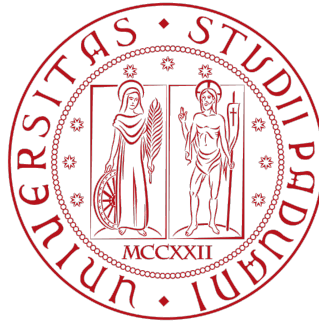


UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA TRIENNALE IN
STATISTICA PER L'ECONOMIA E L'IMPRESA



RELAZIONE FINALE
**VIOLAZIONE DELLE ASSUNZIONI DELL'ANALISI DELLA
VARIANZA: CONFRONTO CON METODI ALTERNATIVI
ATTRAVERSO UNO STUDIO DI SIMULAZIONE IN R**

Relatore: Prof. Erlis Ruli
Dipartimento di Scienze Statistiche

Laureanda : Elena Danese
Matricola 2038094

Anno Accademico 2023/2024

Gutta cavat lapidem non vi, sed saepe cadendo.

(Lucrezio)

Abstract

L'analisi della varianza (ANOVA) è una tecnica statistica ampiamente utilizzata per confrontare le medie di gruppi indipendenti. Tuttavia, chi utilizza metodi statistici deve prestare particolare attenzione alla validità delle assunzioni sottostanti, poiché l'affidabilità dei risultati dipende dal rispetto di tali assunzioni. Ignorare i presupposti può portare a conclusioni errate e decisioni sbagliate.

La seguente tesi esplora le conseguenze delle violazioni dei presupposti dell'ANOVA e, attraverso uno studio di simulazione condotto in R, propone metodi alternativi per gestirle.

In primo luogo, vengono analizzati gli effetti della violazione dell'assunzione di normalità, di indipendenza dei dati e di omoschedasticità. Per quest'ultima, sono esaminati test diagnostici che consentono di rilevare la presenza di eteroschedasticità, come il test di Cochran, di Bartlett, di Hartley, di Levene, di Brown-Forsythe e di Fligner-Killeen.

Per affrontare l'eteroschedasticità, vengono esplorati diversi metodi, tra cui la trasformazione dei dati, la versione non parametrica dell'ANOVA, ossia il test di Kruskal-Wallis, ma sono considerati anche il test di Welch, di Welch con medie troncate e varianze winsorizzate, di Alexander-Govern e di James del secondo ordine. Questi test sono valutati sulla loro efficacia nel gestire la non omogeneità delle varianze.

In presenza di dati dipendenti, sono discusse proposte alternative all'analisi della varianza tradizionale come l'ANOVA a misure ripetute, la versione non parametrica di quest'ultima, ossia il test di Friedman, e i modelli ad effetti misti, che offrono soluzioni affidabili per la gestione della non indipendenza dei dati.

Uno studio di simulazione è stato condotto per valutare l'impatto delle violazioni delle assunzioni sull'ANOVA e per confrontare l'efficacia dei metodi alternativi. Gli scenari considerati includono gruppi bilanciati o sbilanciati, con omoschedasticità o eteroschedasticità, e con osservazioni indipendenti o dipendenti. Inoltre, sono state distinte le situazioni in cui esiste una relazione positiva o negativa tra la numerosità interna ai gruppi e la loro varianza. Nei vari contesti, sono stati calcolati il tasso di errore di I tipo e la potenza dei test confrontati. Sono stati generati grafici per studiare, sotto l'ipotesi nulla, l'uniformità della distribuzione dei p-value e

per confrontare la distribuzione delle statistiche test simulate con quella teorica prevista.

I risultati indicano che le violazioni delle assunzioni possono compromettere significativamente la validità dell'ANOVA, influenzando negativamente l'accuratezza e la robustezza del test. Tuttavia, sono disponibili tecniche alternative che possono efficacemente mitigare questi effetti.

In conclusione, questa tesi offre una comprensione approfondita delle limitazioni dell'ANOVA quando i presupposti di omoschedasticità e indipendenza non sono rispettati e propone soluzioni efficaci per migliorare l'affidabilità delle analisi statistiche in tali condizioni.

Indice

1	ANALISI DELLA VARIANZA	1
1.1	ANOVA ad una via	2
1.1.1	Modello di regressione lineare multipla per il confronto di medie	5
2	CONSEGUENZE DELLA VIOLAZIONE DELLE ASSUNZIONI DELL'ANALISI DELLA VARIANZA	7
2.1	Violazione dell'assunzione di normalità	8
2.2	Violazione dell'assunzione di indipendenza	8
2.3	Violazione dell'assunzione di omoschedasticità	9
2.3.1	Test di Cochran	9
2.3.2	Test di Bartlett	10
2.3.3	Test di Hartley	11
2.3.4	Test di Levene	11
2.3.5	Test di Brown-Forsythe	12
2.3.6	Test di Fligner-Killeen centrato sulla mediana	12
3	TECNICHE PER AFFRONTARE L'ETEROSCHEDASTICITÀ NELL'ANALISI DELLA VARIANZA	15
3.1	Trasformazione dei dati	15
3.2	ANOVA non parametrica: test di Kruskal-Wallis	17
3.3	Test di Welch	18
3.4	Test di Welch con medie troncate e varianze winsorizzate	19
3.5	Test di Alexander-Govern	20
3.6	Test di James del secondo ordine	21
4	METODI PER LA GESTIONE DELLA DIPENDENZA DEI DATI NELL'ANALISI DELLA VARIANZA	23
4.1	ANOVA a misure ripetute	23

4.2	Test di Friedman	26
4.3	Modelli ad effetti misti per risposte normali	26
5	STUDIO DI SIMULAZIONE	29
5.1	Gruppi bilanciati/sbilanciati omoschedastici con osservazioni indipendenti . . .	33
5.2	Gruppi bilanciati/sbilanciati eteroschedastici con osservazioni indipendenti . .	38
5.3	Gruppi bilanciati/sbilanciati omoschedastici con osservazioni dipendenti . . .	52
5.4	Gruppi bilanciati/sbilanciati eteroschedastici con osservazioni dipendenti . . .	57
5.5	Discussione	68
6	CONCLUSIONI	71
	Bibliografia	73
A	TABELLE DEI RISULTATI	75

Capitolo 1

ANALISI DELLA VARIANZA

L'analisi della varianza (ANalysis Of VAriance, ANOVA) è una tecnica statistica che ha l'obiettivo di confrontare i valori medi di una variabile quantitativa (detta variabile dipendente) in corrispondenza delle diverse determinazioni di una o più variabili qualitative e/o quantitative (chiamate fattori o variabili indipendenti).

I fattori possono avere diverse modalità dette livelli e, nel caso della presenza di un solo fattore, tali livelli sono anche definiti trattamenti.

L'ANOVA può essere vista come un'estensione del t-test: mentre il t-test confronta le medie di due gruppi per determinare se esistono differenze significative tra di esse, l'analisi della varianza estende questo concetto per includere più gruppi.

Questo test è stato sviluppato dal famoso statistico britannico Ronald A. Fisher nel 1920 e costituisce ancora oggi una delle tecniche più utilizzate nell'ambito dell'analisi di dati sperimentali.

È necessario sottolineare che tale test non identifica però quali gruppi sono significativamente differenti tra loro, ma solo se almeno una coppia di gruppi ha medie statisticamente diverse.

L'analisi della varianza ad una via (ANOVA a un fattore o ad un criterio di classificazione) serve a saggiare la differenza tra le medie di due o più gruppi quando si ha una sola fonte di variabilità, cioè un solo fattore a cui sono associati diversi livelli.

Se siamo invece in presenza di due o più fattori che influenzano la variabile dipendente, l'ANOVA a due o più vie permette di studiare l'effetto di ciascun fattore separatamente e comprendere anche se e come i fattori interagiscono tra loro.

1.1 ANOVA ad una via

La singola osservazione della variabile dipendente la indichiamo con y_{ij} : essa è riportata con due pedici perché uno è relativo al gruppo o trattamento di appartenenza e l'altro alla posizione occupata all'interno del gruppo. Quindi y_{ij} è l' i -esima osservazione nel j -esimo gruppo con $i = 1, \dots, n_j$ e $j = 1, \dots, J$, dove J è il numero di gruppi e n_j sono le numerosità di ciascun gruppo.

Nell'ANOVA ad una via è preferibile che i vari trattamenti abbiano lo stesso numero di osservazioni. Qualora la numerosità dei gruppi fosse la stessa, si parla di disegno bilanciato, in caso contrario, di disegno sbilanciato.

Al fine di scomporre la devianza complessiva delle osservazioni è essenziale specificare alcune quantità e le rispettive notazioni. Indichiamo con:

- $\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$ la media del j -esimo gruppo, con $j = 1, \dots, J$,
- $\bar{y} = \frac{1}{n} \sum_{j=1}^J \sum_{i=1}^{n_j} y_{ij} = \frac{1}{n} \sum_{j=1}^J n_j \bar{y}_j$ la media generale dove $n = \sum_{j=1}^J n_j$,
- $s_j^2 = \frac{1}{n_j-1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$ la varianza campionaria corretta del j -esimo gruppo, con $j = 1, \dots, J$.

A partire da queste quantità, la devianza totale (Sum of Squares Total, SST) presente nei dati osservati viene scomposta in devianza tra i gruppi (Sum of Squares TRreatment, SSSTR) e devianza entro i gruppi (Sum of Squares Error, SSE).

$$\begin{aligned} \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 &= \sum_{j=1}^J \sum_{i=1}^{n_j} ((y_{ij} - \bar{y}_j) + (\bar{y}_j - \bar{y}))^2 \\ &= \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 + \sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2 + 2 \sum_{j=1}^J (\bar{y}_j - \bar{y}) \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j) \\ &= \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 + \sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2 \\ &= \sum_{j=1}^J s_j^2 (n_j - 1) + \sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2 \end{aligned}$$

ricordando che $\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j) = 0$ per ogni $j = 1, \dots, J$ in quanto la somma degli scostamenti di un insieme di valori dalla loro media aritmetica è nulla.

In definitiva

$$\sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 = \sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2 + \sum_{j=1}^J s_j^2 (n_j - 1)$$

$$SST = SSTR + SSE$$

La $SSTR$ misura quanto le medie dei diversi gruppi si discostano dalla media generale dei dati ed è perciò un indicatore dell'effetto dei trattamenti sulla variabile dipendente.

La SSE rappresenta la variazione della variabile dipendente all'interno di ciascun trattamento infatti misura quanto ciascuna osservazione si distanzia dalla media del proprio gruppo. Questa variazione può essere dovuta a fattori non considerati nel modello o alla casualità dei dati stessi.

Si assume che le y_{ij} siano realizzazioni di $Y_{ij} \sim N(\mu_j, \sigma^2)$, ossia variabili casuali normali e indipendenti con $j = 1, \dots, J$ e $i = 1, \dots, n_j$. Ciò significa che le osservazioni in ogni gruppo devono essere indipendenti e identicamente distribuite e inoltre la varianza deve essere la medesima per tutti i gruppi. Quest'ultima è nota come ipotesi di omoschedasticità.

Sia l'ipotesi nulla $H_0 : \mu_1 = \mu_2 = \dots = \mu_J$ e l'ipotesi alternativa H_1 : almeno due gruppi hanno medie diverse tra loro.

La statistica test per l'ANOVA ad una via è :

$$F = \frac{\frac{SSTR}{(J-1)}}{\frac{SSE}{(n-J)}} \sim F_{J-1, n-J} \quad (1.1)$$

Per determinare se è presente una differenza significativa tra le medie di almeno due gruppi si confronta il valore ottenuto della statistica (1.1) con una distribuzione F di Fisher con $J - 1$ e $n - J$ gradi di libertà. Se il valore della statistica (1.1) calcolato risulta maggiore del corrispondente valore della distribuzione di Fisher, l'ipotesi nulla viene rigettata concludendo che ci sono almeno due gruppi con valori medi diversi. Si può giungere alla medesima conclusione osservando il p-value. Siccome $p\text{-value} = P(F_{J-1, n-J} > F^{oss})$, se esso appare minore del livello di significatività predeterminato α , allora si rifiuta H_0 .

Tipicamente i risultati vengono riassunti nella cosiddetta tabella dell'ANOVA.

Per procedere con la dimostrazione della distribuzione della statistica test F (1.1) sotto l'ipotesi nulla, è necessario inizialmente ricavare la distribuzione di $SSTR$ e SSE e verificare la loro indipendenza.

È utile ricordare alcuni risultati la cui dimostrazione è stata omessa per la necessità di mantenere l'attenzione sull'argomento principale della presente dimostrazione:

Proposizione 1. Sia $S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$ la varianza corretta di un campione casuale semplice di ampiezza n proveniente da una popolazione normale con media μ e varianza σ^2 . Allora la variabile casuale $\frac{(n-1)S^2}{\sigma^2}$ ha distribuzione di probabilità chi-quadrato con $n-1$ gradi di libertà (G.Cicchitelli et al., 2017[1]).

Proposizione 2. Se U_1, U_2, \dots, U_k sono k variabili casuali indipendenti con distribuzione chi-quadrato con rispettivi gradi di libertà n_1, n_2, \dots, n_k , allora $U_1 + U_2 + \dots + U_k$ ha una distribuzione chi-quadrato con $n_1 + n_2 + \dots + n_k$ gradi di libertà.

Proposizione 3. TEOREMA DI COCHRAN

Siano $Z_i \sim N(0, 1)$ per $i = 1, \dots, v$ e $\sum_{i=1}^v Z_i^2 = Q_1 + Q_2 + \dots + Q_s$ dove $s \leq v$ e Q_t ha v_t gradi di libertà per $t = 1, \dots, s$. Allora Q_1, \dots, Q_s sono variabili casuali chi-quadrato indipendenti con rispettivamente v_1, \dots, v_s gradi di libertà se e solo se $v = v_1 + \dots + v_s$ (D.C.Montgomery, 2017 [2]).

Proposizione 4. Se X e Y sono due variabili casuali indipendenti distribuite come χ^2 con rispettivamente d_1 e d_2 gradi di libertà, allora il rapporto $\frac{\frac{X}{d_1}}{\frac{Y}{d_2}}$ segue una distribuzione di Fisher con d_1 gradi di libertà nel numeratore e d_2 gradi di libertà nel denominatore.

Dimostrazione della distribuzione della statistica test F sotto H_0

Dimostrazione. Poiché $Y_{ij} \sim N(\mu_j, \sigma^2)$, con $j = 1, \dots, J$ e $i = 1, \dots, n_j$, la Proposizione 1 garantisce che

$$\frac{\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{\sigma^2} \sim \chi_{n_j-1}^2$$

La Proposizione 2 permette di dichiarare che

$$\frac{SSE}{\sigma^2} = \frac{1}{\sigma^2} \sum_{j=1}^J \sum_{i=1}^{n_j} n_j (y_{ij} - \bar{y}_j)^2 \sim \chi_{n-J}^2$$

infatti $\sum_{j=1}^J J(n_j - 1) = (n_1 + \dots + n_J) - J = n - J$.

Sotto H_0 tutte le osservazioni, data l'assenza di differenza tra trattamenti, sono realizzazioni della stessa variabile casuale $N(\mu, \sigma^2)$, quindi

$$\frac{SST}{\sigma^2} = \frac{1}{\sigma^2} \sum_{j=1}^J \sum_{i=1}^{n_j} n_j (y_{ij} - \bar{y})^2 \sim \chi_{n-1}^2$$

Infine, grazie alla regola della scomposizione della varianza ($SSTR = SST - SSE$), possiamo concludere, se H_0 è vera, che

$$\frac{SSTR}{\sigma^2} \sim \chi_{n-1-(n-J)}^2 = \chi_{J-1}^2$$

Se sommiamo i gradi di libertà di $SSTR$ e SSE otteniamo $n - J + J - 1 = n - 1$ cioè esattamente i gradi di libertà di SST . Per la Proposizione 3 affermiamo perciò che, sotto H_0 , $\frac{SSTR}{\sigma^2}$ e $\frac{SSE}{\sigma^2}$ sono due variabili casuali chi-quadrato indipendenti.

Il concetto di indipendenza è fondamentale per decretare la distribuzione della statistica test F (1.1) per l'ANOVA, infatti essa è definita come il rapporto tra la varianza stimata tra i trattamenti e quella stimata all'interno dei gruppi, normalizzate per i rispettivi gradi di libertà e quindi per la Proposizione 4:

$$F \sim \frac{\frac{\chi_{J-1}^2}{(J-1)}}{\frac{\chi_{n-J}^2}{(n-J)}} = F_{J-1, n-J}$$

Sotto H_0 , cioè sotto l'ipotesi nulla di assenza di differenza significative tra i gruppi, la statistica F (1.1) si distribuisce come una Fisher con $J - 1$ gradi di libertà al numeratore e $n - J$ gradi di libertà al denominatore. \square

1.1.1 Modello di regressione lineare multipla per il confronto di medie

Il modello ANOVA, il confronto tra le medie di J gruppi, può essere riformulato come verifica di ipotesi sui parametri di un modello di regressione lineare multipla.

Consideriamo le osservazioni y_i , per $i = 1, \dots, n$, come realizzazioni indipendenti di $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, con $n = \sum_{j=1}^J n_j$ dove le n_j sono le numerosità dei gruppi, per $j = 1, \dots, J$.

Il modello di regressione risulta $Y_i = f(\text{fattore}) + \epsilon_i$, con $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ anch'esse indipendenti per $i = 1, \dots, n$.

Supponiamo di voler analizzare l'effetto di J diversi trattamenti su una variabile risposta e siamo interessati a verificare se ci sono differenze significative tra i gruppi.

Generalmente se si hanno J gruppi di trattamento, si ricorre a $(J - 1)$ variabili dummy per rappresentarli. Considerando il primo gruppo come trattamento base, le variabili indicatrici sono codificate nel seguente modo: $x_{i2} = 1$ se l' i -esimo elemento è stato sottoposto al secondo trattamento e 0 altrimenti, $x_{i3} = 1$ se l' i -esimo elemento è stato sottoposto al terzo trattamento e 0 altrimenti, e così via, fino ad arrivare a $x_{iJ} = 1$ se l' i -esimo elemento è stato sottoposto al J -esimo, e quindi ultimo, trattamento e 0 altrimenti.

Il modello può essere scritto come $Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_J x_{iJ} + \epsilon_i$, con $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ indipendenti per $i = 1, \dots, n$. Anche per questo modello occorre verificare le stesse assunzioni per la validità dell'ANOVA quindi indipendenza delle osservazioni, normalità all'interno di ciascun trattamento e omoschedasticità dei gruppi.

Si ha perciò che $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ con $\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_J x_{iJ}$ per $i = 1, \dots, n$.

Di conseguenza:

$$Y_i \sim \begin{cases} N(\beta_1, \sigma^2) & \text{se l'}i\text{-esimo elemento appartiene al primo gruppo} \\ N(\beta_1 + \beta_2, \sigma^2) & \text{se l'}i\text{-esimo elemento appartiene al secondo gruppo} \\ \dots & \\ N(\beta_1 + \beta_J, \sigma^2) & \text{se l'}i\text{-esimo elemento appartiene all'ultimo gruppo} \end{cases}$$

I parametri di regressione possono essere interpretati nel seguente modo: β_1 è la media delle osservazioni del primo trattamento, mentre $\beta_1 + \beta_j$ è la media delle osservazioni del j -esimo trattamento, $j = 2, \dots, J$. Allora β_j è la differenza tra la media del j -esimo trattamento con la media del trattamento base, ossia il primo, per $j = 2, \dots, J$.

Siccome siamo interessati a valutare l'omogeneità delle medie, l'ipotesi nulla può essere formulata come $H_0 = \beta_2 = \beta_3 = \dots = \beta_J = 0$.

Per verificare ciò, si effettua il confronto tra il modello corrente $Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_J x_{iJ} + \epsilon_i$ e il modello nullo $Y_i = \beta_1 + \epsilon_i$, con $\epsilon_i \sim N(0, \sigma^2)$ indipendenti per $i = 1, \dots, n$.

La statistica test è:

$$F = \frac{\frac{\hat{\sigma}^2 - \hat{\sigma}^2}{(J-1)}}{\frac{\hat{\sigma}^2}{(n-J)}} \sim F_{J-1, n-J} \quad (1.2)$$

ma $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$, stima di massima verosimiglianza della varianza nel modello nullo, corrisponde a $\frac{SST}{n}$, mentre $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \dots + \hat{\beta}_J x_{iJ})^2$, stima della varianza nel modello corrente, corrisponde a $\frac{SSE}{n}$ (Paragrafo 5.2.2 di M.Grigoletto *et al.*, 2017 [3]).

Di conseguenza la statistica (1.2) può essere riscritta

$$F = \frac{\frac{(SST - SSE)}{\frac{n}{(J-1)}}}{\frac{\frac{SSE}{n}}{(n-J)}} = \frac{\frac{(SST - SSE)}{(J-1)}}{\frac{SSE}{(n-J)}} = \frac{\frac{SSTR}{(J-1)}}{\frac{SSE}{(n-J)}}$$

che coincide esattamente con la statistica (1.1).

Capitolo 2

CONSEGUENZE DELLA VIOLAZIONE DELLE ASSUNZIONI DELL'ANALISI DELLA VARIANZA

Le assunzioni dell'analisi della varianza sono:

- normalità delle distribuzioni della variabile dipendente all'interno di ciascun gruppo,
- indipendenza delle osservazioni,
- omoschedasticità,

e sono importanti da verificare affinché le conclusioni dell'ANOVA risultino affidabili e correttamente interpretabili.

In generale, quando si esegue un test, si possono commettere due tipi di errore :

- errore di I tipo che consiste nel rifiutare erroneamente l'ipotesi nulla in favore di quella alternativa,
- errore di II tipo che consiste nell'accettare l'ipotesi nulla quando quest'ultima è falsa.

Il tasso di errore di I tipo (α), noto anche come livello di significatività, è quindi la probabilità di rifiutare H_0 quando essa è vera, il tasso di errore di II tipo (β) corrisponde alla probabilità di accettare H_0 quando H_1 è vera ed infine, la potenza del test ($1 - \beta$) è la probabilità di rigettare correttamente l'ipotesi nulla [4].

2.1 Violazione dell'assunzione di normalità

Studi di simulazione hanno dimostrato che il test ANOVA è generalmente ritenuto robusto rispetto alle violazioni del presupposto di normalità [5].

Per quanto riguarda il tasso d'errore di I tipo e di II tipo, la forma della distribuzione ha un impatto minimo sul test F (1.1). Tuttavia è necessario porre attenzione alle code della distribuzione: se queste sono più basse di una distribuzione normale parliamo di distribuzione platicurtica (a "code leggere"), mentre se sono più alte parliamo di distribuzione leptocurtica (a "code pesanti"). In caso di distribuzione leptocurtica, il livello di significatività rischia di aumentare, viceversa, in presenza di "code leggere", di diminuire. Tale effetto della non normalità delle osservazioni aumenta quando le dimensioni dei gruppi non sono uguali tra loro e viceversa diminuisce all'aumentare delle numerosità dei campioni [4].

2.2 Violazione dell'assunzione di indipendenza

Il mancato rispetto del presupposto di indipendenza può avere gravi effetti sulla validità della procedura dell'analisi della varianza, pertanto il test F (1.1) non risulta idoneo in queste circostanze.

Se le osservazioni sono dipendenti, il tasso d'errore di I tipo viene gonfiato e la potenza ridotta anche se le dimensioni dei gruppi sono ampie.

La natura della dipendenza, positiva o negativa, ha effetti diversi sul test:

- se positiva, il test risulta più liberale, ossia la probabilità di errore di I tipo aumenta,
- se negativa, il test risulta più conservativo, cioè il tasso di errore di I tipo diminuisce.

Infine, quanto più è elevata la dipendenza tra le osservazioni, tanto più è diverso il livello di significatività osservato rispetto al livello nominale [6].

Per garantire che le osservazioni siano indipendenti è fondamentale che la raccolta dei dati avvenga in modo casuale e indipendente, ciò significa che gli elementi presi in esame devono essere assegnati casualmente ai diversi gruppi assicurandosi che le osservazioni di un gruppo non influenzino e non siano influenzati dagli altri. In caso, invece, di progetti sperimentali in cui i soggetti sono esposti a tutti i livelli della variabile indipendente, è concepibile considerare che gli effetti di un trattamento persistano influenzando la risposta ai trattamenti successivi e quindi necessitano di tecniche statistiche adeguate.

2.3 Violazione dell'assunzione di omoschedasticità

Per quanto riguarda il tasso d'errore di I tipo, il test F (1.1) è sensibile alla violazione del presupposto di omoschedasticità. In particolare, l'impatto sarà maggiore quanto più le varianze dei gruppi differiscono tra loro.

In caso di disegni sbilanciati, il test F (1.1) risulta ancora meno robusto alla violazione dell'assunzione di omoschedasticità e fortemente legato alla relazione tra dimensione del gruppo e rispettiva varianza.

In caso di relazione positiva, ossia il gruppo con numerosità maggiore ha anche la varianza maggiore, il test è troppo conservativo, il tasso d'errore di I tipo risulta inferiore rispetto al valore nominale. Viceversa, nella situazione di relazione negativa in cui il gruppo avente dimensione maggiore ha varianza minore, il test è troppo liberale.

Allo stesso modo, il tasso d'errore di II tipo, in un disegno bilanciato, è meno influenzato rispetto alla situazione in cui i gruppi sono di dimensioni diverse. Se vi è relazione positiva tra numerosità e varianza, il tasso d'errore di II tipo aumenta e di conseguenza la potenza diminuisce, viceversa, in presenza di relazione negativa, il tasso d'errore di II tipo diminuisce e la potenza aumenta [4] [5].

Al fine di evitare di ottenere risultati non validi dall'analisi della varianza, è fondamentale verificare l'omoschedasticità dei gruppi attraverso dei test in modo tale da utilizzare, in caso risultino significativi, metodi statistici alternativi all'ANOVA.

Per la verifica dell'ipotesi di omoschedasticità per J gruppi, l'ipotesi nulla è

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_J^2$$

che viene posta contro l'ipotesi alternativa

$$H_1 : \text{non tutte le varianze sono uguali.} \quad (2.1)$$

Di seguito sono riportati diversi test e la scelta di quale applicare dipende dalle caratteristiche dei dati, ognuno ha i suoi punti di forza e le sue limitazioni che lo rendono adatto in specifiche circostanze.

2.3.1 Test di Cochran

Il test originario proposto da Cochran (1941) della varianza massima è consigliato quando si ipotizza che una varianza sia nettamente maggiore delle altre (successivamente il test è stato esteso con la versione della varianza minima da utilizzare nella situazione contraria).

Il test originale di Cochran verifica che J gruppi abbiano uguale varianza in base al rapporto tra la varianza massima e la somma di tutte le varianze. Affinché il test sia applicabile, è necessario che i J gruppi siano bilanciati, ognuno deve contenere perciò n osservazioni.

La statistica test di Cochran che verifica (2.1) è:

$$C = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_J^2} = \frac{s_{\max}^2}{\sum_{i=1}^J s_i^2}$$

dove s_i^2 , $i = 1, \dots, J$, sono le varianze dei gruppi e s_{\max}^2 è la maggiore di queste ultime.

Il valore C osservato è da confrontare con un valore critico dato da:

$$C(\alpha, n, J) = \left(1 + \frac{J-1}{F_{n-1, (J-1)(n-1); 1-\frac{\alpha}{2}}} \right)^{-1}$$

dove J è il numero di gruppi ognuno contenente n osservazioni e α è il livello di significatività.

Qualora il valore C osservato risultasse maggiore del suo corrispondente valore critico, l'ipotesi nulla verrebbe rigettata concludendo che la varianza massima differisce significativamente rispetto alle altre [7] [8].

Il test è disponibile nel pacchetto *outliers* di R applicando la funzione *cochran.test()*.

2.3.2 Test di Bartlett

Il test di Bartlett (1937) richiede che venga soddisfatta l'assunzione di normalità.

Supponiamo di essere interessati a valutare l'omogeneità delle varianze di J gruppi in cui le osservazioni si distribuiscono normalmente.

Per verificare (2.1), la statistica test di Bartlett è data da:

$$T_B = \frac{(N-J) \ln s^2 - \sum_{g=1}^J (n_g - 1) s_g^2}{1 + \frac{1}{3(J-1)} \left(\sum_{g=1}^J \frac{1}{n_g - 1} - \frac{1}{N-J} \right)}$$

dove n_g , $g = 1, \dots, J$, sono le numerosità dei gruppi, $N = \sum_{g=1}^J n_g$ le osservazioni totali, s_g^2 , $g = 1, \dots, J$, le varianze interne ai gruppi ed infine $s^2 = \frac{1}{N-J} \sum_{g=1}^J (n_g - 1) s_g^2$ la varianza complessiva.

Notiamo che il test permette che i gruppi possano contenere un numero differente di osservazioni, quindi è applicabile anche in caso di disegni sbilanciati. È opportuno però che la numerosità di ciascun trattamento sia sufficientemente grande tale da rendere affidabili gli stimatori s_g^2 , $g = 1, \dots, J$.

Sotto l'ipotesi nulla la statistica T_B si distribuisce come una χ^2_{J-1} . Se il p-value calcolato sotto H_0 risulta minore del livello di significatività stabilito α , l'ipotesi nulla viene respinta stabilendo non omogeneità delle varianze [7] [8] [3].

Box (1953) dimostrò che questo test non è robusto se l'assunzione di normalità viene violata, in tal caso sono più adeguati altri test meno sensibili a tale deviazione.

Il test è disponibile nel pacchetto standard (*stats*) di R richiamando la funzione *bartlett.test()*.

2.3.3 Test di Hartley

Verifica se J gruppi hanno varianze omogenee in base al rapporto tra la varianza massima e quella minima. È sensibile alle deviazioni del presupposto di normalità e necessita che i J gruppi abbiano uguale numerosità.

La statistica test di Hartley (1950) che verifica (2.1) è:

$$F_{max} = \frac{s_{max}^2}{s_{min}^2}$$

dove s_{max}^2 e s_{min}^2 sono rispettivamente la varianza maggiore e minore tra quelle dei trattamenti messi a confronto.

Sotto l'ipotesi nulla, il valore della statistica F_{max} viene confrontato con un valore critico, riportato in una speciale tabella, che dipende dal numero di gruppi e dal grado di libertà dato dalla numerosità comune a ciascun trattamento meno uno. Se F_{max} osservato si rivela inferiore al valore critico corrispondente, non abbiamo evidenza di rifiutare H_0 dichiarando così omogeneità delle varianze [7] [8].

2.3.4 Test di Levene

Il test di Levene è l'equivalente a un'analisi della varianza in cui però la variabile dipendente è il valore assoluto della differenza tra l'osservazione e la media del gruppo a cui essa appartiene.

Dato y_{ij} il valore dell' i -esima osservazione appartenente al j -esimo gruppo, definiamo:

- $Z_{ij} = |y_{ij} - \bar{y}_j|$ dove \bar{y}_j è la media del j -esimo gruppo,
- $\bar{Z}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Z_{ij}$, $j = 1, \dots, J$, media delle Z_{ij} per il j -esimo gruppo di numerosità n_j ,
- $\bar{\bar{Z}} = \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{n_j} Z_{ij}$ media generale di tutte le Z_{ij} dove $N = \sum_{j=1}^J n_j$ è il numero totale di osservazioni.

La statistica test di Levene (1960) che verifica (2.1) è data da:

$$W = \frac{N - J}{J - 1} \frac{\sum_{j=1}^J n_j (\bar{Z}_j - \bar{\bar{Z}})^2}{\sum_{j=1}^J \sum_{i=1}^{n_j} (Z_{ij} - \bar{Z}_j)^2} \quad (2.2)$$

Sotto l'ipotesi nulla, la statistica W si distribuisce come una F di Fisher con $J - 1$ e $N - J$ gradi di libertà. H_0 viene rifiutata se $W^{oss} > F_{J-1, N-J; 1-\alpha}$, cioè se il valore di W calcolato risulta maggiore del valore critico superiore di una F con $J - 1$ e $N - J$ gradi di libertà ad un livello di significatività α , dichiarando così che esiste una differenza tra le varianze dei gruppi confrontati [7] [8].

2.3.5 Test di Brown-Forsythe

Si tratta di un test di Levene a cui è stata applicata una modifica: Brown e Forsythe (1974) proposero di utilizzare la mediana di ciascun gruppo, anziché la media aritmetica, nel calcolo delle Z_{ij} . La statistica test di Brown-Forsythe risulta la stessa di Levene (2.2) ma con $Z_{ij} = |y_{ij} - \tilde{y}_j|$ dove \tilde{y}_j è la mediana del j -esimo gruppo.

Brown e Forsythe proposero anche un'ulteriore versione: sostituire ciascuna osservazione con la deviazione assoluta dalla media troncata del gruppo di appartenenza. L'utilizzo della media troncata consiste nell'eliminare una percentuale prefissata (la scelta della percentuale è arbitraria) di valori estremi. La statistica test risulta la stessa di W (2.2) ma con $Z_{ij} = |y_{ij} - y_j^p|$ dove y_j^p è la media troncata al $p\%$ del j -esimo gruppo.

Il test originario di Levene è più adatto quando si può assumere che i dati sono distribuiti normalmente e non sono influenzati da outliers. La versione con la mediana o con la media troncata sono utili nel momento in cui sono presenti valori estremi che influenzerebbero la media o ci sono evidenze di deviazione dalla normalità. La scelta tra le tre proposte dipende principalmente dalla forma della distribuzione delle osservazioni [7] [8].

Questi test sono disponibili scaricando la libreria *lawstat* di R e applicando la funzione *levene.test()* in cui è presente il parametro *location* che può assumere valori *mean*, *median* o *trim.mean*.

2.3.6 Test di Fligner-Killeen centrato sulla mediana

È un test non parametrico per esaminare l'omogeneità delle varianze di J gruppi basato sui ranghi. Il test originale fu presentato da Fligner e Killeen (1976) e successivamente modificato da Conover e Johnson (1981) che proposero la versione centrata sulla mediana.

I ranghi vengono calcolati sulle deviazioni dalla mediana del gruppo di appartenenza anziché direttamente sulle osservazioni rendendo il test robusto alle distribuzioni non normali.

Definiamo con a_{ij} il rango dell' i -esima osservazione y_{ij} calcolato su $|y_{ij} - \tilde{y}_j|$ dove \tilde{y}_j è la mediana del j -esimo gruppo a cui l'osservazione appartiene.

La statistica test di Fligner-Killeen per verificare (2.1) è:

$$FK = \frac{\sum_{j=1}^J n_j (\bar{a}_j - \bar{a})^2}{V^2}$$

dove n_j , $j = 1, \dots, J$, sono le numerosità dei gruppi, \bar{a}_j è la media dei ranghi del j -esimo gruppo, \bar{a} è la media generale di tutti i ranghi ed infine $V^2 = \frac{1}{N-1} \sum_{i=1}^N (a_{ij} - \bar{a})^2$ con N numero totale di osservazioni.

La statistica FK sotto H_0 ha distribuzione asintotica χ^2 con $J - 1$ gradi di libertà. Se il p-value, ottenuto dal confronto della statistica FK osservata con la distribuzione χ^2 , risulta inferiore al livello di significatività scelto α , l'ipotesi nulla viene rigettata concludendo la non omogeneità delle varianze [7] [9].

Il test è disponibile nel pacchetto standard (*stats*) di R richiamando la funzione *fligner.test()*.

Capitolo 3

TECNICHE PER AFFRONTARE L'ETEROSCHEDASTICITÀ NELL'ANALISI DELLA VARIANZA

Nella ricerca statistica, garantire la validità e l'affidabilità dei risultati è fondamentale per l'interpretazione delle analisi condotte. Uno dei requisiti principali dell'analisi della varianza è l'omogeneità delle varianze dei gruppi ma, in molte situazioni pratiche, ciò può non verificarsi.

Di seguito verrà affrontata tale problematica proponendo metodi che consentono di stabilizzare le varianze e le più comuni tecniche alternative all'ANOVA che permettono di gestire la non omoschedasticità.

3.1 Trasformazione dei dati

Un primo approccio al problema dell'eterogeneità delle varianze consiste nel trasformare i dati in modo da stabilizzarle per poi procedere con la consueta analisi della varianza sui dati trasformati.

Dopo la trasformazione però, può essere più complesso interpretare i risultati dell'ANOVA in termini dei dati originali poiché non sempre le conclusioni che si traggono dai dati trasformati si trasferiscono perfettamente ai dati originali.

La trasformazione può rendere i dati più conformi alle assunzioni della statistica test che si intende applicare ma è necessario porre particolare attenzione alla scelta della trasformazione stessa. Si deve selezionare quella più appropriata nell'affrontare l'eteroschedasticità senza introdurre altre distorsioni ai dati e che allo stesso tempo garantisca una facile interpretabilità e comunicazione dei risultati.

Può risultare utile nella pratica applicare diverse trasformazioni per verificare il loro effetto sulla distribuzione dei dati prima di sceglierne una in modo definitivo [10] [11].

Le trasformazioni più comuni sono:

- la trasformazione logaritmica che è utile quando la deviazione standard è proporzionale alla media dei dati o quando le distribuzioni sono asimmetriche positivamente (ossia hanno la coda più lunga a destra). Infatti, se un insieme di numeri relativamente grandi possiede un valore grande della deviazione standard e, viceversa, un insieme di numeri piccoli ha un valore della deviazione standard piccolo, prendendo i logaritmi, essi ridurranno la deviazione standard del gruppo con numeri grandi più di quanto non faranno ridurre quella del gruppo con numeri piccoli. Non fa differenza quale base del logaritmo viene usata ma, indipendentemente dalla base, ci sono problemi quando i dati sono negativi poiché i logaritmi sono definiti solo per numeri positivi. In questi casi, per poter applicare la trasformazione logaritmica, è necessario aggiungere una costante in modo tale da rendere i dati tutti positivi.
- la trasformazione radice quadrata che è utile quando, come spesso accade con dati sotto forma di conteggio, la media è proporzionale alla varianza e non alla deviazione standard. Siccome varianza e deviazione standard sono altamente correlate tra loro se l'intervallo dei valori non è ampio, è impossibile distinguere ad occhio la loro relazione con la media. In pratica è quindi consigliato provare sia la trasformazione radice quadrata che logaritmica ed utilizzare quella che sembra funzionare meglio in quello specifico contesto.
- la trasformazione di reciprocità che può essere utilizzata quando le varianze dei gruppi diminuiscono all'aumentare della media. Inoltre, se la distribuzione dei dati presenta valori grandi nella coda positiva di destra, questa trasformazione riduce drasticamente l'influenza dei valori estremi. I valori anomali avranno un effetto notevolmente inferiore sulla deviazione standard rispetto a prima della trasformazione.
- la trasformazione arcoseno che è spesso usata quando si lavora con dati che rappresentano proporzioni o percentuali.
- la trasformazione di Box-Cox che richiede tutti dati strettamente positivi. Dipende dal parametro λ che può essere stimato empiricamente per massimizzare l'omogeneità delle varianze dei dati trasformati. Ai dati di ciascun gruppo viene applicata la seguente trasformazione:

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \ln(y) & \text{se } \lambda = 0 \end{cases}$$

dove y rappresenta il dato originale, $y^{(\lambda)}$ il dato trasformato e λ il parametro di trasformazione.

3.2 ANOVA non parametrica: test di Kruskal-Wallis

Il test di Kruskal-Wallis (1952), è l'alternativa non parametrica dell'ANOVA ad un criterio di classificazione, infatti viene utilizzato per confrontare tre o più gruppi indipendenti qualora le assunzioni di normalità e omoschedasticità richieste dall'ANOVA non sono soddisfatte.

Il test, anziché sulle osservazioni, si basa sui ranghi: il rango di un dato è la posizione che occupa in un insieme ordinato dei dati.

Supponiamo di avere N osservazioni y_i , per $i = 1, \dots, N$, divise in J gruppi in base ad un determinato fattore.

Il test di Kruskal-Wallis verifica che i J gruppi provengano dalla stessa popolazione. L'ipotesi nulla è H_0 : tutti i campioni provengono dalla stessa popolazione (stessa distribuzione) posta contro l'ipotesi alternativa H_1 : almeno un campione proviene da una popolazione con una distribuzione diversa.

Il primo passaggio consiste nell'assegnare a ciascuna osservazione il rango indipendentemente dal gruppo di appartenenza. In caso di legami, cioè di due o più dati con lo stesso valore, si assegna loro il rango medio fra i ranghi che occuperebbero in caso di assenza di legami.

Successivamente, sia in presenza di legami che in assenza di essi, si calcola la somma dei ranghi per ciascun gruppo. Se non ci fosse discrepanza tra i gruppi, i ranghi grandi e piccoli dovrebbero essere equamente distribuiti e il rango medio per gruppo dovrebbe essere prossimo alla media di tutti i ranghi.

La statistica test di Kruskal-Wallis è:

$$H = \left(\frac{12}{N(N+1)} \sum_{i=1}^J \frac{R_i^2}{n_i} \right) - 3(N+1)$$

dove n_i , per $i = 1, \dots, J$, è il numero di osservazioni dell' i -esimo gruppo, N è il numero totale delle osservazioni, J è il numero di gruppi considerati nel confronto ed infine R_i è la somma dei ranghi dell' i -esimo gruppo, per $i = 1, \dots, J$.

In caso di dimensioni campionarie ridotte, dove il numero di gruppi confrontati è 3 e in ognuno sono presenti meno di 5 osservazioni, per valutare la significatività della statistica H si

ricorre a delle specifiche tabelle che forniscono i corrispondenti valori critici per un dato numero di gruppi aventi un determinato numero di osservazioni. In tutte le altre situazioni invece, quando le dimensioni campionarie sono sufficientemente grandi, la statistica H si distribuisce asintoticamente come una chi-quadrato con $J - 1$ gradi di libertà.

Come per l'ANOVA, è importante puntualizzare che se il test risulta significativo, ossia se la statistica H osservata appare maggiore del suo valore critico, si conclude che almeno due gruppi tra quelli confrontati differiscono tra loro ma non vengono fornite informazioni su quali effettivamente siano diversi.

Il test è disponibile nel pacchetto standard (*stats*) di R richiamando la funzione *kruskal.test()*.

3.3 Test di Welch

Mentre l'analisi della varianza tradizionale assume l'omoschedasticità, il test di Welch rilassa tale assunzione rendendola una tecnica utile per confrontare medie di più gruppi che non hanno necessariamente varianze uguali.

Sia l'ipotesi nulla $H_0 : \mu_1 = \mu_2 = \dots = \mu_J$ e l'ipotesi alternativa H_1 : almeno due gruppi hanno medie diverse tra loro.

La statistica test di Welch (1951) è:

$$F_W = \frac{\frac{1}{J-1} \sum_{j=1}^J [w_j(\bar{y}_j - \bar{y}')^2]}{1 + \frac{2(J-2)}{J^2-1} \sum_{j=1}^J \left[\left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum_{j=1}^J w_j} \right)^2 \right]}$$

dove J è il numero di gruppi a confronto, \bar{y}_j è la media del j -esimo gruppo di numerosità n_j ,

$w_j = \frac{n_j}{s_j^2}$ con $s_j^2 = \frac{1}{n_j-1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$ ed infine $\bar{y}' = \frac{\sum_{j=1}^J w_j \bar{y}_j}{\sum_{j=1}^J w_j}$, per $j = 1, \dots, J$.

La statistica test F_W si distribuisce approssimativamente come una F di Fisher con $J - 1$ e v gradi libertà dove $v = \frac{J^2-1}{3 \sum_{j=1}^J \left(\frac{1}{n_j-1} \right) \left(1 - \frac{w_j}{\sum_{j=1}^J w_j} \right)^2}$.

Se F_W osservato risulta minore di $F_{(J-1, v; 1-\alpha)}$ dove α è il livello di significatività stabilito, allora non ci sono evidenze per rifiutare l'ipotesi nulla [4] [11] [12].

Il test è disponibile nel pacchetto *onewaytests* di R richiamando la funzione *welch.test()* ma è presente anche nel pacchetto standard (*stats*) di R attraverso la funzione *oneway.test()* specificando il parametro *var.equal=FALSE*.

3.4 Test di Welch con medie troncate e varianze winsorizzate

Questa procedura verifica l'uguaglianza delle medie dei gruppi sostituendo con le medie troncate e le varianze winsorizzate le usuali stime di medie e varianze, risultando non solo meno sensibile all'eteroschedasticità, ma anche robusta agli effetti dei valori anomali.

La media troncata (o trimmed) è una stima della media che esclude una percentuale di osservazioni estreme dai dati di partenza. La varianza winsorizzata è una stima della varianza che tratta gli outliers riducendo la loro influenza senza eliminarli completamente: i valori tagliati vengono sostituiti dal valore più estremo rimanente in ciascuna coda.

Siano $y_{(1)j} \leq y_{(2)j} \leq \dots \leq y_{(n_j)j}$ le osservazioni ordinate del j -esimo gruppo di dimensione n_j e sia $g_j = \lfloor \epsilon n_j \rfloor$ dove ϵ è la proporzione da tagliare in ciascuna coda della distribuzione. La dimensione effettiva di ciascun gruppo dopo il taglio risulta $h_j = n_j - 2g_j$, con $j = 1, \dots, J$.

Definite le seguenti quantità:

- $\bar{y}_{tj} = \frac{1}{h_j} \sum_{i=g_j+1}^{n_j-g_j} y_{(i)j}$ la media troncata del j -esimo gruppo, per $j = 1, \dots, J$,
- $\bar{y}_{wj} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$ media winsorizzata del j -esimo gruppo, con $j = 1, \dots, J$, dove

$$x_{ij} = \begin{cases} y_{(g_j+1)j} & \text{se } y_{ij} \leq y_{(g_j+1)j} \\ y_{ij} & \text{se } y_{(g_j+1)j} < y_{ij} < y_{(n_j-g_j)j} \\ y_{(n_j-g_j)j} & \text{se } y_{ij} \geq y_{(n_j-g_j)j} \end{cases}$$

- $s_{wj}^2 = \frac{1}{n_j-1} \sum_{i=1}^{n_j} (x_{ij} - \bar{y}_{wj})^2$ la varianza winsorizzata del j -esimo gruppo per $j = 1, \dots, J$,

la statistica test di Welch con medie troncate e varianze winsorizzate (1951) è data da:

$$F_{WT} = \frac{\frac{1}{J-1} \sum_{j=1}^J [w_j (\bar{y}_{tj} - \tilde{y})^2]}{1 + \frac{2(J-2)}{J^2-1} \sum_{j=1}^J \left[\left(\frac{1}{h_j-1} \right) \left(1 - \frac{w_j}{\sum_{j=1}^J w_j} \right)^2 \right]}$$

dove $w_j = \frac{1}{q_j}$ con $q_j = \frac{(n_j-1)s_{wj}^2}{h_j(h_j-1)}$ e $\tilde{y} = \frac{\sum_{j=1}^J w_j \bar{y}_{tj}}{\sum_{j=1}^J w_j}$.

La statistica F_{WT} si distribuisce approssimativamente come una F di Fisher con $J-1$ e v' gradi libertà, dove $v' = \frac{J^2-1}{3 \sum_{j=1}^J \left(\frac{1}{h_j-1} \right) \left(1 - \frac{w_j}{\sum_{j=1}^J w_j} \right)^2}$.

Se F_{WT} osservato risulta maggiore di $F_{(J-1, v'; 1-\alpha)}$ dove α è il livello di significatività pre-stabilito, allora si rigetta l'ipotesi nulla $H_0 : \mu_1 = \mu_2 = \dots = \mu_J$ concludendo medie differenti tra i gruppi [11].

Il test è disponibile nel pacchetto *onewaytests* di R applicando la funzione *welch.test()* e modificando il parametro *rate* in base alla percentuale di osservazioni che si vuole escludere.

3.5 Test di Alexander-Govern

Si tratta di un test alternativo all'ANOVA non robusto per dati non normali ma usato quando le varianze dei gruppi non sono omogenee.

Sia l'ipotesi nulla $H_0 : \mu_1 = \mu_2 = \dots = \mu_J$ e l'ipotesi alternativa $H_1 :$ almeno due gruppi hanno medie diverse tra loro.

La statistica test di Alexander-Govern (1994) è:

$$\chi_{AG}^2 = \sum_{j=1}^J z_j^2$$

Per determinare z_j è necessario prima calcolare le seguenti quantità:

- $s'_j = \left[\frac{\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{n_j(n_j-1)} \right]^{0.5}$ deviazione standard del j -esimo gruppo con media \bar{y}_j e numerosità $n_j, j = 1, \dots, J,$
- $w_j = \frac{\frac{1}{s_j'^2}}{\sum_{j=1}^J \frac{1}{s_j'^2}}$ peso del j -esimo gruppo per $j = 1, \dots, J,$
- $t_j = \frac{\bar{y}_j - \tilde{y}}{s'_j}$ statistica del gruppo j -esimo dove $\tilde{y} = \sum_{j=1}^J w_j \bar{y}_j$ per $j = 1, \dots, J.$

Nella statistica $\chi_{AG}^2,$

$$z_j = c + \frac{(c^3 + 3c)}{b} - \frac{(4c^7 + 33c^5 + 240c^3 + 855c)}{(10b^2 + 8bc^4 + 1000b)}$$

dove, dato $v_j = n_j - 1$ e $\alpha = v_j - 0.5, c = [\alpha \ln(1 + \frac{t_j^2}{v_j})]^{0.5}$ e $b = 48\alpha^2.$

La statistica $\chi_{AG}^2,$ sotto l'ipotesi nulla, si distribuisce come una χ^2 con $J - 1$ gradi di libertà [11].

Il test è disponibile nel pacchetto *onewaytests* di R richiamando la funzione *ag.test()*.

3.6 Test di James del secondo ordine

Un'ulteriore alternativa all'analisi della varianza tradizionale è stata proposta da James (1951) da poter utilizzare quando l'assunto di omoschedasticità dell'ANOVA viene violato.

La statistica test di James è data da:

$$J_S = \sum_{j=1}^J t_j^2$$

dove $t_j = \frac{\bar{y}_j - \tilde{y}}{s'_j}$ in cui $s'_j = \left[\frac{\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{n_j(n_j-1)} \right]^{0.5}$ è la deviazione standard del j -esimo gruppo con media \bar{y}_j e numerosità n_j e $\tilde{y} = \sum_{j=1}^J w_j \bar{y}_j$ con $w_j = \frac{\frac{1}{s_j'^2}}{\sum_{j=1}^J \frac{1}{s_j'^2}}$ peso del j -esimo gruppo per $j = 1, \dots, J$.

La statistica J_S osservata viene comparata con un particolare valore critico $h(\alpha)$ dato da:

$$\begin{aligned} h(\alpha) = & r + \frac{1}{2}(3\chi_4 + \chi_2)T + \frac{1}{16}(3\chi_4 + \chi_2)^2 \left(1 - \frac{k-3}{r}\right) T^2 \\ & + \frac{1}{2}(3\chi_4 + \chi_2)(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} - 4R_{11}^2) \\ & + (2R_{23} - 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2)(\chi_2 - 1) \\ & + \frac{1}{4}(-R_{12}^2 + 4R_{12}R_{11} - 2R_{12}R_{10} - 4R_{11}^2 + 4R_{11}R_{10} - R_{10}^2)(3\chi_4 - 2\chi_2 - 1) \\ & + (R_{23} - 3R_{22} + 3R_{21} - R_{20})(5\chi_6 + 2\chi_4 + \chi_2) \\ & + \frac{3}{16}(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} + R_{20})(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2) \\ & + \frac{1}{16}(-2R_{22} + 4R_{21} - R_{20} + 2R_{12}R_{10} - 4R_{11}R_{10} + R_{10}^2)(9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2) \\ & + \frac{1}{4}(-R_{22} + R_{11}^2)(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2) \\ & + \frac{1}{4}(R_{23} - R_{12}R_{11})(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2), \end{aligned}$$

dove, per ogni s e t , $R_{st} = \sum_{j=1}^J (n_j - 1)^{-s} w_j^t$, $\chi_{2s} = r^s / [(J-1)(J+1)\dots(J+2s-3)]$ in cui r è

il centile $(1 - \alpha)$ di una χ^2 con $J - 1$ gradi di libertà ed infine $T = \sum_{j=1}^J (1 - w_j)^2 / (n_j - 1)$.

Se il valore critico corrispondente risulta minore della statistica J_S osservata, allora l'ipotesi nulla $H_0 : \mu_1 = \mu_2 = \dots = \mu_J$ viene rifiutata in favore dell'ipotesi alternativa H_1 : almeno due gruppi hanno medie diverse tra loro.

Lo svantaggio di questo test è la complessità del calcolo del valore critico [11].

Il test è disponibile nel pacchetto *onewaytests* di R applicando la funzione *james.test()*.

Capitolo 4

METODI PER LA GESTIONE DELLA DIPENDENZA DEI DATI NELL'ANALISI DELLA VARIANZA

Quando si devono analizzare dei dati, è fondamentale scegliere il modello statistico più adatto alla struttura dei dati stessi. Nella situazione in cui le osservazioni a nostra disposizione non sono statisticamente indipendenti, l'uso dell'analisi della varianza tradizionale potrebbe non essere appropriata poiché assume l'indipendenza delle osservazioni tra i gruppi a confronto.

Al fine di ottenere conclusioni accurate e correttamente interpretabili, di seguito sono proposte delle alternative all'ANOVA tradizionale da utilizzare quando i dati presentano una struttura di dipendenza.

4.1 ANOVA a misure ripetute

L'ANOVA a misure ripetute è una tecnica statistica progettata per confrontare le medie di tre o più gruppi quando i dati sono osservazioni raccolte da soggetti che sono stati sottoposti a diverse condizioni o misurazioni ripetute nel tempo.

Si supponga di voler analizzare in n individui una certa variabile e che questa venga misurata più di due volte per ogni soggetto. I valori osservati, essendo stati misurati sugli stessi individui, sono tra loro appaiati. Se le misure effettuate fossero state due, il confronto tra i due gruppi si sarebbe potuto eseguire con un t-test per dati appaiati, ma con più di due misurazioni, si ricorre all'ANOVA a una via per misure ripetute.

Questa tecnica, largamente utilizzata nell'ambito delle scienze sperimentali e medico-sanitarie, tiene conto della dipendenza tra le osservazioni consentendo di isolare gli effetti

dei trattamenti diversi o del tempo. Viene così ridotto il rischio di concludere erroneamente differenze significative tra i gruppi migliorando la validità dei risultati rispetto all'ANOVA tradizionale.

Se n soggetti appartenenti al singolo campione sono sottoposti a J misure ripetute, i dati osservati possono essere inseriti in un tabella avente la seguente forma:

SOGGETTO	PRIMA 1	MISURA 2	...	MISURA J	TOTALE
1	y_{11}	y_{12}	...	y_{1J}	$T_{1.}$
2	y_{21}	y_{22}	...	y_{2J}	$T_{2.}$
...
n	y_{n1}	y_{n2}	...	y_{nJ}	$T_{n.}$
TOTALE	$T_{.1}$	$T_{.2}$...	$T_{.J}$	$T_{..}$

Dove:

- y_{ij} è la risposta dell' i -esimo soggetto alla j -esima misurazione, con $i = 1, \dots, n$ e $j = 1, \dots, J$,
- $T_{i.} = \sum_{j=1}^J y_{ij}$ rappresenta la somma totale delle misurazioni dell' i -esimo individuo, $i = 1, \dots, n$,
- $T_{.j} = \sum_{i=1}^n y_{ij}$ è la somma totale delle j -esime osservazioni, $j = 1, \dots, J$,
- $T_{..} = \sum_{i=1}^n \sum_{j=1}^J y_{ij}$ rappresenta la somma totale, con $i = 1, \dots, n$ e $j = 1, \dots, J$.

Siano quindi:

- $\bar{y}_{i.} = \frac{\sum_{j=1}^J y_{ij}}{J}$ la media delle misurazioni del soggetto i -esimo, per $i = 1, \dots, n$,
- $\bar{y}_{.j} = \frac{\sum_{i=1}^n y_{ij}}{n}$ la media nei soggetti per la j -esima misurazione, per $j = 1, \dots, J$,
- $\bar{y}_{..} = \frac{\sum_{i=1}^n \sum_{j=1}^J y_{ij}}{nJ}$ la media generale, con $i = 1, \dots, n$ e $j = 1, \dots, J$.

Per ottenere risultati dell'analisi affidabili è necessario verificare che le assunzioni dell'ANOVA a misure ripetute siano soddisfatte [13]. Il test presuppone che:

- le osservazioni all'interno del gruppo siano indipendenti. Per assicurarsi che tale assunzione venga verificata è sufficiente selezionare gli n soggetti dalla popolazione utilizzando il metodo del campionamento casuale.

- la variabile dipendente sia distribuita, almeno approssimativamente, in maniera normale. Tale assunto può essere verificato graficamente attraverso, ad esempio, il *Q-Q plot* o effettuando il test di Shapiro-Wilk.
- le varianze delle differenze tra tutte le combinazioni di gruppi correlati devono essere uguali. Tale presupposto è definito come ipotesi di sfericità o di omogeneità della varianza delle differenze. Per verificare ciò, si utilizza il test di sfericità di Mauchly.

La varianza totale $SST = \sum_{i=1}^n \sum_{j=1}^J (y_{ij} - \bar{y}_{..})^2$ si può dividere in tre quantità:

- la varianza dovuta alle diverse condizioni $SSTR = n \sum_{j=1}^J (\bar{y}_{.j} - \bar{y}_{..})^2$,
- la varianza dei soggetti $SSS = J \sum_{i=1}^n (\bar{y}_{i.} - \bar{y}_{..})^2$,
- la varianza residua $SSE = \sum_{i=1}^n \sum_{j=1}^J (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$.

Si ha dunque $SST = SSTR + SSS + SSE$.

Sia l'ipotesi nulla $H_0 : \mu_1 = \mu_2 = \dots = \mu_J$ e l'ipotesi alternativa $H_1 : \text{almeno una coppia è significativamente diversa}$.

La statistica test è data da:

$$F_{RM} = \frac{\frac{SSTR}{(J-1)}}{\frac{SSE}{(n-1)(J-1)}}.$$

Se le assunzioni del modello sono verificate, la statistica F_{RM} si distribuisce come una Fisher con $J - 1$ gradi di libertà al numeratore e $(n - 1)(J - 1)$ gradi di libertà al denominatore. Fissato il livello di significatività α , l'ipotesi nulla viene rifiutata se il valore F_{RM} osservato è superiore al valore critico $F_{(J-1), (n-1)(J-1); 1-\alpha}$.

Come si può notare, la statistica test F_{RM} è strutturata in modo tale che le differenze individuali dei soggetti non influiscano né al numeratore né al denominatore. Se esiste un effetto del trattamento (o del tempo), esso agisce solo al numeratore facendo aumentare il valore della statistica test.

Quando il presupposto di sfericità non è soddisfatto, l'approccio da adottare è quello di ridurre i gradi di libertà sia al numeratore che al denominatore moltiplicandoli per ϵ , dove ϵ misura la gravità del problema di sfericità. Di conseguenza, F_{RM} si distribuisce approssimativamente come una $F_{\epsilon(J-1), \epsilon(n-1)(J-1)}$. Poiché il valore di ϵ non è noto, deve essere stimato sulla base dei dati. Il valore di ϵ è compreso tra $1/(J - 1)$ e 1. Quando l'ipotesi di sfericità è soddisfatta, ϵ vale 1 mentre, più grande è la violazione, minore sarà il valore di ϵ . I due metodi più comuni

per stimare ϵ , entrambi calcolabili attraverso software statistici, sono la stima di Huynh-Feldt e quella di Greenhouse-Geisser [14].

Per effettuare l'ANOVA a misure ripetute è possibile utilizzare la funzione *aov()* presente nel pacchetto standard (*stats*) di R oppure applicando la funzione *anova_test()*, inclusa nel pacchetto *rstatix*, in cui nell'output viene riportato automaticamente anche il risultato del test di Mauchly.

4.2 Test di Friedman

Quando le ipotesi dell'ANOVA a una via per misure ripetute non sono soddisfatte, è possibile ricorrere alla sua versione non parametrica, ossia il test di Friedman.

Tale test permette di eseguire, sulla base dei ranghi, il confronto tra più gruppi di misurazioni effettuate sugli stessi soggetti in condizioni diverse o in vari istanti di tempo anche quando i dati non sono distribuiti normalmente.

Un gruppo di n soggetti sono stati sottoposti a J diverse misurazioni. Sia quindi y_{ij} la singola osservazione a nostra disposizione con $i = 1, \dots, n$ e $j = 1, \dots, J$.

Sia l'ipotesi nulla H_0 : non ci sono differenze significative tra i gruppi, posta contro l'ipotesi alternativa H_1 : ci sono differenze tra almeno due gruppi.

I ranghi r_{ij} vengono assegnati ai dati di ciascun soggetto, valutando separatamente ciascun soggetto e successivamente si esegue la somma dei ranghi per ogni gruppo.

La statistica test di Friedman è:

$$\chi_F^2 = \frac{12n}{J(J+1)} \sum_{j=1}^J \frac{R_j^2}{n} - 3(J+1)$$

dove J è il numero di gruppi, n è la quantità di soggetti ed infine R_j è la somma dei ranghi del j -esimo gruppo, per $j = 1, \dots, J$.

La statistica χ_F^2 , quando la numerosità dei dati è abbastanza grande, tende a distribuirsi come una χ^2 con $(J-1)$ gradi di libertà. All'aumentare di n e J , l'approssimazione alla distribuzione della χ^2 migliora. Per piccoli valori di n e J invece, sono disponibili i valori critici corrispondenti tabulati [15].

Il test è disponibile nel pacchetto standard (*stats*) di R richiamando la funzione *friedman.test()*.

4.3 Modelli ad effetti misti per risposte normali

Mentre nell'analisi della varianza a misure ripetute i soggetti con osservazioni mancanti vengono eliminati dall'analisi, l'approccio con modelli ad effetti misti permette di includere nello studio

anche i soggetti con osservazioni mancanti purché siano casualmente mancanti, ciò significa che la probabilità che un'osservazione sia mancante non dipende dal valore delle risposte non osservate.

In generale, un modello lineare con effetti misti per risposte normali può essere scritto nella forma

$$Y_{ij} = x_{ij}\beta + z_{ij}u_i + \epsilon_{ij}$$

dove β è un vettore p -dimensionale di effetti fissi, $u_i \sim N_q(0, \Sigma_u)$ è vettore q -dimensionale di effetti casuali, z_{ij} è un vettore q -dimensionale di covariate associate agli effetti casuali ed infine $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ indipendente da u_i .

Questo modello prevede che $E(Y_{ij}) = \mu_{ij} = x_{ij}\beta$. Il termine $z_{ij}u_i$ descrive la variabilità tra i soggetti, mentre ϵ_{ij} descrive la variabilità interna ai soggetti.

Nello specifico, se n soggetti appartenenti ad un singolo campione, estratti casualmente dalla popolazione di interesse, vengono misurati in J condizioni diverse (o in J istanti di tempo), il modello adeguato risulta

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \epsilon_{ij} \quad (4.1)$$

dove y_{ij} è la risposta dell' i -esimo soggetto alla j -esima misurazione, β_0 è l'intercetta (effetto fisso), $\beta_1 x_{ij}$ è l'effetto fisso del trattamento (o del tempo), u_i rappresenta l'effetto casuale del soggetto i -esimo e ϵ_{ij} è un errore casuale, con $i = 1, \dots, n$ e $j = 1, \dots, J$.

Poiché i soggetti sono estratti casualmente dalla popolazione, gli effetti u_i possono essere considerati realizzazioni di variabili casuali indipendenti e identicamente distribuite con media pari a zero. Si assume perciò $u_i \sim N(0, \sigma_u^2)$ e che gli errori ϵ_{ij} siano anch'essi realizzazioni di variabili casuali indipendenti con distribuzione $N(0, \sigma_\epsilon^2)$. Inoltre u_i e ϵ_{ij} sono indipendenti tra loro al variare di i e j .

Risulta allora che

$$E(y_{ij}) = \beta_0 + \beta_1 x_{ij} \quad \text{e} \quad Var(y_{ij}) = \sigma_u^2 + \sigma_\epsilon^2.$$

Le osservazioni relative ai soggetti diversi sono indipendenti ($Cov(y_{ij}, y_{hj}) = 0$ con $i \neq h$), mentre la covarianza tra misurazioni diverse sullo stesso soggetto è, con $j \neq h$,

$$Cov(y_{ij}, y_{ih}) = E((u_i + \epsilon_{ij})(u_i + \epsilon_{ih})) = \sigma_u^2.$$

Quindi, indipendentemente da j e h , $Cor(y_{ij}, y_{ih}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2} = \rho$.

Di conseguenza si ha

$$Var(y_i) = V_i = (\sigma_u^2 + \sigma_\epsilon^2) \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \text{ con } \rho > 0. \quad (4.2)$$

Il simbolo ρ indica il coefficiente di correlazione intra-classe e rappresenta la quota di varianza totale dovuta alla variabilità delle misure effettuate sullo stesso soggetto.

La matrice V_i in (4.2) è detta con struttura di equicorrelazione o scambiale (A.Salvan *et al.*, 2020 [16]).

Un aspetto interessante dell'approccio con modelli ad effetti casuali è la libertà di usare diverse strutture di correlazione per modellare le misure ripetute su un soggetto. Una possibilità è, appunto, la struttura scambiale, usata quando si presuppone che tutte le correlazioni a coppie siano uguali. Un'altra possibilità consiste nell'utilizzare una struttura autoregressiva: spesso, con osservazioni longitudinali, è ragionevole ipotizzare che la correlazione tra le osservazioni relative ad uno stesso soggetto diminuisca all'aumentare della distanza temporale tra le osservazioni stesse. Con essa, se ρ indica la correlazione per le osservazioni in un tempo precedente, allora ρ^2 è la correlazione per le osservazioni nei due tempi precedenti, ρ^3 nei tre tempi precedenti, e così via. Infine è possibile usare una matrice non strutturata in cui non si fa alcuna ipotesi circa il modello di correlazione.

Per valutare complessivamente se il trattamento (o il tempo) ha un effetto significativo, è possibile confrontare il modello ad effetti misti completo (4.1) con il modello ridotto, cioè senza il fattore trattamento (o tempo), del tipo

$$y_{ij} = \beta_0 + u_i + \epsilon_{ij},$$

utilizzando un test di verosimiglianza.

Confrontando il modello completo con quello ridotto, si può determinare se l'inclusione della variabile trattamento (o tempo) migliora significativamente il modello fornendo evidenze statistiche circa il suo effetto.

Per l'adattamento di modelli lineari normali con effetti misti, è possibile utilizzare la funzione *lme* della libreria *nlme* di R.

Capitolo 5

STUDIO DI SIMULAZIONE

Sebbene l'analisi della varianza sia uno strumento statistico comune per confrontare le medie di più gruppi, esso potrebbe essere inappropriato quando il presupposto di omogeneità della varianza o quello di indipendenza delle osservazioni sono violati.

Il seguente studio di simulazione, eseguito in R (versione 4.4.0), ha l'obiettivo di confrontare diversi test statistici in molteplici circostanze per analizzare il loro comportamento e scoprire il metodo più adatto nei vari contesti.

Il disegno sperimentale prevede l'analisi di situazioni che derivano dalla combinazione dei seguenti fattori:

1. Numero di gruppi

- Saranno considerati scenari con 3, 5 e 10 gruppi.

2. Numerosità dei gruppi

- se gruppi bilanciati, tutti avranno la stessa numerosità (5, 10 o 20 osservazioni),
- se gruppi sbilanciati, la numerosità varia tra i gruppi ($(n_1, n_2, n_3) = (5, 10, 20)$ se i gruppi sono 3, $(n_1, n_2, n_3, n_4, n_5) = (5, 5, 10, 20, 20)$ se sono 5 e $(n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10}) = (5, 5, 5, 5, 10, 10, 10, 20, 20, 20)$ se sono 10).

3. Varianza

- se omoschedasticità, tutti i gruppi hanno la stessa deviazione standard ($\sigma = 1$),
- se eteroschedasticità moderata, alcuni gruppi hanno deviazione standard $\sigma = 1$, altri $\sigma = 1.5$,
- se eteroschedasticità elevata, alcuni gruppi hanno deviazione standard $\sigma = 1$, altri $\sigma = 3$.

4. Relazione tra numerosità e varianza

- se relazione positiva, i gruppi con un maggior numero di osservazioni hanno anche varianza maggiore,
- se relazione negativa, i gruppi con numerosità maggiore hanno varianza minore.

5. Tipo di osservazioni

- se osservazioni indipendenti, ogni osservazione è indipendente dalle altre,
- se misure ripetute, le osservazioni sono misure ripetute sugli stessi soggetti, quindi correlate.

In ogni scenario il test è considerato significativo se il p-value è minore del livello di significatività nominale $\alpha = 0.05$.

Nella situazione in cui l'ipotesi nulla ($H_0 : \mu_1 = \dots = \mu_J$) è vera, il tasso di rifiuto dell'ipotesi nulla sarà considerato come il tasso di errore empirico di I tipo per ciascun metodo statistico applicato. Il test che presenta il tasso di errore di I tipo più vicino al valore nominale $\alpha = 0.05$, è considerato il migliore tra quelli confrontati.

Nella situazione in cui è l'ipotesi alternativa ($H_1 =$ non tutte le medie sono uguali) ad essere vera, il tasso di rifiuto dell'ipotesi nulla sarà considerato come la potenza statistica per ciascun test. Il metodo che ha la potenza maggiore, è considerato il migliore tra quelli comparati [12].

In questo studio, per il calcolo del tasso di errore empirico di I tipo, tutte le medie sono poste uguali a 0, mentre, per il calcolo della potenza, un solo gruppo tra quelli confrontati, sia in caso di 3, 5 e 10 gruppi, avrà la media diversa da 0, ossia 2.

I dati casuali, che hanno distribuzione normale, sono stati generati utilizzando le funzioni *rnorm()*, presente nella libreria standard di R (*stats*), e *rmvnorm()* del pacchetto *mvtnorm*.

Per ogni scenario considerato sono state effettuate 100000 simulazioni.

Se le osservazioni sono indipendenti, i test confrontati sono l'ANOVA tradizionale, il test non parametrico di Kruskal-Wallis, il test di Welch, il test di Welch con medie troncate e varianze winsorizzate, il test di Alexander-Govern e il test di James del secondo ordine. All'interno della legenda dei grafici saranno codificati rispettivamente con 'ANOVA', 'KRUSKAL-WALLIS', 'WELCH', 'WELCH DW', 'ALEXANDER-GOVERN' e 'JAMES'.

Se i dati corrispondono a misure ripetute, si confronterà il metodo dell'analisi della varianza tradizionale con quello dei modelli ad effetti misti per risposte normali. Nei grafici, questi approcci saranno indicati rispettivamente con le diciture "ANOVA" ed "EFFETTI MISTI".

Di seguito vengono descritti gli scenari presi in considerazione nello studio di simulazione:

A1. G gruppi bilanciati di numerosità n , dove $n \in \{5, 10, 20\}$, e deviazioni standard $\sigma_j = 1$ per $j = 1, \dots, G$

- se $G = 3$, i gruppi hanno medie $(\mu_1, \mu_2, \mu_3) = (m, 0, 0)$,
- se $G = 5$, i gruppi hanno medie $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (m, 0, 0, 0, 0)$,
- se $G = 10$, i gruppi hanno medie pari a $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9, \mu_{10}) = (m, 0, 0, 0, 0, 0, 0, 0, 0, 0)$.

Data la $Cov(y_{ij}, y_{ih}) = r$ con $j \neq k$, se $r = 0$ allora si tratta di dati indipendenti, ossia osservazioni relative a soggetti diversi, altrimenti, se $r \neq 0$, i dati sono misurazioni sugli stessi soggetti (misure ripetute) in condizioni o tempi diversi.

A2. G gruppi sbilanciati con deviazioni standard $\sigma_j = 1$ per $j = 1, \dots, G$ e $Cov(y_{ij}, y_{ih}) = r$ con $j \neq k$

- se $G = 3$, i gruppi hanno numerosità $(n_1, n_2, n_3) = (a, b, c)$ e medie $(\mu_1, \mu_2, \mu_3) = (0, m, 0)$,
- se $G = 5$, i gruppi hanno numerosità $(n_1, n_2, n_3, n_4, n_5) = (a, a, b, c, c)$ e medie $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (0, 0, m, 0, 0)$,
- se $G = 10$, i gruppi hanno numerosità $(n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10}) = (a, a, a, a, b, b, b, c, c, c)$ e medie pari a $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9, \mu_{10}) = (0, 0, 0, 0, m, 0, 0, 0, 0, 0)$.

B1. G gruppi bilanciati di numerosità n , dove $n \in \{5, 10, 20\}$, e $Cov(y_{ij}, y_{ih}) = r$ con $j \neq k$

- se $G = 3$, i gruppi hanno medie $(\mu_1, \mu_2, \mu_3) = (m, 0, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3) = (s, 1, 1)$,
- se $G = 5$, i gruppi hanno medie $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (m, 0, 0, 0, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5) = (s, s, 1, 1, 1)$,
- se $G = 10$, i gruppi hanno medie pari a $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9, \mu_{10}) = (m, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8, \sigma_9, \sigma_{10}) = (s, s, s, s, s, 1, 1, 1, 1, 1)$.

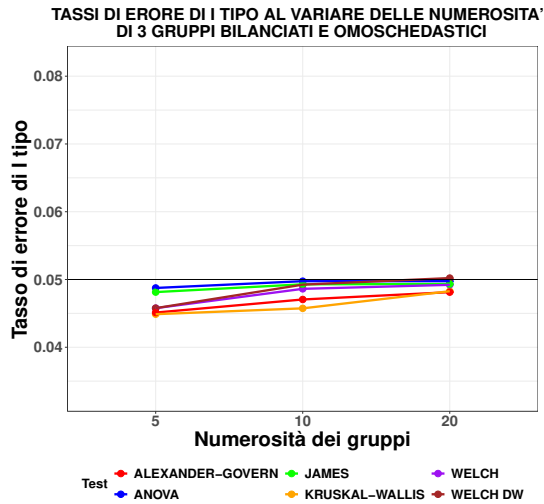
B2. G gruppi sbilanciati con relazione positiva tra varianza e numerosità interna ai gruppi e $Cov(y_{ij}, y_{ih}) = r$ con $j \neq k$

- se $G = 3$, i gruppi hanno numerosità $(n_1, n_2, n_3) = (a, b, c)$, medie $(\mu_1, \mu_2, \mu_3) = (0, m, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3) = (1, 1, s)$,
- se $G = 5$, i gruppi hanno numerosità $(n_1, n_2, n_3, n_4, n_5) = (a, a, b, c, c)$, medie $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (0, 0, m, 0, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5) = (1, 1, 1, s, s)$,
- se $G = 10$, i gruppi hanno numerosità $(n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10}) = (a, a, a, a, b, b, b, c, c, c)$, medie pari a $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9, \mu_{10}) = (0, 0, 0, 0, m, 0, 0, 0, 0, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8, \sigma_9, \sigma_{10}) = (1, 1, 1, 1, 1, 1, s, s, s, s)$.

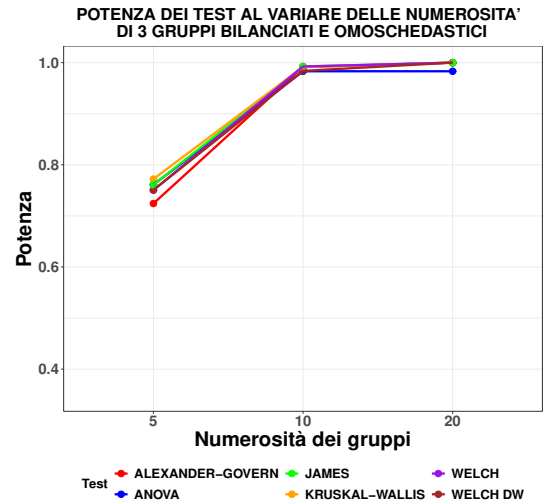
B3. G gruppi sbilanciati con relazione negativa tra numerosità interna ai gruppi e varianza e $Cov(y_{ij}, y_{ih}) = r$ con $j \neq k$

- se $G = 3$, i gruppi hanno numerosità $(n_1, n_2, n_3) = (a, b, c)$, medie $(\mu_1, \mu_2, \mu_3) = (0, m, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3) = (s, 1, 1)$,
- se $G = 5$, i gruppi hanno numerosità $(n_1, n_2, n_3, n_4, n_5) = (a, a, b, c, c)$, medie $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (0, 0, m, 0, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5) = (s, s, 1, 1, 1)$,
- se $G = 10$, i gruppi hanno numerosità $(n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8, n_9, n_{10}) = (a, a, a, a, b, b, b, c, c, c)$, medie pari a $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9, \mu_{10}) = (0, 0, 0, 0, m, 0, 0, 0, 0, 0)$ e deviazioni standard $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8, \sigma_9, \sigma_{10}) = (s, s, s, s, 1, 1, 1, 1, 1, 1)$.

5.1 Gruppi bilanciati/sbilanciati omoschedastici con osservazioni indipendenti

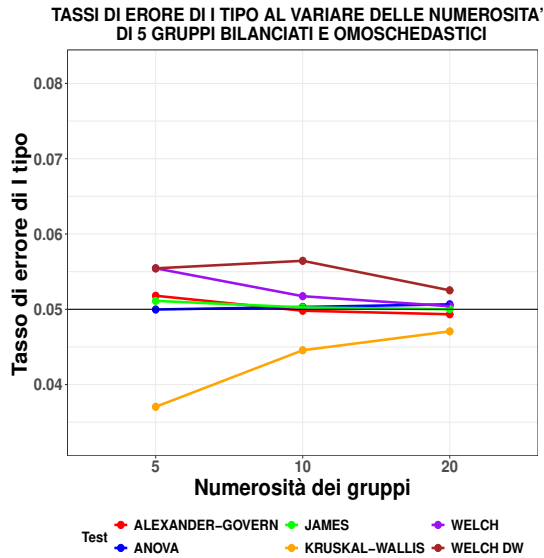


(a) Scenario A1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 0$ e $r = 0$.

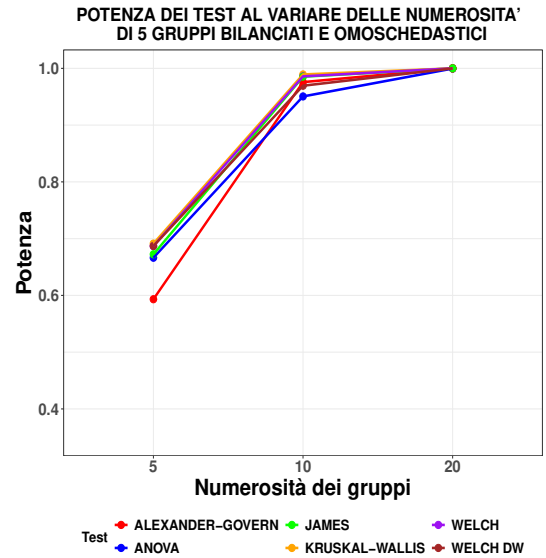


(b) Scenario A1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 2$ e $r = 0$.

Figura 5.1: Tasso di errore di I tipo e potenza in presenza di 3 gruppi bilanciati e omoschedastici al variare della loro numerosità, osservazioni indipendenti.

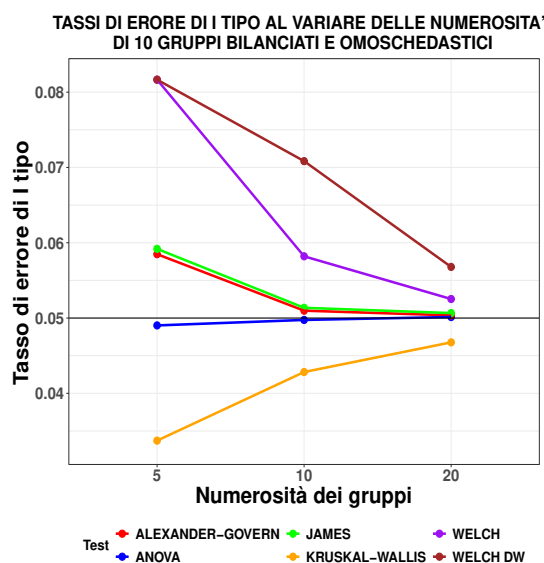


(a) Scenario A1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 0$ e $r = 0$.

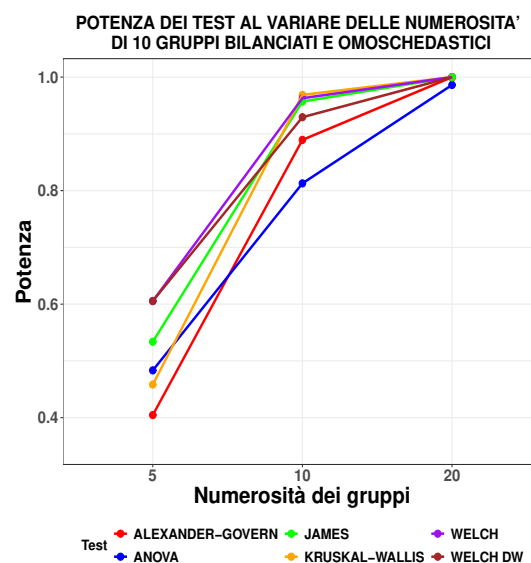


(b) Scenario A1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 2$ e $r = 0$.

Figura 5.2: Tasso di errore di I tipo e potenza in presenza di 5 gruppi bilanciati e omoschedastici al variare della loro numerosità, osservazioni indipendenti.



(a) Scenario A1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 0$ e $r = 0$.



(b) Scenario A1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 2$ e $r = 0$.

Figura 5.3: Tasso di errore di I tipo e potenza in presenza di 10 gruppi bilanciati e omoschedastici al variare della loro numerosità, osservazioni indipendenti.

Osservando la Figura 5.1a, con 3 gruppi bilanciati di varia numerosità e omoschedastici, si nota che il tasso di errore di I tipo di tutti i test considerati è vicino al valore nominale 0.05. Tuttavia, confrontando questo grafico con le Figure 5.2a e 5.3a, emerge che all'aumentare del numero di gruppi la situazione cambia: se le numerosità sono elevate, i test mantengono tassi di errore di I tipo vicini al valore nominale, ma al diminuire delle numerosità, questi tassi si allontanano progressivamente dal valore 0.05, ad eccezione del test ANOVA.

In particolare, il test di Kruskal-Wallis diventa via via più conservativo, mentre i test di James del secondo ordine, di Alexander-Gover e, ancor più marcatamente, quello di Welch e di Welch con medie troncate e varianze winsorizzate, sempre più liberali. Il test ANOVA, invece, mantiene un tasso di errore di I tipo molto prossimo allo 0.05, indipendentemente dal numero di gruppi e dalle loro numerosità.

Per quanto riguarda la potenza (Figure 5.1b, 5.2b e 5.3b), essa per tutti i test aumenta al crescere delle numerosità, a prescindere dal numero di gruppi. Tuttavia, soprattutto per numerosità ridotte, all'incremento del numero di gruppi, la potenza tende a diminuire.

È importante ricordare che la proporzione di gruppi aventi media diversa non è la stessa, essa diminuisce al crescere del numero di gruppi poiché il disegno sperimentale prevede un solo gruppo con media differente sia che i gruppi a confronto siano 3, 5 o 10.

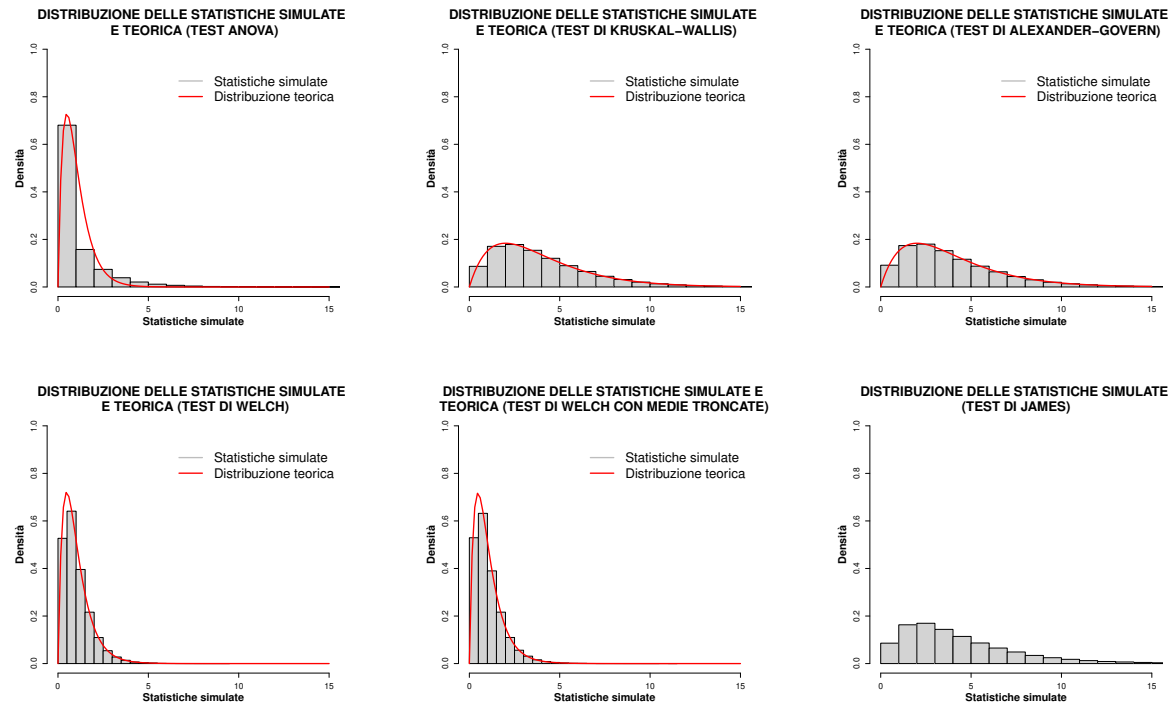


Figura 5.4: Confronto della distribuzione delle statistiche simulate e teorica sotto H_0 dei test, scenario A1 con $G = 5$, $n = 20$, $m = 0$ e $r = 0$.

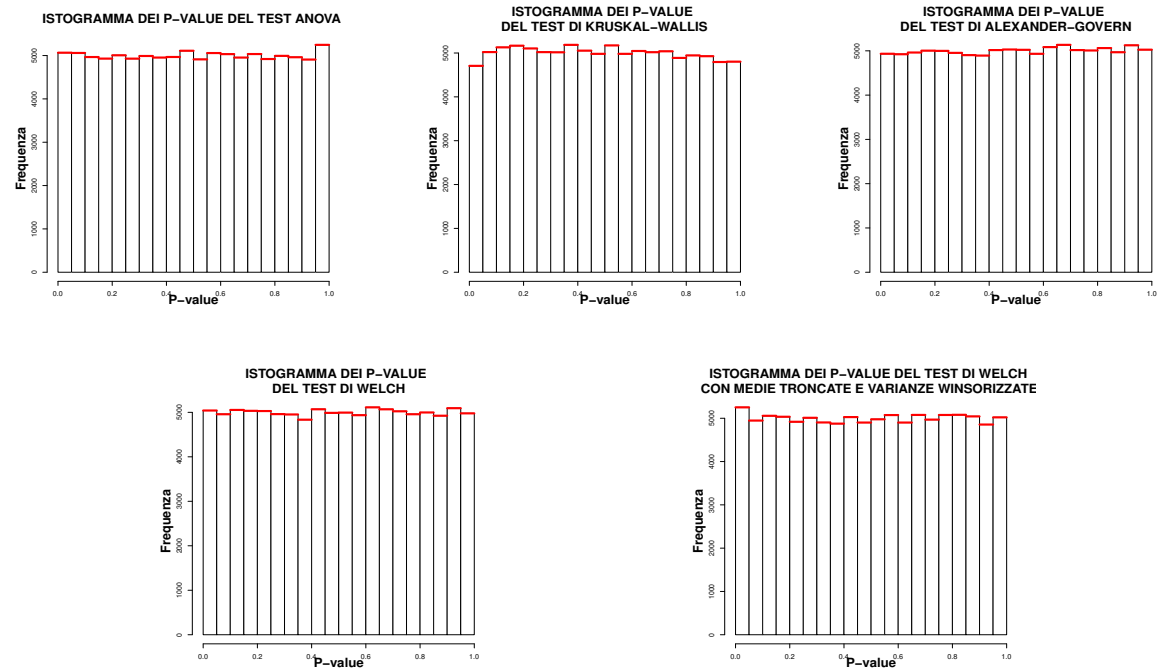
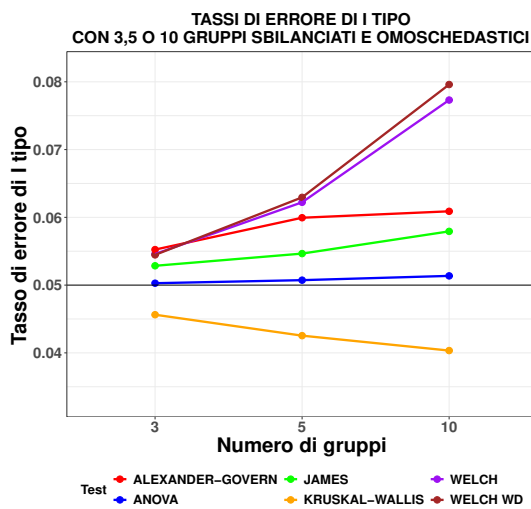


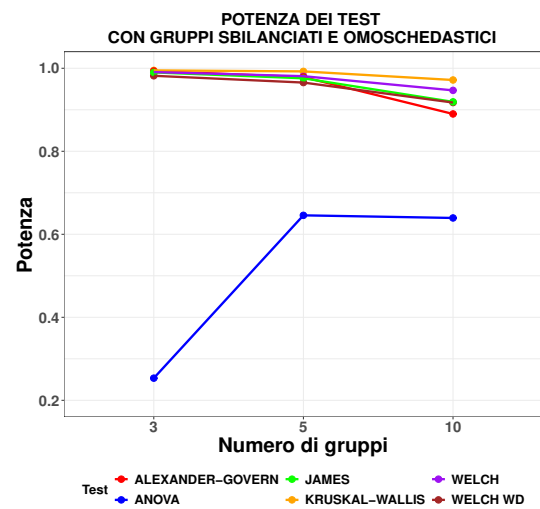
Figura 5.5: Confronto della distribuzione dei p-value sotto H_0 dei test, scenario A1 con $G = 5$, $n = 20$, $m = 0$ e $r = 0$.

Confrontare la distribuzione delle statistiche test simulate con quella teorica permette di verificare se il test si comporta come previsto sotto l'ipotesi nulla. Se le statistiche test seguono la distribuzione teorica prevista, il loro istogramma dovrebbe corrispondere strettamente alla curva della distribuzione teorica sovrapposta. Inoltre, sotto l'ipotesi nulla, i p-value dovrebbero essere distribuiti uniformemente tra 0 e 1. Questa condizione si traduce in un istogramma dei p-value ottenuti dalle simulazioni con barre di altezza simile. Una deviazione significativa da tali situazioni potrebbe indicare che le assunzioni del test non sono soddisfatte e, di conseguenza, le conclusioni che si ricavano potrebbero non essere affidabili. Le due tipologie di grafici forniscono informazioni complementari sulla correttezza del test. Ad esempio, se la distribuzione delle statistiche simulate mostra code più pesanti rispetto a quella teorica sotto l'ipotesi nulla, ciò si riflette in barre più alte all'inizio dell'istogramma dei p-value indicando una maggiore presenza di p-value piccoli rispetto a quanto previsto. Quindi la distribuzione teorica sottostima la probabilità di ottenere valori estremi nelle statistiche test. Del test di James del secondo ordine viene riportato solamente l'istogramma delle statistiche simulate in quanto la statistica test non è associata a una distribuzione teorica specifica sotto l'ipotesi nulla, impedendo così il calcolo dei p-value.

Le Figure 5.5 e 5.4 mostrano che, in presenza di 5 gruppi bilanciati e omoschedastici, tutti i test seguono la distribuzione teorica sotto l'ipotesi nulla e i p-value si distribuiscono uniformemente, suggerendo che le conclusioni che ne ricaviamo in tali condizioni sono affidabili.



(a) Scenario A2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$ e $r = 0$.



(b) Scenario A2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$ e $r = 0$.

Figura 5.6: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati e omoschedastici, osservazioni indipendenti.

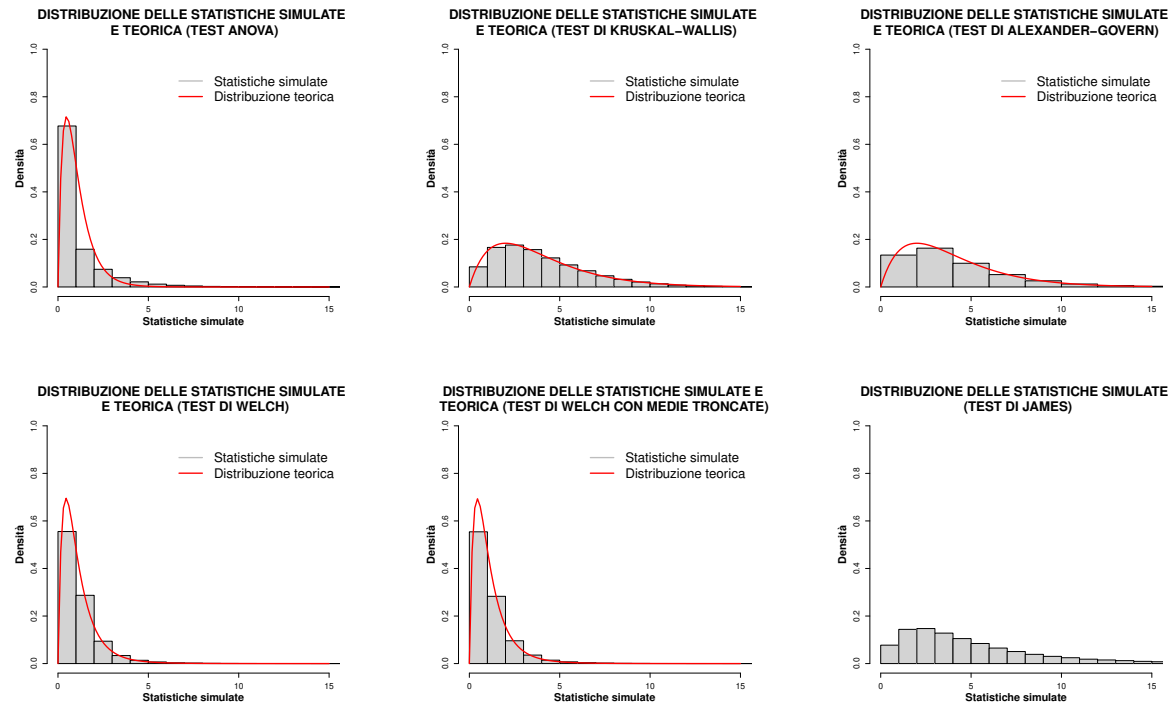


Figura 5.7: Confronto della distribuzione delle statistiche simulate e teorica sotto H_0 dei test, scenario A2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$ e $r = 0$.

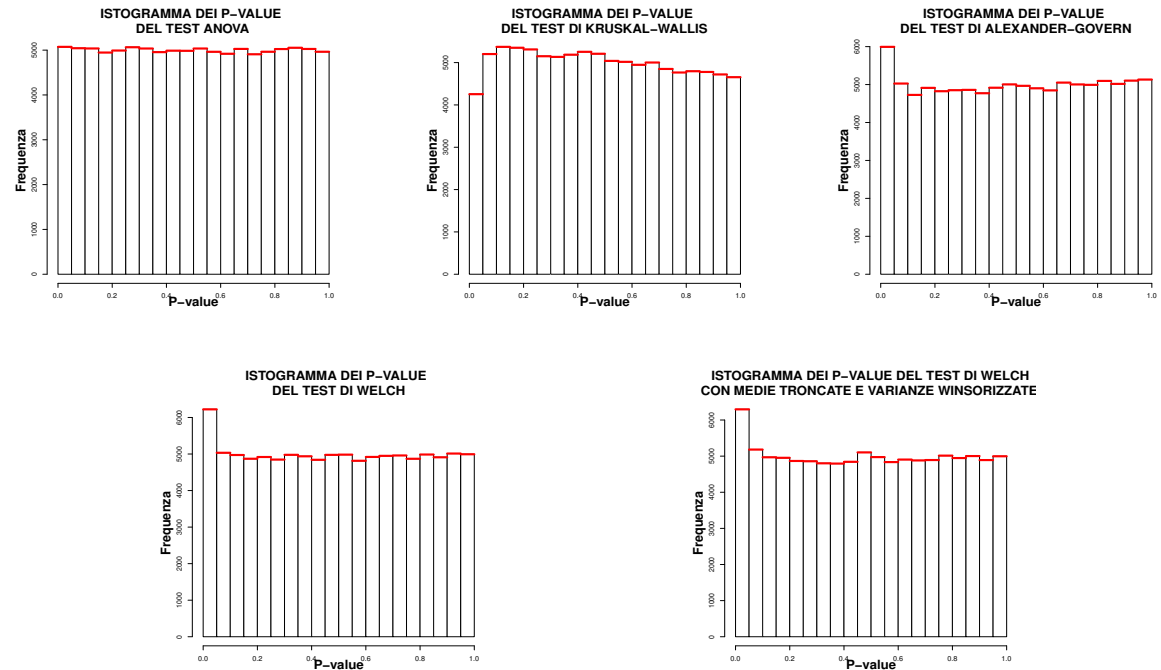


Figura 5.8: Confronto della distribuzione dei p-value sotto H_0 dei test, scenario A2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$ e $r = 0$.

Considerando gruppi omoschedastici ma sbilanciati (Figura 5.6a), con l'aumentare del numero di gruppi, i tassi di errore di I tipo dei test si allontanano dal valore nominale, eccetto quello dell'ANOVA che mantiene un valore prossimo allo 0.05. La potenza del test ANOVA, però, è nettamente inferiore in confronto ai restanti test le cui potenze diminuiscono leggermente al crescere del numero di gruppi (Figura 5.6b).

La distribuzione delle statistiche F simulate rispecchia la distribuzione teorica e i p-value dell'ANOVA sembrano distribuirsi uniformemente tra 0 e 1, a differenza di quanto si verifica per gli altri test. Il test di Alexander-Govern, di Welch e di Welch con medie troncate e varianze winsorizzate mostrano una maggiore quantità di p-value ridotti, mentre il test di Kruskal-Wallis presenta una prima barra inferiore (Figure 5.7 e 5.8).

5.2 Gruppi bilanciati/sbilanciati eteroschedastici con osservazioni indipendenti

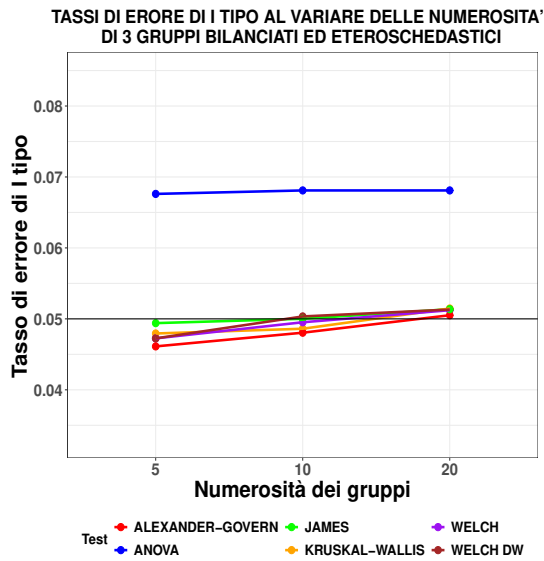
Nel contesto in cui vengono esaminati tre gruppi bilanciati e moderatamente eteroschedastici (Figura 5.9a), il test ANOVA presenta un tasso di errore di I tipo superiore al valore nominale, indicando un test troppo liberale. Tale effetto diminuisce con l'aumentare del numero di gruppi, tanto che con 10 gruppi (Figura 5.11a) il suo valore è prossimo allo 0.05, mentre la numerosità non sembra influenzare significativamente questo valore. Gli altri test, invece, mostrano tassi di errore di I tipo vicini al valore nominale quando i gruppi a confronto sono pochi (Figura 5.9a). Tuttavia, al crescere del numero di gruppi, più diminuiscono le osservazioni all'interno di ciascun gruppo, più i valori dei tassi si allontanano dallo 0.05. In particolare il test di Kruskal-Wallis diventa troppo conservativo, mentre i restanti test diventano troppo liberali.

Quando i gruppi bilanciati confrontati presentano un'elevata eteroschedasticità, gli effetti della violazione dell'ipotesi di omogeneità sul test ANOVA sono i medesimi ma più pronunciati. Infatti, confrontando la Figura 5.9a con la Figura 5.14a, si nota che, al crescere delle differenze tra le varianze, il test ANOVA diventa sempre più liberale, mostrando un tasso di errore di I tipo sempre più grande rispetto al valore nominale.

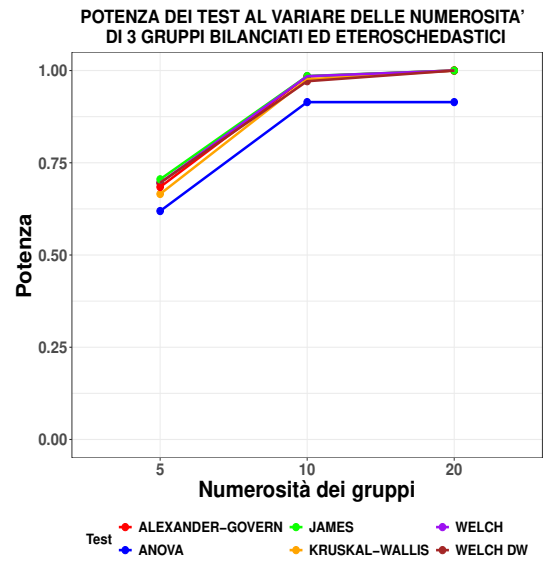
La potenza dell'ANOVA risulta inferiore in relazione a quella dei test alternativi per ogni combinazione di numero di gruppi e numerosità. Comunque, tutti i test riportano un aumento della potenza con il crescere delle numerosità e una diminuzione della potenza al crescere del numero di gruppi, soprattutto se di ridotta numerosità (Figure 5.9b, 5.10b e 5.11b).

In presenza di eteroschedasticità elevata, il test ANOVA continua a mostrare una potenza peggiore, seguito dal test di Kruskal-Wallis che però migliora significativamente al crescere

delle numerosità (Figure 5.14b, 5.15b e 5.16b).

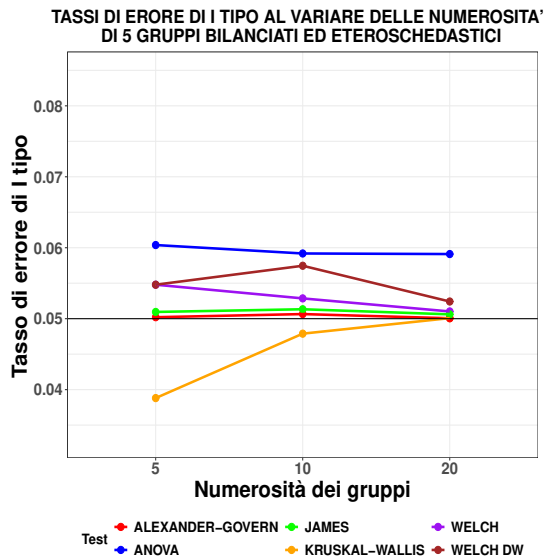


(a) Scenario B1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 1.5$ e $r = 0$.

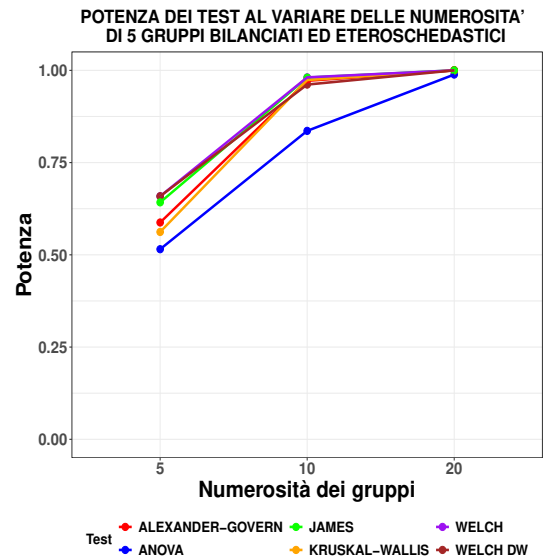


(b) Scenario B1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 1.5$ e $r = 0$.

Figura 5.9: Tasso di errore di I tipo e potenza in presenza di 3 gruppi bilanciati e moderatamente eteroschedastici al variare della loro numerosità, osservazioni indipendenti.

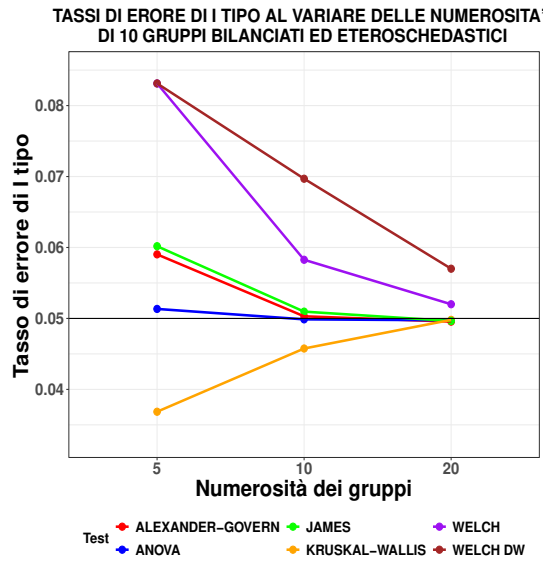


(a) Scenario B1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 1.5$ e $r = 0$.

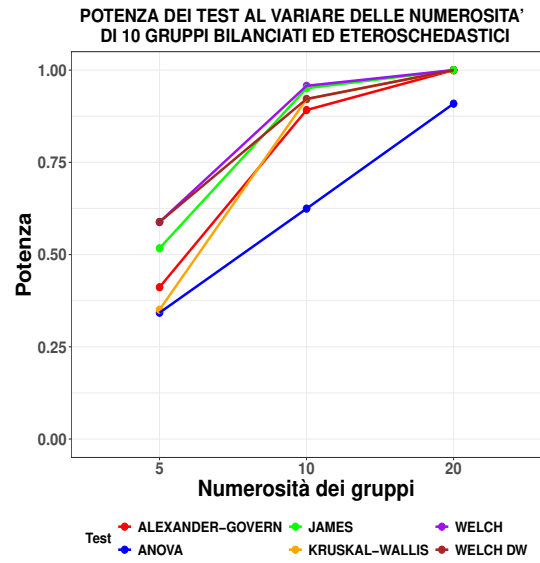


(b) Scenario B1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 1.5$ e $r = 0$.

Figura 5.10: Tasso di errore di I tipo e potenza in presenza di 5 gruppi bilanciati e moderatamente eteroschedastici al variare della loro numerosità, osservazioni indipendenti.



(a) Scenario B1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 1.5$ e $r = 0$.



(b) Scenario B1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 1.5$ e $r = 0$.

Figura 5.11: Tasso di errore di I tipo e potenza in presenza di 10 gruppi bilanciati e moderatamente eteroschedastici al variare della loro numerosità, osservazioni indipendenti.

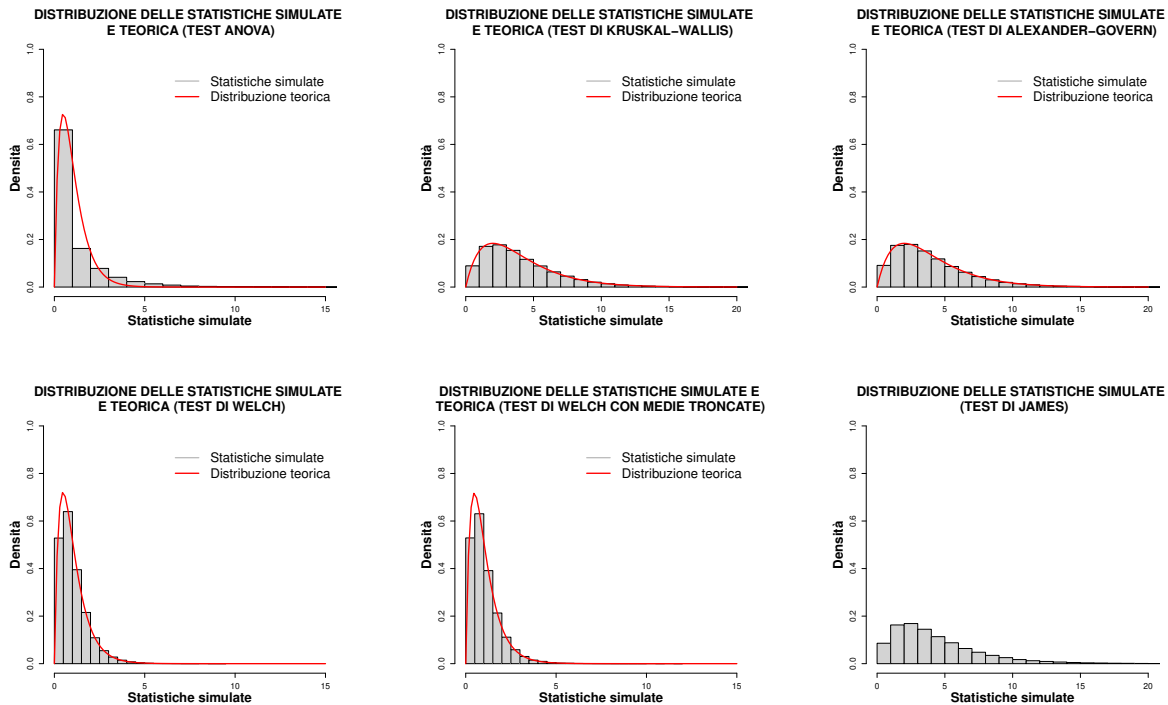


Figura 5.12: Confronto della distribuzione delle statistiche simulate e teorica sotto H_0 dei test, scenario B1 con $G = 5$, $n = 20$, $m = 0$, $s = 1.5$ e $r = 0$.

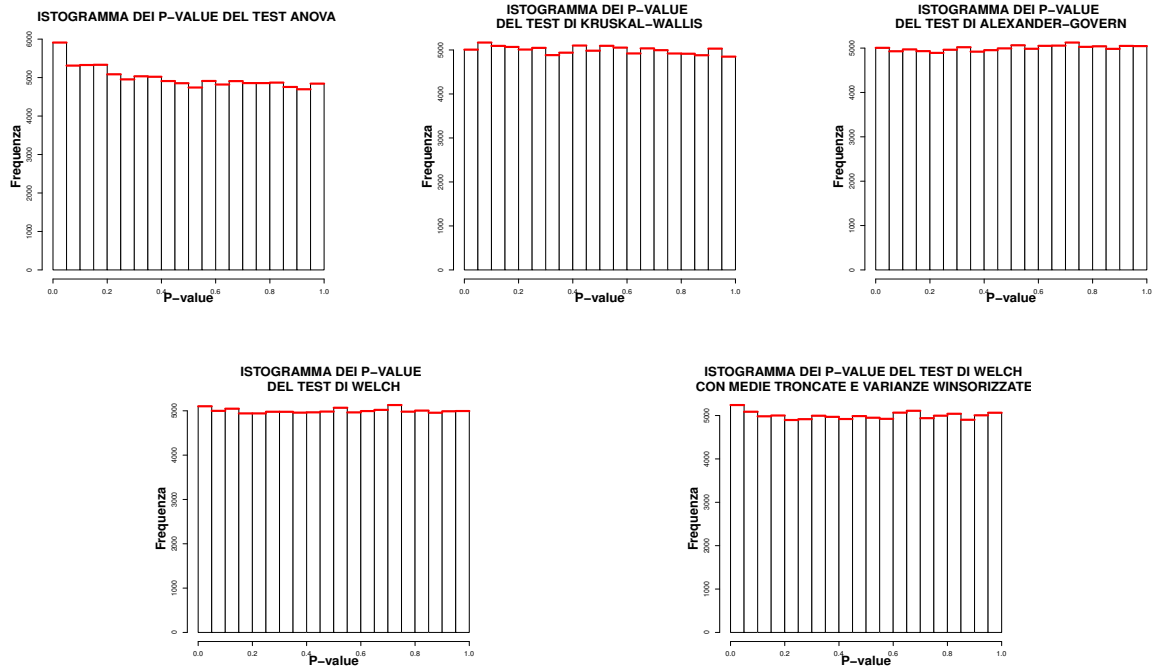
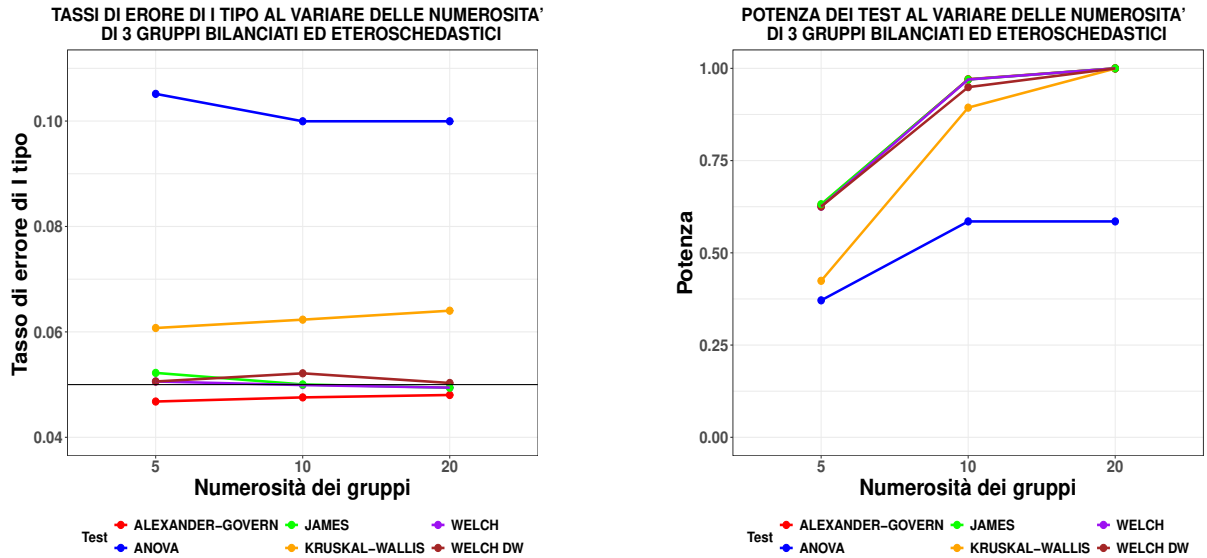


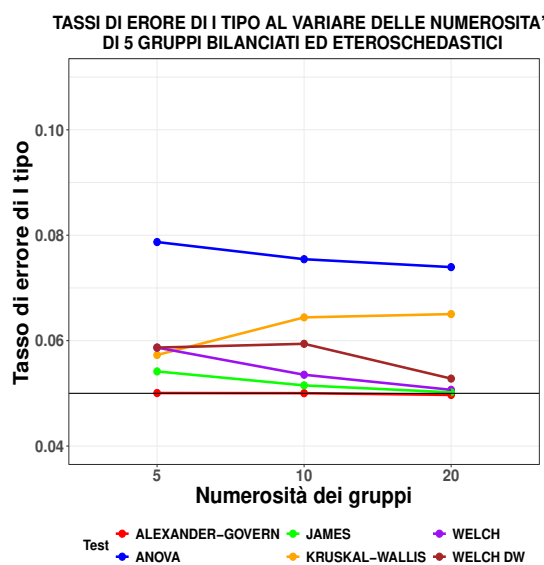
Figura 5.13: Confronto della distribuzione dei p-value sotto H_0 dei test, scenario B1 con $G = 5$, $n = 20$, $m = 0$, $s = 1.5$ e $r = 0$.



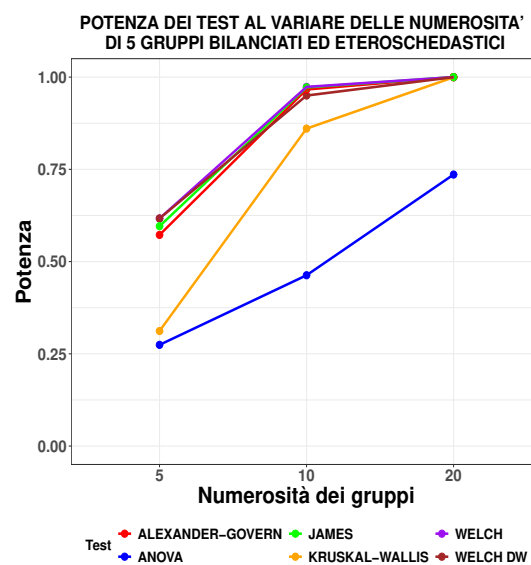
(a) Scenario B1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 3$ e $r = 0$.

(b) Scenario B1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 3$ e $r = 0$.

Figura 5.14: Tasso di errore di I tipo e potenza in presenza di 3 gruppi bilanciati con eteroschedasticità elevata al variare della loro numerosità, osservazioni indipendenti.

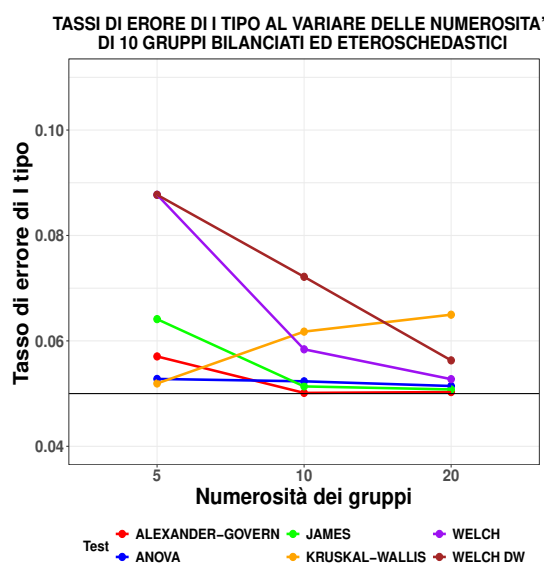


(a) Scenario B1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 3$ e $r = 0$.

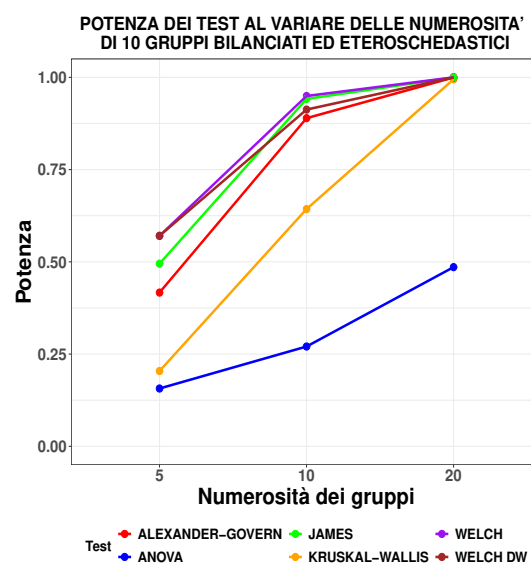


(b) Scenario B1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 3$ e $r = 0$.

Figura 5.15: Tasso di errore di I tipo e potenza in presenza di 5 gruppi bilanciati con eteroschedasticità elevata al variare della loro numerosità, osservazioni indipendenti.



(a) Scenario B1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 3$ e $r = 0$.



(b) Scenario B1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 3$ e $r = 0$.

Figura 5.16: Tasso di errore di I tipo e potenza in presenza di 10 gruppi bilanciati con eteroschedasticità elevata al variare della loro numerosità, osservazioni indipendenti.

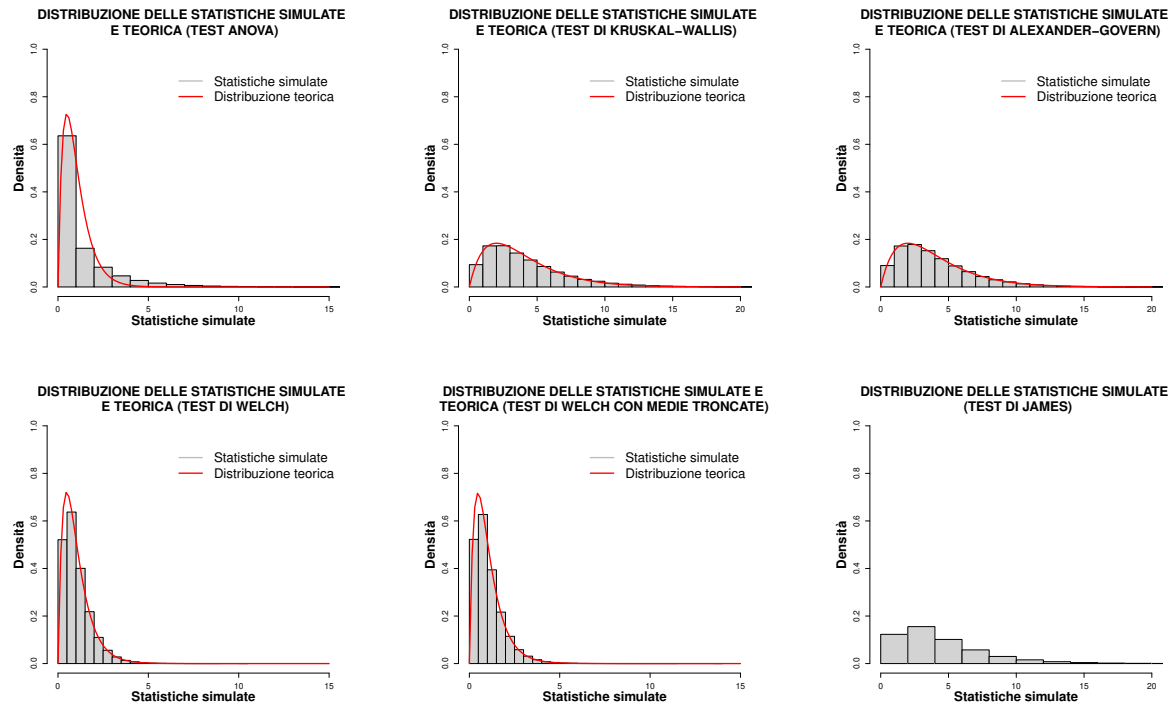


Figura 5.17: Confronto della distribuzione delle statistiche simulate e teorica sotto H_0 dei test, scenario B1 con $G = 5$, $n = 20$, $m = 0$, $s = 3$ e $r = 0$.

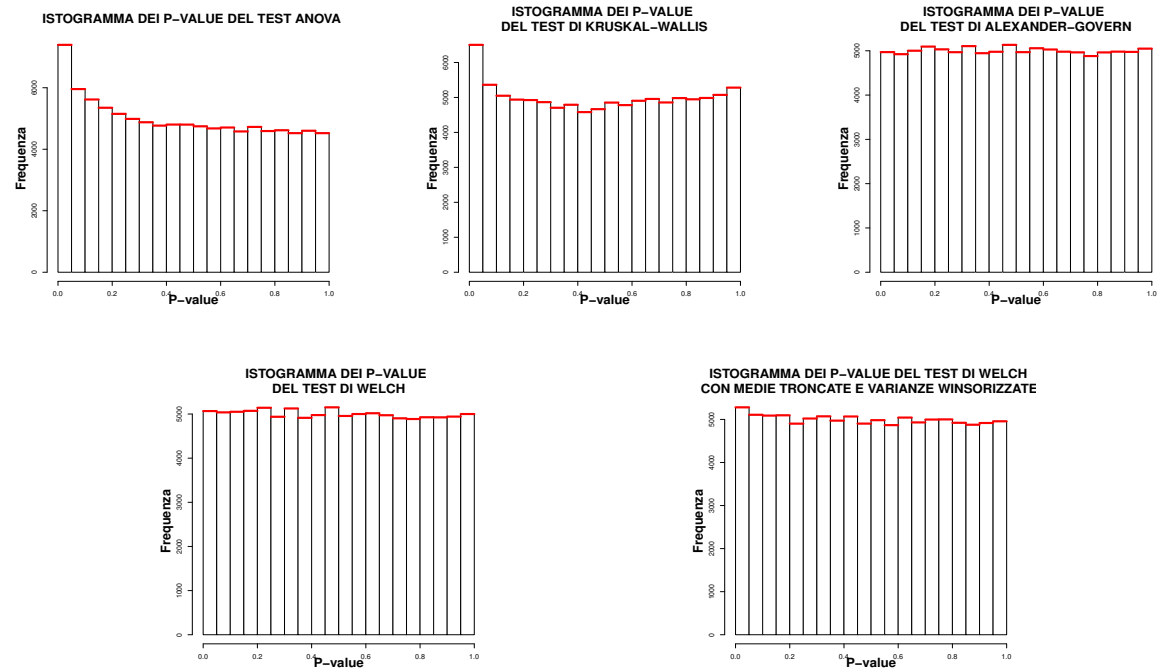
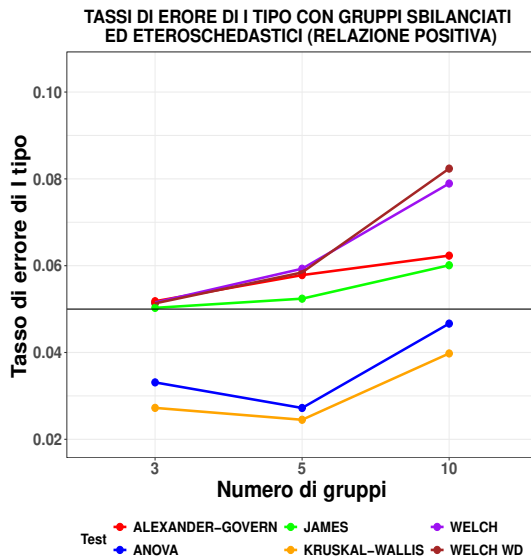


Figura 5.18: Confronto della distribuzione dei p-value sotto H_0 dei test, scenario B1 con $G = 5$, $n = 20$, $m = 0$, $s = 3$ e $r = 0$.

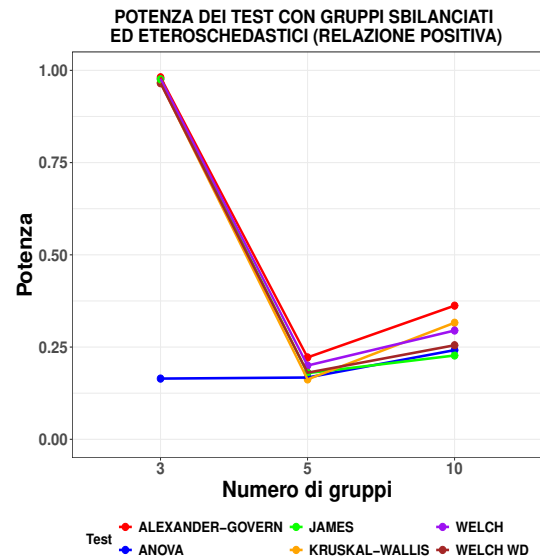
Ponendo l'attenzione sulle Figure 5.12 e 5.13, si osserva che il test ANOVA presenta un eccesso di p-value dal valore ridotto. Infatti, la coda destra della distribuzione delle statistiche F simulate appare più pesante rispetto a quella teorica, suggerendo una quantità maggiore di valori grandi delle statistiche simulate in confronto a quanto previsto sotto l'ipotesi nulla. I restanti test mostrano sia p-value distribuiti in maniera uniforme tra 0 e 1, sia un buon adattamento della distribuzione teorica con quella delle statistiche simulate.

In un contesto di eteroschedasticità elevata (Figure 5.17 e 5.18), la deviazione della distribuzione dei p-value dell'ANOVA è ancora più evidente, e si aggiunge anche quella emersa per il test di Kruskal-Wallis. Il test di Welch, di Welch con medie troncate e varianze winsorizzate e di Alexander-Govern mantengono una buona corrispondenza tra la distribuzione delle statistiche simulate e quella teorica oltre che i p-value uniformi.

Nel contesto in cui i gruppi comparati sono sbilanciati, moderatamente eteroschedastici ed è presente una relazione positiva tra numerosità e varianza (Figura 5.19a), il test ANOVA risulta troppo conservativo così come anche il test di Kruskal-Wallis. Tuttavia, i loro tassi di errore di I tipo, con un numero elevato di gruppi, si avvicinano al valore nominale. Viceversa, per i restanti test, che sono più liberali, all'aumentare del numero di gruppi, i tassi di errore di I tipo si allontanano progressivamente dal valore 0.05, in particolare quello di Welch e di Welch con medie troncate e varianze winsorizzate.



(a) Scenario B2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0$.



(b) Scenario B2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$, $s = 1.5$ e $r = 0$.

Figura 5.19: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati e moderatamente eteroschedastici (relazione positiva), osservazioni indipendenti.

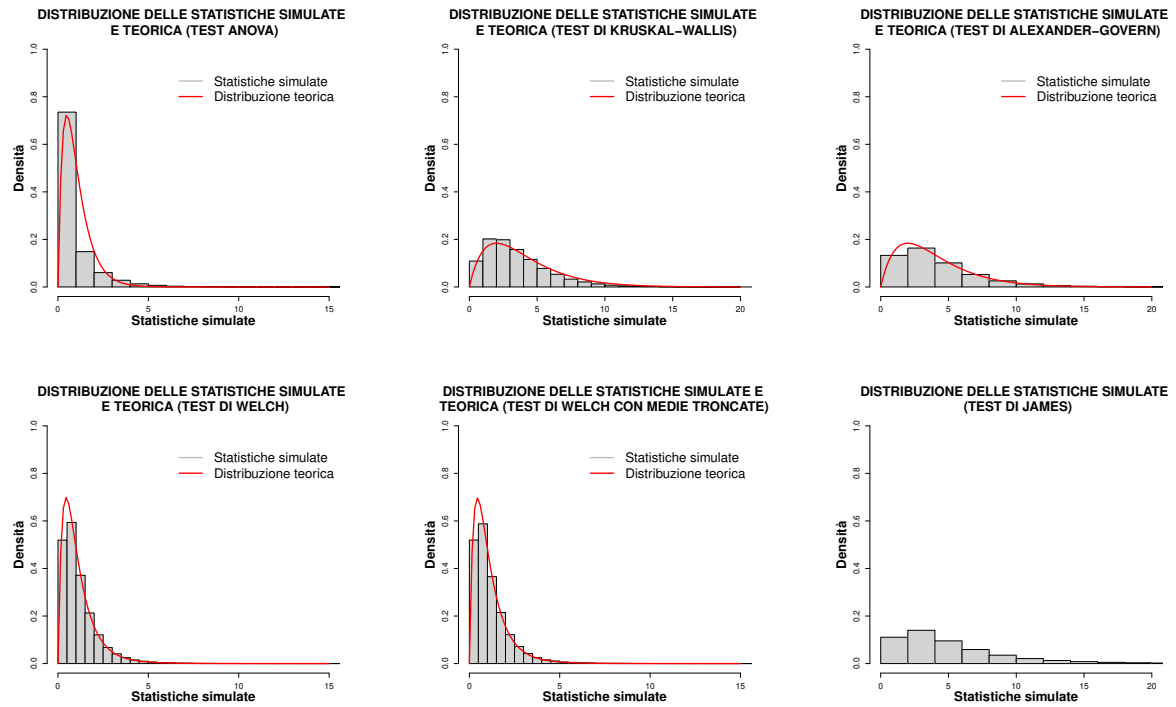


Figura 5.20: Confronto della distribuzione delle statistiche simulate e teorica sotto H_0 dei test, scenario B2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0$.

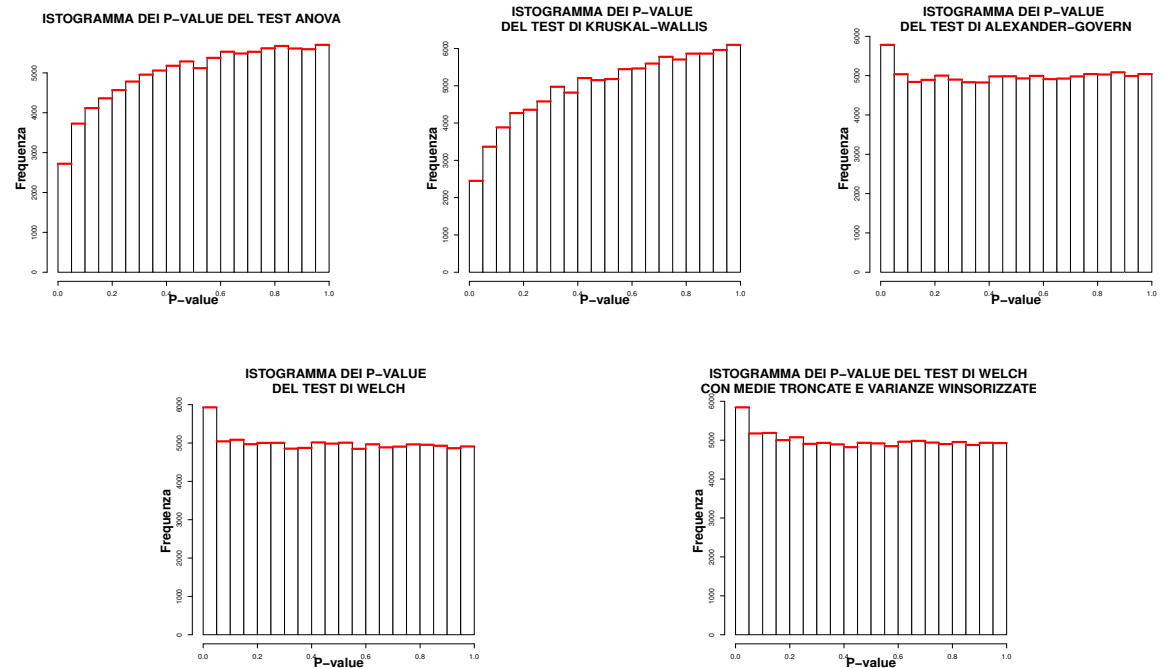
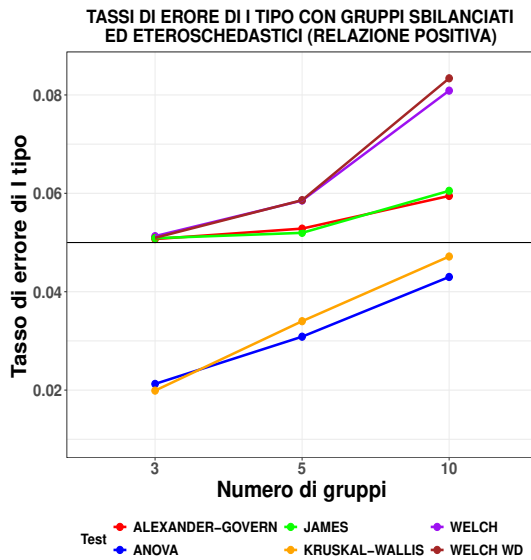


Figura 5.21: Confronto della distribuzione dei p-value sotto H_0 dei test, scenario B2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0$.

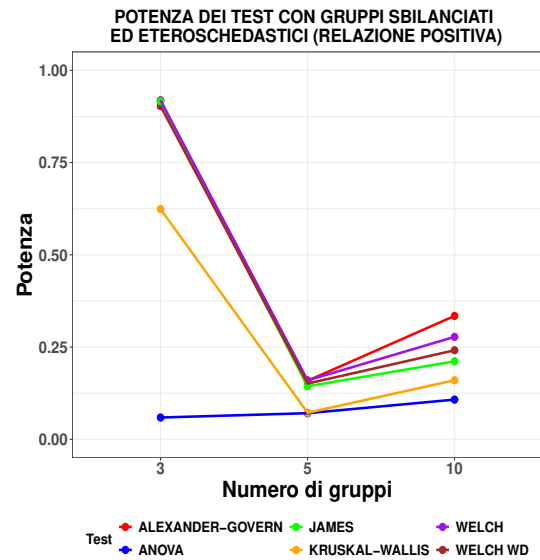
Se l'eteroschedasticità diventa elevata (Figura 5.22a), con un ridotto numero di gruppi considerati l'ANOVA diventa ancora più conservativa. Invece, con un numero elevato di gruppi, il test di Welch e di Welch con medie troncate e varianze winsorizzate ancora più liberali.

Per quanto riguarda la potenza, sia in caso di gruppi con eteroschedasticità moderata (Figura 5.19b) che elevata (Figura 5.22b), quella del test ANOVA assume un valore basso indipendentemente dal numero di gruppi. Diversamente, la potenza degli altri test è alta quando i gruppi comparati sono solo 3, ma si riduce drasticamente con l'aumentare del numero di gruppi.

La Figura 5.21 mostra una distribuzione dei p-value ottenuti tramite simulazione dai test ANOVA e Kruskal-Wallis in presenza di 5 gruppi sbilanciati, moderatamente eteroschedastici e con relazione positiva tra numerosità e varianza (analoghe considerazioni emergono anche in caso di eteroschedasticità elevata 5.24) che si discosta da quella uniforme sotto l'ipotesi nulla. Questo perché, poiché i test sono conservativi, sono meno inclini a rifiutare l'ipotesi nulla. Ciò si riflette in p-value generalmente più alti rispetto a quanto ci si attende, giustificando le prime barre basse dei loro istogramma dei p-value. Al contrario, il test di Alexander-Govern, di Welch e di Welch con medie troncate e varianze winsorizzate mostrano una distribuzione dei p-value che si avvicina a quella uniforme, eccetto per una quantità di p-value ridotti superiore al previsto.



(a) Scenario B2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0$.



(b) Scenario B2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$, $s = 3$ e $r = 0$.

Figura 5.22: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati con eteroschedasticità elevata (relazione positiva), osservazioni indipendenti.

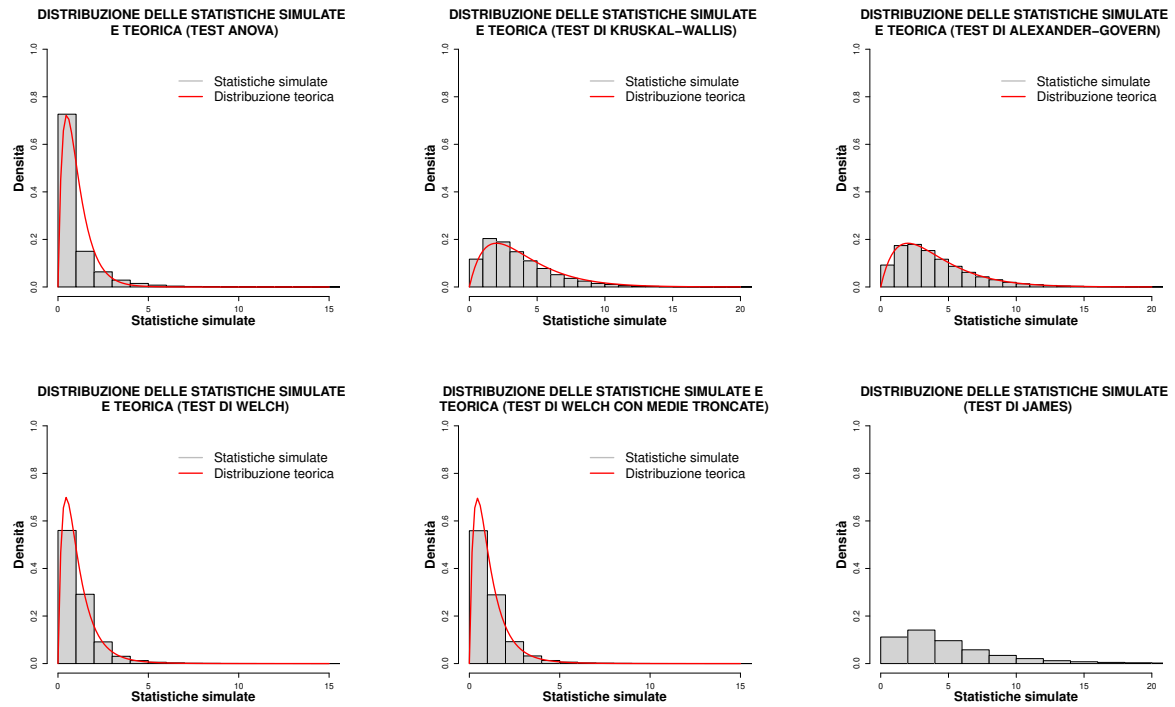


Figura 5.23: Confronto della distribuzione delle statistiche simulate e teorica sotto H_0 dei test, scenario B2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0$.

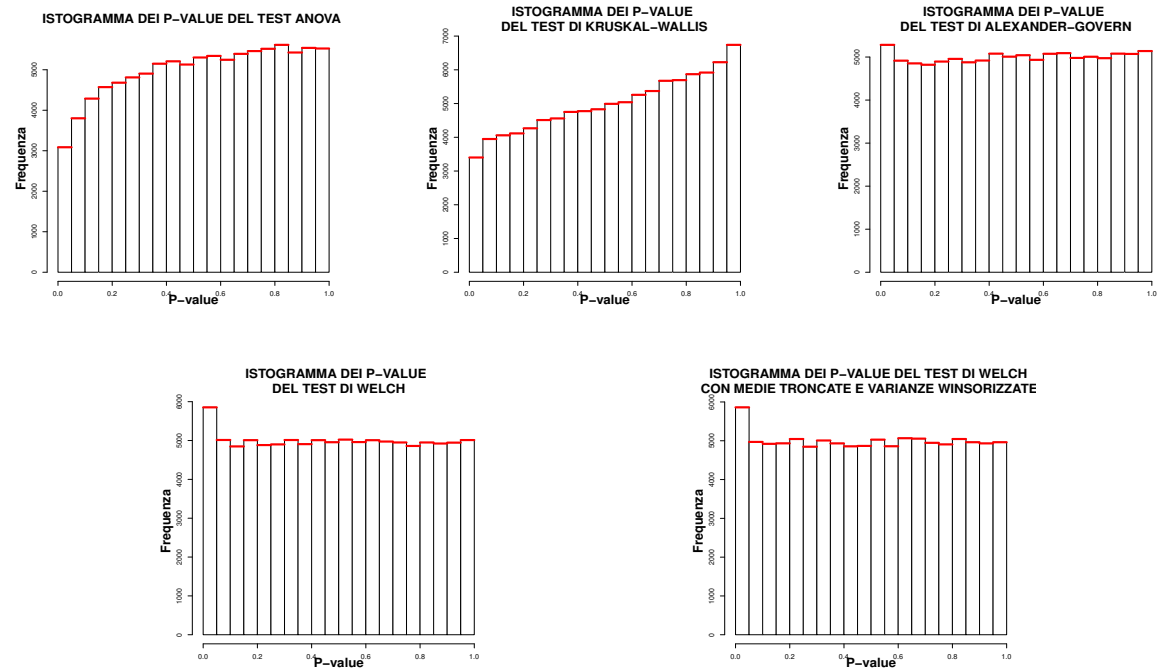


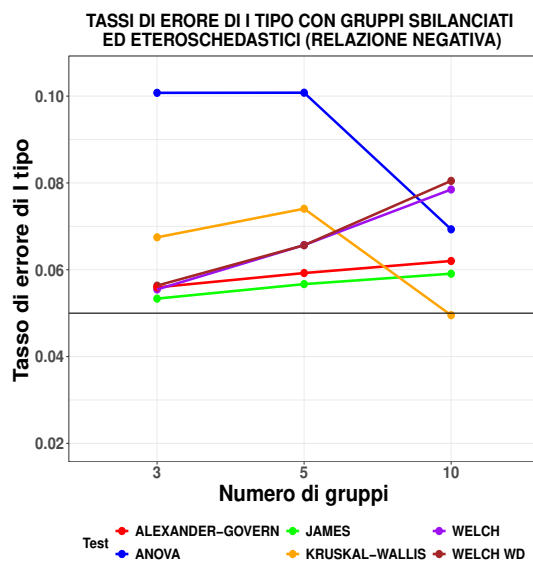
Figura 5.24: Confronto della distribuzione dei p-value sotto H_0 dei test, scenario B2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0$.

Nelle circostanze in cui i gruppi confrontati sono sbilanciati, moderatamente eteroschedastici ed è presente una relazione negativa tra numerosità e varianza (Figura 5.25a), il test di Alexander-Govern, di Welch e di Welch con medie troncate e varianze winsorizzate si comportano come nel caso di relazione positiva tra numerosità e varianza, risultando troppo liberali.

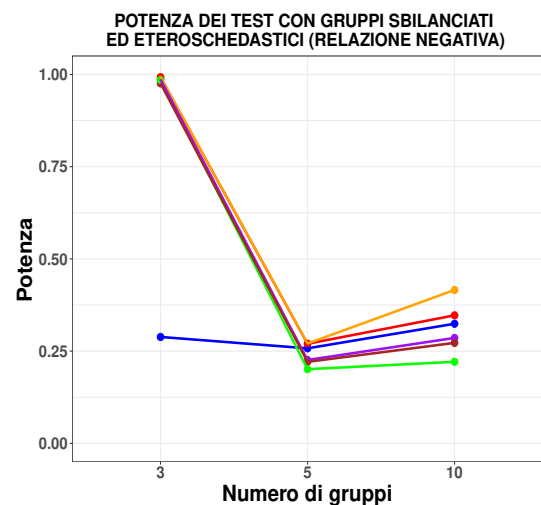
Anche il test ANOVA e quello di Kruskal-Wallis, in questo contesto, risultano troppo liberali. Il loro tasso di errore di I tipo però, in caso di 10 gruppi, si avvicina al valore nominale.

Gli effetti rilevati a causa dell'eteroschedasticità moderata si riscontrano anche in presenza di eteroschedasticità elevata (Figura 5.28a): in particolare, per il test di Kruskal-Wallis, e ancor più per il test ANOVA, il tasso di errore di I tipo assume valori più elevati rispetto al valore di 0.05, soprattutto per campioni di piccola numerosità.

Per quanto riguarda la potenza, le conclusioni osservate in caso di relazione negativa tra numerosità e varianza, sia per eteroschedasticità moderata (Figura 5.25b) che elevata (Figura 5.28b), sono le medesime di quanto emerso in presenza di relazione positiva.



(a) Scenario B3 con $G \in \{3, 5, 10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0$.



(b) Scenario B3 con $G \in \{3, 5, 10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$, $s = 1.5$ e $r = 0$.

Figura 5.25: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati e moderatamente eteroschedastici (relazione negativa), osservazioni indipendenti.

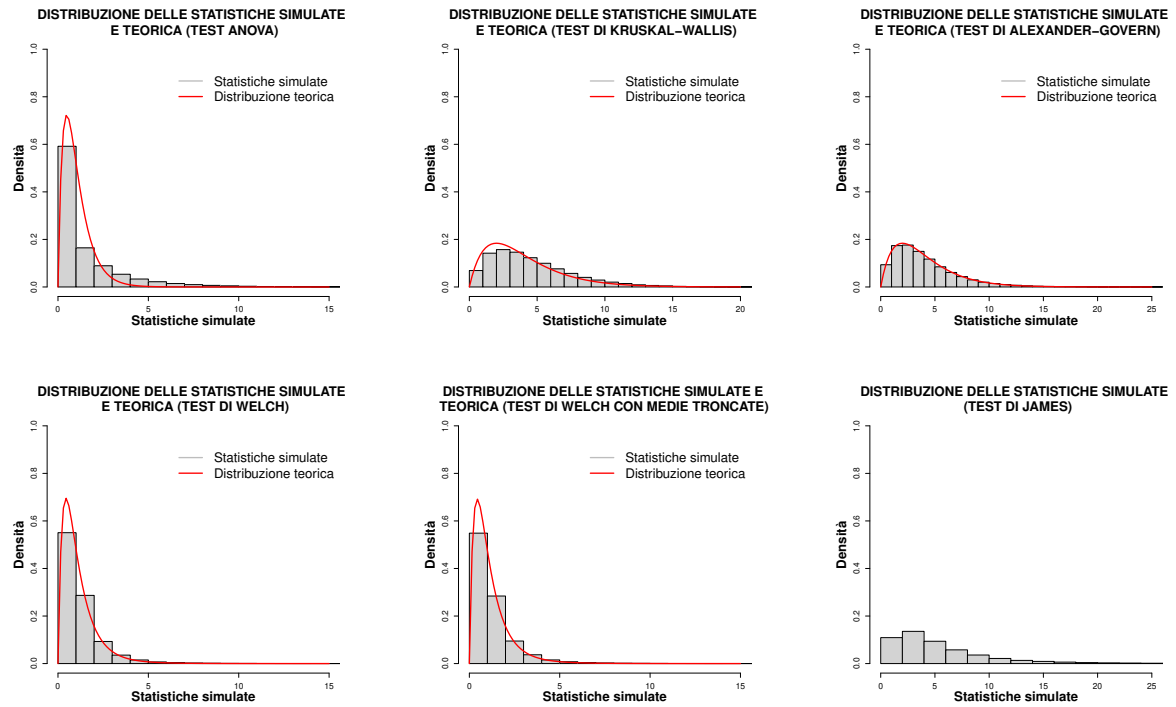


Figura 5.26: Confronto della distribuzione delle statistiche simulate e teorica sotto H_0 dei test, scenario B3 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0$.

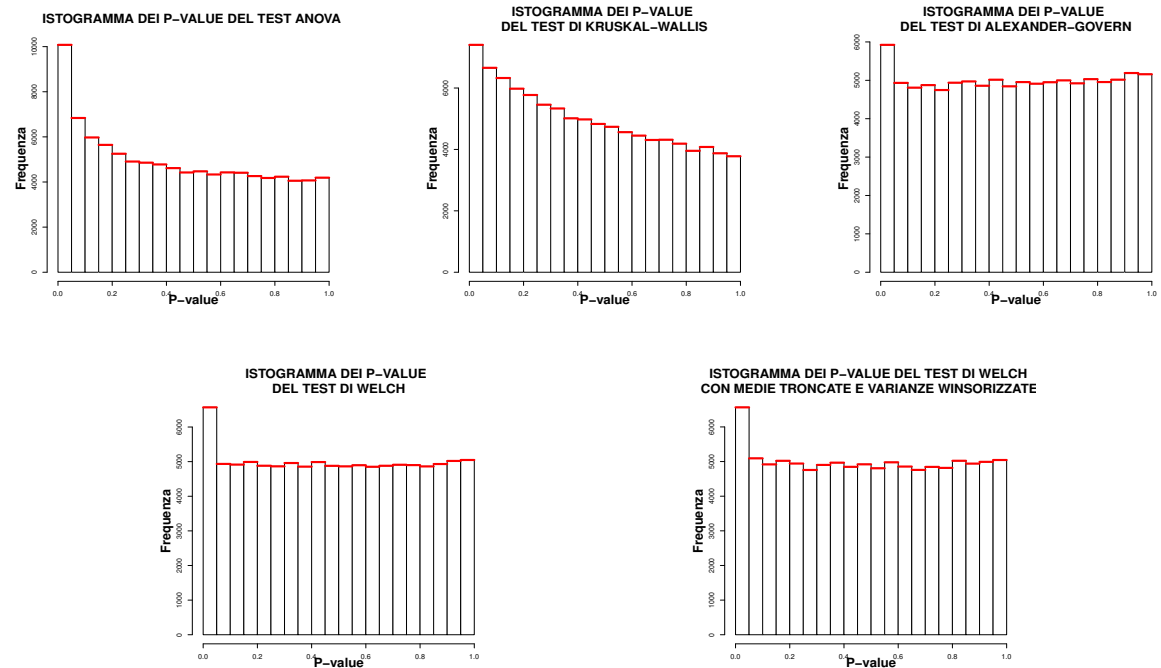
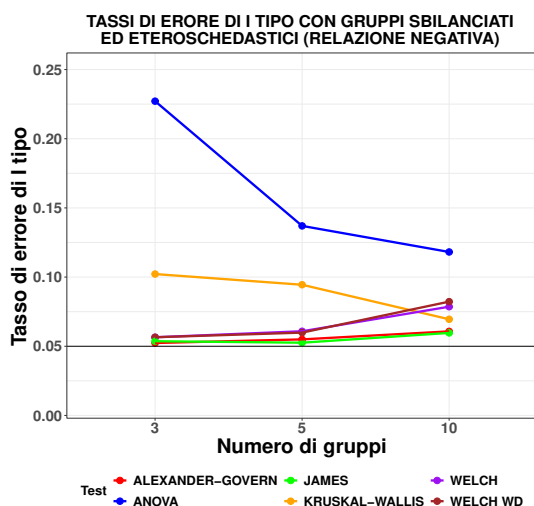
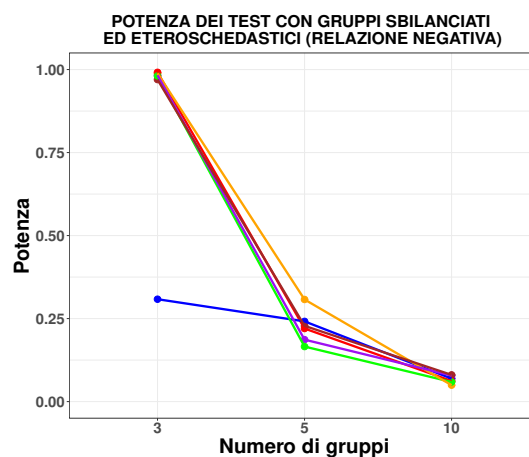


Figura 5.27: Confronto della distribuzione dei p-value sotto H_0 dei test, scenario B3 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0$.



(a) Scenario B3 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0$.



(b) Scenario B3 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$, $s = 3$ e $r = 0$.

Figura 5.28: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati con eteroschedasticità elevata (relazione negativa), osservazioni indipendenti.

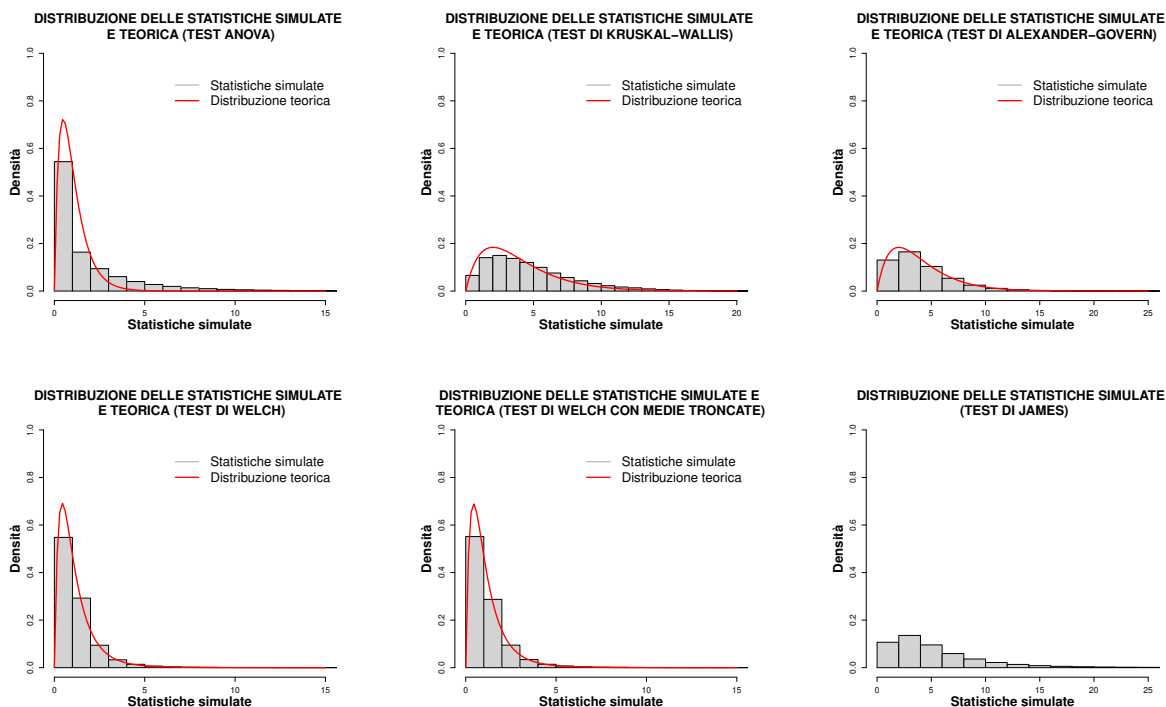


Figura 5.29: Confronto della distribuzione delle statistiche simulate e teorica sotto H_0 dei test, scenario B3 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0$.

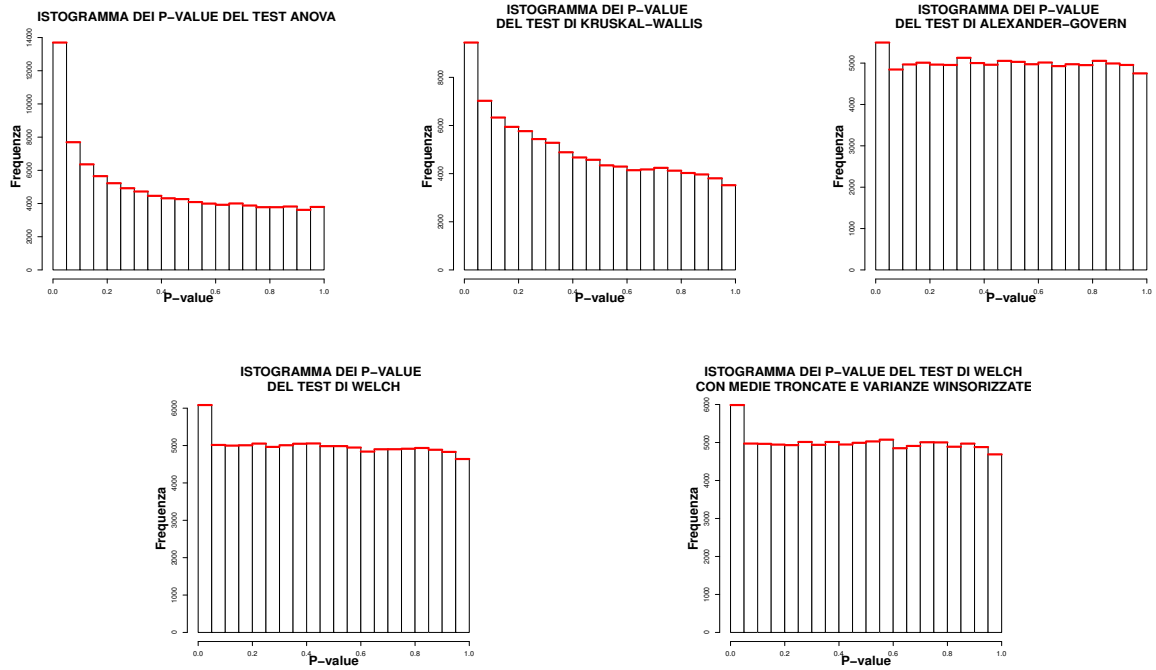


Figura 5.30: Confronto della distribuzione dei p-value sotto H_0 dei test, scenario B3 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0$.

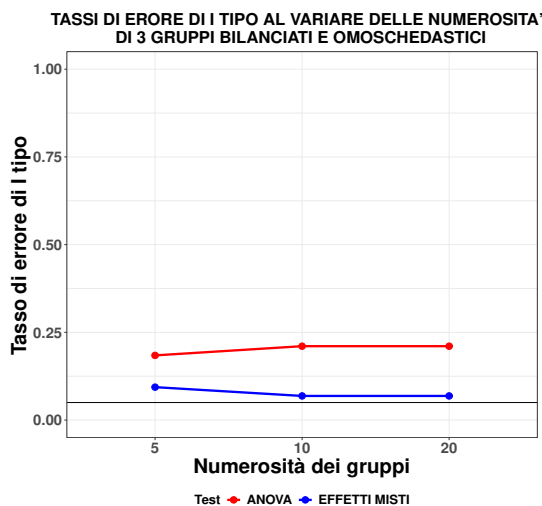
In un contesto di relazione negativa tra numerosità e varianza, tutti i test risultano troppo liberali, sia in caso di eteroschedasticità moderata (Figure 5.26 e 5.27) che elevata (Figure 5.29 e 5.30), infatti gli istogrammi dei p-value riportano le prime barre superiori rispetto alle altre. Questo effetto è particolarmente evidente nei test ANOVA e di Kruskal-Wallis. Questi ultimi hanno una distribuzione teorica che sottostima la probabilità di ottenere valori estremi nelle statistiche test, presentando una coda empirica più pesante rispetto a quella teorica.

5.3 Gruppi bilanciati/sbilanciati omoschedastici con osservazioni dipendenti

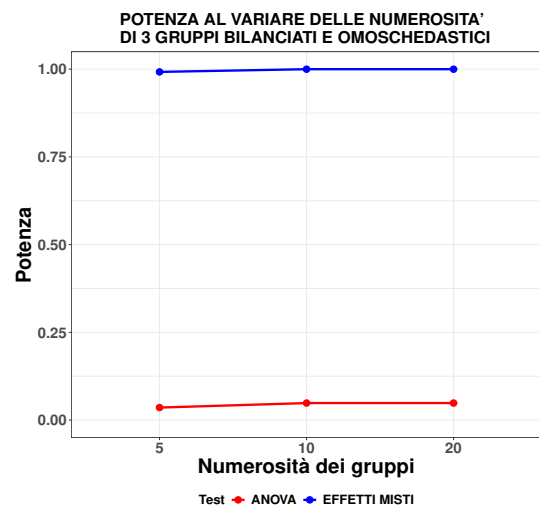
Per adottare l'approccio con modelli ad effetti misti in questo studio di simulazione, in ogni scenario sono stati costruiti due tipi di modelli: il modello completo che include l'effetto fisso del tempo (o del trattamento) e il modello ridotto che non considera l'effetto fisso del tempo ma mantiene l'effetto casuale per i soggetti.

Il confronto tra il modello completo e quello ridotto è stato eseguito utilizzando un test del rapporto di verosimiglianza basato sul metodo della massima verosimiglianza (ML, Maximum Likelihood), anziché sul metodo della massima verosimiglianza ristretta (REML, REstricted Maximum Likelihood). Questo test, condotto tramite la funzione *anova* con l'opzione *method* = "ML", valuta quale dei due modelli si adatta meglio ai dati osservati, confrontando le loro log-verosimiglianze. La statistica del test di verosimiglianza (LRT, Likelihood Ratio Test) segue una distribuzione chi-quadrato con un numero di gradi di libertà pari alla differenza nel numero di parametri stimati tra i due modelli.

Un p-value inferiore al valore nominale prefissato, in questo studio a 0.05, indica che il modello completo si adatta meglio ai dati osservati rispetto al modello ridotto. Di conseguenza, si conclude che i gruppi differiscono in media in modo significativo nel tempo.

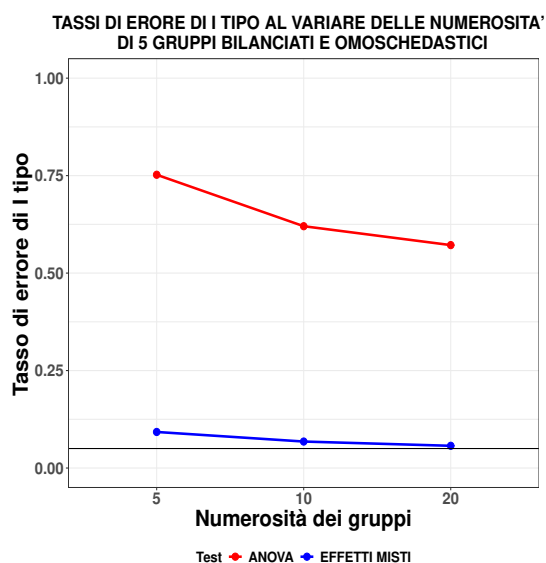


(a) Scenario A1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 0$ e $r = 0.5$.

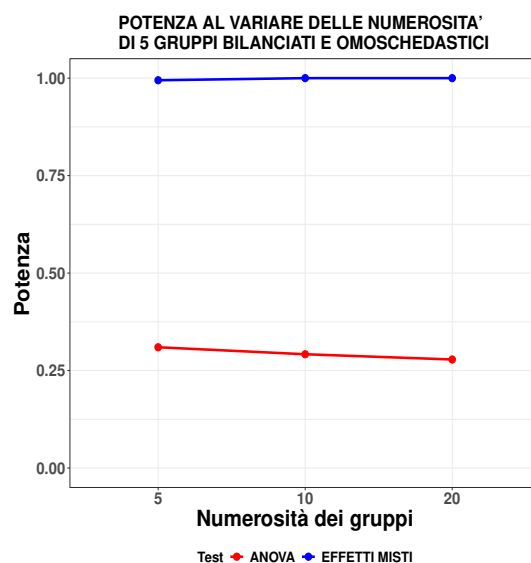


(b) Scenario A1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 2$ e $r = 0.5$.

Figura 5.31: Tasso di errore di I tipo e potenza in presenza di 3 gruppi bilanciati e omoschedastici al variare della loro numerosità, osservazioni dipendenti.

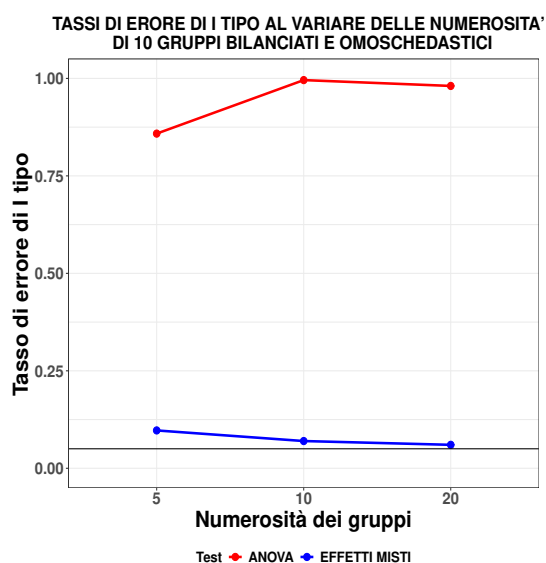


(a) Scenario A1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 0$ e $r = 0.5$.

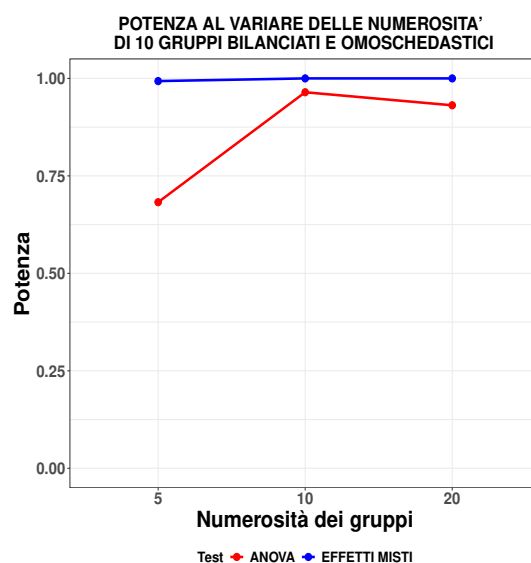


(b) Scenario A1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 2$ e $r = 0.5$.

Figura 5.32: Tasso di errore di I tipo e potenza in presenza di 5 gruppi bilanciati e omoschedastici al variare della loro numerosità, osservazioni dipendenti.



(a) Scenario A1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 0$ e $r = 0.5$.



(b) Scenario A1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 2$ e $r = 0.5$.

Figura 5.33: Tasso di errore di I tipo e potenza in presenza di 10 gruppi bilanciati e omoschedastici al variare della loro numerosità, osservazioni dipendenti.

In un contesto in cui sono state rilevate più misurazioni sugli stessi soggetti e i gruppi di osservazioni sono bilanciati ed omoschedastici, le Figure 5.31a, 5.32a e 5.33a mostrano che il test ANOVA presenta un tasso di errore di I tipo superiore al valore nominale 0.05. Il suo tasso si allontana ulteriormente dal valore nominale all'aumentare del numero di gruppi confrontati. La dipendenza dei dati rende l'ANOVA un test troppo liberale, invece, l'approccio tramite modelli ad effetti misti risulta più adeguato in queste situazioni, indipendentemente dal numero di gruppi confrontati. Il metodo con modelli ad effetti misti mostra un tasso di errore di I tipo prossimo al valore nominale, con solo un leggero aumento quando i gruppi hanno dimensioni ridotte.

L'approccio con modelli ad effetti misti dimostra anche un'elevata potenza statistica, indipendentemente dal numero e dalla dimensione dei gruppi, confermandosi superiore rispetto all'ANOVA, la cui potenza diminuisce drasticamente con la riduzione del numero di gruppi comparati (Figure 5.31b, 5.32b e 5.33b).

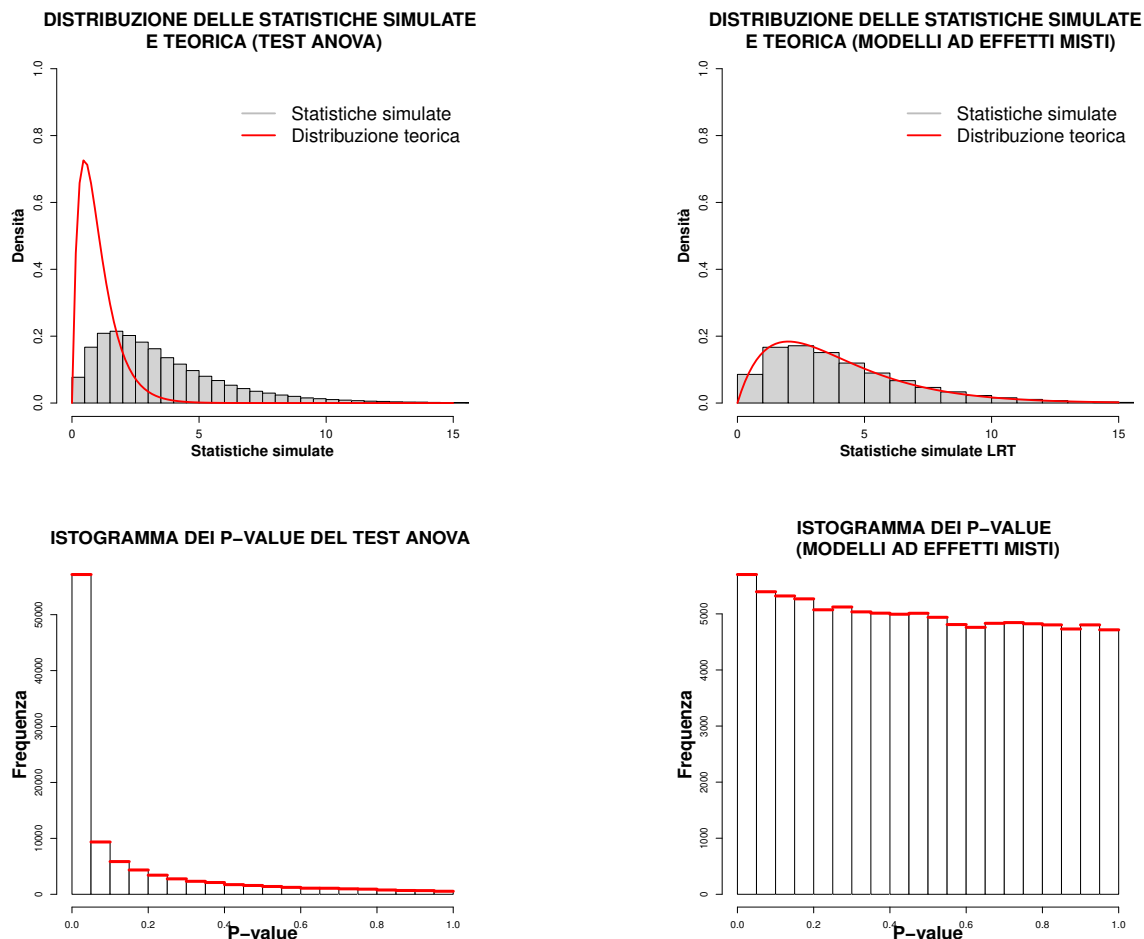
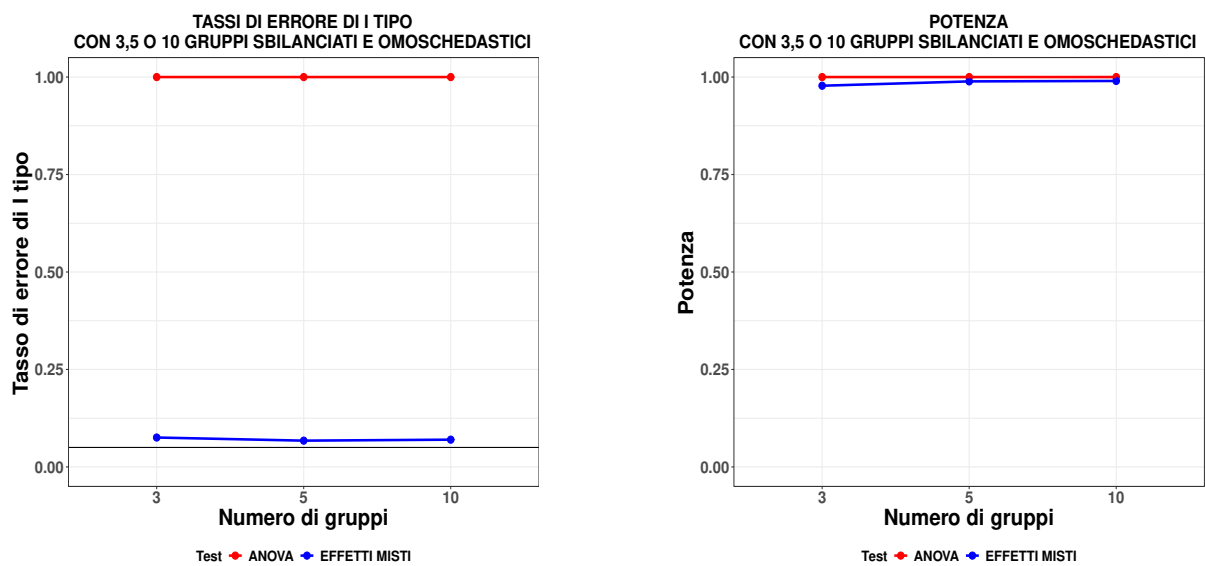


Figura 5.34: Confronto della distribuzione delle statistiche simulate e teorica e dei p-value sotto H_0 , scenario A1 con $G = 5$, $n = 20$, $m = 0$ e $r = 0.5$.

La Figura 5.34 evidenzia come la distribuzione delle statistiche simulate dell'ANOVA non rispecchi quella teorica, mostrando una coda destra empirica più pesante di quella teorica quando i gruppi bilanciati confrontati sono 5 e di dimensione 20. Ciò si riflette in una prima barra dei p-value significativamente più alta rispetto alle altre. Al contrario, i p-value ricavati dal test di verosimiglianza risultano più uniformi rispetto a quelli dell'ANOVA, e le statistiche LRT simulate seguono la distribuzione teorica prevista sotto l'ipotesi nulla.



(a) Scenario A2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$ e $r = 0.5$.

(b) Scenario A2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$ e $r = 0.5$.

Figura 5.35: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati e omoschedastici, osservazioni dipendenti.

Nel caso in cui siano presenti misurazioni mancanti per alcuni soggetti e i gruppi risultino quindi sbilanciati, pur rimanendo omoschedastici, entrambi i metodi mostrano potenze elevate (Figura 5.35b). Tuttavia, l'ANOVA continua ad avere un tasso di errore di I tipo superiore al livello nominale raggiungendo addirittura il valore massimo, mentre l'approccio con modelli ad effetti misti mantiene un valore vicino allo 0.05 (Figura 5.35a).

La non idoneità dell'ANOVA in questo contesto è confermata dai p-value non uniformi sotto l'ipotesi nulla e dalla distribuzione delle sue statistiche test simulate (Figura 5.36), chiaramente distante da quella teorica in presenza di 5 gruppi sbilanciati comparati. Le statistiche LRT simulate seguono la distribuzione teorica, ma i p-value sono meno uniformi rispetto al contesto in cui i gruppi sono bilanciati.

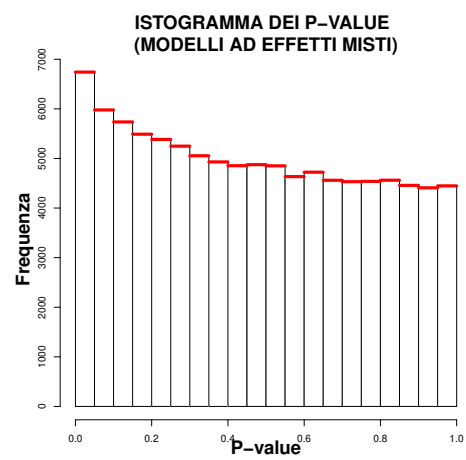
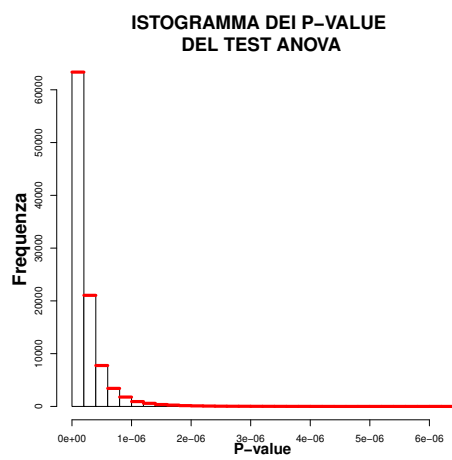
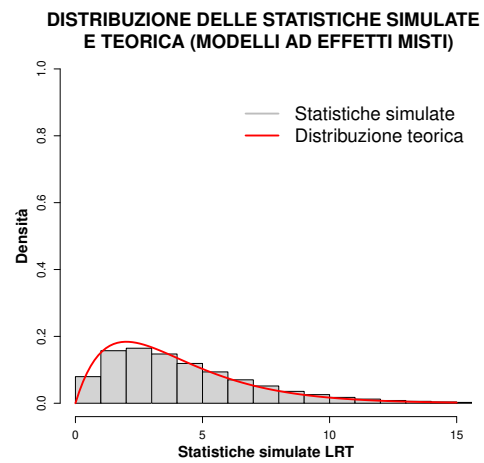
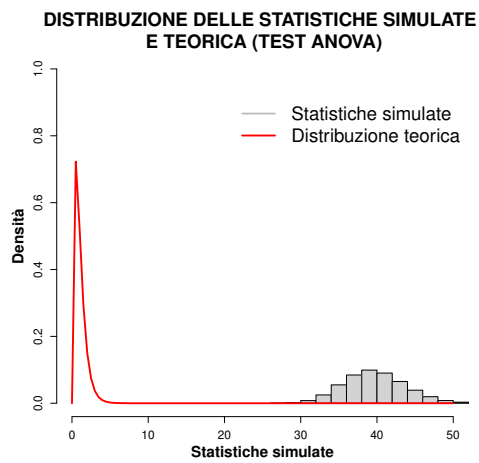
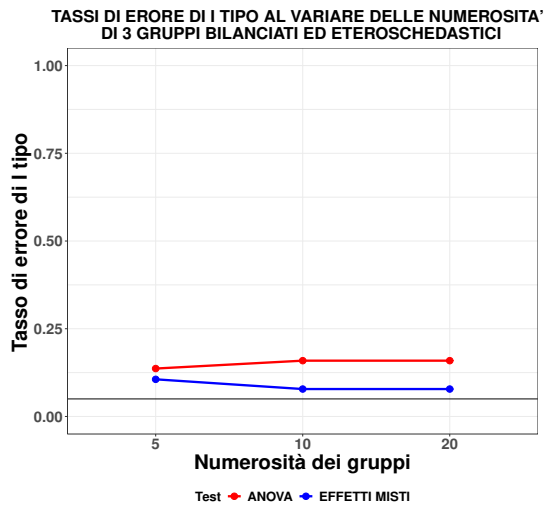
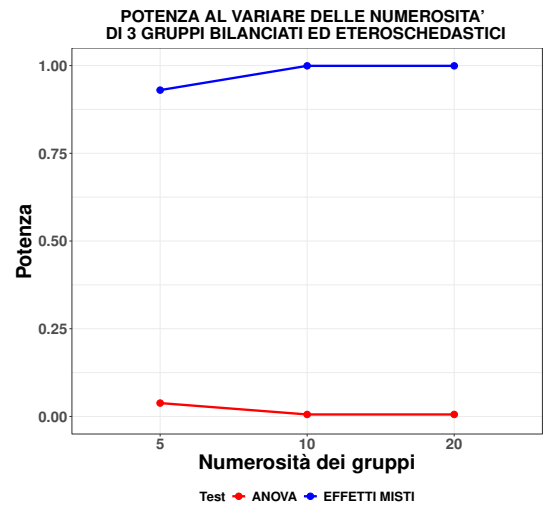


Figura 5.36: Confronto della distribuzione delle statistiche simulate e teorica e dei p-value sotto H_0 , scenario A2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$ e $r = 0.5$.

5.4 Gruppi bilanciati/sbilanciati eteroschedastici con osservazioni dipendenti

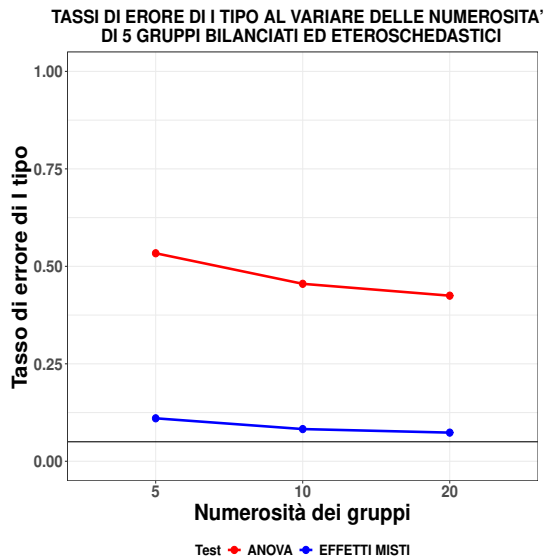


(a) Scenario B1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 1.5$ e $r = 0.5$.

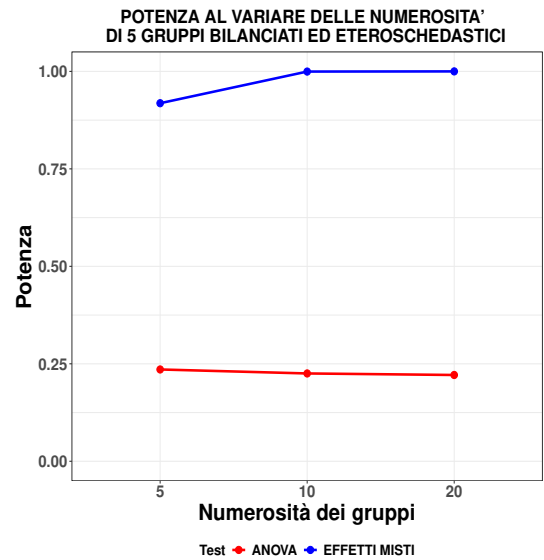


(b) Scenario B1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 1.5$ e $r = 0.5$.

Figura 5.37: Tasso di errore di I tipo e potenza in presenza di 3 gruppi bilanciati e moderatamente eteroschedastici al variare della loro numerosità, osservazioni dipendenti.

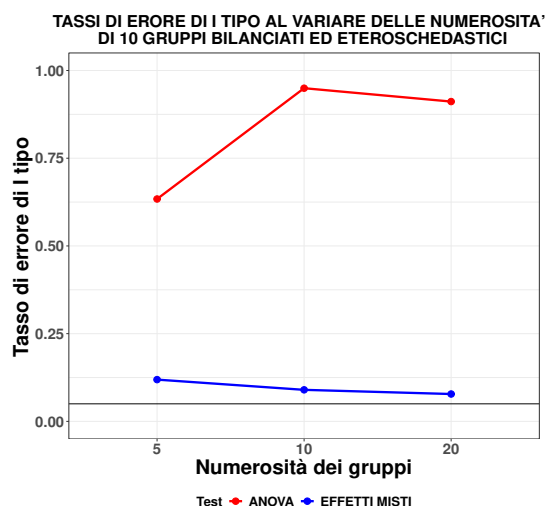


(a) Scenario B1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 1.5$ e $r = 0.5$.

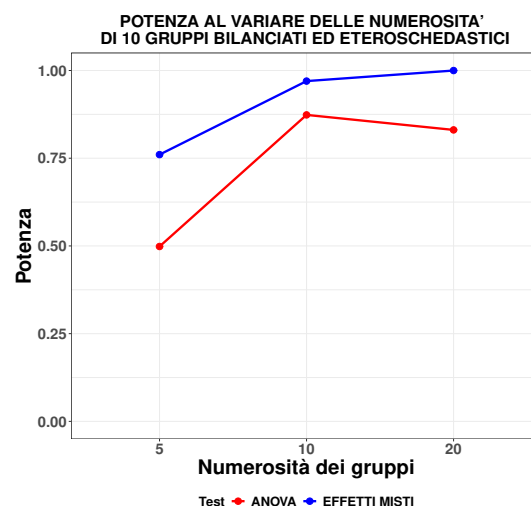


(b) Scenario B1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 1.5$ e $r = 0.5$.

Figura 5.38: Tasso di errore di I tipo e potenza in presenza di 5 gruppi bilanciati e moderatamente eteroschedastici al variare della loro numerosità, osservazioni dipendenti.



(a) Scenario B1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 1.5$ e $r = 0.5$.



(b) Scenario B1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 1.5$ e $r = 0.5$.

Figura 5.39: Tasso di errore di I tipo e potenza in presenza di 10 gruppi bilanciati e moderatamente eteroschedastici al variare della loro numerosità, osservazioni dipendenti.

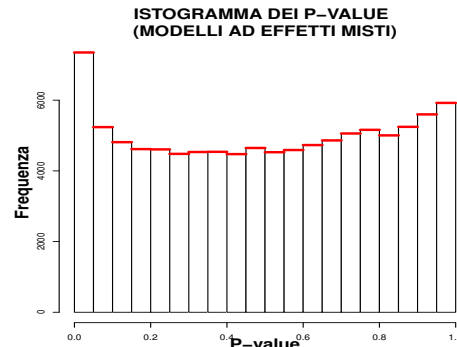
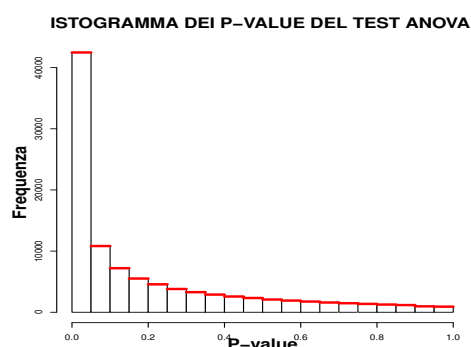
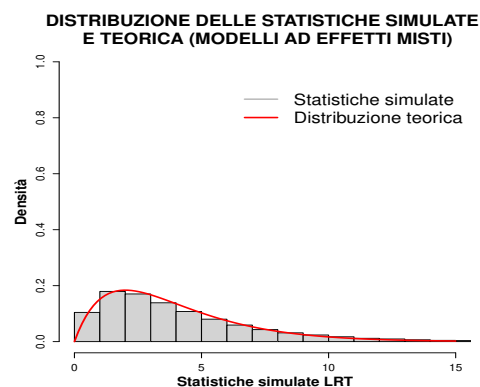
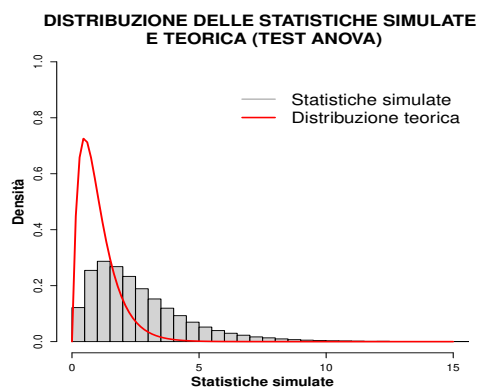
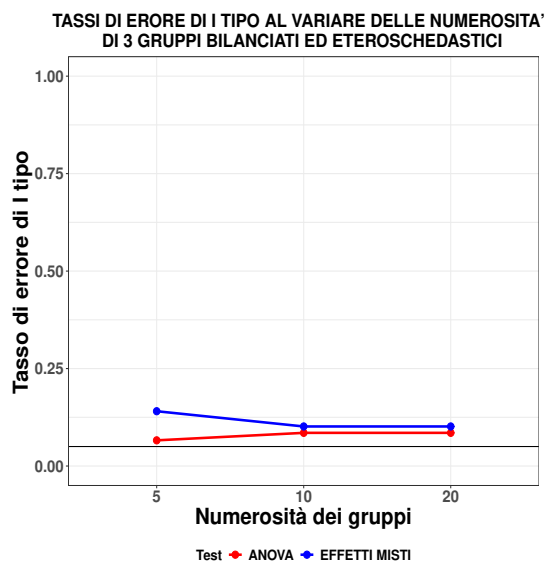
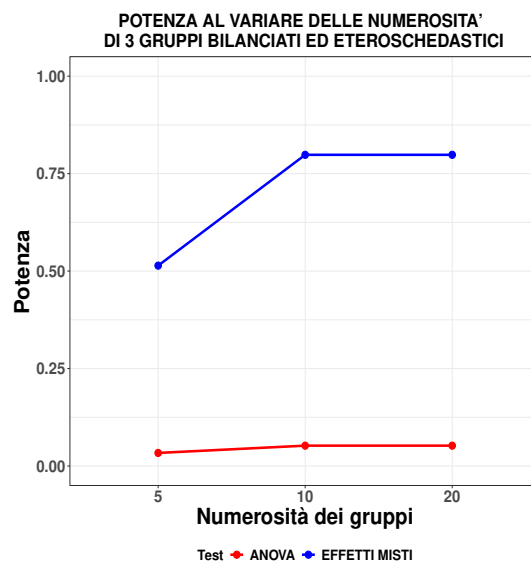


Figura 5.40: Confronto della distribuzione delle statistiche simulate e teorica e dei p-value sotto H_0 , scenario B1 con $G = 5$, $n = 20$, $m = 0$, $s = 1.5$ e $r = 0.5$.

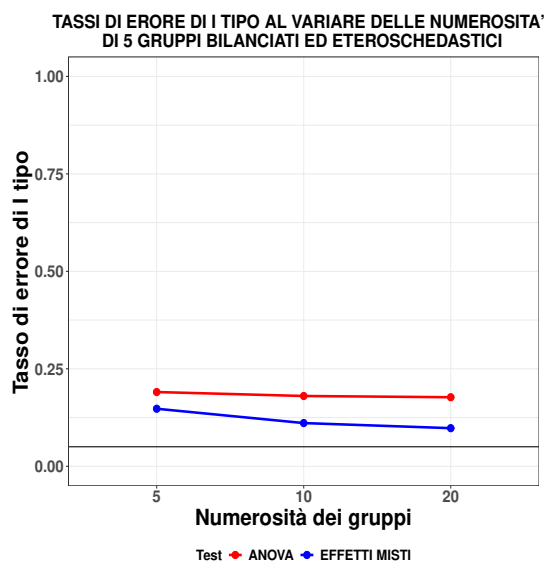


(a) Scenario B1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 3$ e $r = 0.5$.

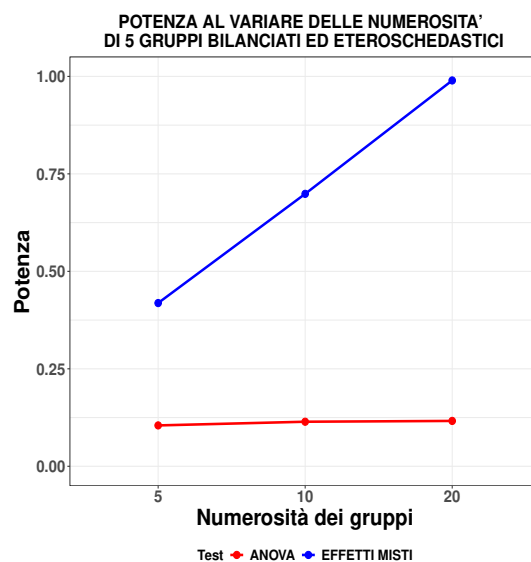


(b) Scenario B1 con $G = 3$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 3$ e $r = 0.5$.

Figura 5.41: Tasso di errore di I tipo e potenza in presenza di 3 gruppi bilanciati con eteroschedasticità elevata al variare della loro numerosità, osservazioni dipendenti.

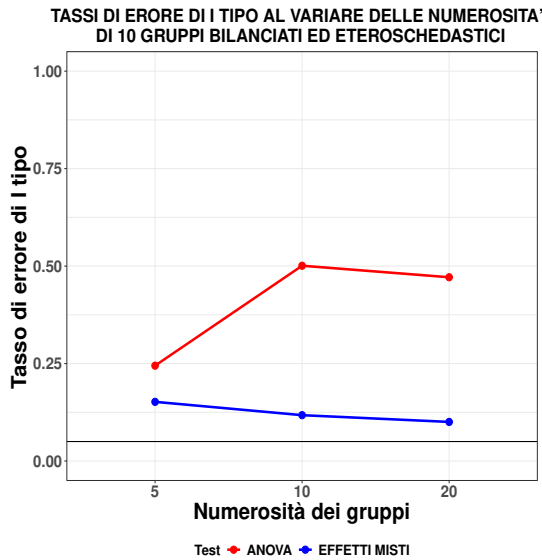


(a) Scenario B1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 3$ e $r = 0.5$.

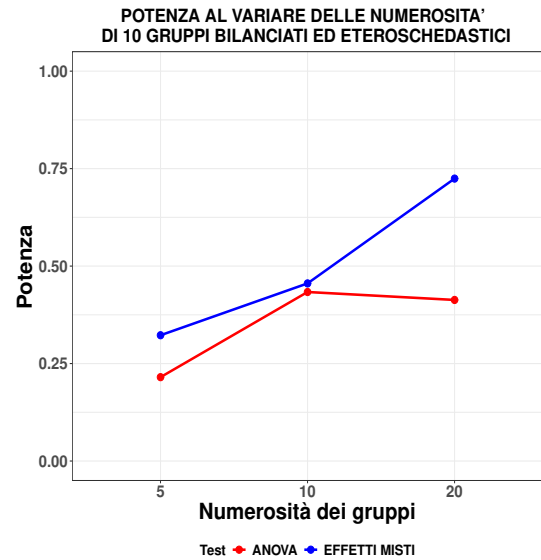


(b) Scenario B1 con $G = 5$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 3$ e $r = 0.5$.

Figura 5.42: Tasso di errore di I tipo e potenza in presenza di 5 gruppi bilanciati con eteroschedasticità elevata al variare della loro numerosità, osservazioni dipendenti.



(a) Scenario B1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 0$, $s = 3$ e $r = 0.5$.



(b) Scenario B1 con $G = 10$, $n \in \{5, 10, 20\}$, $m = 2$, $s = 3$ e $r = 0.5$.

Figura 5.43: Tasso di errore di I tipo e potenza in presenza di 10 gruppi bilanciati con eteroschedasticità elevata al variare della loro numerosità, osservazioni dipendenti.

Nel contesto in cui sono state rilevate più misurazioni sugli stessi soggetti e i gruppi di dati sono bilanciati ma eteroschedastici, l'ANOVA mostra un tasso di errore di tipo I superiore al valore nominale, con un incremento progressivo all'aumentare del numero di gruppi. Invece, l'approccio basato sui modelli ad effetti misti presenta un tasso di errore di tipo I solo leggermente superiore al livello nominale 0.05, avvicinandosi a questo valore man mano che il numero di osservazioni per gruppo aumenta (Figure 5.37a, 5.38a e 5.39a).

Confrontando i dati eteroschedastici con il caso in cui i gruppi sono bilanciati ma omoschedastici, si osserva che l'approccio con modelli ad effetti misti risente dell'eteroschedasticità, poiché il tasso di errore di tipo I aumenta all'aumentare dell'intensità dell'eteroschedasticità (Figure 5.41a, 5.42a e 5.43a). In particolare, questo metodo diventa progressivamente più liberale con l'incremento dell'eteroschedasticità.

Per quanto riguarda la potenza statistica, l'approccio con modelli ad effetti misti dimostra un'elevata potenza in presenza di eteroschedasticità moderata indipendentemente dal numero di gruppi, pur essendo influenzato dalla loro numerosità: la potenza diminuisce per gruppi di dimensioni ridotte (Figure 5.37b, 5.38b e 5.39b). Tuttavia, con l'aumento dell'intensità dell'eteroschedasticità, la potenza relativa al metodo con modelli ad effetti misti si riduce ulteriormente, risentendo sia del numero di gruppi che della loro numerosità. Nello specifico, la potenza diminuisce al crescere del numero di gruppi e al diminuire del numero di osservazioni per gruppo (Figure 5.41b, 5.42b e 5.43b).

L'ANOVA, invece, mostra una potenza che dipende significativamente dal numero di gruppi, aumentando con l'incremento del numero di gruppi. Anche l'ANOVA risente dell'eteroschedasticità, con una potenza inferiore in contesti di elevata eteroschedasticità rispetto a quelli con eteroschedasticità moderata.

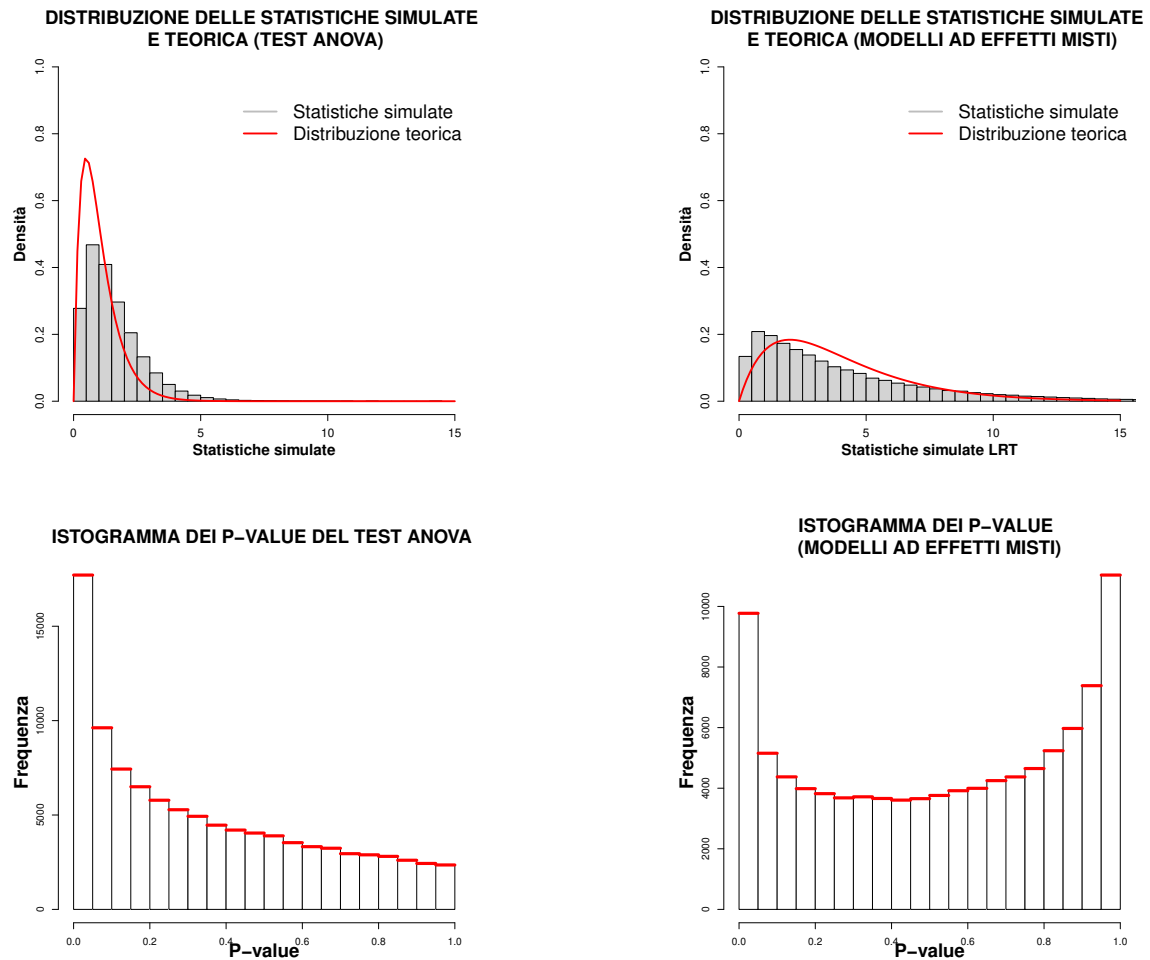
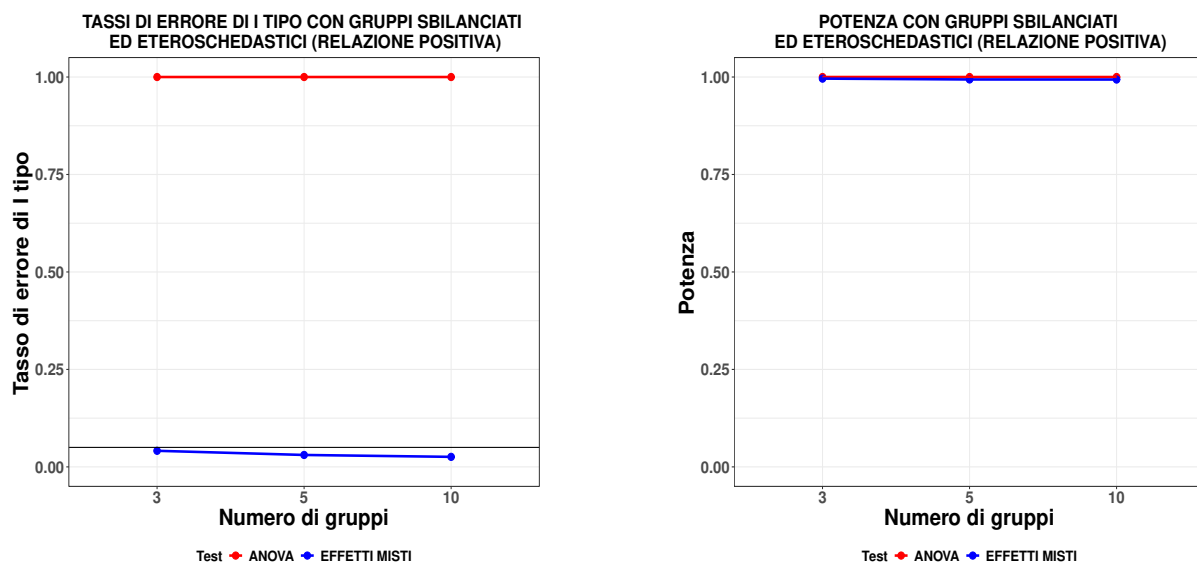


Figura 5.44: Confronto della distribuzione delle statistiche simulate e teorica e dei p-value sotto H_0 , scenario B1 con $G = 5$, $n = 20$, $m = 0$, $s = 3$ e $r = 0.5$.

L'effetto negativo dell'eteroschedasticità sui modelli ad effetti misti si evidenzia anche dalla minore uniformità dei p-value in presenza di cinque gruppi eteroschedastici con numerosità pari a 20 sotto l'ipotesi nulla (Figura 5.40), rispetto allo scenario con gruppi omoschedastici. Tale effetto è ancora più marcato in condizioni di elevata eteroschedasticità (Figura 5.44), dove emergono significativamente alte sia la barra dei p-value ridotti, sia la barra dei p-value estremamente grandi. La distribuzione delle statistiche LRT simulate presenta una coda iniziale meno pesante di quella teorica, risultando poi leggermente superiore nella parte finale.

Tuttavia, questo approccio rimane comunque migliore dell'ANOVA sia in condizioni di eteroschedasticità moderata che elevata, poiché la distribuzione teorica dell'ANOVA sottostima la probabilità di ottenere valori estremi nelle statistiche del test, presentando una coda delle statistiche F simulate più pesante rispetto a quella teorica. Questo spiega la presenza della barra dei p-value così elevata rispetto alle altre: l'ANOVA, essendo più liberale, è maggiormente incline a rifiutare l'ipotesi nulla.



(a) Scenario B2 con $G \in \{3, 5, 10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0.5$.

(b) Scenario B2 con $G \in \{3, 5, 10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$, $s = 1.5$ e $r = 0.5$.

Figura 5.45: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati e moderatamente eteroschedastici (relazione positiva), osservazioni dipendenti.

Quando i gruppi non solo sono moderatamente eteroschedastici, ma anche sbilanciati, con una relazione positiva tra numerosità e varianza (Figura 5.45a), il test ANOVA raggiunge il valore massimo del tasso di errore di tipo I. In pratica, questo significa che il test rifiuta sempre l'ipotesi nulla, indipendentemente dai dati osservati. Questo comportamento è altamente indesiderabile per un test statistico, poiché ne compromette l'utilità pratica.

Al contrario, l'approccio con modelli ad effetti misti diventa troppo conservativo in questo contesto: il suo tasso di errore di tipo I risulta inferiore al valore nominale e diminuisce all'aumentare del numero dei gruppi confrontati. In presenza di elevata eteroschedasticità (Figura 5.47a), il metodo dei modelli ad effetti misti diventa ancora più conservativo, con il tasso di errore di tipo I che si abbassa progressivamente al crescere dell'intensità dell'eteroschedasticità.

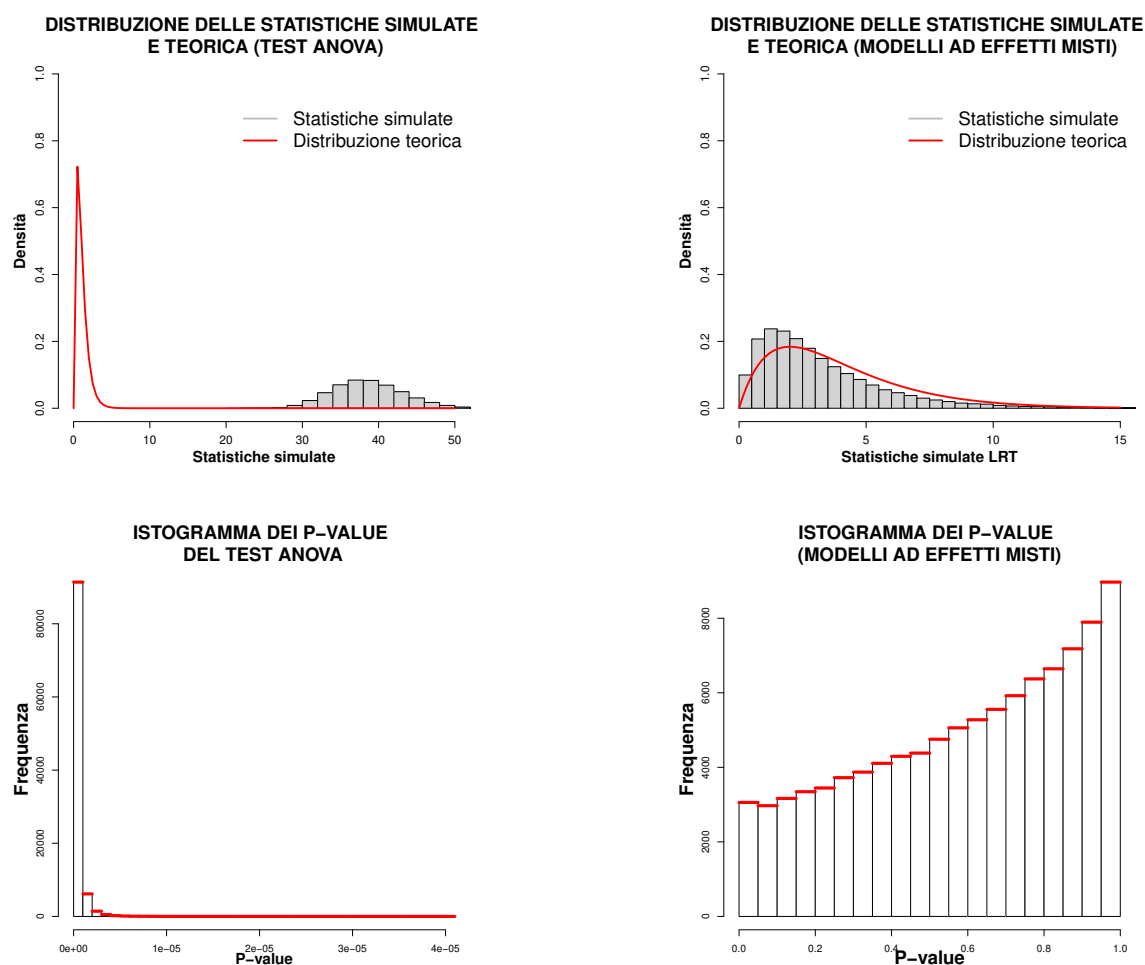


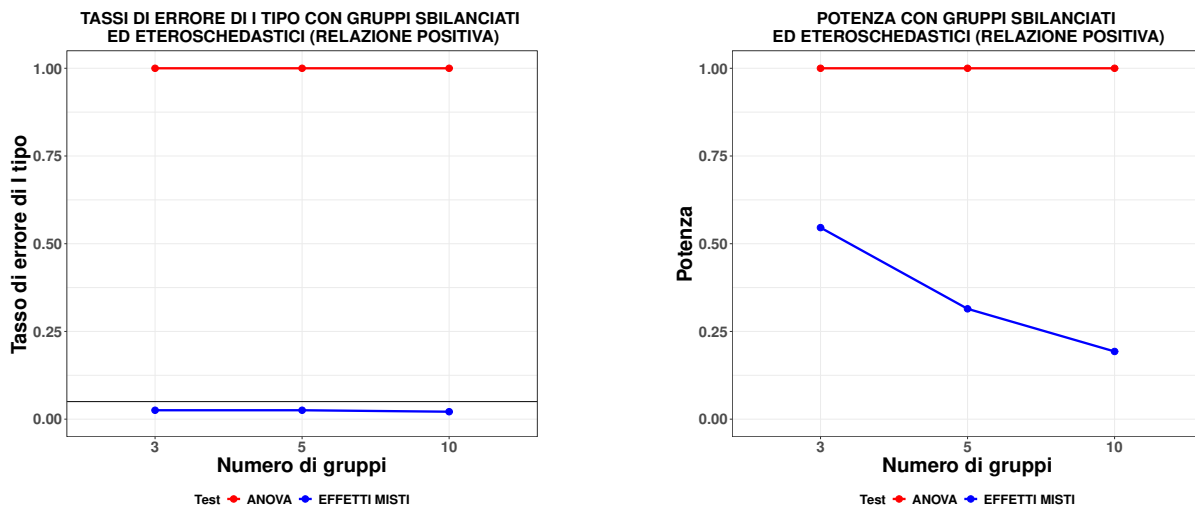
Figura 5.46: Confronto della distribuzione delle statistiche simulate e teorica e dei p-value sotto H_0 , scenario B2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0.5$.

Entrambi i metodi mostrano potenze elevate indipendentemente dal numero di gruppi e dalle loro dimensioni se l'eteroschedasticità è moderata (Figura 5.45b). Tuttavia, se l'eteroschedasticità diventa elevata (Figura 5.47b), la potenza dell'ANOVA rimane invariata mentre quella dei modelli ad effetti misti diminuisce, con una riduzione più marcata se il numero di gruppi confrontati è elevato.

Come illustrato nelle Figure 5.46 e 5.48, la distribuzione delle statistiche F simulate si discosta significativamente dalla distribuzione teorica prevista sotto l'ipotesi nulla quando i 5 gruppi confrontati sono sbilanciati ed eteroschedastici. In questo scenario, i p-value risultanti sono tutti inferiori a 0.05 e, di conseguenza, non si distribuiscono uniformemente tra 0 e 1.

L'approccio con modelli ad effetti misti, essendo conservativo in caso di eteroschedasticità moderata, è meno incline a rifiutare l'ipotesi nulla, quindi produce p-value generalmente più alti rispetto a quelli attesi dall'uniformità. Questo si riflette in una coda delle statistiche LRT simulate meno pesante rispetto a quella teorica (Figura 5.46).

Quando l'eteroschedasticità diventa elevata (Figura 5.48) e il metodo risulta più conservativo, la distribuzione teorica sovrastima ulteriormente la probabilità di ottenere statistiche LRT elevate, aumentando così la concentrazione di p-value alti.



(a) Scenario B2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0.5$.

(b) Scenario B2 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$, $s = 3$ e $r = 0.5$.

Figura 5.47: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati con eteroschedasticità elevata (relazione positiva), osservazioni dipendenti.

In presenza di gruppi non solo moderatamente eteroschedastici ma anche sbilanciati, con una relazione negativa tra numerosità dei gruppi e varianze (Figura 5.49a), l'ANOVA presenta un tasso di errore di tipo I massimo, rifiutando erroneamente sempre l'ipotesi nulla.

In questo scenario, l'approccio con modelli ad effetti misti mostra un tasso di errore di I tipo superiore al valore nominale, diventando progressivamente più liberale all'aumentare del numero di gruppi comparati. Se l'eteroschedasticità diventa elevata (Figura 5.51a), l'ANOVA continua ad avere un tasso di errore di tipo I massimo, mentre l'approccio con modelli ad effetti misti diventa sempre più liberale a parità del numero di gruppi confrontati.

Le potenze risultano alte per entrambe le tecniche indipendentemente dal numero di gruppi, dalle loro dimensioni e dall'intensità dell'eteroschedasticità (Figure 5.49b e 5.51b).

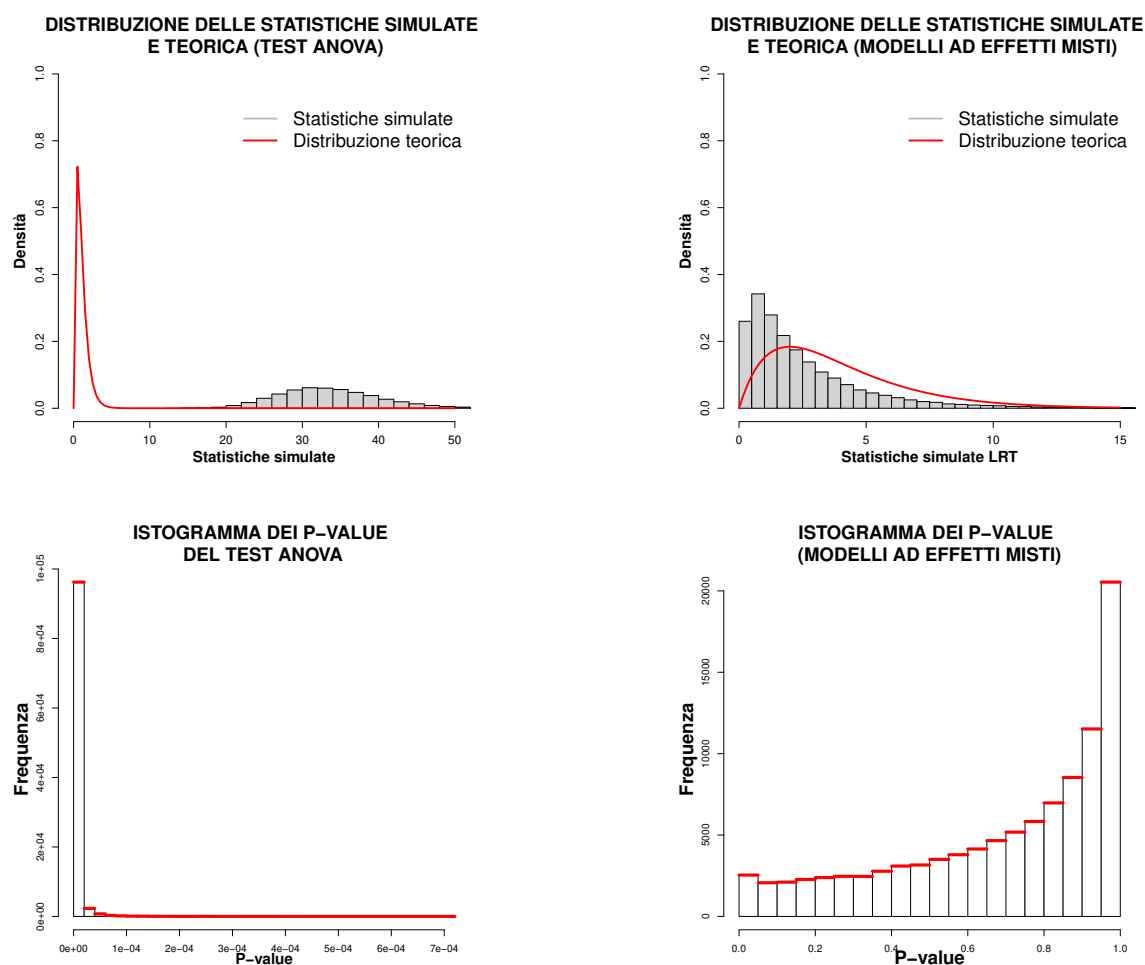
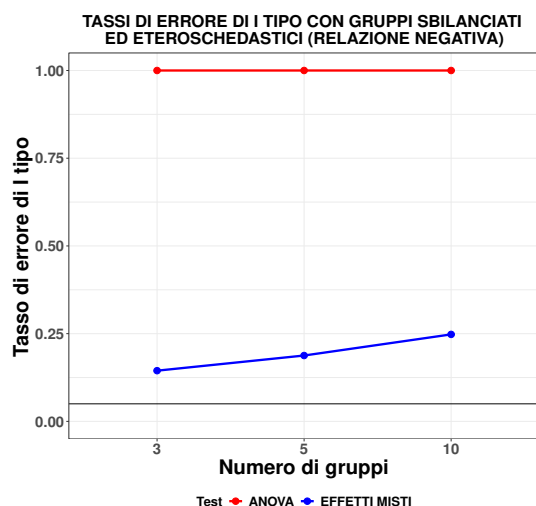


Figura 5.48: Confronto della distribuzione delle statistiche simulate e teorica e dei p-value sotto H_0 , scenario B2 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0.5$.

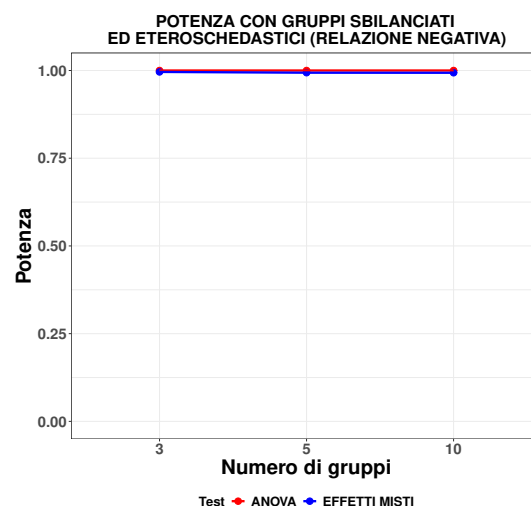
Tutti i p-value dell'ANOVA, avendo un tasso di errore di I tipo pari a 1 sia in presenza di eteroschedasticità moderata che elevata, sono inferiori al valore nominale e non si distribuiscono uniformemente tra 0 e 1 sotto l'ipotesi nulla.

L'approccio con modelli ad effetti misti, essendo troppo liberale e quindi più incline a rifiutare l'ipotesi nulla, presenta una prima barra dei p-value più alta rispetto alle altre e una coda della distribuzione delle statistiche LRT simulate più pesante rispetto a quella teorica prevista sotto l'ipotesi nulla in presenza di cinque gruppi sbilanciati, moderatamente eteroschedastici e con una relazione negativa tra numerosità e varianza (Figura 5.50).

Tale effetto è ancora più marcato quando l'eteroschedasticità diventa elevata (Figura 5.52).



(a) Scenario B3 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0.5$.



(b) Scenario B3 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$, $s = 1.5$ e $r = 0.5$.

Figura 5.49: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati e moderatamente eteroschedastici (relazione negativa), osservazioni dipendenti.

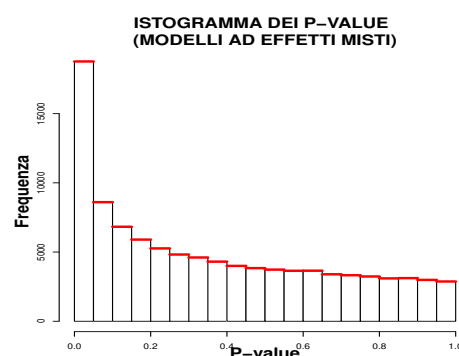
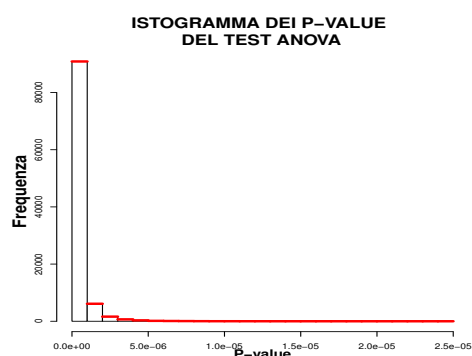
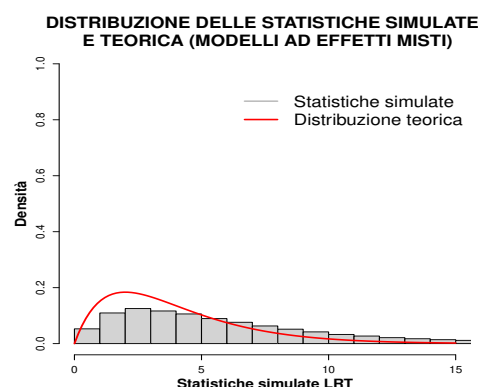
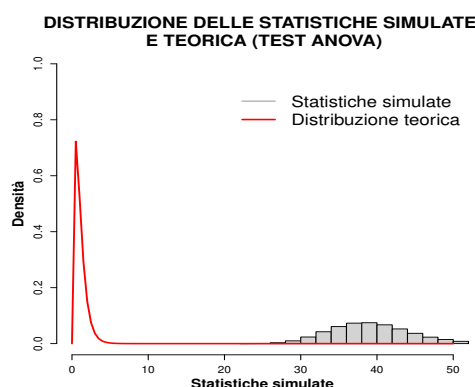
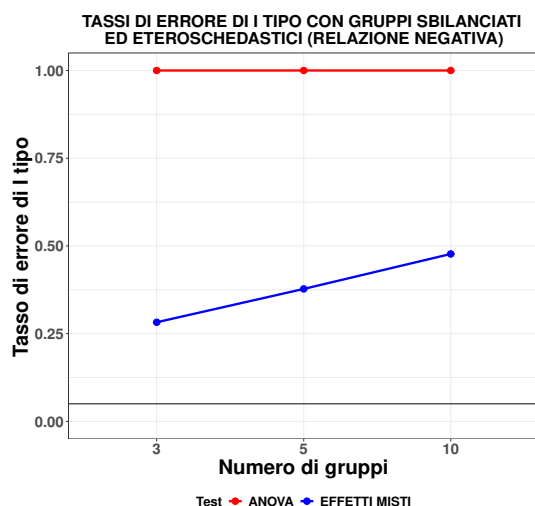
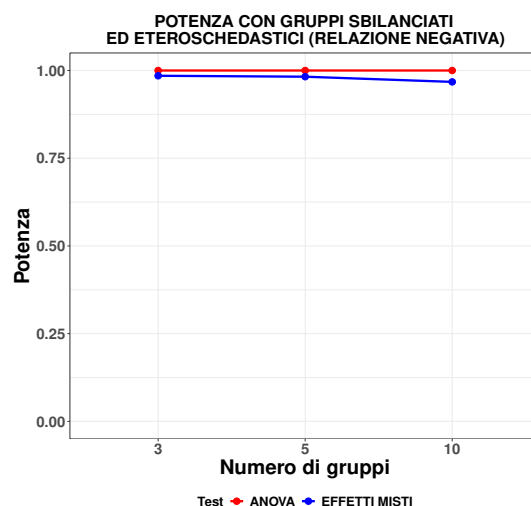


Figura 5.50: Confronto della distribuzione delle statistiche simulate e teorica e dei p-value sotto H_0 , scenario B3 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 1.5$ e $r = 0.5$.



(a) Scenario B3 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0.5$.



(b) Scenario B3 con $G \in \{3,5,10\}$, $a = 5$, $b = 10$, $c = 20$, $m = 2$, $s = 3$ e $r = 0.5$.

Figura 5.51: Tasso di errore di I tipo e potenza in presenza di gruppi sbilanciati con eteroschedasticità elevata (relazione negativa), osservazioni dipendenti.

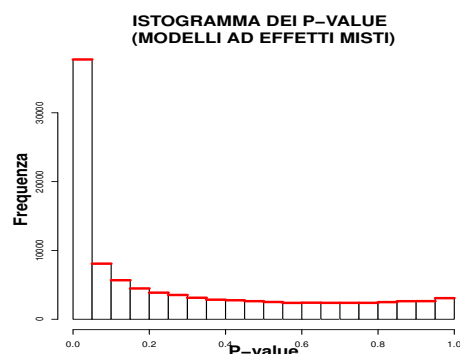
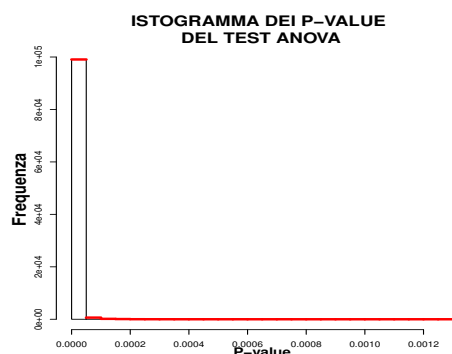
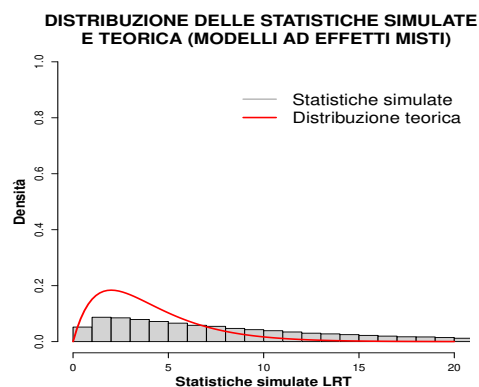
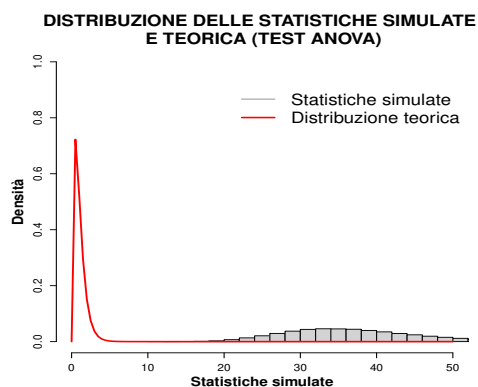


Figura 5.52: Confronto della distribuzione delle statistiche simulate e teorica e dei p-value sotto H_0 , scenario B3 con $G = 5$, $a = 5$, $b = 10$, $c = 20$, $m = 0$, $s = 3$ e $r = 0.5$.

5.5 Discussione

Di seguito vengono riportate le principali osservazioni emerse dall'applicazione dei diversi metodi in base ai risultati ottenuti dallo studio di simulazione. Questo ci permette di mettere in evidenza i punti di forza e le limitazioni di ciascun test nei diversi contesti esaminati. È importante considerare attentamente le caratteristiche dei dati e scegliere il metodo statistico più adatto in modo tale da garantire risultati validi e affidabili.

Nel contesto di gruppi bilanciati e con omoschedasticità, indipendentemente dalla loro quantità e numerosità, il test ANOVA emerge come il più appropriato tra quelli considerati. Tuttavia, se i gruppi hanno un considerevole numero di osservazioni al loro interno, anche altri test mantengono la loro validità, grazie a un livello del tasso di errore di tipo I prossimo allo 0.05 e ad elevate potenze statistiche.

Nel caso in cui i gruppi da confrontare presentino omoschedasticità ma siano sbilanciati, l'ANOVA rimane una scelta adeguata se il numero di gruppi è elevato perché mantiene un tasso di errore di primo tipo prossimo allo 0.05, p-value uniformi tra 0 e 1 e una buona aderenza della distribuzione teorica a quella delle statistiche test simulate. Ciò nonostante, lo sbilanciamento influisce negativamente sulla sua potenza, che risulta inferiore rispetto agli altri metodi. Nel caso di un numero limitato di gruppi, è preferibile utilizzare i test alternativi, come quello di James del secondo ordine o di Kruskal-Wallis, che offrono un miglior compromesso tra il livello del tasso di errore e il valore della potenza.

Se ci sono pochi gruppi da confrontare, che sono bilanciati ma eteroschedastici, l'ANOVA non è il metodo più adatto a causa di un alto tasso di errore di I tipo, di una potenza ridotta e di p-value non uniformi sotto l'ipotesi nulla. Analogamente, il test di Kruskal-Wallis non è appropriato per gli stessi motivi. In queste situazioni, i test di Welch, di Welch con medie troncate e varianze winorizzate, di Alexander-Govern e di James del secondo ordine sono più adatti, anche quando i gruppi sono numerosi e di grandi dimensioni. Quando le numerosità sono basse, è preferibile considerare il test di James del secondo ordine o di Alexander-Govern.

Se i gruppi da confrontare sono sia sbilanciati che eteroschedastici, dallo studio di simulazione effettuato possiamo affermare che né l'ANOVA né il test di Kruskal-Wallis sono appropriati. Indipendentemente dal fatto che ci sia una relazione positiva o negativa tra numerosità e varianza, e tenendo in considerazione sia i tassi di errore di primo tipo che la potenza, i test di James del secondo ordine, di Alexander-Govern, di Welch e di Welch con medie troncate e varianze winsorizzate sono più adeguati se i gruppi da comparare sono pochi. Tuttavia, se i gruppi sono numerosi, le opzioni migliori si restringono ai test di James del secondo ordine e di Alexander-Govern.

Nel contesto di gruppi di dati bilanciati e omoschedastici, ma non indipendenti tra loro a causa della rilevazione di più misurazioni sugli stessi soggetti, l'approccio con modelli ad effetti misti risulta chiaramente più adeguato rispetto al test ANOVA tradizionale. Quest'ultimo mostra un tasso di errore di primo tipo elevato, p-value non uniformi e una distribuzione delle statistiche F simulate che si discosta notevolmente da quella teorica prevista sotto l'ipotesi nulla. Inoltre, il test ANOVA diventa progressivamente più liberale all'aumentare del numero di gruppi confrontati.

Quando i gruppi sono omoschedastici, dipendenti ma sbilanciati a causa di misurazioni mancanti per alcuni soggetti, il tasso di errore di primo tipo dell'ANOVA raggiunge valori massimi, rendendo il test inutile nella pratica, poiché tende a rifiutare sempre l'ipotesi nulla anche quando essa è vera. Al contrario, l'approccio con modelli ad effetti misti, nonostante lo sbilanciamento del disegno sperimentale, mantiene una potenza elevata e un tasso di errore di primo tipo solo leggermente superiore allo 0.05. Tuttavia, l'effetto dello sbilanciamento è evidenziato da p-value meno uniformi rispetto al contesto di gruppi bilanciati.

Nel contesto di gruppi di dati dipendenti, bilanciati ma eteroschedastici, il test ANOVA tradizionale si dimostra troppo liberale, con un tasso di errore di primo tipo che aumenta al crescere del numero di gruppi confrontati. Inoltre, la coda delle statistiche F simulate è più pesante rispetto a quella prevista dalla teoria sotto l'ipotesi nulla. Anche l'approccio con modelli ad effetti misti risulta troppo liberale, ma il suo tasso di errore di primo tipo è più vicino al valore nominale, rendendolo preferibile rispetto all'ANOVA tradizionale in questo contesto. Anche in termini di potenza, il metodo con modelli ad effetti misti si conferma più adatto rispetto all'ANOVA, sebbene la non omoschedasticità abbia comunque un impatto negativo su di esso. All'aumentare dell'intensità dell'eteroschedasticità, il suo tasso di errore di primo tipo si discosta sempre più dal valore nominale e la sua potenza, pur rimanendo superiore a quella dell'ANOVA, diminuisce.

L'approccio con modelli ad effetti misti, sebbene più conservativo, rappresenta la scelta più adatta rispetto all'ANOVA tradizionale anche in contesti di gruppi dipendenti, eteroschedastici e sbilanciati, con relazione positiva tra numerosità e varianza, grazie al mantenimento di un tasso di errore di tipo I più vicino al valore nominale. Infatti, l'ANOVA tradizionale mostra un tasso di errore di tipo I pari a 1, un comportamento altamente indesiderabile per un test statistico poiché rifiuta sempre l'ipotesi nulla indipendentemente dai dati a disposizione. Va sottolineato che l'approccio con modelli ad effetti misti diventa progressivamente più conservativo con l'aumentare dell'intensità dell'eteroschedasticità. Inoltre, tutti i p-value ricavati dal test ANOVA sono inferiori a 0.05, quindi non si distribuiscono uniformemente tra 0 e 1. Invece, l'approccio con modelli ad effetti misti presenta una concentrazione maggiore di p-value elevati, e di conse-

guenza la coda delle statistiche LRT simulate è meno pesante rispetto a quella della distribuzione prevista sotto l'ipotesi nulla.

Anche in presenza di una relazione negativa tra numerosità interna ai gruppi e varianza, l'approccio con modelli ad effetti misti si conferma più idoneo anche in tale contesto. Sebbene il suo tasso di errore di tipo I sia superiore al valore nominale, esso rimane comunque più vicino a questo valore rispetto a quello dell'ANOVA, che raggiunge un tasso di errore massimo, rendendo il test inutile nella pratica. È importante notare che l'approccio con modelli ad effetti misti diventa progressivamente più liberale con l'aumentare del numero di gruppi confrontati e con l'incremento dell'intensità dell'eteroschedasticità. Nonostante questa tecnica sia la migliore tra le due analizzate, la coda della distribuzione delle statistiche LRT simulate risulta più pesante rispetto a quella teorica, specialmente in condizioni di elevata eteroschedasticità.

Capitolo 6

CONCLUSIONI

L'analisi della varianza (ANOVA) presuppone che i dati siano distribuiti normalmente, che le varianze siano omogenee e che le osservazioni siano indipendenti. Quando una o più di queste assunzioni sono violate, l'affidabilità delle conclusioni dell'ANOVA può essere compromessa. In tali situazioni, è importante adottare tecniche alternative che possano gestire efficacemente queste violazioni. Dopo aver discusso teoricamente i test alternativi all'ANOVA, è stato condotto uno studio di simulazione con l'obiettivo di valutare l'impatto delle violazioni delle assunzioni sull'ANOVA stessa e analizzare le prestazioni dei metodi alternativi in contesti specifici.

Dallo studio è emerso che nella situazione in cui i presupposti dell'ANOVA sono soddisfatti, ossia i gruppi sono bilanciati, omoschedastici e indipendenti, il test ANOVA risulta essere il più adatto grazie a un tasso di errore di primo tipo vicino al valore nominale e ad una potenza statistica elevata. Tuttavia, se i gruppi sono sbilanciati, l'ANOVA mantiene il miglior tasso di errore di primo tipo, ma lo sbilanciamento influisce negativamente sulla sua potenza. In tali casi, soprattutto se i gruppi confrontati sono pochi, i test alternativi offrono un miglior compromesso tra tasso di errore e potenza.

Se viene violato solo il presupposto di omoschedasticità, l'ANOVA diventa troppo liberale quando i gruppi sono bilanciati o sbilanciati ma con una relazione negativa tra numerosità e varianza. Diventa invece troppo conservativa in presenza di gruppi sbilanciati con una relazione positiva tra numerosità e varianza. Tali effetti sono più marcati al crescere dell'intensità dell'eteroschedasticità. Il test ANOVA, quindi, non offre una prestazione desiderabile in caso di eteroschedasticità, come anche il test di Kruskal-Wallis. Invece, i test alternativi, come quello di Welch, di Welch con medie troncate e varianze winsorizzate, e ancora di più i test di Alexander-Govern e di James del secondo ordine, sono preferibili.

Se viene violato solo il presupposto di indipendenza perché si rilevano più misurazioni sugli stessi soggetti, il tasso di errore di primo tipo dell'ANOVA aumenta rendendo il test via via più

liberale al crescere del numero di gruppi bilanciati e omoschedastici confrontati. Se i gruppi sono anche sbilanciati, il tasso dell'ANOVA raggiunge il valore 1, rendendo il test inutile nella pratica. Di conseguenza, se i gruppi di dati sono dipendenti, l'ANOVA non è il test adatto per valutare l'omogeneità delle medie, a differenza dell'approccio con modelli ad effetti misti che, nonostante sia leggermente liberale, offre una performance migliore.

Se vengono violati sia il presupposto di omoschedasticità sia quello di indipendenza, il tasso di errore di I tipo dell'ANOVA è molto più alto del valore nominale se i gruppi sono bilanciati e raggiunge il valore massimo quando i gruppi sono sbilanciati. Anche in tale contesto, l'ANOVA non è adeguata, mentre l'approccio con modelli ad effetti misti si conferma migliore. È importante considerare però come cambia il comportamento di quest'ultimo metodo al variare degli scenari. Se i gruppi sono bilanciati o i gruppi sono sbilanciati con una relazione negativa tra numerosità e varianza, esso risulta più liberale. Se i gruppi sono sbilanciati ma la relazione tra numerosità e varianza è positiva, diventa più conservativo. Tali effetti sono più marcati al crescere dell'intensità dell'eteroschedasticità.

Limitazioni e studi futuri

Sebbene il presente studio abbia analizzato diversi test in molteplici situazioni di eteroschedasticità e dipendenza dei dati utilizzando 100000 simulazioni, ci sono delle limitazioni che devono essere considerate per comprendere la generalizzabilità dei risultati ottenuti.

Le condizioni specifiche di eteroschedasticità e dipendenza dei dati simulate rappresentano solo un sottogruppo delle possibili configurazioni che si possono verificare nella pratica, e altre forme di queste condizioni potrebbero portare a risultati differenti.

Un altro aspetto da considerare è che lo studio ha confrontato 3, 5 o 10 gruppi, ma per la valutazione della potenza è stato sempre considerato un solo gruppo con una media diversa. Questo implica che la proporzione di gruppi con media diversa non è stata uniforme nei diversi casi considerati, il che potrebbe influenzare i risultati dei test statistici.

In termini di simulazione, i dati sono stati generati utilizzando solo distribuzioni normali. Tuttavia, in molte applicazioni reali, i dati possono non seguire una distribuzione normale, il che potrebbe influenzare l'accuratezza dei test statistici impiegati. Inoltre, sebbene siano stati compiuti sforzi per verificare accuratamente il codice in R, è possibile che errori di programmazione abbiano influenzato i risultati.

Studi futuri potrebbero espandere quanto già considerato aumentando il numero di gruppi confrontati e le rispettive numerosità. Ulteriori ricerche potrebbero esplorare una gamma più ampia di condizioni, includendo diverse forme di eteroschedasticità, differenti schemi di dipendenza dei dati e distribuzioni dei dati non normali.

Bibliografia

- [1] G. Cicchitelli, P. D'Urso, M. Minozzo et al., «Statistica: principi e metodi (terza edizione),» 2017.
- [2] D. C. Montgomery, *Design and analysis of experiments*. John Wiley & Sons, 2017.
- [3] M. Grigoletto, L. Ventura e F. Pauli, *Modello lineare: teoria e applicazioni con R*. G. Giappichelli Editore, 2017.
- [4] M. Delacre, C. Leys, Y. L. Mora e D. Lakens, «Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA,» *International Review of Social Psychology*, vol. 32, n. 1, p. 13, 2019.
- [5] A. De Beuckelaer, «A closer examination on some parametric alternatives to the ANOVA F-test,» *Statistical Papers*, vol. 37, pp. 291–305, 1996.
- [6] Y. Sheng, «Testing the assumptions of analysis of variance,» *Best practices in quantitative methods*, pp. 324–340, 2008.
- [7] A. Katsileros, N. Antonetsis, P. Mouzaidis, E. Tani, P. J. Bebeli e A. Karagrigoriou, «A comparison of tests for homoscedasticity using simulation and empirical data,» *Communications for Statistical Applications and Methods*, vol. 31, n. 1, pp. 1–35, 2024.
- [8] J. Chukwudi, O. Idochi, I. O. Sylvia et al., «Effect of sample sizes on the empirical power of some tests of homogeneity of variances,» *International Journal of Mathematics Trends and Technology-IJMTT*, vol. 65, 2019.
- [9] I. Adeleke, T. F. Oladeji e O. S. Adesina, «Exploring Robust Methods for Testing Equality of Variances,» *International Journal of Statistics and Economics*, vol. 19, n. 3, pp. 65–74, 2018.
- [10] D. C. Howell, *Statistical methods for psychology*. PWS-Kent Publishing Co, 1992.
- [11] R. Christensen, *Analysis of variance, design, and regression: applied statistical methods*. CRC press, 1996.

- [12] H. Liu, *Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance*. Virginia Commonwealth University, 2015.
- [13] S. ASSUMPTION, «UNDERSTANDING THE REPEATED-MEASURES ANOVA,»
- [14] D Lane, «The assumption of sphericity in repeated-measures designs: What it means and what to do when it is violated,» *Quantitative Methods for Psychology*, vol. 12, n. 2, pp. 114–122, 2016.
- [15] D. G. Pereira, A. Afonso e F. M. Medeiros, «Overview of Friedman’s test and post-hoc analysis,» *Communications in Statistics-Simulation and Computation*, vol. 44, n. 10, pp. 2636–2653, 2015.
- [16] A. Salvan, N. Sartori, L. Pace, A. Salvan, N. Sartori e L. Pace, *Modelli lineari generalizzati*. Springer, 2020.
- [17] O. Dag, A. Dolgun e N. M. Konar, «Onewaytests: An R Package for One-Way Tests in Independent Groups Designs.,» *R Journal*, vol. 10, n. 1, 2018.
- [18] C. Schuster e A. Von Eye, «The relationship of ANOVA models with random effects and repeated measurement designs,» *Journal of Adolescent Research*, vol. 16, n. 2, pp. 205–220, 2001.
- [19] L. N. Muhammad et al., «Guidelines for repeated measures statistical analysis approaches with basic science research considerations,» *The Journal of Clinical Investigation*, vol. 133, n. 11, 2023.
- [20] A. C. Bathke, S. Friedrich, M. Pauly et al., «Testing mean differences among groups: multivariate and repeated measures analysis with minimal assumptions,» *Multivariate Behavioral Research*, vol. 53, n. 3, pp. 348–359, 2018.
- [21] H. Keselman, J. Algina e R. K. Kowalchuk, «The analysis of repeated measures designs: a review,» *British Journal of Mathematical and Statistical Psychology*, vol. 54, n. 1, pp. 1–20, 2001.

Appendice A

TABELLE DEI RISULTATI

F = test ANOVA

K-W = test di Kruskal-Wallis

WELCH = test di Welch

W DW = test di Welch con medie troncate e varianze winsorizzate

A-G = test di Alexander-Govern

JAMES = test di James del secondo ordine

MISTI = modelli ad effetti misti

Gruppi	Numerosità	F	K-W	WELCH	W DW	A-G	JAMES
3	5	0.04875	0.04488	0.04577	0.04577	0.04512	0.04814
3	10	0.04975	0.04574	0.04862	0.04924	0.04705	0.04929
3	20	0.04975	0.04827	0.04921	0.05021	0.04813	0.04933
5	5	0.04997	0.03705	0.05544	0.05544	0.05180	0.05112
5	10	0.05031	0.04456	0.05174	0.05644	0.04981	0.05025
5	20	0.05066	0.04707	0.05041	0.05252	0.04934	0.05001
10	5	0.04902	0.03372	0.08166	0.08166	0.05848	0.05919
10	10	0.04975	0.04283	0.05820	0.07084	0.05098	0.05136
10	20	0.05014	0.04677	0.05253	0.05679	0.05036	0.05066

Tabella A.1: Tassi di errore empirici di I tipo con gruppi bilanciati, omoschedastici e osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.05028	0.04563	0.04862	0.05446	0.05524	0.05285
5	0.05073	0.04254	0.05174	0.06295	0.05994	0.05465
10	0.05136	0.04034	0.04921	0.07960	0.06089	0.05793

Tabella A.2: Tassi di errore empirici di I tipo con gruppi sbilanciati, omoschedastici e osservazioni indipendenti

Gruppi	Numerosità	F	K-W	WELCH	W DW	A-G	JAMES
3	5	0.76113	0.77210	0.75053	0.75053	0.72419	0.76069
3	10	0.98311	0.99145	0.99240	0.98385	0.99128	0.99252
3	20	0.98311	0.99999	1.00000	0.99999	1.00000	1.00000
5	5	0.66639	0.69126	0.68694	0.68694	0.59326	0.67241
5	10	0.95052	0.98924	0.98590	0.96941	0.97572	0.98533
5	20	0.99966	1.00000	1.00000	0.99999	1.00000	1.00000
10	5	0.48331	0.45835	0.60552	0.60552	0.40461	0.53381
10	10	0.81276	0.96866	0.96280	0.92935	0.88937	0.95692
10	20	0.98616	0.99998	0.99999	0.99994	0.99991	0.99999

Tabella A.3: Potenza dei test con gruppi bilanciati, omoschedastici, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.25339	0.99489	0.99077	0.98186	0.99337	0.98998
5	0.64564	0.99245	0.98094	0.96553	0.97876	0.97581
10	0.63931	0.97172	0.94676	0.91735	0.88991	0.91908

Tabella A.4: Potenza dei test con gruppi sbilanciati, omoschedastici, osservazioni indipendenti.

Gruppi	Numerosità	F	K-W	WELCH	W DW	A-G	JAMES
3	5	0.06761	0.04794	0.04725	0.04725	0.04612	0.04940
3	10	0.06810	0.04861	0.04952	0.05034	0.04805	0.05001
3	20	0.06810	0.05146	0.05122	0.05131	0.05051	0.05136
5	5	0.06038	0.03881	0.05479	0.05479	0.05021	0.05095
5	10	0.05919	0.04789	0.05286	0.05746	0.05065	0.05133
5	20	0.05911	0.05008	0.05102	0.05242	0.05006	0.05061
10	5	0.05134	0.03685	0.08312	0.08312	0.05903	0.06018
10	10	0.04987	0.04576	0.05827	0.06969	0.05030	0.05096
10	20	0.04974	0.04980	0.05200	0.05701	0.04955	0.04964

Tabella A.5: Tassi di errore empirici di I tipo con gruppi bilanciati, eteroschedasticità moderata, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.03311	0.02724	0.05152	0.05132	0.05181	0.05028
5	0.02722	0.02450	0.05930	0.05844	0.05783	0.05240
10	0.04666	0.03978	0.07893	0.08237	0.06232	0.06010

Tabella A.6: Tassi di errore empirici di I tipo con gruppi sbilanciati, eteroschedasticità moderata, relazione positiva tra numerosità e varianza, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.02401	0.03134	0.05126	0.05126	0.05213	0.05054
5	0.10079	0.07406	0.06568	0.06565	0.05924	0.05669
10	0.06931	0.04953	0.07849	0.08048	0.06202	0.05909

Tabella A.7: Tassi di errore empirici di I tipo con gruppi sbilanciati, eteroschedasticità moderata, relazione negativa tra numerosità e varianza, osservazioni indipendenti.

Gruppi	Numerosità	F	K-W	WELCH	W DW	A-G	JAMES
3	5	0.61916	0.66541	0.69558	0.69558	0.68432	0.70523
3	10	0.91439	0.97596	0.98474	0.97103	0.98396	0.98501
3	20	0.91439	0.99998	1.00000	0.99995	1.00000	1.00000
5	5	0.51525	0.56198	0.65875	0.65875	0.58767	0.64220
5	10	0.83599	0.97384	0.98103	0.96144	0.97170	0.98022
5	20	0.98865	0.99997	1.00000	0.99996	1.00000	1.00000
10	5	0.34237	0.35059	0.58830	0.58830	0.41145	0.51710
10	10	0.62457	0.92135	0.95738	0.92175	0.89174	0.95083
10	20	0.90897	0.99997	0.99999	0.99990	0.99996	0.99999

Tabella A.8: Potenza dei test con gruppi bilanciati, eteroschedasticità moderata, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.16457	0.96668	0.97712	0.96515	0.98121	0.97575
5	0.16745	0.16230	0.19996	0.18054	0.22198	0.17933
10	0.24174	0.31594	0.29483	0.25494	0.36223	0.22717

Tabella A.9: Potenza dei test con gruppi sbilanciati, eteroschedasticità moderata, relazione positiva tra numerosità e varianza, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.28846	0.99293	0.98650	0.97567	0.99226	0.98509
5	0.25760	0.27061	0.22602	0.22097	0.26993	0.20106
10	0.32422	0.41582	0.28580	0.27239	0.34705	0.22109

Tabella A.10: Potenza dei test con gruppi sbilanciati, eteroschedasticità moderata, relazione negativa tra numerosità e varianza, osservazioni indipendenti.

Gruppi	Numerosità	F	K-W	WELCH	W DW	A-G	JAMES
3	5	0.10516	0.06074	0.05059	0.05059	0.04679	0.05222
3	10	0.09996	0.06232	0.04991	0.05214	0.04757	0.05003
3	20	0.09996	0.06402	0.04942	0.05031	0.04803	0.04941
5	5	0.07871	0.05728	0.05868	0.05868	0.05006	0.05416
5	10	0.07545	0.06441	0.05353	0.05940	0.05003	0.05151
5	20	0.07395	0.06504	0.05067	0.05280	0.04970	0.05019
10	5	0.05278	0.05192	0.08772	0.08772	0.05705	0.06413
10	10	0.05233	0.06176	0.05840	0.07216	0.05013	0.05136
10	20	0.05144	0.06497	0.05274	0.05630	0.05029	0.05079

Tabella A.11: Tassi di errore empirici di I tipo con gruppi bilanciati, eteroschedasticità elevata, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.02126	0.01990	0.05130	0.05096	0.05071	0.05088
5	0.03086	0.03401	0.05854	0.05861	0.05283	0.05197
10	0.04300	0.04716	0.08088	0.08338	0.05947	0.06051

Tabella A.12: Tassi di errore empirici di I tipo con gruppi sbilanciati, eteroschedasticità elevata, relazione positiva tra numerosità e varianza, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.22715	0.10219	0.05651	0.05651	0.05235	0.05364
5	0.13699	0.09446	0.06084	0.05986	0.05497	0.05257
10	0.11815	0.06952	0.07858	0.08225	0.06076	0.05964

Tabella A.13: Tassi di errore empirici di I tipo con gruppi sbilanciati, eteroschedasticità elevata, relazione negativa tra numerosità e varianza, osservazioni indipendenti.

Gruppi	Numerosità	F	K-W	WELCH	W DW	A-G	JAMES
3	5	0.37117	0.42420	0.62527	0.62527	0.63027	0.63168
3	10	0.58513	0.89353	0.96975	0.94902	0.97045	0.96982
3	20	0.58513	0.99940	0.99995	0.99987	0.99995	0.99995
5	5	0.27429	0.31177	0.61683	0.61683	0.57230	0.59588
5	10	0.46293	0.86047	0.97348	0.95001	0.96646	0.97207
5	20	0.73587	0.99950	0.99998	0.99995	0.99998	0.99998
10	5	0.15666	0.20404	0.57049	0.57049	0.41668	0.49504
10	10	0.27046	0.64300	0.94947	0.91271	0.88936	0.94153
10	20	0.48562	0.99467	0.99990	0.99973	0.99974	0.99990

Tabella A.14: Potenza dei test con gruppi bilanciati, eteroschedasticità elevata, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.05903	0.62412	0.91897	0.90306	0.91686	0.91624
5	0.07074	0.07202	0.15976	0.15105	0.15867	0.14353
10	0.10760	0.16003	0.27761	0.24130	0.33447	0.21147

Tabella A.15: Potenza dei test con gruppi sbilanciati, eteroschedasticità elevata, relazione positiva tra numerosità e varianza, osservazioni indipendenti.

Gruppi	F	K-W	WELCH	W DW	A-G	JAMES
3	0.30835	0.98866	0.98266	0.96981	0.99162	0.98078
5	0.24159	0.30755	0.18657	0.22813	0.22008	0.16576
10	0.25141	0.38979	0.21388	0.26717	0.24636	0.16487

Tabella A.16: Potenza dei test con gruppi sbilanciati, eteroschedasticità elevata, relazione negativa tra numerosità e varianza, osservazioni indipendenti.

Gruppi	Numerosità	F	MISTI
3	5	0.18436	0.09384
3	10	0.21059	0.06882
3	20	0.21059	0.06882
5	5	0.75222	0.09250
5	10	0.62023	0.06794
5	20	0.57180	0.05702
10	5	0.85843	0.09708
10	10	0.99566	0.06987
10	20	0.98071	0.06017

Tabella A.17: Tassi di errore empirici di I tipo con gruppi bilanciati, omoschedastici e osservazioni dipendenti.

Gruppi	F	MISTI
3	1.00000	0.07540
5	1.00000	0.06741
10	1.00000	0.06992

Tabella A.18: Tassi di errore empirici di I tipo con gruppi sbilanciati, omoschedastici e osservazioni dipendenti

Gruppi	Numerosità	F	MISTI
3	5	0.03567	0.99229
3	10	0.04850	1.00000
3	20	0.04850	1.00000
5	5	0.30974	0.99461
5	10	0.29182	1.00000
5	20	0.27838	1.00000
10	5	0.68248	0.99316
10	10	0.96451	1.00000
10	20	0.93105	1.00000

Tabella A.19: Potenza con gruppi bilanciati, omoschedastici, osservazioni dipendenti.

Gruppi	F	MISTI
3	1.00000	0.979999
5	1.00000	0.989999
10	1.00000	1.00000

Tabella A.20: Potenza con gruppi sbilanciati, omoschedastici, osservazioni dipendenti

Gruppi	Numerosità	F	MISTI
3	5	0.13658	0.10583
3	10	0.15901	0.07815
3	20	0.15901	0.07815
5	5	0.53357	0.11023
5	10	0.45519	0.08262
5	20	0.42477	0.07350
10	5	0.63379	0.11901
10	10	0.94953	0.08989
10	20	0.91145	0.07798

Tabella A.21: Tassi di errore empirici di I tipo con gruppi bilanciati, eteroschedasticità moderata e osservazioni dipendenti.

Gruppi	F	MISTI
3	1.00000	0.04129
5	1.00000	0.03059
10	1.00000	0.02576

Tabella A.22: Tassi di errore empirici di I tipo con gruppi sbilanciati, eteroschedasticità moderata, relazioni positiva tra numerosità e varianza, osservazioni dipendenti

Gruppi	F	MISTI
3	1.00000	0.14449
5	1.00000	0.18768
10	1.00000	0.24787

Tabella A.23: Tassi di errore empirici di I tipo con gruppi sbilanciati, eteroschedasticità moderata, relazioni negativa tra numerosità e varianza, osservazioni dipendenti

Gruppi	Numerosità	F	MISTI
3	5	0.03817	0.93010
3	10	0.00574	0.99926
3	20	0.00574	0.99926
5	5	0.23548	0.91841
5	10	0.22525	0.99959
5	20	0.22134	1.00000
10	5	0.49840	0.76025
10	10	0.87346	0.96981
10	20	0.83066	0.99987

Tabella A.24: Potenza con gruppi bilanciati, moderatamente eteroschedastici, osservazioni dipendenti.

Gruppi	F	MISTI
3	1.00000	0.99588
5	1.00000	0.98370
10	1.00000	0.99351

Tabella A.25: Potenza con gruppi sbilanciati, moderatamente eteroschedastici, relazione positiva tra numerosità e varianza, osservazioni dipendenti

Gruppi	F	MISTI
3	1.00000	0.99987
5	1.00000	0.99994
10	1.00000	0.99997

Tabella A.26: Potenza con gruppi sbilanciati, moderatamente eteroschedastici, relazione negativa tra numerosità e varianza, osservazioni dipendenti

Gruppi	Numerosità	F	MISTI
3	5	0.06587	0.14057
3	10	0.08511	0.10148
3	20	0.08511	0.10148
5	5	0.19044	0.14757
5	10	0.18027	0.11081
5	20	0.17711	0.09773
10	5	0.24462	0.15181
10	10	0.50093	0.11762
10	20	0.47139	0.10032

Tabella A.27: Tassi di errore empirici di I tipo con gruppi bilanciati, eteroschedasticità elevata e osservazioni dipendenti.

Gruppi	F	MISTI
3	1.00000	0.02540
5	1.00000	0.02542
10	1.00000	0.02124

Tabella A.28: Tassi di errore empirici di I tipo con gruppi sbilanciati, eteroschedasticità elevata, relazioni positiva tra numerosità e varianza, osservazioni dipendenti

Gruppi	F	MISTI
3	1.00000	0.28249
5	1.00000	0.37746
10	1.00000	0.47714

Tabella A.29: Tassi di errore empirici di I tipo con gruppi sbilanciati, eteroschedasticità elevata, relazioni negativa tra numerosità e varianza, osservazioni dipendenti

Gruppi	Numerosità	F	MISTI
3	5	0.03355	0.51400
3	10	0.05225	0.79824
3	20	0.05225	0.79824
5	5	0.10480	0.41866
5	10	0.11428	0.69871
5	20	0.11632	0.98959
10	5	0.21512	0.32265
10	10	0.43353	0.45559
10	20	0.41317	0.72444

Tabella A.30: Potenza con gruppi bilanciati, eteroschedasticità elevata, osservazioni dipendenti.

Gruppi	F	MISTI
3	1.00000	0.54592
5	1.00000	0.31446
10	1.00000	0.19294

Tabella A.31: Potenza con gruppi sbilanciati, eteroschedasticità elevata, relazione positiva tra numerosità e varianza, osservazioni dipendenti

Gruppi	F	MISTI
3	1.00000	0.99496
5	1.00000	0.98231
10	1.00000	0.96752

Tabella A.32: Potenza con gruppi sbilanciati, eteroschedasticità elevata, relazione negativa tra numerosità e varianza, osservazioni dipendenti

