



***Modelli statistici per
l'analisi dei dati e la
valutazione d'efficacia***

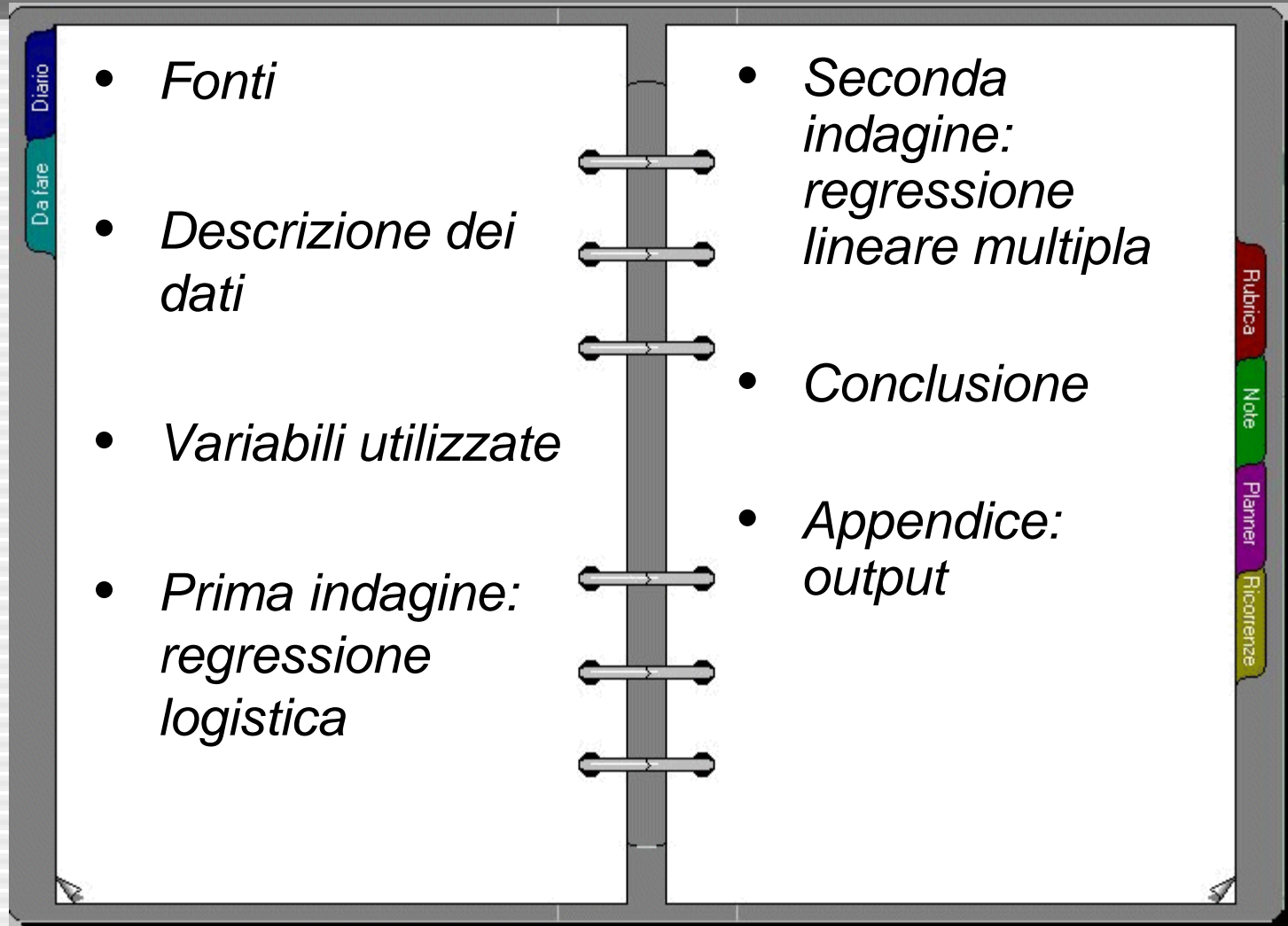
Il caso del Comune di Perugia

Alessandra Pelliccia

Matteo Cataldi

Matteo Filippo Donadi

AGENDA



Le fonti

- database appartenente ad una ricerca effettuata dall'ISTAT nel marzo 2009: è stato sottoposto un questionario ad un **campione di 192 cittadini** residenti all'interno del Comune di Perugia;
- Il questionario si focalizza sulle **aree di maggiore interesse** per il Comune stesso:
 - Utilizzo di uffici anagrafici comunali circoscrizionali;
 - Conoscenza e utilizzo di servizi di certificazione e semplificazione procedurale;
 - Trasporti;
 - Sicurezza
 - Vita quotidiana.

Descrizione dei dati

- Dato che i soggetti sottoposti al questionario in esame sono stati casualmente selezionati all'interno della popolazione residente nel Comune di Perugia, è possibile classificare questa analisi come **“sperimentale”**;
- Si è deciso di estrarre dal database originario esclusivamente dati relativi a sicurezza e vita quotidiana, oltre a quelli anagrafici;
- I dati raccolti sono sia di tipo continuo, che di tipo discreto.

Variabili

- Età
- Sesso
- Professione
- Titolo di studio
- Sezione censimento
- Vita quotidiana
- Sicurezza percepita
- Tipo di area
- Tipo di zona
- Lavoro, extracomunitari

Variabile “Tipo di zona”

In particolare:

- per la variabile “Tipo di zona” sono state ridotte a **3 categorie** le numerose sezioni censimento presenti nel database originario, suddivise in base alla loro distanza dal centro.
- Si è divisa una mappa della città in **tre aree circolari concentriche**:
 - la più interna ha un raggio di 4,5 km (centro)
 - la successiva ha un raggio pari a 9 km (zone limitrofe)
 - la più esterna ha un raggio di oltre 9 km (campagna)

Prima indagine

- Dei 192 cittadini del campione, 107 percepiscono la disoccupazione come problema più urgente: → **quali sono le variabili rilevanti per spiegare tale tipo di percezione?**
- Abbiamo utilizzato la variabile “vita quotidiana” come VD (dicotomica con valori 0 = altro; 1= disoccupazione);
- Si è provato a spiegarla attraverso le VI “titolo di studio”, “età”, “sesso” e “professione”.



Regressione logistica

- Grazie all'utilizzo di una regressione logistica, abbiamo verificato che:
 - Le variabili con p-value maggiore di 0,05 (e quindi non significative nel modello) sono: “età”, “sesso” e “titolo di studio”;
 - L'unica variabile che si è rilevata significativa è stata quella relativa alla “**professione**”. Questa, presa nella sua totalità (prescindendo dai suoi outcome), è risultata significativa, in quanto riporta un P-value = 0,036 ($<0,05$).

Variabile “Professione”

→ Ad ogni outcome PROFESSIONE(n) (con $n=1,2,3,4,5,6,7,8$) corrisponde una tra le professioni contenute nel database;

- PROFESSIONE(7), corrispondente alla categoria “operai”, è risultata essere la più significativa (P-value minore di 0,05, odds-ratio $\text{Exp}(B)$ pari a 21,667);



→ Nel caso in cui il soggetto del campione sia un operaio, è risultato più probabile che questi sia preoccupato primariamente dal problema disoccupazione.

Variabile “Professione”

- Un'altra VI che è risultata significativa, avendo un P-value $< 0,05$, è PROFESSIONE(3), corrispondente alla categoria “**disoccupati**”;
- Se il soggetto del campione è un disoccupato, la probabilità che questi sia preoccupato dal problema disoccupazione aumenta del 11,9% (Exp(B)) rispetto al caso in cui non sia disoccupato.

Variabile “Lavoro”

Introduzione nel modello la VI “**lavoro**”, che indica il livello percepito di difficoltà di trovare lavoro relativamente al solo Comune di Perugia e non al Paese intero;

→ Lo scopo è stato quello di verificare se il problema disoccupazione viene percepito come rilevante in quanto vissuto “in prima persona” dai cittadini del campione oppure in quanto più generalmente percepito all’interno del Paese.

Variabile “Lavoro”

- La variabile è risultata significativa, in quanto riporta un P-value molto basso;
- Tuttavia, l’incremento di probabilità che il soggetto sia sensibile al problema disoccupazione, conseguente all’incremento unitario nella scala relativa al livello di difficoltà percepito nel trovare lavoro all’interno del Comune, si è rivelato modesto (solo 1,719%)
- Dal LR test emerge che il modello completo (con la VI “lavoro”) è significativamente migliore di quello ridotto.

Seconda indagine



Tra i dati del questionario in esame ci si è poi proposti di indagare sui fattori anagrafici e territoriali che incidono sulla percezione del **livello di sicurezza** da parte dei cittadini del campione;

- Abbiamo considerato come VD “Sicurezza percepita”, che è stata trattata come una variabile continua con un minimo di 1 e un massimo di 6;
- Le VI utilizzate sono: “Età”, “Tipo zona”, “Area” ed “Extracomunitari”.

Regressione lineare multipla

- Per procedere in questa analisi abbiamo utilizzato una regressione lineare multipla;
- Le uniche VI significative del modello sono risultate essere “età” e “tipo zona”;
- Si è osservato, analizzando i rispettivi Beta, che risultano entrambe correlate negativamente;

“Età” e “Tipo di zona”

- Per la variabile “**Età**” questo indica che, all’aumentare dell’età dei soggetti del campione, diminuisce il livello di sicurezza percepita: i soggetti più anziani sono quindi risultati essere relativamente più “timorosi”;
- Per la variabile “**Tipo di zona**” un Beta uguale a $-0,425$ indica che i residenti nel centro città si sentono meno sicuri rispetto ai residenti nelle altre zone.
- Non sono presenti problemi di *collinearità* tra le variabili

Conclusioni

- Dall'analisi effettuata sui dati fornitici dal Comune di Perugia si possono trarre diverse conclusioni;
- La preoccupazione per il problema della disoccupazione a livello nazionale è significativamente influenzata dalla professione dei cittadini del campione, in particolare se operai o disoccupati.
- I soggetti del campione risultano preoccupati riguardo al problema della disoccupazione anche in quanto vivono tale problema in prima persona, all'interno del Comune stesso;

Conclusioni

- Rispetto ai soggetti più giovani, quelli relativamente più anziani percepiscono un livello di sicurezza minore, in quanto questi generalmente evitano di uscire durante le ore notturne proprio perché non si sentono al sicuro.
- I soggetti residenti nel centro città tendono a sentirsi meno sicuri rispetto ai soggetti residenti al di fuori del centro. Questo risultato potrebbe essere spiegato dalla presenza nel centro cittadino di numerosi luoghi usualmente frequentati anche durante le ore notturne.

Output regressione logistica

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	240,739 ^a	,112	,151

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Classification Table^a

Observed			Predicted		
			Vita quotidiana 2		Percentage Correct
			0	1	
Step 1	Vita quotidiana 2	0	46	39	54,1
		1	28	79	73,8
Overall Percentage					65,1

a. The cut value is ,500

Output regressione logistica

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	ETAanni	-,008	,014	,359	1	,549	,992
	SESSO(1)	,361	,344	1,101	1	,294	1,434
	PROFESSIONE			16,519	8	,036	
	PROFESSIONE(1)	1,285	1,065	1,458	1	,227	3,616
	PROFESSIONE(2)	2,226	1,139	3,822	1	,051	9,262
	PROFESSIONE(3)	2,476	1,003	6,098	1	,014	11,899
	PROFESSIONE(4)	1,111	,912	1,484	1	,223	3,037
	PROFESSIONE(5)	2,061	1,084	3,615	1	,057	7,858
	PROFESSIONE(6)	1,616	,929	3,025	1	,082	5,033
	PROFESSIONE(7)	3,076	,977	9,913	1	,002	21,677
	PROFESSIONE(8)	1,250	1,022	1,496	1	,221	3,489
	TIT.DISTUDIO	,033	,067	,235	1	,628	1,033
	Constant	-1,261	,892	1,999	1	,157	,283

a. Variable(s) entered on step 1: ETAanni, SESSO, PROFESSIONE, TIT.DISTUDIO.

Output regressione logistica

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a			12,498	8	,130	
PROFESSIONE						
PROFESSIONE(1)	,785	1,112	,498	1	,480	2,193
PROFESSIONE(2)	1,606	1,171	1,880	1	,170	4,981
PROFESSIONE(3)	1,873	1,033	3,286	1	,070	6,505
PROFESSIONE(4)	,552	,941	,344	1	,557	1,737
PROFESSIONE(5)	1,705	1,120	2,319	1	,128	5,501
PROFESSIONE(6)	1,066	,959	1,237	1	,266	2,905
PROFESSIONE(7)	2,444	1,012	5,825	1	,016	11,515
PROFESSIONE(8)	,608	1,071	,322	1	,570	1,836
TIT.DISTUDIO	,046	,071	,423	1	,516	1,047
ETAanni	,000	,015	,000	1	,992	1,000
SESSO(1)	,361	,363	,992	1	,319	1,435
Lavoro	,542	,137	15,728	1	,000	1,719
Constant	-3,456	1,083	10,182	1	,001	,032

a. Variable(s) entered on step 1: PROFESSIONE, TIT.DISTUDIO, ETAanni, SESSO, Lavoro.

[Torna indietro](#)

Output regressione lineare multipla

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,318 ^a	,101	,082	1,29122

a. Predictors: (Constant), tipozona, ETA' (anni), Extracomunitari, Area

b. Dependent Variable: sicurezza2

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35,176	4	8,794	5,275	,000 ^a
	Residual	311,777	187	1,667		
	Total	346,953	191			

a. Predictors: (Constant), tipozona, ETA' (anni), Extracomunitari, Area

b. Dependent Variable: sicurezza2

Output regressione lineare multipla

Coefficients^a

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	5,445	,533		10,214	,000		
ETA' (anni)	-,020	,005	-,286	-4,112	,000	,992	1,008
Extracomunitari	-,086	,073	-,083	-1,177	,241	,977	1,023
Area	,317	,357	,065	,886	,377	,890	1,123
tipozona	-,425	,213	-,147	-1,999	,047	,884	1,131

pendent Variable: sicurezza2