



Sensorveiledning:
Vanligvis ikke trekk for følgefeil forutsatt at det som er gjort gir mening. Delvis uttelling kan gis dersom noe er gjort delvis riktig så det kan utvises skjønn innenfor grensene spesifisert nedenfor. Karaktergrenser etter prosentvurderingsmetoden men med grense for E satt ned til 37 poeng.

10 poeng: Automatisk rettet, 2 poeng per punkt.

TMA4245, juni 2023, løsningskisse
(versjon 8. august 2023, 10:18)

Oppgave 1

a) $P(-2 < X < 0.5) = F(0.5) - F(-2) = 0.5^3 - 0 = 0.125.$

b) Tettheten til X blir $f(x) = 3x^2$ for $0 < x < 1$ slik at $EX = \int_0^1 xf(x)dx = \int_0^1 3x^3 dx = 3/4 = 0.75.$

c) Vi finner først $E(X^2) = \int_0^1 x^2 f(x)dx = \int_0^1 3x^4 dx = 3/5$ som gir at $\text{Var } X = E(X^2) - (EX)^2 = 3/5 - 9/16 = 0.0375.$

d) Standardavviket er dermed $\text{Std } X = \sqrt{0.0375} = 0.194.$

e) Øvre %5-kvantil q tilfredstiller ligningen

$$P(X > q) = 0.05$$

som betyr at

$$1 - F(q) = 0.05$$

slik at

$$q = F^{-1}(1 - 0.05) = (1 - 0.05)^{1/3} = 0.9830$$

1d) Poenggivningen er overstyrt slik at de som her har beregnet standardavviket til $\sqrt{0.038} = 0.1949359 \approx 0.195$ også har fått 2 poeng for dette.

Oppgave 2

Som oppgave 1

a) $P(Y < 0) = P(Z < \frac{0-10}{5}) = \Phi(-2) = 0.0228.$

b) Siden Y og dermed Y^2 begge er kontinuerlig fordelt er sannsynligheten for Y^2 tar eksakt verdien 100 lik null.

c) Legg merke til at $Y^2 - 20Y$ er et annengradspolynom med positiv annengrads-koeffisient og nullpunkter i $Y = 0$ og i $Y = 20$. Dermed er $P(Y^2 - 20Y < 0) = P(0 < Y < 20) = P(\frac{0-10}{5} < Z < \frac{20-10}{5}) = \Phi(2) - \Phi(-2) = 0.9772 - 0.0228 = 0.954$

d) Siden $\text{Var } Y = E(Y^2) - (EY)^2$ er $E(Y^2) = \text{Var } Y + (EY)^2 = 5^2 + 10^2 = 125.$

e) Tettheten til en normalfordelt variabel er gitt ved $f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}5} e^{-1/2} = 0.0483$ i $y = 15$ når $\sigma = 5$ og $\mu = 10.$

Oppgave 3

Som oppgave 1

- a) Lov om totalsannsynlighet gir at $P(Y = 2) = P(Y = 2 \cap X = 0) + P(Y = 2 \cap X = 1) = 0.1 + 0 = 0.1$
- b) Definisjon av beginget sannsynlighet og lov om totalsannsynlighet gir at $P(Y = 0 | X = 1) = \frac{P(Y=0 \cap X=1)}{P(X=1)} = \frac{0.4}{0.4+0.2+0} = 0.667$
- c) Definisjon av forventning til diskret stokastisk variabel gir $E(Y) = \sum_{y=0}^2 yP(Y = y) = 0 \cdot 0.5 + 1 \cdot 0.4 + 2 \cdot 0.1 = 0.6$
- d) Setning om forventning til funksjoner av to diskrete stokastiske variable gir $E(XY) = \sum_{x=0}^1 \sum_{y=0}^2 xyP(X = x \cap Y = y) = \dots + 1 \cdot 1 \cdot 0.2 + \dots = 0.2$
- e) Videre er $EX = 0.6$ og $\text{Var } X = 0.6 \cdot 0.4 = 0.24$ (siden X er Bernoulli), $\text{Cov}(X, Y) = E(XY) - (EX)(EY) = 0.2 - 0.6 \cdot 0.6 = -0.16$, $E(Y^2) = 0.4 + 4 \cdot 0.1 = 0.8$ slik $\text{Var } Y = E(Y^2) - (EY)^2 = 0.8 - 0.6^2 = 0.44$. Dette gir at

$$\text{corr}(5X - 2, 10Y + 3) = \text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var } X} \sqrt{\text{Var } Y}} = \frac{-0.16}{\sqrt{0.44} \cdot \sqrt{0.24}} = -0.492.$$

Oppgave 4

- a) Betrakter vi standardavviket som kjent og lik σ_0 har vi på vanlig måte at $\bar{X}_n \sim N(\mu, \sigma_0^2/n)$ slik at pivotalen

$$\frac{\bar{X}_n - \mu}{\sigma_0/\sqrt{n}}$$

er standard normalfordelt. Dermed er

$$P(-z_{\alpha/2} < \frac{\bar{X}_n - \mu}{\sigma_0/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$$

og det følger da at

$$\bar{X}_n \pm z_{\alpha/2} \sigma_0 / \sqrt{n} \quad (1)$$

er et $(1-\alpha)$ - konfidens intervall for μ . Innsetting gir intervallet $(1991.1, 1998.9)\text{mAh}$.

Det kan vises at hendelsen at intervallet inneholder μ_0 er identisk med hendelsen at vi ikke forkaster $H_0 : \mu = \mu_0$ i en tosidig test. Vi ville dermed forkastet H_0 .

- b) Mobiltelefonprodusenten ønsker å finne støtte for den alternativ hypotesen $H_1 : \sigma^2 > \sigma_0^2$ (hvor $\sigma_0 = 10\text{mAh}$) mens batteriprodusentens påstand her spiller rollen som nullhypotesen $H_0 : \sigma^2 = \sigma_0^2$. Under H_0 har vi at testobservatoren

$$\chi^2 = \frac{S_n^2(n-1)}{\sigma_0^2} \quad (2)$$

Poeng:

- 1: Riktig fordeling $\bar{X}$
 2: Riktig standard normal pivotale
 1: Riktig utgangspunkt
 2: Riktig intervall
 1: Riktig tallsvar
 3: Forkaste H_0 samme hendelse

Poeng:

- 2: Riktig H_0 og H_1
 3: Riktig testobservator
 2: Riktig forkastningsområdet
 1: Riktig observert verdi
 2: Riktig konklusjon

er kji-kvadrat med $n - 1$ frihetsgrader. Under H_1 vil χ^2 tendere til å ta større verdier enn forventet under H_0 . Vi forkaster dermed dersom

$$\chi^2 > \chi_{\alpha, n-1}^2$$

hvor den kritiske verdien i vårt tilfelle er $\chi_{0.05, 25-1}^2 = 36.415$. Observert verdi av χ^2 blir $15.48^2 \cdot (25 - 1)/10^2 = 57.51$ og vi forkaster H_0 .

c) Under H_1 er

$$\frac{S_n^2(n-1)}{\sigma^2} \quad (3)$$

i stedet for (2) kji-kvadratfordelt. Dette gir at teststyrken kan uttrykkes som

$$\begin{aligned} \gamma &= P(\text{Forkaste } H_0 | H_1) \\ &= P\left(\frac{S_n^2(n-1)}{\sigma_0^2} > \chi_{\alpha, n-1}^2 | H_1\right) \\ &= P\left(\frac{S_n^2(n-1)}{\sigma^2} > \frac{\sigma_0^2}{\sigma^2} \chi_{\alpha, n-1}^2 | H_1\right) \\ &= 1 - F_{\chi_{n-1}^2}\left(\frac{\sigma_0^2}{\sigma^2} \chi_{\alpha, n-1}^2\right) \\ &= 1 - F_{\chi_{24}^2}(16.18). \end{aligned}$$

Poeng:

2: $P(\text{forkastningsregel} | H_1)$

5: Riktig restandardisering

2: Riktig algebraisk uttrykk

1: Riktig etter innsetting

d) Siden

$$\bar{X}_n - \bar{X}_N = \sum_{i=1}^n \left(\frac{1}{n} - \frac{1}{N}\right) X_i - \sum_{i=n+1}^N \frac{1}{N} X_i$$

er $\bar{X}_n - \bar{X}_N$ en lineærkombinasjon av $X_1, X_2, \dots, X_N$. Siden $X_1, X_2, \dots, X_N$ er normalfordelte er dermed $\bar{X}_n - \bar{X}_N$ også normalfordelt. Forventningen blir

$$E(\bar{X}_n - \bar{X}_N) = \mu - \mu = 0.$$

Regneregler for varians til lineærkombinasjon av uavhengige variable gir at

$$\begin{aligned} \text{Var}(\bar{X}_n - \bar{X}_N) &= \text{Var}\left(\sum_{i=1}^n \left(\frac{1}{n} - \frac{1}{N}\right) X_i - \sum_{i=n+1}^N \frac{1}{N} X_i\right) \\ &= \sum_{i=1}^n \left(\frac{1}{n} - \frac{1}{N}\right)^2 \text{Var } X_i + \sum_{i=n+1}^N \frac{1}{N^2} \text{Var } X_i \\ &= \sigma^2 \left(n \left(\frac{1}{n} - \frac{1}{N}\right)^2 + (N-n) \frac{1}{N^2} \right) \\ &= \sigma^2 \left(n \left(\frac{1}{n^2} - \frac{2}{nN} + \frac{1}{N^2} \right) + \frac{1}{N} - \frac{n}{N^2} \right) \\ &= \sigma^2 \left(\frac{1}{n} - \frac{2}{N} + \frac{n}{N^2} + \frac{1}{N} - \frac{n}{N^2} \right) \\ &= \sigma^2 \left(\frac{1}{n} - \frac{1}{N} \right) \\ &= \frac{\sigma^2}{n} \left(1 - \frac{n}{N} \right). \end{aligned}$$

Poeng:

2: Riktig omskriving til eller argument for lin.komb.

1: Normalfordeling og riktig argument

1: Riktig forventning

3: Riktig utledning av varians

2: Riktig pivotal (t-fordelt) og

start utledning av intervall. 1

poeng trekk hvis standardnormal

pivotal basert på kjent σ i stedet

er brukt med mindre dette er en

følgefeil av å ikke ha forkastet

$H_0 : \sigma = \sigma_0$ i punkt b)

1: Riktig intervall

Et enklere alternativ er å bruke setning om varians til lineærkombinasjon av avhengige variable og setning om kovarians mellom lineærkombinasjoner av stokastiske variable som gir

$$\begin{aligned}\text{Var}(\bar{X}_n - \bar{X}_N) &= \text{Var} \bar{X}_n + \text{Var} \bar{X}_N - 2 \text{Cov}(\bar{X}_n, \bar{X}_N) \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2}{N} - 2n \frac{\sigma^2}{nN} \\ &= \sigma^2 \left(\frac{1}{n} - \frac{1}{N} \right).\end{aligned}$$

Det følger at

$$\frac{\bar{X}_n - \bar{X}_N}{\sigma \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right)}}$$

er standard normalfordelt, og uavhengig av (3) som er kji-kvadrat med $n - 1$ frihetsgrader, slik at pivotalen

$$\frac{\bar{X}_n - \bar{X}_N}{S_n \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right)}} = \frac{\frac{\bar{X}_n - \bar{X}_N}{\sigma \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right)}}}{\sqrt{\frac{S_n^2(n-1)}{\sigma^2} / (n-1)}}$$

er på formen $\frac{Z}{\sqrt{V/\nu}}$ og dermed t -fordelt med $n - 1$ frihetsgrader. Dette gir at

$$P \left(-t_{\alpha/2, n-1} < \frac{\bar{X}_n - \bar{X}_N}{S_n \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right)}} < t_{\alpha/2, n-1} \right) = 1 - \alpha$$

som kan omskrives til

$$P \left(\bar{X}_n - t_{\alpha/2, n-1} S_n \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right)} < \bar{X}_N < \bar{X}_n + t_{\alpha/2, n-1} S_n \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right)} \right) = 1 - \alpha$$

som betyr at $\bar{X}_n \pm t_{\alpha/2, n-1} S_n \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right)}$ er et $(1 - \alpha)$ -prediksjonsintervall for $\bar{X}_N$. Det kan bemerkes at intervallet som forventet blir identisk det vanlige konfidensintervallet for μ for ett-utvalgs normalfordelte data med ukjent varians i grensen når $N \rightarrow \infty$ mens det får lengde null når $n = N$.

Innsetting gir intervallet (1990.5, 1999.5)mAh.

Oppgave 5

- a) Vi har en binomisk forsøksrekke hvor feil ("suksesser") inntreffer uavhengig og med samme sannsynlighet q . Når Y_i er antall delforsøk til og med andre feil inntreffer er Y_i dermed negativt binomisk fordelt med parametere $r = 2$ og q . For $r = 2$ er sannsynlighetsmassefunksjonen

$$f(y) = (y-1)q^2(1-q)^{y-2}$$

og for $q = 1/10$ blir $P(Y_i = 5) = f(5) = (5-1)/10^2(9/10)^3 = 0.02916$.

Poeng:

2: Uavhengige delforsøk med to utfall

2: Samme sannsynlighet

2: Antall forsøk til r 'te suksess

2: Riktig uttrykk for $f(5)$

2: Riktig tallsvar

b) Rimelighetsfunksjonen blir

$$L(q) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n (y_i - 1) q^2 (1 - q)^{y_i - 2} = q^{2n} (1 - q)^{\sum_{i=1}^n y_i - 2n} \prod_{i=1}^n (y_i - 1)$$

og

$$\ln L(q) = 2n \ln q + \left(\sum_{i=1}^n y_i - 2n \right) \ln(1 - q) + \sum_{i=1}^n \ln(y_i - 1)$$

som har sitt maksimum der

$$\frac{d \ln L}{dq} = \frac{2n}{q} - \frac{\sum_{i=1}^n y_i - 2n}{1 - q} = 0.$$

Løst m.h.p. q får vi at SME av q er

$$\hat{q} = \frac{2n}{\sum_{i=1}^n Y_i}$$

Innsatt får vi estimatet $\hat{q} = 20/1601 = 0.0125$

c) En negativt binomisk variabel med parametre r og q kan betraktes som en sum av r uavhengige (geometrisk fordelte) variable med samme endelige forventning og varians. I grensen når $r \rightarrow \infty$ gjelder dermed sentralgrenseteoremet. For store r (dog kanskje ikke $r = 2$) kan fordelingen til Y_i derfor tilnærmes med en normalfordeling med forventning r/q og varians $r(1 - q)/q^2$ (formelsamling). Dette gir at

$$P(Y_i > m) \approx P\left(Z > \frac{m - r/q}{\sqrt{r(1 - q)/q^2}}\right)$$

Innsatt $m = 500$, $r = 2$ og $\hat{q} = 0.0125$ får vi

$$P(\widehat{Y_i > m}) = P(Z > \frac{500 - 160}{112}) = 1 - \Phi(3.02) = 0.0012.$$

Lar vi X være totalt antall feil på samtlige ladere produsert på en gitt produksjonsdag er $X \sim \text{bin}(500, q)$. Vi innser da at hendelsen $Y_i > 500$ er samme hendelse som at vi har færre enn to feil i løpet av hele produksjonsdagen finner vi at, d.v.s. hendelsen $X < 2$. Dermed er

$$P(Y_i > m) = P(X < 2) = P(X = 0) + P(X = 1) = (1 - q)^m + mq(1 - q)^{m-1}$$

Innsatt gir dette

$$P(\widehat{\widehat{Y_i > m}}) = 0.0136$$

som tyder på at normaltilnærming fungerer dårlig i dette tilfellet.

Poeng:

1: $L(q) = \prod f(y_i)$

2: Riktig $L(q)$

1: Riktig $\ln L(q)$

2: Riktig derivert

2: Riktig videre utledning

2: Riktig tallsvar

Poeng:

3: Sum av geometrisk (uavhengige identiske med samme forventning og varians)

1: Sentralgrenseteoremet

2: Riktig standardisering

1: Riktig tallsvar

3: Riktig eksakt svar via kumulativ for binomisk som i LF eller via

$P(Y_i > m) = 1 - \sum_{y=2}^m P(Y_i = y)$.