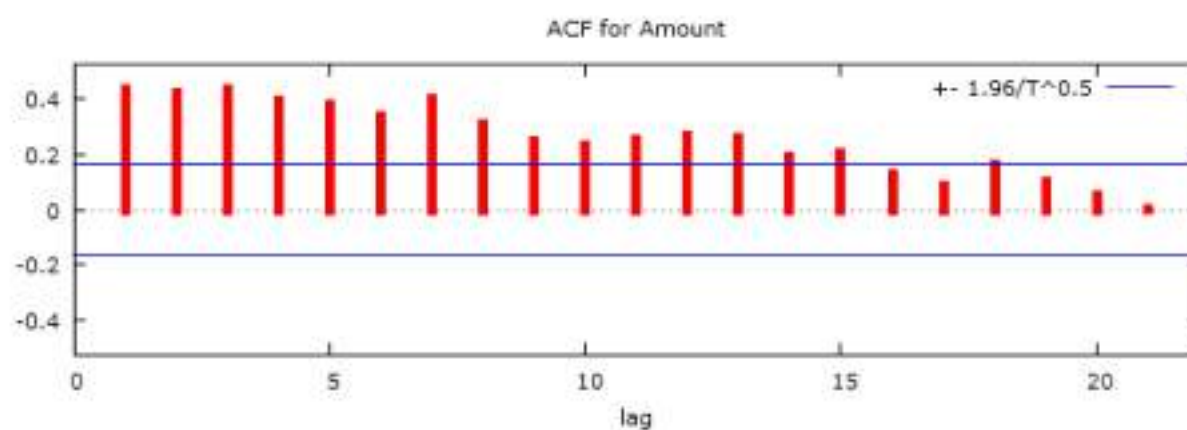


- n) Hvordan kan samspill mellom skift og tidsperiode innarbeides i regresjonsmodellen? Hvor mange ekstra variabler må eventuelt inkluderes?

Parkeringsselskapet ønsker å lage en enkel modell for å predikere framtidige inntekter. Figuren under viser autokorrelasjonsfunksjonen for innsamlet parkeringsavgift i ulike uker (skift 1).



- o) Vil en ARMA(0,1)-modell eller en ARMA(1,0)-modell beskrive denne prosessen best?

## Løsningsforslag

Oppgavesettet og løsningsforslaget er utarbeidet av Jarle Møen. Takk til Jostein Lillestøl som har bidratt med datasettet og problemstillingen. Det anbefales at sensorene gir inntil 10 poeng per delspørsmål.

### Oppgave 1

- Dette er T-test for å sammenligne gjennomsnittet i to uavhengige utvalg med lik varians. (Det siste ser man enklest fra frihetsgradstallet)
- Den lave P-verdien gir grunn til å mistenke at det er en utro tjener på skift 2.
- Antall uker:  $5 \cdot 52 + 42 = 300$ . Skift 2 jobber i halvparten av ukene. Estimert tap per uke: 10845  
Estimert samlet tap:  $150 \cdot 10845 = 1\,626\,750$ .
- Det vil være 95% sannsynlighet for at sann differanse ligger over nedre grense i et 90% konfidensintervall.  
 $90\% \text{ CI: } (147207 - 136362) \pm 1,64 \cdot 2934,06 = 10845 \pm 4811,86 = [6033, 15657]$ .  
Det gjennomsnittlige ukentlige tapet ligger altså med 95% sannsynlighet over 6033.
- $\sqrt{\text{Var}(\bar{X} - \bar{Y})} = \sqrt{\text{Var}(\bar{X}) + \text{Var}(\bar{Y}) - 2\text{Cov}(\bar{X}, \bar{Y})} =$   
 $\sqrt{1994,36^2 + 2152,03^2 - 2 \cdot 0,5 \cdot 1994,36 \cdot 2152,03} = 2077,7$ . [ $\text{Cov}(\bar{X}, \bar{Y}) = r_{xy} \cdot \sigma_{\bar{X}} \cdot \sigma_{\bar{Y}}$ ]
- Vi vet at rangsummen er tilnærmet normalfordelt med  
$$\mu_1 = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{149(149 + 149 + 1)}{2} = 22.275,5 \quad \text{og}$$
$$\sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{149^2 (149 + 149 + 1)}{12}} = 743,76.$$
Testobservatoren blir da  $\frac{25.000,5 - 22.275,5}{743,76} = 3,66$ .  
Kritisk grense på 5%-nivå er 1,96. Følgelig forkaster vi nullhypotesen om ingen forskjell mellom de to skiftene.
- Vi ser at begge faktorene er klart signifikante. Spesifikt kan vi forkaste en nullhypotese om at forventet innsamlet ukentlig beløp er det samme i alle de fem periodene.
- Vi ser at ukentlig innsamlet beløp er signifikant høyere i periode 3 enn i alle de andre periodene. Vi ser også at det ikke er signifikante forskjeller mellom noen av de andre periodene. (P-verdiene tar hensyn til at de parvise sammenligningene ikke er uavhengige og at det totale signifikansnivået for de 10 parvise sammenligningene skal være 95%.)
- Dataene gir grunnlag for å påstå at det er samspill mellom skift og periode. Samspill innebærer i denne sammenheng at forskjellen mellom skift 1 og skift 2 ikke er lik i alle perioder. (Det kan være flere forklaringer på dette. Underslaget startet kanskje etter 1.1 1983, vedkommende ble grådigere over tid, eller omfanget av tyveriet har variert med vedkommendes personlige behov for penger i hver periode.)
- Resultatet av regresjonsanalysen samsvarer godt med det vi har funnet ut i de øvrige analysene. Forventet ukentlig innsamlet beløp er 10.846 kroner lavere for skift 2 enn for skift

1. Forskjellen er klart signifikant. Med hensyn til tidsdimensjonen er innsamlet beløp signifikant høyere i periode 3 enn i de andre periodene. Den estimerte forskjellen mot periode 1 er nesten 31.000. Forklaringsgraden er 45% hvilket må sies å være bra med så få og grovinndelte forklaringsvariabler.
- k) Vi ser at residualene har klokkefasong, men de er noe skjevfordelte og vi kan forkaste en nullhypotese om normalfordeling ved bruk av kjiqvadrattest. Med 298 observasjoner er det imidlertid ingen grunn til bekymring for moderate avvik fra normalfordelingen. Diagnoseplottet til høyre viser klart positiv autokorrelasjon. Dette samsvarer også godt med Durbin-Watson-observatoren i utskriften som er 1,5. Autokorrelasjon påvirker inferensresultatene, selv om de estimerte koeffisientene er forventningsrette. En grundigere analyse burde modellert autokorrelasjonen eller i det minste korrigert standardvvikene (men dette ligger utenfor INT010-pensum).
- l) Forventing for skift 1 i periode 1: kr 132.426  
Forventing for skift 2 i periode 2: kr  $132.426 - 10846 + 4933 = 126.513$
- m) 95% CI:  $10845,7 \pm 1,96 \cdot 2236,1 = 10845,7 \pm 4382,76 = [6463, 15228]$ . Vi ser at punktestimatet er nær identisk, men at konfidensintervallet er smalere. (Vi har noe bedre presisjon enn i analyse 1 fordi vi nå kontrollerer for systematiske forskjeller i forventet innsamlet beløp mellom de ulike periodene.)
- n) Samspill må modelleres som interaksjonsledd mellom Dshift\_2 og de fire inkluderte periodedummyene. Vi må altså inkludere fire nye dummyvariabler. (Fra analyse 5 ser vi at vi også der bruker opp fire frihetsgrader på å inkludere samspillseffekter.)
- o) Vi ser at det er sterk autokorrelasjon som dør ut sakte. Dette er forenelig med en  $ARMA(1,0)=AR(1)$ , men ikke med en  $ARMA(0,1)=MA(1)$ . Sistnevnte vil bare ha positiv autokorrelasjon for lag 1.