Esame di MQAT + Statistica del 3-2-10

Nome / eMail : ....

Un campione di $n = 50$ famiglie, a cui e' stato chiesto se abitano in Periferia/Centro ed hanno la casa in Affitto/Proprieta', ha generato la seguente tabelle di frequenze:

Y \ X	Periferia	Centro
Affitto	8	2
Proprieta'	24	16

7) Si chiede di calcolare l'indice di Dipendenza di Pearson e di fare il test del χ^2. Calcolare anche l'indice di Cramer V e commentarlo.

7) Assegnare valori *numerici* appropriati alle variabili X, Y e calcolare l'indice di Correlazione r_{xy}. Confrontare il suo valore con V.

3) Usando le frequenze relative, calcolare la probabilita' condizionata di avere una casa in Proprieta', in Periferia.

3) Supponendo che X, Y siano *sommabili*, calcolare la Varianza della trasformazione $Z = (2 - X + Y)$ (occhio alla dipendenza).

5) Calcolare le proporzioni campionarie $\hat{\pi}_x$, $\hat{\pi}_y$ delle case in Proprieta' e di quelle in Perifieria (totali) e verificare se sono uguali.

5) Stimare il modello di regressione $Y_i = \alpha + \beta\, X_i + e_i$, e verificare se il coefficente $\hat{\beta}$ e' statisticamente significativo.

Soluzioni

1) Si procede al calcolo della matrice di frequenze sotto condizioni di indipendenza $n_{ij}^* = n_i \, n_j / n$ ottenendo n_{11}^*=10*32/50=6.4, n_{12}^*=10*18/50 ... poi si sommano le contingenze quadratiche $c_{ij} = (n_{ij} - n_{ij}^*)^2 / n_{ij}^*$, ottenendo la statistica di Pearson $\hat{W}_n$=1.39. Il valore e' basso, in particolare rispetto al valore tabulato $\chi^2_{0.95}(1)$ =3.84, si accetta quindi l'ipotesi di Indipendenza tra X, Y, ed il valore dell'indice di Cramer $V_n = \sqrt{W_n/n}$=0.17 e' basso.

2) Il modo piu' semplice e' di assegnare i valori dicotomici X=0,1 e Y=0,1, per cui usando le frequenze marginali le medie diventano $\bar{X} = 1 * 18/50 = 0.36$, $\bar{Y} = 0.8$, ed utilizzando le medie quadratiche si hanno le varianze $S_x^2 = 1^2 18/50 - .36^2 = 0.23$ e $S_y^2 = 0.16$. La covarianza diventa $S_{xy} = 1 * 1 * 16/50 - .36 * .8 = 0.032$ ed infine la correlazione $R_{xy} = 0.17$ bassa, equivalente all'indice di Cramer.

3) La probabilita' e' P($Y > 0 | X < 1$)=P[$(Y = 1) \cap (X = 0)$]/P($X = 0$)=24/32

4) Var(Z)=Var(-X)+Var(Y)+2Cov(-X, Y)=0.23+0.16-2*0.032

5) Avendo $\hat{\pi}_y$=40/50 e $\hat{\pi}_x$=32/50 si calcola l'intervallo di confidenza approssimato per la differenza delle proporzioni al 95%

$$\hat{\delta}_{1,2} \approx (0.8 - 0.64) \pm 2\sqrt{.8 * .2/50 + .64 * .36/50} = [-0.017, +0.34]$$

poiche' include il valore 0 si accetta l'ipotesi $\pi_x = \pi_y$

6) Le stime dei parametri sono $\hat{\beta}$=.032/.23=0.14, $\hat{\alpha}$=.8-.14*.36=0.75 e $\hat{\sigma}_e^2$=.16-.14^2*.23=0.15. Si calcola l'intervallo di confidenza approssimato al 95% per β

$$\hat{\beta}_{1,2} \approx 0.14 \pm 2\sqrt{.15/(50 * .23)} = [-0.09, 0.37]$$

per cui si accetta l'ipotesi $\beta = 0$, il modello non ha una validita' statistica.