



EKSAMENSOPPGAVE I STA-0001

Eksamen i : Sta - 0001 Brukerkurs i statistikk 1
Dato : Fredag 28. september 2007
Tid : Kl. 09.00 – 13.00
Sted : Aud. max. Administrasjonsbygget
Tillatte hjelpemidler : Alle trykte og skrevne hjelpemidler samt kalkulator.

Oppgavesettet er på 3 sider inklusive forside.

Kontaktperson: Sigrunn H. Sørbye, tlf. 92 43 97 22.

Oppgave 1

Et tilfeldig valgt ungekull fra en gnagerart består av X unger, der $2 \leq X \leq 5$. Sannsynlighetsfordelingen til X er gitt ved

x	2	3	4	5
$P(X = x)$	0.05	0.25	0.50	0.20

Vi antar at antall unger i ulike kull er uavhengige av hverandre.

- Finn den kumulative fordelingsfunksjonen $F(x) = P(X \leq x)$ og skisser denne. Hva er sannsynligheten for at et tilfeldig valgt kull har minst 4 unger?
- Finn forventningsverdien og standardavviket til X .
- Vi undersøker tre tilfeldig valgte ungekull. Hva er sannsynligheten for at minst ett av disse 3 kullene har 5 unger? Hva er sannsynligheten for at det tilsammen er nøyaktig 13 unger i de tre kullene?

Oppgave 2

Ved en fabrikk har man anskaffet en pakkemaskin som pakker hvetemel i 2.5 kg's pakker. Vekten X (i kg) av en vilkårlig melpakke antas å være normalfordelt med forventning μ og standardavvik $\sigma = 0.05$. Når maskinen er i orden, er $\mu = 2.57$. Vekten av ulike pakker antas å være uavhengige stokastiske variabler.

Anta i punkt a) og b) at maskinen er i orden.

- Hva er sannsynligheten for at en vilkårlig pakke veier mindre enn 2.5 kg? Hva er sannsynligheten for at en vilkårlig pakke veier mindre enn 2.5 kg gitt at den veier mindre enn 2.57 kg?
- Anta at vi trekker ut melpakker tilfeldig og veier dem inntil vi finner en pakke som veier mindre enn 2.5 kg. La Y angi antall melpakker som veies inntil vi finner en pakke som veier mindre enn 2.5 kg. Hvilken sannsynlighetsfordeling har Y ? Hva blir $P(Y > 4)$? Hva blir $E(Y)$?

For å kontrollere om melpakkene er for lette vil bedriften gjennomføre en hypotesetest der

$$H_0 : \mu = 2.57, \quad H_1 : \mu < 2.57.$$

Testen skal utføres med signifikansnivå $\alpha = 0.01$ og bedriften ønsker at sannsynligheten for å forkaste nullhypotesen er minst 0.99 dersom den sanne verdien for μ er lik 2.49.

- Hvor mange melpakker må minst kontrolleres for å gjennomføre testen etter bedriftens ønsker? Hva blir forkastningsområdet for denne testen?

Vi veier nå 10 tilfeldig valgte melpakker og får følgende resultater

2.51, 2.60, 2.55, 2.48, 2.53, 2.58, 2.46, 2.47, 2.54, 2.58.

- Finn et 99% konfidensintervall for μ basert på disse observasjonene der standardavviket fremdeles antas kjent lik 0.05. Forklar hva slags sammenheng vi har mellom dette konfidensintervallet og en hypotesetest for μ .

Oppgave 3

Norges årlige CO₂ utslipp har i perioden 1973 - 2006 økt fra 30.6 millioner tonn i 1973 til 43.3 millioner tonn i 2006. Den årlige mengden CO₂ utslipp i denne 34 års-perioden ligger på nettsidene til Statistisk Sentralbyrå (<http://www.ssb.no/emner/01/04/10/klimagassn/tab-2007-05-11-04.html>) og gjengis ikke her, men en lineær regresjonsanalyse i Excel for disse 34 årene gav følgende resultater:

SAMMENDRAG (UTDATA)

Regresjonsstatistikk					
Multippel R	0.930				
R-kvadrat	0.865				
Justert R-kvadrat	0.861				
Standardfeil	1.753				
Observasjoner	34				

Variansanalyse					
	fg	SK	GK	F	Signifikans-F
Regresjon	1	630.611	630.611	205.304	0.00
Residualer	32	98.291	3.072		
Totalt	33	728.903			

	Koeffisienter	Standardfeil	t-Stat	P-verdi
Skjæringspunkt	-837.129	60.953	-13.734	0.000
X-variabel 1	0.439	0.031	14.328	0.000

- a) Sett opp en lineær regresjonsmodell og forklar hva de ulike parametrene i modellen beskriver. Begrunn utfra Excel-resultatene hvorvidt en lineær modell er passende for å beskrive Norges årlige CO₂ utslipp i den gitte tidsperioden. Har det vært en signifikant økning i årlig CO₂ utslipp i denne perioden? Formuler hypoteser og gjennomfør testen på 1% nivå.

Oppgave 4

To målemetoder for bestemmelse av pH-verdi i oppløsninger skal sammenlignes. Vi har i alt 12 ulike typer oppløsninger. For hver oppløsning foretas en pH-måling med hver metode. La D_i angi differansen (metode I - metode II) i målt pH-verdi mellom de to metodene for $i = 1, \dots, 12$. Vi antar at $D_1, \dots, D_{12}$ er uavhengige og identisk fordelte, mens de to målingene for hver oppløsning er avhengige. Forsøket gav følgende resultater for de 12 oppløsningene:

Forsøk nr. (i)	1	2	3	4	5	6	7	8	9	10	11	12
Differanse (d_i)	-0.36	-0.07	0.11	-0.15	-0.05	-0.45	-0.40	-0.36	-0.46	0.12	-0.35	0.22

La S angi antallet av de 12 differansene som er positive.

- a) Begrunn hvorfor S vil være binomisk fordelt. Hvordan vil du formulere hypoteser for å teste om de to metodene gjennomgående gir samme verdier eller ikke når testen gjennomføres med S som testobservator? Finn p-verdien for denne testen og konkluder hvorvidt nullhypotesen bør forkastes eller ikke utfra et signifikansnivå på 5%.

Vi vil nå anta at differansene $D_1, \dots, D_{12}$ er identisk normalfordelt. Det oppgis at $\sum_{i=1}^{12} d_i = -2.20$ og $\sum_{i=1}^{12} d_i^2 = 1.06$.

- b) Hvilken test vil du nå benytte for å sammenligne metodene? Formuler hypoteser for denne testen. Gjennomfør testen og konkluder utfra et signifikansnivå på 5%. Kommenter resultatet sammenlignet med punkt a).