بسم الله الرحمن الرحيم

داده‌های طولی و پانلی

ارائه کنندگان:
پویا علی نیان
رضا یار محمدی
محمد مهدی کیانی

فهرست:

۱- معرفی داده‌های طولی
۲- تفاوت داده‌ی طولی با سری و رگرسیون
۳- مثال برای داده‌های طولی
۴- مدل بندی داده‌های طولی
۵- ساختارهای مختلف واریانس کواریانس
۶- مدل با عرض از مبدا تصادفی
۷- مثال با استفاده از نرم افزار spss

داده طولی: داده هایی که در طول زمان و به طور مکرر برای واحد آزمایشی (نفر،کشور و ...) جمع‌آوری می‌شوند.

داده های پنلی: داده هایی که به طور مکرر برای واحد آزمایشی جمع آوری می‌شود.

تفاوت داده های طولی وپنلی: زمان در داده های پنلی مطرح نیست.

داده های طولی و پانلی و تفاوت آن با داده های رگرسیونی و سری زمانی:

در تحلیل رگرسیونی داده ها در یک مقطع از زمان بررسی می‌شوند ولی در تحلیل داده های طولی داده ها در طول زمان بررسی می‌شود.

تفاوت سری زمانی با داده های طولی این است که در داده های سری زمانی متغیر توضیحی مرسوم نیست که به کار برده شود و تعداد مشاهدات و موضوع کم در بازه ی زمانی زیاد مورد بررسی قرار می‌گیرد.

انواع همبستگی ها

۱- تاثیرپذیری از واحد آزمایشی i ام (وابستگی به علت واحد آزمایشی)

۲- همبستگی پیایی (وابستگی به علت زمان)

۱- تاثیرپذیری از واحد آزمایشی i ام

متغیر پاسخ : جمعیت استان

متغیر توضیحی : وسعت استان

واحد آزمایشی: کشورهای جهان

از هر کشور M استان به تصادف انتخاب می‌کنیم

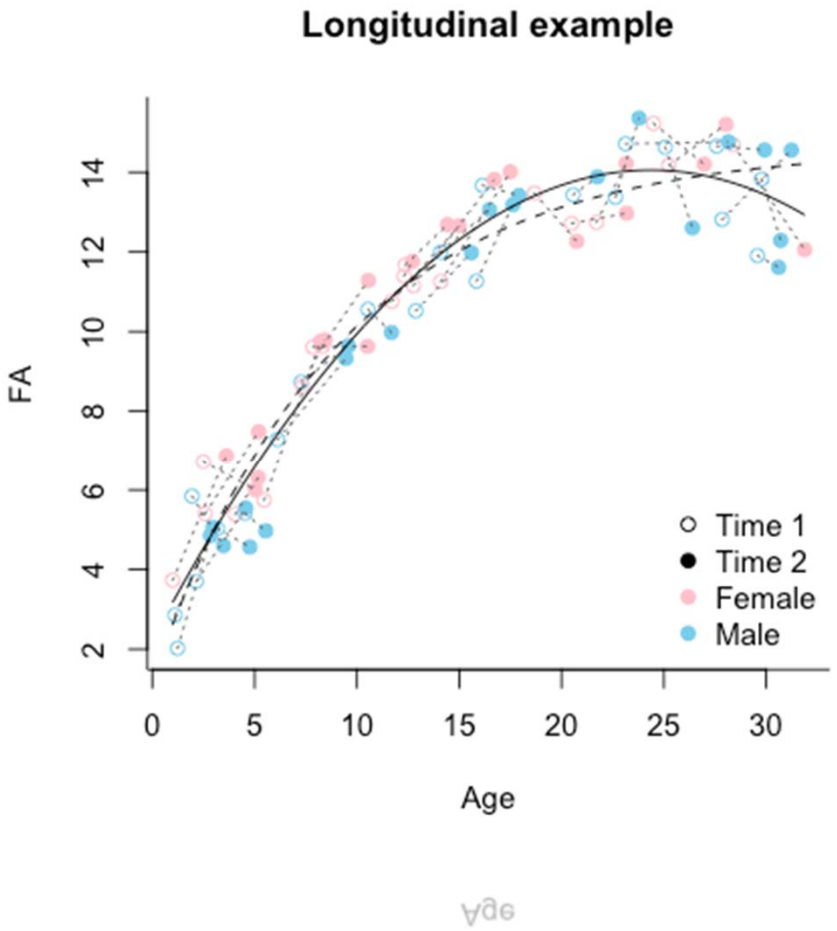
۲- همبستگی پیایی

متغیر پاسخ : رشد اقتصادی

متغیر توضیحی: توسعه‌ی اقتصادی

واحد آزمایشی : کشورهای مجمع جهانی اقتصاد

از هر کشور به مدت T سال رشد و توسعه‌ی اقتصادی آن‌ها را اندازه می‌گیریم



مدل کردن داده‌های طولی

۱) روی هم نریختن داده‌ها

۲) عدم مدل بندی به صورت مدل رگرسیونی ساده به علت برقرار نبودن فرض ثبات واریانس و ناهمبسته بودن خطاها

مدل داده‌های بالانس

$$Y = X\beta + \varepsilon \qquad Y_{nT\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

مشاهدات

زمان

تمام مشاهدات فرد اول در طول زمان (یا به طور کلی واحد آزمایشی)

تمام مشاهدات فرد nام در طول زمان (یا به طور کلی واحد آزمایشی)

$$X_{nT\times(K+1)} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ \vdots & & & \\ 1 & X_{n1} & \cdots & X_{nk} \end{bmatrix}$$

تعداد متغیر های توضیحی

متغیر توضیحی kام برای
فرد nام در طول زمان

مشاهدات فرد n ام در متغیر توضیحی k ام در طول زمان

$$X_{nk} = \begin{bmatrix} X_{nk1} \\ X_{nk2} \\ \vdots \\ X_{nkt} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\varepsilon_i \sim N_T(0, V_i)$$

$$cov(\varepsilon_i, \varepsilon_j) = 0 \qquad \text{for j} \neq i$$

علت وجود همبستگی پیاپی

$$var(\varepsilon_i) = V_i = \begin{bmatrix} var(\varepsilon_{i1}) & cov(\varepsilon_{i1}, \varepsilon_{i2}) & cov(\varepsilon_{i1}, \varepsilon_{it}) \\ cov(\varepsilon_{i2}, \varepsilon_{i1}) & var(\varepsilon_{i2}) & cov(\varepsilon_{i2}, \varepsilon_{it}) \\ cov(\varepsilon_{in}, \varepsilon_{i1}) & cov(\varepsilon_{i1}, \varepsilon_{i2}) & var(\varepsilon_{it}) \end{bmatrix}$$

$$V^* = \mathrm{V}ar(\varepsilon) = \begin{bmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_n \end{bmatrix}$$

$$\varepsilon \sim N_{nT}(0, V^*)$$

تعداد مجهولات معادله در حالت کلی:

$$\frac{n(T * (T+1))}{2} + (K+1)$$

ساختارهای مختلف ماتریس واریانس کواریانس

۱- ساختار نامشخص

ضعف : مشکل برآورد در تعداد تکرار زیاد

-2

$$\mathrm{Cov}(Y_{it},Y_{it'})=0 \quad \forall i,\forall t \neq t' \qquad \mathrm{Var}(Y_{it})=\sigma^2 \qquad \forall i,\forall t$$

بین مشاهدات افراد در طول زمان همبستگی وجود ندارد. (رگرسیون)

compound symmetry -3

یکسان بودن میزان وابستگی فرد i ام در مشاهده‌ی اول با دوم با میزان وابستگی مشاهده ی اول با مشاهده‌ی T ام

مثال کشورها

$$V_i = \mathrm{V}ar(\varepsilon) = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & & \\ \vdots & & \ddots & \vdots \\ \sigma_1^2 & & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ & & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

۴- Autoregressive

وابستگی به فاصله‌ی زمانی مشاهدات افراد بستگی دارد.

مثال رشد و توسعه‌ی اقتصادی

$$V_i = \mathrm{V}ar(\varepsilon) = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho^{T-1} \\ \sigma^2\rho & \sigma^2 & & \\ \vdots & & \ddots & \vdots \\ \sigma^2\rho^{T-1} & & \cdots & \begin{matrix} \sigma^2 & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 \end{matrix} \end{bmatrix}.$$

مدل های با اثرات آمیخته (Mixed effect models)

مدل compound symmetry (تنها وابستگی درون گروهی)

مدل با ماتریس واریانس و کواریانس ۳ را می‌توان به صورت مدل به صورت عرض از مبدا تصادفی بیان کرد.

$$Y_{it} = \beta_{\circ i} + \beta_1 X_{it} + U_{it}$$

$$Y_{it} = \beta_{\circ} + \alpha_i + \beta_1 X_{it} + U_{it} = \beta_0 + \beta_1 X_{it} + \varepsilon_i$$

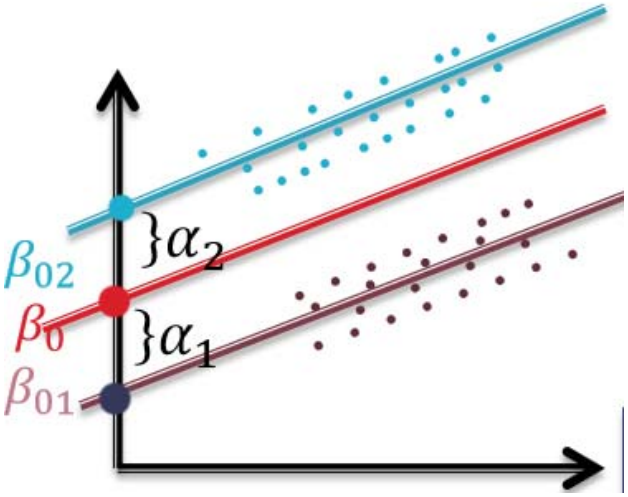
$$U_{it} \sim N(0, \sigma_u^2)$$

indep

$$\alpha_i \sim N(0, \sigma_\alpha^2)$$

خطوط رگرسیونی برای افراد مختلف دارای شیب یکسان و عرض از مبدا متفاوت می‌باشد.

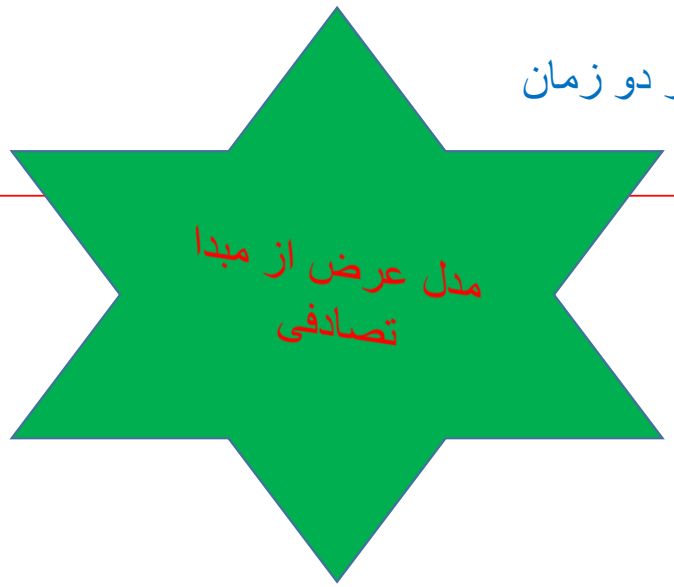
با استفاده از خطوط کلی می‌توان برای فرد جدید اضافه شده به نمونه، پیشبینی کرد.



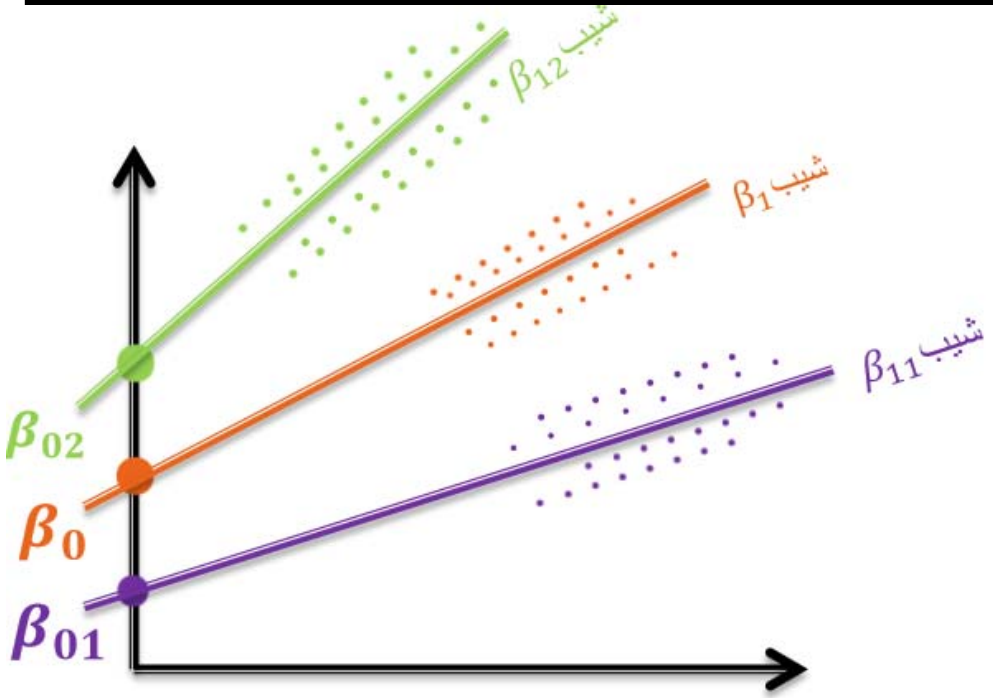
$$Cov(Y_{it},Y_{it'}) = Cov(X_{it}\beta + \alpha_i + u_{it}\ ,\ X_{it'}\beta + \alpha_i + u_{it'}) = \text{Cov}(\alpha_i,\ \alpha_i) + \text{Cov}(\alpha_i,u_{it'})$$

$$+ \text{Cov}(u_{it'},\alpha_i) + Cov(u_{it},u_{it'}) = \sigma_\alpha^2$$

حفظ وجود وابستگی بین پاسخ یک فرد در دو زمان



حالت شیب و عرض از مبدا تصادفی



برآورد پارامترهای مدل به روش MLE

$$\hat{\beta} = \sum_{i=1}^{n} (X_i' V^{-1} X_i)^{-1} X_i V^{-1} Y_i$$

$$\hat{\sigma}_{\alpha}^2 = \frac{T * SSA - \frac{SSE}{T-1}}{nT}$$

$$\hat{\sigma}_{u}^2 = \frac{SST - T * SSA}{nT}$$

References;

1-Zeger, S.C., Liang, K.-Y., and Albert, P.S. (1988) Models for longitudinal data: a generalized estimating equation approach. Biometrics,

2- Verbeke, G. and Lesaffre, E. (1996a) A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association, 91, 217–221.

3- Longitudinal Data Analysis, Handbook (2009)