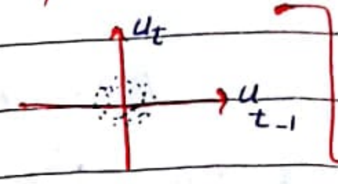
Year : Month : Date :

☑ نقض فرض دوم :

$$E(U_t U_s) = 0 \quad t \neq s$$



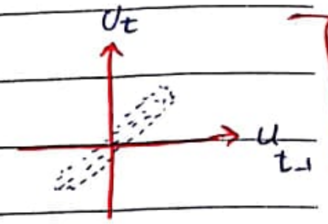
عدم وجود خودهمبستگی (پراکندگی کروی)

یعنی هیچ گونه رابطه ای بین جملات اخلال وجود ندارد

☑ وقتی این فرض نقض میشه یعنی ما دچار مشکل خودهمبستگی هستیم - زمانی وجود می آید که یک متغیر مهم را حذف کنیم.

☑ ماهیت خودهمبستگی چیست ؟

وقتی کواریانس خطاها صفره ما یعنی هیچ رابطهٔ مشخص و سیستماتیک بین خطاها وجود ندارد. یعنی خطای مربوط به دورهٔ t ام هیچ رابطه ای با خطای دیگر ندارد.

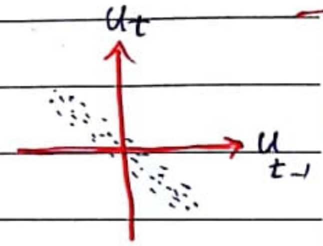


خودهمبستگی مثبت

یعنی یک خطای مثبت پشت سرش یک خطای مثبت است / و یک خطای منفی یک خطای منفی پشت سرش هست

ضریبی برای خودهمبستگی

$$\rho\ (\text{رو}) = \frac{cov(U_t , U_{t-1})}{Var\ (U_t)}$$



$$-1 \leqslant \rho \leqslant +1$$

☑ اگر ρ به سمت عدد یک حرکت کند خودهمبستگی مثبت است

☑ اگر ρ صفر باشد (کواریانس صفر باشد) یعنی خودهمبستگی هیچ رابطه ای بین خطاها وجود ندارد (شکل کروی) یعنی فرض نقض نشده

☑ اگر ρ به سمت -1 حرکت کند خودهمبستگی منفی است.

☑ هرچه به سمت -1 و $+1$ حرکت کنیم مشکل حادتر می شود.

خودهمبستگی منفی

یعنی یک خطای منفی یک خطای مثبت پشت سرش است

☑ ρ دقیقاً مثل $\hat{\beta}$ است.

$$\hat{\beta} = \frac{\Sigma (X_t - \bar{X})(Y_t - \bar{Y})}{\Sigma (X_t - \bar{X})^2}$$

درواقع $\hat{\beta}$ دقیقاً یک ضریب خودهمبستگی است

در β خودهمبستگی بین x و y است و در ρ خودهمبستگی بین خطاها است. ولی تفاوت دراینست که β هر علامتی می تواند باشد اما ρ نمی تواند و فقط بین $+1$ و -1 است

☑ مهمترین دلیل بروز مشکل خودهمبستگی ← حذف یک متغیر مهم از مدل رگرسیون است.

و اینکه اگر بخواهیم متغیری را اضافه کنیم درجه آزادی مدل کاهش پیدا می کند و باعث میشه در محاسبات بعدی با کمبود اطلاعات روبرو شویم و دوران مواجه شویم ... ورودی واریانس اثر می گذارد.

Amin

که اگه مدل‌مون جمله اخلالش میانگینش صفره اما دچار مشکل خودهمبستگی است ولی هم‌چنان واریانس هم ثابت است

فرض اول ادامه — فرض دوم نقض شده — فرض سوم ادامه

در این حالت می‌آییم یک فرآیندی را برقرار می‌کنیم عبارت اخلال ما تابعی از ~~[illegible]~~ گذشته‌اش هست + جمله اخلال خوش رفتار

Subject:
Year: Month: Date:

☑ نقض فرض دوم چه اثری روی ویژگی برآوردگرها می‌گذارد؟

۱) یک مدل رگرسیون داریم که فرض می‌کنیم دچار مشکل خودهمبستگی شده (یعنی خطای ما با گذشته‌اش رابطه دارد مثلاً خطای دوره t با خطای دوره t-1 رابطه داره)

$$E(U_t U_s) \neq 0$$

۲) حالا جمله اخلالی (V_t) را به جمله اضافه می‌کنیم

$$Y_t = \beta_0 + \beta_1 X_t + U_t$$

۳) فرض می‌کنیم خطایی که رخ داده (یا خودهمبستگی که در جملات اخلال وجود دارد) از یک فرآیند (Auto Regressive) تبعیت می‌کند.

یعنی خودش روی خودش رگرس شده ← یعنی خطاش با گذشته خودش رابطه داره

$$U_t = \rho u_{t-1} + V_t$$

جمله اخلال

۴) این اتورگرسیو می‌تواند k مرحله بود (تابعی از یک دوره تا k دوره) حالا (اینجا) فرض می‌کنیم که این اتورگرسیو از مرتبه اول است (برای سادگی) ← هرچه بیشتر شود محاسبات بیشتر می‌شود.

۵) پس فرض می‌کنیم خودهمبستگی که اتفاق افتاده از نوع AR(1) باشد (خودهمبستگی از نوع مارکوف)

خودهمبستگی از نوع مارکوف: فرآیندهایی که مقدار جاری آن متغیر تنها به یک دوره قبلش بستگی دارد.

یعنی فرآیندهای تصادفی که از نوع مارکوف هستند تنها با یک دوره گذشته خودش رابطه دارد پس جملات اخلال هم از این فرآیند تبعیت می‌کند

خوش رفتار است ← یعنی

$$\begin{cases} E(V_t) = 0 \rightarrow \text{یعنی هم‌چنان میانگینش صفر است} \\ E(V_t V_s) = 0 \rightarrow \text{یعنی خودهمبستگی ندارد (دچار مشکل خودهمبستگی نیست)} \\ E(V_t)^2 = \sigma^2 \rightarrow \text{واریانسش با یک مقدار ثابت } \sigma^2 \text{ برابر است.} \end{cases}$$

☑ با چنین شرایطی آیا برآوردگرها هم BLUE هستند یا نه؟؟

خب پس باید امید U_t را حساب کنیم واریانس و کوواریانس اش را نیز حساب کنیم

☑ روش بازگشتی (برای این کار از روش بازگشتی استفاده می‌کنیم)

یعنی هی به عقب برگردیم و از مقادیر گذشته در اینجا جای گذاری کنیم.

$$\begin{cases} U_t = \rho u_{t-1} + V_t \\ U_{t-1} = \rho u_{t-2} + V_{t-1} \\ U_{t-2} = \rho u_{t-3} + V_{t-2} \end{cases} \Longrightarrow \begin{cases} U_t = \rho(\rho u_{t-2} + V_{t-1}) + V_t \\ U_t = \rho^2 U_{t-2} + \rho V_{t-1} + V_t \\ U_t = \rho^2(\rho U_{t-3} + V_{t-2}) + \rho V_{t-1} + V_t \end{cases}$$

$$U_t = \rho^3 U_{t-3} + \rho^2 V_{t-1} + \rho V_{t-1} + V_t$$

$$\boxed{U_t = \rho^s U_{t-s} + \sum_{i=0}^{s-1} \rho^i V_{t-i}}$$ ← شکل کلی جملات اخلال

حالا در صفحه بعد از این شکل کلی جملات اخلال:
① امید ریاضی می‌گیریم
② کواریانس را حساب می‌کنیم
③ واریانس را حساب می‌کنیم

تا بینیم ۳ فرض برقرار بیانه؟

$$u_t = \rho u_{t-s} + \sum_{i=0} \rho^i v_{t-1}$$

Subject :

Year : **Month :** **Date :**

امید ریاضی

$$E(u_t)=0$$

$$E\left(\rho^s u_{t-s} + \sum_{i=1}^{s-1} \rho^i v_{t-1}\right) = 0$$

واریانس

$$Var(u_t) = E\left(\rho u_{t-s} + \sum \rho^i . v_{t-1}\right)^2$$

واریانس یک عدد می شود امید آن عدد خطاهای میانگین اش به توان ۲ که در اینجا میانگین صفر است

$$Var(u_t) = E\left[\rho^{2s} u_{t-s}^2 + v_t^2 + \rho^2 u_{t-1}^2 + \rho^4 v_{t-2}^2 + \dots\right]$$

اگر $S=1$ امید می گیریم

$$= E\left[\rho^2 u_{t-1}^2 + v_t^2 + 2\rho u_{t-1} . v_t\right]$$

خودش

امید این عبارت صفر

$$\sigma_u^2 = \rho^2 \sigma_u^2 + \sigma_v^2 \Rightarrow \sigma_u^2 = \frac{\sigma_v^2}{1-\rho^2}$$

ضریب همبستگی بین خطاها را نشان می دهد

$$\sigma_u^2 - \rho^2 \sigma_u^2 = \sigma_v^2 \Rightarrow$$

$$\sigma_u^2 (1-\rho^2) = \sigma_v^2 \Rightarrow \sigma_u^2 = \frac{\sigma_v^2}{1-\rho^2}$$

هرچه ضریب خودهمبستگی از صفر فاصله بگیره مخرج کوچکتر میشه کسر بزرگتر میشه و واریانس خطاها بزرگتر می شه پس ویژگی حداقل واریانس نقض میشه

کواریانس

$$E(u_t u_{t-1})$$

برای یک دوره حساب می کنیم و تعمیم می دیم به دوره های دیگر

$$E\left[(u_t)(\rho u_{t-2} + v_{t-1})\right]$$

$$E\left[(\rho u_{t-1} + v_t)(\rho v_{t-2} + v_{t-1})\right]$$

چون V_t خوش رفتاره ۰ است

امید صفر است

$$E\left[\rho^2 u_{t-1} v_{t-2} + \rho u_{t-1} v_{t-1} + \rho v_{t-2} v_t + v_t + v_{t-1}\right]$$

$$Cov(u_t u_{t-1}) = \rho^2 Cov(u_{t-1} v_{t-2}) + \rho \sigma_v^2$$

توان ρ به تفاوت $t-2$ و $t-1$ بستگی دارد

$$Cov(u_{t-i}\, u_{t-j}) = \frac{\rho^{i-j} \sigma_v^2}{1-\rho^2}$$

$j-1$

هرچه وقفه ها بیشتر بشه ρ کوچکتر شه

هرچه وقفه بیشتر مشکل ما از بین می رود

نتیجه: اگر $\rho=0$ → کواریانس وجود نداره → مشکل وجود نداره

اگر ρ ماکزیمم (۱) → i وقفه ...

هرچه وقفه های ... $i-j$ ...

۱) نقض فرض دوم (عدم وجود خودهمبستگی)؛ روی ویژگی خطی بودن (بدون تأثیر) است.

۲) نقض فرض دوم روی ویژگی بدون تورش تأثیر ندارد، چون همچنان $E(u)$ صفر است

برآوردگرهای β_0 و β_1 همچنان بدون تورش هستند

۳) نقض فرض دوم روی ویژگی حداقل واریانس:

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

$$Var(\hat{\beta}_0) = E\left[\sum h_t u_t\right]^2$$

چون $Var\,\hat{\beta}_0$ یک مقدار اضافه شده
بنابراین دیگر حداقل نیست و کارا و بهترین نیست.

$$Var(\hat{\beta}_0) = E\left[h_1^2 u_1^2 + h_2^2 u_2^2 + \dots + 2h_1 h_2 u_1 u_2\right]$$

$$\sigma^2\left[\frac{1}{T} - \frac{\bar{X}^2}{\sum x_t^2}\right]$$

$$\frac{\rho\sigma^2 v}{1-\rho^2}$$

واریانس در این صورت برابر این مقدار + یک

مقدار $\frac{\rho\sigma^2 v}{1-\rho^2}$ مثبت دیگر اضافه شده

بنابراین واریانس دیگر حداقل نیست

و کارا نیست

در مورد $\hat{\beta}_1$ هم دقیقاً همین‌طور:

$$Var(\hat{\beta}_1) = E\left[k_1^2 u_1^2 + k_2^2 u_2^2 + \dots + 2k_1 k_2 u_1 u_2\right]$$

(*) وقتی یک مشکل خودهمبستگی موجود است
یک رابطه ای بین خطاها به وجود آمده ... آن ...
طی یک فرآیند آتو رگرسیو ...

Subject :
Year : Month : Date :

حالت نوداری نرمال
نمودار خودهمبستگی مثبت
منفی

۱۳

جملات اخلال ما از یک فرآیند اتورگرسیو از مرتبه اول تبعیت می‌کند

برای تشخیص خودهمبستگی :
- روش ترسیمی ـ نموداری
- (آماره) روش دوربین ـ واتسون
- h-Durbin

آماره دوربین واتسون

$$D\text{-}W = \frac{\sum (U_t - U_{t-1})^2}{\sum_{t=1} U_t^2}$$

اختلاف خطاها از دوره قبل مجذور است سپس جمع می‌شود
مجموع اختلاف جملات اخلال با دوره قبل است

صورت کسر را باز می‌کنیم :

جمله اول به توان ۲ — جمله دوم به توان ۲ — دو برابر ...

$$D\text{-}W = \frac{\sum U_t^2 + \sum U_{t-1}^2 - 2\sum U_t U_{t-1}}{\sum U_t^2}$$

مخرج به توان ۲

فقط یک داده فاصله / تمرکز $\sum U_t$ دارد

فرض می‌کنیم که $\sum U_t^2 \simeq \sum U_{t-1}^2$ در این صورت داریم :

$$D\text{-}W = \frac{2\sum U_t^2 - 2\sum U_t U_{t-1}}{\sum U_t^2} \Rightarrow D\text{-}W = 2 - 2\frac{\sum U_t U_{t-1}}{\sum U_{t-1}^2}$$

شبیه $\hat{\beta}_1 = \frac{\sum x_t y_t}{\sum x_t^2}$ یا همان ضریب همبستگی یا همان ρ (رو) است بنابراین داریم :

(نسبت کوواریانس دو متغیر دارای ...)

آماره دوربین واتسون

$$D\text{-}W = 2 - 2\rho$$

بنابراین نتیجه می‌گیریم که :

$$D\text{-}W \begin{cases} \rho = 1 \text{ اگر} \Rightarrow \text{خودهمبستگی مثبت وجود دارد (شدیدتر)} \Rightarrow D\text{-}W = 0 \\ \rho = 0 \text{ اگر} \Rightarrow \text{خودهمبستگی وجود ندارد} \Rightarrow D\text{-}W = 2 \\ \rho = -1 \text{ اگر} \Rightarrow \text{خودهمبستگی منفی وجود دارد (شدیدتر)} \Rightarrow D\text{-}W = 4 \end{cases}$$

Subject :
Year : Month : Date :

☑ شرایط استفاده از آماره D-W :

۱) در سری‌های داده مفقوده نباید داشته باشیم و (در داده‌های بانک مرکزی هیچ دارای NA نیست)

۲) خودهمبستگی حتماً باید از نوع AR(1) باشد
چنانچه این خودهمبستگی از مراتب بالاتر باشد
نمی‌توانیم از آماره D-W استفاده کنیم

Y_t	X_t	U_t
۲	۲	-۱
۳	۱	۰/۲
۵	NA	
۱	۷	-۰/۵
۴	۶	
۵	۲	۰/۱

۳) Xها یا متغیرهای مستقل حتماً باید ثابت و برون‌زا باشد (یعنی از یک نمونه به نمونه دیگر تغییر نکند)

۴) اگر در معادله رگرسیون

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + U_t$$

⊛ اگر وقتی متغیر وابسته به عنوان یکی از متغیرهای مستقل در معادله رگرسیون ظاهر شود، دیگر نمی‌توانیم از معادله دوربین واتسون (D-W) استفاده کنیم.

☑ روش سوم ← برای تشخیص خودهمبستگی :

آماره (h-Durbin) : برای شرایطی است که وقفه‌ای در معادله رگرسیون ظاهر شده

$$y_t - y_0 = e^{-\lambda_t}$$

تعداد کل مشاهدات

$$h\text{-}Durbin = \rho \sqrt{\frac{T}{1 - Var(\beta_2)}}$$

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + U_t$$

واریانس این β را باید قرار دهیم در

ضریب خودهمبستگی

$\sim AN(0,1)$

نورمال مجانبی

نکته داخل پرانتز از کلاس :
در نظریه همگرایی : اگر موقعیت اولیه کشورها به لحاظ تولید که متفاوت باشند متعادل ها شد چقدر طول می‌کشد که این فاصله از بین برود؟

$\alpha = ٪۵$

$\pm ۱,۹۶$

⊛ وقتی ما وقفه متغیر وابسته را جزء متغیرهای مستقل بیاوریم به خاطر اینکه خودهمبستگی یا رابطه بلندمدت بین متغیرها را نشان بدهیم مثل سوال‌های رشد

Amin

* در واقع ما باید ارتباطی که بین خطاها بوجود آمده را از بین ببریم برای اینکار باید ρ ما یا صفر شود یا خیلی به صفر نزدیک شود

Subject : ۱۴
Year : Month : Date :

☑ روش های رفع مشکل خودهمبستگی :

☑ روش اول : تفاضل مرتبهٔ اول :

یعنی از هر دو طرف معادله یک دوره کم کنیم.
و ما نیاز داریم که $\rho = 1$ باشد.

$$Y_t = \beta_0 + \beta_1 X_t + U_t$$

$$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + U_{t-1}$$

$\rho = 1$

$$(Y_t - Y_{t-1}) = \beta_1 (X_t - X_{t-1}) + (U_t - 1U_{t-1})$$

$$\Delta y_t = \beta_1 \Delta X_t + V_t$$

در روش تفاضل مرتبهٔ اول نیاز به اینست که حتماً ضریب همبستگی صفر باشد.
اگر ضریب همبستگی ۱ باشد می‌توانیم جملهٔ اخلال U را حذف کنیم و به جملهٔ اخلال V برسیم که این جملهٔ اخلال مشکل خودهمبستگی ندارد. اما این فرض یک فرض خیلی سختی است و براحتی نقض میشود.

☑ روش کوکران - اورکات :

① معادله رگرسیون $= \quad Y_t = \beta_0 + \beta_1 X_t + U_t$

جملهٔ اخلال U

- گام اول تخمین می‌زنیم
- گام دوم جملات اخلال را استخراج می‌کنیم (رگرسیون میشه)

$$U_t \to U_{t-1} \Rightarrow \rho$$

- گام سوم جملات اخلال را دوره‌ای قبل رگرسیون می‌کنیم تا ρ آن را بدست می‌آوریم
- یک وقفه می‌دهیم به آنها
- سپس طرفین را در ρ بدست آمده ضرب می‌کنیم

② $\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{t-1} + \rho U_{t-1}$

- حالا دو معادله را از هم کم می‌کنیم: ① و ②

$$(Y_t - \rho Y_t) = \beta_0 (1-\rho) + \beta_1 (X_t - \rho X_{t-1}) + (U_t - \rho U_{t-1})$$

$$Y_t^* = \beta_0^* + \beta_1 X_t^* + V_t$$

(*) در این روش چون ρ قبلش محاسبه شده و β_0 و β_1 قبلاً محاسبه شده، در معادله آخر Y_t^* براحتی مشکل خودهمبستگی برطرف میشه.

Year : Month : Date :

☑ روش سوم : روش کوکران اورکات تکراری ؟

تمام مراحل روش دوم (کوکران ـ اورکات) را بدست می آوریم و $\hat{\beta}_2$ و $\hat{\beta}_1$ را در مرحله آخر جدید تخمین می زنیم و باز هم در مرحله اول قرار می دهیم و مجدداً باز هم مراحل را تکرار می کنیم تا جایی که دیگر اختلافی بین $\hat{\beta}_2$ و $\hat{\beta}_1$ وجود نداشته باشد در این صورت متوجه می شویم که دیگر مشکل خود همبستگی مرتفع شده.

☑ روش GLS ← (ماتریس وزنی در دو طرف معادله ضرب میشود)

مرور جلسات قبل :

هدف اقتصادسنجی اینکه رابطه‌ای (یک) را مورد بررسی قرار دهد. مثلاً رابطه‌ی بین درآمد و مصرف را بررسی می کند (مثلاً نظریه درآمد ـ مصرف کینز را بررسی می کند) با استفاده از آماره‌های ریاضی

معادلاتی داریم که ارتباط بین متغیرها نشان می دهد. کار اقتصادسنجی اینست که ارتباط بین این متغیرها را بتوانند پیدا کنند. یعنی مثلاً بگوییم چه ارتباطی بین درآمد و مصرف وجود دارد

از تکنیک‌های مختلف استفاده می کنیم که پارامترهایی را که در معادله رگرسیون وجود دارند، تخمین بزنیم

وقتی می خواهیم پارامترها را تخمین بزنیم آن معادله را به دو بخش تقسیم می کنیم :

$$Y = \beta_1 + \beta_2 X + u$$

بخش تصادفی ـ بخش سیستماتیک

بخش سیستماتیک رابطه بین X و Y را نشان می دهد و

بخش تصادفی که جمله اخلال بود نماینده تمام آن عوامل غیر شناخته‌هاست

در واقع هر آنچه که قابل شناخت باشد در دل X قرار می دهیم و هر آنچه که غیر قابل محاسبه و شناخت در دل u قرار می دهیم

حال آمدیم معادله رگرسیون را تخمین زدیم تا $\hat{\beta}_1$ و $\hat{\beta}_2$ را بدست آوریم برای اینکار از روشی تحت عنوان روش حداقل مربعات معمولی استفاده می کنیم که فرض عمده این روش این بود که از طریق min کردن مجموع مجذورات خطاها

پارامترهای β_1 و β_2 را تخمین می زنیم و $\hat{\beta}_1$ و $\hat{\beta}_2$ را بدست آوردیم پس گفتیم هدف روش OLS دو موضوع است : (تورش بودن)

1 پارامترهایی را بدست می آوریم که این پارامترها ویژگی‌های خوب داشته باشند BLUE باشد. یعنی بهترین تخمین زننده

حال برای اینکه بفهمیم این برآوردگرها BLUE هستند ۳ فرض باید برقرار باشد ۱) میانگین خطاها در (جامعه) صفر باشد ۲) هیچگونه ارتباطی بین جملات اخلال نباشد ۳) واریانس یک مقدار ثابت که باشد.

2 بتوانیم با استفاده از برآوردگرهایی که بدست آوردیم ، دست به استنتاج آماری بزنیم (یعنی بتوانیم آزمون فرضیه انجام دهیم) که نیاز صدر این بود که با استفاده از آماره T و F معناداری کل رگرسیون را چک کنیم.

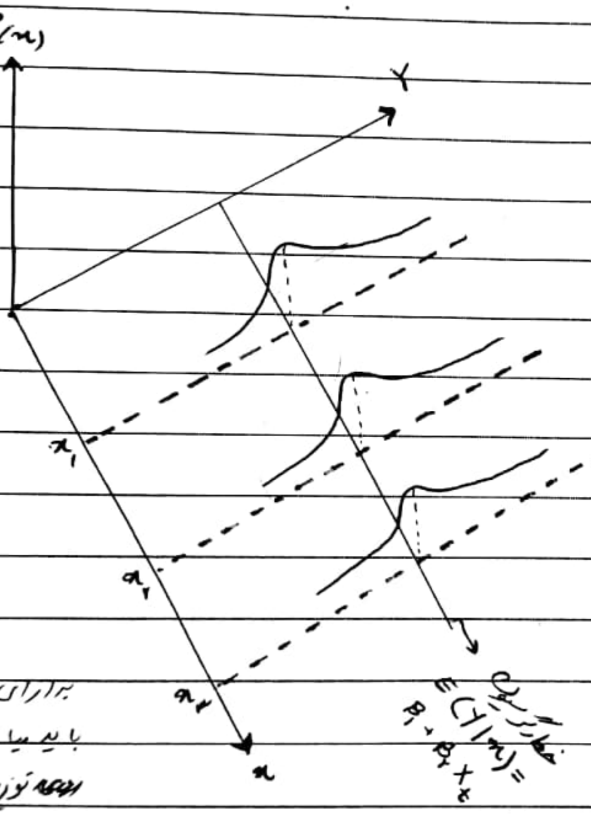
حالا باید چک کنیم ببینیم مدل ما خوب است یا نه ؟ (یعنی مثلاً دستگاهی را که ساخته ایم باید امتحان کنیم و ببینیم که خوب عمل می کند یا خیر؟) گفتیم برای اینکه بدونیم این برآوردگرها مال خوب هستند یا خیر باید ۳ فرض برقرار باشه

حالا فرض ۵ را بررسی می کنیم : اگر فرض اول نقض شود

☑ واریانس همسانی بودن :

یعنی توزیع یا پراکندگی مربوط به خطاها از یک نمونه به نمونه دیگر مقدار ثابتی باشد.
(با تغییری که در متغیر مستقل اتفاق می افتد ، واریانس (یا پراکندگی) متغیر وابسته یا جملات اخلال مقدار ثابت است)
مشاهدات

☑ واریانس ناهمسانی : (هر چه سرعت بالاتر بیشتر شود میزان اشتباه بیشتر میشه)
یعنی با تغییری که در x اتفاق می افتد واریانس خطاها و متغیر وابسته کم و زیاد می شود.
از یک نمونه به نمونه دیگر ، واریانس کم و زیاد میشود
برخلاف مسئله خودهمبستگی که بیشتر در داده های سری زمانی اتفاق می افتد ، واریانس ناهمسانی بیشتر در داده های سری مقطعی رخ می دهد.



به ازای هر x و تغییراتی که در x ها بوجود می آید
توزیع مربوط به جملات اخلال ها و به تبع آن
توزیع متغیر وابسته مان پراکندگی اش ها
یک مقدار ثابت و مشخص است .
و خط رگرسیون ما از میانگین این توزیع ها
عبور می کند .

برابر این ها باید میانگین توزیع ها
مشاهدات باید روی
خط رگرسیون بیفتد

در مسئله واریانس همسانی :
با تغییری که در مشاهدات مربوط به متغیر
مستقل اتفاق می افتد :
میزان واریانس یا پراکندگی خطاها
در متغیر وابسته مان یک مقدار ثابت است

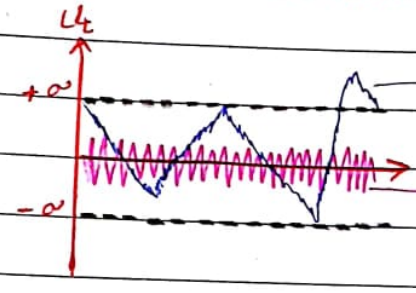
☑ مشکل خودهمبستگی به چیدمان مشاهدات برمی گردد ، درحالیکه مشکل واریانس ناهمسانی به ایجاد کردن داده ها
(فرآیند ایجاد داده ها DGP)

☑ تفاوت ایجاد و چیدمان : در مشکل خودهمبستگی میگوییم که یک خطا با خطای دیگری
رابطه داشته باشد ، فرض میکنیم در امتحان دانشجوهایی که در کنار هم نشستند قاعدتاً
نمره هایشان رابطه ای با هم دارد ← این مشکل خودهمبستگی است چون بستگی به چیدمان افراد دارد
چون اگر جای افراد تغییر کند دیگر نمره ها باهم رابطه ای نخواهند داشت

Month : Date :

در اینجا فرض کن خودآشکار دانشجویانی را که به لحاظ درسی قوی‌تر هستند را کنار هم و دانشجویانی که ضعیف‌تر هستند را کنار هم می‌نشانند، در این صورت پراکندگی بسیار بزرگی اتفاق می‌افتد که عده‌ای نمره‌ای ۲۰ می‌گیرند و عده‌ای نمره ۱۰ ! پس پراکندگی بسیار بالاست ← به این شکل می‌گویند واریانس ناهمسانی یا ایجاد اطلاعات (یعنی با توجه به اطلاعاتی که داده شده سبب این واریانس ناهمسانی شده است)

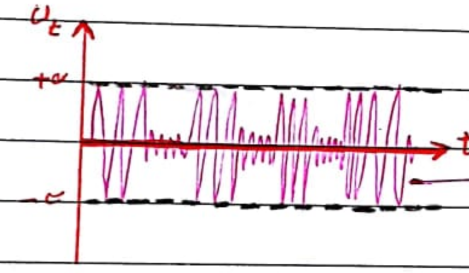
پس تفاوت خودهمبستگی و واریانس همسانی در ایجاد چیدمان داده‌هاست.



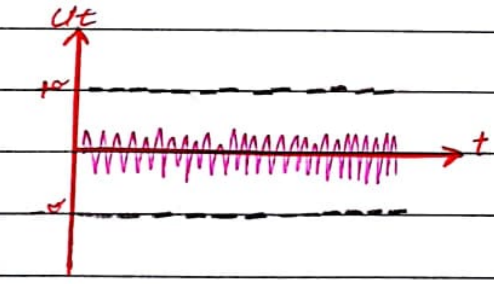
← شکل خودهمبستگی

← عدم وجود خودهمبستگی

$u_t \sim iid(0, \sigma^2)$

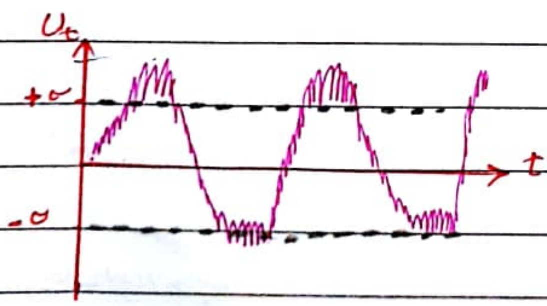


← واریانس ناهمسانی (عدم واریانس همسانی)



۱) عدم خودهمبستگی

۲) واریانس همسانی



۱) خودهمبستگی

۲) واریانس ناهمسانی

تبعات مشکل واریانس ناهمسانی

تبعات مشکل واریانس ناهمسانی: باشد خود چیست؟ یعنی برآوردگرهای روش حداقل مربعات معمولی (OLS) زمانی که دچار مشکل واریانس ناهمسانی هستند همچنان خطی و بدون تورش است ولی واریانس حداقل نیست یعنی ناکارا است.

واریانس ثبات است اول

$$E(U_t)^2 = \sigma^2 \implies$$ هرگاه شکل اینجوری نباشد بیانگر مشکل واریانس ناهمسانی است.

مانند

$$E(U_t)^2 = \sigma^2 \Omega \quad \left\{ \begin{array}{l} \sigma^2 t \\ \text{یا} \\ \dfrac{\sigma^2}{\sqrt{X_t}} \end{array} \right.$$

امگا

واریانس همسانی (یا مشکل عدم واریانس ناهمسانی)

* زمانی که دچار مشکل واریانس ناهمسانی هستیم دو مسئله برای ما مهم است
  1. عامل ایجاد کننده واریانس ناهمسانی
  2. فرم تبعی (شکل) مشکل واریانس ناهمسانی به چه صورت است

* برای فهم بهتر از به جای فرم خطی از فرم ماتریسی استفاده می‌کنیم.

رگرسیون K متغیره ماتریسی:

برداری از متغیر وابسته داریم با T مشاهده

K مشاهده داریم

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1K} \\ X_{21} & X_{22} & \cdots & X_{2K} \\ \vdots & & & \\ X_{T1} & X_{T2} & \cdots & X_{TK} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_T \end{bmatrix}$$

$T \cdot 1$ $T \cdot K$ $K \cdot 1$ $T \cdot 1$

سطر (T) ستون (1)

ماتریس ترانهاده

$$U' = [U_1 \cdots U_T]_{1 \cdot T}$$

$$U \cdot U' = T \cdot T \quad \times$$

$$U' \cdot U = 1 \cdot 1$$

خطی: $Y_t = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + U_t$

ماتریسی: $Y = X\beta + U \longrightarrow U = (Y - X\beta)$ خطاها

* در روش OLS گفتیم که ما باید Min کنیم مجموع مجذورات خطاها را پس در قالب ماتریسی هم همین کار را می‌کنیم. $\Sigma(\hat{u}^2)$

* در قالب خطی می‌گفتیم $\Sigma(\hat{u}^2)$ یعنی دوتا u را در هم ضرب می‌کردیم در قالب ماتریسی هم برای اینکه $1 \cdot 1$ بدست بیاید باید u را در ترانهاده اش ضرب کنیم.

$$\text{Min}(U'U) = (Y - X\beta)'(Y - X\beta)$$

$$= Y'Y - Y'X\beta - X'\beta'Y + X'\beta'X\beta$$

$$= Y'Y - \underbrace{Y X'\beta - Y X'\beta}_{} + \beta X'X\beta$$

$$= Y'Y - 2X'Y\beta + \beta X'X\beta$$

* X ماتریس مربعی و متقارن و دارای رتبه کامل است. (رتبه کامل: یعنی اگر رتبه اش کامل باشد دترمینانش مخالف صفر است) (ماتریسی که تعداد سطر و ستون هایش از هم مستقل باشد)

* β اسکالر (بردار) است. $Y' = Y$ | $X' = X$ | $\beta' = \beta$

مشتق از

$$\frac{\partial(U'U)}{\partial \beta} = 0 \Rightarrow -2X'Y + 2X'X\beta = 0 \Rightarrow \frac{X'Y}{X'X} = \frac{X'X}{X'X}\beta \xrightarrow{(X'X)^{-1}} \boxed{(X'X)^{-1}X'Y = \beta}$$

فرم کلی β چون فرم ماتریسی است

داشتیم $AA^{-1} = I$

$$\hat{\beta}_1 = \frac{\Sigma x_t y_t}{\Sigma x_t^2}$$

Subject :

Year : Month : Date :

$Y = X\beta + U$ $E(U) = 0$

$\beta = (X'X)^{-1}X'Y$ $E(UU') = 0$

$E(U)^2 = \sigma^2$

$E(\hat{\beta}) = \beta \rightarrow$ اگر برقرار باشد یعنی بدون تورش است

می‌خواهیم نشان دهیم که این برآوردگر بدون تورش هست یا نه؟

$\beta = (X'X)^{-1}X'(X\beta + U)$

$\beta = \underbrace{(X'X)^{-1}X'X\beta}_{X'X} + (X'X)^{-1}X'U \rightarrow \frac{(X'X)^{-1}X'X\beta}{X'X}$

$\beta = \beta + (X'X)^{-1}X'U$

امید ریاضی خودش است

از طرفین امید می‌گیریم

$E(\beta) = \beta + (X'X)^{-1}X'\cancel{E(U)}^{0}$

$E(\beta) = \beta \rightsquigarrow$ پس این برآوردگر یک برآوردگر بدون تورش است

بررسی واریانس برآوردگر:

واریانس اش را حساب می‌کنیم

$E(\beta - E(\beta))^2 = E[(X'X)^{-1}X'UU'X(X'X)^{-1}]$

$= E[\underbrace{(X'X)X'X}_{I}(UU')(X'X)^{-1}] = \sigma^2(X'X)^{-1}$

همان $Var(\hat{\beta})$ می‌باشد

$Var(\hat{\beta}) = \frac{\sigma^2}{\sum x_t^2}$

$$\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_T \end{bmatrix} [U_1 \cdots U_T] = \begin{bmatrix} U_1U_1 & U_1U_2 & \cdots & U_1U_T \\ U_2U_1 & \cdots & \cdots & U_2U_T \\ \vdots & & & \\ U_TU_1 & & & U_TU_T \end{bmatrix} = \underbrace{\begin{bmatrix} \sigma^2 & 0 & 0 & \cdot \\ 0 & \sigma^2 & & \\ 0 & & \ddots & \\ 0 & & & \\ 0 & \cdots & & \sigma^2 \end{bmatrix}}_{\text{اگر از کل ماتریس فاکتور بگیریم میشه}}$$

U_1U_1, U_2U_2, U_3U_3 این همان (UU') $\sigma^2 I$

همان σ^2 هستند

یعنی اگر امید UU' را بگیریم میشه $\sigma^2 I$ که I را نمی‌نویسیم.

Scanned by CamScanner

$$E(UU') = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \sigma^2 I$$

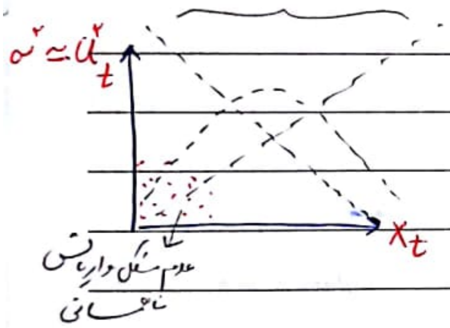
اگر قطر اصلی σ^2 نباشد
و بقیه مقادیر نیز صفر نباشد
علاوه بر مشکل واریانس ناهمسانی
مشکل خودهمبستگی هم داریم.
یعنی اگر بقیه عناصر فرعی صفر نباشد
مشکل خودهمبستگی هم بوجود می آید.

$$\Omega = \begin{bmatrix} \frac{1}{\lambda_i} & 0 & 0 & 0 \\ 0 & \frac{1}{\lambda_i} & 0 & 0 \\ 0 & 0 & \frac{1}{\lambda_i} & 0 \\ 0 & 0 & 0 & \frac{1}{\lambda_i} \end{bmatrix} \qquad \lambda_i \neq 0$$

روش های تشخیص واریانس ناهمسانی

1) روش ترسیمی
2) روش پارک
3) روش گلد فلد-کوانت
4) روش آزمون وایت

این شکل ها مشکل واریانس ناهمسانی دارند



1) روش ترسیمی:

اگر u را به توان ۲ برسانیم و در مقابل X ها رسم کنیم
نقطه های پرکنش اگر به صورت محدوده عدم مشکل واریانس ناهمسانی میشود.
و اگر هر شکلی غیر از مورد بالایی باشد یعنی مشکل واریانس ناهمسانی داریم

2) آزمون پارک:

$\sigma_i^2 = \sigma^2 X_t^{\alpha}$ → عامل ایجاد واریانس ناهمسانی

از طرفین لگاریتم میگیریم $\Rightarrow Ln(\sigma_i^2) = Ln(\sigma^2) + \alpha Ln(X_t)$

اگر این معادله را تخمین بزنیم و α به لحاظ آماری معنادار باشد یعنی اگر مقدار آماره t اش از ۲ بزرگتر باشد

در این صورت بین X و σ^2 رابطه وجود دارد ← پس دچار مشکل واریانس ناهمسانی شده ایم

Subject :
Year : Month : Date :

۳) آزمون گلدفلد ـ کوانت :

مشکل داریم که کدام یک از متغیرها (Xها) موجب مشکل واریانس ناهمسانی شده باید تشخیص بدهیم که کدام X بوده پس

$Y_t = \beta_1+\beta_2 X_t + u_t$ نمونه کوچک

$Y_t = \beta_1+\beta_2 X_t + u_t$ نمونه بزرگ

مجموع مجذورات خطاها را حساب می کنیم

$\sum \hat{u}_t^2 = RSS_1$

$\sum \hat{u}_t^2 = RSS_2$

☑ گام اول: حدس می زنیم که کدام X عامل ایجاد مشکل بوده

☑ گام دوم: تمامی مشاهدات را برحسب آن عامل از کوچک به بزرگ مرتب می کنیم

☑ گام سوم: تعداد c مشاهده را حذف می کنیم. مثلاً ۱۰۰تا مشاهده داریم از وسط ۲۰ تا شو حذف می کنیم.

☑ حالا دو نمونه داریم. نمونه ای که دارای اعداد بزرگ است و نمونه ای که دارای اعداد کوچک است.

$$\lambda = \frac{RSS_2}{RSS_1}$$

دارای توزیع کای دو χ^2 است

چون حین صورت و مخرجش از واریانس است

☑ گام چهارم: این دو نمونه را با دو رگرسیون تخمین می زنیم پس مجموع مجذور خطاهایشان را حساب می کنیم و اینکه قطعاً خطای نمونه بزرگ، بزرگتر است

☑ λ را حساب می کنیم. هرچه قدر λ از عدد ۱ بزرگتر باشد یعنی واریانسشان متفاوت است (RSSها متفاوت است) (خطاهایشان متفاوت است) پس یعنی نمونه ها توی ناهمسانی مختلف داشتند و ما X را درست تشخیص دادیم و مشکل واریانس ناهمسانی داریم.

و هرچقدر λ به ۱ یا کوچکتر از ۱ نزدیک شود یعنی واریانس ناهمسانی اتفاق نیفتاده (تنها ایراد در انتخاب X ها بوده است)

۴) آزمون وایت :

$\sigma_t^2 \sim X_1 X_2 \quad X_1^2 X_2^2 \quad X_1 X_2$

برای تشخیص مشکل واریانس ناهمسانی باید:

$$\sigma_t^2 = \alpha + \beta_1X_1+\beta_2X_2+\beta_3X_1^2+\beta_4X_2^2+\beta_5X_1X_2$$

u^2 را روی خود متغیرها و حاصل ضربشان (توان دوم آن) رگرسیون می کنیم

$R^2 \to T \cdot R^2 \sim \chi^2 >$ مقدار بحرانی جدول

T: تعداد مشاهدات — χ^2: ۲ یا ۵

هرکدام از این ضرایب که معنادار به دست بیاید (به لحاظ آماری) می تواند موجب مشکل واریانس ناهمسانی شود. حالا از این رگرسیون R^2 به دست می آید.

حالا تعداد مشاهدات را در R^2 ضرب می کنیم حالا این دارای توزیع χ^2 است حالا اگر این مقدار از مقدار بحرانی جدول بزرگتر بدست بیاید مشکل واریانس ناهمسانی داریم

فرض کنید که با مشکل واریانس ناهمسانی روبرو شده‌ایم. واریانس جملات اخلال ما به صورت X_t شده. یعنی با تغییری که در X اتفاق می‌افتد

Subject :
Year : Month : Date :

واریانس جملات اخلال هر چه زیادتر شود باید به I تبدیل اش کنیم.

روش‌های رفع مشکل واریانس ناهمسانی :

ماتریسی $Y = X\beta + u$

خطی $Y_t = \beta_0 + \beta_1 X_t + u_t$

اگر به هر دلیلی واریانس جملات اخلال ما به جای $\sigma^2 I$ به Ω تبدیل شود (که این Ω می‌تواند هر ماتریسی باشد) باید کاری کنیم (عملیاتی انجام دهیم) تا تأثیر آن عامل مزاحم از بین برود. یعنی ماتریس Ω به ماتریس I تبدیل شود، که عناصر روی قطر اصلی اش یک عدد ثابت شود.

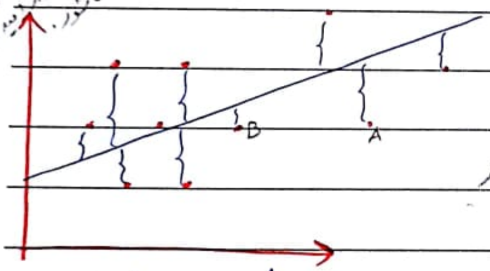
واریانس جملات اخلال $E(u_t)^2 = \sigma^2 X_t \rightarrow \sigma^2 I$

$$\frac{Y_t}{\sqrt{X_t}} = \frac{\beta_0}{\sqrt{X_t}} + \frac{\beta_1 X_t}{\sqrt{X_t}} + \frac{u_t}{\sqrt{X_t}} \Rightarrow \text{Min} \sum u_t^2 = \sum \left(\frac{Y_t}{\sqrt{X_t}} - \frac{\beta_0}{\sqrt{X_t}} - \frac{\beta_1 X_t}{\sqrt{X_t}} \right)^2$$

$$E\left(\frac{u_t}{\sqrt{X_t}}\right)^2 = E\left(\frac{u_t^2}{X_t}\right) = \frac{E(u_t)^2}{X_t} = \frac{\sigma^2 \cancel{X_t}}{\cancel{X_t}} = \sigma^2$$

(*) امید ریاضی نقطه به مقادیر تصادفی جواب نمی‌گیرد و امید یک عدد ثابت خودش می‌شود.

جمله تصادفی است پس باید به آن توجه کرد و انتظار ریاضی گرفته شود.



فلسفه این روش :

در روش OLS برای ما خطاهای A و B مثل هم بودن اما وقتی مشکل واریانس ناهمسانی داریم، خطاهای دور از خط رگرسیون ما را دچار مشکل می‌کنند پس باید با توجه به اینکه خطا چه مقدار در خط رگرسیون اهمیت دارد به آن وزن می‌دهیم.

$OLS = \text{Min} \sum \hat{u}_t$

$Y = X\beta + u$

$Y_t = \beta_0 + \beta_1 X_t + u_t$

$$E(u_t)^2 = \frac{\sigma^2}{t}$$

$$Y_t \times \sqrt{t} = \beta_0 \cdot \sqrt{t} + \beta_1 X_t \cdot \sqrt{t} + u_t \cdot \sqrt{t}$$

$$E(u_t \cdot \sqrt{t})^2 = E(u_t^2 \cdot t) = E(u_t^2) \cdot t = \frac{\sigma^2}{t} \cdot t = \sigma^2$$

زمانی که از مدل رگرسیون ۲ متغیره به یک مدل رگرسیون چند متغیره تغییر می‌دهیم مدلمان را، یک نکتهٔ اساسی که وجود دارد این است که متغیرهای مستقل باید کاملاً از هم مستقل باشند یا به عبارتی نباید هیچ گونه رابطهٔ خطی با هم داشته باشند

Subject :

Year : Month : Date :

☑ مسئله هم‌خطی: Xها (متغیرهای مستقل با هم رابطهٔ خطی داشته باشند)

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

← یه چیزی تابعی از چیز دیگر

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

X_{1t} و X_{2t} (متغیرهای مستقل) باید مستقل باشند و هیچ‌گونه رابطهٔ خطی نداشته باشند.

☑ استقلال خطی و وابستگی خطی

استقلال خطی / (در بحث بردارها (ماتریس‌ها) داشتیم)

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0$$

if $\lambda_i \neq 0$ → اگر یکی از λهای بالایی ما حداقل صفر باشد و ترکیبات بالا مساوی صفر باشد می‌گوییم که دارای استقلال خطی هستند.

دارای استقلال خطی

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots \lambda_n X_n \neq 0 \Rightarrow$$ دارای وابستگی خطی

۱) هم‌خطی کامل: $X_2 = 2X_1$ (ضریبی از دیگری)

۲) هم‌خطی ناقص: $X_t = \delta_s X_s + \varepsilon$

ضریبی از X — عدد تصادفی

X_t ضریبی از X_s است به اضافه یک مقدار تصادفی

فرض می‌کنیم if $X_2 = \delta X_1$

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_2 + u_t \Longrightarrow Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 \delta X_{1t} + u_t$$

if $X_2 = \delta X_1$

فاکتور

$$\Rightarrow Y_t = \beta_0 + \underbrace{(\beta_1 + \beta_2 \delta)} X_{1t} + u_t \Longrightarrow \boxed{Y_t = \beta_0^* + \beta_1^* X_{1t} + u_t}$$

تخمین

مشکل ما این است که چقدر از β_1^* چقدرش β_1 و چقدر β_2 وجود دارد؟!

مشکلی که هم‌خطی برای ما ایجاد می‌کنه ← در مقابل تفکیک اثرات دچار مشکل می‌شویم.

* چیزی که برای ما دردسرساز است ← هم‌خطی کامل است.

وقتی مشکل هم‌خطی رخ بدهد ← ما در تشخیص برآوردگرها دچار مشکل می‌شویم

β ای که در فرم ماتریسی بدست آوردیم

اگر همین مسئله صفحه قبل را بخواهیم در فرم ماتریسی نشان بدهیم

*

Subject :

Year : Month : Date :

$$\beta = (X'X)^{-1} X'Y$$

زمانی که می خواستیم معکوس ماتریس را حساب کنیم شرط اصلی آن این است که det مخالف صفر باشد به زمانی $\neq$ det 0 میشود که تمام عناصر سطر و ستون ها مستقل از هم باشند.

$$\text{ماتریس } X \quad \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1T} \\ X_{21} & X_{22} & \cdots & X_{2T} \\ \vdots & & & \\ X_{T1} & \cdots & \cdots & X_{TT} \end{bmatrix}$$

$$A^{-1} = \frac{1}{|det|} \cdot cof$$

در شکل هم خطی به یک سطر یا یک ستون تابعی از دیگری است → در این صورت det صفر میشه یعنی معکوس نمیشه

دیگه β قابل محاسبه نیست

۲) زمانی که مسئله هم خطی بوجود می آید بحث اصلی روی واریانس برآوردگرها است

r_{ij}^2 → ضریب همبستگی بین آن دو متغیری است که ما شک کردیم که مشکل هم خطی داشته باشند

پس تا محدوده همبستگی شون 1+ تا 1- میشه. هر چه این دو متغیر هم خطی بالا داشته باشند هم بستگی ↑ → مخرج ↓ → کل کسر ↑

$$Var(\hat{\beta}_i) = \frac{\sigma^2_{\beta}}{1 - r^2_{ij}}$$

واریانس ↑

واریانس نا یعنی خطا یا ناکارآمد بودن

زمانی که دچار مشکل هم خطی می شویم به دلیل اینکه آن دو متغیر مورد نظر هم بستگی بالایی با یکدیگر پیدا می کنند باعث می شود انحراف معیار برآوردگر β افزایش پیدا کنند در این صورت

هرچه var ↑ → برآوردگرها → خطا ↑ استفاده از آماره ها کم تر می شوند

آماره های T و F اعتبار نخواهد داشت.

آماره T $\frac{\hat{\beta}}{se\hat{\beta}}$ اگر واریانس زیاد شه یعنی انحراف معیار هم ↑ پس کل عبارت ↓

$$t = \frac{\hat{\beta}}{se(\hat{\beta})} > \pm 1{,}96 \quad \text{رد } H_0$$

↓ یعنی آماره T اش مارو گمراه می کنه

(*) زمانی که مسئله هم خطی بوجود می آید دو تا مشکل داریم:

① برآوردگرها تمایل محاسبه نیستند

② انحراف معیار برآوردگرها افزایش پیدا می کند

③ استفاده از آماره T و F اعتبار خودشان را از دست می دهند

روش های تخمین مسئله هم خطی:

۱ روش اول: ضریب هم بستگی مرتبه صفر

در این روش دو به دو روی متغیرها روی هم دیگر رگرس میشوند ضریب هم بستگی آن دو متغیر را با مجذور متغیرهای مستقل

R^2 (r_{ij}^2) و $\sqrt{R^2}$ (r_{ij}) مقایسه می کنیم → اگر ضریب هم بستگی یا مجذورش بیشتر از ضریب تعیین یا جذرش باشد میگوییم هم خطی شدید به وجود آمده است.

Subject :
Year : Month : Date :

روش سوم : R^2 جزئی :

متغیر وابسته — متغیر مستقل

فرض کنیم که یک متغیر وابسته Y و سه تا متغیر مستقل X_1 و X_2 و X_3 داریم اگر Y را روی

$Y \overset{رگرس}{\sim} X_1 X_2 X_3 \rightarrow R^2_T$

آن‌ها رگرس کنیم یک ضریب تعیین R^2_T به ما می‌دهد

1) $Y \sim X_1 X_2 \rightarrow R^2_{12}$

حالا هر بار یکی از X ها را حذف می‌کنیم و سپس عمل رگرسی را

2) $Y \sim X_1 X_3 \rightarrow R^2_{13}$

انجام می‌دهیم و هر بار یک ضریب تعیین به ما می‌دهد.

3) $Y \sim X_2 X_3 \rightarrow R^2_{23}$

☑ هرچه R^2 ناشی از حذف یک متغیر با R^2 اصلی به هم نزدیک باشند یعنی می‌تونیم بفهمیم که اونی که حذف شده با بقیه هم‌خطی داشته.

روش چهارم : آزمون فاروس - گلوبر

X ها k تا داریم

$\frac{R^2_{kx}}{R^2_Y} > 1$

هر بار یکی از X ها را برمی‌داریم و رگرس می‌کنیم روی بقیه X ها و یک R^2 به ما می‌دهد

یعنی Y رگرس شده روی تمامی X ها ← خود رگرسیون

چنانچه آماره این آزمون عدد بزرگتر از ۱ را نشان دهد یعنی آن x_k که بر روی سایر متغیرهای مستقل رگرس شده دارای هم‌خطی با سایر X ها بوده و می‌تونیم برای رفع مشکل هم‌خطی باید آن را حذف کنیم

$X_1 \sim X_2 X_3 \cdots X_k \rightarrow R^2$

$X_2 \sim X_1 X_3 \cdots X_k \rightarrow R^2$

$X_k \sim X_1 X_2 \cdots X_{k-1} \rightarrow R^2_{kx}$

اگر $R^2_{kx} > R^2_Y$ باشه ⇐ $\frac{R^2_{kx}}{R^2_Y}$ بزرگتر از یک می‌شود

☑ اگر R^2 (ناشی از رگرسیون X بر روی بقیه متغیرها (یعنی صورت)) بزرگتر از R^2 کل رگرسیون (یعنی مخرج) باشد نشان می‌دهد که اون با بقیه رابطه هم‌خطی داشته و باید حذفش کنیم.

مهمترین روش : VIF

$VIF = \frac{1}{1-R^2_k}$ از ۱ همواره بیشتر

ضریب تعیین ناشی از رگرسیون X_k روی سایر متغیرها.

هرچه قدر R^2 عدد بزرگتری را نشان دهد یعنی X ها

هرچه ~~[illegible]~~ R^2 عدد ۱ یا کوچکتر را نشان دهد یعنی هم‌خطی اتفاق نیفتاده.

بین ۱ تا ۱۰ ← تقریباً هم‌خطی وجود ندارد

بین ۱۰ تا ۳۰ ← هم‌خطی ناقص

از ۳۰ بالاتر ← هم‌خطی شدید

Amfa

Subject :

Year : Month : Date :

☑ روش های رفع مشکل هم‌خطی :

۱) استفاده از حالت لگاریتمی م.ت ها

~~مسئله هم‌خطی معمولاً در نمونه رخ می‌دهد و می‌توانیم با تغییر در نمونه مشکل هم‌خطی از بین برود.~~

۲) تغییر در نمونه (یا به عبارتی حجم م.ت ها را کم یا زیاد کنیم)

۳) Ridge Regression → حاشیه دار کردن / لبه دار کردن

$$\min \sum (Y - XB) + \lambda_i \sum B_j$$

یه جزء یا عبارتی است که اضافه میشه

زمانی که داریم مدل رگرسیونی را min می‌کنیم جزئی را به آن اضافه می‌کند مانند (٤) که اضافه می‌کنیم یک مقدار کوچکی را به م.ت ها که باعث می‌شه که تقریباً شبیه هم‌خطی ناقص شده که مشکل رفع می‌کند