Løsningsforslag (ST1201/ST6201 2015, kontinuasjonseksamen)

1.

a) Rimelighetsfunksjonen blir

$$L(\beta) = \beta^{-2n} e^{-\frac{1}{\beta}\sum_{i=1}^{n} X_i} \left(\prod_{i=1}^{n} X_i \right).$$

Logaritme

$$\ln L = -2n \ln \beta - \frac{1}{\beta} \sum_{i=1}^{n} X_i + \ln \left(\prod_{i=1}^{n} X_i \right).$$

Deriverer

$$\frac{\partial \ln L}{\partial \beta} = -\frac{2n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} X_i.$$

Løser ligningen

$$\frac{1}{\beta} \sum_{i=1}^{n} X_i = 2n.$$

Løsningen er SME:

$$\hat{\beta} = \frac{1}{2n} \sum_{i=1}^{n} X_i.$$

b) Her kan man benytte transformasjonsformelen, elle man kan observere at fordelingen til X_i er en gammafordeling med parametre $\alpha = 2$ og β. Derfor kjenner man momentgenererende funksjonen til X_i:

$$M_{X_i}(t) = \left(\frac{1}{1 - \beta t} \right)^2$$

for $t < 1/\beta$. Regneregler for momentgenererende funksjoner gir at

$$M_{Z_i}(t) = \left(\frac{1}{1 - 2t} \right)^2$$

som er momentgenererende funksjon for en χ^2-fordelt stokastisk variabel med fire frihetsgrader. Videre har man at

$$\frac{4n\hat{\beta}}{\beta} = \frac{4n \sum_{i=1}^{n} X_i}{2n\beta} = \frac{2}{\beta} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \frac{2X_i}{\beta}$$

som er en sum av uavhengige χ^2-fordelte stokastiske variabler med fire frihetsgrader hver. Derfor, har vi at $4n\hat{\beta}/\beta \sim \chi^2_{4n}$.

c) Vi skal teste

$$H_0 : \beta = \beta_0 \text{ mot } H_1 : \beta > \beta_0$$

Siden $4n\hat{\beta}/\beta \sim \chi^2_{4n}$ når H_0 er sann, benytter vi $4n\hat{\beta}/\beta$ som testobservator og beslutningen blir:

$$\text{“Forkast } H_0 \text{ hvis } 4n\hat{\beta}/\beta > \chi^2_{\alpha,4n}\text{”}$$

Med de gitte tallene blir $4n\hat{\beta}/\beta = 25.87$ og $\chi^2_{\alpha,4n} = 55.758$. H_0 forkastes ikke.

d) Betegn styrkefunksjonen $\pi(\beta)$. Da

$$\pi(\beta) = P_\beta(4n\hat{\beta}/\beta_0 > \chi^2_{\alpha,4n}) = \pi(\beta) = P_\beta(4n\hat{\beta}/\beta > \chi^2_{\alpha,4n}\beta_0/\beta) = 1 - F_{4n}(\chi^2_{\alpha,4n}\beta_0/\beta)$$

der $F_{4n}(x)$ er kumulativ fordelingsfunksjon foe en χ^2-fordeling med $4n$ frihetsgrader. Med de gitte tallene blir styrkefunksjonen lik

$$\pi(\beta) = 1 - F_{40}(53.018/\beta),$$

og fra tabellene finner vi at teststyrken blir 0.99 hvis $53.018/\beta = \chi^2_{0.01,40} = 22.164$, og β må bli 5.031.

Hvis $\beta = 5.031$ er sannsynligheten for å konkludere med at H_0 skal forkastes lik 0.99.

2.

a) Dette er en ANOVA-tabell for k-utvalg med $k = 4$ og $n_j = 6$ for $j = 1, 2, 3, 4$. Den fullstendige ANOVA-tabellen blir

Kilde	df	SS	MS	F
Betong	$k - 1 = 3$	47203.13	15734.38	2.9
Error	$6 \cdot 4 - 4 = 20$	108671.50	5433.58	
Total	$6 \cdot 4 - 1 = 23$	155874.63		

der

$$\text{SSTR} = \text{MSTR} \cdot 3 = 47203.14,$$

$$\text{MSE} = \text{SSE}/20 = 10861.5/20 = 5433.58,$$

$$\text{SSTOT} = \text{SSTR} + \text{SSE} = 47203.13 + 108671.5 = 155874.63$$

og

$$F = MSTR/MSE = 15734.38/5433.58 = 2.9.$$

b) Testobservatoren F relaterer seg til hypotesene

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \text{ mot } H_1 : \text{ikke alle er like},$$

der μ_i, for $i = 1, 2, 3, 4$, er forventet opptatt fuktighet for betong av type nummer i. Når H_0 er riktig er F Fisher fordelt med 3 og 20 frihetsgrader. Finner kritisk verdi for $\alpha = 0.05$ fra tabell til a være $f_{0.05,3,20} = 3.10$. Beslutningsregelen blir dermed at vi skal forkaste H_0 når $F > 3.10$. Betongdataene gav $F = 2.90 < 3.10$ slik at konklusjonen blir at vi skal ikke forkaste H_0.

c) En to-utvalg t-test baserer seg på at man har observasjoner av stokastiske variabler $X_1, ..., X_n$ og $Y_1, ..., Y_m$ der alle X-ene og Y-ene er uavhengige av hverandre,

$$X_i \sim \mathrm{N}(\mu_X, \sigma^2), \; Y_i \sim \mathrm{N}(\mu_Y, \sigma^2).$$

Man benytter da testobservatoren

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

som har t-fordeling med $n + m - 2$ frihtsgrader under H_0. Variansestimatoren S_p^2 er gitt ved formalen

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_Y^2}{n + m - 2}.$$

Vi lar X-ene og Y-ene være henholdvis data for betong av type 3 og 4. Ved å benytte oppgitte verdier S_X^2 og S_Y^2 i tabellen over får man $s_p^2 = 3648.89$, $t = 3.53$. Man må her benytte en tosidig test slik at kritisk verdi blir $t_{\alpha/2, n+m-2} = t_{0.025,10} = 2.228$. Beslutningsregelen blir dermed at man skal forkaste H_0 dersom $T < -2.228$ eller $T > 2.228$. Vi observerte $t = 3.53 > 2.228$, slik at konklusjonen blir at vi forkaster H_0.

Det er ikke urimelig at vi i punkt **b)** konkluderer med at det ikke er signifikant foskjell mellom forventningsverdiene til de fire utvalgene, mens vi her i punkt **c)** konkluderer med at det er signifikant forskjell mellom forventningsverdiene til utvalg 3 og 4. Vi kan spesielt legge merke til at vi her i punkt **c)** sammenligner de to av de fire utgavene som har størst avvik i gjennosnittsverdi. Vi kan også legge merke til at empirisk varians for utvalg nummer 1 er betydelig større enn for de andre tre utgavene. ANOVA-analysen baserer seg som kjent på antagelsen om lik varians for alle utvalg. Den store empiriske variansen for utvalg nummer 1 vil dermed føre til at estimatert ("pooled") varians i ANOVA-analysen blir betydelig større enn tilsvarende størrelse i t-testen.

3.

a) Rimelighetsfunksjonen blir

$$L(\alpha, \beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2 x_i}} \exp\left(-\frac{(Y_i - \alpha - \beta x_i)^2}{2\sigma_0^2 x_i}\right) =$$

$$= \frac{1}{(2\pi)^{n/2}\sigma_0^n \prod_{i=1}^{n} \sqrt{x_i}} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} \frac{(Y_i - \alpha - \beta x_i)^2}{x_i}\right).$$

Logaritme

$$\ln L = -\frac{n}{2}\ln(2\pi) - \ln\sigma_0 - \frac{1}{2}\sum_{i=1}^{n}\ln x_i - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}\frac{(Y_i-\alpha-\beta x_i)^2}{x_i}.$$

Partiellderiverer med hensyn på hver av α og β og får

$$\frac{\partial \ln L}{\partial \alpha} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}\frac{(Y_i-\alpha-\beta x_i)}{x_i},$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(Y_i-\alpha-\beta x_i).$$

Vi setter de deriverte lik null

$$\sum_{i=1}^{n}\frac{(Y_i-\alpha-\beta x_i)}{x_i} = 0$$

$$\sum_{i=1}^{n}(Y_i-\alpha-\beta x_i) = 0,$$

løser ligningene og får

$$\hat{\beta} = \frac{\bar{Y}\sum_{i=1}^{n}\frac{1}{x_i} - \sum_{i=1}^{n}\frac{Y_i}{x_i}}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i} - n},$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}.$$

b)

$$E(\hat{\beta}) = E\left(\frac{\bar{Y}\sum_{i=1}^{n}\frac{1}{x_i} - \sum_{i=1}^{n}\frac{Y_i}{x_i}}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i} - n}\right) =$$

$$= \frac{1}{\bar{x}\sum_{i=1}^{n}\frac{1}{x_i} - n}\left(E(\bar{Y})\sum_{i=1}^{n}\frac{1}{x_i} - \sum_{i=1}^{n}\frac{E(Y_i)}{x_i}\right) = \beta$$

fordi $E(Y_i) = \alpha + \beta x_i$ og $E(\bar{Y}) = \alpha + \beta\bar{x}$.

c) $\hat{\beta}$ er en lineærkombinasjon av uavhengige normalfordelte stokastiske variabler dvs Y-ene, derfor er $\hat{\beta}$ normalfordelt:

$$\hat{\beta} \sim N(\beta, \mathrm{Var}(\hat{\beta})).$$

Da

$$\frac{\hat{\beta}-\beta}{\sqrt{\mathrm{Var}(\hat{\beta})}} \sim N(0,1)$$

og

$$P\left(-z_{\delta/2} \leq \frac{\hat{\beta}-\beta}{\sqrt{\mathrm{Var}(\hat{\beta})}} \leq z_{\delta/2}\right) = 1-\delta$$

eller

$$P(\hat{\beta}) - z_{\delta/2}\sqrt{\mathrm{Var}(\hat{\beta})} \leq \beta \leq \hat{\beta}) + z_{\delta/2}\sqrt{\mathrm{Var}(\hat{\beta})}) = 1-\delta.$$

Ved å sette inn uttriket for $\mathrm{Var}(\hat{\beta})$ får vi at konfidensintervallet blir

$$\left[\hat{\beta} - z_{\delta/2}\frac{\sigma_0}{\sqrt{n}}\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}{\bar{x}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)-1}}, \hat{\beta} + z_{\delta/2}\frac{\sigma_0}{\sqrt{n}}\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}{\bar{x}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)-1}}\right].$$