifiutata, non di molto, con un alpha fissato a 0.05 ($W = 0.91176$, $p\text{-value} = 0.03851$).

2.2 Analisi bivariate

2.2.1 Body Mass vs Work Level

Per valutare l'effetto marginale della variabile **Body Mass** sulla variabile **Work Level** viene riportato il diagramma di dispersione (Figura 4).

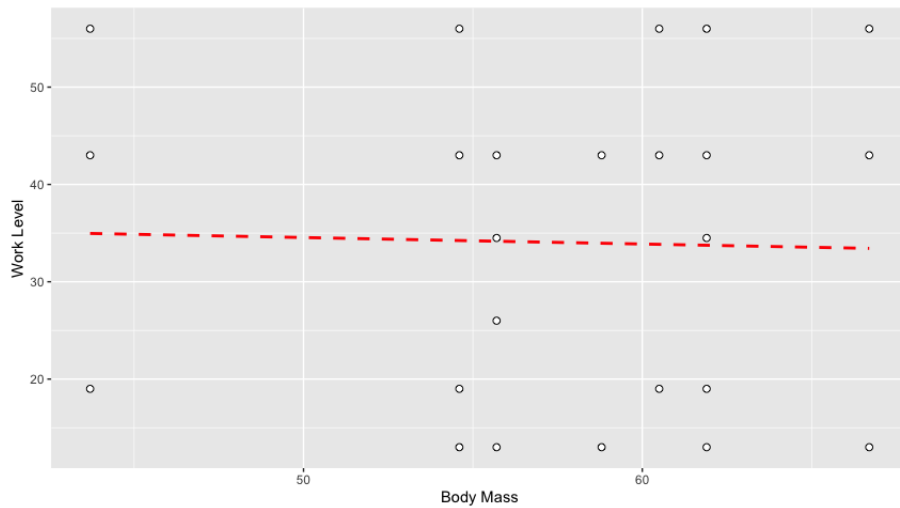


Figura 4: Diagramma di dispersione tra le variabili **Body Mass** e **Work Level**. In rosso retta ai minimi quadrati.

Dal grafico in Figura 4 non sembrano esserci andamenti sistematici e non sembra nemmeno esserci una relazione tra le variabili **Body Mass** e **Work Level**. L'indice di correlazione non parametrico di Spearman evidenzia che non vi è correlazione tra le due variabili ($\rho = -0.004054354$, $p\text{-value} = 0.985$).

2.2.2 Body Mass vs Heat Output

Per valutare l'effetto marginale della variabile **Body Mass** sulla variabile **Heat Output** viene riportato il diagramma di dispersione (Figura 5).

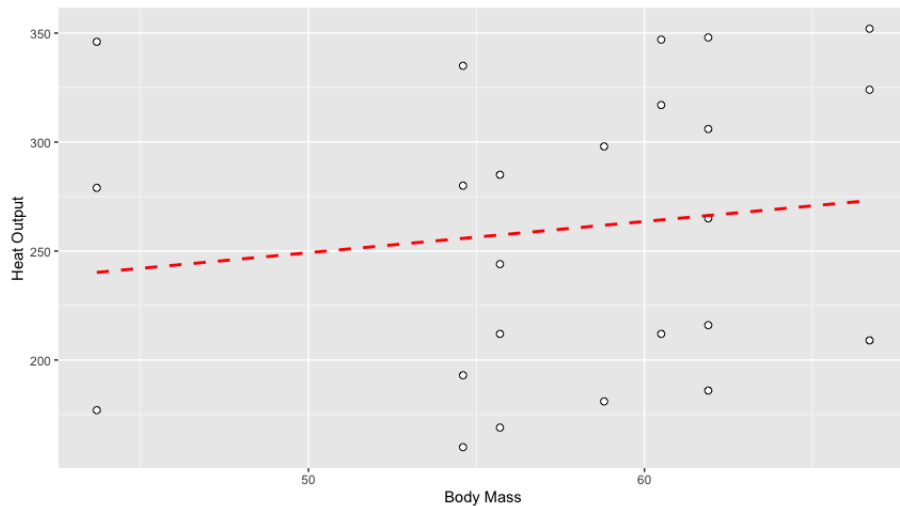


Figura 5: Diagramma di dispersione tra le variabili **Body Mass** e **Heat Output**. In rosso retta ai minimi quadrati.

Dal grafico in Figura 5 non sembrano esserci andamenti sistematici e non sembra nemmeno esserci una relazione tra le variabili **Body Mass** e **Heat Output**. L'indice di correlazione non parametrico di Spearman evidenzia che non vi è correlazione tra le due variabili ($\rho = 0.2611258$, p-value = 0.2178).

2.2.3 Work Level vs Heat Output

Per valutare l'effetto marginale della variabile **Work Level** sulla variabile **Heat Output** viene riportato il diagramma di dispersione (Figura 6).

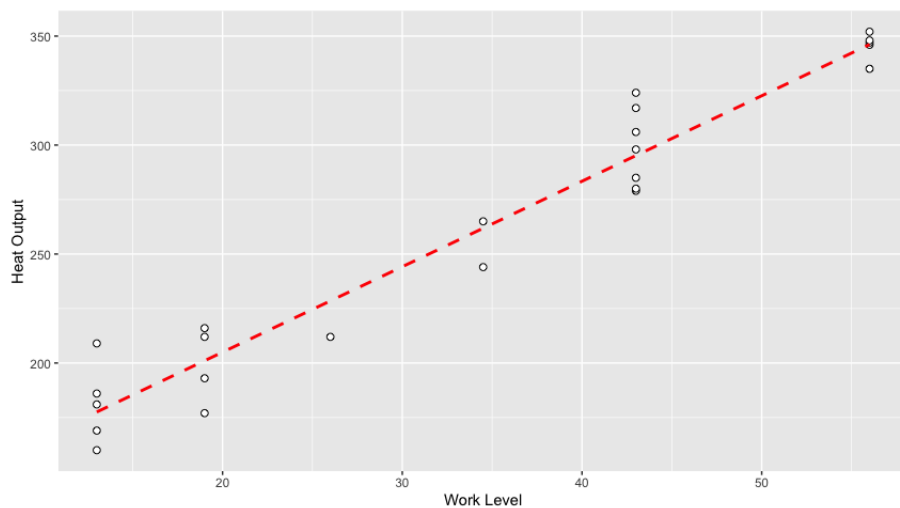


Figura 6: Diagramma di dispersione tra le variabili **Work Level** e **Heat Output**. In rosso retta ai minimi quadrati.

Dal grafico in Figura 6 sembra esserci un andamento sistematico con una relazione crescente tra le variabili **Body Mass** e **Heat Output** (infatti la retta stimata ai minimi quadrati ha coefficiente angolare stimato pari a 3.9212). L'indice di correlazione non parametrico di Spearman evidenzia che c'è correlazione tra le due variabili ($\rho = 0.9573013$, p-value = 2.432e-13).

3 Stima dei modelli

Per modellare l'**Heat Output** (*consumo calorico*) in funzione della **Body Mass** (*massa corporea*) e del **Work Level** (*intensità di lavoro*), si adattano due diversi modelli di regressione sulla base di studi precedentemente svolti.

3.1 Modello lineare

Come primo modello viene adattato un modello di regressione lineare normale multipla che assume l'indipendenza fra tutte le misurazioni.

$$H_i = \alpha_0 + \alpha_1 M_i + \alpha_2 W_i + \epsilon_i \text{ con } \epsilon_i \sim N(0, \sigma^2) \text{ indipendenti}$$

dove H_i rappresenta l'**Heat Output** (*consumo calorico*) per l'*i-esimo* soggetto, M_i rappresenta la **Body Mass** (*massa corporea*) per l'*i-esimo* soggetto e W_i rappresenta il **Work Level** (*intensità di lavoro*) per l'*i-esimo* soggetto, per $i = 1, \dots, 24$. $\alpha = (\alpha_0 \ \alpha_1 \ \alpha_2)^T$ è il vettore dei parametri ignoti di regressione e ϵ_i rappresenta il termine di errore.

La tabella di adattamento del modello risulta:

Parametro	Stima	SE	t value	p-value
α_0	28.3126	20.0806	1.410	0.173
α_1	1.6965	0.3355	5.057	5.24e-05 ***
α_2	3.9395	0.1351	29.153	< 2e-16 ***

Tutte le stime dei parametri del modello, ad eccezione di α_0 (*l'intercetta*), sono significative (con il livello pre-fissato pari a 0.05).

Il modello stimato risulta quindi il seguente:

$$H = 28.3126 + 1.6965M + 3.9395W$$

- Effetto della variabile **Body Mass**: fissata la variabile **Work Level** con il valore medio, pari a 34.04 *cal/ora*, il *consumo calorico* (**Heat Output**) stimato aumenta di 1.6965 *cal* per ogni incremento unitario (1 *kg*) di *cal*.
- Effetto della variabile **Work Level**: fissata la variabile **Body Mass** con il valore medio, pari a 57.54 *kg*, il *consumo calorico* (**Heat Output**) stimato aumenta di 3.9395 *cal* per ogni incremento unitario (1 *cal/ora*) di *cal*.

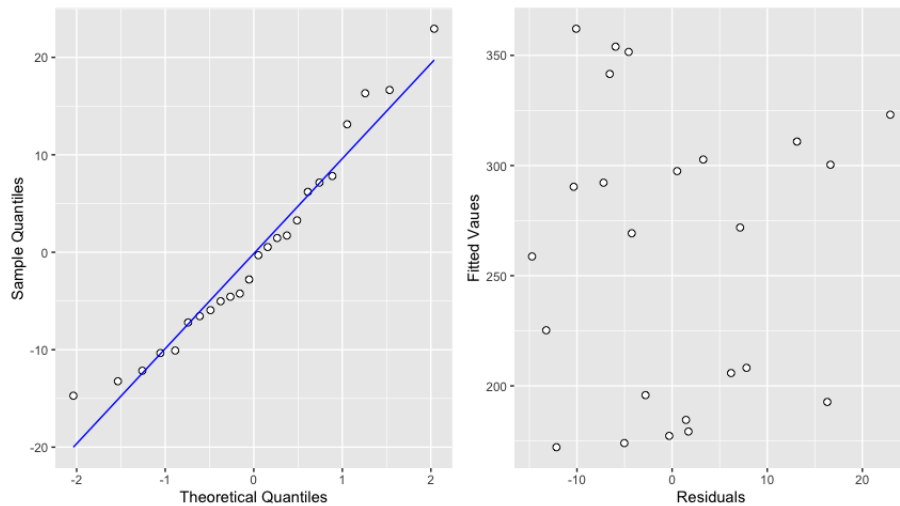


Figura 7: Analisi dei residui del modello di regressione lineare multipla.

Sia il diagramma quantile-quantile (Figura 7 a sinistra) sia il test di normalità di Shapiro-Wilk ($W = 0.95349$, $p\text{-value} = 0.3219$) confermano l'ipotesi di normalità per i residui studentizzati del modello stimato. Il grafico dei residui studentizzati rispetto ai valori stimati (Figura 7 a destra) non mostra andamenti sistematici pertanto l'ipotesi di omoschedasticità non viene rifiutata.

Per valutare la bontà di adattamento del modello stimato possiamo confrontare i valori osservati della variabile **Heat Output** rispetto ai valori stimati con il modello (Figura 8).

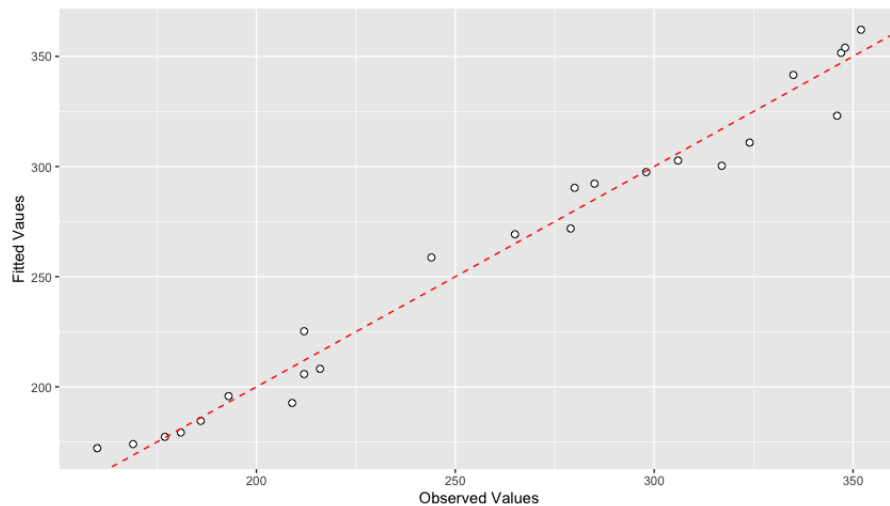


Figura 8: Diagramma di dispersione tra i valori osservati della variabile che misura l'**Heat Output** e i valori stimati con il modello lineare normale

multiplo. La bisettrice è in colore rosso e tratteggiata.

Dal grafico si nota che i punti sono disposti lungo la bisettrice indicando un buon adattamento del modello ai dati. Infatti non si notano sottostime e sovrastime sistematiche.

3.2 Modello non lineare

Come secondo modello viene adattato un modello di regressione non lineare nei parametri. Si prende a riferimento il modello teorico proposto da Glazebrook e Dye [3].

$$H_i = \beta_0 + \beta_1 M_i + \frac{W_i}{\beta_2 + \beta_3 M_i} + \epsilon_i \text{ con } \epsilon_i \sim N(0, \sigma^2) \text{ indipendenti}$$

dove H_i rappresenta l'**Heat Output** (*consumo calorico*) per l'*i-esimo* soggetto, M_i rappresenta la **Body Mass** (*massa corporea*) per l'*i-esimo* soggetto e W_i rappresenta il **Work Level** (*intensità di lavoro*) per l'*i-esimo* soggetto, per $i = 1, \dots, 24$. $\beta = (\beta_0 \ \beta_1 \ \beta_2 \ \beta_3)^T$ è il vettore dei parametri ignoti di regressione e ϵ_i rappresenta il termine di errore. La tabella di adattamento del modello risulta:

Parametro	Stima	SE	t value	p-value
β_0	-117.0967	33.37	-3.509	0.00221 **
β_1	4.221904	0.5755	7.336	4.33e-07 ***
β_2	0.03119462	0.04104	0.760	0.45608
β_3	0.003925117	0.000758	5.178	4.57e-05 ***

Si osserva che la stima del parametro β_2 non è significativa (rispetto al livello pre-fissato pari a 0.05).

Il modello stimati risulta quindi il seguente:

$$H = -117.0967 + 4.221904M + \frac{W}{0.03119462 + 0.003925117M}$$

- Effetto della variabile **Body Mass**: fissata la variabile **Work Level** con il valore medio, pari a 34.04 *cal/ora*, il *consumo calorico* stimato è non lineare al variare della massa corporea e per un'intensità di lavoro fissata.
- Effetto della variabile **Work Level**: fissata la variabile **Body Mass** con il valore medio, pari a 57.54 *kg*, il *consumo calorico* (**Heat Output**) stimato aumenta di 3.890356 *cal* per ogni incremento unitario (1 *cal/ora*) di *cal*.

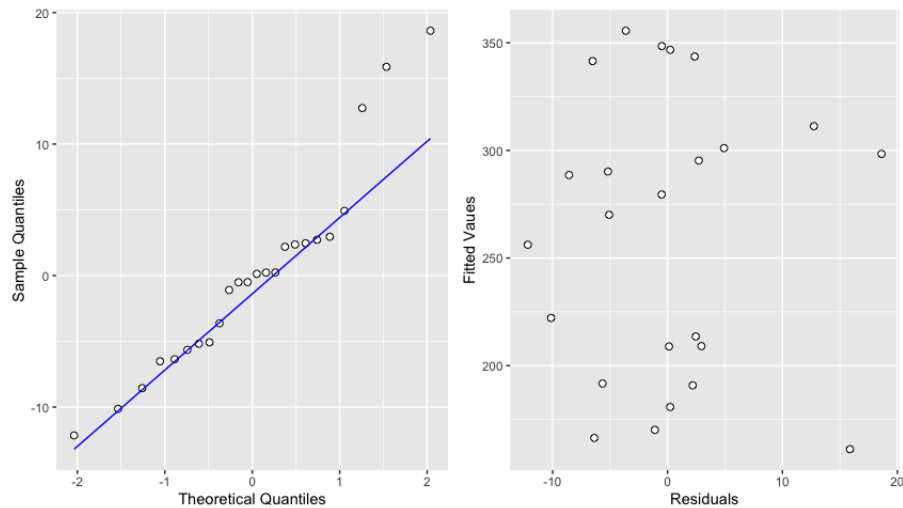


Figura 9: Analisi dei residui del modello non lineare.

Sia il diagramma quantile-quantile (Figura 9 a sinistra) sia il test di normalità di Shapiro-Wilk ($W = 0.92441$, $p\text{-value} = 0.07316$) confermano l'ipotesi di normalità per i residui studentizzati del modello stimato. I tre valori nella coda superiore che si discostano da tale retta non portano quindi al rifiuto della normalità. Il grafico dei residui studentizzati rispetto ai valori stimati (Figura 9 a destra) non mostra andamenti sistematici pertanto l'ipotesi di omoschedasticità non viene rifiutata.

Per valutare la bontà di adattamento del modello stimato possiamo confrontare i valori osservati della variabile **Heat Output** rispetto ai valori stimati con il modello (Figura 10).

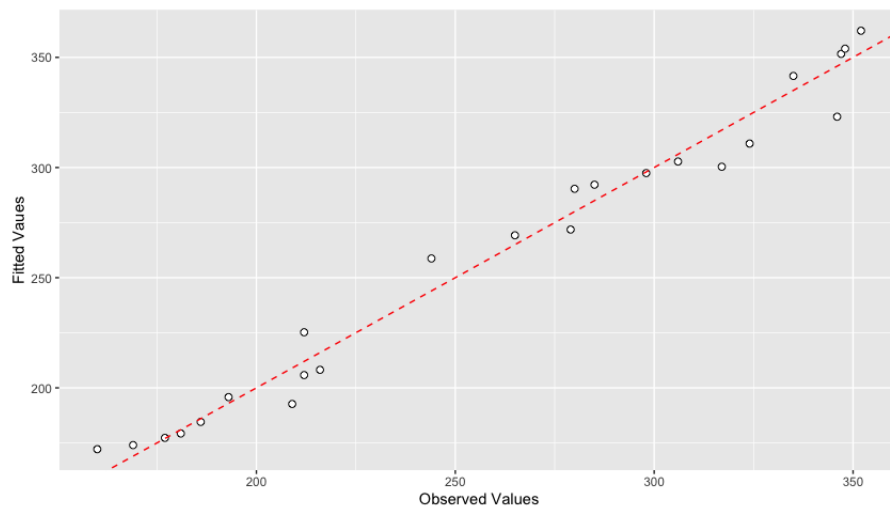


Figura 10: Diagramma di dispersione tra i valori osservati della variabile che misura l'**Heat Output** e i valori stimati con il modello lineare normale multiplo. La bisettrice è in colore rosso e tratteggiata.

Dal grafico si nota che i punti sono disposti lungo la bisettrice indicando un buon adattamento del modello ai dati. Infatti non si notano sottostime e sovrastime sistematiche.

4 Ulteriori modelli

Come ulteriore modello viene adattato un modello di regressione lineare normale multipla introducendo un termine di interazione.

$$H_i = \gamma_0 + \gamma_1 M_i + \gamma_2 W_i + \gamma_3 M_i W_i + \epsilon_i \text{ con } \epsilon_i \sim N(0, \sigma^2) \text{ indipendenti}$$

dove H_i rappresenta l'**Heat Output** (*consumo calorico*) per l'*i-esimo* soggetto, M_i rappresenta la **Body Mass** (*massa corporea*) per l'*i-esimo* soggetto e W_i rappresenta il **Work Level** (*intensità di lavoro*) per l'*i-esimo* soggetto, per $i = 1, \dots, 24$. $\gamma = (\gamma_0 \ \gamma_1 \ \gamma_2 \ \gamma_3)^T$ è il vettore dei parametri ignoti di regressione e ϵ_i rappresenta il termine di errore.

La tabella di adattamento del modello risulta:

Parametro	Stima	SE	t value	p-value
γ_0	-85.74896	42.37416	-2.024	0.05658 .
γ_1	3.66528	0.72758	5.038	6.30e-05 ***
γ_2	6.95045	1.02897	6.755	1.43e-06 ***
γ_3	-0.05200	0.01766	-2.945	0.00801 **

Tutte le stime dei parametri del modello, ad eccezione di γ_0 (*l'intercetta*), sono significative (con il livello pre-fissato pari a 0.05).

Il modello stimato risulta quindi il seguente:

$$H = -85.74896 + 3.66528M + 6.95045W - 0.05200MW$$

- Effetto della variabile **Body Mass**: fissata la variabile **Work Level** con il valore medio, pari a 34.04 *cal/ora*, il *consumo calorico* (**Heat Output**) stimato aumenta di 1.8952 *cal* per ogni incremento unitario (1 *kg*) di *cal*.
- Effetto della variabile **Work Level**: fissata la variabile **Body Mass** con il valore medio, pari a 57.54 *kg*, il *consumo calorico* (**Heat Output**) stimato aumenta di 3.95837 *cal* per ogni incremento unitario (1 *cal/ora*) di *cal*.

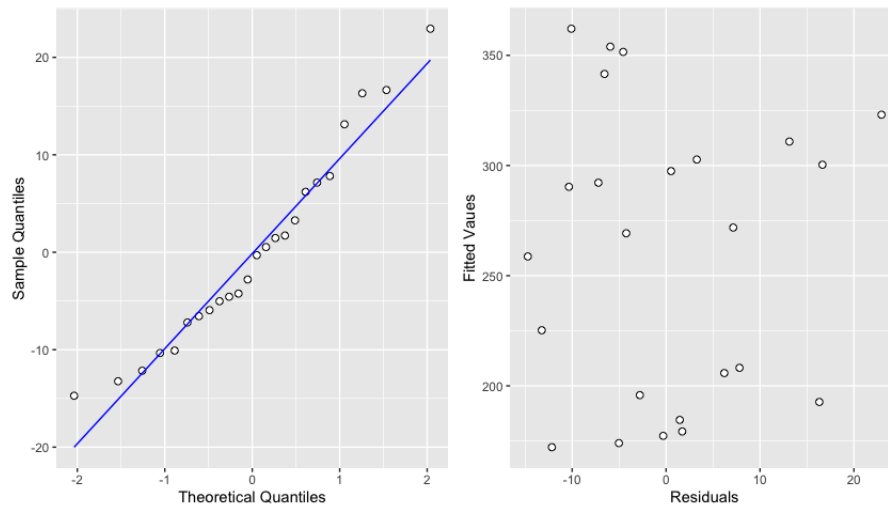


Figura 11: Analisi dei residui del modello di regressione lineare multipla.

Sia il diagramma quantile-quantile (Figura 11 a sinistra) sia il test di normalità di Shapiro-Wilk ($W = 0.96612$, $p\text{-value} = 0.5729$) confermano l'ipotesi di normalità per i residui studentizzati del modello stimato. Il grafico dei residui studentizzati rispetto ai valori stimati (Figura 9 a destra) non mostra andamenti sistematici pertanto l'ipotesi di omoschedasticità non viene rifiutata.

Per valutare la bontà di adattamento del modello stimato possiamo confrontare i valori osservati della variabile **Heat Output** rispetto ai valori stimati con il modello (Figura 12).

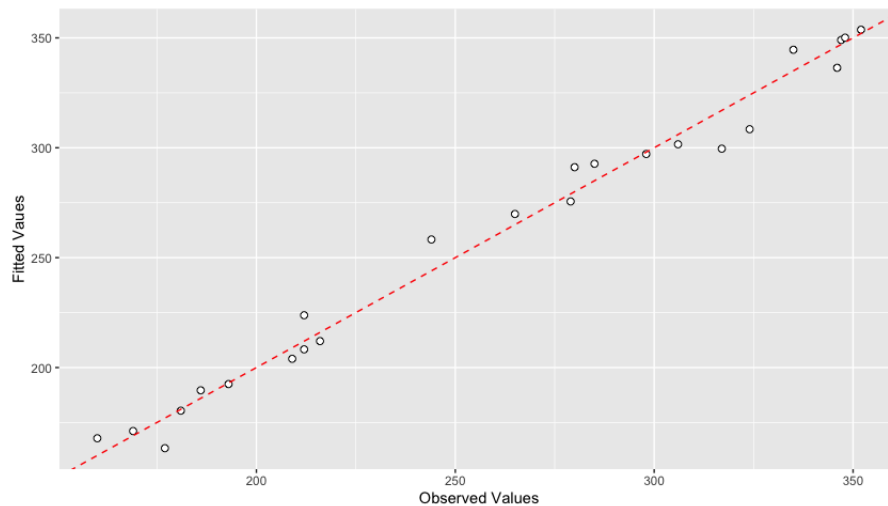


Figura 12: Diagramma di dispersione tra i valori osservati della variabile che misura l'**Heat Output** e i valori stimati con il modello lineare normale

multiplo. La bisettrice è in colore rosso e tratteggiata.

Dal grafico si nota che i punti sono disposti lungo la bisettrice indicando un buon adattamento del modello ai dati. Infatti non si notano sottostime e sovrastime sistematiche.

5 Conclusioni

Dai 3 modelli considerati si nota come l'**Heat Output** (*consumo calorico*) dipende significativamente sia dalla **Body Mass** (*massa corporea*) sia dal **Work Level** (*intensità di lavoro*). Tutti i modelli presentano intercetta negativa, stimano quindi *consumi calorici* negativi per valori molto piccoli della *massa corporea* e dell'*intensità di lavoro*, per questo non si ritiene adeguato il loro utilizzo per valori di massa e intensità distanti da quelli osservati. Confrontando i valori stimati dei 3 modelli considerati con i valori osservati, si nota come i modelli assumono valori molto simili (Figura 13).

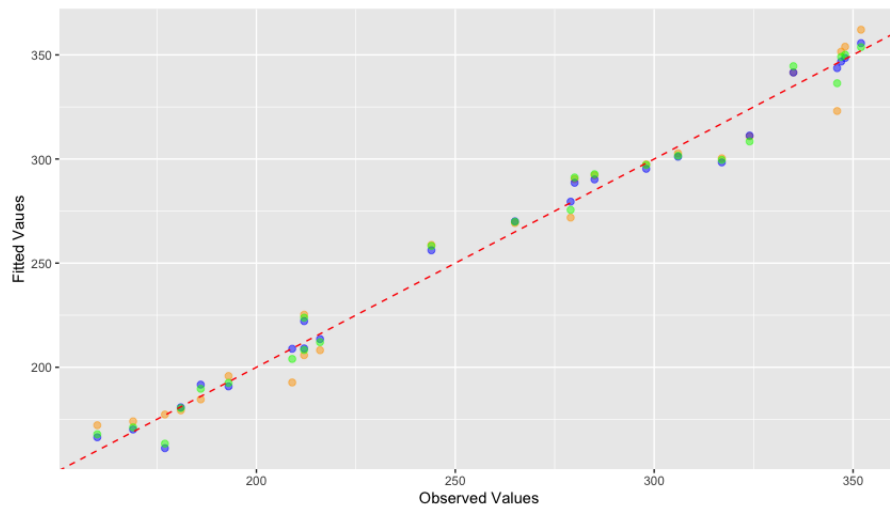


Figura 13: Diagramma di dispersione tra i valori osservati della variabile che misura l'**Heat Output** e i valori stimati dei 3 diversi modelli.

- Il modello di regressione lineare normale multipla senza interazione è di colore arancio.
- Il modello di regressione non lineare è di colore blu.
- Il modello di regressione lineare normale multipla con interazione è di colore verde.
- La bisettrice è in colore rosso e tratteggiata.

Confrontando gli AIC dei tre modelli si nota che il modello non lineare ha un AIC leggermente inferiore ($AIC = 174.3675$) rispetto al modello lineare normale multiplo senza interazione ($AIC = 186.2213$) e al modello lineare normale multiplo con interazione ($AIC = 179.5768$). Si preferisce pertanto, in base al criterio di Akaike, il modello non lineare.

Riferimenti bibliografici

- [1] M. Greenwood (1918), "*On the Efficiency of Muscular Work*"
- [2] L. Ventura, W. Racugno (2017), *Biostatistica*, Casi di studio in R. Egea
- [3] Ibid., vol.87, p. 311 (1914)