Lösningsförslag till valda uppgifter i
SANNOLIKHETSTEORI och STATISTIKTEORI med TILLÄMPNINGAR
av Blom, Enger, Englund, Grandell & Holst.

Version 28 februari 2005

Fel i lösningarna mottages tacksamt till mattsson@math.kth.se.

Notera att lösningarna på vissa ställen utnyttjar andra, mer fullständiga, tabeller än vad som normalt är tillgängliga för studenterna. Därför kan t.ex. kvantiler i normalfördelningen och $t(n)$-fördelningar i lösningarna vara bestämda med mycket god nogrannhet.

11.1 Definitionen är att en skattning θ^*_{obs} är väntevärdevärdesriktig skattning av en parameter θ om $E(\theta^*) = \theta$.

1. Nej, skattningen $\theta^*_{\text{obs}} = \theta^*(x_1, \ldots, x_n)$ kan vara vilken funktion av $x_1, \ldots, x_n$ som helst, inte nödvändigtvis ett medelvärde.

2. Nej, inte heller ett viktat medelvärde. Tag som exempel skattningen s^2 av variansen för en stokastisk variabel.

3. Ja, en tolkning av väntvärdet $E(\theta^*)$ är att det är det genomsnittliga värdet av ett stort antal försök. (Stora talens lag säger att medelvärdet konvergerar mot väntevärdet.)

4. Nej, av flera skäl. Ett — det finns ingenting i väntevärdesriktigheten som säger att det är ett väntevärde som skattas. Två — skattningen θ^*_{obs} är ett utfall av en stokastisk variabel och ger olika skattningar i olika försök.

5. Nej, skattningen θ^*_{obs} är ett utfall av en stokastisk variabel och ger olika skattningar i olika försök.

6. Ja, detta är definitionen av väntevärdesriktighet.

11.2 Låt $x_1, \ldots, x_n$ vara batteriernas uppmätta livslängder. Vi modellerar dessa som utfall av oberoende och likafördelade stokastiska variabler $X_1, \ldots, X_n$, med $\mu = E(X_i)$ och $\sigma^2 = V(X_i)$.

Väntevärdet μ skattas med

$$\mu^*_{\text{obs}} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 5.2$$

och variansen σ^2 med

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2\right) = 1.7,$$

det vill säga, standardavvikelsen σ skattas med $s = \sqrt{1.7} = 1.3038$.

Skattningen $\mu^*_{\text{obs}} = \bar{x}$ är ett utfall av stickprovsvariabeln

$$\mu^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

som har väntevärde

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu,$$

vilket visar att skattningen μ^*_{obs} är väntevärdesriktig, och varians

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{\sigma^2}{n},$$

det vill säga $D\left(\overline{X}\right) = \sigma/\sqrt{n}$. Standardavvikelsen $D\left(\overline{X}\right) = \sigma/\sqrt{n}$ skattas med $s/\sqrt{n} = 0.5831$ och kallas för medelfelet för $\bar{x}$.

11.3 Låt X beskriva uppmätt halt av ämnet. Modell:

$$X = \text{halt} + \text{mätfel} = \mu + \epsilon$$

där ϵ är en stokastisk variabel med $E(\epsilon) = 0$ och $D(\epsilon) = \sigma = 0.5$. Då är

$$E(X) = E(\mu + \epsilon) = \mu + E(\epsilon) = \mu$$

och $V(X) = V(\mu + \epsilon) = V(\epsilon) = \sigma^2$. Skattningen $\bar{x}$ av μ beskrivs av stickprovsvariabeln

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

som har

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu = \mu,$$

det vill säga $\bar{x}$ är en väntevärdesriktig skattning av μ, och

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{V(X_i)}_{=\sigma^2} = \frac{\sigma^2}{n}.$$

Om n är stor så är $\sum_{i=1}^{n} X_i$ approximativt normalfördelad, det vill säga,

$$\overline{X} \text{ är approximativt } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\begin{aligned} P\left(|\overline{X} - \mu| \leq 0.25\right) &= P\left(-0.25 \leq \overline{X} - \mu \leq 0.25\right) = P\left(\frac{-0.25}{\sigma/\sqrt{n}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.25}{\sigma/\sqrt{n}}\right) \\ &\approx \Phi\left(0.5\sqrt{n}\right) - \Phi\left(-0.5\sqrt{n}\right) = 2\Phi\left(0.5\sqrt{n}\right) - 1 = 0.99. \end{aligned}$$

Alltså är $1 - \Phi(0.5\sqrt{n}) = 0.005$ men $1 - \Phi(\lambda_{0.005}) = 0.005$ där $\lambda_{0.005} = 2.5758$, vilket ger

$$0.50\sqrt{n} = \lambda_{0.005} \quad \text{eller} \quad n = \left(\frac{\lambda_{0.005}}{0.50}\right)^2 = 26.54,$$

alltså välj $n \geq 27$.

Om man inte kan förutsätta normalfördelning ger Tjebychovs olikhet en övre gräns för vad n behöver vara.

$$0.01 = P\left(|\overline{X} - \mu| > 0.25\right) \leq \frac{V\left(\overline{X}\right)}{0.25^2} = \frac{\sigma^2/n}{0.25^2}$$

Vilket ger

$$n \leq \frac{\sigma^2}{0.25^2 \cdot 0.01} = 400,$$

det vill säga oavsett fördelning behöver inte mer än 400 mätningar göras.

Det är viktigt att se skillnaden på μ, parametern i fördelningen, $\bar{x}$, skattningen av parameterns värde, och $\overline{X}$, den stokastiska variabel som beskriver skattningen.

11.4

11.5 Låt X_1 och X_2 vara oberoende och binomialfördelade, X_i är $\text{Bin}(1, p)$, $i = 1, 2$, med observerade värden x_1 respektive x_2. Då är $E(X_i) = p$ och $V(X_i) = p(1-p)$.

Låt
$$p^*_{\text{obs}} = x_1 \quad \text{och} \quad \hat{p}_{\text{obs}} = \frac{x_1 + x_2}{2}.$$
Dessa skattningar modelleras av
$$p^* = X_1 \quad \text{och} \quad \hat{p} = \frac{X_1 + X_2}{2}.$$
Möjliga värden på X_i är $\{0,1\}$ så möjliga värden på p^* ges av $\{0,1\}$ medan för $\hat{p}$ är det $\{0,1/2,1\}$.

b) Vi har att
$$E(p^*) = E(X_1) = p \qquad E(\hat{p}) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) = \frac{p}{2} + \frac{p}{2} = p$$
så båda skattningarna p^*_{obs} och $\hat{p}_{\text{obs}}$ är väntevärdesriktiga.

c) Vidare så är, eftersom X_1 och X_2 är oberoende,
$$V(p^*) = V(X_1) = p(1-p) \qquad V(\hat{p}) = V\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{4}V(X_1) + \frac{1}{4}V(X_2) = \frac{p(1-p)}{2}.$$
Eftersom $V(\hat{p}) < V(p^*)$ för alla $p \in (0,1)$ är skattningen $\hat{p}_{\text{obs}}$ effektivare än skattningen p^*_{obs}.

11.6 Låt $x_1, \ldots, x_n$ vara utfall av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $X_1, \ldots, X_n$. Väntevärdet μ skattas dels med
$$\mu^*_{\text{obs}} = \bar{x} \qquad \text{och dels med} \qquad \hat{\mu}_{\text{obs}} = \frac{x_1 + x_2}{2}.$$

a) Båda dessa skattningar är väntevärdesriktiga eftersom
$$E(\mu^*) = E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}n\mu = \mu$$
och
$$E(\hat{\mu}) = E\left(\frac{1}{2}(X_1 + X_n)\right) = \frac{1}{2}E(X_1) + \frac{1}{2}E(X_n) = \frac{\mu}{2} + \frac{\mu}{2} = \mu.$$

b) Eftersom $X_1, \ldots, X_n$ är oberoende så är
$$V(\mu^*) = V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$
och
$$V(\hat{\mu}) = V\left(\frac{1}{2}(X_1 + X_n)\right) = \frac{1}{4}V(X_1) + \frac{1}{4}V(X_n) = \frac{\sigma^2}{2}.$$
Om $n > 2$ är $V(\mu^*) < V(\hat{\mu})$ och skattningen μ^*_{obs} är effektivare än skattningen $\hat{\mu}_{\text{obs}}$. Notera även att $V(\mu^*) \to 0$ då $n \to \infty$ medan $V(\hat{\mu})$ är konstant, så skattningen μ^*_{obs} är en konsistent skattning av μ men $\hat{\mu}$ är det inte.

11.7 Den stokastiska variabeln X är binomialfördelad, X är $\mathrm{Bin}(n,p)$ och Y är hypergeometriskt fördelad, Y är $\mathrm{Hyp}(N,n,p)$. Låt $p^* = X/n$ och $\hat{p} = Y/n$.

a) Då är
$$E(p^*) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}np = p$$
och
$$V(p^*) = V\left(\frac{1}{n}X\right) = \frac{1}{n^2}V(X) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$$

Vidare så är

$$E(\widehat{p}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{1}{n}np = p$$

och

$$\begin{aligned} V(\widehat{p}) &= V\left(\frac{1}{n}Y\right) = \frac{1}{n^2}V(Y) = \frac{1}{n^2}np(1-p)\frac{N-n}{N-1} = \frac{p(1-p)}{n}\cdot\frac{N-n}{N-1} = V(p^*)\underbrace{\frac{N-n}{N-1}}_{\leq 1} \\ &\leq V(p^*), \end{aligned}$$

med likhet endast om $n = 1$.

b) För $n > 1$ är $V(\widehat{p}) < V(p^*)$ och skattningen $\widehat{p}_{\text{obs}}$ är effektivare än skattningen p^*_{obs}.

c) Med observationer erhålls skattningarna $p^*_{\text{obs}} = \widehat{p}_{\text{obs}} = 23/100 = 0.23$ av andelen. Stickprovsvariablernas varianser skattas med hjälp av skattningarna av p. Det vill säga,

$$V(p^*) = \frac{p(1-p)}{n} \quad \text{skattas med} \quad \frac{0.23\cdot(1-0.23)}{100} = 0.001771$$

och

$$V(\widehat{p}) = \frac{p(1-p)}{n}\,\frac{N-n}{N-1} \quad \text{skattas med} \quad \frac{0.23\cdot(1-0.23)}{100}\,\frac{1000-100}{1000-1} = 0.00160.$$

Detta medför att standardavvikelserna $D(p^*)$ och $D(\widehat{p})$ skattas med 0.0421 respektive 0.0399. Dessa skattningar kallas för medelfelen för p^*_{obs} respektive $\widehat{p}_{\text{obs}}$.

11.8 Att θ^*_{obs} och $\hat{\theta}_{\text{obs}}$ är väntevärdesriktiga skattningar av en parameter θ betyder att motsvarande stickprovsvariabler uppfyller $E(\theta^*) = \theta$ och $E\left(\hat{\theta}\right) = \theta$. Med skattningen

$$\tilde{\theta}_{\text{obs}} = a\theta^*_{\text{obs}} + (1-a)\hat{\theta}_{\text{obs}}$$

så är

$$E\left(\tilde{\theta}\right) = E\left(a\theta^* + (1-a)\hat{\theta}\right) = a\underbrace{E(\theta^*)}_{=\theta} + (1-a)\underbrace{E\left(\hat{\theta}\right)}_{=\theta} = \theta$$

för alla värden a, så $\tilde{\theta}_{\text{obs}}$ är en väntevärdesriktig skattning av θ oavsett värdet på a. Vidare så är

$$\begin{aligned} V\left(\tilde{\theta}\right) &= V\left(a\theta^* + (1-a)\hat{\theta}\right) = \{\text{oberoende}\} = a^2\underbrace{V(\theta^*)}_{=\sigma_1^2} + (1-a)^2\underbrace{V\left(\hat{\theta}\right)}_{=\sigma_2^2} \\ &= a^2\sigma_1^2 + (1-a)^2\sigma_2^2 = g(a), \end{aligned}$$

en funktion av a. Den effektivaste skattningen erhålls om a väljes så att variansen $g(a)$ minimeras. Detta a bestäms som lösningen till

$$0 = \frac{d}{da}g(a) = \sigma_1^2\cdot 2a + \sigma_2^2\cdot 2(1-a)(-1) = 2\left[a(\sigma_1^2+\sigma_2^2) - \sigma_2^2\right],$$

där lösningen är

$$a = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}.$$

Teckenstudium av andraderivatan visar att det är ett minimum som vi fått fram. Alltså, den effektivaste skattningen fås om

$$\tilde{\theta}_{\text{obs}} = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta^*_{\text{obs}} + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\hat{\theta}_{\text{obs}}.$$

11.9 Låt $x_1, \ldots, x_n$ vara utfallen av de stokastiska variablerna $X_1, \ldots, X_n$ som antas vara oberoende och $\mathrm{N}(\mu, \sigma)$-fördelade.

a) Om μ är känd så kan variansen σ^2 skattas med $(\hat{\sigma}^2)_{\text{obs}}$ där

$$(\hat{\sigma}^2)_{\text{obs}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Denna skattning är väntevärdesriktig eftersom

$$E\left(\hat{\sigma}^2\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right) = \frac{1}{n}\sum_{i=1}^{n}\underbrace{E\left((X_i - \mu)^2\right)}_{=\sigma^2} = \frac{1}{n}n\sigma^2 = \sigma^2.$$

Med de $n = 4$ observationerna $x_1, \ldots, x_4$ och $\mu = 1457.0$ så är

$$(\hat{\sigma}^2)_{\text{obs}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - 1457.0)^2 = 0.8175.$$

b) Om μ är okänd måste väntevärdet skattas. En väntevärdesriktig skattning av σ^2 är s^2 där

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

vilket, då μ skattas med $\overline{x} = 1457.4$, ger skattningen $s^2 = 0.84917$.

Att s^2 är väntevärdesriktig ges av följande uträkning.

$$\begin{aligned}
E\left(S^2\right) &= E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}X_i^2 - 2\overline{X}X_i + \overline{X}^2\right)\\
&= E\left(\frac{1}{n-1}\left(\left(\sum_{i=1}^{n}X_i^2\right) - 2\overline{X}\left(\sum_{i=1}^{n}X_i\right) + n\overline{X}^2\right)\right)\\
&= E\left(\frac{1}{n-1}\left(\left(\sum_{i=1}^{n}X_i^2\right) - 2\overline{X}n\overline{X} + n\overline{X}^2\right)\right) = E\left(\frac{1}{n-1}\left(\left(\sum_{i=1}^{n}X_i^2\right) - n\overline{X}^2\right)\right)\\
&= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n}E\left(X_i^2\right)\right) - nE\left(\overline{X}^2\right)\right) = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n}\mu^2 + \sigma^2\right) - n(\mu^2 + \frac{\sigma^2}{n})\right)\\
&= \frac{1}{n-1}\left(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2\right) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2.
\end{aligned}$$

c) Nej, av uträkningarna ovan framgår att endast $E(X_i)$ och $V(X_i)$ har utnyttjats.

11.10 Den stokastiska variabeln X har sannolikhetsfunktion $p_X(k) = (1-\theta)^{k-1}\theta$ för $k = 1, 2, 3, \ldots$ där $0 < \theta < 1$. Låt $x_1, \ldots, x_n$ vara ett stickprov av X.

a) Då observationerna är utfall av oberoende stokastiska variabler blir Likelihoodfunktionen

$$\begin{aligned}
L(\theta) &= p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1)\cdots p_{X_n}(x_n) = (1-\theta)^{x_1-1}\theta\cdots(1-\theta)^{x_n-1}\theta\\
&= \theta^n(1-\theta)^{\sum_{i=1}^{n}x_i - n}.
\end{aligned}$$

b) Vi söker det värde på θ som maximerar $L(\theta)$. Det är samma värde som maximerar den logaritmerade likelihoodfunktionen

$$\ln(L(\theta)) = n\ln(\theta) + \ln(1-\theta)\left(\sum_{i=1}^{n}x_i - n\right).$$

Detta maximum bestäms som lösningen till

$$
\begin{aligned}
0 &= \frac{d}{d\theta}\ln(L(\theta)) = \frac{d}{d\theta}\left[n\ln(\theta) + \ln(1-\theta)\left(\sum_{i=1}^{n} x_i - n\right)\right] = \frac{n}{\theta} - \frac{1}{1-\theta}\left(\underbrace{\sum_{i=1}^{n} x_i}_{=n\overline{x}} - n\right) \\
&= \frac{n}{\theta(1-\theta)}\left(1-\theta\overline{x}\right).
\end{aligned}
$$

Alltså är $\theta = 1/\overline{x}$ det värde som ger maximum. (Kontroll av andraderivatan ger att det verkligen är ett maximum som erhållits.) Alltså, ML-skattningen av θ är

$$\theta^*_{\text{obs}} = \frac{1}{\overline{x}}.$$

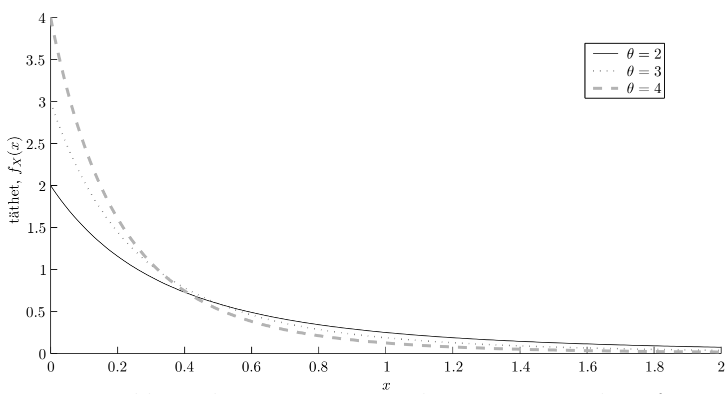
Med $(x_1, \ldots, x_n) = (4,\ 5,\ 4,\ 6,\ 4,\ 1)$ är $\overline{x} = 4$ och $\theta^*_{\text{obs}} = 1/\overline{x} = 1/4$.

Notera att X är ffg(θ)-fördelad och har väntevärde $E\,(X) = 1/\theta$. Om detta väntevärde skattas med $\overline{x}$ fås $\theta^*_{\text{obs}} = 1/\overline{x}$ som skattning av θ.

11.11 Den stokastiska variabeln X har täthetsfunktion

$$f_X(x) = \theta(1+x)^{-(\theta+1)}, \quad x \geq 0,$$

där möjliga värden på parametern θ är 2, 3 eller 4. De tre möjliga tätheterna för X visas i figuren nedan.



De tre olika täthetern a motsvarande parametervärdena $\theta = 2, 3, 4$.

Med stickprovet x_1, x_2 blir likelihoodfunktionen

$$L(\theta) = p_{X_1,X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2) = \theta(1+x_1)^{-(\theta+1)}\theta(1+x_2)^{-(\theta+1)} = \theta^2\left[(1+x_1)(1+x_2)\right]^{-(\theta+1)}.$$

Med observationerna $(x_1, x_2) = (0.2,\ 0.8)$ blir $L(\theta) = \theta^2 \cdot (2.16)^{-(\theta+1)}$ och

Parameter	$L(\theta)$
$\theta = 2$	0.39692
$\theta = 3$	0.41345
$\theta = 4$	0.34029

så valet $\theta = 3$ ger maximum. Alltså, ML-skattningen av θ är $\theta^*_{\text{obs}} = 3$.

11.12 Låt $x_1, \ldots, x_n$ vara antalet telefonsamtal under olika dagar. Vi ansätter modellen att $x_1, \ldots, x_n$ är utfall av oberoende Poisson(μ)-fördelade stokastiska variabler $X_1, \ldots, X_n$.

Maximum likelihoodskattningen av μ är det värde på μ som maximerar

$$L(\mu) = p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1)\cdots p_{X_n}(x_n) = \frac{\mu^{x_1}}{x_1!}e^{-\mu}\cdots\frac{\mu^{x_n}}{x_n!}e^{-\mu} = \frac{\mu^{\sum_{i=1}^{n} x_i}}{x_1!\cdots x_n!}e^{-n\mu}.$$

Det är samma μ som maximerar

$$\ln(L(\mu)) \quad = \quad \ln\left(\frac{\mu^{\sum_{i=1}^{n} x_i}}{x_1!\cdots x_n!}\mathrm{e}^{-n\mu}\right) = \ln(\mu)\sum_{i=1}^{n} x_i - n\mu - \sum_{i=1}^{n}\ln(x_i!).$$

Detta maximum bestäms som det μ som löser

$$0 = \frac{d}{d\mu}\ln(L(\mu)) = \frac{d}{d\mu}\left[\ln(\mu)\sum_{i=1}^{n} x_i - n\mu - \sum_{i=1}^{n}\ln(x_i!)\right] = \frac{1}{\mu}\sum_{i=1}^{n} x_i - n$$

vilket ger $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$. Kontroll av andraderivatan ger att detta är ett maximum. Skattningen $\mu^*_{\text{obs}} = \bar{x}$ av μ beskrivs av stickprovsvariabeln $\mu^* = \overline{X}$ som har

$$E\left(\overline{X}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu,$$

vilket ger att $\bar{x}$ är en väntevärdesriktig skattning av μ, och

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{V\left(X_i\right)}_{=\mu} = \frac{\mu}{n},$$

det vill säga $D\left(\overline{X}\right) = \sqrt{\mu/n}$.

Med observationer $x_1, \ldots, x_8$:

115 82 108 106 118 87 99 92

fås skattningen $\bar{x} = 100.88$. En skattning av $D\left(\overline{X}\right) = \sqrt{\mu/n}$ ges av

$$\sqrt{\frac{\mu^*_{\text{obs}}}{n}} = \sqrt{\frac{100.88}{8}} = 3.551$$

och kallas för skattningens medelfel.

11.13 Låt $x_1, \ldots, x_n$ vara de uppmätta tiderna mellan fel hos den komplicerade tekniska utrustningen. Med modellen att $x_1, \ldots, x_n$ är utfall av oberoende exponential(λ)-fördelade stokastiska variabler $X_1, \ldots, X_n$ är likelihoodfunktionen

$$L(\lambda) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n) = \lambda\mathrm{e}^{-\lambda x_1}\cdots\lambda\mathrm{e}^{-\lambda x_n} = \lambda^n\mathrm{e}^{-\lambda\sum_{i=1}^{n} x_i}.$$

Det värde på λ som maximerar $L(\lambda)$ är samma som maximerar

$$\ln(L(\lambda)) = \ln\left(\lambda^n\mathrm{e}^{-\lambda\sum_{i=1}^{n} x_i}\right) = n\ln(\lambda) - \lambda\underbrace{\sum_{i=1}^{n} x_i}_{=n\overline{x}} = n\ln(\lambda) - \lambda n\bar{x}.$$

Maximum bestäms som det λ som löser

$$0 = \frac{d}{d\lambda}\ln(L(\lambda)) = \frac{d}{d\lambda}\left[n\ln(\lambda) - \lambda n\bar{x}\right] = \frac{n}{\lambda} - n\bar{x} = n\left(\frac{1}{\lambda} - \bar{x}\right),$$

det vill säga $\lambda = 1/\bar{x}$. Kontroll av andraderivatan ger att detta är ett maximum. Skattningen $\lambda^*_{\text{obs}} = 1/\bar{x}$ av λ beskrivs av $\lambda^* = 1/\overline{X}$. Den har för $t \geq 0$ fördelningsfunktion

$$F_{\lambda^*}(t) = P\left(\lambda^* \leq t\right) = P\left(\frac{1}{\overline{X}} \leq t\right) = P\left(\sum_{i=1}^{n} X_i \geq \frac{n}{t}\right) = 1 - F_{\sum X_i}(n/t)$$

det vill säga

$$f_{\lambda^*}(t) = \frac{d}{dt}F_{\lambda^*}(t) = \frac{n}{t^2}f_{\sum X_i}(n/t).$$

En summa av n stycken oberoende exp(λ)-fördelade stokastiska variabler har täthet

$$f_{\sum X_i}(x) = \frac{\lambda^n x^{n-1}}{\Gamma(n)}\mathrm{e}^{-\lambda x}, \quad x \geq 0,$$

så

$$f_{\lambda^*}(t) = \frac{n}{t^2}\frac{\lambda^n (n/t)^{n-1}}{\Gamma(n)}\mathrm{e}^{-\lambda n/t}, \quad t \geq 0.$$

Alltså är

$$\begin{aligned} E(\lambda^*) &= \int_{-\infty}^{\infty} t f_{\lambda^*}(t)\,dt = \int_0^{\infty}\frac{\lambda^n (n/t)^n}{\Gamma(n)}\mathrm{e}^{-\lambda n/t}\,dt = \{u = n/t\} = n\int_0^{\infty}\frac{\lambda^n u^{n-2}}{\Gamma(n)}\mathrm{e}^{-\lambda u}\,du \\ &= \{\Gamma(n) = (n-1)\Gamma(n-1)\} = \frac{n}{n-1}\lambda\int_0^{\infty}\underbrace{\frac{\lambda^{n-1}u^{n-2}}{\Gamma(n-1)}\mathrm{e}^{-\lambda u}}_{=f_Y(u)}\,du = \frac{n}{n-1}\lambda \neq \lambda \end{aligned}$$

vilket visar att skattningen inte är väntevärdesriktig, men att $(n-1)\lambda^*_{\text{obs}}/n$ är det. Tätheten $f_Y(u)$ var tätheten för en summa av $n-1$ stycken oberoende exp(λ)-fördelade stokastiska variabler och den integreras till 1 över intervallet $[0, \infty)$.

11.14 Låt $x_1, \ldots, x_n$ vara utfall av $X_1, \ldots, X_n$ där de stokastiska variablerna är oberoende och har täthetsfunktion

$$f_X(x) = \theta x^{\theta-1}, \quad 0 < x < 1.$$

Likelihoodfunktionen blir

$$L(\theta) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \{\text{ober.}\} = f_{X_1}(x_1)\cdots f_{X_n}(x_n) = \theta x_1^{\theta-1}\cdots\theta x_n^{\theta-1} = \theta^n\left(x_1\cdots x_n\right)^{\theta-1}.$$

Det är samma θ som maximerar

$$\ln(L(\theta)) = \ln\left(\theta^n\left(x_1\cdots x_n\right)^{\theta-1}\right) = n\ln(\theta) + (\theta-1)\sum_{i=1}^{n}\ln(x_i).$$

Maximum bestäms som det θ som löser

$$0 = \frac{d}{d\theta}\ln(L(\theta)) = \frac{d}{d\theta}\left[n\ln(\theta) + (\theta-1)\sum_{i=1}^{n}\ln(x_i)\right] = \frac{n}{\theta} + \sum_{i=1}^{n}\ln(x_i)$$

det vill säga $\theta = -n/\sum_{i=1}^{n}\ln(x_i)$. Kontroll av andraderivatan ger att detta är ett maximum. (Notera att $\ln(x_i) < 0$ för $0 < x_i < 1$.) Alltså, maximum-likelihoodmetodens skattning av θ är $\theta^*_{\text{obs}} = -n/\sum_{i=1}^{n}\ln(x_i)$.

11.15 Låt $x_1, \ldots, x_n$ vara utfall av $X_1, \ldots, X_n$ där de stokastiska variablerna är oberoende och Rayleigh(a)-fördelade, det vill säga har täthetsfunktion

$$f_X(x) = (x/a)\mathrm{e}^{-x^2/2a} \quad x \geq 0.$$

Likelihoodfunktionen blir

$$\begin{aligned} L(a) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \{\text{ober.}\} = f_{X_1}(x_1)\cdots f_{X_n}(x_n) = \frac{x_1}{a}\mathrm{e}^{-x_1^2/2a}\cdots\frac{x_n}{a}\mathrm{e}^{-x_n^2/2a} \\ &= \frac{x_1\cdots x_n}{a^n}\mathrm{e}^{-\frac{1}{2a}\sum_{i=1}^{n}x_i^2}. \end{aligned}$$

Vi söker det värde på a som maximerar $L(a)$, och det är samma a som maximerar

$$\ln(L(a)) = \ln\left(\frac{x_1\cdots x_n}{a^n}\mathrm{e}^{-\frac{1}{2a}\sum_{i=1}^{n}x_i^2}\right) = -n\ln(a) + \sum_{i=1}^{n}\ln(x_i) - \frac{1}{2a}\sum_{i=1}^{n}x_i^2.$$

Detta maximum bestäms som det a som löser

$$0 = \frac{d}{da}\ln(L(a)) = \frac{d}{da}\left[-n\ln(a) + \sum_{i=1}^{n}\ln(x_i) - \frac{1}{2a}\sum_{i=1}^{n}x_i^2\right] = -\frac{n}{a} + \frac{1}{2a^2}\sum_{i=1}^{n}x_i^2 = \frac{-1}{a}\left[n - \frac{1}{2a}\sum_{i=1}^{n}x_i^2\right]$$

vilket medför att $a = \frac{1}{2n}\sum_{i=1}^{n}x_i^2$. Kontroll av andraderivatan ger att detta är ett maximum. Således, maximum-likelihoodskattningen av a är $a^*_{\text{obs}} = \frac{1}{2n}\sum_{i=1}^{n}x_i^2$.

11.16 Låt $X_1,\ldots,X_n$ vara oberoende och likformigt fördelade på intervallet $[-\theta,\theta]$, $\theta > 0$, med observerade värden $x_1,\ldots,x_n$, det vill säga de stokastiska variablerna har täthetsfunktion

$$f_X(x) = \frac{1}{2\theta} \qquad \text{för } -\theta \le x \le \theta.$$

Maximum likelihoodskattningen är det värde på θ som maximerar

$$L(\theta) = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \{\text{oberoende}\} = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n}\frac{1}{2\theta} = (2\theta)^{-n}$$

för $-\theta \le x_1,\ldots,x_n \le \theta$. Eftersom $L(\theta)$ är avtagande i θ skall man göra θ så liten som möjligt. Kravet $-\theta \le x_1,\ldots,x_n \le \theta$, det vill säga $|x_1|,\ldots,|x_n| \le \theta$, gör att det minsta möjliga värdet på θ ges av $\theta = \max(|x_1|,\ldots,|x_n|)$. Alltså, maximum-likelihoodskattningen av θ är $\theta^*_{\text{obs}} = \max(|x_1|,\ldots,|x_n|)$.

Eftersom denna skattning omöjligen kan överskatta θ kommer den ha ett systematiskt fel, $E(\theta^*) < \theta$.

11.17 Den stokastiska variabeln X har täthetsfunktion

$$f_X(x) = \theta(1+x)^{-(\theta+1)}, \quad x \ge 0,$$

där möjliga värden på parametern θ är 2, 3 eller 4. Nu är

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^{\infty} x\cdot\frac{\theta}{(1+x)^{\theta+1}}\,dx = \left[\frac{-x}{(1+x)^{\theta}}\right]_0^{\infty} + \int_0^{\infty}\frac{1}{(1+x)^{\theta}}\,dx \\ &= \left[\frac{-1/(\theta-1)}{(1+x)^{\theta-1}}\right]_0^{\infty} = \frac{1}{\theta-1}. \end{aligned}$$

Med $n = 2$ observationer $(x_1,x_2) = (0.2,\ 0.8)$ är minsta-kvadratskattningen av θ är det värde som minimerar

$$Q(\theta) = \sum_{i=1}^{n}(x_i - E(X_i))^2 = \left(0.2 - \frac{1}{\theta-1}\right)^2 + \left(0.8 - \frac{1}{\theta-1}\right)^2$$

För de olika värdena på θ erhålles

Parameter	$Q(\theta)$
$\theta = 2$	0.68
$\theta = 3$	0.18
$\theta = 4$	0.24

så valet $\theta = 3$ ger minimum. Alltså, MK-skattningen av θ är $\theta^*_{\text{obs}} = 3$.

11.18 Låt X_1,X_2,X_3 beskriva mätningarna på vinkeln AOC och X_4,X_5 på vinkeln AOB. Modell: $X_1,\ldots,X_n$ är oberoende och

$$E(X_1) = E(X_2) = E(X_3) = \theta_1 + \theta_2 \qquad E(X_4) = E(X_5) = \theta_1$$

och $V(X_i) = \sigma^2$. Minsta kvadratmetodens skattning av (θ_1,θ_2) är det värde på (θ_1,θ_2) som minimerar

$$\begin{aligned} Q(\theta_1,\theta_2) = \sum_{i=1}^{n}(x_i - E(X_i))^2 &= (x_1-(\theta_1+\theta_2))^2 + (x_2-(\theta_1+\theta_2))^2 + (x_3-(\theta_1+\theta_2))^2 \\ &\quad + (x_4-\theta_1)^2 + (x_5-\theta_1)^2. \end{aligned}$$

Derivering med avseende på θ_1 och θ_2 ger

$$\begin{aligned}
\frac{\partial}{\partial\theta_1}Q(\theta_1,\theta_2) &= -2(x_1-(\theta_1+\theta_2))-2(x_2-(\theta_1+\theta_2))-2(x_3-(\theta_1+\theta_2)) \\
&\quad -2(x_4-\theta_1)-2(x_5-\theta_1) \\
&= -2\left[\sum_{i=1}^{5}x_i-5\theta_1-3\theta_2\right] \\
\frac{\partial}{\partial\theta_2}Q(\theta_1,\theta_2) &= -2(x_1-(\theta_1+\theta_2))-2(x_2-(\theta_1+\theta_2))-2(x_3-(\theta_1+\theta_2)) \\
&= -2\left[x_1+x_2+x_3-3\theta_1-3\theta_2\right]
\end{aligned}$$

Sätts derivatorna till 0 fås ekvationssystemet

$$\begin{cases} 5\theta_1+3\theta_2 = x_1+x_2+x_3+x_4+x_5 \\ 3\theta_1+3\theta_2 = x_1+x_2+x_3 \end{cases}$$

med lösningen

$$(\theta_1^*)_{\text{obs}} = \frac{x_4+x_5}{2} \qquad (\theta_2^*)_{\text{obs}} = \frac{x_1+x_2+x_3}{3}-(\theta_1^*)_{\text{obs}}.$$

b) Skattningen $(\theta_1^*)_{\text{obs}}$ är väntevärdesriktig ty

$$E\left(\theta_1^*\right) = E\left(\frac{1}{2}X_4+\frac{1}{2}X_5\right) = \frac{1}{2}E\left(X_4\right)+\frac{1}{2}E\left(X_5\right) = \frac{1}{2}\theta_1+\frac{1}{2}\theta_1=\theta_1.$$

Vidare så är

$$\begin{aligned}
E\left(\theta_2^*\right) &= E\left(\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3-\theta_1^*\right) = \frac{1}{3}E\left(X_1\right)+\frac{1}{3}E\left(X_2\right)+\frac{1}{3}E\left(X_3\right)-E\left(\theta_1^*\right) \\
&= \frac{1}{3}(\theta_1+\theta_2)+\frac{1}{3}(\theta_1+\theta_2)+\frac{1}{3}(\theta_1+\theta_2)-\theta_1=\theta_2
\end{aligned}$$

så även $(\theta_2^*)_{\text{obs}}$ är väntevärdesriktig.

c) Skattningarna $(\theta_1^*)_{\text{obs}}$ och $(\theta_2^*)_{\text{obs}}$ är utfall av stokastiska variabler med varians

$$V\left(\theta_1^*\right) = V\left(\frac{1}{2}X_4+\frac{1}{2}X_5\right) = \{\text{oberoende}\} = \frac{1}{2^2}\sigma^2+\frac{1}{2^2}\sigma^2 = \frac{\sigma^2}{2}.$$

respektive

$$\begin{aligned}
V\left(\theta_2^*\right) &= V\left(\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3-\theta_1^*\right) = \{\text{ober.}\} = \frac{1}{3^2}\sigma^2+\frac{1}{3^2}\sigma^2+\frac{1}{3^2}\sigma^2+(-1)^2V\left(\theta_1^*\right) = \frac{\sigma^2}{3}+\frac{\sigma^2}{2} \\
&= \frac{5}{6}\sigma^2.
\end{aligned}$$

11.19 Låt x_1, x_2, x_3 vara mätningar på vinkeln vid B och $x_4, \ldots, x_7$ de vid C. Minsta-kvadratmetodens skattning av θ är det värde på θ som minimerar

$$\begin{aligned}
Q(\theta) &= \sum_{i=1}^{7}(x_i-E\left(X_i\right))^2 = \sum_{i=1}^{3}(x_i-\theta)^2+\sum_{i=4}^{7}(x_i-(90-\theta))^2 \\
&= (x_1-\theta)^2+(x_2-\theta)^2+(x_3-\theta)^2 \\
&\quad +(x_4-(90-\theta))^2+(x_5-(90-\theta))^2+(x_6-(90-\theta))^2+(x_7-(90-\theta))^2
\end{aligned}$$

Funktionens maximipunkt hittas genom att derivatan sätts till noll.

$$\begin{aligned}
\frac{d}{d\theta}Q(\theta) &= -2\sum_{i=1}^{3}(x_i-\theta)-2\sum_{i=4}^{7}(90-x_i-\theta) \\
&= -2(x_1+x_2+x_3+(90-x_4)+\cdots+(90-x_7)-7\theta)=0.
\end{aligned}$$

Detta ger

$$\theta^*_{\text{obs}} = \frac{x_1 + x_2 + x_3 + (90 - x_4) + \cdots + (90 - x_7)}{7},$$

eller, med siffror, $\theta^*_{\text{obs}} = 61.17$. Minsta-kvadratskattningen av vinkeln vid C blir då $90 - 61.17 = 28.83$.

Skattningen beskrivs av stickprovsvariabeln

$$\theta^* = \frac{X_1 + X_2 + X_3 + (90 - X_4) + (90 - X_5) + (90 - X_6) + (90 - X_7)}{7}.$$

Den har väntevärde

$$\begin{aligned} E(\theta^*) &= E\left(\frac{X_1 + X_2 + X_3 + (90 - X_4) + (90 - X_5) + (90 - X_6) + (90 - X_7)}{7}\right) \\ &= \frac{1}{7}\big(\underbrace{E(X_1)}_{=\theta} + \cdots + \underbrace{E(X_3)}_{=\theta} + (90 - \underbrace{E(X_4)}_{=90-\theta}) + \cdots + (90 - \underbrace{E(X_7)}_{90-\theta})\big) = \frac{1}{7} \cdot 7\theta = \theta, \end{aligned}$$

vilket innebär att skattningen θ^*_{obs} är väntevärdesriktig. Vidare,

$$\begin{aligned} V(\theta^*) &= V\left(\frac{X_1 + X_2 + X_3 + (90 - X_4) + (90 - X_5) + (90 - X_6) + (90 - X_7)}{7}\right) \\ &= \frac{1}{7^2}\big(\underbrace{V(X_1)}_{=\sigma^2} + \cdots + \underbrace{V(X_3)}_{=\sigma^2} + (-1)^2 \underbrace{V(X_4)}_{=\sigma^2} + \cdots + (-1)^2 \underbrace{V(X_7)}_{=\sigma^2}\big) = \frac{\sigma^2}{7} \end{aligned}$$

så $D(\theta^*) = \sigma/\sqrt{7} = 0.1/\sqrt{7} = 0.0378$. Skattningen $90 - \theta^*_{\text{obs}}$ är ett utfall från en stokastisk variabel med samma standardavvikelse.

11.20 Låt $x_1, \ldots, x_n$ vara mätningarna av våglängden θ. Dessa mätvärden är utfall av de oberoende stokastiska variablerna $X_1, \ldots, X_n$ där $E(X_i) = \theta$. Minsta-kvadratskattningen av θ är det värde på θ som minimerar

$$Q(\theta) = \sum_{i=1}^{n} \frac{1}{V(X_i)}(x_i - E(X_i))^2.$$

Med mätningarna

x_i	79.1	80.0	81.3	81.9	81.7
$100D(X_i)$	2	1	2	3	1

fås

$$10000\, Q(\theta) = \frac{1}{2^2}(x_1 - \theta)^2 + \frac{1}{1^2}(x_2 - \theta)^2 + \frac{1}{2^2}(x_3 - \theta)^2 + \frac{1}{3^2}(x_4 - \theta)^2 + \frac{1}{1^2}(x_5 - \theta)^2$$

Lösning av

$$0 = \frac{d}{d\theta}[10000Q(\theta)] = 2\left[\frac{1}{4}x_1 + x_2 + \frac{1}{4}x_3 + \frac{1}{9}x_4 + x_5 - \frac{47}{18}\theta\right] = 2\left[210.9 - \frac{47}{18}\theta\right]$$

ger skattningen $\theta^*_{\text{obs}} = 210.9 \cdot 18/47 = 80.77$ Ångström.

11.21

11.22

11.23 Låt $x = 16$ vara ett utfall av en binomialfördelad stokastisk variabel X, X är $\text{Bin}(n, p)$ där $n = 25$.

ML-skattning: Den logaritmerade likelihoodfunktionen är

$$\ln(L(p)) = \ln\left(\binom{n}{x}\right) + x\ln(p) + (n - x)\ln(1 - p)$$

så lösning av

$$0 = \frac{d}{dp}\ln(L(p)) = \frac{x}{p} - \frac{n - x}{1 - p} = \frac{1}{p(1 - p)}[x - np]$$

ger, då $p(1 - p) \neq 0$, att $p = x/n$. Alltså, maximum-likelihoodskattningen är $p^*_{\text{obs}} = x/n$.

MK-skattning: Väntevärdet $E(X) = np$ så minsta-kvadratskattningen av p är det värde på p som minimerar

$$Q(p) = (x - E(X))^2 = (x - np)^2.$$

Eftersom $(x - np)^2 \geq 0$ med likhet om $x = np$ fås skattningen $p^*_{\text{obs}} = x/n$.

Med värden fås skattningen $p^*_{\text{obs}} = x/n = 16/25 = 0.64$

b) Skattningen $p^*_{\text{obs}} = x/n$ är ett utfall av en stokastisk variabel med varians

$$V(p^*) = V\left(\frac{1}{n}X\right) = \frac{1}{n^2}\underbrace{V(X)}_{=np(1-p)} = \frac{p(1-p)}{n}$$

så standardavvikelsen är $D(p^*) = \sqrt{p(1-p)/n}$.

c) En skattning av $D(p^*) = \sqrt{p(1-p)/n}$ är

$$\sqrt{\frac{p^*_{\text{obs}}(1-p^*_{\text{obs}})}{n}} = \sqrt{\frac{0.64(1-0.64)}{25}} = 0.096$$

och kallas medelfelet för skattningen p^*_{obs}.

11.24 Låt x_1 och x_2 vara antalet färgblinda som forskare A respektive B har funnit. Vi ansätter modellen att x_1 och x_2 är utfall av oberoende stokastiska variabler X_1 och X_2 där X_1 är $\text{Bin}(n_1, p) = \text{Bin}(1000, p)$ och X_2 är $\text{Bin}(n_2, p) = \text{Bin}(2000, p)$.

Likelihoodfunktionen är

$$L(p) = p_{X_1,X_2}(x_1, x_2) = \{\text{oberoende}\} = p_{X_1}(x_1)p_{X_2}(x_2) = \binom{n_1}{x_1}p^{x_1}(1-p)^{n_1-x_1}\binom{n_2}{x_2}p^{x_2}(1-p)^{n_2-x_2}$$

så den logaritmerade likelihoodfunktionen är

$$\ln(L(p)) = \ln\left(\binom{n_1}{x_1}\binom{n_2}{x_2}\right) + (x_1 + x_2)\ln(p) + (n_1 + n_2 - x_1 - x_2)\ln(1-p).$$

Lösning av

$$0 = \frac{d}{dp}\ln(L(p)) = \frac{x_1 + x_2}{p} - \frac{(n_1 + n_2) - (x_1 + x_2)}{1-p} = \frac{1}{p(1-p)}\left[(x_1 + x_2) - (n_1 + n_2)p\right]$$

ger skattningen $p^*_{\text{obs}} = (x_1 + x_2)/(n_1 + n_2) = 201/3000 = 0.067$. Notera att detta kan ses som att man slår samman de två observationsserierna till en och betraktar den relativa frekvensen av färgblinda i den sammanslagna serien.

11.25 Låt $x_1, \ldots, x_n$ vara antalet fartyg som passerat Helsingborg under tidsperioder av längder $t_1, \ldots, t_n$. Vi ansätter modellen att $x_1, \ldots, x_n$ är utfall av oberoende Poissonfördelade stokastiska variabler $X_1, \ldots, X_n$ där $E(X_i) = \lambda t_i$.

Likelihoodfunktionen blir

$$L(\lambda) = p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \{\text{oberoende}\} = p_{X_1}(x_1)\cdots p_{X_n}(x_n) = \frac{(\lambda t_1)^{x_1}}{x_1!}\mathrm{e}^{-\lambda t_1}\cdots\frac{(\lambda t_n)^{x_n}}{x_n!}\mathrm{e}^{-\lambda t_n}.$$

Det λ som maximerar $L(\lambda)$ är samma λ som maximerar

$$\ln(L(\lambda)) = \sum_{i=1}^{n}\left(x_i\ln(\lambda) + x_i\ln(t_i) - \ln(x_i!) - \lambda t_i\right).$$

Detta maximum bestäms som det λ som löser

$$0 = \frac{d}{d\lambda}\ln(L(\lambda)) = \sum_{i=1}^{n}\left(\frac{x_i}{\lambda} - t_i\right) = \frac{1}{\lambda}\left(\sum_{i=1}^{n}x_i\right) - \left(\sum_{i=1}^{n}t_i\right)$$

vilket ger skattningen

$$\lambda^*_{\text{obs}} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n t_i}.$$

Med de $n = 3$ observationerna

Observationstid, t_i	30	30	40
Antal fartyg x_i	10	12	18

fås skattningen $\lambda^*_{\text{obs}} = 40/100 = 0.40$ fartyg per minut.

b) Skattningen λ^*_{obs} är ett utfall av en stokastisk variabel med varians

$$V(\lambda^*) = V\left(\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n t_i}\right) = \{\text{oberoende}\} = \frac{1}{\left(\sum_{i=1}^n t_i\right)^2}\sum_{i=1}^n \underbrace{V(X_i)}_{=\lambda t_i} = \frac{\lambda}{\sum_{i=1}^n t_i},$$

det vill säga med standardavvikelse

$$D(\lambda^*) = \sqrt{\frac{\lambda}{\sum_{i=1}^n t_i}} = \frac{\sqrt{\lambda}}{10}.$$

11.26 Låt $x = 32$ vara antalet bostadssökande av $n = 50$ undersökta som redan har en bostad. Ansätt modellen att x är ett utfall av en hypergeometriskt fördelad stokastisk variabel X. Låt N beteckna antalet personer i bostadskö och p andelen av dessa som redan har en bostad.

Väntevärdet $E(X) = np$ så minsta-kvadratskattningen av p är det värde på p som minimerar

$$Q(p) = (x - E(X))^2 = (x - np)^2.$$

Eftersom $(x - np)^2 \geq 0$ med likhet om $x = np$ fås skattningen $p^*_{\text{obs}} = x/n$.

a) Med värden fås skattningen $p^*_{\text{obs}} = x/n = 32/50 = 0.64$

b) Eftersom $V(X) = np(1-p)(N-n)/(N-1)$ är skattningen p^*_{obs} är ett utfall av en stokastisk variabel med varians

$$V(p^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{p(1-p)}{n}\,\frac{N-n}{N-1}.$$

Standardavvikelsen $D(p^*) = \sqrt{\frac{p(1-p)}{n}\frac{N-n}{N-1}}$ skattas då $N = 100$ med

$$\sqrt{\frac{p^*_{\text{obs}}(1-p^*_{\text{obs}})}{n}\,\frac{N-n}{N-1}} = \sqrt{\frac{0.64(1-0.64)}{50}\,\frac{100-50}{100-1}} = 0.048242$$

c) och då $N = 1000$ med

$$\sqrt{\frac{p^*_{\text{obs}}(1-p^*_{\text{obs}})}{n}\,\frac{N-n}{N-1}} = \sqrt{\frac{0.64(1-0.64)}{50}\,\frac{1000-50}{1000-1}} = 0.066197.$$

11.27

11.28

11.29 Låt $x_1, \ldots, x_n$ vara resultaten av mätningarna av kvadratens sida. Dessa antas vara utfall av oberoende $N(\sqrt{\theta}, \sigma)$-fördelade stokastiska variabler där σ är känd och θ är kvadratens (okända) area.

Maximum-likelihoodskattningen av θ är det θ som maximerar

$$\begin{aligned} L(\theta) &= f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \{\text{oberoende}\} = f_{X_1}(x_1)\cdots f_{X_n}(x_n) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x_1-\sqrt{\theta})^2/2\sigma^2}\cdots\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x_n-\sqrt{\theta})^2/2\sigma^2}. \end{aligned}$$

Det är samma θ som maximerar

$$\ln(L(\theta)) = \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x_i-\sqrt{\theta})^2/2\sigma^2}\right) = -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\sqrt{\theta})^2$$

och detta maximum bestäms som lösningen till

$$0 = \frac{d}{d\theta}\ln(L(\theta)) = \frac{1}{2\sqrt{\theta}\sigma^2}\sum_{i=1}^{n}(x_i-\sqrt{\theta}) = \frac{1}{2\sqrt{\theta}\sigma^2}\left(\underbrace{\sum_{i=1}^{n} x_i}_{=n\overline{x}} - n\sqrt{\theta}\right) = \frac{n}{2\sqrt{\theta}\sigma^2}\left(\overline{x}-\sqrt{\theta}\right),$$

vilket ger $\sqrt{\theta} = \overline{x}$ eller $\theta = \overline{x}^2$. Alltså, ML-skattningen av θ är $\theta^*_{\text{obs}} = (\overline{x})^2$.

b) Skattningen θ^*_{obs} är inte väntevärdesriktig eftersom

$$E(\theta^*) = E\left(\overline{X}^2\right) = V\left(\overline{X}\right) + \left(E\left(\overline{X}\right)\right)^2 = \frac{\sigma^2}{n} + (\sqrt{\theta})^2 = \frac{\sigma^2}{n} + \theta \neq \theta,$$

men med $\hat{\theta}_{\text{obs}} = \overline{x}^2 - \sigma^2/n$ fås en väntevärdesriktig skattning.

11.30 Låt X beskriva antalet dragna kulor med olika färg tills två med samma färg erhålles. Möjliga värden på X ges av $S_X = \{1, 2, \ldots, N\}$. (Notera att antalet dragna kulor är $X+1$.)

Händelsen $X=1$ är att de två första kulorna har samma färg, det vill säga $P(X=1) = 1/N$. Händelsen $X=k$, $k>1$ är att de första k kulorna har olika färg och kula $k+1$ har någon av de tidigare k dragna färgerna. Alltså,

$$P(X=k) = \underbrace{\frac{N}{N}\cdot\frac{N-1}{N}\cdots\frac{N-(k-1)}{N}}_{k \text{ st}}\cdot\frac{k}{N} = \frac{N!}{(N-k)!N^k}\cdot\frac{k}{N}, \quad k = 1, 2, \ldots, N.$$

Med utfallet $x=3$ blir likelihoodfunktionen

$$L(N) = \frac{N!}{(N-x)!N^x}\cdot\frac{x}{N} = \frac{(N-1)(N-2)3}{N^3}$$

vilken är avtagande för stora N. För olika värden på N fås

N	$L(N)$
3	0.22222
4	0.28125
5	0.28800
6	0.27778
7	0.26239
8	0.24609

Maximum ges med valet $N=5$.

Anmärkning. Om man vill få en uppfattning om var detta maximum ligger någonstans kan man för tillfället låta N vara kontinuerlig i uttrycket $\frac{(N-1)(N-2)3}{N^3}$. Logaritmering och derivering ger att maximum nås i det N som löser

$$0 = \frac{d}{dN}\ln(L(N)) = \frac{d}{dN}\left[\ln(N-1)+\ln(N-2)+\ln(3)-3\ln(N)\right] = \frac{-(N-3-\sqrt{3})(N-3+\sqrt{3})}{N(N-1)(N-2)}$$

det vill säga $N = 3+\sqrt{3} = 4.73$. Alltså erhålls tillåtet maximum då $N=4$ eller $N=5$. Av dessa två är $L(N)$ störst då $N=5$.

11.31

12.1 Låt X vara χ^2-fördelad med 24 frihetsgrader. Bestäm a så att $P(X < a) = 0.95$. Då är

$$0.05 = 1 - P(X \leq a) = 1 - F_X(a) = 1 - F_X(\chi^2_{0.05})$$

det vill säga $a = \chi^2_{0.05} = 36.415$.

Bestäm b och c, $b < c$, så att $P(b < X < c) = 0.95$. Vi antar att händelsen $X \notin (b, c)$ är sådan att $P(X \leq b) = 0.025$ och $P(X \geq c) = 0.025$.

Då är

$$0.025 = 1 - P(X \leq c) = 1 - F_X(c) = 1 - F_X(\chi^2_{0.025})$$

dvs $c = \chi^2_{0.025} = 39.364$ och

$$0.975 = 1 - P(X \leq b) = 1 - F_X(b) = 1 - F_X(\chi^2_{0.975})$$

dvs $b = \chi^2_{0.975} = 12.401$.

12.2 Om $Z, Z_1, Z_2, \ldots, Z_n$ är oberoende $\mathrm{N}(0, 1)$ så är

$$X = \sum_{i=1}^{n} Z_i^2$$

χ^2-fördelad med n frihetsgrader, $X \sim \chi^2(n)$. För Z är

$$E\left(Z^2\right) = V(Z) + (E(Z))^2 = 1 + 0^2 = 1$$

så

$$E(X) = E\left(\sum_{i=1}^{n} Z_i^2\right) = \sum_{i=1}^{n} \underbrace{E\left(Z_i^2\right)}_{=1} = n.$$

Vidare, enligt ledning är $E\left(Z^4\right) = 3$ vilket medför att

$$V\left(Z^2\right) = E\left(Z^4\right) - \left(E\left(Z^2\right)\right)^2 = 3 - 1^2 = 2$$

och

$$V(X) = V\left(\sum_{i=1}^{n} Z_i^2\right) = \{\text{oberoende}\} = \sum_{i=1}^{n} \underbrace{V\left(Z_i^2\right)}_{=2} = 2n.$$

12.3 Om $Z, Z_1, Z_2, \ldots, Z_n$ är oberoende $\mathrm{N}(0, 1)$ så är

$$X = \sum_{i=1}^{n} Z_i^2$$

$\chi^2(n)$-fördelad. Ur uppgift 12.2 är

$$E(X) = n \qquad V(X) = 2n,$$

och enligt centrala gränsvärdessatsen är X approximativt normalfördelad, X är approximativt. $\mathrm{N}(n, \sqrt{2n})$. Det vill säga

$$1 - \Phi(\lambda_\alpha) = \alpha = P\left(X > \chi^2_\alpha\right) = P\left(\frac{X - n}{\sqrt{2n}} > \frac{\chi^2_\alpha - n}{\sqrt{2n}}\right) \approx 1 - \Phi\left(\frac{\chi^2_\alpha - n}{\sqrt{2n}}\right)$$

ger att

$$\lambda_\alpha \approx \frac{\chi^2_\alpha - n}{\sqrt{2n}}$$

eller

$$\chi^2_\alpha \approx n + \sqrt{2n}\lambda_\alpha.$$

12.4

12.5 Låt x beteckna antalet konfidensintervall som innehåller den därför avsedda konstanten. Då är x ett utfall av en binomialfördelad stokastisk variabel X, X är $\mathrm{Bin}(n, p)$, där $n = 15$ är antalet konfidensintervall och $p = 0.90$ är konfidensgraden för varje intervall.

a) Med X som $\mathrm{Bin}(n, p) = \mathrm{Bin}(15, 0.90)$ är $P(X = n) = p^n = 0.90^{15} = 0.20589$.

b) Typvärdet k, det värde som maximerar $P(X = k)$, är $k = 14$ då $P(X = 14) = 15p^{14}(1-p) = 0.34315$, alltså 14 av 15 intervall träffar och ett intervall missar.

12.6 Låt glödlampornas livslängder beskrivas av den stokastiska variabeln X med täthetsfunktion

$$f_X(x) = \frac{1}{\theta}\mathrm{e}^{-x/\theta}, x \geq 0.$$

Detta är exponentialfördelningen med intensitet $1/\theta$, det vill säga väntevärde θ. Fördelningsfunktionen är för $t \geq 0$

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\,dx = \int_0^t \frac{1}{\theta}\mathrm{e}^{-x/\theta}\,dx = 1 - \mathrm{e}^{-t/\theta}.$$

Nu är för $c > 0$

$$P(\theta < cX) = P\left(X > \frac{\theta}{c}\right) = 1 - F_X\left(\frac{\theta}{c}\right) = \mathrm{e}^{-(\theta/c)/\theta} = \mathrm{e}^{-1/c}.$$

Alltså, med $c = -1/\ln(0.975)$ är $P(\theta < cX) = 0.975$ och med $c = -1/\ln(0.025)$ är $P(\theta < cX) = 0.025$. Alltså,

$$P\left(\frac{X}{-\ln(0.025)} < \theta < \frac{X}{-\ln(0.0975)}\right) = 0.95$$

och ett 95% konfidensintervall för θ är

$$\frac{x}{-\ln(0.025)} < \theta < \frac{x}{-\ln(0.0975)}.$$

Med observationen $x = 1000$ erhålls intervallet

$$271.09 \leq \theta \leq 39498 \qquad (95\%).$$

12.7 Konfidensgraden väljes fortfarande till 95%. Med $c = -1/\ln(0.05)$ är $P(\theta < cX) = 0.05$ vilket ger

$$P\left(\frac{X}{-\ln(0.05)} < \theta\right) = 0.95$$

och ett 95% konfidensintervall för θ är

$$\frac{x}{-\ln(0.05)} < \theta.$$

Med observationen $x = 1000$ erhålls intervallet

$$333.81 \leq \theta \qquad (95\%)$$

eller, omformulerat, $\theta \in [333.81,\ \infty)$.

12.8

12.9 Låt X beskriva resultatet av en avståndsmätning. Modell:

$$X = \text{avstånd} + \text{mätfel} = \mu + \epsilon$$

där ϵ är $\mathrm{N}(0,\sigma)$, $\sigma = 5\cdot 10^{-3}$. Då är X $\mathrm{N}(\mu,\sigma)$. Låt $X_1,\ldots,X_n$ beskriva resultaten av n oberoende avståndsmätningar med observerade värden $x_1,\ldots,x_n$. Väntevärdet $\mu = E(X)$ skattas med $\bar{x}$ som beskrivs av $\overline{X}$, där

$$\overline{X} \text{ är } \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$$

så med sannolikhet $1-\alpha$ är

$$-\lambda_{\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \lambda_{\alpha/2}$$

eller, omformat, med sannolikhet $1-\alpha$ är

$$\overline{X} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Med $n = 4$ observationer fås $\bar{x} = 1132.155$. Konfidensgrad $1-\alpha = 0.95$ ger $\lambda_{\alpha/2} = \lambda_{0.025} = 1.9600$ och intervallet

$$\mu \in \bar{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1132.155 \pm 1.96\,\frac{5\cdot 10^{-3}}{\sqrt{4}} = 1132.155 \pm 0.0048999 \qquad (95\%)$$

eller

$$1132.150 \leq \mu \leq 1132.160 \qquad (95\%).$$

12.10 Låt $x_1,\ldots,x_n$ vara utfall av oberoende $\mathrm{N}(\mu,\sigma)$-fördelade stokastiska variabler $X_1,\ldots,X_n$ där $\sigma = 2$. Väntevärdet $\mu = E(X)$ skattas med $\bar{x}$ som beskrivs av $\overline{X}$, där

$$\overline{X} \text{ är } \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$$

så med sannolikhet $1-\alpha$ är

$$-\lambda_{\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \lambda_{\alpha/2}$$

eller, omformat, med sannolikhet $1-\alpha$ är

$$\overline{X} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Med $n = 4$ observationer fås $\bar{x} = 45.2$. Konfidensgrad $1-\alpha = 0.95$ ger $\lambda_{\alpha/2} = \lambda_{0.025} = 1.9600$ och intervallet

$$\mu \in \bar{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 45.2 \pm 1.96\,\frac{2}{\sqrt{4}} = 45.2 \pm 1.96 \qquad (95\%)$$

eller

$$43.24 \leq \mu \leq 47.16 \qquad (95\%).$$

12.11 Låt $x_1,\ldots,x_n$ vara utfall av oberoende $\mathrm{N}(\mu+\Delta,\sigma)$-fördelade stokastiska variabler $X_1,\ldots,X_n$ där $\sigma = 0.05$ och $\Delta = 0.10$. Om väntevärdet $\mu+\Delta$ skattas med $\bar{x}$ kan pH-halten μ skattas med $\bar{x}-\Delta$. Denna skattning beskrivs av $\overline{X}-\Delta$, där

$$\overline{X}-\Delta \text{ är } \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Alltså är

$$\frac{(\overline{X}-\Delta)-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$$

så med sannolikhet $1-\alpha$ är

$$-\lambda_{\alpha/2} \le \frac{(\overline{X}-\Delta)-\mu}{\sigma/\sqrt{n}} \le \lambda_{\alpha/2}$$

eller, omformat, med sannolikhet $1-\alpha$ är

$$(\overline{X}-\Delta)-\lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le (\overline{X}-\Delta)+\lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Med $n = 4$ observationer fås $\bar{x} = 8.2$. Konfidensgrad $1-\alpha = 0.99$ ger $\lambda_{\alpha/2} = \lambda_{0.005} = 2.5758$ och intervallet

$$\mu \in (\bar{x}-\Delta) \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (8.2-0.1) \pm 2.5758\,\frac{0.05}{\sqrt{4}} = 8.1 \pm 0.0644 \qquad (99\%)$$

eller

$$8.0356 \le \mu \le 8.1644 \qquad (99\%).$$

12.12 Konfidensintervallet för väntevärdet med konfidensgrad $1-\alpha$ ges av

$$\bar{x} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

det vill säga har bredden

$$L_{1-\alpha,n} = 2\lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Med konfidensgrad $1-\alpha = 0.90$ fås $\lambda_{\alpha/2} = \lambda_{0.05} = 1.6449$.

a) Vi söker n så att $L(0.90,n) = L(0.90,5)/2$ [alternativt $L(0.90,5)/10$].

$$2\lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 2\lambda_{\alpha/2}\frac{\sigma}{\sqrt{5}}\Big/\,2$$

ger $\sqrt{n} = 2\sqrt{5}$, eller $n = 4\cdot 5 = 20$. Med en tiondel så brett fås $\sqrt{n} = 10\sqrt{5}$ eller $n = 100\cdot 5 = 500$.

b) Vi söker n så att $L(0.99,n) = L(0.90,5)$.

$$2\lambda_{0.005}\frac{\sigma}{\sqrt{n}} = 2\lambda_{0.05}\frac{\sigma}{\sqrt{5}}$$

ger $\sqrt{n} = \sqrt{5}\lambda_{0.005}/\lambda_{0.05}$ eller, med $\lambda_{0.005} = 2.5758$, $n = 5(\lambda_{0.005}/\lambda_{0.05})^2 = 12.26$, det vill säga 13 observationer.

c) Vi söker n så att $L(0.99,n) = L(0.90,5)/2$ [alternativt $L(0.90,5)/10$].

$$2\lambda_{0.005}\frac{\sigma}{\sqrt{n}} = 2\lambda_{0.05}\frac{\sigma}{\sqrt{5}}\Big/\,2$$

ger $\sqrt{n} = 2\sqrt{5}\lambda_{0.005}/\lambda_{0.05}$ eller, med $\lambda_{0.005} = 2.5758$, $n = 4\cdot 5(\lambda_{0.005}/\lambda_{0.05})^2 = 49.047$, dvs. 50 stycken. Med en tiondel så brett fås $n = (10)^2\cdot 5(\lambda_{0.005}/\lambda_{0.05})^2 = 1226.2$, dvs. 1227 stycken.

Notera att i denna lösning utnyttjade vi aldrig det uppmätta intervallet [7.02, 7.14] eftersom vi inte behövde bestämma $\bar{x}$ eller σ.

12.13 Låt $x_1,\ldots,x_n$ vara avkastningen i ton under $n = 10$ dagar. Dessa modelleras som utfall av oberoende $\mathrm{N}(\mu,\sigma)$-fördelade stokastiska variabler $X_1,\ldots,X_n$. Alltså är $\overline{X} \sim \mathrm{N}(\mu,\sigma/\sqrt{n})$ och

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \text{ är } t(n-1)\text{-fördelad}$$

där S^2 ges av

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Således, med sannolikhet $1-\alpha$ är

$$-t_{\alpha/2} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2},$$

vilket omskrivet ger att med sannolikhet $1-\alpha$ är

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}.$$

Med $n = 10$ observationer och konfidensgrad $1-\alpha = 0.95$ fås ur $t(n-1) = t(9)$-tabell att $t_{\alpha/2} = t_{0.025} = 2.26$. Vidare beräknas

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 7.51 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.1543$$

$s = \sqrt{s^2} = 0.3929$, så konfidensintervallet blir

$$\mu \in \bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 7.51 \pm 0.281 = [7.229,\ 7.791] \qquad (95\%).$$

12.14 Från uppgift 12.11 vet vi att

$$\frac{(\overline{X}-\Delta)-\mu}{\sigma/\sqrt{n}} \;\sim\; \mathrm{N}(0,1).$$

Standardavvikelsen σ skattas med

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0.042426$$

så

$$\frac{(\overline{X}-\Delta)-\mu}{S/\sqrt{n}} \;\sim\; t(n-1)\text{-fördelad.}$$

Med sannolikhet $1-\alpha$ är

$$(\overline{X}-\Delta) - t_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq (\overline{X}-\Delta) + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

där kvantilen $t_{\alpha/2}$ fås ur $t(n-1)$-tabell.

Med $n = 4$ observationer fås $\bar{x} = 8.2$. Konfidensgrad $1-\alpha = 0.99$ ger $t_{\alpha/2} = t_{0.005} = 5.8409$ och intervallet

$$\mu \in (\bar{x}-\Delta) \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = (8.2-0.1) \pm 5.8409\,\frac{0.042426}{\sqrt{4}} = 8.1 \pm 0.1239 \qquad (99\%)$$

eller

$$7.9761 \leq \mu \leq 8.2239 \qquad (99\%).$$

12.15

12.16 Baserat på observationer $x_1, \ldots, x_n$ av oberoende $\mathrm{N}(\mu,\sigma)$-fördelade stokastiska variabler ges ett konfidensintervall, med konfidensgrad $1-\alpha$, för μ av

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}.$$

Intervallet $(\bar{x} - s,\ \bar{x} + s)$ har konfidensgrad åtminstone 99% om n är sådan att $t_{0.005}/\sqrt{n} < 1$.

n	$t_{0.005}$	$t_{0.005}/\sqrt{n}$	n	$t_{0.005}$	$t_{0.005}/\sqrt{n}$
1	63.657	63.657	7	3.4995	1.3227
2	9.9248	7.0179	8	3.3554	1.1863
3	5.8409	3.3723	9	3.2498	1.0833
4	4.6041	2.3020	10	3.1693	1.0022
5	4.0321	1.8032	11	3.1058	0.93644
6	3.7074	1.5136	12	3.0545	0.88177

Ur tabellen får vi att $n \geq 11$.

12.17 Skattningen

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

av σ^2 beskrivs av S^2 som är sådan att $(n-1)S^2/\sigma^2$ är en $\chi^2(n-1)$-fördelad stokastisk variabel. Alltså kan man ur $\chi^2(n-1)$-tabeller bestämma kvantilerna $\chi^2_{1-\alpha/2}$ och $\chi^2_{\alpha/2}$ så att

$$P\left(\chi^2_{1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2}\right) = 1 - \alpha$$

eller, omformulerat, med sannolikhet $1 - \alpha$ är

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}.$$

Med $n = 8$ observationer och skattningen $s = 5.2$ fås ur $\chi^2(7)$-tabell att $\chi^2_{0.975} = 1.69$ och $\chi^2_{0.025} = 16.0$. Alltså är ett 95% konfidensintervall för σ^2

$$11.821 = \frac{(8-1)(5.2)^2}{16.0} \leq \sigma^2 \leq \frac{(8-1)(5.2)^2}{1.69} = 112.01.$$

Motsvarande intervall för σ är

$$3.44 = \sqrt{11.821} \leq \sigma \leq \sqrt{112.01} = 10.6 \qquad (95\%).$$

12.18 Låt $x_1, \ldots, x_n$ vara furuplankornas uppmätta längder. Dessa modelleras som utfall av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $X_1, \ldots, X_n$. Väntevärdet μ skattas med $\bar{x}$ som beskrivs av $\overline{X}$ som är $\mathrm{N}(\mu, \sigma/\sqrt{n})$. Alltså är

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \quad t(n-1)\text{-fördelad.}$$

och med sannolikhet $1 - \alpha$ är

$$-t_{\alpha/2}(n-1) \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2}(n-1)$$

eller, omformulerat, med sannolikhet $1 - \alpha$ är

$$\mu \in \overline{X} \pm t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}.$$

Med observationer $x_1, \ldots, x_n$, $n = 16$:

5.8	5.9	5.1	3.5	4.2	4.9	5.3	5.3
4.7	3.9	4.5	4.1	4.0	4.2	4.7	4.8

fås skattningarna

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 4.6812$$

och

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{0.46962} = 0.68529.$$

Med konfidensgrad $1-\alpha = 0.95$ fås att $t_{\alpha/2}(15) = t_{0.025}(15) = 2.1314$ och konfidensintervallet blir

$$\mu \in 4.6812 \pm 2.1314\,\frac{0.68529}{\sqrt{16}} = 4.6812 \pm 0.36517 = [4.3161,\ 5.0464] \quad (95\%).$$

b) Vi vet att

$$\frac{(n-1)S^2}{\sigma^2} \quad \text{är} \quad \chi^2(n-1)\text{-fördelad}$$

om $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ där X_i är oberoende $\mathrm{N}(\mu, \sigma)$. Alltså är

$$P\left(\chi^2_{1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2}\right) = 1 - \alpha.$$

Med $\alpha = 0.05$ fås ur $\chi^2(n-1) = \chi^2(15)$-tabeller

$$P\left(6.26 \le \frac{(n-1)S^2}{\sigma^2} \le 27.5\right) = 0.95.$$

Alltså med sannolikhet 95% är

$$\frac{(n-1)S^2}{27.5} \le \sigma^2 \le \frac{(n-1)S^2}{6.26}.$$

Med skattningen $s^2 = 0.46962$ fås intervallet

$$0.25627 = \frac{(n-1)s^2}{27.5} \le \sigma^2 \le \frac{(n-1)s^2}{6.26} = 1.1249 \qquad (95\%)$$

eller

$$0.50623 = \sqrt{0.25627} \le \sigma \le \sqrt{1.1249} = 1.0606 \qquad (95\%).$$

12.19

12.20

12.21 Låt $X_1, \ldots, X_{n_x}$ beskriva de uppmätta överhöjningarna för betongelement från fabrik A och $Y_1, \ldots, Y_{n_y}$ motsvarande för fabrik B. Modell: alla stokastiska variabler är oberoende och X_i är $\mathrm{N}(\mu_x, \sigma)$ och Y_i är $\mathrm{N}(\mu_y, \sigma)$. Då är

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{S\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \quad t(n_x + n_y - 2)\text{-fördelad,}$$

där

$$S^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}.$$

Alltså är med sannolikhet $1 - \alpha$

$$\mu_x - \mu_y \in \overline{X} - \overline{Y} \pm t_{\alpha/2} S\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}.$$

Med observationer

$$\bar{x} = 18.1,\ s_x = 5.0,\ n_x = 9 \qquad \bar{y} = 14.6,\ s_y = 7.1,\ n_y = 16$$

fås

$$s = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} = \sqrt{41.572} = 6.4476.$$

Ur $t(n_x + n_y - 2) = t(23)$-tabeller fås att med konfidensgrad $1 - \alpha = 0.99$ är $t_{\alpha/2} = t_{0.005} = 2.8073$ så konfidensintervallet blir

$$\mu_x - \mu_y \in 18.1 - 14.6 \pm 2.8073 \cdot 6.4476\sqrt{\frac{1}{9} + \frac{1}{16}} = 3.5 \pm 7.54 \qquad (99\%)$$

eller intervallet $[-4.04,\ 11.0]$.

12.22 Blodtrycksmätningarna före och efter behandling sammanfattas i tabellen:

Person nr, i	1	2	3	4	5	6	7	8	9	10
Blodtryck före, x_i	75	70	75	65	95	70	65	70	65	90
Blodtryck efter, y_i	85	70	80	80	100	90	80	75	90	100
Förändring, $w_i = y_i - x_i$	10	0	5	15	5	20	15	5	25	10

Vi modellerar x_i som utfall av en $\mathrm{N}(\mu_i, \sigma_1)$-fördelad stokastiska variabel X_i, $i = 1, \ldots, n$, och y_i som ett utfall av en $\mathrm{N}(\mu_i + \Delta, \sigma_2)$-fördelad stokastiska variabel Y_i, $i = 1, \ldots, n$. Parametrarna μ_i är personens blodtryck före behandling och Δ är den genomsnittliga behandlingseffekten. Blodtrycksförändringarna $w_i = y_i - x_i$ är utfall av oberoende $N(\Delta, \sigma)$-fördelade stokastiska variabler $W_1, \ldots, W_n$, $i = 1, \ldots, n$.

Vi skattar Δ med $\Delta^*_{\text{obs}} = \overline{w} = 11$ och σ med

$$s_w = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(w_i - \overline{w})^2} = 7.746.$$

Ur $t(n-1) = t(9)$-tabell fås att $t_{0.025} = 2.2622$ och ett 95% konfidensintervall för Δ ges av

$$\Delta \in \overline{w} \pm t_{0.025}\frac{s}{\sqrt{n}} = 11 \pm 2.26 \cdot \frac{7.746}{\sqrt{10}} = 11 \pm 5.5 \quad (95\%).$$

12.23 Mätningarna $x_1, \ldots, x_n$ på lösningen med okänt pH-värde modelleras av oberoende $\mathrm{N}(\mu+\Delta, \sigma)$-fördelade stokastiska variabler $X_1, \ldots, X_n$.

Vidare, låt $y_1, \ldots, y_m$ vara de sex bestämningarna på en lösning med det kända pH-värdet 4.84. Dessa modelleras som utfall av oberoende $\mathrm{N}(4.84 + \Delta, \sigma)$-fördelade stokastiska variabler $Y_1, \ldots, Y_m$.

En skattning av Δ är $\Delta^*_{\text{obs}} = \bar{y} - 4.84 = 4.7 - 4.84 = -0.14$. Detta är en väntevärdesriktig skattning ty

$$E\left(\Delta^*\right) = E\left(\overline{Y} - 4.84\right) = E\left(\overline{Y}\right) - 4.84 = 4.84 + \Delta - 4.84 = \Delta.$$

Vidare så är

$$V\left(\Delta^*\right) = V\left(\overline{Y} - 4.84\right) = V\left(\overline{Y}\right) = \frac{\sigma^2}{m}.$$

a) En skattning av $\mu + \Delta$ är $\bar{x} = 4.285$ så en skattning av μ fås som $\mu^*_{\text{obs}} = \bar{x} - \Delta^*_{\text{obs}} = 4.425$. Denna är väntevärdesriktig ty

$$E\left(\mu^*\right) = E\left(\overline{X} - \Delta^*\right) = E\left(\overline{X}\right) - E\left(\Delta^*\right) = (\mu + \Delta) - \Delta = \mu.$$

b) Här är

$$V\left(\mu^*\right) = V\left(\overline{X} - \Delta^*\right) = \{\text{oberoende}\} = V\left(\overline{X}\right) + (-1)^2 V\left(\Delta^*\right) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \frac{(n+m)\sigma^2}{nm}$$

så $D(\mu^*) = \sigma\sqrt{\frac{n+m}{nm}}$.

c) För att skatta $D(\mu^*)$ måste σ skattas. Från $x_1, \ldots, x_n$ har vi skattningen

$$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0.072342$$

och från $y_1, \ldots, y_m$ har vi att σ kan skattas med

$$s_y = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(y_i - \overline{y})^2} = 0.089443.$$

Båda dessa skattningar kombineras till s, där

$$s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{(n-1)+(m-1)} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2 + \sum_{i=1}^{m}(y_i - \overline{y})^2}{n+m-2} = 0.0069625$$

så $s = 0.083442$. Standardavvikelsen $D(\mu^*) = \sigma\sqrt{\frac{n+m}{nm}}$ skattas med

$$s\sqrt{\frac{n+m}{nm}} = 0.083442\sqrt{\frac{4+6}{24}} = 0.053861,$$

det vill säga medelfelet $d(\mu^*) = 0.054$.

d) Ur $t(n+m-2) = t(8)$-tabeller fås $t_{0.025} = 2.306$ och ett 95% konfidensintervall för μ ges av

$$\mu \in \mu^*_{\text{obs}} \pm t_{0.025}\, d(\mu^*) = 4.425 \pm 2.306 \cdot 0.053861 = 4.425 \pm 0.124 \qquad (95\%).$$

12.24 Observationerna $x_1, \ldots, x_n$ och $y_1, \ldots, y_n$ sammanfattas av tabellen:

Förare, i	1	2	3	4	5
Förslitning däcktyp A, x_i	1.0	0.9	0.7	1.5	0.5
Förslitning däcktyp B, y_i	0.9	0.7	0.8	1.2	0.5
Förslitningsskillnad, $w_i = x_i - y_i$	0.1	0.2	−0.1	0.3	0

De uppmätta parvisa förslitningsskillnaderna $w_1, \ldots, w_n$ modelleras som utfall av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $W_1, \ldots, W_n$ där μ mäter hur mycket mer i genomsnitt A-däck slits än B-däck.

Parametrarna μ och σ^2 skattas med

$$\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i = 0.10 \quad \text{respektive} \quad s_w^2 = \frac{1}{n-1}\sum_{i=1}^{n}(w_i - \overline{w})^2 = 0.025$$

så σ skattas med $s_w = \sqrt{0.025} = 0.1581$.

Med $n = 5$ observationer $w_1, \ldots, w_n$ och konfidensgrad $1 - \alpha = 0.95$ fås ur $t(n-1) = t(4)$-tabell att $t_{\alpha/2} = t_{0.025} = 2.7764$, så det observerade intervallet för den genomsnittliga skillnaden i däckförslitning blir

$$\mu \in \overline{w} \pm t_{\alpha/2}\frac{s_w}{\sqrt{n}} = 0.10 \pm 0.1963 = [-0.0963,\ 0.2963] \qquad (95\%).$$

12.25 a) Observationerna $x_1, \ldots, x_{n_1}$ av levervärden för personer utan medicinering modelleras som utfall av $\mathrm{N}(\mu_x, \sigma)$-fördelade stokastiska variabler. Observationerna $y_1, \ldots, y_{n_2}$ av levervärden för personer med behandling modelleras som utfall av $\mathrm{N}(\mu_y, \sigma)$-fördelade stokastiska variabler. Samtliga stokastiska variabler förutsätts vara oberoende.

Med $n_1 = 50$ och $n_2 = 25$ erhöll man

$$\bar{x} = 148.2 \quad \bar{y} = 151.7 \quad s_x = 10.0 \quad s_y = 8.0$$

så skillnaden $\mu_y - \mu_x$ skattas med $\bar{y} - \bar{x} = 3.5$.

Standardavvikelsen σ skattas med s där

$$s^2 = \frac{(n_1-1)s_x^2 + (n_2-1)s_y^2}{(n_1-1)+(n_2-1)} = \frac{\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{i=1}^{n_2}(y_i-\bar{y})^2}{n_1+n_2-2} = 88.164$$

så $s = 9.3896$. Eftersom $D\left(\overline{Y} - \overline{X}\right) = \sigma\sqrt{\frac{1}{n_2} + \frac{1}{n_1}}$ så är medelfelet $d(\overline{Y} - \overline{X}) = s\sqrt{\frac{1}{n_2} + \frac{1}{n_1}} = 2.3$.

Ur $t(n_1 + n_2 - 2) = t(73)$-tabeller fås $t_{0.025} = 1.993$ och ett 95% konfidensintervall för $\mu_y - \mu_x$ ges av

$$\mu_y - \mu_x \in (\bar{y} - \bar{x}) \pm t_{0.025}\, d(\overline{Y} - \overline{X}) = 3.5 \pm 1.993 \cdot 2.3 = 3.5 \pm 4.5 = [-1.08,\ 8.08] \quad (95\%).$$

b) Observationerna $x_1, \ldots, x_n$ av levervärden för personer före medicinering och $y_1, \ldots, y_n$ motsvarande efter medicinering, sammanfattas av

$$n = 25 \quad \bar{x} = 149.0 \quad \bar{y} = 150.9.$$

Förändringarna $z_i = y_i - x_i$ modelleras av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $Z_1, \ldots, Z_n$. Den genomsnittliga höjningen μ skattas med $\bar{z} = \bar{y} - \bar{x} = 1.9$ och σ skattas med $s_z = 1.6$.

Eftersom $D\left(\overline{Z}\right) = \sigma/\sqrt{n}$ så är medelfelet $d(\overline{Z}) = s_z/\sqrt{n} = 0.32$.

Ur $t(n-1) = t(24)$-tabeller fås $t_{0.025} = 2.064$ och ett 95% konfidensintervall för μ ges av

$$\mu \in \bar{z} \pm t_{0.025}\, d(\overline{Z}) = 1.9 \pm 2.064 \cdot 0.32 = 1.9 \pm 0.66 = [1.24,\ 2.56] \quad (95\%).$$

12.26

12.27 Låt $x_1, \ldots, x_{n_1}$ vara utfall av $\mathrm{N}(\mu_1, \sigma)$-fördelade stokastiska variabler och $y_1, \ldots, y_{n_2}$ vara utfall av $\mathrm{N}(\mu_2, \sigma)$-fördelade stokastiska variabler. Alla stokastiska variabler antas vara oberoende.

Skillnaden $\mu_1 - \mu_2$ skattas med $\bar{x} - \bar{y} = 49.2 - 37.4 = 11.8$. Vidare är $D\left(\overline{X} - \overline{Y}\right) = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ där σ skattas med

$$s = \sqrt{\frac{(n_1-1)s_x^2 + (n_2-1)s_y^2}{(n_1-1)+(n_2-1)}} = \sqrt{\frac{\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{i=1}^{n_2}(y_i-\bar{y})^2}{n_1+n_2-2}} = 2.157.$$

Alltså är medelfelet $d\left(\overline{X} - \overline{Y}\right) = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.0785$. Ur $t(n_1+n_2-2) = t(16)$-tabeller fås $t_{0.05} = 1.7459$ och ett 90% konfidensintervall för $\mu_1 - \mu_2$ ges av

$$\mu_1 - \mu_2 \in \bar{x} - \bar{y} \pm t_{0.05}\, d(\overline{X} - \overline{Y}) = 11.8 \pm 1.7459 \cdot 1.0785 = 11.8 \pm 1.88 = [9.92,\ 11.8] \quad (90\%).$$

12.28

12.29 Låt $x_1, \ldots, x_n$ vara studietiderna för de $n = 25$ personerna som undersökts. Dessa modelleras som utfall av oberoende och likafördelade stokastiska variabler $X_1, \ldots, X_n$ med $E(X) = \mu$ och $D(X) = \sigma$. Väntevärdet skattas med $\bar{x} = 49.3$ och σ med $s = 9.3$. Skattningen $\bar{x}$ har medelfel $s/\sqrt{n} = 1.86$.

Med Centrala gränsvärdessatsen är $\sum_{i=1}^{n} X_i$ approximativt normalfördelad och således är $\overline{X}$ approximativt $\mathrm{N}(\mu, \sigma/\sqrt{n})$. Ett konfidensintervall för μ ges av

$$\mu \in \bar{x} \pm \lambda_{0.025}\frac{s}{\sqrt{n}} = 49.3 \pm 1.96 \cdot 1.86 = 49.3 \pm 3.65 = [45.654,\ 52.946] \quad (\approx 95\%).$$

12.30 Låt $x_1, \ldots, x_n$, $n = 30$, vara de uppmätta föroreningshalterna uppströms och $y_1, \ldots, y_r$, $r = 40$, motsvarande nedströms. Vi ansätter modellen att $x_1, \ldots, x_n$ är utfall av likafördelade stokastiska variabler $X_1, \ldots, X_n$ med $E(X) = \mu_x$ och $D(X) = \sigma_x$. På motsvarande sätt är $y_1, \ldots, y_r$ utfall av likafördelade $Y_1, \ldots, Y_r$ där $E(Y) = \mu_y$ och $D(Y) = \sigma_y$. Alla stokastiska variabler förutsätts vara oberoende.

Skillnaden $\mu_y - \mu_x$ skattas med $\bar{y} - \bar{x} = 86.1 - 13.2 = 72.9$. Skattningen $\bar{y} - \bar{x}$ är ett utfall av en stokastisk variabel $\overline{Y} - \overline{X}$ med standardavvikelse $D\left(\overline{Y} - \overline{X}\right) = \sqrt{\frac{\sigma_y^2}{r} + \frac{\sigma_x^2}{n}}$. Alltså har skattningen $\bar{y} - \bar{x}$ medelfel

$$d\left(\overline{Y} - \overline{X}\right) = \sqrt{\frac{s_y^2}{r} + \frac{s_x^2}{n}} = \sqrt{\frac{(38.7)^2}{40} + \frac{(2.80)^2}{30}} = 6.1403.$$

Med Centrala gränsvärdessatsen är $\sum_{i=1}^n X_i$ och $\sum_{i=1}^r Y_i$ approximativt normalfördelade och således är $\overline{Y} - \overline{X}$ approximativt N $\left(\mu_y - \mu_x, \sqrt{\frac{\sigma_y^2}{r} + \frac{\sigma_x^2}{n}}\right)$. Ett konfidensintervall för skillnaden $\mu_y - \mu_x$ ges av

$$\mu_y - \mu_x \in \bar{y} - \bar{x} \pm \lambda_{0.025}\sqrt{\frac{s_y^2}{r} + \frac{s_x^2}{n}} = 72.9 \pm 1.96 \cdot 6.1403 = 72.9 \pm 12.04 = [60.87,\ 84.94] \quad (\approx 95\%).$$

12.31 Av $n = 600$ undersökta enheter befanns $x = 24$ vara felaktiga. Om partiet är mycket stort kan vi modellera x som ett utfall av en binomialfördelad stokastisk variabel X, det vill säga X är $\text{Bin}(n, p)$ där p är andelen felaktiga enheter i partiet.

Vi skattar p med $p^*_{\text{obs}} = x/n = 0.04$. Skattningen p^*_{obs} beskrivs av $p^* = X/n$ som har

$$V(p^*) = V\left(\frac{1}{n}X\right) = \frac{1}{n^2}\underbrace{V(X)}_{np(1-p)} = \frac{p(1-p)}{n}$$

och standardavvikelse $D(p^*) = \sqrt{p(1-p)/n}$. Medelfelet för skattningen p^*_{obs} är

$$d(p^*) = \sqrt{\frac{p^*_{\text{obs}}(1 - p^*_{\text{obs}})}{n}} = \sqrt{\frac{0.040(1 - 0.040)}{600}} = 0.008.$$

Eftersom $V(X) = np(1-p)$ skattas med $np^*_{\text{obs}}(1 - p^*_{\text{obs}}) = 23.04$ vilket är större än 10 kan binomialfördelningen approximeras med normalfördelningen. Alltså är även $p^* = X/n$ approximativt normalfördelad. Med $1 - \alpha = 0.95$ är $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$ och ett konfidensintervall för p ges av

$$p \in p^*_{\text{obs}} \pm \lambda_{0.025} d(p^*) = 0.040 \pm 1.96 \cdot 0.008 = 0.040 \pm 0.0157 = [0.024,\ 0.056] \quad (\approx 95\%).$$

12.32

12.33 Låt x_1 och x_2 vara antalet sympatisörer med det borgerliga blocket i oktober och november. Vi modellerar x_1 och x_2 som utfall av oberoende binomialfördelade stokastiska variabler X_1 och X_2, där X_1 är $\text{Bin}(n_1, p_1)$ och X_2 är $\text{Bin}(n_2, p_2)$. Förändringen $p_2 - p_1$ skattas med $(p_2)^*_{\text{obs}} - (p_1)^*_{\text{obs}} = 0.456 - 0.465 = -0.009$. Eftersom

$$V(p_2^* - p_1^*) = \{\text{ober.}\} = V(p_2^*) + (-1)^2 V(p_1^*) = \frac{1}{n_2^2}V(X_2) + \frac{1}{n_1^2}V(X_1) = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$$

fås medelfelet för $(p_2)^*_{\text{obs}} - (p_1)^*_{\text{obs}}$ till

$$\begin{aligned} d(p_2^* - p_1^*) &= \sqrt{\frac{(p_2)^*_{\text{obs}}(1 - (p_2)^*_{\text{obs}})}{n_2} + \frac{(p_1)^*_{\text{obs}}(1 - (p_1)^*_{\text{obs}})}{n_1}} = \sqrt{\frac{0.456(1 - 0.456)}{1689} + \frac{0.465(1 - 0.465)}{1704}} \\ &= 0.017113. \end{aligned}$$

Varianserna $V(X_1) = n_1p_1(1-p_1)$ och $V(X_2) = n_2p_2(1-p_2)$ skattas med $n_1(p_1)^*_{\text{obs}}(1-(p_1)^*_{\text{obs}}) = 423.91$ respektive $n_2(p_2)^*_{\text{obs}}(1-(p_2)^*_{\text{obs}}) = 418.98$. Eftersom båda är större än 10 (med råge!) kan binomialfördelningarna approximeras med normalfördelningar och även $p_2^* - p_1^*$ kan antas vara approximativt normalfördelad.

Med $1-\alpha = 0.95$ fås $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$ och ett konfidensintervall för förändringen $p_2 - p_1$ ges av

$$p_2 - p_1 \in (p_2)^*_{\text{obs}} - (p_1)^*_{\text{obs}} \pm \lambda_{0.025} d(p_2^* - p_1^*) = -0.009 \pm 1.96 \cdot 0.017 = -0.009 \pm 0.0335 \quad (\approx 95\%).$$

12.34 Låt $x = 36$ vara antalet defekta bland de $n = 1000$ undersökta enheterna. Vi modellerar x som ett utfall av en hypergeometriskt fördelad stokastisk variabel X, X är Hypergeo(N, n, p) där $N = 100000$ och p är andelen defekta.

Andelen p skattas med $p^*_{\text{obs}} = x/n = 0.036$. Eftersom $V(X) = np(1-p)(N-n)/(N-1)$ skattas till

$$np^*_{\text{obs}}(1-p^*_{\text{obs}})\frac{N-n}{N-1} = 1000 \cdot 0.036(1-0.036) \cdot 0.99 = 34.357$$

vilket är större än 10 kan den hypergeometriska fördelningen approximeras med en normalfördelning och även $p^* = X/n$ är approximativt normalfördelad. Medelfelet för p^*_{obs} är

$$d(p^*) = \sqrt{\frac{p^*_{\text{obs}}(1-p^*_{\text{obs}})}{n} \cdot \frac{N-n}{N-1}} = \sqrt{\frac{0.036(1-0.036)}{1000} \cdot 0.99} = 0.0058615.$$

Med $1-\alpha = 0.95$ fås $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$ och ett konfidensintervall för andelen defekta p ges av

$$p \in p^*_{\text{obs}} \pm \lambda_{0.025} d(p^*) = 0.036 \pm 1.96 \cdot 0.0058615 = 0.036 \pm 0.0115 \quad (\approx 95\%).$$

b) Det totala antalet defekta i partiet Np skattas med $Np^*_{\text{obs}} = 3600$. Denna skattning har medelfel

$$d(Np^*) = N\sqrt{\frac{p^*_{\text{obs}}(1-p^*_{\text{obs}})}{n} \cdot \frac{N-n}{N-1}} = 586.15$$

Ett konfidensintervall för antalet defekta ges av

$$Np \in Np^*_{\text{obs}} \pm \lambda_{0.025} d(Np^*) = 3600 \pm 1.96 \cdot 586.15 = 3600 \pm 1149 \quad (\approx 95\%).$$

12.35

12.36

12.37 Låt $x_1, \ldots, x_n$ vara antal samtal under det aktuella tidsintervallet för de olika dagarna. Dessa modelleras som utfall av oberoende Poisson(μ)-fördelade stokastiska variabler $X_1, \ldots, X_n$. Parametern μ (väntevärdet) skattas med

$$\mu^*_{\text{obs}} = \bar{x} = 100.88 \text{ samtal.}$$

Skattningen beskrivs av

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

där

$$Y = X_1 + \cdots + X_n$$

är Poissonfördelad med väntevärde

$$E(Y) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = n\mu$$

som skattas med $\sum x_i = 807$ vilket är större än 15 med råge. Alltså kan Poissonfördelningen för Y approximeras med normalfördelning, men då är även $\overline{X}$ approximativt normalfördelad. Alltså,

$$\overline{X} \text{ är approximativt N}\left(\mu, \sqrt{\frac{\mu}{n}}\right)$$

Ett konfidensintervall för m med konfidensgrad approximativt $1-\alpha = 0.95$ ges av

$$\mu \in \bar{x} \pm \lambda_{\alpha/2}\sqrt{\frac{\bar{x}}{n}} = 100.88 \pm 6.9599 = [93.9,\ 107.8] \qquad (\approx 95\%).$$

12.38

12.39

12.40

13.1 Låt x vara antalet gånger som Pål svarar rätt.

a) Om Pål chansar ansätter vi modellen att är x ett utfall av en binomialfördelad stokastisk variabel X, X är Bin(n,p) med $n = 15$ och $p = 0.5$.

b) I binomialfördelningen är

$$P(X \geq 11) = \sum_{k=11}^{15} \binom{n}{k} p^k (1-p)^{n-k} = 0.05923.$$

Om man har tillgång till fördelningsfunktionen för binomialfördelningen kan sannolikheten erhållas enligt

$$P(X \geq 11) = 1 - P(X < 11) = 1 - P(X \leq 10) = 1 - F_X(10) = 1 - 0.94077 = 0.05923.$$

13.2

13.3

13.4 Låt X beskriva en glödlampas lystid i timmar. Med modellen att X är exponentialfördelad med väntevärde $\theta > 0$ har X täthetsfunktion

$$f_X(x) = \frac{1}{\theta}\mathrm{e}^{-x/\theta}, \quad x \geq 0.$$

Fördelningsfunktionen för X blir då för $t > 0$

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(x)\,dx = \int_0^t \frac{1}{\theta}\mathrm{e}^{-x/\theta}\,dx = 1 - \mathrm{e}^{-t/\theta}.$$

a) Ekvationen $\alpha = P(X < a) = F_X(a) = 1 - \mathrm{e}^{-a/\theta}$ ger $a = -\theta \ln(1-\alpha)$ och med $\theta = 1000$ erhålls $a = -1000\ln(1-\alpha)$.

b) Med $\alpha = 0.05$ är $a = -1000\ln(1-0.05) = 51.293$. Vi förkastar hypotesen $\theta = 1000$ till förmån för hypotesen $\theta < 1000$ på signifikansnivå 5% om $x < 51.293$. Här är $x = 75$ och hypotesen $\theta = 1000$ förkastas ej.

c) Med $x = 50 < 51.293$ förkastas hypotesen $\theta = 1000$ på nivå 5% till förmån för hypotesen $\theta < 1000$.

d) Om a väljes så att $a = x = 45$ är

$$P(X < a) = F_X(a) = 1 - \mathrm{e}^{-a/\theta} = 1 - \mathrm{e}^{-45/1000} = 0.044003.$$

Alltså, den lägsta signifikansnivå hypotesen $\theta = 1000$ kan förkastas på, givet observationen $x = 45$, är 4.4%.

13.5 Låt x vara antalet spel till första vinst. Vi modellerar x som ett utfall av den ffg(p)-fördelade stokastiska variabeln X, det vill säga

$$p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \dots$$

Då är

$$P(X > k) = (1-p)^k, \qquad k = 0, 1, 2, 3, \dots$$

Med $H_0 : p = 0.2$ och $H_1 : p < 0.2$ så förkastas H_0 till förmån för H_1 för stora värden på x. Med förkastelseområde $\{x, x+1, x+2, \dots, x\}$ är signifikansnivån för testet

$$P(X \geq x | H_0) = P(X > x-1 | H_0) = (1-p)^{x-1}\big|_{p=0.20} = (0.80)^{x-1}.$$

Med utfallet $x = 11$ förkastas H_0 på (lägsta) signifikansnivå

$$P(X \geq 11 | H_0) = P(X > 10 | H_0) = (0.80)^{10} = 0.1074.$$

Alltså, vi förkastar inte H_0 på nivå 10%.

13.6 Med $x_1, \dots, x_n$ som utfall av oberoende N(μ, σ)-fördelade stokastiska variabler $X_1, \dots, X_n$ är $\bar{x}$ ett utfall av $\overline{X}$ som är N($\mu, \sigma/\sqrt{n}$)-fördelad. Ett test av

$$H_0 : \mu = 0.5 \text{ (oskyldig)} \quad \text{mot} \quad H_1 : \mu > 0.5 \text{ (skyldig)}$$

görs på signifikansnivå $\alpha = 1\%$ om H_0 förkastas då $\bar{x} > 0.5 + \lambda_{0.01}\sigma/\sqrt{n}$. Eftersom P (Förkasta sann H_0) $\leq \alpha$ kommer sannolikheten för att en oskyldig person förklaras skyldig vara högst 1%, dvs alternativ 2 är en riktig tolkning.

I alternativ 3 beskrivs sannolikheten för fel av andra slaget, det vill säga P (ej förkasta falsk H_0) och den begränsas inte av signifikansnivån. Alternativ 1 och 4 är en form av omvända betingade sannolikheter: givet att hypotesen förkastas, vad är sannolikheten...

13.7 Vi modellerar de uppmätta planklängderna $x_1, \dots, x_n$ som utfall av oberoende N(μ, σ)-fördelade stokastiska variabler $X_1, \dots, X_n$. Vi vill testa hypotesen

$$H_0 : \sigma = 0.4 \quad \text{mot} \quad H_1 : \sigma \neq 0.4$$

på signifikansnivå $\alpha = 0.05$.

Alternativ 1. Med skattningen $s^2 = 0.46962$ fås ett konfidensintervall för σ med konfidensgrad $1 - \alpha = 0.95$ av (uppgift 12.18b)

$$0.50623 = \sqrt{\frac{(n-1)s^2}{27.5}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{6.26}} = 1.0606 \qquad (95\%).$$

Eftersom 0.4 inte ingår i intervallet förkastas $H_0 : \sigma = 0.4$ till förmån för H_1 på risknivå $\alpha = 0.05$.

Alternativ 2. Vi förkastar H_0 då s är mycket större eller mycket mindre än 0.4, alternativt då

$$\frac{(n-1)s^2}{0.4^2}$$

är väldigt stor eller väldigt liten. Då H_0 är sann så är

$$\frac{(n-1)S^2}{0.4^2} \quad \text{en} \quad \chi^2(n-1)\text{-fördelad stokastisk variabel.}$$

Alltså, med $\alpha = 0.05$ så är

$$P\left(6.26 \leq \frac{(n-1)S^2}{0.4^2} \leq 27.5\right) = 0.95.$$

Vi förkastar H_0 till förmån för H_1 om vi observerar att $(n-1)s^2/0.4^2$ mindre än 6.26 eller större än 27.5. Här:

$$\frac{(n-1)s^2}{0.4^2} = 44.027$$

så H_0 förkastas på nivå $\alpha = 0.05$.

13.8 Vi modellerar den uppmätta grumligheten $x_1, \ldots, x_n$ som utfall av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $X_1, \ldots, X_n$ där $\sigma = 0.2$. Vi vill testa hypotesen

$$H_0 : \mu = 4.0 \quad \text{mot} \quad H_1 : \mu > 4.0$$

på signifikansnivå $\alpha = 0.05$.

Alternativ 1: Vi förkastar H_0 för stora värden på $\bar{x}$, alternativ, då $(\bar{x}-4)/(\sigma/\sqrt{n})$ är stor. Då H_0 är sann är

$$\frac{\overline{X}-4}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$$

så $P\left((\overline{X}-4)/(\sigma/\sqrt{n}) > \lambda_{0.05} | H_0\right) = 0.05$.

Alltså, H_0 förkastas om $(\bar{x}-4)/(\sigma/\sqrt{n}) > \lambda_{0.05}$. Här är $\lambda_{0.05} = 1.6449$ och $\bar{x} = 4.1$ och

$$\frac{\bar{x}-4}{\sigma/\sqrt{n}} = \frac{4.1-4}{0.2/\sqrt{10}} = 1.5811 < \lambda_{0.05},$$

så H_0 förkastas ej på nivå 5%.

Alternativ 2: Vi förkastar H_0 för stora värden på $\bar{x}$. Då H_0 är sann är

$$P\left(\overline{X} > x | H_0\right) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > \frac{x-\mu}{\sigma/\sqrt{n}} \middle| H_0\right) = 1 - \Phi\left(\frac{x-4.0}{0.2/\sqrt{10}}\right).$$

Med observationen $\bar{x} = 4.1$ fås att den lägsta signifikansnivån H_0 kan förkastas på är

$$P\left(\overline{X} > 4.1 | H_0\right) = 1 - \Phi\left(\frac{4.1-4.0}{0.2/\sqrt{10}}\right) 1 - \Phi\left(1.5811\right) = 0.056923 > 0.05.$$

Alltså kan vi inte förkasta H_0 på nivå 5%.

13.9 Teststorheten

$$\frac{\overline{X}-4}{\sigma/\sqrt{n}}$$

är normalfördelad med väntevärde

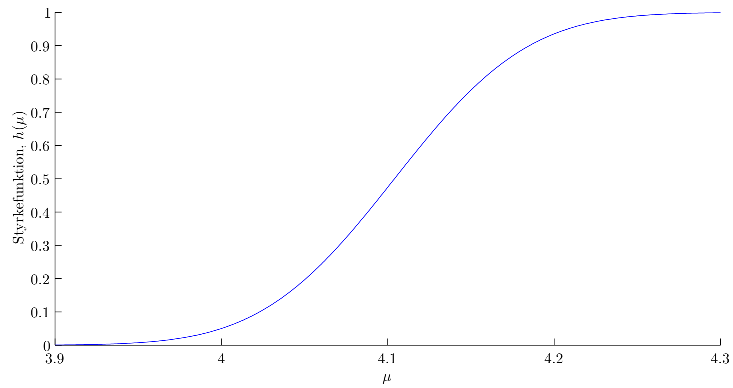
$$E\left(\frac{\overline{X}-4}{\sigma/\sqrt{n}}\right) = \frac{E\left(\overline{X}\right)-4}{\sigma/\sqrt{n}} = \frac{\mu-4}{\sigma/\sqrt{n}}$$

och varians

$$V\left(\frac{\overline{X}-4}{\sigma/\sqrt{n}}\right) = \frac{V\left(\overline{X}\right)}{\sigma^2/n} = \frac{\sigma^2/n}{\sigma^2/n} = 1.$$

Styrkefunktionen blir då

$$\begin{aligned} h(\mu) &= P\left(\text{Förkasta } H_0\right) = P\left(\frac{\overline{X}-4}{\sigma/\sqrt{n}} > \lambda_{0.05}\right) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > \lambda_{0.05} + \frac{4-\mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\lambda_{0.05} + \frac{4-\mu}{\sigma/\sqrt{n}}\right). \end{aligned}$$



Styrkefunktionen $h(\mu)$, det vill säga sannolikheten att förkasta H_0 som funktion av μ.

$$\begin{aligned} h(3.8) &= 1 - \Phi\left(1.6449 + \frac{4-3.8}{0.2/\sqrt{10}}\right) = 1 - \Phi\left(4.8071\right) = 7.7 \cdot 10^{-7} \\ h(4.3) &= 1 - \Phi\left(1.6449 + \frac{4-4.3}{0.2/\sqrt{10}}\right) = 1 - \Phi\left(-3.0986\right) = 0.99903. \end{aligned}$$

13.10 Låt $x_1, \ldots, x_n$ vara de uppmätta smältpunkterna. Vi modellerar $x_1, \ldots, x_n$ som utfall av oberoende och likafördelade stokastiska variabler $X_1, \ldots, X_n$, där

$$X = \text{smältpunkt} + \text{mätfel} = \mu + \epsilon$$

där ϵ är $\mathrm{N}(0, \sigma)$, $\sigma = 2.3$. Då är X_i $\mathrm{N}(\mu, \sigma)$. Vi vill testa:

$$H_0 : \mu = \mu_0 = 1050^\circ \quad \text{mot} \quad H_1 : \mu \neq \mu_0$$

på nivå $\alpha = 0.05$.

Vi förkastar H_0 för stora värden på $|\overline{x} - \mu_0|$, alternativt stora värden på $|z|$ där

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

som då H_0 är sann är ett utfall på Z, en $\mathrm{N}(0, 1)$-fördelad stokastisk variabel.

Alternativ 1: Alltså, om H_0 är sann så är $P\left(|Z| > \lambda_{\alpha/2}\right) = \alpha$. Med $\alpha = 0.05$ är $\lambda_{\alpha/2} = 1.96$ och vi förkastar H_0 till förmån för H_1 om vi observerar ett utfall $|z| > 1.96$. Här fås

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1050.9 - 1050}{2.3/\sqrt{10}} = 1.2649$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

Alternativ 2: Vi observerar utfallet

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1050.9 - 1050}{2.3/\sqrt{10}} = 1.2649.$$

Om H_0 är sann är

$$P\left(|Z| > 1.2649\right) = 2(1 - \Phi\left(1.2649\right)) = 0.2059$$

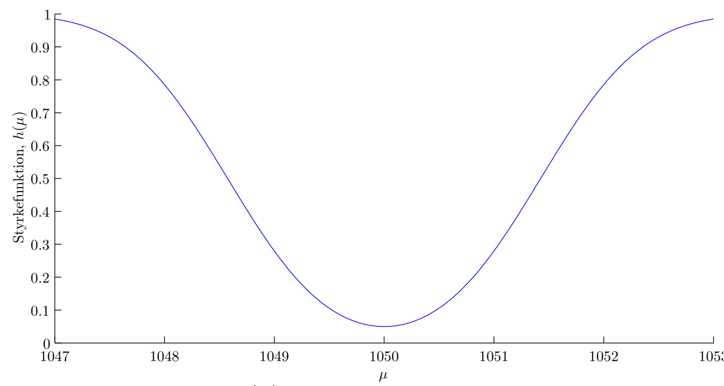
så den lägsta signifikansnivå vi skulle förkasta H_0 på är 0.2059, det vill säga vi förkastar inte H_0 på nivå 5%.

Styrkefunktionen $h(\mu)$ kan bestämmas som

$$\begin{aligned}
h(\mu) &= P\,(\text{förkasta } H_0) = P\left(\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| > \lambda_{\alpha/2}\right) = 1 - P\left(-\lambda_{\alpha/2} \le \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \le \lambda_{\alpha/2}\right) \\
&= 1 - P\left(-\lambda_{\alpha/2} + \frac{\mu_0-\mu}{\sigma/\sqrt{n}} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \lambda_{\alpha/2} + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right) \\
&= 1 - \left(\Phi\left(\lambda_{\alpha/2} + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right) - \Phi\left(-\lambda_{\alpha/2} + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right)\right)
\end{aligned}$$

Vi beräknar

$$h(1051) = 0.27968 \qquad h(1053) = 0.9848.$$



Styrkefunktionen $h(\mu)$, det vill säga sannolikheten att förkasta $H_0 : \mu = 1050^\circ$ som funktion av μ.

13.11

13.12 Om σ inte är känd så skattas den med stickprovsstandardavvikelsen $s = 2.028$.

Vi förkastar H_0 för stora värden på $|\overline{x} - \mu_0|$, alternativt stora värden på $|t|$ där

$$t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}}$$

som då H_0 är sann är ett utfall av en $t(n-1)$-fördelad stokastisk variabel T. Alltså, om H_0 är sann så är $P\left(|T| > t_{\alpha/2}\right) = \alpha$. Med $\alpha = 0.05$ är $t_{\alpha/2} = 2.2622$ och vi förkastar H_0 till förmån för H_1 om vi observerar $|t| > 2.2622$. Här fås

$$t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}} = \frac{1050.9-1050}{2.028/\sqrt{10}} = 1.4345$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

13.13 Ur uppgift 12.24 fås att ett konfidensintervall för den genomsnittliga förslitningsskillnaden μ är

$$\mu \in \overline{w} \pm t_{\alpha/2}\frac{s_w}{\sqrt{n}} = 0.10 \pm 0.1963 = [-0.0963,\ 0.2963] \qquad (95\%).$$

Eftersom $\mu = 0$ inte är ett orimligt värde på förslitningsskillnaden, 0 ingår i konfidensintervallet, kan inte hypotesen $H_0 : \mu = 0$ förkastas till förmån för $H_1 : \mu \neq 0$ på nivå 5%.

13.14

13.15 Låt $x_1, \ldots, x_n$ vara de uppmätta Hg-halterna i gäddorna. Dessa modelleras som utfall av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $X_1, \ldots, X_n$.

a) Vi vill med $\mu_0 = 0.9$ testa hypotesen

$$H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu > \mu_0$$

på nivå $\alpha = 0.05$.

Vi förkastar H_0 för stora värden på

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

som då H_0 är sann är ett utfall av en $t(n-1)$-fördelad stokastisk variabel T. Alltså, om H_0 är sann så är $P(T > t_\alpha) = \alpha$. Med $\alpha = 0.05$ är $t_\alpha = 1.8331$ och vi förkastar H_0 till förmån för H_1 om vi observerar $t > 1.8331$. Här fås

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{0.97 - 0.9}{0.33015/\sqrt{10}} = 0.67048$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

b) Vi vill med $\mu_0 = 1.1$ testa hypotesen

$$H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu < \mu_0$$

på nivå $\alpha = 0.05$.

Vi förkastar H_0 för små värden på

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

som då H_0 är sann är ett utfall av en $t(n-1)$-fördelad stokastisk variabel T. Alltså, om H_0 är sann så är $P(T < t_{1-\alpha}) = \alpha$. Med $\alpha = 0.05$ är $t_{1-\alpha} = -1.8331$ och vi förkastar H_0 till förmån för H_1 om vi observerar $t < -1.8331$. Här fås

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{0.97 - 1.1}{0.33015/\sqrt{10}} = -1.2452$$

så H_0 förkastas ej till förmån för H_1 på nivå $\alpha = 0.05$.

13.16 Observationerna $x_1, \ldots, x_n$ och $y_1, \ldots, y_n$ sammanfattas av tabellen:

Person, i	1	2	3	4	5	6	7	8
Morgon, x_i	172	168	180	181	160	163	165	177
Kväll, y_i	172	167	177	179	159	161	166	175
Skillnad, $w_i = x_i - y_i$	0	1	3	2	1	2	−1	2

De uppmätta parvisa längdskillnaderna $w_1, \ldots, w_n$ modelleras som utfall av oberoende $\mathrm{N}(\mu, \sigma)$-fördelade stokastiska variabler $W_1, \ldots, W_n$ där μ mäter hur mycket längre i genomsnitt personerna är på morgonen än på kvällen.

Parametrarna μ och σ^2 skattas med

$$\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i = 1.25 \quad \text{respektive} \quad s_w^2 = \frac{1}{n-1}\sum_{i=1}^{n}(w_i - \overline{w})^2 = 1.6429$$

så σ skattas med $s_w = \sqrt{1.6429} = 1.2817$.

Vi vill testa

$$H_0 : \mu = 0 \quad \text{mot} \quad H_1 : \mu \neq 0$$

på nivå $\alpha = 0.05$.

Alternativ 1: Vi förkastar H_0 för stora värden på $|t|$ där

$$t = \frac{\overline{w} - 0}{s/\sqrt{n}}$$

som då H_0 är sann är ett utfall av en $t(n-1)$-fördelad stokastisk variabel T. Alltså, om H_0 är sann så är $P\left(|T| > t_{\alpha/2}\right) = \alpha$. Med $\alpha = 0.05$ är $t_{\alpha/2} = 2.3646$ och vi förkastar H_0 till förmån för H_1 om vi observerar $|t| > 2.3646$. Här fås

$$t = \frac{\overline{w} - 0}{s/\sqrt{n}} = \frac{1.25 - 0}{1.2817/\sqrt{8}} = 2.7584$$

så H_0 förkastas till förmån för H_1 på nivå $\alpha = 0.05$.

Alternativ 2: Med $n = 8$ observationer $w_1, \ldots, w_n$ och konfidensgrad $1-\alpha = 0.95$ fås ur $t(n-1) = t(7)$-tabell att $t_{\alpha/2} = t_{0.025} = 2.3646$, så konfidensintervallet för den genomsnittliga längdskillnaden blir

$$\mu \in \overline{w} \pm t_{\alpha/2}\frac{s_w}{\sqrt{n}} = 1.25 \pm 1.072 = [0.17844,\ 2.3216] \text{ cm} \qquad (95\%).$$

Eftersom $\mu = 0$ inte är ett rimligt värde enligt konfidensintervallet förkastas hypotesen $H_0 : \mu = 0$ till förmån för $H_1 : \mu \neq 0$ på nivå 5%.

13.17 Låt $x_1, \ldots, x_n$ vara mätningarna av molekylvikt A och $y_1, \ldots, y_m$ motsvarande för B. Dessa värden modelleras som utfall av $\mathrm{N}(\mu_1, \sigma)$-fördelade stokastiska variabler $X_1, \ldots, X_n$ respektive $\mathrm{N}(\mu_2, \sigma)$-fördelade stokastiska variabler $Y_1, \ldots, Y_m$. Alla stokastiska variabler antas vara oberoende.

Vi vill testa:

$$H_0 : \mu_1 = \mu_2 \quad \text{mot} \quad H_1 : \mu_1 \neq \mu_2.$$

Dessa hypoteser formuleras som

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{mot} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Vi förkastar H_0 till förmån för H_1 för stora värden på $|t|$ där

$$t = \frac{\overline{x} - \overline{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

och s är skattningen av σ som ges av

$$s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{(n-1) + (m-1)} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2 + \sum_{i=1}^{m}(y_i - \overline{y})^2}{n + m - 2}.$$

Om H_0 är sann så är t ett utfall på en $t(n+m-2)$-fördelad stokastisk variabel T. Ur $t(6+8-2) = t(12)$-tabell fås att $P\left(|T| > t_{0.025}\right) = 0.05$ för $t_{0.025} = 2.1788$. Alltså, om H_0 förkastas då $|t| > 2.1788$ så har testet signifikansnivå 5%.

b) Med observationer

$x_1, \ldots, x_6$	174.18	174.30	174.23	174.29	174.36	174.25		
$y_1, \ldots, y_8$	174.19	174.40	174.20	174.35	174.32	174.14	174.27	174.34

fås

$$\overline{x} = 174.27,\ s_x = 0.062423 \qquad \overline{y} = 174.28,\ s_y = 0.091486$$

vilket ger

$$s = 0.080659 \quad \text{och} \quad t = \frac{\overline{x} - \overline{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{174.27 - 174.28}{0.080659\sqrt{\frac{1}{6} + \frac{1}{8}}} = -0.18174.$$

Eftersom $|t| < 2.1788$ kan inte H_0 förkastas på nivå 5%.

13.18

13.19

13.20 Låt $x_1, \ldots, x_n$ vara de uppmätta alkoholhalterna i flaskorna. Vi modellerar $x_1, \ldots, x_n$ som utfall av oberoende N(μ, σ)-fördelade stokastiska variabler $X_1, \ldots, X_n$, där $\sigma = 0.10$. Vi vill testa:

$$H_0 : \mu = 3.0\% = \mu_0 \quad \text{mot} \quad H_1 : \mu < \mu_0$$

på nivå $\alpha = 0.01$. Signifikansnivån begränsar sannolikheten för att förkasta en korrekt nollhypotes, det vill säga att dra slutsatsen att en flaska har liten alkoholhalt fast den egentligen är hög.

Vi förkastar H_0 då för små värden på

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

som om H_0 är sann är ett utfall på en $N(0,1)$-fördelad stokastisk variabel Z.

Alternativ 1: Ur $N(0,1)$-tabell fås att $\lambda_{0.01} = 2.3263$ och $P(Z < -2.3263) = 0.01$. Vi observerar utfallet

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.98 - 3.0}{0.10/\sqrt{10}} = -0.632 > -\lambda_{0.01}.$$

Vi förkastar inte H_0 på nivå 1%.

Alternativ 2: Vi observerar utfallet

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.98 - 3.0}{0.10/\sqrt{10}} = -0.632.$$

Vi har att då H_0 är sann så är

$$P(\text{Förkasta } H_0) = P(Z < -0.632) = \Phi(-0.632) = 1 - \Phi(0.632) = 0.26354$$

vilket är den lägsta signifikansnivå H_0 kan förkastas på. Vi kan inte förkasta H_0 på nivå 1%.

13.21 Låt $x_1, \ldots, x_n$ och $y_1, \ldots, y_n$ vara de slumpmässiga stickproven av de 10 + 10 oberoende N(μ_1, σ_1)- respektive N(μ_2, σ_2)-fördelade stokastiska variablerna $X_1, \ldots, X_n$ och $Y_1, \ldots, Y_n$, där $\sigma_1 = 0.3$ och $\sigma_2 = 0.4$.

Vi vill pröva hypotesen

$$H_0 : \mu_1 = \mu_2 \quad \text{mot} \quad H_1 : \mu_1 \neq \mu_2$$

på nivå $\alpha = 0.01$.

a) Vi förkastar H_0 för stora värden på $|z|$ där

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}$$

som om H_0 är sann är ett utfall av en N$(0,1)$-fördelad stokastisk variabel Z. På nivå $\alpha = 0.01$ förkastas H_0 om $|z| > \lambda_{0.01/2} = 2.5758$.

Om $\mu_1 - \mu_2 = 0.6$ är sannolikheten för att H_0 förkastas

$$\begin{aligned}
P(|Z| > \lambda_{0.005}) &= 1 - P(-\lambda_{0.005} \leq Z \leq \lambda_{0.005}) = 1 - P\left(-\lambda_{0.005} \leq \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \lambda_{0.005}\right) \\
&= 1 - P\left(-\lambda_{0.005} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \lambda_{0.005} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}\right) \\
&= 1 - P\left(-\lambda_{0.005} - 3.7947 \leq \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \leq \lambda_{0.005} - 3.7947\right) \\
&= 1 - (\Phi(-1.2189) - \Phi(-6.3706)) = 0.8886.
\end{aligned}$$

b) Ekvationen

$$1 - P\left(-\lambda_{0.005} - \frac{\mu_1-\mu_2}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}} \le \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}} \le \lambda_{0.005} - \frac{\mu_1-\mu_2}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}}\right) = 0.99$$

ger approximativt att

$$P\left(\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}} \le \lambda_{0.005} - \frac{\mu_1-\mu_2}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}}\right) \approx 0.01$$

det vill säga

$$\lambda_{0.005} - \frac{\mu_1-\mu_2}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}} \approx -\lambda_{0.01}$$

eller

$$\sqrt{n} \approx \frac{(\lambda_{0.01}+\lambda_{0.005})\sqrt{\sigma_1^2+\sigma_2^2}}{\mu_1-\mu_2} = \frac{(2.3263+2.5758)\sqrt{0.3^2+0.4^2}}{0.6} = 4.0851$$

eller $n \approx 16.68$. Alltså, minst 17 observationer krävs för att erhålla önskvärd styrka. (Styrkan är då 0.99115).

13.22

13.23

13.24

13.25 Låt $x_1, \ldots, x_n$ vara observationer på oberoende Po(μ)-fördelade stokastiska variabler $X_1, \ldots, X_n$.

Vi vill testa

$$H_0 : \mu = 0.2 = \mu_0 \quad \text{mot} \quad H_1 : \mu > \mu_0.$$

Vi förkastar H_0 för stora värden på $y = \sum_{i=1}^n x_i$ som om H_0 är sann är ett utfall på en Po$(n\mu_0)$ = Po(10)-fördelad stokastisk variabel Y. Vi observerar utfallet $y = 19$. Om H_0 är sann är

$$P\left(\text{Förkasta } H_0\right) = P\left(Y \ge 19\right) = 1 - P\left(Y \le 18\right) = 1 - \sum_{k=0}^{18} \frac{10^k}{k!}\mathrm{e}^{-10} = 1 - 0.99281 = 0.00719.$$

Detta är den lägsta signifikansnivå som H_0 kan förkastas på och vi ser att vi kan förkasta H_0 på nivå 1%.

13.26

13.27

13.28 Låt $x_1, \ldots, x_r$ vara antalet marsvinsungar med färg motsvarande kategori $1, \ldots, r$. Dessa modelleras som utfall av de (beroende) Bin(n, p_i)-fördelade stokastiska variablerna $X_1, \ldots, X_r$.

Enligt den genetiska modellen är p_i, sannolikheten att en unge får färg i, 9/16, 3/16 och 4/16 för $i = 1, 2, 3$. Formellt, vi vill testa

$$H_0 : p_1 = \frac{9}{16},\ p_2 = \frac{3}{16},\ p_3 = \frac{4}{16}$$

på nivå $\alpha = 0.05$. Enligt modellen så är

	Kategori (färg)			
	1 (röd)	2 (svart)	3 (vit)	
Observerat antal: x_i	43	10	34	87
Hypotes: p_i	9/16	3/16	4/16	1
Förväntat antal: np_i	48.938	16.312	21.750	87

Vi förkastar H_0 för stora värden på

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i}$$

som om H_0 är sann är ett utfall på en (approximativt) $\chi^2(r-1)$-fördelad stokastisk variabel. Ur $\chi^2(3-1) = \chi^2(2)$-tabeller får man att $\chi^2_{0.05} = 5.99$. Med utfallet $q = 10.063 > \chi^2_{0.05}$ så förkastas H_0 på nivå 5%. (Vi förkastar även på 1%, testets p-värde är 0.65%.)

13.29 Låt $x_1, \ldots, x_r$ vara antalet utlånade böcker under de olika veckodagarna, med n som det totala antalet utlånade böcker under veckan,

$$n = \sum_{i=1}^{r} x_i = 623.$$

Notera att n egentligen borde modelleras som ett utfall av en stokastisk variabel, men vi kommer att räkna som om n vore fix. Då kan $x_1, \ldots, x_r$ modelleras som utfall av de (beroende) $\text{Bin}(n, p_i)$-fördelade stokastiska variablerna $X_1, \ldots, X_r$.

Vi vill testa en hypotes om att utlåningen är densamma oavsett veckodag mot att den varierar mellan veckodagar. Formellt: vi vill testa

$$H_0 : p_1 = p_2 = \cdots = p_r = \frac{1}{r} \quad \text{mot} \quad H_1 : \text{inte } H_0$$

på signifikansnivå $\alpha = 0.05$. Enligt H_0 så är

	Kategori (veckodag)					
	1 (mån)	2 (tis)	3 (ons)	4 (tor)	5 (fre)	
Observerat antal: x_i	135	108	120	114	146	623
Hypotes: p_i	1/5	1/5	1/5	1/5	1/5	1
Förväntat antal: np_i	124.6	124.6	124.6	124.6	124.6	623

Vi förkastar H_0 för stora värden på

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i}$$

som om H_0 är sann är ett utfall på en (approximativt) $\chi^2(r-1)$-fördelad stokastisk variabel. Ur $\chi^2(5-1) = \chi^2(4)$-tabeller får man att $\chi^2_{0.05} = 9.49$. Med utfallet $q = 7.8266 < \chi^2_{0.05}$ så förkastas inte H_0 på nivå 5%.

På nivå $\alpha = 0.10$ får man ur $\chi^2(4)$-tabeller att $\chi^2_{0.10} = 7.78 < q$. Alltså förkastar vi H_0 på nivå 10%. (Testets p-värde är 9.81%.)

13.30 En stokastisk variabel X kan anta värdena $S_X = \{0, 1, 2, 3\}$. Vi vill testa hypotesen

$$H_0 : X \text{ är } \text{Bin}(3, \frac{1}{4}) \qquad \text{mot} \qquad H_1 : X \text{ är ej } \text{Bin}(3, \frac{1}{4})$$

på nivå $\alpha = 0.01$ med hjälp av $n = 4096$ oberoende observationer av X. Med de möjliga värdena på X som kategorier är enligt H_0 sannolikheten att få en observation i kategori $i \in S_X$

$$p_i = P(X = i) = \binom{3}{i} p^i (1-p)^{3-i} = \binom{3}{i} \left(\frac{1}{4}\right)^i \left(1 - \frac{1}{4}\right)^{3-i}, \; i = 0, \ldots, 3.$$

Låt x_i vara antalet utfall av X i kategori i, $i \in S_X$. Om H_0 är sann så är

i	0	1	2	3	
observerad frekvens, x_i	1764	1692	552	88	4096
Hypotes, p_i	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$	1
Förväntat antal, np_i	1728	1728	576	64	4096

Vi förkastar H_0 för stora värden på

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i}$$

som om H_0 är sann är ett utfall på en (approximativt) $\chi^2(r-1)$-fördelad stokastisk variabel. Ur $\chi^2(4-1) = \chi^2(3)$-tabeller får man att $\chi^2_{0.01} = 11.35$. Med utfallet $q = 11.5 > \chi^2_{0.01}$ så förkastas H_0 på nivå 1%.

13.31 Låt x_{ij} vara antalet observationer i kategori j, $j = 1, \ldots, r$, i population i, $i = 1, \ldots, s$. Vi modellerar x_{ij} som ett utfall av en $\text{Bin}(n_i, p_j^{(i)})$-fördelad stokastisk variabel X_{ij}.

x_{ij}	Kategori (kön)		
	1 (män)	2 (kvinnor)	
Population 1:	46	54	$100 = n_1$
Population 2:	78	72	$150 = n_2$
Population 3:	143	107	$250 = n_3$
Totalt:	$m_1 = 267$	$m_2 = 233$	$N = 500$

Vi vill testa om fördelningen av män och kvinnor skiljer sig mellan populationerna. Med H_0 som hypotesen att det inte finns någon skillnad,

$$H_0 : p_j^{(1)} = p_j^{(2)} = \cdots = p_j^{(s)} \quad \text{för alla } j = 1, \ldots, r,$$

så skattas p_j, sannolikheten att en utvald person har kön motsvarande kategori j, med

$$(p_j)^*_{\text{obs}} = \frac{1}{N} \sum_{i=1}^{s} x_{ij} = \frac{m_j}{N} \qquad (p_1)^*_{\text{obs}} = \frac{267}{500},\ (p_2)^*_{\text{obs}} = \frac{233}{500}.$$

Vi kan då skatta det förväntade antalet observationer i kategori j i serie i, $E(X_{ij}) = n_i p_j$, med $n_i (p_j)^*_{\text{obs}} = \frac{n_i m_j}{N}$.

$\frac{n_i m_j}{N}$	Kategori (kön)		
	1 (män)	2 (kvinnor)	
Population 1:	53.4	46.6	100
Population 2:	80.1	69.9	150
Population 3:	133.5	116.5	250
Totalt:	267	233	500

Vi jämför de observerade antalen x_{ij} med de skattade förväntade antalen $\frac{n_i m_j}{N}$ med

$$q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{\left(x_{ij} - \frac{n_i m_j}{N}\right)^2}{\frac{n_i m_j}{N}}$$

som om H_0 är sann är ett utfall av en (approximativt) $\chi^2((s-1)(r-1))$-fördelad stokastisk variabel. Vi förkastar H_0 för stora värden på q. Ur $\chi^2((3-1)(2-1)) = \chi^2(2)$-tabeller får man att $\chi^2_{0.10} = 4.61$. Vi observerar utfallet $q = 3.7694 < \chi^2_{0.10}$ så vi förkastar inte H_0 på nivå 10%. Det är på nivå 10% ingen signifikant skillnad i könsfördelning mellan populationerna.

13.32 De $n = 100$ observationerna på den förmodat ffg-fördelade stokastiska variabeln delas in i 9 kategorier:

	Kategori (värde)									
	1	2	3	4	5	6	7	8	≥ 9	
Observerad frekvens: x_i	42	23	10	11	8	2	3	1	0	$100 = n$

Notera att den sista kategorin är bara implicit given i uppgiften.

Låt p_i vara sannolikheten att en observation hamnar i kategori i enligt statistikerns modell. Om ffg(1/2)-antagandet stämmer är $p_i = (1 - 1/2)^(1/2)^{i-1} = 2^{-i}$, $i < 9$, och $p_9 = 2^{-8}$, och det förväntade antalet observationer i kategori i, $E(X_i) = np_i$.

	Kategori (värde)									
	1	2	3	4	5	6	7	8	≥ 9	
Observerad frekvens: x_i	42	23	10	11	8	2	3	1	0	100
Hypotes: p_i	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}	2^{-8}	1
Förväntat antal: np_i	50	25	12.5	6.25	3.125	1.5625	0.78125	0.39062	0.39062	100

Kategorier med ett lågt förväntat antal, $np_i < 5$, slås samman för att öka det förväntade antalet. Här gör vi sammanslagningar till $r = 5$ kategorier:

	Kategori (värde)					
	1	2	3	4	≥ 5	
Observerad frekvens: x_i	42	23	10	11	14	100
Hypotes: p_i	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-4}	1
Förväntat antal: np_i	50	25	12.5	6.25	6.25	100

Vi förkastar hypotesen om att observationerna kommer från en ffg(1/2)-fördelning för stora värden på

$$q = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i}$$

som då hypotesen stämmer är ett utfall från en (approximativt) $\chi^2(r-1)$-fördelad stokastisk variabel. Ur $\chi^2(5-1) = \chi^2(4)$-tabeller får man att $\chi^2_{0.05} = 9.4877$. Vi observerar utfallet $q = 15.16 > \chi^2_{0.05}$ så vi förkastar hypotesen på nivå 5%.

På nivå 1% fås $\chi^2_{0.01} = 13.277$ så hypotesen förkastas även på nivå 1%. (Testets p-värde är 0.438%.)

13.33 De $n = 300$ observationerna på den förmodat Poisson(μ)-fördelade stokastiska variabeln delas in i 4 kategorier beroende på dess värde:

	Kategori (värde)				
	0	1	2	≥ 3	
Observerad frekvens: x_i	249	42	9	0	300

Låt p_i vara sannolikheten att en observation hamnar i kategori i. Om Poisson(μ)-antagandet stämmer skattas μ med

$$\mu^*_{\text{obs}} = 0 \cdot \frac{249}{300} + 1 \cdot \frac{42}{300} + 2 \cdot \frac{9}{300} = 0.2$$

Med denna skattning erhålles skattningarna

$$(p_0)^*_{\text{obs}} = \frac{(\mu^*_{\text{obs}})^0}{0!} \mathrm{e}^{-\mu^*_{\text{obs}}} \quad (p_1)^*_{\text{obs}} = \frac{(\mu^*_{\text{obs}})^1}{1!} \mathrm{e}^{-\mu^*_{\text{obs}}} \quad (p_2)^*_{\text{obs}} = \frac{(\mu^*_{\text{obs}})^2}{2!} \mathrm{e}^{-\mu^*_{\text{obs}}} \quad (p_3)^*_{\text{obs}} = 1 - \sum_{i=0}^{2} (p_i)^*_{\text{obs}}.$$

De (skattade) förväntade antalet observationer i kategori i är $n(p_i)^*_{\text{obs}}$.

	Kategori (värde)				
	0	1	2	≥ 3	
Observerad frekvens: x_i	249	42	9	0	$300 = n$
Hypotes: $(p_i)^*_{\text{obs}}$	0.81873	0.16375	0.016375	0.0011485	1
Förväntat antal: $n(p_i)^*_{\text{obs}}$	245.62	49.124	4.9124	0.34454	300

Kategorier med ett lågt (skattat) förväntat antal, $n(p_i)^*_{\text{obs}}$, slås samman för att öka det förväntade antalet. Här gör vi sammanslagningar till $r = 3$ kategorier:

	Kategori (värde)			
	0	1	≥ 2	
Observerad frekvens: x_i	249	42	9	$300 = n$
Hypotes: $(p_i)^*_{\text{obs}}$	0.81873	0.16375	0.017523	1
Förväntat antal: $n(p_i)^*_{\text{obs}}$	245.62	49.124	5.2569	300

Vi förkastar en hypotes om Poisson(μ)-fördelning för stora värden på

$$q = \sum_{i=1}^{r} \frac{(x_i - n(p_i)^*_{\text{obs}})^2}{n(p_i)^*_{\text{obs}}}$$

som då hypotesen stämmer är ett utfall av en (approximativt) $\chi^2(r-1-1)$-fördelad stokastisk variabel. En extra frihetsgrad har försvunnit då parametern μ ersattes med skattningen μ^*_{obs}. Ur $\chi^2(3-1-1) = \chi^2(1)$-tabeller får man att $\chi^2_{0.05} = 3.8415$. Vi observerar utfallet $q = 3.7448 < 3.8415$ så vi förkastar inte hypotesen på nivå 5%. Vi kan inte utesluta att observationerna kommer från en Poissonfördelning.

På nivå 10% fås $\chi^2_{0.10} = 2.7055 < q$ så hypotesen förkastas på nivå 10%. (Testets p-värde är 5.30%.)

13.34

13.35 Låt x_{ij} vara antalet observationer i kategori (i, j), $j = 1, \ldots, r$, $i = 1, \ldots, s$. Beteckna det totala antalet skadade med $N = \sum_{i=1}^{s} \sum_{j=1}^{r} x_{ij} = 500$. Notera att N egentligen borde modelleras som ett utfall av en stokastisk variabel, men vi kommer att räkna som om N vore fix. Då kan x_{ij} modelleras som utfall av en $\text{Bin}(N, p_{ij})$-fördelad stokastisk variabel X_{ij}.

x_{ij}		Säkerhetsbälte		
		1 (användes)	2 (användes ej)	
Personskador	1 (lätta)	101	143	$n_1 = 244$
	2 (svåra)	58	198	$n_2 = 256$
		$m_1 = 159$	$m_2 = 341$	$N = 500$

Vi skattar fördelningen för personskador med

$$(p_i)^*_{\text{obs}} = \frac{n_i}{N} \quad (p_1)^*_{\text{obs}} = \frac{244}{500},\ (p_2)^*_{\text{obs}} = \frac{256}{500}.$$

Vi skattar fördelningen för användandet av säkerhetsbälte med

$$(q_i)^*_{\text{obs}} = \frac{m_i}{N} \quad (q_1)^*_{\text{obs}} = \frac{159}{500},\ (q_2)^*_{\text{obs}} = \frac{341}{500}.$$

Om graden av personskada är oberoende av användandet av säkerhetsbälte, $p_{ij} = p_i q_j$, kan p_{ij} skattas med $(p_i)^*_{\text{obs}} \cdot (q_j)^*_{\text{obs}}$ och det förväntade antalet observationer med

$$N\,(p_i)^*_{\text{obs}} \cdot (q_j)^*_{\text{obs}} = \frac{n_i\, m_j}{N}.$$

$\frac{n_i m_j}{N}$		Säkerhetsbälte		
		1 (användes)	2 (användes ej)	
Personskador	1 (lätta)	77.592	166.408	244
	2 (svåra)	81.408	174.592	256
		159	341	500

Vi förkastar en hypotes om oberoende mellan grad av personskada och användande av säkerhetsbälte för stora värden på

$$q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{\left(x_{ij} - \frac{n_i m_j}{N}\right)^2}{\frac{n_i m_j}{N}}$$

som om hypotesen stämmer är ett utfall av en (approximativt) $\chi^2((s-1)(r-1))$-fördelad stokastisk variabel. Ur $\chi^2((2-1)(2-1)) = \chi^2(1)$-tabeller får man att $\chi^2_{0.01} = 6.63$. Vi observerar utfallet $q = 20.2235 > \chi^2_{0.01}$ så vi förkastar hypotesen om oberoende på nivå 1%. Användandet av säkerhetsbälte och graden av personskada är inte oberoende. (Testets p-värde är $\approx 6.9 \cdot 10^{-6}$.)

14.1 Låt $y_1, \ldots, y_n$ vara de uppmätta ljusextinktionerna vid koncentrationerna $x_1, \ldots, x_n$. Vi ansätter modellen att $y_1, \ldots, y_n$ är utfall av oberoende $\text{N}(\alpha + \beta x_i, \sigma)$-fördelade stokastiska variabler $Y_1, \ldots, Y_n$.

Koncentration, $x_1, \ldots, x_n$	0.40	0.70	1.00	1.20	1.40	1.70	2.00
Extinktion, $y_1, \ldots, y_n$	0.23	0.34	0.42	0.55	0.61	0.77	0.84.

Observationerna sammanfattas av storheterna

$$\bar{x} = 1.20, \quad \bar{y} = 0.537$$

$$S_{xx}=\sum_{i=1}^{n}(x_i-\overline{x})^2=1.86,\ S_{xy}=\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=0.741,\ S_{yy}=\sum_{i=1}^{n}(y_i-\overline{y})^2=0.298.$$

Parametrarna β och α skattas med

$$\beta^*_{\text{obs}}=\frac{S_{xy}}{S_{xx}}=0.398 \quad \text{respektive} \quad \alpha^*_{\text{obs}}=\overline{y}-\beta^*_{\text{obs}}\overline{x}=0.0591$$

Variansen σ^2 skattas med

$$s^2=\frac{1}{n-2}\left[S_{yy}-\beta^*_{\text{obs}}S_{xy}\right]=0.0013888,$$

det vill säga σ skattas med $s=0.025052$.

a) Per enhet ökar extinktionen med faktorn β som skattas med $\beta^*_{\text{obs}}=0.398$.

b) En skattning av $E\left(Y(1.20)\right)=\alpha+\beta\cdot 1.20$ är $\alpha^*_{\text{obs}}+\beta^*_{\text{obs}}\cdot 1.20=0.53714$. Skattningen $\alpha^*_{\text{obs}}+\beta^*_{\text{obs}}\cdot 1.20$ är ett utfall av en stokastisk variabel med varians

$$\begin{aligned}V\left(\alpha^*+\beta^*\cdot 1.2\right) &= V\left(\overline{Y}-\beta^*\overline{x}+\beta^*\cdot 1.2\right)=V\left(\overline{Y}+\beta^*(1.2-\overline{x})\right)=\{\text{oberoende}\}\\ &= V\left(\overline{Y}\right)+(1.2-\overline{x})^2V\left(\beta^*\right)=\frac{\sigma^2}{n}+(1.2-\overline{x})^2\frac{\sigma^2}{S_{xx}}.\end{aligned}$$

Alltså, standardavvikelsen är

$$D\left(\alpha^*+\beta^*\cdot 1.2\right)=\sigma\sqrt{\frac{1}{n}+\frac{(1.2-\overline{x})^2}{S_{xx}}}=\sigma\sqrt{\frac{1}{7}+\frac{(1.2-1.2)^2}{S_{xx}}}=\sigma/\sqrt{7}.$$

c) En skattning av $E\left(Y(0)\right)=\alpha$ är $\alpha^*_{\text{obs}}=0.0591$. Skattningen α^*_{obs} är ett utfall av en stokastisk variabel med varians

$$V\left(\alpha^*\right) = V\left(\overline{Y}-\beta^*\overline{x}\right)=\{\text{oberoende}\}=V\left(\overline{Y}\right)+(-\overline{x})^2V\left(\beta^*\right)=\frac{\sigma^2}{n}+\overline{x}^2\frac{\sigma^2}{S_{xx}}.$$

Alltså, standardavvikelsen är

$$D\left(\alpha^*\right)=\sigma\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{S_{xx}}}=\sigma\sqrt{\frac{1}{7}+\frac{(1.2)^2}{1.86}}=0.9576\sigma.$$

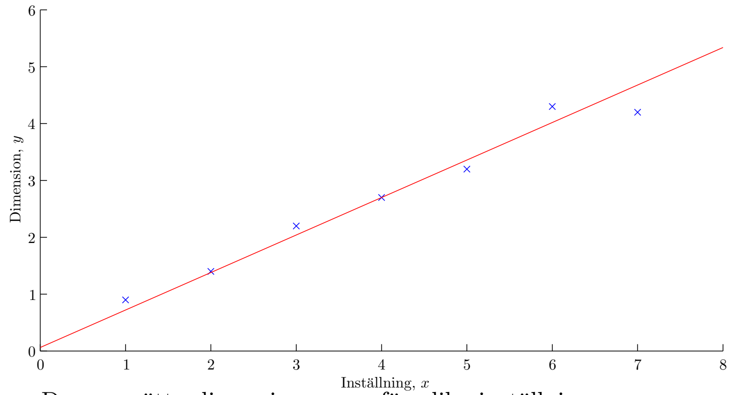
d) En skattning av σ är $s=0.025052$.

14.2

14.3

14.4 Låt $y_1,\ldots,y_n$ vara de uppmätta dimensionerna vid inställningarna $x_1,\ldots,x_n$. Vi ansätter modellen att $y_1,\ldots,y_n$ är utfall av oberoende N$(\alpha+\beta x_i,\sigma)$-fördelade stokastiska variabler $Y_1,\ldots,Y_n$.

Inställning, $x_1,\ldots,x_n$	1.0	2.0	3.0	4.0	5.0	6.0	7.0
Dimension, $y_1,\ldots,y_n$	0.9	1.4	2.2	2.7	3.2	4.3	4.2



De uppmätta dimensionerna y för olika inställningar x markerade med kryss med den skattade regressionslinjen inritad.

Observationerna sammanfattas av storheterna

$$\bar{x} = 4.0, \quad \bar{y} = 2.7$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 28, \; S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 16.7, \; S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 10.24.$$

b) Parametrarna β och α skattas med

$$\beta^*_{\text{obs}} = \frac{S_{xy}}{S_{xx}} = 0.59643 \quad \text{respektive} \quad \alpha^*_{\text{obs}} = \bar{y} - \beta^*_{\text{obs}}\bar{x} = 0.31429.$$

Variansen σ^2 skattas med

$$s^2 = \frac{1}{n-2}\left[S_{yy} - \beta^*_{\text{obs}} S_{xy}\right] = 0.055929,$$

det vill säga σ skattas med $s = 0.23649$.

c) Om man önskar att $2.5 = E\left(Y(x)\right) = \alpha + \beta x$ skall x väljas så att $x = (2.5 - \alpha)/\beta$. Denna storhet skattas med $(2.5 - \alpha^*_{\text{obs}})/\beta^*_{\text{obs}} = 3.6647$.

d) Skattningen α^*_{obs} är ett utfall av den normalfördelade stokastiska variabeln α^*, där

$$\alpha^* \quad \text{är} \quad \mathrm{N}\left(\alpha, \sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\right).$$

Ur $t(n-2) = t(7-2)$-tabell erhålles att $t_{0.025} = 2.57$. Alltså ges ett konfidensintervall för α av

$$\alpha \in \alpha^*_{\text{obs}} \pm t_{0.025}\, s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} = 0.31429 \pm 2.57 \cdot 0.19987 = 0.314 \pm 0.514 \quad (95\%).$$

Skattningen β^*_{obs} är ett utfall av den normalfördelade stokastiska variabeln β^*, där

$$\beta^* \quad \text{är} \quad \mathrm{N}\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right).$$

Alltså ges ett konfidensintervall för β av

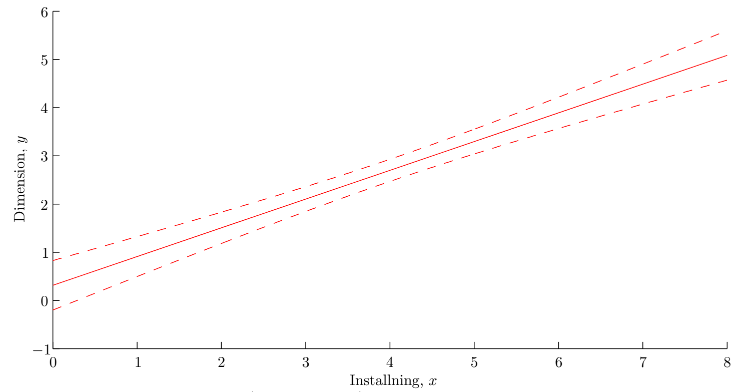
$$\beta \in \beta^*_{\text{obs}} \pm t_{0.025}\frac{s}{\sqrt{S_{xx}}} = 0.59643 \pm 2.57 \cdot 0.044693 = 0.596 \pm 0.115 \quad (95\%).$$

e) För en punkt x_0 skattas $E\left(Y(x_0)\right) = \alpha + \beta x_0$ av $\alpha^*_{\text{obs}} + \beta^*_{\text{obs}} x_0$ som är ett utfall av $\alpha^* + \beta^* x_0$, där

$$\alpha^* + \beta^* x_0 \quad \text{är} \quad \mathrm{N}\left(\alpha + \beta x_0, \sigma\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\right).$$

Alltså ges ett konfidensintervall för $\alpha + \beta x_0$ av

$$\alpha + \beta x_0 \in \alpha^*_{\text{obs}} + \beta^*_{\text{obs}} x_0 \pm t_{0.025} s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}.$$



Regressionslinjen (heldragen) med konfidensintervallet för väntevärdet $E\,(Y(x))$ som funktion av x (streckade linjer).

14.5 Vi gör mätningar i punkter x svarande mot $x = 1/\lambda^2$ och får följande observationer:

λ_i	6232	5571	5100
$x_i = 1/\lambda_i^2$	$2.5748 \cdot 10^{-8}$	$3.2221 \cdot 10^{-8}$	$3.8447 \cdot 10^{-8}$
y_i	19.38	25.62	30.10

Data kan sammanfattas med

$$\bar{x} = 3.2138 \cdot 10^{-8} \quad \bar{y} = 25.033$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = 8.0638 \cdot 10^{-17} \quad S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = 6.8137 \cdot 10^{-8} \quad S_{yy} = \sum (y_i - \bar{y})^2 = 57.975.$$

Vi ansätter modellen att $y_1, \ldots, y_n$ är utfall av oberoende $\mathrm{N}(\alpha + \beta x_i, \sigma)$-fördelade stokastiska variabler $Y_1, \ldots, Y_n$.

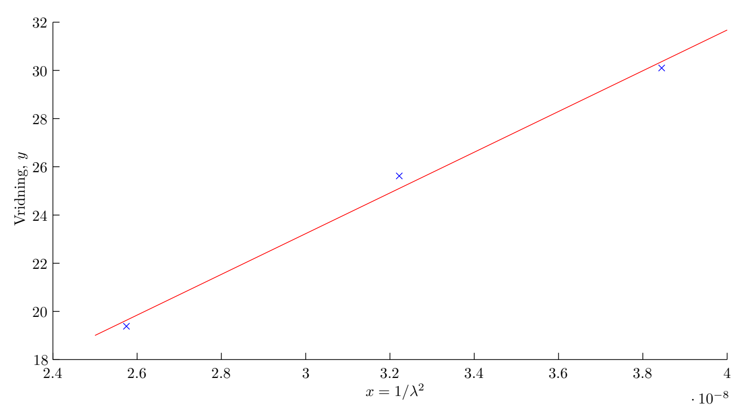
a) Parametrarna β och α skattas med

$$\beta^*_{\text{obs}} = \frac{S_{xy}}{S_{xx}} = 8.45 \cdot 10^8 \quad \text{respektive} \quad \alpha^*_{\text{obs}} = \bar{y} - \beta^*_{\text{obs}} \bar{x} = -2.123.$$

Variansen σ^2 skattas med

$$s^2 = \frac{1}{n-2} \left[S_{yy} - \beta^*_{\text{obs}} S_{xy} \right] = 0.40141,$$

det vill säga σ skattas med $s = 0.63357$.



De tre mätningarna på vridningen som funktion av $1/\lambda^2$, där λ är ljusets våglängd, med inritad regressionslinje.

c) Skattningen β^*_{obs} är ett utfall av den normalfördelade stokastiska variabeln β^*, där

$$\beta^* \quad \text{är} \quad \mathrm{N}\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right).$$

Ur $t(n-2) = t(3-2)$-tabell erhålles att $t_{0.025} = 12.706$ (!). Alltså ges ett konfidensintervall för β av

$$\beta \in \beta^*_{\text{obs}} \pm t_{0.025}\frac{s}{\sqrt{S_{xx}}} = 8.45 \cdot 10^8 \pm 12.706 \cdot 7.0555 \cdot 10^7 = (8.45 \pm 8.96) \cdot 10^8 \quad (95\%).$$

14.6

14.7 a) Om hypotesen $\beta_1 = 0$ förkastas till förmån för $\beta_1 \neq 0$ har testet p-värde $8.3 \cdot 10^{-6}$. Detta är den lägsta signifikansnivå hypotesen kan förkastas på och speciellt förkastas den på nivå 5%.

Om hypotesen $\beta_2 = 0$ förkastas till förmån för $\beta_2 \neq 0$ har testet p-värde 0.132. Detta är den lägsta signifikansnivå hypotesen kan förkastas på. Alltså, hypotesen förkastas ej på nivå 5%.

b) Förändringen $\beta_1 + \beta_2$ skattas med $(\beta_1)^*_{\text{obs}} + (\beta_2)^*_{\text{obs}} = 1.423 - 0.176 = 1.247$.

c) Skattningen $(\beta_1)^*_{\text{obs}} + (\beta_2)^*_{\text{obs}}$ är ett utfall av en stokastisk variabel $\beta_1^* + \beta_2^*$ med varians

$$V\left(\beta_1^* + \beta_2^*\right) = V\left(\beta_1^*\right) + V\left(\beta_2^*\right) + 2C\left(\beta_1^*, \beta_2^*\right),$$

det vill säga standardavvikelsen skattas till

$$d(\beta_1^* + \beta_2^*) = \sqrt{(0.247)^2 + (0.112)^2 + 2 \cdot (-0.00248)} = 0.2619.$$

14.8