

ELABORATO PALMIERI LORENZO 233036

Descrizione del Dataset

Il dataset `mtcars` è un classico set di dati di R che contiene informazioni tecniche e di prestazione su 32 modelli di automobili, principalmente dagli anni '70. Il dataset è stato raccolto dai test automobilistici della rivista Motor Trend negli anni '70 ed è stato originariamente pubblicato nel 1974 da Henderson e Velleman come parte del loro studio statistico sui dati delle auto. All'interno di questo progetto viene scelto di effettuare un'analisi multidimensionale di tipo esplorativo.

Obiettivo Finale dell'Analisi Fattoriale Esplorativa

L'obiettivo dell'analisi fattoriale esplorativa (EFA) sul dataset `mtcars` è di identificare e comprendere le strutture latenti sottostanti tra le variabili osservate. Specificamente, miriamo a:

1. **Ridurre la Dimensionalità:** Sintetizzare le 11 variabili originali in un numero inferiore di fattori latenti, mantenendo quanta più informazione possibile. Questo aiuta a semplificare il modello e a facilitare l'interpretazione delle relazioni tra le variabili.
2. **Identificare le Variabili Latenti:** Comprendere l'interrelazione tra variabili latenti e manifeste per stimare il fenomeno oggetto di studio.
3. **Interpretazione dei Fattori:** Fornire un'interpretazione significativa dei fattori estratti.
4. **Utilizzare i Risultati per le Decisioni:** Applicare i risultati dell'EFA per prendere decisioni informate, come identificare i modelli di auto con le migliori caratteristiche complessive o determinare quali variabili sono più importanti per predire le prestazioni del veicolo.

Lettura del dataset

Si esegue una prima lettura del dataset:

```
> str(mtcars)
'data.frame': 32 obs. of 11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Composizione del Dataset

Il dataset `mtcars` include 32 unità statistiche (diversi modelli di auto) e 11 variabili numeriche che descrivono le varie caratteristiche delle automobili. Una panoramica delle variabili presenti nel dataset:

- **mpg:** Miles per Gallon (miglia per gallone) – efficienza del carburante.
- **cyl:** Numero di cilindri – indica la configurazione del motore.
- **disp:** Displacement (cilindrata) in pollici cubici – volume totale dei cilindri del motore.
- **hp:** Gross horsepower (cavalli vapore) – potenza del motore.

- **drat**: Rear axle ratio (rapporto al ponte posteriore) – un indicatore delle prestazioni di trasmissione.
- **wt**: Peso dell'auto (in 1000 libbre).
- **qsec**: 1/4 mile time (tempo sul quarto di miglio) – tempo in secondi per percorrere un quarto di miglio.
- **vs**: V/S (motore a V contro motore in linea) – un indicatore binario della configurazione del motore.
- **am**: Trasmissione (0 = automatica, 1 = manuale).
- **gear**: Numero di marce – indica il numero di rapporti della trasmissione.
- **carb**: Numero di carburatori.

Vengono visualizzate le prime righe del dataset e delle principali statistiche di sintesi:

```
> head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

```
> summary(mtcars)
```

mpg		cyl		disp		hp		drat		wt		qsec		vs	
Min.	:10.4	Min.	:4.00	Min.	: 71.1	Min.	: 52.0	Min.	:2.76	Min.	:1.51	Min.	:14.5	Min.	:0.000
1st Qu.	:15.4	1st Qu.	:4.00	1st Qu.	:120.8	1st Qu.	: 96.5	1st Qu.	:3.08	1st Qu.	:2.58	1st Qu.	:16.9	1st Qu.	:0.000
Median	:19.2	Median	:6.00	Median	:196.3	Median	:123.0	Median	:3.69	Median	:3.33	Median	:17.7	Median	:0.000
Mean	:20.1	Mean	:6.19	Mean	:230.7	Mean	:146.7	Mean	:3.60	Mean	:3.22	Mean	:17.8	Mean	:0.438
3rd Qu.	:22.8	3rd Qu.	:8.00	3rd Qu.	:326.0	3rd Qu.	:180.0	3rd Qu.	:3.92	3rd Qu.	:3.61	3rd Qu.	:18.9	3rd Qu.	:1.000
Max.	:33.9	Max.	:8.00	Max.	:472.0	Max.	:335.0	Max.	:4.93	Max.	:5.42	Max.	:22.9	Max.	:1.000

am		gear		carb	
Min.	:0.000	Min.	:3.00	Min.	:1.00
1st Qu.	:0.000	1st Qu.	:3.00	1st Qu.	:2.00
Median	:0.000	Median	:4.00	Median	:2.00
Mean	:0.406	Mean	:3.69	Mean	:2.81
3rd Qu.	:1.000	3rd Qu.	:4.00	3rd Qu.	:4.00
Max.	:1.000	Max.	:5.00	Max.	:8.00

Spiegazione Dettagliata dell'Output

L'output fornisce una panoramica iniziale del dataset `mtcars`, che include sia la visualizzazione delle prime sei righe del dataset (`head(mtcars)`) sia un sommario statistico delle variabili (`summary(mtcars)`).

Le prime sei righe del dataset `mtcars` mostrano i valori delle variabili per sei modelli di auto specifici. Ogni riga rappresenta un modello di auto, e ogni colonna rappresenta una variabile che descrive le caratteristiche di quel modello.

Il sommario statistico fornisce informazioni descrittive sulle variabili del dataset. Per ciascuna variabile, vengono forniti i seguenti valori:

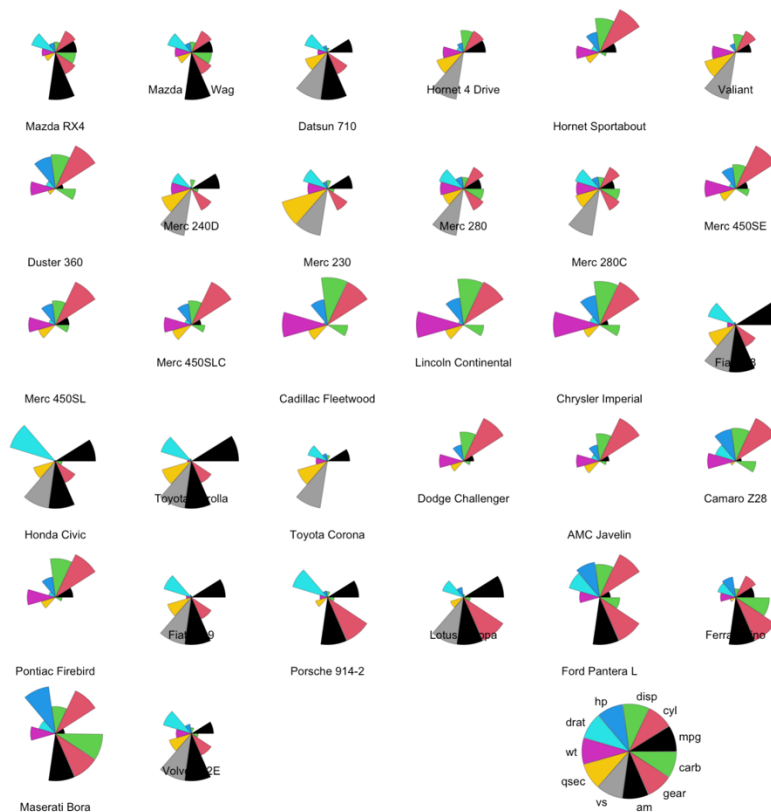
- **Min**: Valore minimo osservato.
- **1st Qu**: Primo quartile (25° percentile), rappresenta il valore sotto il quale cade il 25% delle osservazioni.
- **Median**: Mediana (50° percentile), rappresenta il valore centrale delle osservazioni.
- **Mean**: Media, il valore medio delle osservazioni.
- **3rd Qu**: Terzo quartile (75° percentile), rappresenta il valore sotto il quale cade il 75% delle osservazioni.
- **Max**: Valore massimo osservato.

ANALISI GRAFICA

Si è scelto di iniziare l'analisi con il supporto di alcune metodologie grafiche per comprendere, in un primo momento, il comportamento e i legami che definiscono le diverse variabili all'interno del dataset di riferimento. Si considera come approccio grafico il "Grafico a stella".

Il grafico a stella è uno strumento di visualizzazione multivariata che rappresenta ciascuna osservazione (in questo caso, ciascun modello di auto) come una stella, con ciascun raggio che rappresenta una delle variabili del dataset. Questo tipo di grafico è utile per confrontare rapidamente le caratteristiche di diverse osservazioni.

Visualizzazione delle variabili mtcars



Struttura del Grafico a Stella

- **Ogni Stella:** Ogni stella rappresenta un modello di automobile.
- **Raggi della Stella:** Ogni raggio rappresenta una variabile. La lunghezza di ogni raggio è proporzionale al valore della variabile per quel modello di auto.
- **Colori:** I colori distinti per ogni raggio aiutano a distinguere le variabili e a visualizzare meglio le differenze tra i modelli.

Esempi di Interpretazione

- **Mazda RX4:** Osservando la stella della Mazda RX4, notiamo che ha valori medi per mpg, hp, e wt, indicando una efficienza del carburante e una potenza del motore moderata, con un peso dell'auto nella media.
- **Datsun 710:** Ha una stella con un raggio lungo per mpg e qsec, indicando un'alta efficienza del carburante e un buon tempo sul quarto di miglio, con un basso numero di cilindri (cyl) e cavalli vapore (hp).
- **Hornet Sportabout:** Mostra un raggio lungo per hp e disp, indicando una alta potenza del motore e cilindrata, ma con un mpg più corto, suggerendo una minore efficienza del carburante.

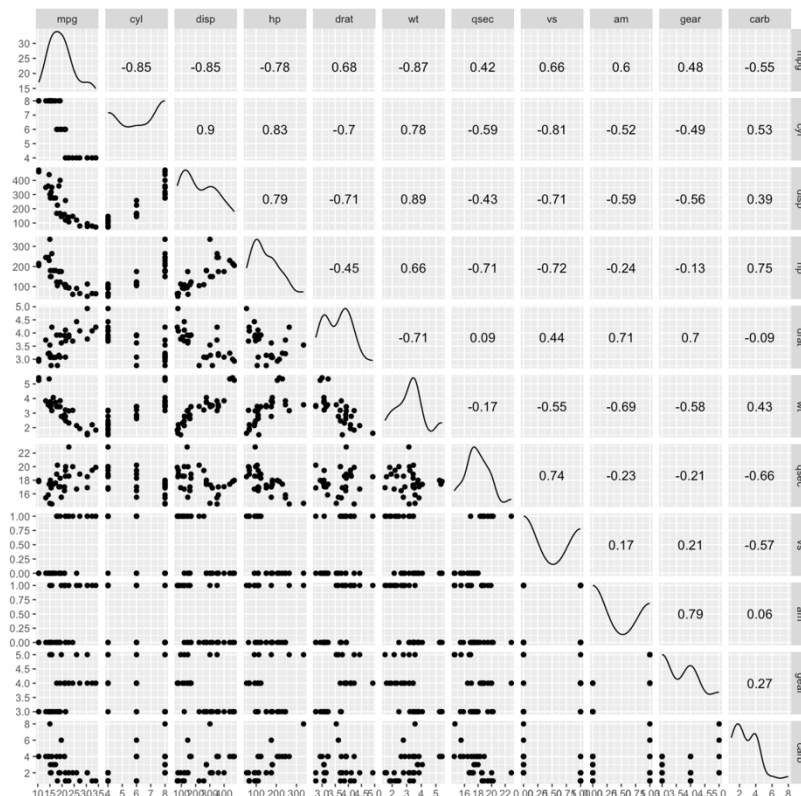
Confronto tra Modelli

Il grafico a stella permette di confrontare facilmente diversi modelli di auto:

- **Efficienza del Carburante:** Confrontando i raggi `mpg` tra le stelle, possiamo vedere quali modelli hanno una maggiore efficienza del carburante.
- **Potenza del Motore:** I raggi `hp` mostrano rapidamente quali modelli hanno una potenza del motore più elevata.
- **Peso dell'Auto:** I raggi `wt` ci permettono di visualizzare quali modelli di auto sono più pesanti o più leggeri.

MATRICE DI CORRELAZIONE E INDICI

È opportuno considerare se esista una correlazione tra queste variabili al fine di fornire un'analisi più approfondita del dataset. Si procederà con l'analisi della “Matrice degli scatterplot”.



Spiegazione Dettagliata della Matrice degli Scatterplot

La matrice degli scatterplot è uno strumento visivo utile per esaminare le relazioni bivariate tra le variabili in un dataset. Nell'output fornito, ogni cella della matrice rappresenta un grafico di dispersione (scatterplot) tra due variabili specifiche del dataset `mtcars`.

Struttura della Matrice degli Scatterplot

- **Diagonale Principale:** Contiene gli istogrammi delle distribuzioni delle singole variabili. Gli istogrammi mostrano la frequenza dei valori delle variabili.
- **Sopra la Diagonale:** Ogni cella contiene il coefficiente di correlazione tra le due variabili corrispondenti. Questo valore fornisce una misura quantitativa della forza e della direzione della relazione lineare tra le variabili.
- **Sotto la Diagonale:** Ogni cella contiene un grafico di dispersione che mostra la relazione tra due variabili specifiche. I punti del grafico rappresentano le osservazioni del dataset.

Analisi delle Relazioni tra le Variabili

1. mpg (Miles per Gallon):

- **mpg vs. cyl:** Coefficiente di correlazione -0.85, indicando una forte correlazione negativa. Le auto con più cilindri tendono ad avere un'efficienza del carburante inferiore.
- **mpg vs. disp:** Coefficiente di correlazione -0.85, indicando una forte correlazione negativa. Le auto con maggiore cilindrata tendono ad avere un'efficienza del carburante inferiore.
- **mpg vs. hp:** Coefficiente di correlazione -0.78, indicando una forte correlazione negativa. Le auto con maggiore potenza tendono ad avere un'efficienza del carburante inferiore.
- **mpg vs. wt:** Coefficiente di correlazione -0.87, indicando una forte correlazione negativa. Le auto più pesanti tendono ad avere un'efficienza del carburante inferiore.

2. cyl (Numero di Cilindri):

- **cyl vs. disp:** Coefficiente di correlazione 0.90, indicando una forte correlazione positiva. Le auto con più cilindri tendono ad avere una cilindrata maggiore.
- **cyl vs. hp:** Coefficiente di correlazione 0.83, indicando una forte correlazione positiva. Le auto con più cilindri tendono ad avere una potenza maggiore.

3. disp (Cilindrata):

- **disp vs. hp:** Coefficiente di correlazione 0.79, indicando una forte correlazione positiva. Le auto con maggiore cilindrata tendono ad avere una potenza maggiore.
- **disp vs. wt:** Coefficiente di correlazione 0.89, indicando una forte correlazione positiva. Le auto con maggiore cilindrata tendono ad essere più pesanti.

4. hp (Cavalli Vapore):

- **hp vs. wt:** Coefficiente di correlazione 0.66, indicando una moderata correlazione positiva. Le auto con maggiore potenza tendono ad essere più pesanti.

5. drat (Rapporto al Ponte Posteriore):

- **drat vs. mpg:** Coefficiente di correlazione 0.68, indicando una moderata correlazione positiva. Le auto con un maggiore rapporto al ponte posteriore tendono ad avere una migliore efficienza del carburante.

6. wt (Peso dell'Auto):

- **wt vs. qsec:** Coefficiente di correlazione 0.09, indicando una bassa correlazione positiva. Il peso dell'auto ha una bassa correlazione con il tempo sul quarto di miglio.

7. qsec (Tempo sul Quarto di Miglio):

- **qsec vs. am:** Coefficiente di correlazione 0.71, indicando una forte correlazione positiva. Le auto con trasmissione manuale tendono ad avere tempi migliori sul quarto di miglio.

Conclusioni: Le forti correlazioni positive e negative sono particolarmente utili per comprendere come le caratteristiche delle auto influenzano le prestazioni e l'efficienza. Ad esempio, sappiamo che l'efficienza del carburante (mpg) è fortemente influenzata dal numero di cilindri (cyl), dalla cilindrata (disp), dalla potenza del motore (hp), e dal peso dell'auto (wt).

Dopo aver confermato la correlazione tra le variabili analizzate si procede con ulteriori test per valutare l'applicabilità dell'Analisi fattoriale. Si utilizza il criterio di KMO (Kaiser-Meyer-Olkin) e successivamente si effettua anche il test di sfericità Bartlett.

Il **Criterio di Kaiser-Meyer-Olkin (KMO)** è utilizzato per valutare l'adeguatezza dei dati per l'analisi fattoriale. L'indice KMO varia da 0 a 1, con valori più alti che indicano che l'analisi fattoriale è appropriata. In generale, un valore KMO superiore a 0.6 è considerato accettabile.

Output del Test KMO

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = mtcars)

Overall MSA = 0.83

MSA for each item =

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0.93	0.90	0.76	0.84	0.95	0.74	0.74	0.91	0.88	0.85	0.62

L'output mostra due informazioni principali:

1. Overall MSA (Measure of Sampling Adequacy):

Un valore globale che indica l'adeguatezza complessiva del campione per l'analisi fattoriale.

2. **MSA per ciascuna variabile:** Valori specifici per ciascuna variabile che indicano quanto ciascuna variabile è adeguata all'analisi fattoriale.

Interpretazione dell'Output

1. Overall MSA = 0.83:

- Questo valore indica che il campione complessivo è molto adeguato per l'analisi fattoriale. Valori superiori a 0.8 sono considerati molto buoni, e 0.83 è un'ottima indicazione che i dati sono adatti per procedere con l'analisi fattoriale.

2. MSA per ciascuna variabile:

- **mpg: 0.93** - Molto adeguato.
- **cyl: 0.90** - Molto adeguato.
- **disp: 0.76** - Adeguato.
- **hp: 0.84** - Molto adeguato.
- **drat: 0.95** - Molto adeguato.
- **wt: 0.74** - Adeguato.
- **qsec: 0.74** - Adeguato.
- **vs: 0.91** - Molto adeguato.
- **am: 0.88** - Molto adeguato.
- **gear: 0.85** - Molto adeguato.
- **carb: 0.62** - Accettabile.

Il test di sfericità Bartlett:

Il test di sfericità di Bartlett è utilizzato per verificare che la matrice di correlazione è pari ad una matrice identità, indicando l'assenza di correlazione tra le variabili. Se il test rifiuta l'ipotesi nulla H_0 , significa che le variabili nel dataset sono correlati, giustificando l'uso del modello di analisi fattoriale per descrivere la struttura dei dati.

Output del Test di Sfericità di Bartlett:

L'output fornisce tre informazioni principali:

1. **Chi-Square Test Statistic (\$chisq):** La statistica del test chi-quadro.
2. **p-value (\$p.value):** Il valore p associato al test chi-quadro.
3. **Degrees of Freedom (\$df):** I gradi di libertà del test.

Interpretazione dell'Output

\$chisq

[1] 408.0116

\$p.value

[1] 2.226927e-55

\$df

[1] 55

1. Chi-Square Test Statistic (\$chisq = 408.0116):

○ La statistica del test chi-quadro è 408.0116. Questo valore è confrontato con una distribuzione chi-quadro con 55 gradi di libertà per determinare il valore p.

2. p-value (\$p.value = 2.226927e-55):

○ Il valore p è estremamente piccolo (2.226927e-55), molto inferiore a qualsiasi livello di significatività comunemente usato (ad esempio, 0.05). Un valore p così basso indica che l'ipotesi nulla può essere rifiutata con molta certezza.

3. Degrees of Freedom (\$df = 55):

- I gradi di libertà del test sono 55. Questo è calcolato come $\frac{q(q-1)}{2}$, dove q è il numero di variabili (qui 11 variabili).

Conclusioni

- **Rifiuto dell'Ipotesi Nulla:** Il p-value estremamente basso indica che esistono correlazioni significative tra le variabili del dataset, rendendole adatte per l'analisi fattoriale.

Scelta del numero di fattori

Il numero di fattori su cui effettuare l'analisi è una scelta che va fatta in modo molto accurato. Questa viene effettuata attraverso la valutazione di un test e del suo relativo p-value, contenente sotto l'ipotesi nulla l'affermazione che k fattori siano sufficienti a spiegare le informazioni presenti nel dataset.

nf = 1	nf = 2	nf = 3	nf = 4	nf = 5	nf = 6
1.496220e-17	4.047510e-04	2.051923e-01	8.181113e-01	9.785961e-01	9.341394e-01

L'output fornito mostra i p-value associati ai test di significatività per un numero crescente di fattori, da 1 a 6.

Interpretazione del Test

L'ipotesi nulla (H_0) del test è che il numero specificato di fattori sia sufficiente per spiegare le informazioni presenti nel dataset. Il p-value deve essere maggiore di un livello di significatività (pari ad 0.05), per accettare l'ipotesi nulla.

1. nf = 1:

- p-value = 1.496220e-17
- Il p-value è estremamente basso, indicando che un singolo fattore non è sufficiente.

2. nf = 2:

- p-value = 4.047510e-04
- Il p-value è ancora molto basso, indicando che due fattori non sono sufficienti.

3. nf = 3:

- p-value = 2.051923e-01
- Il p-value è 0.205, che è superiore a 0.05. Questo suggerisce che tre fattori potrebbero essere sufficienti per spiegare la varianza osservata.

4. nf = 4:

- p-value = 8.181113e-01
- Il p-value è 0.818, molto superiore a 0.05. Questo indica che quattro fattori sono sufficienti.

5. **nf = 5:**

- p-value = 9.785961e-01
- Il p-value è 0.979, suggerendo che cinque fattori sono più che sufficienti.

6. **nf = 6:**

- p-value = 9.341394e-01
- Il p-value è 0.934, indicando che sei fattori sono sufficienti.

Conclusioni

- **Numero Sufficiente di Fattori:** Poiché il p-value diventa maggiore di 0.05 per la prima volta con tre fattori (nf = 3, p-value = 0.205), possiamo concludere che tre fattori sono sufficienti per spiegare la varianza osservata nel dataset `mtcars`.

Concludiamo che tre fattori siano sufficienti per spiegare le informazioni contenuti nel dataset.

Possiamo procedere con l'analisi fattoriale esplorativa utilizzando tre fattori (k=3). L'output ottenuto:

```
Call:
factanal(x = mtcars, factors = 3)

Uniquenesses:
  mpg   cyl  disp    hp  drat    wt  qsec    vs  am  gear  carb
0.135 0.055 0.090 0.127 0.290 0.060 0.051 0.223 0.208 0.125 0.158

Loadings:
      Factor1 Factor2 Factor3
mpg   0.643   -0.478   -0.473
cyl  -0.618    0.703    0.261
disp -0.719    0.537    0.323
hp   -0.291    0.725    0.513
drat  0.804   -0.241
wt   -0.778    0.248    0.524
qsec -0.177   -0.946   -0.151
vs    0.295   -0.805   -0.204
am    0.880
gear  0.908          0.224
carb  0.114    0.559    0.719

      Factor1 Factor2 Factor3
SS loadings   4.380   3.520   1.578
Proportion Var 0.398   0.320   0.143
Cumulative Var 0.398   0.718   0.862

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 30.53 on 25 degrees of freedom.
The p-value is 0.205
```

L'output ottenuto dal software è suddiviso in 4 parti:

1) **Uniquenesses:** Le unicità rappresentano la proporzione della varianza di ciascuna variabile che non è spiegata dai fattori estratti. Nell'analisi in questione tali quote risultano essere abbastanza basse, il che è indice del fatto che i fattori riescono a spiegare una parte preponderante di informazioni del dataset.

2) Loadings: I pesi fattoriali (loadings) mostrano quanto ciascuna variabile manifesta dipenda dalle variabili latenti

3) SS Loadings, Proportion Var, Cumulative Var:

- **Varianza cumulativa** spiegata dai 3 fattori, è 86.2%.

4) Test sull'ipotesi: Testa se il numero di fattori è sufficiente per spiegare la varianza osservata.

- **Chi-Square Statistic:** 30.53
- **Degrees of Freedom (df):** 25
- **p-value:** 0.205

Conclusioni

L'analisi fattoriale esplorativa con 3 fattori sul dataset `mtcars` ha prodotto risultati significativi che spiegano l'86.2% della varianza totale. I fattori identificati sono interpretabili e rappresentano:

1. **Factor 1:** Caratteristiche della trasmissione e delle marce.
2. **Factor 2:** Potenza del motore e configurazione del motore.
3. **Factor 3:** Efficienza del carburante e cilindrata.

Il test dell'ipotesi conferma che 3 fattori sono sufficienti per spiegare la varianza osservata. Pertanto, possiamo concludere che l'analisi fattoriale esplorativa con 3 fattori fornisce un modello adeguato e interpretabile per il dataset `mtcars`.

Comunalità, Unicità e Matrice dei Residui

Si procede con il calcolo delle comunalità. Le comunalità sono date dalla somma dei pesi fattoriali elevati al quadrato. Rappresentano la quota di varianza di ciascuna variabile che è spiegata dai fattori comuni. L'output fornisce le unicità e le comunalità per ciascuna variabile del dataset `mtcars`.

```
> # Calcolo delle comunalità
> mtcars.f$uniquenesses
      mpg      cyl      disp      hp      drat      wt      qsec      vs      am      gear      carb
0.13493750 0.05548969 0.08978442 0.12678036 0.28999718 0.05958649 0.05146955 0.22337819 0.20838125 0.12474069 0.15788615
> apply(mtcars.f$loadings^2, 1, sum) # Communalities
      mpg      cyl      disp      hp      drat      wt      qsec      vs      am      gear      carb
0.8650617 0.9445105 0.9102147 0.8732196 0.7100058 0.9404137 0.9485306 0.7766178 0.7916147 0.8752584 0.8421159
```

Le *Unicità* rappresentano la proporzione della varianza di ciascuna variabile che non è spiegata dai fattori estratti. Valori più bassi indicano che una maggiore proporzione della varianza è spiegata dai fattori comuni:

- **mpg:** 0.1349 (il 13.49% della varianza di `mpg` non è spiegata dai fattori).
- **cyl:** 0.0555 (il 5.55% della varianza di `cyl` non è spiegata dai fattori).
- **disp:** 0.0898 (l'8.98% della varianza di `disp` non è spiegata dai fattori).
- ...e così via per le altre variabili.

Comunalità: Valori più alti indicano che una maggiore proporzione della varianza è spiegata dai fattori comuni:

- **mpg:** 0.8651 (l'86.51% della varianza di `mpg` è spiegata dai fattori).
- **cyl:** 0.9445 (il 94.45% della varianza di `cyl` è spiegata dai fattori).
- **disp:** 0.9102 (il 91.02% della varianza di `disp` è spiegata dai fattori).
- ...e così via per le altre variabili.

Questi risultati indicano che l'analisi fattoriale con tre fattori è appropriata e fornisce una buona rappresentazione dei dati.

Matrice dei Residui

Si procede con la stima della matrice Sigma. Facendo la differenza tra la matrice di correlazione e la matrice stimata Sigma, viene calcolata la matrice dei residui. La matrice dei residui, mostra quanto i fattori riescono a spiegare la correlazione tra le variabili. Valori residui piccoli (vicini a zero) indicano che i fattori estratti spiegano bene le relazioni tra le variabili.

Output della Matrice dei Residui

L'output della matrice dei residui è presentato come una tabella dove ogni cella rappresenta il residuo tra una coppia di variabili:

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	0.000001	0.004632	0.024298	0.000024	0.016494	-0.001781	0.008884	-0.007074	0.032488	0.012696	-0.017277
cyl	0.004632	0.000000	-0.004436	0.008927	-0.015234	-0.009277	0.003690	-0.009197	-0.016389	-0.004687	0.017030
disp	0.024298	-0.004436	0.000001	0.026432	0.020070	0.026066	-0.004504	0.000015	0.024595	0.013836	-0.055424
hp	0.000024	0.008927	0.026432	0.000000	-0.004473	-0.016163	0.003085	0.051062	-0.003302	0.008641	0.009363
drat	0.016494	-0.015234	0.020070	-0.004473	-0.000003	0.008780	-0.004908	-0.005241	0.019433	-0.010468	0.001389
wt	-0.001781	-0.009277	0.026066	-0.016163	0.008780	0.000000	0.001342	-0.018700	0.018725	0.000246	0.001241
qsec	0.008884	0.003690	-0.004504	0.003085	-0.004908	0.001342	0.000000	0.004744	-0.004371	0.001719	0.000848
vs	-0.007074	-0.009197	0.000015	0.051062	-0.005241	-0.018700	0.004744	0.000004	-0.039217	0.000820	-0.006596
am	0.032488	-0.016389	0.024595	-0.003302	0.019433	0.018725	-0.004371	-0.039217	0.000004	0.013459	-0.025740
gear	0.012696	-0.004687	0.013836	0.008641	-0.010468	0.000246	0.001719	0.000820	0.013459	0.000001	-0.002446
carb	-0.017277	0.017030	-0.055424	0.009363	0.001389	0.001241	0.000848	-0.006596	-0.025740	-0.002446	-0.000002

Interpretazione della Matrice dei Residui

- **Valori Residui Piccoli:** Un valore residuo vicino a zero indica che i fattori estratti spiegano bene la correlazione tra le due variabili. Ad esempio, il residuo tra `mpg` e `hp` è 0.000024, che è molto vicino a zero, indicando che la correlazione tra `mpg` e `hp` è ben spiegata dai fattori estratti.
- **Valori Residui Più Grandi:** Un valore residuo lontano da zero indica che i fattori estratti non spiegano bene la correlazione tra le due variabili. Ad esempio, il residuo tra `disp` e `carb` è -0.055424, suggerendo che la correlazione tra `disp` e `carb` non è ben spiegata dai fattori estratti.

Conclusioni

- **Qualità del Modello:** In generale, la matrice dei residui mostra valori piccoli, suggerendo che il modello fattoriale con 3 fattori spiega bene la maggior parte delle correlazioni tra le variabili. Ciò conferma la validità del modello fattoriale estratto.

Analisi fattoriale senza rotazione

L'analisi fattoriale è una metodologia di analisi che non propone un'unica soluzione, in quanto, applicando una rotazione ortogonale dei fattori, si ottiene una soluzione ancora valida per il modello a k fattori. La rotazione dei fattori è una tecnica utilizzata nell'analisi fattoriale per semplificare e facilitare l'interpretazione dei risultati.

Tipi di Rotazione

1. **Rotazione Ortogonale:** Dopo aver effettuato la rotazione, i fattori non saranno più legati tra loro, quindi non vi sarà ridondanza di informazioni.
 - *Varimax*
 - *Quartimax*
2. **Rotazione Obliqua:** Permette ai fattori di essere correlati.
 - *Promax*
 - *Oblimin*

Si andranno a confrontare le diverse soluzioni tenendo conto dell'ausilio delle rotazioni, così da decidere quale sarà il risultato migliore.

L'output ottenuto dall'analisi fattoriale senza rotazione è il seguente:

```
Call:
factanal(x = mtcars, factors = 3, rotation = "none")

Uniquenesses:
   mpg   cyl  disp    hp  drat    wt   qsec    vs    am  gear  carb
0.135 0.055 0.090 0.127 0.290 0.060 0.051 0.223 0.208 0.125 0.158

Loadings:
      Factor1 Factor2 Factor3
mpg  -0.910   0.137  -0.136
cyl   0.962         -0.135
disp  0.937  -0.174
hp    0.875   0.292   0.147
drat  -0.689   0.453   0.175
wt    0.858  -0.382   0.242
qsec  -0.591  -0.754   0.177
vs    -0.809  -0.309   0.164
am    -0.522   0.719
gear  -0.459   0.729   0.365
carb   0.594   0.517   0.471

      Factor1 Factor2 Factor3
SS loadings    6.448    2.465    0.565
Proportion Var  0.586    0.224    0.051
Cumulative Var  0.586    0.810    0.862

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 30.53 on 25 degrees of freedom.
The p-value is 0.205
```

Conclusioni

L'analisi fattoriale esplorativa senza rotazione con 3 fattori sul dataset `mtcars` ha fornito un modello robusto e interpretabile che spiega l'86.2% della varianza cumulata.

Sebbene l'analisi senza rotazione fornisca una buona comprensione delle strutture latenti, l'applicazione di tecniche di rotazione (come Varimax o Promax) potrebbe migliorare ulteriormente l'interpretabilità dei fattori.

Analisi fattoriale con rotazione Varimax

L'output ottenuto è:

```
Call:
factanal(x = mtcars, factors = 3, rotation = "varimax")

Uniquenesses:
  mpg  cyl  disp   hp  drat   wt  qsec   vs   am  gear  carb
0.135 0.055 0.090 0.127 0.290 0.060 0.051 0.223 0.208 0.125 0.158

Loadings:
      Factor1 Factor2 Factor3
mpg   0.643  -0.478  -0.473
cyl  -0.618   0.703   0.261
disp -0.719   0.537   0.323
hp   -0.291   0.725   0.513
drat  0.804  -0.241
wt   -0.778   0.248   0.524
qsec -0.177  -0.946  -0.151
vs    0.295  -0.805  -0.204
am    0.880
gear  0.908           0.224
carb  0.114   0.559   0.719

      Factor1 Factor2 Factor3
SS loadings    4.380    3.520    1.578
Proportion Var  0.398    0.320    0.143
Cumulative Var  0.398    0.718    0.862

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 30.53 on 25 degrees of freedom.
The p-value is 0.205
```

L'analisi fattoriale esplorativa con rotazione Varimax sul dataset `mtcars` ha prodotto risultati che spiegano l'86.2% della varianza cumulata con tre fattori (la varianza cumulata spiegata con questo tipo di rotazione rimane sempre uguale) ed un p-value pari a 0.205, quindi si potrebbe accettare questa rotazione dei fattori.

Interpretazione dei Fattori

- Factor 1:** Caratteristiche della trasmissione e dell'efficienza del carburante.
 - Variabili fortemente correlate: `mpg`, `drat`, `wt`, `am`, `gear`.
 - Implicazioni:** Le caratteristiche della trasmissione e l'efficienza del carburante sono interconnesse, suggerendo che migliorare una di queste caratteristiche può influenzare positivamente le altre. Ad esempio, ridurre il peso del veicolo potrebbe migliorare l'efficienza del carburante e le prestazioni generali della trasmissione.
- Factor 2:** Potenza del motore e prestazioni.
 - Variabili fortemente correlate: `cyl`, `disp`, `hp`, `qsec`.

- **Implicazioni:** Questo fattore suggerisce che aumentare la potenza del motore (attraverso un maggiore numero di cilindri o una maggiore cilindrata) può migliorare le prestazioni complessive del veicolo, riducendo i tempi sul quarto di miglio. Tuttavia, potrebbe anche influire negativamente su altre caratteristiche come l'efficienza del carburante.
3. **Factor 3:** Configurazione del motore e distribuzione del peso.
- Variabili fortemente correlate: mpg, hp, wt, carb.
 - **Implicazioni:** Questo fattore sottolinea il compromesso tra potenza del motore e efficienza del carburante. Veicoli con motori potenti e pesanti tendono ad essere meno efficienti dal punto di vista del carburante. Questo suggerisce che migliorare l'efficienza del carburante potrebbe richiedere compromessi in termini di potenza e configurazione del motore.

Analisi fattoriale con rotazione Promax

L'output ottenuto è:

```
Call:
factanal(x = mtcars, factors = 3, rotation = "promax")

Uniquenesses:
   mpg   cyl  disp    hp  drat    wt   qsec    vs    am  gear  carb
0.135 0.055 0.090 0.127 0.290 0.060 0.051 0.223 0.208 0.125 0.158

Loadings:
      Factor1 Factor2 Factor3
mpg   0.539   0.156  -0.431
cyl  -0.525  -0.576
disp -0.646  -0.317   0.194
hp   -0.112  -0.481   0.475
drat  0.834   0.121
wt   -0.709   0.177   0.544
qsec -0.357   1.066
vs    0.169   0.787
am    0.958  -0.306
gear  1.045           0.450
carb  0.352  -0.202   0.872

      Factor1 Factor2 Factor3
SS loadings    4.484    2.625    1.723
Proportion Var  0.408    0.239    0.157
Cumulative Var  0.408    0.646    0.803

Factor Correlations:
      Factor1 Factor2 Factor3
Factor1  1.000  -0.387   0.691
Factor2 -0.387  1.000  -0.443
Factor3  0.691  -0.443  1.000

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 30.53 on 25 degrees of freedom.
The p-value is 0.205
```

La rotazione Promax permette ai fattori comuni di essere correlati tra loro, abbandonando l'ipotesi di incorrelazione. Questo riflette la realtà che le caratteristiche automobilistiche non sono completamente indipendenti.

Conclusioni

L'analisi fattoriale esplorativa con rotazione Promax sul dataset mtcars ha prodotto risultati che spiegano l'80.3% della varianza cumulata con tre fattori, che risulta essere inferiore in corrispondenza degli altri casi.

Scelta della Rotazione: Promax

Motivo della Scelta

Correlazione tra Fattori:

La rotazione Promax consente ai fattori di essere correlati tra loro, riflettendo meglio la realtà delle caratteristiche automobilistiche. Questo è evidente nel dataset mtcars, dove le caratteristiche del

veicolo tendono a influenzarsi reciprocamente. La correlazione tra i fattori fornisce una visione più realistica delle interrelazioni tra le variabili.

Essa facilita l'identificazione di gruppi distinti di variabili che caricano fortemente su ciascun fattore, permettendo di cogliere meglio le relazioni complesse nel dataset.

Varianza Spiegata:

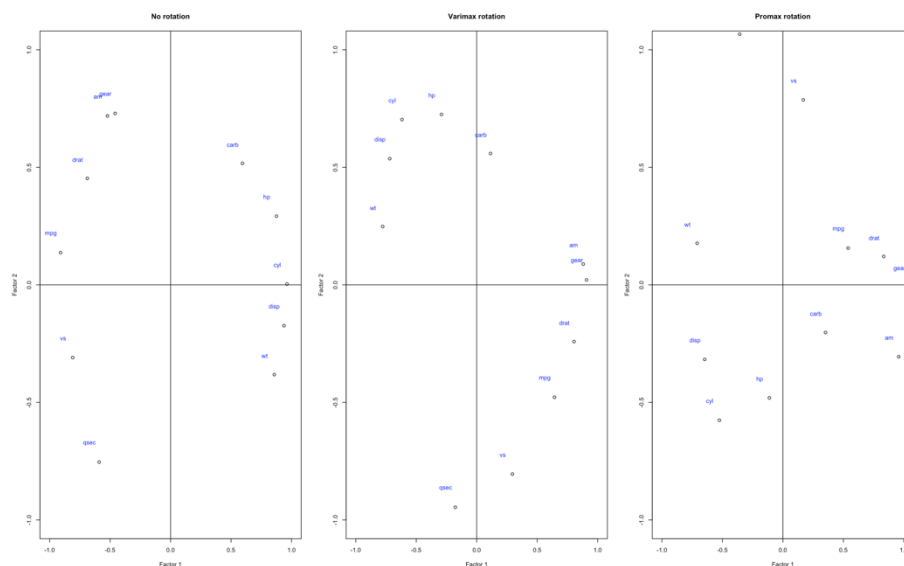
Sebbene la rotazione Varimax spieghi una percentuale leggermente superiore della varianza totale (86.2% rispetto all'80.3% della Promax), la differenza non è significativa. La maggiore capacità interpretativa e il realismo della rotazione Promax compensano ampiamente questa lieve differenza.

Conclusione

Rotazione Promax è la scelta migliore per l'analisi fattoriale del dataset mtcars perché riflette meglio la realtà delle interrelazioni tra le variabili, facilitando una comprensione più accurata e dettagliata delle caratteristiche automobilistiche.

SCATTERPLOT

Per confermare o implementare le interpretazioni ottenute, i pesi fattoriali vengono rappresentati su dei grafici appositi, chiamati scatterplot. La rappresentazione dei pesi fattoriali nel piano cartesiano consente di visualizzare come le variabili caricano su ciascun fattore. Questo aiuta a comprendere meglio le relazioni tra le variabili e i fattori.



Qui di seguito è riportata la spiegazione dettagliata degli scatterplot per le tre rotazioni:

- senza rotazione
- rotazione Varimax
- rotazione Promax.

Scatterplot Senza Rotazione

Interpretazione:

- Senza rotazione, l'interpretazione dei fattori può risultare meno chiara poiché le variabili possono caricare su più fattori in modo confuso.

Scatterplot con Rotazione Varimax

Interpretazione:

La rotazione Varimax semplifica l'interpretazione, rendendo chiaro quali variabili caricano fortemente su ciascun fattore. È ideale per identificare gruppi distinti di variabili che sono fortemente associate a singoli fattori.

Scatterplot con Rotazione Promax

Interpretazione:

La rotazione Promax riflette meglio la realtà delle relazioni tra le variabili, mostrando che alcune variabili sono correlate con più fattori. Questo fornisce una visione più dettagliata e realistica delle interrelazioni tra le variabili.

Conclusione dell'Analisi Fattoriale Esplorativa sul Dataset mtcars

L'analisi fattoriale esplorativa (EFA) condotta sul dataset mtcars ha fornito una visione approfondita delle relazioni tra le variabili del dataset, permettendo di identificare e interpretare le strutture latenti che influenzano le caratteristiche delle automobili.

Sintesi dei Risultati

1. Scelta del Numero di Fattori:

- L'analisi ha determinato che tre fattori sono sufficienti per spiegare la varianza osservata nel dataset. Il p-value per il test con tre fattori è risultato essere 0.205, superiore alla soglia di significatività di 0.05, indicando che tre fattori sono adeguati.

2. Rotazioni dei Fattori:

Rotazione Promax: Ha permesso di catturare meglio le interrelazioni tra le variabili, riflettendo la realtà delle caratteristiche automobilistiche. D'altro canto, ha mostrato una distribuzione dei pesi più diffusa, permettendo ai fattori di essere correlati tra loro. Questo ha facilitato una comprensione più realistica delle relazioni tra le variabili.

Risultati Ottenuti:

1. Riduzione della Dimensionalità:

- È stato possibile ridurre le 11 variabili originali in tre variabili latenti principali. Questo ha facilitato l'interpretazione delle relazioni tra le variabili, semplificando il modello e mantenendo una quantità significativa di informazione.

2. Identificazione dei Fattori Latenti:

- L'analisi ha rivelato che le variabili possono essere raggruppate in fattori che rappresentano concetti sottostanti come l'efficienza del carburante, le prestazioni del motore e la configurazione del motore e del peso del veicolo.

3. Interpretazione dei Fattori:

- I tre fattori identificati sono stati chiaramente interpretati in termini di caratteristiche delle automobili, come le prestazioni del motore e l'efficienza del carburante. Questo ha fornito un quadro significativo delle interrelazioni tra le variabili.