

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK 1000 — Innføring i anvendt statistikk.

Eksamensdag: Fredag 1. juni 2007.

Tid for eksamen: 09.00 – 12.00.

Oppgavesettet er på 4 sider.

Vedlegg: Tabell over normalfordeling, tabell over t -fordeling, tabell over binomisk fordeling.

Tillatte hjelpemidler: Lærebok: Moore & McCabe "Introduction to the practice of statistics", 3., 4. el. 5. utgave. Ordliste for bruk i STK 1000, kalkulator.

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Oppgave 1.

Et lite bilutleiefirma disponerer i alt tre biler og leier ut biler for en dag av gangen. La X betegne antall biler som blir leid ut på en dag. Det har vist seg at X kan antas å ha følgende fordeling:

Verdi	0	1	2	3
Sannsynlighet	0.1	0.2	0.3	0.4

a) Beregn forventning μ_X og standardavvik σ_X i fordelingen.

(Fortsettes side 2.)

La Y betegne antall dager i løpet av en uke (7 dager) hvor ingen bil er utleid. Antall biler som leies ut på ulike dager antas å være stokastisk uavhengige.

- b) Finn sannsynlighetsfordelingen til Y . Hva er sannsynligheten for at firmaet ikke leier ut en eneste bil to eller flere dager i løpet av en uke?

Oppgave 2.

Mutasjoner i genet LRP5 er mistenkt å være assosiert med endret benmasse. Et eksperiment med mus er utført for å belyse problemstillingen. 60 forsøksmus ble manipulert til å ha LRP5-mutasjon. Disse ble sammenlignet med 60 normale forsøksmus.

Enkeltmålinger er ikke tilgjengelige, men tabellen nedenfor oppsummerer resultatene for benmasse og kroppsvekt for de tilsammen 120 musene i forsøket.

	Normale		Muterte	
	\bar{x}_1	s_1	\bar{x}_2	s_2
benmasse (mg)	12.6	1.3	11.6	2.5
vekt (g)	22.5	1.2	22.7	1.3

Målinger av benmasse for normale mus kan antas å komme fra en normalfordeling med forventning μ_1 og standardavvik σ_1 . Tilsvarende kan benmasse for muterte mus antas å følge en normalfordeling med forventning μ_2 og standardavvik σ_2 .

- a) Forklar hvorfor vi velger å teste hypotesene

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{mot} \quad H_a : \mu_1 - \mu_2 \neq 0$$

for å finne ut om det er forskjell i benmasse mellom normale og LRP5-muterte mus. Sett opp passende testobservator for hypotesene, og beregn dennes numeriske verdi.

- b) Vis at tilnærmet P-verdi for testen blir mindre enn 0.01 og formuler en konklusjon. Angi spesielt hvor mange frihetsgrader du har brukt ved beregning av P-verdien.
- c) Finn et 90% konfidensintervall for forskjellen mellom de to gruppene i forventet benmasse.
- d) En forsker foreslår at forskjellen i benmasse kan forklares ved forskjell i kroppsvekt. Vis at dette ikke kan stemme fordi det ikke er signifikant forskjell i vekt mellom de to gruppene. Bruk en test der du antar at standardavviket for kroppsvekt i de to gruppene er identiske.
- e) Hadde konklusjonene i punktene ovenfor vært troverdige selv om normalantakelsene skulle vise seg å ikke stemme? Begrunn svaret.

(Fortsettes side 3.)

Oppgave 3.

Målinger av forekomsten av giftstoffet PCB står sentralt i konflikten rundt dypvannsdeponiet ved Malmøykalven i indre Oslofjord. Det finnes 209 “typer” PCB og totalforekomsten av PCB finnes som summen av forekomsten av alle typene. Det er svært kostbart å måle alle typene, og derfor ønskelig å estimere total PCB ved bare å bruke målinger av noen få typer PCB.

I forbindelse med dypvannsdeponiet i Oslofjorden har Statens forurensningstilsyn (SFT) basert sine PCB-målinger på 7 ulike typer PCB, mens den svenske forskeren som motstanderorganisasjonen Neptun har engasjert, har brukt målinger av 79 typer PCB.

Et uavhengig datamateriale fra USA består av målinger av alle 209 typer PCB i 69 forskjellige vannprøver. Vi forsøker å bruke fem utvalgte typer PCB i en multippel regresjon for å predikere total PCB på grunnlag av dette datasettet. Utskrift fra analysen følger vedlagt. Av plasshensyn inkluderer vi kun et residualplott der residualene er plottet mot forklaringsvariabel PCB153. Residualer plottet mot de andre forklaringsvariablene ser tilsvarende ut.

- Formuler modell og antakelser for en slik multippel lineær regresjon. Kommenter antakelsene i lys av vedlagte plott. Finn estimater for parametrene i modellen fra Minitab-utskriften.
- Forklar hvilken hypotese som testes for PCB153 i utskriften under og angi hvordan t-observator (T) og P-verdi (P) er beregnet.
- Gi en begrunnet vurdering av de fem utvalgte PCB-typenes evne til å predikere total PCB-forekomst.

Regression Analysis: PCB versus PCB138, PCB153, PCB180, PCB52, PCB118

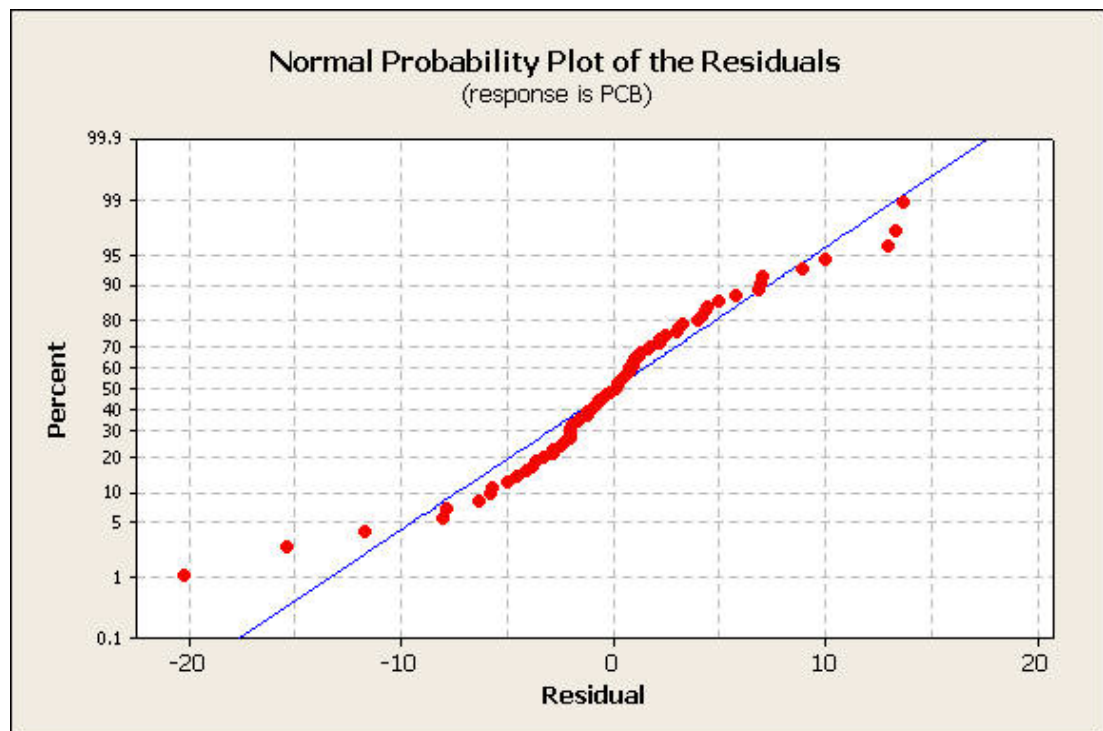
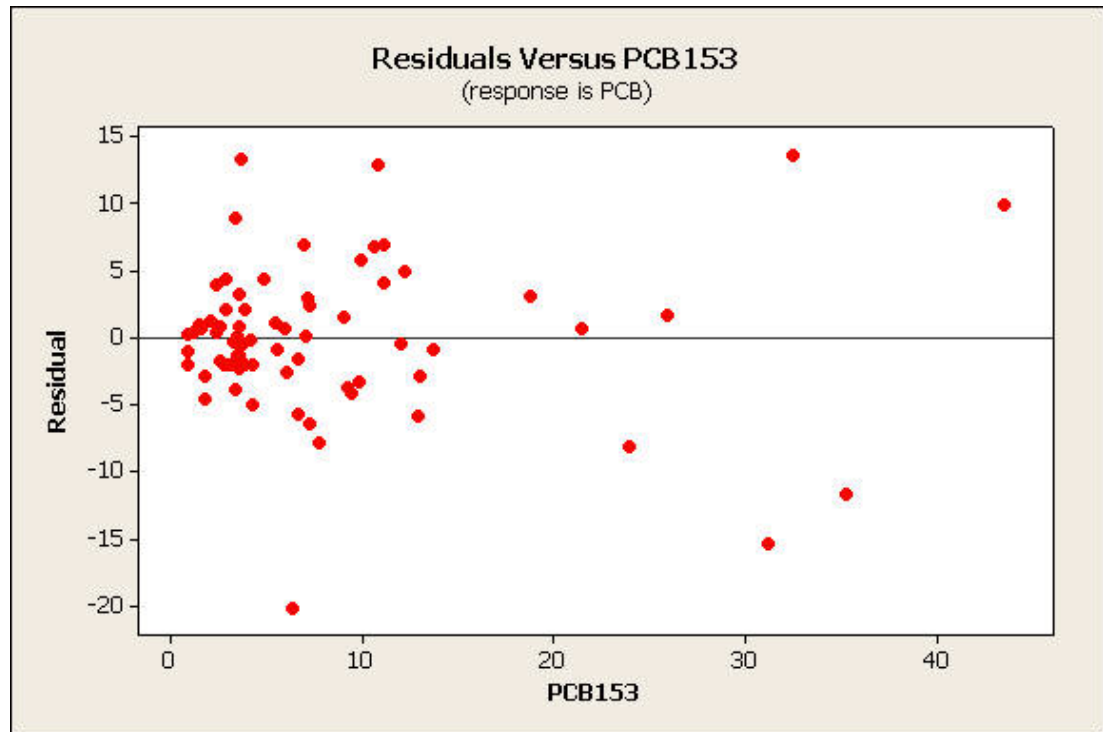
The regression equation is

PCB = 1.74 + 3.49 PCB138 + 0.865 PCB153 + 3.47 PCB180 + 11.8 PCB52 + 3.09 PCB118

Predictor	Coef	SE Coef	T	P
Constant	1.741	1.168	1.49	0.141
PCB138	3.4875	0.4781	7.29	0.000
PCB153	0.8653	0.2607	3.32	0.002
PCB180	3.4726	0.4549	7.63	0.000
PCB52	11.7521	0.6789	17.31	0.000
PCB118	3.0890	0.6308	4.90	0.000

S = 5.93473 R-Sq = 99.1% R-Sq(adj) = 99.0%

(Fortsettes side 4.)



SLUTT