Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

LØSNINGSFORSLAG TIL EKSAMEN I FAG TMA4240 STATISTIKK
Torsdag 9.desember 2004

Oppgave 1 Bronsebolter

a) Vi jobber med X som er normalfordelt med forventning $\mu_x = 85$ gram og varians $\sigma^2 = 1$ gram^2.

Hva er sannsynligheten for at kobberinnholdet i en tilfeldig valgt bronsebolt er mindre enn 84 gram?

$$\begin{aligned} P(X < 84) &= P(X \leq 84) = P(\frac{X-85}{1} \leq \frac{84-85}{1}) = P(Z \leq -1) \\ &= \Phi(-1) = \underline{\underline{0.1587}} \end{aligned}$$

Finn et tall, k, slik at sannsynligheten er 0.01 for at kobberinnholdet i en tilfeldig valgt bronsebolt er større enn k.

$$\begin{aligned} P(X > k) &= 0.01 \\ P(\frac{X-85}{1} > \frac{k-85}{1}) &= 0.01 \\ \frac{k-85}{1} &= z_{0.01} = 2.326 \\ k &= 2.326 \cdot 1 + 85 = \underline{\underline{87.326}} \end{aligned}$$

Vi ser på kobberinnholdet i to tilfeldig valgte og uavhengige bronsebolter, som vi kaller X_1 og X_2, og skal se på differansen $X_1 - X_2$. Siden X_1 og X_2 begge er normalfordelte og uavhengige så er også $X_1 - X_2$ normalfordelt, med forventningsverdi og varians som følger:

$$\begin{aligned} \mathrm{E}(X_1 - X_2) &= \mathrm{E}(X_1) - \mathrm{E}(X_2) = \mu_x - \mu_x = 0 \\ \mathrm{Var}(X_1 - X_2) &= \mathrm{Var}[X_1 + (-1) \cdot X_2] = \mathrm{Var}(X_1) + (-1)^2 \cdot \mathrm{Var}(X_2) \\ &= \sigma^2 + \sigma^2 = 2\sigma^2 = 2 \end{aligned}$$

da $\sigma^2 = 1$.

Hva er sannsynligheten for at kobberinnholdet i de to bronseboltene avviker med mer enn 1.5 gram?

$$\begin{aligned} P(|X_1 - X_2| > 1.5) &= P(X_1 - X_2 < -1.5) + P(X_1 - X_2 > 1.5) \\ &= 2 \cdot P(X_1 - X_2 < -1.5) = 2 \cdot P(\frac{X_1 - X_2 - 0}{\sqrt{2}} < \frac{-1.5 - 0}{\sqrt{2}}) \\ &= 2 \cdot \Phi(\frac{-1.5}{\sqrt{2}}) = 2 \cdot \Phi(-1.06) = 2 \cdot 0.1446 = \underline{\underline{0.2892}} \end{aligned}$$

b) En god estimator $\hat{\theta}$ er en estimator som er

- forventningsrett, dvs. $E(\hat{\theta}) = \theta$, og
- har liten varians, dvs. $\mathrm{Var}(\hat{\theta})$ er liten.

Vi liker veldig godt hvis variansen minker når antall observasjoner som estimatoren er basert på øker, og at variansen går mot 0 når antallet observasjoner går mot uendelig (konsistens).

To aktuelle estimatorer for σ^2 er $\widehat{\sigma}^2$ og S^2. Begge estimatorene er funksjoner av $\sum_{i=1}^{n}(X_i - \bar{X})^2$ og for å beregne forveningsverdi og varians til estimatorene så bruker vi at

$$V = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

er kji-kvadratfordelt med $(n-1)$ frihetsgrader (se formelsamlingen side 27). Vi vet videre at $\mathrm{E}(V) = (n-1)$ og $\mathrm{Var}(V) = 2 \cdot (n-1)$.

Vi kan nå uttrykke S^2 og $\hat{\sigma}^2$ som funksjoner av V;

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{V\sigma^2}{n} \\ S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{V\sigma^2}{n-1} \end{aligned}$$

Dermed forventning til de to estimatorene:

$$\begin{aligned} \mathrm{E}(\widehat{\sigma}^2) &= \mathrm{E}(\frac{V\sigma^2}{n}) = \frac{\sigma^2}{n}\mathrm{E}(V) = \frac{n-1}{n}\sigma^2 \\ \mathrm{E}(S^2) &= \mathrm{E}(\frac{V\sigma^2}{n-1}) = \frac{\sigma^2}{n-1}\mathrm{E}(V) = \frac{n-1}{n-1}\sigma^2 = \sigma^2 \end{aligned}$$

og videre variansen:

$$\begin{aligned} \mathrm{Var}(\widehat{\sigma}^2) &= \mathrm{Var}(\frac{V\sigma^2}{n}) = \frac{\sigma^4}{n^2}\mathrm{Var}(V) = \frac{2 \cdot (n-1)}{n^2}\sigma^4 \\ \mathrm{Var}(S^2) &= \mathrm{Var}(\frac{V\sigma^2}{n-1}) = \frac{\sigma^4}{(n-1)^2}\mathrm{Var}(V) = \frac{2 \cdot (n-1)}{(n-1)^2}\sigma^4 = \frac{2 \cdot \sigma^4}{(n-1)} \end{aligned}$$

Vi ser at S^2 er forventningsrett, mens $\widehat{\sigma}^2$ er skjev. Siden $\frac{n-1}{n} < 1$ vil $\widehat{\sigma}^2$ i det lange løp underestimere σ^2.

Variansen til begge estimatorene avtar når antall observasjoner øker. Estimatoren $\widehat{\sigma}^2$ har mindre varians enn S^2 for alle verdier av n.

Begge estimatorene er konsistente.

Hvis vi legger mest vekt på at vi ønsker en estimator som er forventningsrett så ville vi velge S^2, mens er det viktigst med minst mulig varians så ville vi velge $\widehat{\sigma}^2$.

c) Vi jobber med to uavhengige normalfordelte utvalg med samme varians.

Vi har Bronsespesialisten: $X_1, X_2, ..., X_n$ u.i.f. normal med $\mathrm{E}(X_i) = \mu_x$ og $\mathrm{Var}(X_i) = \sigma^2$, og Metalleksperten: $Y_1, Y_2, ..., Y_m$ u.i.f. normal med $\mathrm{E}(Y_j) = \mu_y$ og $\mathrm{Var}(Y_j) = \sigma^2$.

Vi ønsker å undersøke om forventningsverdien til kobberinnholdet i bronsebolter fra Metalleksperten er lavere enn kobberinnholdet i bronsebolter fra Bronseeksperten.

Null- og alternativ hypotese:

$$\begin{aligned} H_0 &: \mu_x = \mu_y \qquad & H_1 &: \mu_x > \mu_y \\ H_0 &: \mu_x - \mu_y = 0 \qquad & H_1 &: \mu_x - \mu_y > 0 \end{aligned}$$

Det er også mulig å sette $\mu_x \leq \mu_y$ som H_0, men alle beregningene blir uforandret.

De ukjente parameterene er μ_x, μ_y og σ^2, og vi setter opp følgende estimatorer:

$$\begin{aligned} \hat{\mu}_x &= \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\ \hat{\mu}_y &= \frac{1}{m}\sum_{j=1}^{n} Y_j = \overline{Y} \\ S_p^2 &= \frac{1}{n+m-2}[\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{m}(Y_j - \overline{Y})^2] \end{aligned}$$

Vi vet at under H_0 så er

$$T_0 = \frac{(\overline{X} - \overline{Y}) - 0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \quad t\text{-fordelt med } (n+m-2) \text{ frihetsgrader.}$$

Vi vil forkaste H_0 når $T_0 \geq k$, der konstanten k finnes slik at Type-I feilen er kontrollert på nivå α.

$$\begin{aligned} P(T_0 \geq k | H_0 \text{ sann}) &\leq \alpha \\ k &\leq t_{\alpha,(n+m-2)} \end{aligned}$$

der $t_{\alpha,(n+m-2)}$ er α-kvantilen i en t-fordeling med $n + m - 2$ frihetsgrader.

Forkastningsmråde: Forkast H_0 når $T_0 \geq t_{\alpha,(n+m-2)}$.

Når $\alpha = 0.05$ og $n = 10$, $m = 10$, er $t_{0.05,18} = 1.734$. Innsatt data fra tabell 1 i oppgaveteksten har vi:

$$\begin{aligned} \overline{x} &= \frac{850.75}{10} = 85.075 \\ \overline{y} &= \frac{842.10}{10} = 84.21 \\ s_p^2 &= \frac{1}{18}[\sum_{i=1}^{10}(x_i - \overline{x})^2 + \sum_{j=1}^{10}(y_j - \overline{y})^2] = \frac{1}{18}(8.19 + 9.70) = 0.994 \\ t_0 &= \frac{\overline{x} - \overline{y} - 0}{s_p\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{85.075 - 84.21 - 0}{\sqrt{0.994} \cdot \sqrt{\frac{1}{10} + \frac{1}{10}}} = 1.94 \end{aligned}$$

Siden $t_0 = 1.94 > t_{0.05,18} = 1.734$ så forkaster vi H_0 på nivå $\alpha = 0.05$, og konkluderer med at kobberinnholdet i bronseboltene fra Metalleksperten er lavere enn kobberinnholdet i bronseboltene fra Bronsespesialisten.

Når vi forkastet H_0 på nivå 0.05 så betyr det at p-verdien må være mindre enn 0.05. P-verdien er gitt som

$$P(T_0 > t_0 | H_0 \text{ sann}) = P(T_0 > 1.94 | \mu_x - \mu_y = 0) = 1 - P(T_0 \leq 1.94 | \mu_x - \mu_y = 0)$$

Fra tabell 2 i oppgaven så slår vi opp på $P(T \leq t)$ med $t = 1.94$ og $\nu = 18$, og finner 0.966, som gir p-verdi $1 - 0.966 = \underline{\underline{0.034}}$.

d) Et 95 % prediksjonintervall er et intervall som med 95 % sannsynlighet inneholder en ny observasjon. En ny observasjon fra Bronsespesialisten, kalt X_0, er normalfordelt med forventning μ_x og varians σ^2. Vi benytter $\hat{\mu}_x = \overline{X}$ som estimator for μ_x og S_p^2 som estimator for σ^2.

For å utlede prediksjonsintervallet starter vi med $\overline{X} - X_0$, som er normalfordelt med

$$\begin{aligned} \mathrm{E}(\overline{X} - X_0) &= \mathrm{E}(\overline{X}) - \mathrm{E}(X_0) = 0 \\ \mathrm{Var}(\overline{X} - X_0) &= \mathrm{Var}(\overline{X}) + \mathrm{Var}(X_0) = \frac{\sigma^2}{n} + \sigma^2 = (1 + \frac{1}{n})\sigma^2 \end{aligned}$$

Hvis σ^2 var kjent ville dermed

$$\frac{\overline{X} - X_0 - 0}{\sqrt{1 + \frac{1}{n}}\sigma}$$

vært standard normalfordelt. Nå er σ^2 ukjent, og vi benytter estimatoren S_p^2. Vi ser da på

$$\frac{\overline{X} - X_0 - 0}{\sqrt{1+\frac{1}{n}}S_p}$$

som er t-fordelt med $n+m-2$ frihetsgrader.

Et $(1-\alpha)100\%$ predisjonsintervall for X_0 blir da:

$$P(-t_{\frac{\alpha}{2},(n+m-2)} < \frac{\overline{X} - X_0 - 0}{\sqrt{1+\frac{1}{n}}S_p} < t_{\frac{\alpha}{2},(n+m-2)}) = 1-\alpha$$

$$P(\overline{X} - t_{\frac{\alpha}{2},(n+m-2)}\sqrt{1+\frac{1}{n}}S_p < X_0 < \overline{X} + t_{\frac{\alpha}{2},(n+m-2)}\sqrt{1+\frac{1}{n}}S_p) = 1-\alpha$$

der $P(T > t_{\frac{\alpha}{2},(n+m-2)}) = \frac{\alpha}{2}$ i t-fordelingen med $n+m-2$ frihetsgrader.

NB: vi hadde fått det samme intervallet hvis vi hadde startet med $X_0 - \overline{X}$ på grunn av symmetri.

Med $n+m-2 = 18$ og $\alpha = 0.05$ blir $t_{\frac{\alpha}{2},(n+m-2)} = 2.101$ og numeriske verdier for intervallet blir $[82.88, 87.27]$, da

$$\overline{x} - t_{\frac{\alpha}{2},(n+m-2)}\sqrt{1+\frac{1}{n}}s_p = 85.075 - 2.101\sqrt{1+\frac{1}{10}}0.997 = 82.88$$

$$\overline{x} + t_{\frac{\alpha}{2},(n+m-2)}\sqrt{1+\frac{1}{n}}s_p = 85.075 + 2.101\sqrt{1+\frac{1}{10}}0.997 = 87.27$$

Det tilsvarende intervallet for Metalleksperten blir $[82.01, 86.41]$

$$\overline{y} - t_{\frac{\alpha}{2},(n+m-2)}\sqrt{1+\frac{1}{m}}s_p = 84.21 - 2.101\sqrt{1+\frac{1}{10}}0.997 = 82.01$$

$$\overline{y} + t_{\frac{\alpha}{2},(n+m-2)}\sqrt{1+\frac{1}{m}}s_p = 84.21 + 2.101\sqrt{1+\frac{1}{10}}0.997 = 86.41$$

Bronsebolten som ble brukt til å knuse vinduet hos Metalleksperten ble målt til å ha et kobberinnhold på 86.30 gram. Kan du ut fra intervallene du har laget over si noe om hvilken produsent som kan ha laget bronsebolten?

95% prediksjonsintervallet for en ny observasjon fra Metalleksperten og 95% prediksjonsintervallet for en ny observasjon fra Bronsespesialisten inneholder begge verdien 86.30. Det er dermed mulig at bronsebolten kan komme fra begge produsentene.

Det vil ikke bli trukket i poeng hvis intervallet for Bronsespesialisten er laget med $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ som estimator for σ^2 og for Metalleksperten med $S_y^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \bar{Y})^2$ som estimator for σ^2. Men da vil man jobbe med en t-fordeling med hhv. $n-1$ og $m-1$ frihetsgrader, og intervallene ville blitt litt bredere.

Oppgave 2 Pyramidespillet

a) En geometrisk fordeling beskrives ved :

- Vi skal utføre et (på forhånd) ukjent antall forsøk.
- Hvert forsøk har to mulige utfall, "suksess" eller "fiasko".
- Sannsynligheten for "suksess" er den samme i alle forsøk.
- Forsøkene er uavhengige av hverandre.
- Vi avslutter første gang vi oppnår suksess.

I vårt tilfelle ser vi på X, som er antall personer Ole Petter må spørre inntil den første personen seier ja til å bli med i pyramidespillet. Vi må ha at:

- Vi vet ikke på forhånd hvor mange personer vi må spørre, altså antall forsøk er ukjent.
- Enten vil en person som blir spurt om å melde seg inn gjøre det (suksess) eller ikke (fiasko).
- Sannsynligheten for at en tilfeldig valgt person svarer ja er p, og denne sannsynligheten er den samme for alle personer vi spør.
- Personer som blir spurt om å bli med, må svare uten påvirkning fra andre, slik at hver avgjørelse blir uavhengig. (Vi spør hver person bare en gang.)
- Vi avslutter når en person sier seg villig til å bli med

Vi finner at det er rimelig å anta at disse betingelsene er oppfylt, og dermed er X er geometrisk fordelt med parameter p.

Vi finner i formelsamlingen at

$$E(X) = \frac{1}{p}$$

Med $p = 1/3$ får vi dermed

$$E(X) = \frac{1}{p} = \frac{1}{1/3} = \underline{\underline{3}}$$

Punktsannsynligheten i en geometrisk fordeling er oppgitt ved

$$f(x) = p(1-p)^{x-1}\ ;\ x = 1, 2, \ldots$$

Vi vil finne sannsynligheten for at X er større enn fem, når $p = 1/3$.

$$
\begin{aligned}
P(X>5) &= 1-P(X\le 5) = 1-\sum_{x=1}^{5} f(x)\\
&= 1-\left[p(1-p)^{1-1}+p(1-p)^{2-1}+p(1-p)^{3-1}+p(1-p)^{4-1}+p(1-p)^{5-1}\right]\\
&= 1-\left[\frac{1}{3}\left(\frac{2}{3}\right)^{0}+\frac{1}{3}\left(\frac{2}{3}\right)^{1}+\frac{1}{3}\left(\frac{2}{3}\right)^{2}+\frac{1}{3}\left(\frac{2}{3}\right)^{3}+\frac{1}{3}\left(\frac{2}{3}\right)^{4}\right]\\
&= 1-\left(\frac{1}{3}+\frac{2}{9}+\frac{4}{27}+\frac{8}{81}+\frac{16}{243}\right)=\underline{\underline{\frac{32}{243}\approx 0.1317}}
\end{aligned}
$$

b) Vi skal finne sannsynlighetsmaksimeringsestimatoren (SME) for p basert på n uavhengige observasjoner $X_1, X_2, \ldots, X_n$.

Siden vi har uavhengige observasjoner blir rimelighetsfunksjonen produktet av alle punktsannsynlighetene

$$
\begin{aligned}
L(p) &= L(p;x_1,x_2,\ldots,x_n) = f(x_1,x_2,\ldots,x_n)\\
&= f(x_1;p)f(x_2;p)\cdots f(x_n;p)\\
&= \prod_{i=1}^{n} f(x_i;p) = \prod_{i=1}^{n} p(1-p)^{x_i-1}\\
&= p^n(1-p)^{(\sum_{i=1}^{n} x_i)-n}
\end{aligned}
$$

Det er enklere å finne maksimum av den naturlige logaritmen av rimelighetsfunksjonen enn til rimelighetsfunksjonen direkte (disse har samme maksimum), dermed

$$
\begin{aligned}
l(p;x_1,x_2,\ldots,x_n) &= \ln L(p;x_1,x_2,\ldots,x_n)\\
&= \ln\left(p^n(1-p)^{(\sum_{i=1}^{n} x_i)-n}\right)\\
&= n\ln p + (\sum_{i=1}^{n} x_i - n)\ln(1-p)
\end{aligned}
$$

Maksimum finner vi så ved å derivere $l(p;x_1,x_2,\ldots,x_n)$ med hensyn på p,

$$
\begin{aligned}
\frac{\partial l(p;x_1,x_2,\ldots,x_n)}{\partial p} &= \frac{\partial}{\partial p}\left(n\ln p + (\sum_{i=1}^{n} x_i - n)\ln(1-p)\right)\\
&= n\frac{1}{p} + (\sum_{i=1}^{n} x_i - n)\frac{1}{1-p}(-1)\\
&= \frac{n}{p} - \frac{\sum_{i=1}^{n} x_i - n}{1-p}
\end{aligned}
$$

og sette det uttrykket vi får lik null

$$
\begin{aligned}
\frac{\partial l}{\partial p} &= 0\\
\frac{n}{p} - \frac{\sum_{i=1}^{n} x_i - n}{1-p} &= 0\\
n(1-p) &= (\sum_{i=1}^{n} x_i - n)p\\
n - np &= p\sum_{i=1}^{n} x_i - np\\
p &= \frac{n}{\sum_{i=1}^{n} x_i}
\end{aligned}
$$

dvs. sannsynlighetsmaksimeringsestimatoren er

$$
\underline{\underline{\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i}}}
$$

Dersom vi kun har en observasjon X_1, vil sannsynlighetsmaksimeringsestimatoren bli $\hat{p} = \frac{1}{X_1}$. Vi vil finne ut om den er forventningsrett, dvs. finne forventningsverdien til $\frac{1}{X_1}$ når X_1 er geometrisk fordelt.

$$
\begin{aligned}
E(\hat{p}) &= E\left(\frac{1}{X_1}\right) = \sum_{x_1=1}^{\infty}\frac{1}{x_1}f(x_1;p) = \sum_{x_1=1}^{\infty}\frac{1}{x_1}p(1-p)^{x_1-1}\\
&= p\sum_{x_1=1}^{\infty}\frac{1}{x_1}\frac{(1-p)^{x_1}}{1-p} = \frac{p}{1-p}\sum_{x_1=1}^{\infty}\frac{1}{x_1}(1-p)^{x_1}\\
&= \frac{p}{1-p}[-\ln p] = -\frac{p}{1-p}\ln p
\end{aligned}
$$

Der siste overgang kommer frem ved å bruke formelen oppgitt i oppgaven, dvs.

$$
\sum_{x_1=1}^{\infty}\frac{1}{x_1}(1-p)^{x_1} = -\ln p
$$

Siden $\mathrm{E}(\hat{p}) \neq p$ så er $\hat{p}$ ikke forventningsrett.

Oppgave 3 Test nasjonen

Kommentar: Vi ser i denne oppgaven på IQ som en kontinuerlig variabel. I IQ-tester så oppgis IQ-score som heltall, og vi vil i bedømmelsen av besvarelsene ikke trekke i poeng hvis man har valgt å regne med heltallig IQ eller har lagt til en slags "kontinuitetskorreksjon", såfremt dette er begrunnet. Svarene man da kommer frem til avviker svært lite fra svarene som er oppgitt i denne løsningsskissen.

a) La X være IQ-score til tilfeldig valgt person. Vi har at X er normalfordelt med $\mathrm{E}(X) = 100$ og $\mathrm{Var}(X) = 15^2$.

Hva er sannsynligheten for at en tilfeldig valgt person skal få en IQ-score på minst 122?

$$\begin{aligned} P(X \geq 122) &= 1 - P(X < 122) = 1 - P(X \leq 122) = 1 - P(\frac{X-100}{15} \leq \frac{122-100}{15}) \\ &= 1 - \Phi(1.47) = 1 - 0.9292 = \underline{\underline{0.0708}} \end{aligned}$$

Hvis vi tester et representativt utvalg på 270 personer, hva er da forventet antall personer som får en IQ-score på minst 122?

Dette er en binomisk situasjon med $n = 270$ og $p = 0.0708$. Forventet antall personer med IQ-score på minst 122 blir $n \cdot p = 270 \cdot 0.0708 = \underline{\underline{19.2}}$. Vi forventer at rundt 19 personer får en IQ-score på minst 122.

Hva er sannsynligheten for at maksimal IQ-score i et tilfeldig utvalg av størrelse 270 vil være større enn 122?

$$\begin{aligned} P(\max_{i=1,\ldots,270} X_i > 122) &= 1 - P(\max_{i=1,\ldots,270} X_i \leq 122) \\ &= 1 - P(X_1 \leq 122 \cap X_2 \leq 122 \cap \cdots X_{270} \leq 122) \\ &= 1 - [P(X_i \leq 122)]^{270} = 1 - (0.9292)^{270} = \underline{1} \end{aligned}$$

b) Vi antar at IQ-score til en tilfeldig valgt person er normalfordelt med ukjent forventningsverdi μ, men kjent standardavvik 15.

Vi har observasjoner $X_1, X_2, ..., X_n$ der $n = 42$. En estimator for μ er $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$. Her er $\sigma = 15$ kjent, slik at

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

er standard normalfordelt.

Vi setter opp et $(1-\alpha) \cdot 100\%$ konfidensintervall:

$$\begin{aligned} P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) &= 1 - \alpha \\ P(-z_{\frac{\alpha}{2}} < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}) &= 1 - \alpha \\ P(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \end{aligned}$$

Med $\alpha = 0.05$ og innsatt tall fra de tidligere Reality-deltakerne blir intervallet $\underline{\underline{[89.46{,}98.54]}}$, da

$$\begin{aligned} \overline{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} &= 94 - 1.960 \cdot \frac{15}{\sqrt{42}} = 89.46 \\ \overline{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} &= 94 + 1.960 \cdot \frac{15}{\sqrt{42}} = 98.54 \end{aligned}$$

Etter programmet har det blitt stilt spørsmål om oppgavene i IQ-testen var for vanskelige, slik at IQ-scorene som ble oppnådd var lavere enn man kunne forvente. Hvis vi antar at IQ-score til en tilfeldig valgt person er normalfordelt med forventningsverdi 100 og standardavvik 15, hva er da sannsynligheten for at man i et tilfeldig utvalg på 42 personer oppnår en gjennomsnittlig IQ-score på mindre eller lik 94?

Gjennomsnittet av IQ-score et tilfeldig utvalg av størrelse 42 er normalfordelt med forventning $\mu = 100$ og standardavvik $\frac{\sigma}{\sqrt{42}} = \frac{15}{\sqrt{42}} = 2.31$

$$\begin{aligned} P(\overline{X} \leq 94) &= P(\frac{\overline{X} - 100}{2.31} \leq \frac{94 - 100}{2.31}) \\ &= \Phi(-2.59) = \underline{\underline{0.0048}} \end{aligned}$$

Denne sannsynligheten er lik p-verdien i en test av $H_0 : \mu = 100$ vs $H_1 : \mu < 100$ basert på tallene fra de tidligere Realitydeltakerne.