

Delluaðhvarf=spurious-regression

Helgi Tómasson

18. september 2014

Sjónhverfing: Enska biskupakirjan stórhættuleg



Sjónhverfing: Enska biskupakirjan stórhættuleg

- Englendingar hafa langa menningarsögu

Sjónhverfing: Enska biskupakirjan stórhættuleg

- Englendingar hafa langa menningarsögu
- Virðulegar gamlar stofnanir, til dæmis landlæknisembætti og Ensku kirkjuna.

Sjónhverfing: Enska biskupakirjan stórhættuleg

- Englendingar hafa langa menningarsögu
- Virðulegar gamlar stofnanir, til dæmis landlæknisembætti og Ensku kirkjuna.
- Þessar stofnanir safna vönduðum mælingum.

Sjónhverfing: Enska biskupakirjan stórhættuleg

- Englendingar hafa langa menningarsögu
- Virðulegar gamlar stofnanir, til dæmis landlæknisembætti og Ensku kirkjuna.
- Þessar stofnanir safna vönduðum mælingum.
- Til dæmis hefur landlæknisembættið safna gögnum um „mortalitet“, dánir á ári per 1000 íbúa og kirkjan hefur talið hver mörg hjónabönd á ári per 1000 hjónabönd eru á vegum kirkjunnar.

Sjónhverfing: Enska biskupakirjan stórhættuleg

- Englendingar hafa langa menningarsögu
- Virðulegar gamlar stofnanir, til dæmis landlæknisembætti og Ensku kirkjuna.
- Þessar stofnanir safna vönduðum mælingum.
- Til dæmis hefur landlæknisembættið safna gögnum um „mortalitet“, dánir á ári per 1000 íbúa og kirkjan hefur talið hver mörg hjónabönd á ári per 1000 hjónabönd eru á vegum kirkjunnar.
- Hver skyldu tengsl þessara hagstærða vera?

Sjónhverfing: Enska biskupakirjan stórhættuleg

- Englendingar hafa langa menningarsögu
- Virðulegar gamlar stofnanir, til dæmis landlæknisembætti og Ensku kirkjuna.
- Þessar stofnanir safna vönduðum mælingum.
- Til dæmis hefur landlæknisembættið safna gögnum um „mortalitet“, dánir á ári per 1000 íbúa og kirkjan hefur talið hver mörg hjónabönd á ári per 1000 hjónabönd eru á vegum kirkjunnar.
- Hver skyldu tengsl þessara hagstærða vera?
- Reiknum fylgnistuðul og fáum:

Sjónhverfing: Enska biskupakirjan stórhættuleg

- Englendingar hafa langa menningarsögu
- Virðulegar gamlar stofnanir, til dæmis landlæknisembætti og Ensku kirkjuna.
- Þessar stofnanir safna vönduðum mælingum.
- Til dæmis hefur landlæknisembættið safna gögnum um „mortalitet“, dánir á ári per 1000 íbúa og kirkjan hefur talið hver mörg hjónabönd á ári per 1000 hjónabönd eru á vegum kirkjunnar.
- Hver skyldu tengsl þessara hagstærða vera?
- Reiknum fylgnistuðul og fáum:
Svar: 0.95

Sjónhverfing: Enska biskupakirjan stórhættuleg

- Engendingar hafa langa menningarsögu
- Virðulegar gamlar stofnanir, til dæmis landlæknisembætti og Ensku kirkjuna.
- Þessar stofnanir safna vönduðum mælingum.
- Til dæmis hefur landlæknisembættið safna gögnum um „mortalitet“, dánir á ári per 1000 íbúa og kirkjan hefur talið hver mörg hjónabönd á ári per 1000 hjónabönd eru á vegum kirkjunnar.
- Hver skyldu tengsl þessara hagstærða vera?
- Reiknum fylgnistuðul og fáum:
Svar: 0.95
- Ályktum, **sterkt samband** eða jafnvel **sterkt marktækt samband** milli markaðshlutdeildar kirkjunnar í brúðkaupum og „mortalitets“.

Varúð:

Hugtakið marktækt hefur ekkert með mikilvægi að gera.

Varúð:

Hugtakið marktækt hefur ekkert með mikilvægi að gera.

Marktækt þýðir eingöngu: $p < \alpha$, þar sem p þýðir líkur á fenginni útkomu eða ótrúlegri gefið að tiltekið tölfræðilegt líkan sé rétt.

Varúð:

Hugtakið marktækt hefur ekkert með mikilvægi að gera.

Marktækt þýðir eingöngu: $p < \alpha$, þar sem p þýðir líkur á fenginni útkomu eða ótrúlegri gefið að tiltekið tölfræðilegt líkan sé rétt.

Ekkert vit í gagnagreiningu (ályktunum) án tölfræðilegs líkans. Ég tel fullt vit í reikningshaldi/bókhaldi/bókasafnsfræði og annari skráningu staðreynda.

Varúð:

Hugtakið marktækt hefur ekkert með mikilvægi að gera.

Marktækt þýðir eingöngu: $p < \alpha$, þar sem p þýðir líkur á fenginni útkomu eða ótrúlegri gefið að tiltekið tölfræðilegt líkan sé rétt.

Ekkert vit í gagnagreiningu (ályktunum) án tölfræðilegs líkans. Ég tel fullt vit í reikningshaldi/bókhaldi/bókasafnsfræði og annari skráningu staðreynda.

Í þessum útreikningi hefur það gleymst að gögnin eru tímaraðir og því kallar tölfræðileg greining á tímaraðalíkön/tímaraðaaðferðir

Varúð:

Hugtakið marktækt hefur ekkert með mikilvægi að gera.

Marktækt þýðir eingöngu: $p < \alpha$, þar sem p þýðir líkur á fenginni útkomu eða ótrúlegri gefið að tiltekið tölfræðilegt líkan sé rétt.

Ekkert vit í gagnagreiningu (ályktunum) án tölfræðilegs líkans. Ég tel fullt vit í reikningshaldi/bókhaldi/bókasafnsfræði og annari skráningu staðreynda.

Í þessum útreikningi hefur það gleymst að gögnin eru tímaraðir og því kallar tölfræðileg greining á tímaraðalíkon/tímaraðaaðferðir

Mjög mikið af haggögnum eru tímaraðir.

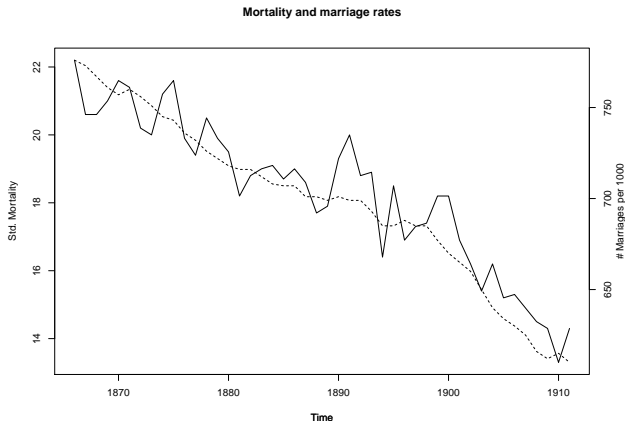
Tímaraðir eru mælingar á hreyfimyngstri(dynamic system)

Tímaraðir eru mælingar á hreyfimyngstri(dynamic system)

Það verður að taka tillit hreyfimyngstursins í allri gagnagreiningu

Motivating example

- Yule (1926): Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series.



Mynd : Yule(1926) data on mortality and marriages.



- Í ýmsum vísindum myndi fólk segja, marktæk fylgni

$$r = 0.95, \quad p < 0.001^{***}$$

eða jafnvel bara marktækt samband milli giftingar í kirkju og mortalitets ($p < 0.001$).

- Í ýmsum vísindum myndi fólk segja, marktæk fylgni

$$r = 0.95, \quad p < 0.001^{***}$$

eða jafnvel bara marktækt samband milli giftingar í kirkju og mortalitets ($p < 0.001$).

- Þetta er að sjálfsögðu galið því líkanið er ekki tilgreint.

- Í ýmsum vísindum myndi fólk segja, marktæk fylgni

$$r = 0.95, \quad p < 0.001^{***}$$

eða jafnvel bara marktækt samband milli giftingar í kirkju og mortalitets ($p < 0.001$).

- Þetta er að sjálfsögðu galið því líkanið er ekki tilgreint.
- Hér hefur verið metið líkanið:

$$y_t = \mu + \beta x_t + \varepsilon_t$$

- Í ýmsum vísindum myndi fólk segja, marktæk fylgni

$$r = 0.95, \quad p < 0.001^{***}$$

eða jafnvel bara marktækt samband milli giftingar í kirkju og mortalitets ($p < 0.001$).

- Þetta er að sjálfsögðu galið því líkanið er ekki tilgreint.
- Hér hefur verið metið líkanið:

$$y_t = \mu + \beta x_t + \varepsilon_t$$

- Og enginn tímastrúkstúr.

- Í ýmsum vísindum myndi fólk segja, marktæk fylgni

$$r = 0.95, \quad p < 0.001^{***}$$

eða jafnvel bara marktækt samband milli giftingar í kirkju og mortalitets ($p < 0.001$).

- Þetta er að sjálfsögðu galið því líkanið er ekki tilgreint.
- Hér hefur verið metið líkanið:

$$y_t = \mu + \beta x_t + \varepsilon_t$$

- Og enginn tímastrúkstúr.
- Granger & Newbold (1974) kynntu hugmyndir Yule og markaðsettu í ekonómétríunni.



- Granger & Newbold (1974) Rifjuð upp reglu Bartlett (1946) sem segir að varíans úrtaksfylgni tveggja óháðra stationary ferla er flókið fall af sjálfygnifalli þeirra. Ef sjálfygnin er mikil og röðin stutt getur dreifing fylgnistuðulsins orðið U-laga.

- Granger & Newbold (1974) Rifjuð upp reglu Bartlett (1946) sem segir að varíans úrtaksfylgni tveggja óháðra stationary ferla er flókið fall af sjálfygnifalli þeirra. Ef sjálfygnin er mikil og röðin stutt getur dreifing fylgnistuðulsins orðið U-laga.
- Dæmi sem ég hef látið nemendur gera undanfarin 20 ár. Hermið X_t og Y_t , þ.a.

$$X_t = \phi X_{t-1} + \varepsilon_t$$

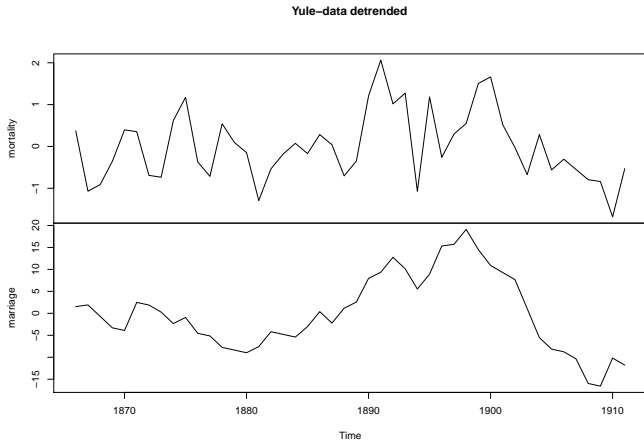
$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t, \varepsilon_t \text{ óháðir white-noise}$$

Hafið $\phi = 0.95 - 1$ og skoðið dreifingu r =úrtaksfylgni X_t og Y_t . Athugið hversu oft kenningunni H_0 :Engin fylgni, er hafnað. Útkoman: Kenningunni er hafnað alltof oft.

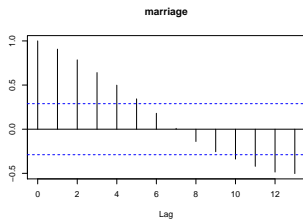
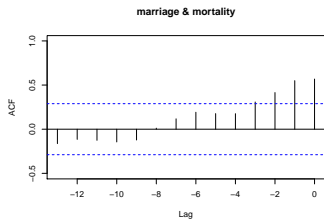
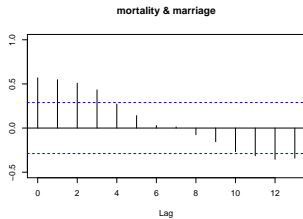
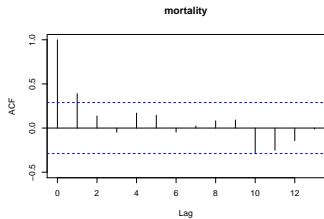
Hvað segir hefðbundin tímaraðagreiningu um gögn Yule?

- 1 Fjarlægið trend/season og aðra deterministíska þætti.
- 2 Skoðið pre-whitened útgáfu af röðunum.
- 3 Skoðið cross-correlation/cross-spectur af prewhitened röðum.
- 4 Metið hugsanlegt transfer-function fall til að lýsa sambandi þeirra.

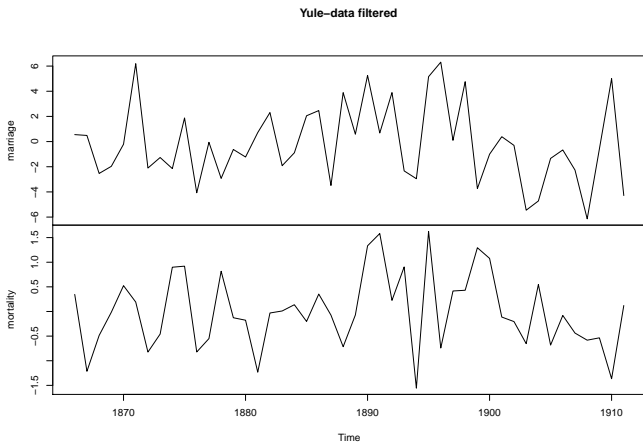
Geri það.



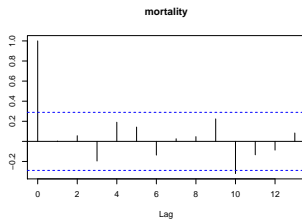
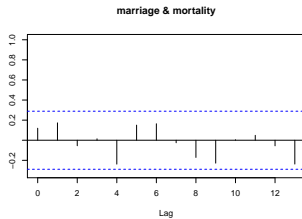
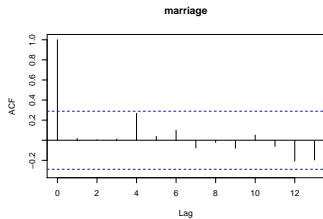
Mynd : Trendhreinsuð Yule-gögn.



Mynd : Sjálffylgni í trendhreinsum Yule-gögnum.



Mynd : Trend og sjálffylgihreinsuð Yule-gögn.



Mynd : Sjálffylgni í trend- og sjálffylgnihreinsuðum Yule-gögnum.



- Niðurstaða: Ekkert sem bendir til þess að samband sé milli mortalitets og markaðshlutdeildar Ensku biskupakirkjunnar í brúðkaupum.

- Niðurstaða: Ekkert sem bendir til þess að samband sé milli mortalitets og markaðshlutdeildar Ensku biskupakirkjunnar í brúðkaupum.
- Nútímamenn myndu e.t.v. meta tvívítt VAR með trendum og prófa fyrir Granger causality.

- Niðurstaða: Ekkert sem bendir til þess að samband sé milli mortalitets og markaðshlutdeildar Ensku biskupakirkjunnar í brúðkaupum.
- Nútímamenn myndu e.t.v. meta tvívítt VAR með trendum og prófa fyrir Granger causality.
- Ekónómetrían var slæm með svona fram yfir 1980. Hugsanlega vegna þess að Havelmaa sló í gegn með simultan líkönum. Hermann Wold sagði: simultan líkön skemmtileg en hagfræði ekki simultan vísindagrein.

- Niðurstaða: Ekkert sem bendir til þess að samband sé milli mortalitets og markaðshlutdeildar Ensku biskupakirkjunnar í brúðkaupum.
- Nútímamenn myndu e.t.v. meta tvívítt VAR með trendum og prófa fyrir Granger causality.
- Ekónómetrían var slæm með svona fram yfir 1980. Hugsanlega vegna þess að Havelmaa sló í gegn með simultan líkönum. Hermann Wold sagði: simultan líkön skemmtileg en hagfræði ekki simultan vísindagrein.
- Granger cointegration kúltur gjörbylti hvernig ekónómetrían tekur á tímaraðagögnum.

- Niðurstaða: Ekkert sem bendir til þess að samband sé milli mortalitets og markaðshlutdeildar Ensku biskupakirkjunnar í brúðkaupum.
- Nútímamenn myndu e.t.v. meta tvívítt VAR með trendum og prófa fyrir Granger causality.
- Ekónómétrían var slæm með svona fram yfir 1980. Hugsanlega vegna þess að Havelmaa sló í gegn með simultan líkönum. Hermann Wold sagði: simultan líkön skemmtileg en hagfræði ekki simultan vísindagrein.
- Granger cointegration kúltur gjörbylti hvernig ekónómétrían tekur á tímaraðagögnum.
- Mér vitanlega er ekónómétrían eina tölfræðigreinin sem hefur tekið formlega á því hvernig bera að álykta um non-stationary tímaraðir

- Bartlett, M. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal the Royal Statistical Society, series B*, 8, 27–41.
- Granger, C. & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 38, 189–203.
- Yule, G. (1926). Why do we sometimes get nonsense-correlations between time-series? a study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89(1), 1–63.