

8. Dane: *SPzwroty.csv*

Indeks giełdowy Standard&Poors 500 (S&P 500) jest wskaźnikiem jakości rynku giełdowego, opartym o notowania 400 przedsiębiorstw przemysłowych, 40 finansowych, 40 usługowych i 20 transportowych. Dane pokazują wartość na końcu każdego roku od 1871 do 1999 sumy 1\$ zainwestowanej w 1871 roku na każdą spółkę S&P 500. Zakłada się, że wszystkie dywidendy są reinwestowane.

- Narysuj wykres zwrotów jako funkcji czasu. Zauważ nieliniowy charakter funkcji trendu. Uzasadnij (wykres, model liniowy), że dane te będą zlinearyzowane gdy zwroty przedstawi się w skali logarytmicznej.
- Oblicz reszty z modelu liniowego logarytmu zwrotów jako funkcji roku. Narysuj wykres.
- Narysuj wykresy funkcji autokorelacji (acf) i funkcji autokorelacji cząstkowej (pacf). Która z nich wskazuje, że reszty są procesem autokorelacji rzędu 1?
- Przeprowadź filtrację procesu resztowego wraz z filtracją czasu używając współczynnika autokorelacji rzędu 1.
- Zbadaj model liniowy między przefiltrowanym procesem resztowym a przefiltrowanym czasem i porównaj z modelem liniowym przed filtracją.
- Narysuj wykres autokorelacji procesu resztowego z ostatniego modelu liniowego i przekonaj się, że jest tu możliwy zwykły model regresji.
- Ile razy błąd standardowy współczynnika przy roku w modelu reszt po filtracji jest większy od błędu standardowego współczynnika przy roku dla modelu reszt przed filtracji? Porównaj ten iloraz ze współczynnikiem, zależnym od autokorelacji, podanym na wykładzie.
- Oszacuj jaka jest typowa wartość (mediana) parametru, określającego o ile procent rocznie rosną zwroty S&P 500 (w latach 1871-1999). Oblicz 95% przedział ufności dla tego parametru. Wartości lewego i prawego końca podaj w % z dokładnością do jednego miejsca po przecinku. Wyjaśnij, dlaczego wyznaczanie tego przedziału jest błędne, gdy opieramy się na modelu regresji z resztami zależnymi.

9. Dane: *czerniak plamy.csv*

wiele czynników wskazuje na zależność częstości występowania czerniaka od nasłonecznienia: występuje częściej w strefie równikowej i u osób z jasną karnacją skóry
Dane zawierają:
Zmienna **czerniak** - częstość (liczba przypadków zachorowania w populacji 100 000 osób) występowania czerniaka, skorygowana względem wieku. Dane pochodzą od chorych mężczyzn, zapisanych w Connecticut Tumor Registry w latach 1936 - 1972. Nasłonecznienie reprezentuje roczna liczba plam na Słońcu (zmienna **plamy**).

Źródło: Houghton, A., Munster, E.W., Viola, M.V., *Increased Incidence of Malignant Melanoma After Peaks of Sunspot Activity*, Lancet, (April 8, 1978, 759-60)

- Narysuj wykres punktowy częstości występowania czerniaka w zależności od liczby plam na Słońcu (wykres synchroniczny). Czy widzisz tu jakąś zależność?
- Narysuj wykres korelacji wzajemnej (w R funkcja `ccf{stats}`) częstości występowania czerniaka w zależności od liczby plam na Słońcu (wykres synchroniczny). Zauważ, że istotny związek między występowaniem czerniaka a liczbą plam na Słońcu występuje przy przesunięciu o 12 lat¹.
- Oszacuj współczynnik korelacji wzajemnej między występowaniem czerniaka a liczbą plam na Słońcu przy przesunięciu o 12 lat (w R funkcja `ccf{stats}` z opcją `plot=F`)

¹ Badania nad okresowością plam na Słońcu pokazują, że występują one cyklicznie co 11 lat.

d) Zbuduj model liniowy częstości występowania czerniaka w zależności od liczby plam na Słońcu w momencie o 12 lat wcześniejszym.

e) Narysuj wykres punktowy częstości występowania czerniaka w zależności od liczby plam na Słońcu (wykres asynchroniczny z przesunięciem o 12 lat) wraz z nałożoną na wykres linią regresji. Skomentuj wykres i adekwatność modelu liniowego.

f) Oblicz reszty tego modelu i narysuj wykres autokorelacji cząstkowej. Zauważ, że reszty są procesem autoregresji rzędu 1 (procesem $AR(1)$)² z dużą wartością autokorelacji. Co to może oznaczać?

g) Przeprowadź filtrację (tu $rm1$ jest autokorelacją cząstkową rzędu 1 z punktu 9.f

```
czەر.f <- czerniak[14:n]-rm1*czerniak[13:(n-1)]  
plamy.f <- plamy[2:(n-12)]-rm1*plamy[1:(n-13)]
```

oraz zbuduj model $\text{lm}(czەر.f \sim plamy.f)$

h) Sprawdź, badając reszty tego modelu, że jest to już model zwykłej regresji.

i) Podsumuj punkty 9.a – 9.h opisując model związku plam na Słońcu i zachorowań na czerniaka.

² Mówi się, że występuje *autokorelacja seryjna*