

GRUPPO QUATTRO RUOTE



Alessandro Tondo 19632
Laura Lavazza 19758
Matteo Scordo 19813
Alessandro Giosa 19894

3.1 Analisi fattoriale (continua)

		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Optionals e velocità	imp_stereo	0.80717
	cerchi_lega	0.75698
	speed_1	0.72524
	tett_apr	0.67617	.	0.51080
	sedili_pelle	0.59835	0.47713
Affidabilità	amp_bag_1	.	0.67939
	sicurezza_1	.	0.67109
	aff_post_1	.	0.65183
	aff_extra_1	0.54423	0.55578	-0.35540
Optionals tecnologici	navigatore	.	.	0.76147
	com_vol	0.47876	.	0.67509
	bluetooth	0.44647	.	0.55573	.	.	0.35428	.
	ruota_scort	.	0.42156	0.47475
Intensità utilizzo	ore_guida_lav	.	.	.	0.88569	.	.	.
	ore_guida_nlav	.	.	.	0.84538	.	.	.
Praticità	dim_rid_1	0.76303	.	.
	cons_rid_1	0.66147	.	.
	risp_amb_1	0.42650	.	.	.	0.50165	.	.
Caratteristiche strutturali	comf_int_1	0.87646	.
	design_1	0.55074	0.69326	.
Sensori di parcheggio	sens_park	0.81650

Confrontando dal punto di vista interpretativo le soluzioni a 6 e 7 fattori, abbiamo rilevato che quella a sette fattori appare decisamente più significativa.

In particolare abbiamo interpretato i fattori ottenuti nel seguente modo:



Factor1	Optionals e velocità
Factor2	Affidabilità
Factor3	Optionals tecnologici
Factor4	Intensità utilizzo
Factor5	Praticità
Factor6	Caratteristiche strutturali
Factor7	Sensori di parcheggio

3. Regressione lineare (continua)

La procedura REG
Modello: MODEL1
Variabile dipendente: spesa_max spesa_max

Numero osservazioni lette	200
Numero osservazioni usate	199
Numero osservazioni con valori mancanti	1

La procedura REG
Modello: MODEL1
Variabile dipendente: spesa_max spesa_max

Numero osservazioni lette	200
Numero osservazioni usate	199
Numero osservazioni con valori mancanti	1

Analisi della varianza					
Origine	DF	Somma dei quadrati	Media quadratica	Valore F	Pr > F
Modello	3	1723858315	574619438	14.28	<.0001
Errore	195	7844350228	40227437		
Totale corretto	198	9568208543			

Radice dell'MSE	6342.51031	R-quadro	0.1802
Media dip.	12284	R-quadro corr	0.1676
Var coeff	51.63263		

Stime dei parametri							
Variabile	Etichetta	DF	Stima dei parametri	Errore standard	Valore t	Pr > t	Stima standardizzata
Intercept	Intercept	1	12284	449.60864	27.32	<.0001	0
Factor4		1	1446.57850	450.74258	3.21	0.0016	0.20809
Factor5		1	-1883.94495	450.74258	-4.14	<.0001	-0.26813
Factor6		1	1771.85637	450.74258	3.93	0.0001	0.25489

- Dopo aver lanciato la procedura proc reg, utilizzando come input i 7 fattori ottenuti con la proc factor, abbiamo verificato che un'osservazione non è stata utilizzata, quindi esse non sono più 200 ma 199, quindi vi è un valore missing. Dalla procedura stepwise (usando come slentry=0,05 e slstay=0,05) risultano significativi solo 3 fattori: factor4, factor5 e factor6.
- Per quanto riguarda il **Test F** possiamo verificare che il modello ha capacità esplicativa, il P-value associato al test F è < 0.05 (<0.0001).
- Per quanto riguarda il coefficiente di determinazione **R-quadro** per valutare la capacità esplicativa del modello possiamo dire che le nostre variabili non sono in grado di spiegare in modo soddisfacente la relazione tra la variabile dipendente e i regressori (R-quadro 0.1802).
- Siamo ricorsi al **Test t** per valutare la significatività dei singoli fattori. In questo caso il P-value del test è molto piccolo, quindi tutti i fattori rimasti sono rilevanti per spiegare la variabile dipendente. (Pr > t)

3. Regressione lineare (continua)

A questo punto ci siamo focalizzati sulle **osservazioni influenti** e abbiamo preso in considerazione la **Distanza di Cook** e il **Leverage H**. Abbiamo così individuato ed eliminato 1 osservazione influente.

Il valore di soglia del Leverage utilizzato è 0.22, considerando la formula $H > 2 * (p+1) / n$, dove **p** è il numero di regressori ed **n** il numero di osservazioni.

```
NOTE: There were 199 observations read from the data set CORSO.INFLUENCE.  
      WHERE (cook<=1) or (leverage<=0.22);  
NOTE: The data set CORSO.FACTORS_NEW has 199 observations and 66 variables.  
NOTE: DATA statement used (Total process time):  
      real time           0.19 seconds  
      user cpu time       0.01 seconds  
      system cpu time     0.01 seconds  
      memory              941.31k  
      OS Memory           30112.00k  
      Timestamp           10/01/2015 10:16:04 m.  
      Step Count          41  Switch Count  41  
      Page Faults         0  
      Page Reclaims       153  
      Page Swaps          0  
      Voluntary Context Switches 343  
      Involuntary Context Switches 82  
      Block Input Operations 32  
      Block Output Operations 264
```

Precedentemente le osservazioni erano 200

3. Regressione lineare (continua)

Dopo aver eliminato le osservazioni influenti, abbiamo rieseguito la regressione lineare. Anche in questo caso un'osservazione non viene utilizzata, quindi non sono più 199 ma 198 (ancora un missing). Dalla procedura stepwise (usando come slentry=0,05 e slstay=0,05) risultano significativi solo 2 fattori : factor5 e factor6. Durante l'analisi non abbiamo utilizzato l'opzione VIF poiché come regressori della proc reg abbiamo utilizzato solamente i fattori derivanti dall'analisi fattoriale. Per costruzione non può esserci multicollinearità.

La procedura REG
Modello: MODEL1
Variabile dipendente: spesa_max spesa_max

Numero osservazioni lette	199
Numero osservazioni usate	198
Numero osservazioni con valori mancanti	1

Selezione stepwise: Passo 1

La procedura REG
Modello: MODEL1
Variabile dipendente: spesa_max spesa_max

Numero osservazioni lette	199
Numero osservazioni usate	198
Numero osservazioni con valori mancanti	1

Analisi della varianza

Origine	DF	Somma dei quadrati	Media quadratica	Valore F	Pr > F
Modello	2	1332901348	666450674	19.10	<.0001
Errore	195	6805620117	34900816		
Totale corretto	197	8138521465			

Radice dell'MSE	5907.67433	R-quadro	0.1638
Media dip.	12093	R-quadro corr	0.1552
Var coeff	48.85026		

Stime dei parametri

Variabile	Etichetta	DF	Stima dei parametri	Errore standard	Valore t	Pr > t	Stima standardizzata
Intercept	Intercept	1	12088	419.88515	28.79	<.0001	0
Factor5		1	-1452.59548	424.65288	-3.42	0.0008	-0.22405
Factor6		1	2151.28907	423.93830	5.07	<.0001	0.33238

- Per quanto riguarda il **Test F** possiamo verificare che il modello ha capacità esplicativa, il P-value associato al test F è < 0.05 (<0.0001).
- Per quanto riguarda il coefficiente di determinazione **R-quadro** per valutare la capacità esplicativa del modello possiamo dire che le nostre variabili non sono in grado di spiegare in modo soddisfacente la relazione tra la variabile dipendente e i regressori (R-quadro 0.1638).
- Siamo ricorsi al **Test t** per valutare la significatività dei *singoli* fattori. In questo caso il P-value del test è molto piccolo, quindi tutti i fattori rimasti sono rilevanti per spiegare la variabile dipendente.(Pr>t)

3. Regressione lineare (continua)

La procedura REG

Modello: MODEL1

Variabile dipendente: spesa_max spesa_max

Numero osservazioni lette	199
Numero osservazioni usate	198
Numero osservazioni con valori mancanti	1

Analisi della varianza					
Origine	DF	Somma dei quadrati	Media quadratica	Valore F	Pr > F
Modello	3	1891942444	563980815	16.97	<.0001
Errore	194	8446579021	33229789		
Totale corretto	197	8138521465			

Radice dell'MSE	5764.52849	R-quadro	0.2079
Media dip.	12093	R-quadro corr	0.1956
Var coeff	47.86660		

Stime dei parametri								
Variabile	Etichetta	DF	Stima dei parametri	Errore standard	Valore t	Pr > t	Stima standardizzata	Inflazione varianza
Intercept	Intercept	1	14724	900.63488	16.35	<.0001	0	0
Factor5		1	-1418.79366	414.49072	-3.42	0.0008	-0.21884	1.00105
Factor6		1	1919.21881	419.64762	4.57	<.0001	0.29652	1.02958
anni_change	anni_change	1	-312.16478	94.96763	-3.29	0.0012	-0.21316	1.02993

Poiché l'R-quadro così basso rendeva il modello non esplicativo, abbiamo deciso di considerare anche le variabili psicografiche nella proc reg, in particolare la variabile che illustra il numero di anni che intercorrono prima del cambio della macchina. Aggiungendo questa variabile il modello diventa più esplicativo con un R-quadro dello 0,2079 e un R-quadro adjusted di 0,1956 (con una soglia di significatività di 0,2 – 0,3).

Eseguendo l'opzione VIF, dopo l'aggiunta della variabile, i valori trovati sono prossimi a 1, quindi possiamo affermare l'assenza di multicollinearità.

Le considerazioni relative al Test F rimangono invariate. Le tre variabili rimaste dalla stepwise sono significative rispetto al Test T.

3. Regressione lineare: analisi economica

Al termine della regressione lineare possiamo affermare che le variabili indipendenti che spiegano il modello sono il fattore 5, il fattore 6 e il numero di anni che intercorre prima del cambio della macchina. In particolare:

- Il **fattore 5** identifica la **praticità** e l'**utilizzabilità** della vettura (**dimensioni ridotte, consumi ridotti** e il **rispetto dell'ambiente**). Il coefficiente standardizzato è -0,21884 (relazione negativa), quindi coloro che richiedono una vettura pratica, di piccole dimensioni e che consumi poco, ossia una macchina adatta all'utilizzo in città, si aspettano di pagare un prezzo abbastanza contenuto.
- Il **fattore 6** identifica le **caratteristiche strutturali della vettura** (**comfort degli interni** e il **design**). Il coefficiente è pari a 0,29652 (relazione positiva). Coloro che attribuiscono un punteggio elevato al comfort degli interni e al design sono disposti a pagare un prezzo più elevato per la vettura. Questo è il coefficiente più rilevante.
- Anche l'**ultima variabile (anni che intercorrono tra l'acquisto di una vettura e l'altra)** comporta una relazione negativa (coefficiente pari a -0.21316), quindi la diminuzione del tempo che intercorre tra un acquisto e l'altro è indice di una maggiore disponibilità economica e una maggiore propensione al cambiamento frequente della vettura. Dato che siamo in presenza di una relazione negativa, la disponibilità a pagare aumenta.

Stime dei parametri								
Variabile	Etichetta	DF	Stima dei parametri	Errore standard	Valore t	Pr > t	Stima standardizzata	Inflazione varianza
Intercept	Intercept	1	14724	900.63488	16.35	<.0001	0	0
Factor5		1	-1418.79366	414.49072	-3.42	0.0008	-0.21884	1.00105
Factor6		1	1919.21881	419.64762	4.57	<.0001	0.29652	1.02958
anni_change	anni_change	1	-312.16478	94.96763	-3.29	0.0012	-0.21316	1.02993