Exam

10 mc questions → pass

4 problems. → 10/10 | 2 hours

offline | proofs: online.

---

time

SPSS output for PCA + Varimax.

KMO Bartlett's test:

KMO (Kaiser - Meyer-Olkin) statistics

original variables.

X

↳ correlations { $r_{ij} = \dfrac{s\mathrm{Cov}(x_i, x_j)}{\sqrt{s\mathrm{Var}(x_i)\cdot s\mathrm{Var}(x_j)}}$

s = sample

$$s\mathrm{Cov}(x_i, x_j) = \frac{\sum_{t=1}^{n} (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j)}{n-1}$$

$$s\mathrm{Var}(x_i) = s\mathrm{Cov}(x_i, x_i)$$

→ partial correlations → p_{ij}

$$KMO = \frac{\sum_{i\neq j} r_{ij}^2}{\sum_{i\neq j} r_{ij}^2 + \sum_{i\neq j} p_{ij}^2}$$

"the overall quality of all sample regression"

KMO measure of sampling adequacy.

$r_{ab} = 0.6$?

$\hat{a}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot b_i$

$\hat{\beta}_2 = 2.6$

on average a goes up by 2.6 when b goes by 1

$\hat{b}_i = \hat{\gamma}_1 + \hat{\gamma}_2 a_i$

$\gamma_2 \neq 2.6$

slope coefficient for stand variables.

Interpr:

If I scale a and b.

$a_i^* = \frac{a_i - \bar{a}}{se(a)}$ $\quad b_i^* = \frac{b_i - \bar{b}}{se(b)}$

$se(a) = \sqrt{\frac{\Sigma(a_i - \bar{a})^2}{n-1}}$

and apply OLS

$\hat{a}_i^* = \underset{\text{interc}}{0} + \underbrace{0.6}_{r_{ab}} \cdot b_i^*$ $\quad \hat{b}_i^* = \underset{\text{interc}}{0} + \underbrace{0.6}_{r_{ab}} \cdot a_i^*$

$\boxed{0.6}$: if b goes up by 1 se(b) then a (on average) goes up by 0.6 se(a).

Stat: $r_{ab}^2 = \hat{\beta}_2(a \text{ on } b) \cdot \hat{\beta}_2 (b \text{ on } a)$

$\hat{a}_i = \hat{\beta}_1^{a \to b} + \hat{\beta}_2^{a \to b} \cdot b_i$

$\hat{b}_i = \hat{\beta}_1^{b \to a} + \hat{\beta}_2^{b \to a} a_i$

$\hat{\beta}_2^{a \to b} \cdot \hat{\beta}_2^{b \to a} = r_{ab}^2$

r_{ab} - geom mean of two regression coefficients.

$\text{sign}(r_{ab}) = \text{sign}(\hat{\beta}_2^{a \to b}) = \text{sign}(\hat{\beta}_2^{b \to a})$

$r_{ab}^2 = 0{,}36$ ← R^2 in all 4 regressions!

$\hat{a}_i = \hat{\beta}_1 + \hat{\beta}_2 b_i$ — $R^2(a \to b)$

$\hat{b}_i = \hat{\gamma}_1 + \hat{\gamma}_2 a_i$ — $R^2(b \to a)$

$\hat{a}^*_i = 0 + 0{,}6\, b^*_i$ — $R^2(a^* \to b^*)$

$\hat{b}^*_i = 0 + 0.6\, a^*_i$ — $R^2(b^* \to a^*)$

(any comb gives the same R^2)

$R^2(b^* \to a)$

$R^2(a^* \to b)$

$$\frac{sVar(\hat{a})}{sVar(a)} = \frac{sVar(\hat{b})}{sVar(b)} =$$

$$= \frac{sVar(\hat{a}^*)}{\underbrace{sVar(a^*)}_{=1}} = sVar(\hat{a}^*) = \frac{sVar(\hat{b}^*)}{\underbrace{sVar(b^*)}_{=1}} = sVar(\hat{b}^*)$$

$$\begin{bmatrix} | \\ x_1 \\ | \end{bmatrix} \begin{bmatrix} | \\ x_2 \\ | \end{bmatrix} \cdots \begin{bmatrix} | \\ x_k \\ | \end{bmatrix}$$

$x_i \ [n \times 1]$

$X \ [n \times k]$

x_{-i} – all var-s except x_i $[n \times (k-1)]$

$x_{-i,j}$ – all var-s except x_i, x_j $[n \times (k-2)]$

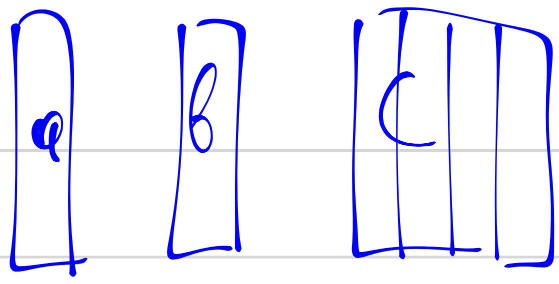
def 1. partial correlation

$$p_{ij} = sCorr(\hat{u}[x_i \to x_{-ij}], \hat{u}[x_j \to x_{-ij}])$$

residuals in regression of x_i on all other variables except x_i and x_j

Clean out x_i from everything else $\hat{u}[x_i \to x_{-ij}]$

Clean out x_j from everything else $\hat{u}[x_j \to x_{-ij}]$



$a = x_i$ $\quad$ $C = x_{-ij}$

$b = x_j$

regress a on C, get residuals $\hat{u}_a$

regress b on C, get residuals $\hat{u}_b$.

$$p_{ab} = sCorr(\hat{u}_a, \hat{u}_b)$$

var-le b clean from variables C.

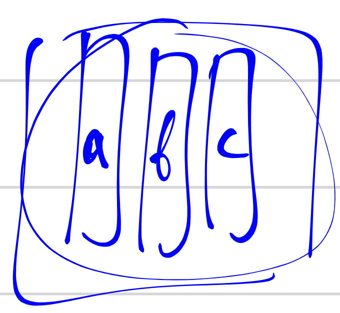
var-le a clean from variables in C

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} p_{ij}^2} \in [0;1]$$

$KMO \approx 0$

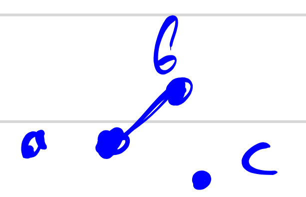
$$\sum_{i \neq j} p_{ij} >> \sum_{i \neq j} r_{ij}^2$$

$KMO \approx 1$



$$\sum r_{ij}^2 >> \sum p_{ij}^2$$

$\boxed{\text{case 1}}$



$$KMO \approx \frac{1+0+0}{1+0+0+1+0+0} = \frac{1}{2}$$

$\hat{a} \approx 0 + 0 \cdot c$ $\quad$ $\hat{u}_a \approx a$

$\hat{b} = 0 + 0 \cdot c$ $\quad$ $\hat{u}_b \approx b$

$\hat{c} \approx 0 + 0 \cdot a$ $\quad$ $\hat{u}_c \approx c$

$\hat{b} \approx \hat{\beta}_1 + \hat{\beta}_2 a$ $\quad$ $\hat{u}_b = b - \hat{\beta}_1 - \hat{\beta}_2 a$

case 2. $a_i + b_i + c_i = 2022$

[no perfect linear dep between
a and b.
b and c.
a and c]

? $\rho_{ab} \approx$ {intuit} $= 1$

Ex ? $r_{ab} \approx -\frac{1}{2}$!!

$$KMO = \frac{\frac{1}{4}+\frac{1}{4}+\frac{1}{4}}{\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+1+1+1}$$
$\Leftarrow$ low

n observ-ts.

$x_i \sim N(100,1)$

$y_i \sim N(100,1)$

$z_i \sim N(100,1)$

$$a_i = \frac{x_i}{x_i+y_i+z_i} \quad b = \frac{y_i}{x_i+y_i+z_i} \quad c_i = \frac{z_i}{x_i+y_i+z_i}$$

$a_i + b_i + c_i = 1$!!

$$Cov(c_i, \underbrace{a_i + b_i + c_i}_{= const}) = 0$$

$$Cov(c_i, a_i) + Cov(c_i, b_i) + Cov(c_i, c_i) = 0$$

by symm

$$2\,Cov(c_i, a_i) = -Cov(c_i, c_i)$$

$$Corr(a_i, c_i) = \frac{Cov(a_i, c_i)}{\sqrt{Var(a_i)Var(c_i)}} = \frac{Cov(a_i, c_i)}{Var(c_i)} = -\frac{1}{2}$$

equal

$$\hat{a}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot b_i$$

$$\frac{sCov(a,b)}{sVar(b)} \approx -\frac{1}{2} \qquad \frac{sCov(x,y)}{sVar(x)}$$

$$\hat{u}_a = a - \hat{a} \approx a - \hat{\beta}_1 + \frac{1}{2} b$$

$$\hat{c}_i = \hat{\gamma}_1 + \hat{\gamma}_2 \cdot b_i$$

$$\hat{u}_c = c - \hat{c} \approx c - \hat{\gamma}_1 + \frac{1}{2} b.$$

$(\hat{u}_a + \hat{u}_c) = [a + b + c - const] = const$

$$sCorr(\hat{u}_a, \hat{u}_c) \approx -1$$