

## NORGES HANDELSHØYSKOLE

### Eksamen vår 2014

Kurskode: INT010      Tittel: Anvendt metode

Dato: 27.5.2014      Kl. 09.00-12.00

Faglærer går ikke rundt i eksamenslokalene, men kan kontaktes av eksamensvakten på  
tlf. 59612/99 69 93 05

### ***Hjelpemidler til eksamen:***

Hjelpemidler tillatt: JA, alle trykte/egenskrevne

Kalkulator: JA [I tråd med retningslinjer for bruk av kalkulator, jf.  
utfyllende bestemmelser til Forskrift om eksamen ved Norges  
Handelshøyskole (fulltidsstudiene).]

***Alle delspørsmål i oppgavesettet teller likt.  
Husk at dere må skrive slik at sensor kan lese det som er skrevet.***

Eksamensoppgaven består av 6 sider, inkludert denne.

## Oppgave 1

Parkeringsselskapet i Bergen hadde i årene 1983 til 1988 to skift som tømte parkometre og billettautomater for penger. Skift 1 jobbet i ulike uker og skift 2 i like uker.

Data for ukentlig omsetning (variabelnavn «Amount») er tilgjengelig fra og med uke 1 i 1983 til og med uke 42 i 1988. Datasettet inneholder 149 observasjoner av hvert skift. Det mangler data fra enkelte uker, blant annet knyttet til en streik to uker våren 1986.

Både personalet og tømmerutinene var stabile fram til uke 42 i 1988. Da kom det en varslingssak hvor ledelsen i parkeringsselskapet fikk vite at utstyret for å tømme parkometrene hadde en svakhet. Det gikk rykter om at en utro tjener utnyttet dette.

En statistiker ved NHH ble engasjert som sakkyndig. I denne oppgaven får du se en del analyser som er gjort for å utforske datasettet (utskrifter fra Excel og Gretl er lett redigert for økt lesbarhet).

Statistikeren startet ut med følgende enkle test:

### ANALYSE 1

Null hypothesis: Difference of means = 0

Sample 1 (Shift 1):

n = 149, mean = 147207, s.d. = 24344,3  
standard error of mean = 1994,36  
95% confidence interval for mean: 143266 to 151148

Sample 2 (Shift 2):

n = 149, mean = 136362, s.d. = 26268,8  
standard error of mean = 2152,03  
95% confidence interval for mean: 132109 to 140614

Test statistic:  $t(296) = (147207 - 136362) / 2934,06 = 3,69648$

Two-tailed p-value = 0,0002604

(one-tailed = 0,0001302)

- Hvilken test er dette?
- Gir testen grunn til å anta at parkeringsselskapet har en utro tjener?
- Regn ut et beste anslag for det samlede tapet som parkeringsselskapet har lidt dersom de har en utro tjener fra og med uke 1 i 1983 til og med uke 42 i 1988.
- Gi et nedre anslag for forskjellen mellom de to skiftene i forventet ukentlig innsamlet parkeringsavgift slik at det er 95 % sannsynlighet for at den sanne forskjellen ligger over dette anslaget.

La  $t$  være en periodelengde på 14 dager slik at det inngår en uke for hvert skift i hver periode  $t$ . La videre  $X_t$  være innsamlet parkeringsavgift for skift 1 og  $Y_t$  være innsamlet parkeringsavgift for skift 2. I utvalget er korrelasjonskoeffisienten mellom  $X_t$  og  $Y_t$ ,  $r_{xy} = r_{\bar{x}\bar{y}} = 0,50$ .

- e) Regn ut standardavviket til forskjellen mellom de to skiftene i gjennomsnittlig ukentlig innsamlet parkeringsavgift ( $\bar{X} - \bar{Y}$ ).

Som en sjekk av robustheten til resultatet i analyse 1 gjennomfører også statistikeren Wilcoxons test for to utvalg. Deler av Excelutskriften er gjengitt under.

## ANALYSE 2

### Wilcoxon Rank Sum Test

	Rank Sum	Observations
Amount shift 1	25007.5	149
Amount shift 2	19543.5	149

- f) Fullfør testen i analyse 2.

Parkeringsavgiften økte 1.1.1984, 1.1.1987 og 1.1.1988. Dessuten var det en reduksjon i antall parkeringsplasser 30.6.1984. Dette definerer fem tidsperioder hvor forventet ukentlig omsetning kan antas å være lik innenfor hver periode.

Statistikeren ønsker å undersøke om det finnes systematisk variasjon i innsamlet parkeringsavgift knyttet til endringer i pris og antall parkeringsplasser. Han gjør derfor en toveis variansanalyse med «skift» som den ene faktoren (to nivåer) og «periode» som den andre faktoren (fem nivåer). Se analyse 3, som er gjort i Gretl ved hjelp av R-konsollen.

## ANALYSE 3

```
aov(Amount)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Shift	1	8.763e+09	8.763e+09	23.52	2.01e-06 ***
Period	4	8.107e+10	2.027e+10	54.41	< 2e-16 ***
Residuals	292	1.088e+11	3.725e+08		

```
Signif. codes:  0. '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- g) Drøft kort utskriften.

For å få en bedre forståelse av hvordan innsamlet parkeringsavgift varierer mellom de fem periodene som datasettet er delt inn i, gjør statistikeren multiple sammenligninger og tester differansen mellom de ulike periodene med Tukeys test, se analyse 4. Første tall i kolonnen «diff» (4932,615) er eksempelvis gjennomsnittlig innsamlet parkeringsavgift i periode 2 minus gjennomsnittlig innsamlet parkeringsavgift i periode 1. Programmet oppgir nedre og øvre konfidensgrenser og testens P-verdier.

#### ANALYSE 4

Tukey multiple comparisons of means  
95% family-wise confidence level

Period		diff	lwr	upr	p adj
2-1	4932.615	-9760.835	19626.065	0.8884746	
3-1	30735.618	19492.235	41979.000	0.0000000	
4-1	-5794.962	-18519.862	6929.939	0.7219132	
5-1	-2232.540	-15452.757	10987.677	0.9904781	
3-2	25803.003	14559.620	37046.385	0.0000000	
4-2	-10727.577	-23452.478	1997.324	0.1433228	
5-2	-7165.156	-20385.373	6055.061	0.5713121	
4-3	-36530.579	-45041.706	-28019.452	0.0000000	
5-3	-32968.158	-42203.434	-23732.882	0.0000000	
5-4	3562.421	-7428.474	14553.316	0.9005781	

h) Hva forteller utskriften fra analyse 4 oss?

Det neste statistikeren tester er eksistensen av samspillseffekter mellom faktoren skift og faktoren periode. Dette er gjort i analyse 5.

#### ANALYSE 5

aov(Amount)

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Shift	1	8.763e+09	8.763e+09	24.816	1.09e-06 ***
Period	4	8.107e+10	2.027e+10	57.391	< 2e-16 ***
Shift*Period	4	7.071e+09	1.768e+09	5.006	0.000649 ***
Residuals	288	1.017e+11	3.531e+08		

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

i) Hva er tolkningen av en eventuell samspillseffekt, og hva forteller utskriften av analyse 5 oss?

Til slutt foretar statistikeren en regresjonsanalyse.

Dshift\_2 er en dummyvariabel som er 1 hvis observasjonen er fra skift 2 og null ellers

Dperiod\_2 er en dummyvariabel som er 1 hvis observasjonen er fra periode 2 og null ellers.

Dperiod\_3 er en dummyvariabel som er 1 hvis observasjonen er fra periode 3 og null ellers, osv.

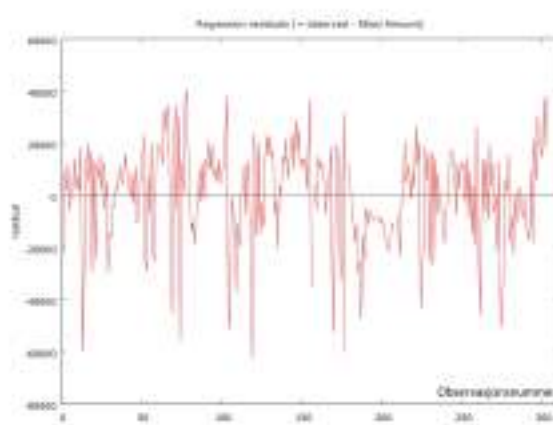
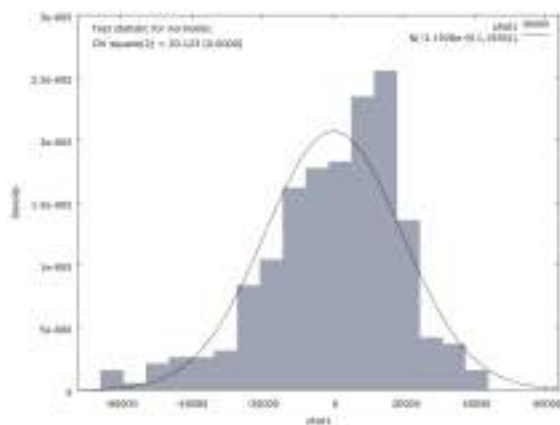
## ANALYSE 6

Model 1: OLS, using observations 1-298

Dependent variable: Amount

	coefficient	std. error	t-ratio	p-value
const	132426	3946.81	33.55	3.39e-102 ***
DShift_2	-10845.7	2236.10	-4.850	2.01e-06 ***
DPeriod_2	4932.62	5353.00	0.9215	0.3576
DPeriod_3	30735.6	4096.10	7.504	7.54e-013 ***
DPeriod_4	-5794.96	4635.83	-1.250	0.2123
DPeriod_5	2232.54	4816.28	-0.4635	0.6433

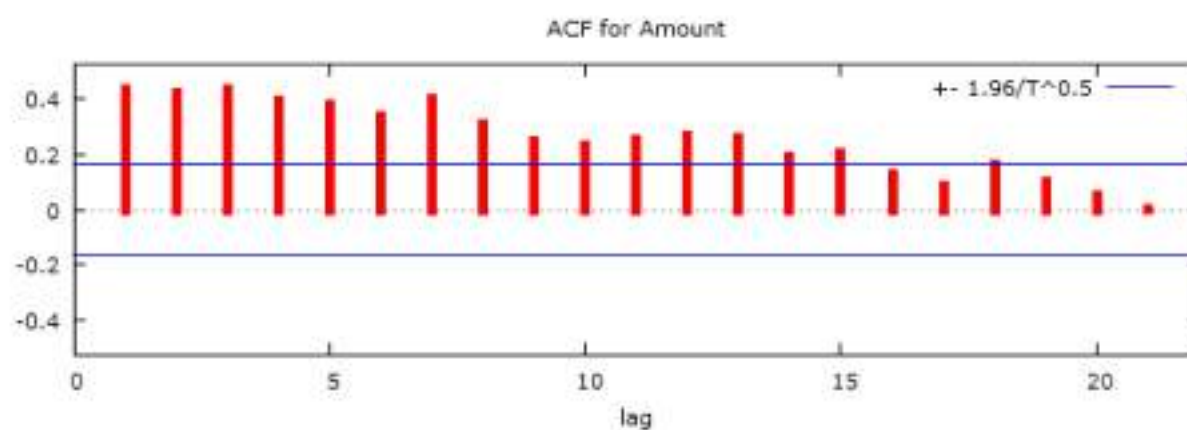
Mean dependent var	141784.5	S.D. dependent var	25859.15
Sum squared resid	1.09e+11	S.E. of regression	19300.52
R-squared	0.452309	Adjusted R-squared	0.442931
F(5, 292)	48.22944	P-value (F)	2.77e-36
rho	0.219739	Durbin-Watson	1.548300



- Drøft kort utskriften.
- Drøft kort de to diagnoseplottene. [Histogram av residualene til venstre og residualene plottet mot observasjonsnummeret (uke) til høyre.]
- Hva er estimert innsamlet ukentlig parkeringsavgift for skift 1 i periode 1 og skift 2 i periode 2?
- Regn ut et 95 % konfidensintervall for forskjellen mellom de to skiftene i forventet ukentlig innsamlet parkeringsavgift.

- 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?