

Bokmål

Institutt:	IKBM	
Eksamen i:	Stat100	Statistikk
Tid:	Tirsdag 19. mai, 2009	09.00-12.30
Emneansvarlig:	Trygve Almøy 95141344	

Oppgaveteksten er på: 4 sider

Tillatte hjelpemiddel: C3: alle typer kalkulatorer, alle andre hjelpemiddel

Alle deloppgaver teller likt.

For Alle tester du utfører må du skrive opp både nullhypotese og alternativ.

Oppgave 1

Et firma ønsker å teste ut en ny hamburgertype. De trekker ut 10 personer tilfeldig og lar hver av disse smake på den nye hamburgeren, for deretter å gi poeng på en skala fra 0 til 10. For å ha noe å sammenligne med lar de også de samme personene smake på en hamburger av den nåværende typen og gi poeng til denne. (10 poeng er best)

Person	1	2	3	4	5	6	7	8	9	10
ny type	10	7	8	7	6	8	7	9	9	8
nåværende type	8	5	7	4	7	5	6	7	10	6

- a) Sett opp en modell for forsøket.
Lag et 95 % konfidensintervall for forventet forskjell i poeng mellom disse to typene.
- b) Det ble samtidig utført en statistisk hypotesetest der alternativet var om folk likte den nye typen best. Denne testen fikk en p-verdi på 0,006.
Sett opp nullhypotese og alternativ.
Forklar hvordan P-verdien ble regnet ut i dette tilfelle (skriv opp det matematiske uttrykket).
Forklar hamburgerprodusenten, som ikke har kunnskap i statistikk, hva denne P-verdien betyr i praksis.
- c) En annen måte å teste den nye hamburgertypen på, er å telle hvor mange som sier den nye typen er best. La X være antallet av de 10 som likte ny type best.
Hva slags sannsynlighetsfordeling har X ?
Sett opp en statistisk hypotesetest basert på X , der alternativet er om folk foretrekker den nye typen.
Denne testen får en p-verdi på 0,055 (det trenger du ikke vise). Hvorfor tror du at p-verdien blir større i denne testen jamfør den i oppgave b)?
- d) De resultatene du har sett til nå er bare en del av en større undersøkelse der firmaet ville undersøke om alder har betydning for hva folk synes om den nye hamburgertypen. Firmaet spurte 294 personer om hva slags type hamburger disse personene foretrekker, samtidig registrerte de om det var en ung, en middelaldrende eller en gammel som svarte. Dette, gav følgende resultat:
80 av 118 unge likte den nye best,
70 av 111 middelaldrende likte den nye best,
29 av 65 eldre likte den nye best.
Test på 5 % nivå om smak og aldersgruppe er avhengig.

Oppgave 2

I 2001 ble det undersøkt om gravidens røykevaner påvirker vekta av den nyfødte. (S. Bonnelie. *Pediatric and Perinatal Epidemiology* vol 15). Mødrene ble delt inn i tre kategorier; ikke-røyker, festrøyker (røyker av og til) og vanerøyker. Det ble trukket tilfeldig ut 40 nybakte mødre i hver kategori, og vekta av disse nyfødte barn ble registrert.

a) Sett opp en modell for forsøket.

Bruk utskriften under til å utføre en test på om mødres røyking har effekt på vekta av den nyfødte.

Source	DF	SS	MS	F
Røykevaner	*	4220902	*	*
Error	*	75803576	*	*
Total	*	80024477	*	*

Gruppe	N	Mean	StDev
Ikke-røyker	40	3434,2	642,5
Fest-røyker	40	3215,1	996,8
Vanerøyker	40	2974,9	732,9

Pooled StDev = 804,9

b) Er det sikkert at vi får redusert spedbarnsvekt dersom ikke-røykere blir festrøykere?
Er det sikkert at vi får redusert spedbarnsvekt dersom festrøykere går over til å bli vanerøykere?

c) Forsøket i a) ble utført bare i USA. Som et tenkt forsøk kunne en i stedet for ha trukket ut 120 gravide over hele verden, og delt dem inn i kategorier som tidligere (40 i hver kategori). Dersom data hadde resultert i den samme gjennomsnittsverdien for hver kategori, men at standardavviket i hver gruppe hadde (for å forenkle rekninga) blitt 1500, ville nedre del av tabellen i a) sett slik ut:

Gruppe	N	Mean	StDev
Ikke-røyker	40	3434,2	1500
Fest-røyker	40	3215,1	1500
Vanerøyker	40	2974,9	1500

Hva ville nå resultatet av testen bli?

Hva er det med dette forsøksopplegget som gjør F-verdien mindre i forhold til det opprinnelige forsøket?

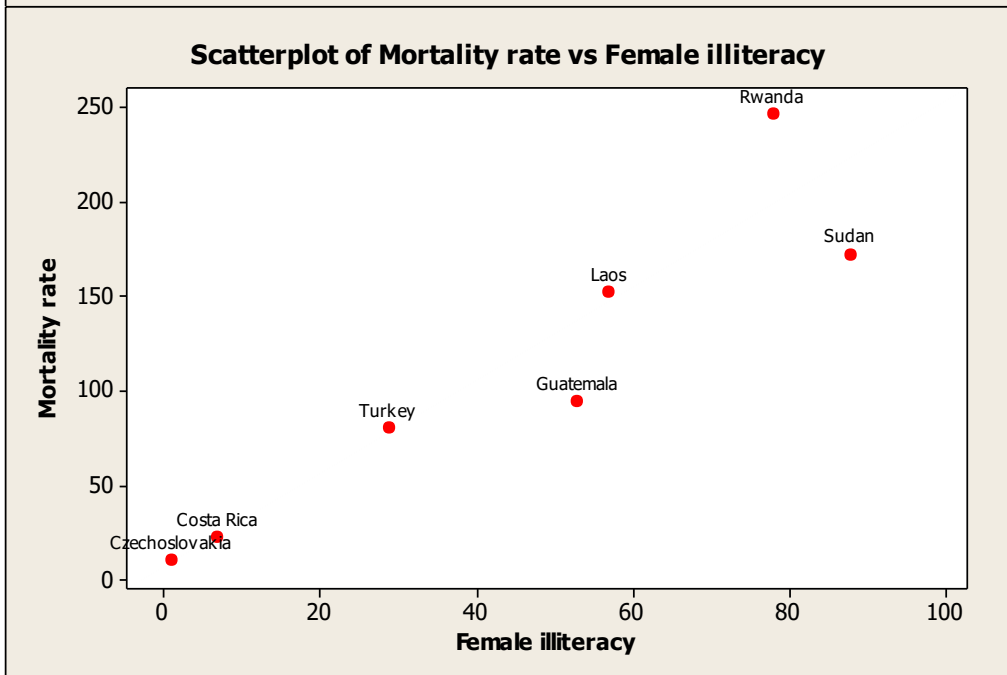
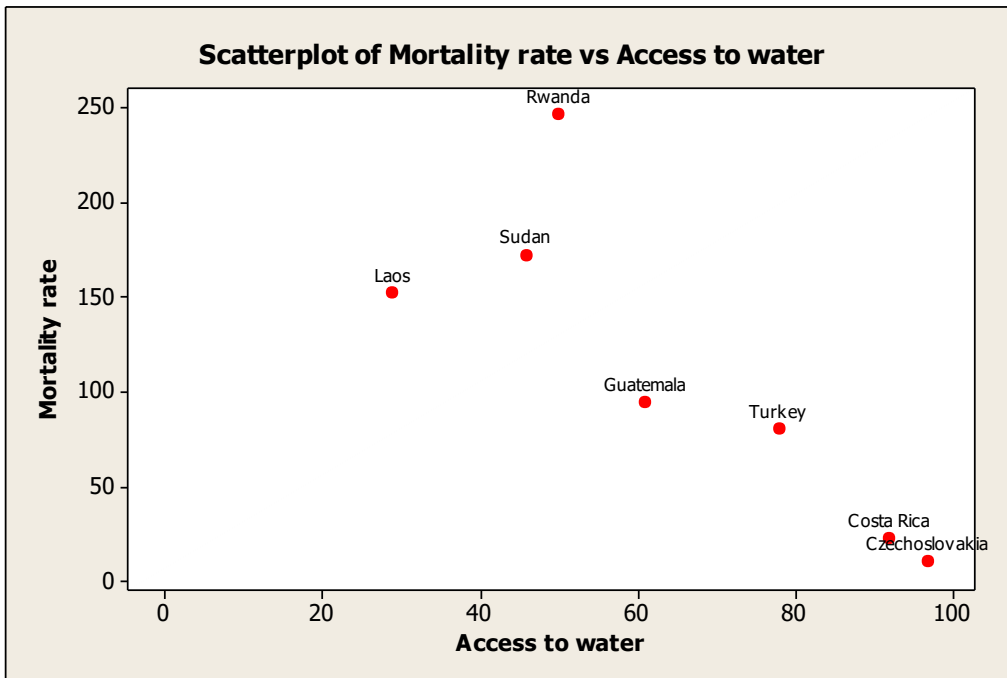
d) Et tilsvarende forsøk gjort i Norge ga (tallene er forenklet):

Gruppe	N	Mean	StDev
Ikke-røyker	20	3400	500
Fest-røyker	30	3100	600
Vanerøyker	40	3000	550

Lag en fullstendig ANOVA- tabell og test om røyking har effekt på spedbarnsvekt.

Oppgave 3

Årlig publiserer FN en rapport som kalles «Tilstand for verdens barn». Fra 2007-utgaven har vi plottet informasjon fra sju svært ulike land:



Mortality rate er årlige antall på døde barn under 5 år per 10 000 levende fødte. Senere i oppgava kalt barnedødelighet.

Access to water er prosentandel av befolkningen som har tilgang til rent drikkevann.

Female illiteracy er prosentandel av kvinner over 15 år som *ikke* kan lese.

Vi vil ha *Mortality rate* som responsvariabel, men prøver med to ulike forklaringsvariable.

Modell 1: La *access to water* være forklaringsvariabel

Modell 2: La *Female illiteracy* være forklaringsvariabel.

På slutten av denne oppgava finner du to Minitabutskrifter

a) For modell 1: Sett opp en lineær regresjonsmodell og tolk alle parametrene.

For modell 2: Sett opp en lineær regresjonsmodell.

Hvordan vil du estimere standardavviket i barnedødelighet for alle land med samme prosentandel kvinnelige analfabeter?

Hvordan ville du estimert forventet barnedødelighet dersom alle kvinner kunne lese?

- b) Test påstanden om at det er sammenheng mellom tilgang på rent vann og barnedødelighet.
 Test påstanden om at det er sammenheng mellom kvinnelig analfabetisme og barnedødelighet.
- c) I Vietnam har 42 % av befolkningen tilgang på rent vann mens 84 % av kvinnene kan lese.
 For begge modeller anslå barnedødelighet for Vietnam.
 Hvorfor tror du modellene gir så forskjellige anslag?
 Hva er den viktigste årsaken til at prediksjonsintervallet blir smalest for Modell 2?
 I Norge har 99 % tilgang til rent vann. Hvorfor vil et prediksjonsintervall for Modell 1 bli bredere for Norge enn for Vietnam?
- d) Finn determinasjonskoeffisienten (R^2) for Modell 2.
 Hva betyr det at R^2 er større i Modell 2 enn i Modell 1?

Modell 1					
Predictor	Coef	StDev	T	P	
Constant	290.97	57.66	5.05	0.004	
Access to water	-2.7832	*	*	0.021	
S = 51.79 R-Sq = 68.8 %					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	*	29586	11.03	0.021
Error	5	13413	2683		
Total	6	42999			
Predicted Values					
Fit	95,0% CI		95,0% PI		
*	(103,9; 244,3)		(23,6; 324,6)		Vietnam

Modell 2					
Predictor	Coef	StDev	T	P	
Constant	6.97	23.66	0.29	0.780	
Female Illiteracy	*	0.4342	5.35	*	
S = 35.76 R-Sq = *					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	36606	36606	*	*
Error	5	*	*		
Total	6	*			
Predicted Values					
Fit	95,0% CI		95,0% PI		
*	(-3,1; 91,4)		(-59,2; 147,5)		Vietnam