<u>Løsningsforslag INT010 høst 2009</u>

Oppgave 1

a) Forutsetninger:
   a. Begge populasjonene må være normalfordelt
   b. Populasjonene må være uavhengige
   c. Tilfeldig utvalg

b) F-test. Testobservator $F = s_1^2 / s_2^2 = 0{,}12^2 / 0{,}11^2 =$ **1,19**. Forkastningsområde på 5% signifikansnivå med 13 og 8 frihetsgrader: $F > F_{0{,}025,\ 13,\ 8} =$ **4,2** eller $F < 1/F_{0{,}025,8,13} = 1/3{,}39 =$ **0,29**. **Aksepter H_0** om lik varians.

c) Testobservator $t = (x_1 - x_2) / \sqrt{[s_p^2\ (1/n_1 + 1/n_2)]} = (0{,}317\text{-}0{,}210) / \sqrt{[\ 0{,}014 \cdot (1/14 + 1/9)]}$ = **2,15**, da $s_p^2 = (13 \cdot 0{,}12^2 + 8 \cdot 0{,}11^2) /21 = 0{,}014$. Kritisk grense på 5% nivå er med $n_1 + n_2 - 2 = 21$ frihetsgrader $t > t_{0{,}025,\ 21} =$ **2,08** eller $t < - t_{0{,}025,\ 21} = -2{,}08$. **Forkast H_0** (men med liten margin) og påstå at forskjellen i månedlige økning i KPI er signifikant.

Oppgave 2

a) Modell: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$, der ε er normalfordelt med forventning null og standardavvik lik σ.

X_4	X_5	Tilstand
1	0	Dårlig
0	1	Gjennomsnittlig
0	0	God

b) Koeffisienten for bilens alder er -1132$. Dvs. gitt alt annet konstant vil bilens verdi bli redusert med 1132$ per år. Ny bil: $x_1 = 0$, $x_2 = 0$, $x_3 = 1$, $x_4 = x_5 = 0$, da blir prediksjonen y = 17357 + 776 = **18133$**. Modellen synes **urimelig**, bilen vil neppe gå ned i verdi med et konstant beløp hvert år. Videre er nybilprisen i oppgaven satt til 25000$, dermed er prediksjonen for en ny bil dårlig.

c) Plottene på venstre side kan tyde på at residualene ikke er normalfordelte. Videre viser grafen øverst i høyre hjørne antydninger til heteroskedastisitet (variansen til residualene er økende).

d) Når y er venstresidevariabel vil prisen falle lineært med kjørelengde (x_2), dvs. med et fast kronebeløp for hvert år. Med ln(y) som venstresidevariabel vil β_2 gi oss (omtrentlig) **prosentvis endring** i prisen ved endring i kjørelengden (dette fordi at små endringer i ln(y) tilsvarer ca. prosentvis endring).

e) $\ln y = y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$, der ε er normalfordelt med forventning null og standardavvik lik σ. Prediksjon for ny bil: $y = \exp(10 - 0{,}178 \cdot x_1 - 0.00376 \cdot x_2 + 0{,}122 \cdot x_3 - 0{,}217 \cdot x_4 - 0.0651 \cdot x_5) = \exp(10+0{,}122) =$ **24884$**. Denne prediksjonen er mer korrekt i forhold til den vi fant i b). [Vær oppmerksom på at det egentlig ikke er så rett frem å regne om forventningen til prisen basert på forventningen til ln(pris), fordi E(pris) $\neq e^{E(\ln(pris))}$, men vi forventer ikke at studentene utdyper dette].

f) Estimat for Obamas bil : $x_1 = 2$, $x_2 = 21$, $x_3 = 0$, $x_4=x_5=0$, da blir prediksjonen y = exp(10 – 0,178·2 – 0.00376·21) = **14216$**. **PI** er prediksjonsintervallet for verdien av en bestemt bil, derfor vil dette intervallet være relevant for Obama. Dette intervallet gir (exp(9,2528), exp(9,8713) = (10434, 19366), dvs. at 17500$ **ikke** er **urimelig** høyt.

g) 95% konfidensintervall for koeffisienten til x_5: -0,06508 ± 2,06 · 0,07195 = **[-0,2133; 0,0831]**, der vi har benyttet at $t_{0,025,31-6} = t_{0,025,25}$ = **2,06**. Hvis vi tester H_0: $\beta_5 = 0$ mot H_1: $\beta_5 \neq 0$, vil vi dermed ikke forkaste H_0 på 5% nivå. Derfor er det rimelig å si at β_5 ikke har signifikant betydning.

Oppgave 3

a) I en kontingenstabell bør forventet verdi være minst 5 i hver celle ('rule-of-five'). Det er ikke tilfredstilt i den oppgitte tabellen. Ny tabell:

Tilstand / Kategori	GG	Dårlig	SUM
Privat	4 (9,29)	14 (8,71)	18
Forhandler	12 (6,71)	1 (6,29)	13
SUM	16	15	31

De forventete frekvensene er oppgitt i parantes, og beregnet med formelen: e_{ij} = (Rad i total · Kolonne j total) / utvalgsstørrelse, for eksempel er e_{11} = (18·16)/31 = 9,29. Tilsvarende for de andre cellene (trenger bare regne ut e_{12} i tillegg).

b) Benytter testobservator $\chi^2 = \sum_{i=1}^{4} (f_i - e_i)^2 / e_i = (4-9,28)^2 / 9,28 + (12 - 6,71)^2 / 6,71 + (14 - 8,71)^2 / 8,71 + (1-6,29)^2 / 6,29$ = **14,85**. Forkast om $\chi^2 > \chi^2_{0,01,1} = 6,63$ (antall frihetsgrader er $\nu = (2-1)\cdot(2-1) = 1$) dvs. vi **forkaster**, og påstår at det er signifikant forskjell mellom tilstanden til biler som selges gjennom forhandler og privat.

c) **E(T)** = $n_1 (n_1 + n_2 +1) / 2$ = 13 (13 + 18 + 1) / 2 = **208**. $\sigma_T = \sqrt{[\, n_1 n_2(n_1 + n_2 +1) / 12 \,]} = \sqrt{[13\cdot18\cdot(13+18+1) / 12]}$ = **25**. Testobservator Z = T – E(T) / σ_T = (126,5 – 208) / 25 = **-3,26**. Kritisk verdi er Z < - $z_{0,05}$ = - 1,645. **Forkast H_0**.

Oppgave 4

a) Sett for eksempel $Y = X_1 + X_2$, da vet vi at $E(Y) = \mu_1 + \mu_2$ og $Var(Y) = \sigma_1 + \sigma_2 + 2\sigma_{12}$. Bruker nå approksimasjonsformlene; $E(Y) \approx f(\mu_1, \mu_2) = \mu_1 + \mu_2$. Ser at dette holder eksakt. For variansen ser vi at alle de partielle deriverte er lik 1, og dermed er også denne approksimasjonen eksakt.

b) Formlene gir: $E(Y) \approx \mu_1 + \mu_2 \cdot\mu_3$. De partielle deriverte er $\frac{\partial f}{\partial X_1} = 1, \frac{\partial f}{\partial X_2} = \mu_3, \frac{\partial f}{\partial X_3} = \mu_2$, dermed blir $Var(Y) \approx \sigma_1^2 + \mu_3\sigma_2^2 + \mu_2\sigma_3^2 + 2\mu_3\sigma_{12} + 2\mu_2\sigma_{13} + 2\mu_2\mu_3\sigma_{23}$.