

Convegno 'Statistica in azienda Statistici in azienda', Padova 15/06/10

Metodi statistici non convenzionali per la customer satisfaction

Mauro Gasparini Dipartimento di Matematica - Politecnico di Torino
mauro.gasparini@polito.it
<http://calvino.polito.it/~gasparin>

in collaborazione con **Franco Pellerey**, Politecnico di Torino
Mauro Proietti, Vodafone Italia

1. Introduzione alla Customer Satisfaction

2. Metodi psicometrici

3. Il dati Vodafone

4. Il modello normale

5. Il modello binario

6. Implementazione software

7. Analisi spazio-temporale

Le parole chiave

Impresa

Management

Marketing

Qualità

Marca

Vantaggio competitivo

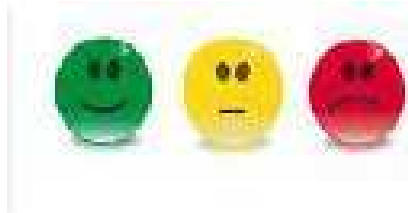
Dati di soddisfazione del consumatore (CS)

Raccolta tramite sondaggi, questionari, feedback, telefono.

Dati **ordinali**, cioè interi da 1 a K ($K=2$ oppure $K=3, 4, 5, 9, 10, 12$) rappresentanti diversi livelli ordinati di soddisfazione.

Se $K=2$, i dati si dicono anche **binari**.

Esempio con $K=3$: iniziativa “**Mettiamoci la faccia**” del ministro Brunetta



I dati del caso Vodafone

Dati di CS raccolti dalla Vodafone tramite l'*Interactive Voice Recognition system*.

Valutazioni da 1 a 9 per sfruttare la tastiera del telefonino.

PROV	ass1	ass2	ass3	ass4	ass5	OVR
prov1	5	6	6	4	6	6
prov1	9	4	9	9	9	9
					
					
provK	7	9	9	7	8	8

PROV è una sigla per la [provincia](#), OVR è una [valutazione complessiva](#) di CS e ass1 ... ass5 sono [valutazioni parziali](#) relative alla copertura indoor, copertura outdoor, qualità fonica, capacità della rete di mantenere le chiamate e di prendere la linea al primo colpo (non in questo ordine per anonimizzarle).

1. Introduzione alla Customer Satisfaction
2. **Metodi psicometrici**
3. Il dati Vodafone
4. Il modello normale
5. Il modello binario
6. Implementazione software
7. Analisi spazio-temporale

Impostazione psicometrica

L'analisi usuale (→ consulenti) tratta la CS come una variabile psicometrica.

Comune anche in altri ambiti sociometrici. CS come intelligenza, socialità, comunicatività...

Esige una modellizzazione a **variabili latenti**, in quanto l'assunto fondamentale é che **la soddisfazione non é una variabile osservabile**.

“...associare ad un concetto, attinente a manifestazioni fenomeniche, un **costrutto concettuale** e dei processi di misurazione atti a collegarlo, tramite ‘**variabili latenti**’, con variabili osservabili o ‘manifeste’ o indicatori.” (A. Zanella (2001) Valutazione e modelli interpretativi di customer satisfaction: una presentazione di insieme. Università Cattolica, E.P. 105.)

Strumenti dell'impostazione psicometrica

SEM, structural equation models, nella versione di covariance structure models (CSM) (→ LISREL)

Kaplan, D.(2009). *Structural Equation Modeling.*

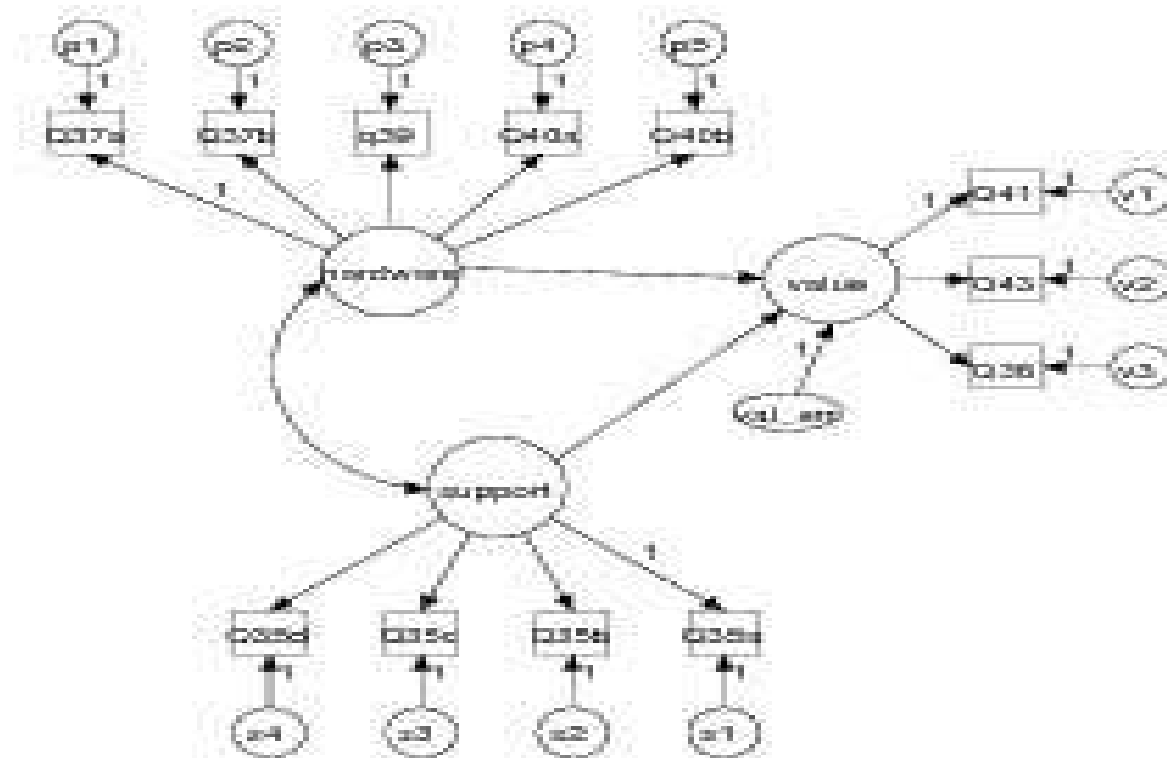
e partial least squares (PLS path analysis) (→ PLS software)

Vinzi, V.E., Chin W.W., Henseler J. and Wang, H. (eds.) (2009).

*Handbook of Partial Least Squares:
Concepts, Methods and Applications
in Marketing and Related Fields.* Springer.

Modelli grafici

In questa impostazione, modelli grafici descrivono le relazioni tra variabili latenti e variabili manifeste (o indicatori). Estensione dell'analisi fattoriale.

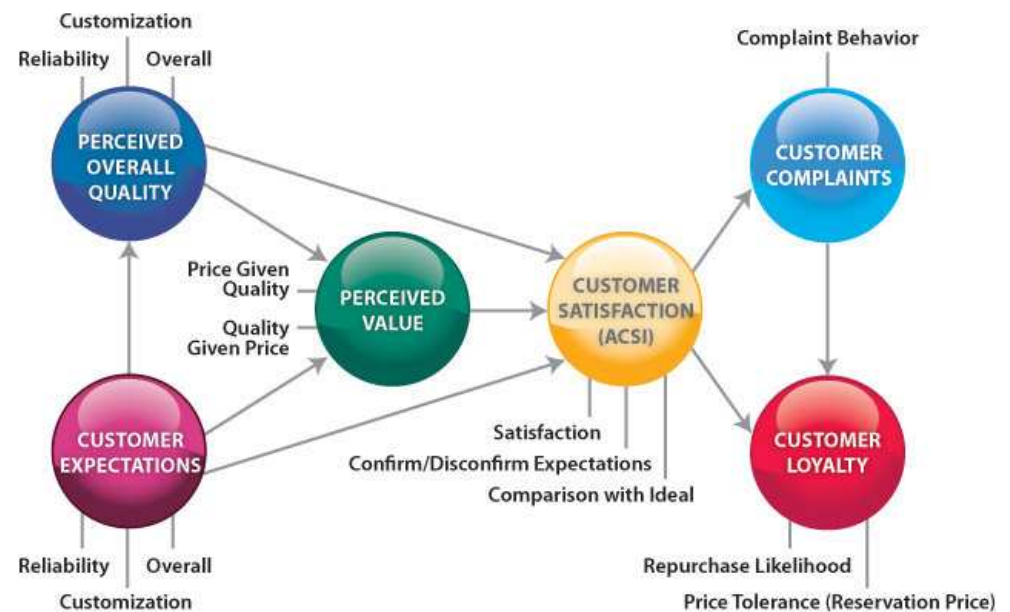


Da Bacon LD (1999). Using LISREL and PLS to measure customer satisfaction.

The American Customer Satisfaction Index (ACSI)

Esempi di costrutti concettuali e variabili latenti: l'**ACSI**, *American Customer Satisfaction Index* e **ECSI**, equivalente europeo.

Il costrutto è usato non solo a livello nazionale, ma anche a livello di imprese o settori con fini di posizionamento. Esempio: Türkyilmaz, A. and Özkan, C (2007). Development of a customer satisfaction Index model. An application to the Turkish mobile phone sector.



I nostri scopi, un poco differenti

Nei dati Vodafone viene richiesto esplicitamente di manifestare la soddisfazione sia globale sia su diversi aspetti del servizio telefonico. Inoltre abbiamo un questionario molto breve (5 domande).

Gli scopi del dipartimento tecnico dell'azienda sono il miglioramento del servizio ([miglioramento della qualità](#)), più che il posizionamento rispetto ai competitori o la messa a punto di un indice assoluto di gradimento.

Decisione di usare i valori così come sono (*face value*), senza l'uso di variabili latenti che esprimano costrutti concettuali.

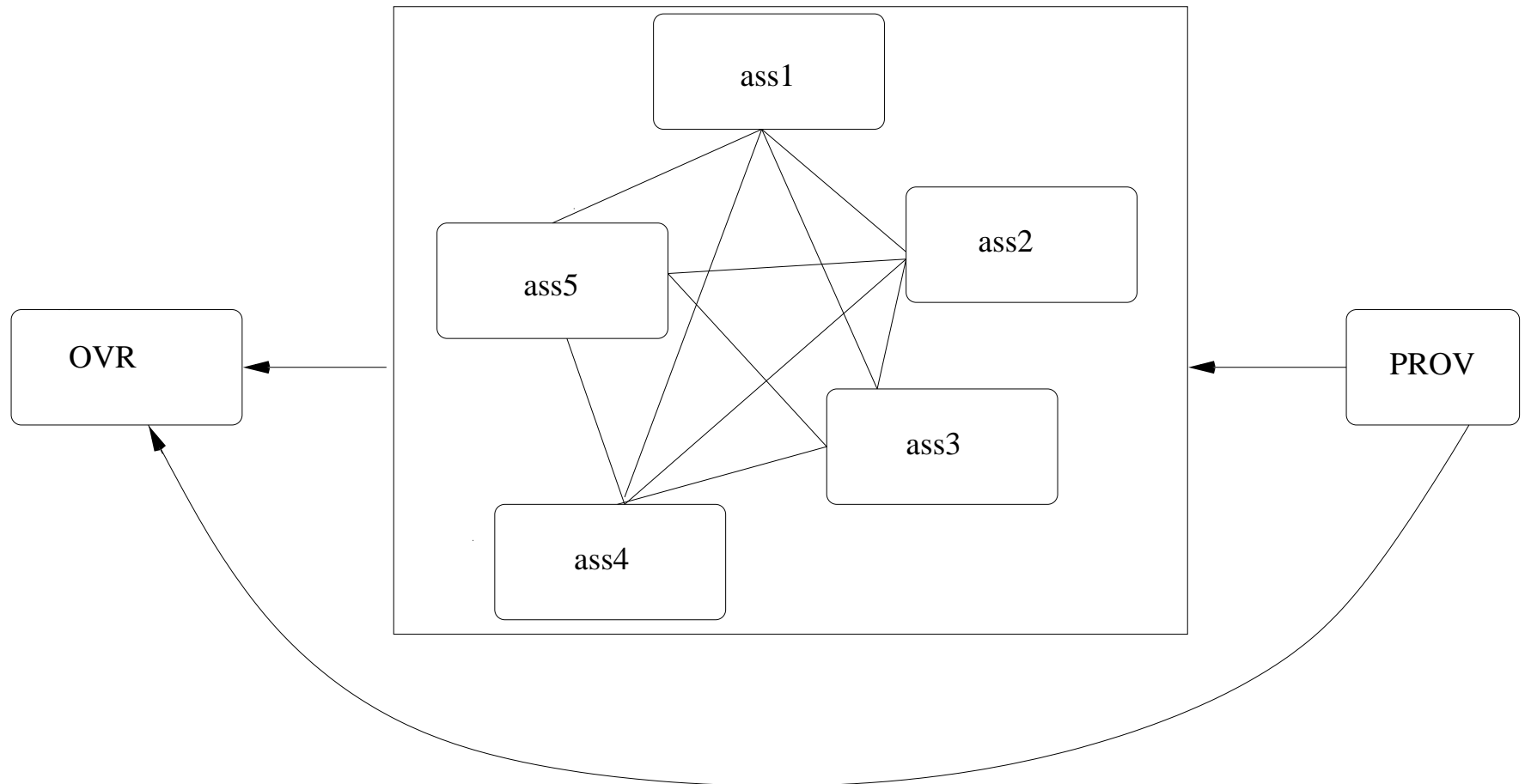
I nostri "metodi non convenzionali", in realtà sono un trattamento statistico convenzionale di [variabili osservabili](#).

1. Introduzione alla Customer Satisfaction
2. Metodi psicometrici
3. I dati Vodafone
4. Il modello normale
5. Il modello binario
6. Implementazione software
7. Analisi spazio-temporale

I dati del caso Vodafone

PROV	ass1	ass2	ass3	ass4	ass5	OVR
prov1	5	6	6	4	6	6
prov1	9	4	9	9	9	9
					
					
provK	7	9	9	7	8	8

Grafo delle variabili osservabili



OVR = valutazione complessiva di CS

PROV = sigla provincia

ass1 ... ass5 = valutazioni parziali

Interpretazione del grafo

- un SEM con sole variabili osservabili (no latenti)
- un *joint-response chain graph of a multivariate regression type*
- la base di un DAG bayesiano (grafo orientato aciclico), a cui aggiungere i nodi rappresentanti i parametri per ottenere il DAG completo ben noto ai bayesiani

Dati normali, ordinali o binari?

Le valutazioni, sia parziali che complessive, sono numeri da 1 a 9. Nei metodi standard, si assume per esse un **modello normale**, che chiaramente è una forzatura.

In prima approssimazione, lo imponiamo anche noi.

Successivamente, dicotomizziamo le valutazioni in basse/alte e utilizziamo invece un **modello binario (bernoulliano)**.

Se avessimo un chiaro criterio per partizionare più finemente, potremmo usare anche → **modelli per dati ordinali**.

1. Introduzione alla Customer Satisfaction
2. Metodi psicometrici
3. I dati Vodafone
4. Il modello normale
5. Il modello binario
6. Implementazione software
7. Analisi spazio-temporale

Il modello normale in forma algebrica: verosimiglianza

Le valutazioni di CS **parziali** $\mathbf{X}_{k,j}$ (vettore 5-dimensionale) del j -esimo cliente, $j = 1, \dots, n_k$ della k -esima provincia, $k = 1, \dots, K$ sono normali

$$\mathbf{X}_{k,j} = (X_{k,j,1}, X_{k,j,2}, X_{k,j,3}, X_{k,j,4}, X_{k,j,5})' \stackrel{\text{i.i.d.}}{\sim} \text{Normale}_5(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$\boldsymbol{\mu}_k = (\mu_{k,1}, \mu_{k,2}, \mu_{k,3}, \mu_{k,4}, \mu_{k,5})', \quad \boldsymbol{\Sigma}_k = \begin{bmatrix} \sigma_k^2 & \gamma_k & \gamma_k & \gamma_k & \gamma_k \\ \gamma_k & \sigma_k^2 & \gamma_k & \gamma_k & \gamma_k \\ \gamma_k & \gamma_k & \sigma_k^2 & \gamma_k & \gamma_k \\ \gamma_k & \gamma_k & \gamma_k & \sigma_k^2 & \gamma_k \\ \gamma_k & \gamma_k & \gamma_k & \gamma_k & \sigma_k^2 \end{bmatrix}$$

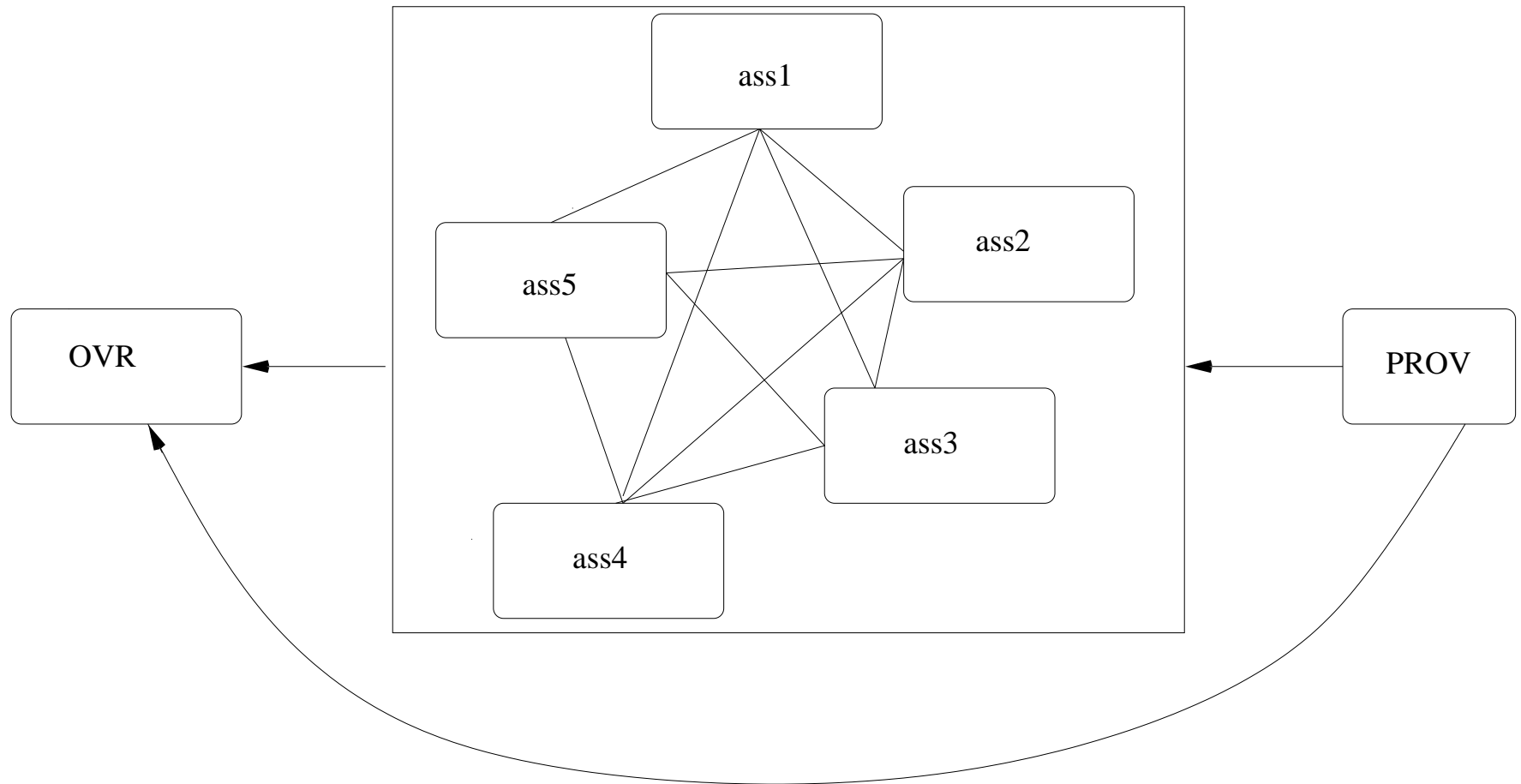
cioè il vettore $\mathbf{X}_{k,j}$ è uguale al vettore di medie $\boldsymbol{\mu}_k$ più **errori scambiabili**.
Per quanto riguarda la corrispondente valutazione **complessiva** $Y_{k,j}$:

$$Y_{k,j} | \mathbf{X}_{k,j} \sim \text{Normal} (E(Y_{k,j}), \tau_k^2)$$

$$E(Y_{k,j}) = \beta_{k,0} + \beta_{k,1}X_{k,j,1} + \beta_{k,2}X_{k,j,2} + \beta_{k,3}X_{k,j,3} + \beta_{k,4}X_{k,j,4} + \beta_{k,5}X_{k,j,5}$$

(regressione lineare semplice condizionata alle $\mathbf{X}_{k,j}$)

Grafo delle variabili osservabili

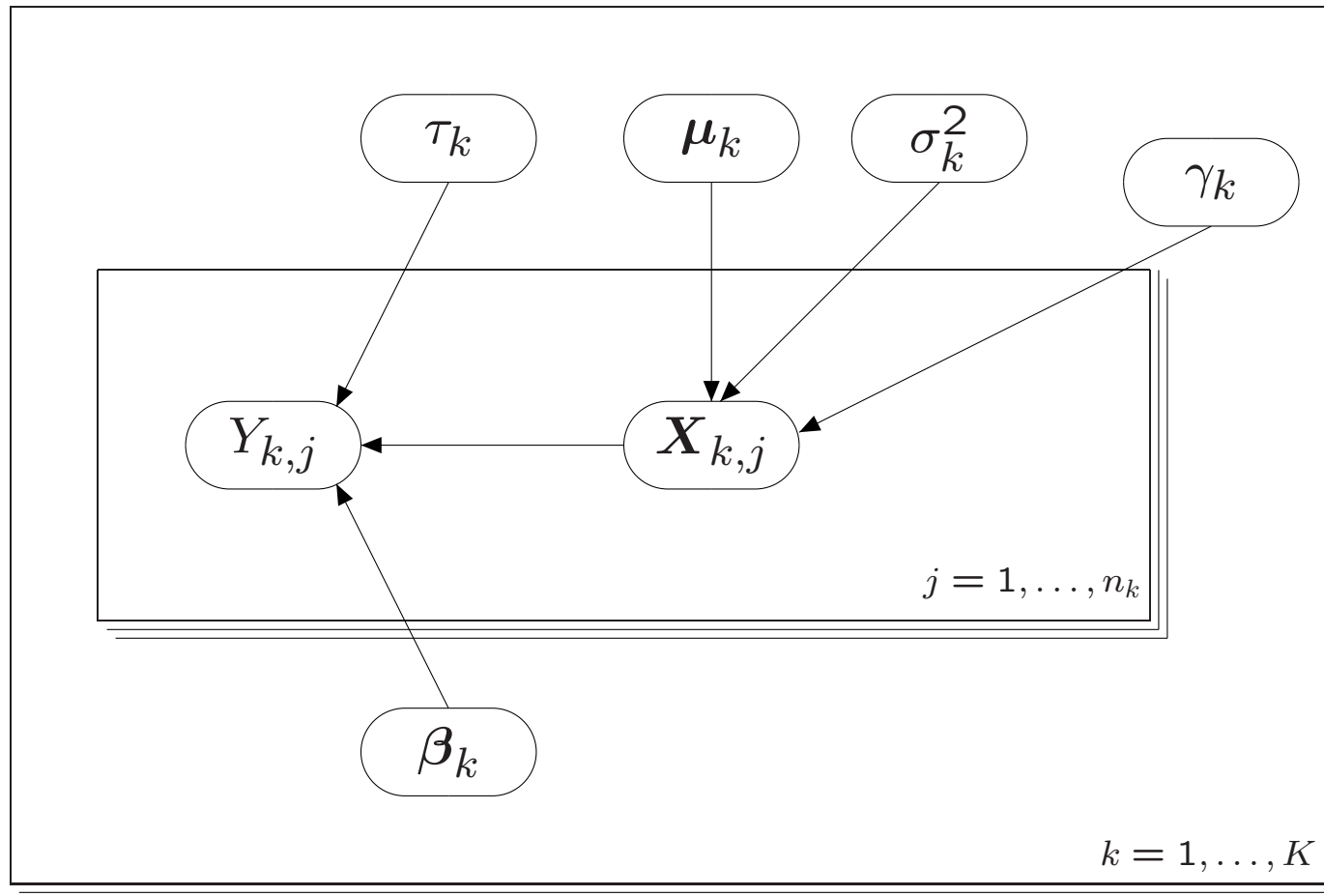


OVR = valutazione complessiva di CS

PROV = sigla provincia

ass1 ... ass5 = valutazioni parziali

DAG espanso per il modello normale



A priori piatte per il modello normale

$$\begin{aligned}\mu_{k,v} &\stackrel{\text{i.i.d.}}{\sim} \text{Normal}(5, L) \\ \beta_{k,0} &\stackrel{\text{i.i.d.}}{\sim} \text{Normal}(5, L) \\ \beta_{k,v} &\stackrel{\text{i.i.d.}}{\sim} \text{Normal}(0, L) \\ 1/\tau_k^2 &\stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\epsilon, \epsilon) \\ \log(\sigma_k^2) &\stackrel{\text{i.i.d.}}{\sim} \text{Normal}(0, L) \\ \gamma_k | \sigma_k^2 &\sim \text{Uniform}(0, S\sigma_k^2)\end{aligned}$$

$k = 1, \dots, K, v = 1, 2, 3, 4, 5$

L un numero opportunamente grande

ϵ un numero opportunamente piccolo

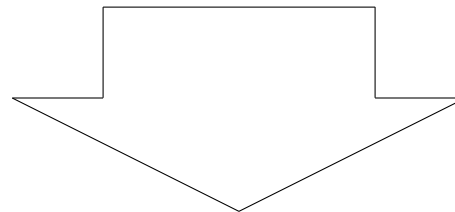
S un numero vicino a 1

vincolo necessario: $\gamma_k \leq \sigma_k^2$)

1. Introduzione alla Customer Satisfaction
2. Metodi psicometrici
3. I dati Vodafone
4. Il modello normale
5. **Il modello binario**
6. Implementazione software
7. Analisi spazio-temporale

Dicotomizzazione

PROV	ass1	ass2	ass3	ass4	ass5	OVR
prov1	5	6	6	4	6	6
prov1	9	4	9	9	9	9
					
					
provK	7	9	9	7	8	8



PROV	ass1	ass2	ass3	ass4	ass5	OVR
prov1	0	1	1	0	1	1
prov1	1	0	1	1	1	1
					
					
provK	1	1	1	1	1	1

Il modello binario in forma algebrica: verosimiglianza

Dicotomizzazione: $I_{k,j,l} = (X_{k,j,l} > 5)$, $Z_{k,j,l} = (Y_{k,j} > 5)$

Verosimiglianza bernoulliana per le valutazioni parziali:

$$I_{k,j,l} \sim \text{Bernoulli}(p_{k,l}), \quad l = 1, 2, 3, 4, 5 \quad j = 1, \dots, n_k$$

$$\mathbf{p}_k = (p_{k,1}, p_{k,2}, p_{k,3}, p_{k,4}, p_{k,5})' \quad (\text{analogo a } \boldsymbol{\mu}_k)$$

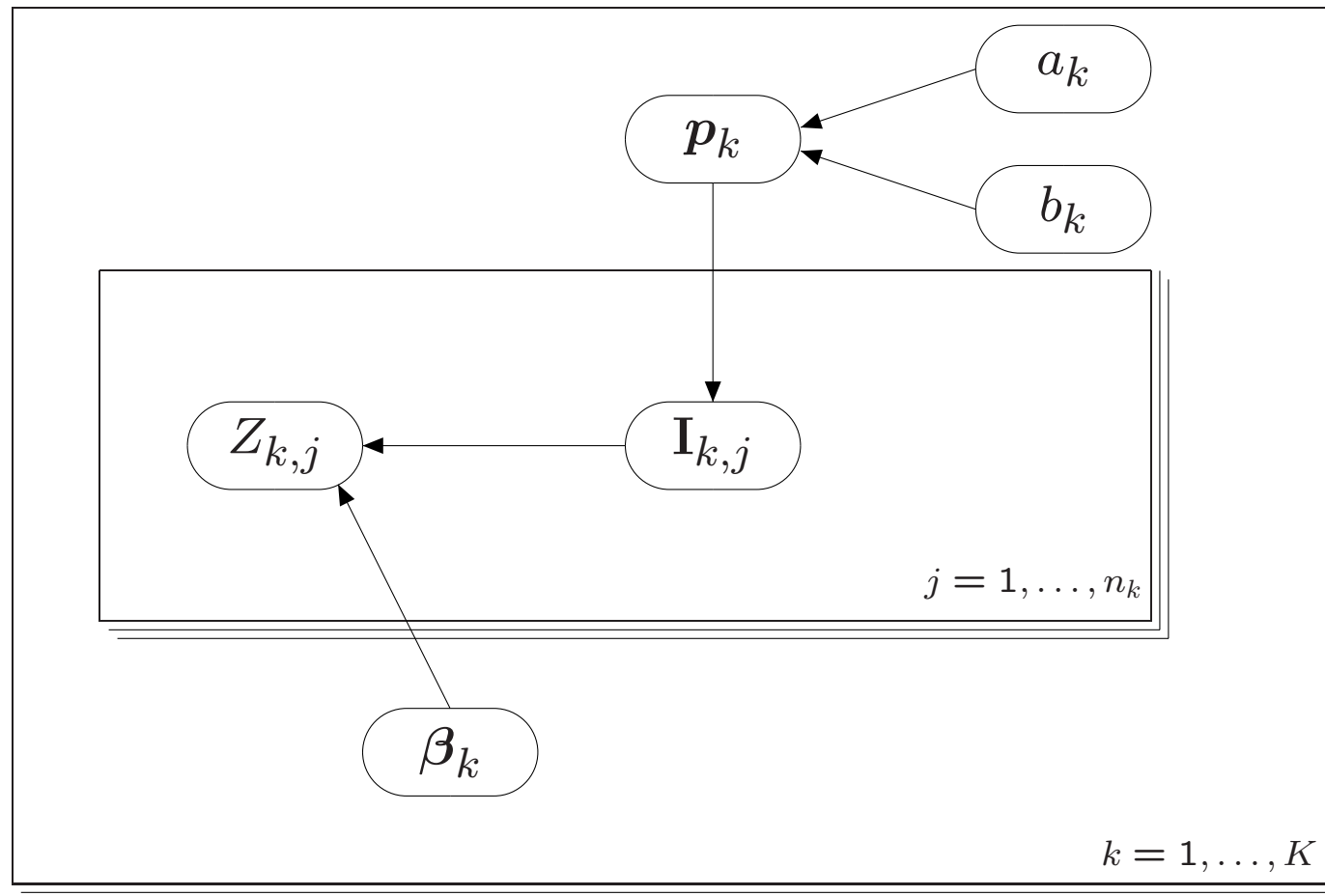
Verosimiglianza bernoulliana per la valutazione complessiva:

$$Z_{k,j} | \mathbf{I}_{k,j} \sim \text{Bernoulli}(q_{k,j})$$

$$\log \frac{q_{k,j}}{1 - q_{k,j}} = \beta_{k,0} + \beta_{k,1}I_{k,j,1} + \beta_{k,2}I_{k,j,2} + \beta_{k,3}I_{k,j,3} + \beta_{k,4}I_{k,j,4} + \beta_{k,5}I_{k,j,5}$$

(regressione logistica condizionata agli $I_{k,j,l}$)

DAG espanso per il modello binario



A priori piatte per il modello binario

$$\begin{aligned} p_{k,1}, p_{k,2}, p_{k,3}, p_{k,4}, p_{k,5} &\stackrel{\text{i.i.d.}}{\sim} \text{Beta}(a_k, b_k) \\ \beta_{k,v} &\stackrel{\text{i.i.d.}}{\sim} \text{Normal}(0, L) \\ a_k &\stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\epsilon, \epsilon) \\ b_k &\stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\epsilon, \epsilon) \end{aligned}$$

$$k = 1, \dots, K, v = 1, 2, 3, 4, 5$$

L un numero opportunamente grande

ϵ un numero opportunamente piccolo

1. Introduzione alla Customer Satisfaction
2. Metodi psicometrici
3. I dati Vodafone
4. Il modello normale
5. Il modello binario
6. Implementazione software
7. Analisi spazio-temporale

Implementazione software

Il modello, sia normale che binario, è stato stimato in WinBUGS, un programma di calcolo bayesiano basato sulla simulazione MCMC delle distribuzioni a posteriori. (In entrambi i casi, qualche progresso si può fare anche in forma chiusa ma non ne vale la pena.)

Per sviluppi recenti del programma BUGS, si veda Lunn D., Spiegelhalter D., Thomas A. and Best N. (2009) The BUGS project: Evolution, critique and future directions *Statistics in Medicine*, 28:3049-3067



→ OPENBUGS

WinBUGS può essere invocato (per ora solo in ambiente Windows) dall'interno di R con la routine R2WinBUGS.

Dettagli: il codice BUGS

```
#####  
# exch4.txt - by MG, February 2009  
#           WINBUGS for CS problem  
#####  
  
# the model  
model  
{  
##### PRIORS  
# there are K=7 provinces  
for (k in 1:K){  
  
# prior on intercept  
  b0[k] ~ dnorm( 0,1.0E-6)  
  
# priors on regression coefficients  for each province  
  for (j in 1:5) {b[k,j] ~ dnorm( 0.0,1.0E-6)}  
  
# priors on means of the partial CS indices  
# mean around 5, possibly 7  
  for (j in 1:5) {media[k,j] ~ dnorm( 5,1.0E-6)}  
  
# priors on precisions  
  prec.ovr[k] ~ dgamma( 1.0E-1, 1.0E-1)  
  
# priors on variances and covariances of the partial CS indices  
varianza[k] ~ dlnorm(0,1.0E-10)  
upco[k]  <- 0.95*varianza[k]
```

```

covarianza[k] ~ dunif(0, upco[k])

Sigma[k,1,1] <- varianza[k]
for (j in 2:5) {Sigma[k, j , j] <- varianza[k]
  for (r in 1:(j-1)) {Sigma[k, j, r] <- covarianza[k]}
  for (r in 1:(j-1)) {Sigma[k, r, j] <- covarianza[k]}
}
Omega[k,1:5,1:5] <- inverse(Sigma[k, , ])
}

##### DATA
### reading data
for (i in 1:N) {

# y[,1] is an indicator for province
# CT=1, GE=2, MI=3, NA=4, RM=5, VE=6, VR=7
  ind[i] <- y[i,1]
# y[,2:6] are the partial cs variables
  y[i, 2:6] ~ dnorm( media[ind[i], 1:5 ], Omega[ind[i], 1:5, 1:5])

#mean of the overall CS variable
  mean.ovr[i] <- b0[ind[i]] + b[ind[i],1]*y[i,2]+ b[ind[i],2]*y[i,3]+
    b[ind[i],3]*y[i,4]+ b[ind[i],4]*y[i,5]+b[ind[i],5]*y[i,6]

# y[,7] is the overall CS variable
  y[i,7] ~ dnorm(mean.ovr[i], prec.ovr[ind[i]])
}

}

```

Dettagli: l'interfaccia R2WinBUGS

```
#
# Rexch.R - by LP and MG, February 2009
#           interface between R and Winbugs
#           in the cs problem
#

library("R2WinBUGS")

iniz<-function(){list(b0=rep(0,7),
                      P=c(0,0,0,0,0,0,0),
                      V=c(0,0,0,0,0,0),
                      G=0,
                      prec.ovr=rep(1,7),
                      covarianza=c(0,0,0,0,0,0,0),
                      varianza=c(5,5,5,5,5,5,5),
                      b=structure(.Data=rep(0,35),.Dim=c(7,5)),
                      E=structure(.Data=rep(0,35),.Dim=c(7,5)))}

dati <- list( K = 7,      # no. provinces
              N = 1615, # total no. customers
              y = y
            )

BUGS.DIR<-"C:/Programmi/WINBUGS14" ### set wherever WINBUGS is

modelname<-"exch4.txt"

burn<-1
iter<-10000
```

```
results<- bugs (data=dati,  
               init=iniz,  
               parameters.to.save=c("media","P","V","b0","b","varianza","covarianza","prec.ovr"),  
               model.file=modelname,  
               n.chains= 1,  
               n.iter=iter,  
               n.burnin=burn,  
               bugs.directory=BUGS.DIR,  
               debug=TRUE)  
  
#sink("file.txt")  
print(results,digits.summary=3)  
#sink()  
  
##### roughly, Table 3  
round(results$mean$media,3)  
round(results$sd$media,3)  
  
##### roughly, Table 4  
round(cbind(results$mean$b0,results$mean$b),3)  
round(cbind(results$sd$b0,results$sd$b),3)  
  
##### roughly, Table 5  
round(cbind(results$mean$varianza,results$mean$covarianza,results$mean$prec.ovr),3)  
round(cbind(results$sd$varianza,results$sd$covarianza,results$sd$prec.ovr),3)
```


Risultati

I risultati di un'analisi bayesiana sono vari riassunti della distribuzione a posteriori, che in uno scenario MCMC sono statistiche dei valori simulati dei parametri.

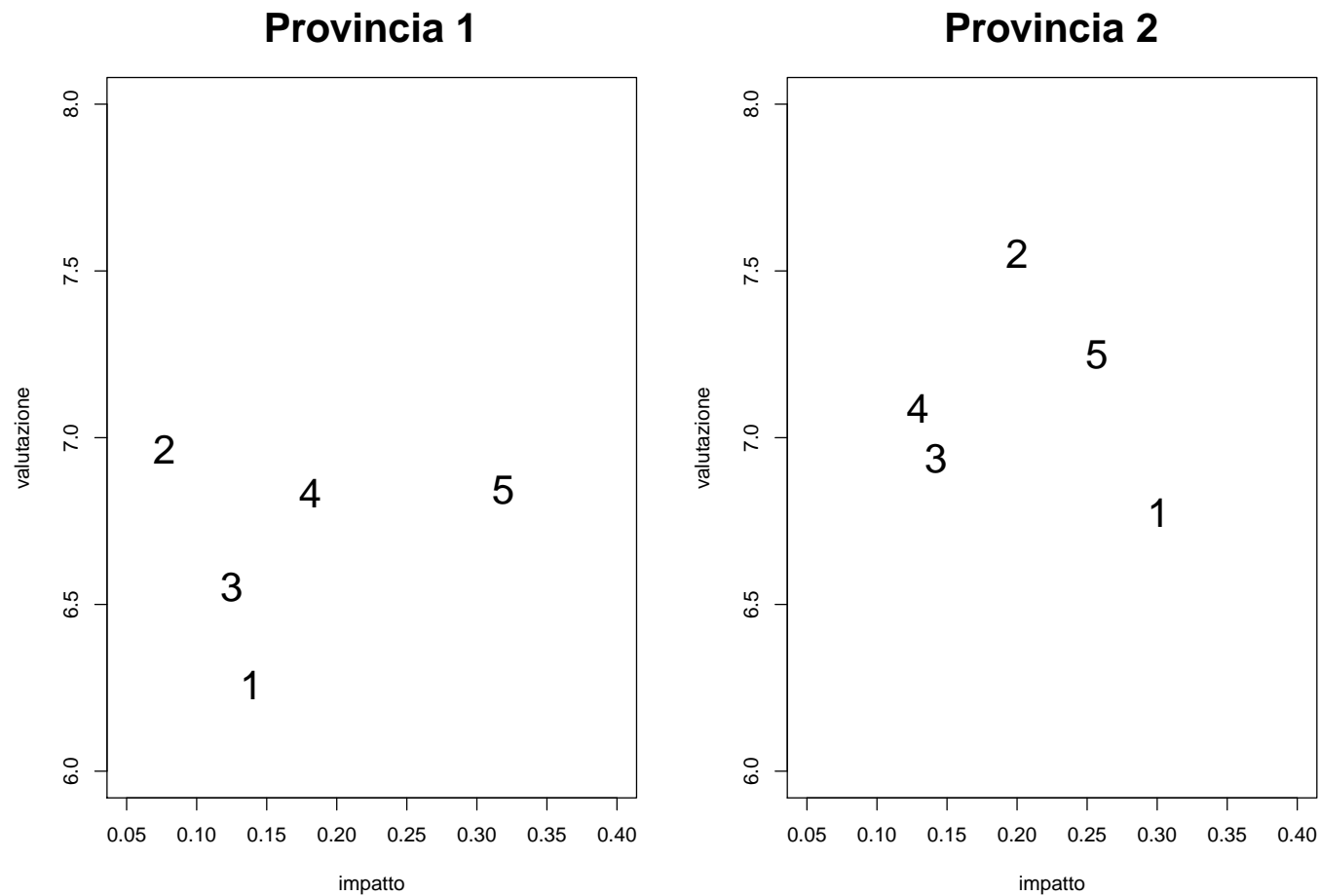
Esempio: medie (deviazioni standard) a posteriori dei μ per il caso normale:

k	$\mu_{k,1}$	$\mu_{k,2}$	$\mu_{k,3}$	$\mu_{k,4}$	$\mu_{k,5}$
1	6.258 (0.147)	6.965 (0.145)	6.551 (0.147)	6.834 (0.146)	6.842 (0.146)
2	6.774 (0.139)	7.552 (0.140)	6.937 (0.138)	7.087 (0.140)	7.248 (0.139)
3	7.091 (0.126)	7.617 (0.126)	7.193 (0.126)	7.057 (0.122)	7.299 (0.125)
4	6.377 (0.147)	7.235 (0.145)	6.478 (0.146)	7.181 (0.145)	7.081 (0.144)
5	6.651 (0.126)	7.441 (0.125)	6.809 (0.126)	7.095 (0.126)	7.093 (0.127)
6	6.836 (0.120)	7.625 (0.120)	7.028 (0.120)	7.105 (0.119)	7.334 (0.119)
7	6.506 (0.137)	7.239 (0.137)	7.025 (0.135)	6.979 (0.135)	7.016 (0.135)

1. Introduzione alla Customer Satisfaction
2. Metodi psicometrici
3. I dati Vodafone
4. Il modello normale
5. Il modello binario
6. Implementazione software
7. **Analisi spazio-temporale**

Grafici valutazione su impatto del primo tipo

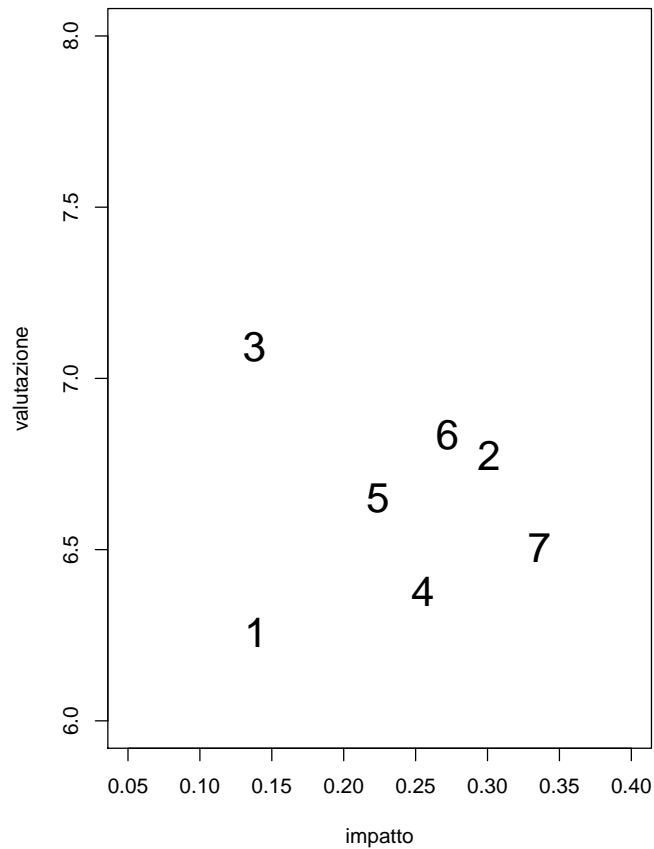
Sono grafici con $\beta_{k,v}$ (impatto) in ascissa e $\mu_{k,v}$ (valutazione media) in ordinata, per v (valutazione parziale) fissata.



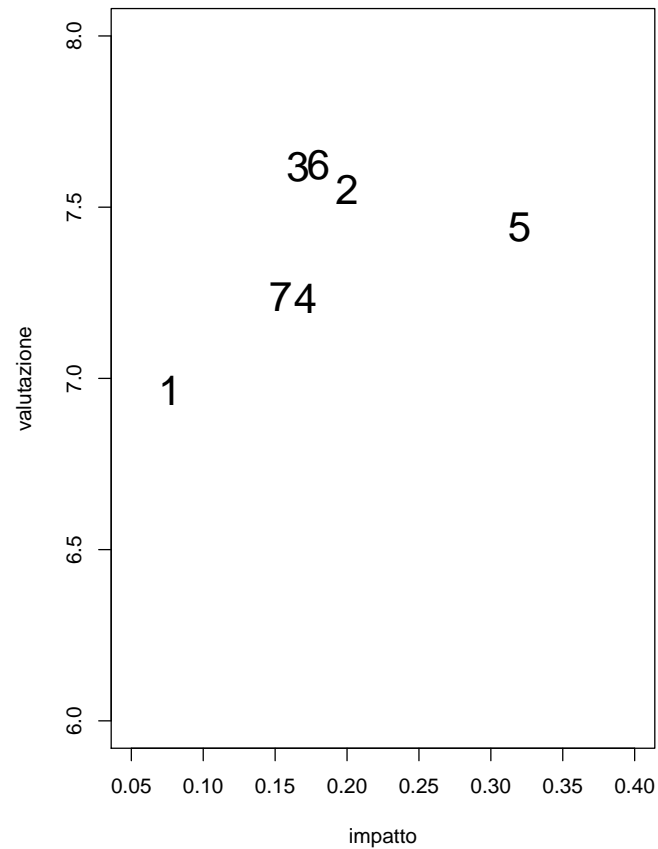
Grafici valutazione su impatto del secondo tipo

Sono grafici con $\beta_{k,v}$ (impatto) in ascissa e $\mu_{k,v}$ (valutazione media) in ordinata, per k (provincia) fissata.

Valutazione parziale 1



Valutazione parziale 2



Confronti temporali

Confrontiamo Novembre e Dicembre 2009 in termini di stime e intervalli di credibilità.

