Høgskolen i Telemark
Avdeling for allmennvitenskaplige fag

EKSAMEN

I

STATISTIKK II 037/5551

7.12.05

Tid:	9-13 (4 timer).
Fagkode:	5551/037.
Sidetall:	6.
Hjelpemidler:	Alle trykte og skrevne samt kalkulator.
Målform:	Bokmål/nynorsk.

BOKMÅL

OPPGAVE 1

Vi lar θ_1 være andelen av kvinner i Grenland som minst én gang i løpet av siste uke har benyttet seg av kollektiv transport. θ_2 er tilsvarende andel blant menn.

Av 500 tilfeldig uttrukne kvinner i fylket svarte $Y_1 = 74$ at de hadde benyttet kollektiv transport minst én gang sist uke. Tilsvarende antall av 500 uttrukne menn var $Y_2 = 57$.

a) Angi et tilnærmet 95% konfidensintervall for $\delta = \theta_1 - \theta_2$

b) Er det flere kvinner enn menn som benytter kollektiv transport i Grenland?
Formuler dette spørsmålet som en hypotesetest. Gjennomfør testingen med signifikansnivå 5%. Finn signifikanssannsynligheten (tilnærmet) for testen.

OPPGAVE 2

En idrettsforsker har undersøkt såkalt motorisk ferdighet for 8 år gamle gutter fra 3 skoler i henholdsvis Halden, Sarpsborg og Nøtterøy. Siktemålet med undersøkelsene er å finne ut om det er forskjeller i motorisk ferdighetsnivå for barn som har vokst opp i ulike sosiale og geografiske miljø.

Guttene som deltok i undersøkelsen, gjennomgikk en grovmotorisk test bestående av 4 deloppgaver. På grunnlag av resultatene på de 4 oppgavene ble det regnet ut en motorisk kvotient (MQ).

Vi lar X_{ij} være målt motorisk kvotient for gutt nr j fra skole nr i, der $i = 1$ er skolen i Halden, $i = 2$ er skolen i Sarpsborg og $i = 3$ er skolen på Nøtterøy. Antall deltakere i undersøkelsen var $n_1 = 18$ (Halden), $n_2 = 23$ (Sarpsborg) og $n_3 = 22$ (Nøtterøy).

Resultatene er gitt til slutt i oppgaven.

Vi antar at X_{ij} er normalfordelt med forventning μ_i og varians σ^2 ($i = 1, 2, 3$ og $j = 1, 2, 3, \ldots, n_i$). Alle observasjonene er uavhengige.

a) Angi estimatorer for de ukjente parametrene i modellen (μ_1, μ_2, μ_3 og σ^2). Finn et 95% konfidensintervall for forventet motorisk ferdighetsnivå for 8 år gamle gutter på skolen i Halden.

b) Tyder forsøket på at det er forskjeller i motorisk ferdighetsnivå hos 8 år gamle gutter på de 3 skolene? Formuler dette spørsmålet som en hypotesetest. Gjennomfør testingen og angi konklusjonen når vi velger signifikansnivå 5%.

c) Gjør kort rede for valg av modell/analyse for følgende problemstillinger knyttet til motorisk ferdighetsnivå for barn:

1) Resultatene fra de 3 skolene skal sammenlignes både for gutter og for jenter. En skal både se på om det er signifikante forskjeller mellom skoler og mellom gutter og jenter. Som i a) og b) antar vi at målte motoriske kvotienter for barn er uavhengige og normalfordelte variabler.

2) Problemstillingen er den samme som i b): Vi skal sammenligne gutter fra 3 skoler. I motsetning til i b) antar vi nå at motorisk ferdighet for hver gutt er klassifisert og registrert som enten **god**, **middels** eller **dårlig**. Hvordan kan en nå analysere problemstillingen i b)?

Resultater

X_{1j} (Halden): 100 86 71 80 95 77 104 96 91 82 85 72 92 97 82 85 71 78

$n_1 = 18$ $\quad \bar{X}_1 = 85.78$ $\quad \sum (X_{1j} - \bar{X}_1)^2 = 1744.5$

X_{2j} (Sarpsborg): 90 75 111 80 105 91 69 105 108 79 80 83 119 97 105 111 79 76 60 113 80 102 95

$n_2 = 23$ $\quad \bar{X}_2 = 91.87$ $\quad \sum (X_{2j} - \bar{X}_2)^2 = 5745.2$

X_{3j} (Nøtterøy): 90 114 110 127 102 72 80 94 93 96 96 85 106 108 108 83 89 89 91 99 111 118

$n_3 = 22$ $\quad \bar{X}_3 = 98.23$ $\quad \sum (X_{3j} - \bar{X}_3)^2 = 3810.3$

Totalt gjennomsnitt: $\bar{X} = 92.35$

OPPGAVE 3

Vi skal studere utviklingen av folketallet i Bø kommune ved hjelp av en enkel regresjonsmodell

$$Y = \beta_0 + \beta_1 x + U$$

der Y er folketall på tidspunkt nr x ($x = 1, 2, \ldots, 8$) og $U_1, U_2, \ldots, U_8$ er uavhengige og normalfordelte variabler med forventning 0 og standardavvik σ.

Registrering av folketall er foretatt hver tredje måned, første gang 01.01.2004 og åttende og foreløpig siste gang 01.10.2005. Se Minitab-utskriften til slutt i oppgaven. Dataene er hentet fra Statistisk sentralbyrå.

a) Skriv opp den estimerte regresjonsmodellen og forklar kort hva de estimerte parametrene uttrykker.

Har det vært en signifikant økning i folketallet i Bø i den aktuelle perioden? Formuler selv nullhypotese og alternativ hypotese for å teste dette spørsmålet. Angi konklusjonen når signifikansnivået velges lik 1%.

Ledelsen i kommunen ønsker et intervall med sikkerhet 95% for folketallet i Bø 01.01.2006. Vil det i dette tilfelle være et konfidensintervall eller et prediksjonsintervall som er et naturlig svar på kommunens ønske? Begrunn svaret kort og regn ut det aktuelle intervallet.

Utskrift fra Minitab

Data Display

Dato	Tidspunkt	Folketall
01.01.2004	1	5173
01.04.2004	2	5168
01.07.2004	3	5182
01.10.2004	4	5241
01.01.2005	5	5249
01.04.2005	6	5244
01.07.2005	7	5210
01.10.2005	8	5288

Regression Analysis: Folketall versus Tidspunkt

```
The regression equation is
Folketall = 5155 + 14,4 Tidspunkt

Predictor        Coef     SE Coef          T        P
Constant      5154,61       20,70     248,99    0,000
Tidspunk       14,393       4,100       3,51    0,013

S = 26,57       R-Sq = 67,3%     R-Sq(adj) = 61,8%
```

NYNORSK

OPPGÅVE 1

Vi lar θ_1 vere andelen av kvinner i Grenland som minst ein gong i løpet av siste veke har nytta kollektiv transport. θ_2 er tilsvarande andel blant menn.

Av 500 tilfeldig uttrekte kvinner i fylket svarte $Y_1 = 74$ at dei hadde nytta kollektiv transport minst ein gong siste veke. Tilsvarande tal av 500 uttrekte menn var $Y_2 = 57$.

a) Gi eit tilnærma 95% konfidensintervall for $\delta = \theta_1 - \theta_2$

b) Er det fleire kvinner enn menn som nyttar kollektiv transport i Grenland?
Formuler dette spørsmålet som ein hypotesetest. Gjennomfør testinga med signifikansnivå 5%. Finn signifikanssannsynligheten (tilnærma) for testen.

OPPGÅVE 2

Ein idrettsforskar har undersøkt såkalla motorisk dugleik for 8 år gamle gutar frå 3 skular i høvesvis Halden, Sarpsborg og Nøtterøy. Siktemålet med undersøkingane er å finne ut om det er forskjeller i motorisk dugleik for barn som har vakse opp i ulike sosiale og geografiske miljø.

Gutane som deltok i undersøkinga, gikk gjennom ein grovmotorisk test som besto av 4 deloppgåver. På grunnlag av resultata på dei 4 oppgåvene blei det rekna ut ein motorisk kvotient (MQ).

Vi lar X_{ij} vere målt motorisk kvotient for gut nr j frå skule nr i, der $i = 1$ er skulen i Halden, $i = 2$ er skulen i Sarpsborg og $i = 3$ er skulen på Nøtterøy. Antall deltakarar i undersøkinga var $n_1 = 18$ (Halden), $n_2 = 23$ (Sarpsborg) og $n_3 = 22$ (Nøtterøy).

Resultata er gitt til slutt i oppgåva (neste side).

Vi antar at X_{ij} er normalfordelt med forventning μ_i og varians σ^2 ($i = 1, 2, 3$ og $j = 1, 2, 3, \ldots, n_i$). Alle observasjonane er uavhengige.

a) Gi estimatorar for dei ukjente parametrane i modellen (μ_1, μ_2, μ_3 og σ^2). Finn eit 95% konfidensintervall for forventa motorisk dugleik for 8 år gamle gutar på skulen i Halden.

b) Tyder forsøket på at det er forskjeller i motorisk dugleik hos 8 år gamle gutar på dei 3 skulane? Formuler dette spørsmålet som ein hypotesetest. Gjennomfør testinga og gi konklusjonen når vi vel signifikansnivå 5%.

c) Gjer kort greie for val av modell/analyse for følgjande problemstillingar knytta til motorisk dugleik for barn:

1) Resultata frå dei 3 skulane skal samanliknas både for gutar og for jenter. Ein skal både sjå på om det er signifikante forskjeller mellom skular og mellom gutar og jenter. Som i a) og b) antar vi at målte motoriske kvotienter for barn er uavhengige og normalfordelte variablar.

2) Problemstillinga er den same som i b): Vi skal samanlikne gutar frå 3 skular.
I motsetning til i b) antar vi nå at motorisk dugleik for kvar gut er klassifisert og registrert som enten **god**, **middels** eller **dårleg**. Korleis kan ein nå analysere problemstillinga i b)?

Resultat

X_{1j} (Halden): 100 86 71 80 95 77 104 96 91 82 85 72 92 97 82 85 71 78

$n_1 = 18 \quad \overline{X}_1 = 85.78 \quad \sum (X_{1j} - \overline{X}_1)^2 = 1744.5$

X_{2j} (Sarpsborg): 90 75 111 80 105 91 69 105 108 79 80 83 119 97 105 111 79 76 60 113 80 102 95

$n_2 = 23 \quad \overline{X}_2 = 91.87 \quad \sum (X_{2j} - \overline{X}_2)^2 = 5745.2$

X_{3j} (Nøtterøy): 90 114 110 127 102 72 80 94 93 96 96 85 106 108 108 83 89 89 91 99 111 118

$n_3 = 22 \quad \overline{X}_3 = 98.23 \quad \sum (X_{3j} - \overline{X}_3)^2 = 3810.3$

Totalt gjennomsnitt: $\overline{X} = 92.35$

OPPGÅVE 3

Vi skal studere utviklinga av folketalet i Bø kommune ved hjelp av ein enkel regresjonsmodell

$$Y = \beta_0 + \beta_1 x + U$$

der Y er folketal på tidspunkt nr x ($x = 1, 2, \ldots, 8$) og $U_1, U_2, \ldots, U_8$ er uavhengige og normalfordelte variablar med forventning 0 og standardavvik σ.

Registrering av folketal er foretatt kvar tredje månad, første gong 01.01.2004 og åttande og førebels siste gong 01.10.2005. Sjå Minitab-utskrifta til slutt i oppgåva (neste side). Dataene er henta frå Statistisk sentralbyrå.

a) Skriv opp den estimerte regresjonsmodellen og forklar kort kva dei estimerte parametrane uttrykker.

Har det vært signifikant auke i folketallet i Bø i den aktuelle perioden? Formuler sjølv nullhypotese og alternativ hypotese for å teste dette spørsmålet. Gi konklusjonen når signifikansnivået er valt lik 1%.

Leiinga i kommunen ønsker eit intervall med sikkerhet 95% for folketalet i Bø 01.01.2006. Vil det i dette tilfelle vere eit konfidensintervall eller eit prediksjonsintervall som er eit naturleg svar på kommunens ønske? Grunngi svaret kort og rekn ut det aktuelle intervallet.

Utskrift frå Minitab

Data Display

Dato	Tidspunkt	Folketall
01.01.2004	1	5173
01.04.2004	2	5168
01.07.2004	3	5182
01.10.2004	4	5241
01.01.2005	5	5249
01.04.2005	6	5244
01.07.2005	7	5210
01.10.2005	8	5288

Regression Analysis: Folketall versus Tidspunkt

The regression equation is
Folketall = 5155 + 14,4 Tidspunkt

Predictor	Coef	SE Coef	T	P
Constant	5154,61	20,70	248,99	0,000
Tidspunk	14,393	4,100	3,51	0,013

S = 26,57 R-Sq = 67,3% R-Sq(adj) = 61,8%