Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1910/SF1925 TILLÄMPAD STATISTIK,
TORSDAG 9 JANUARI 2025 KL 8.00–13.00.

Examinator: Björn-Olof Skytt, 08-790 86 49

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik (utdelas vid tentamen), miniräknare.

Tentamen består av två delar, benämnda del I och del II.
Del I består av uppgifterna 1–12 och varje korrekt svar ger 1 poäng. På denna del ska endast svar anges, i form av val av ett av de möjliga svarsalternativen. Svaren ska anges på svarsblanketten (utdelas vid tentamen). Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1–3, utan får tillgodoräkna sig dessa tre uppgifter (i svarsblanketten anges ordet ”Bonus”). Studenter som är godkända på den andra datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges ordet ”Bonus”). Gränsen för godkänt är 9 poäng. Möjlighet att komplettera ges för tentander med 8 poäng.

Del II består av uppgifterna 13–16 och varje korrekt lösning ger 10 poäng. Del II rättas bara för studenter som är godkända på eller får komplettera del I och poäng på del II krävs för högre betyg än E. På denna del ska resonemang och uträkningar vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar ska förklaras och definieras samt numeriska svar ska anges med minst tre värdesiffrors noggrannhet. Studenter som är godkända på den andra datorlaborationen får 3 bonuspoäng på del II.

Tentamen kommer att vara rättad inom tre arbetsveckor från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Del I

Uppgift 1

Händelserna B och C är oberoende. Händelserna A och C är disjunkta. $\mathrm{P}(C) = 0.2$. $\mathrm{P}(A \mid B) = 0.6$.
Bestäm den betingade sannolikheten $\mathrm{P}(A \cup C \mid B)$.

A: 0.2

B: 0.4

C: 0.6

D: 0.8

Uppgift 2

Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion

$$f_X(x) = \begin{cases} 2x, & \text{för } 0 \leq x \leq 1, \\ 0, & \text{annars.} \end{cases}$$

och sätt $Y = X^2$. Bestäm variansen $V(Y)$.

A: 0.083

B: 0.29

C: 0.33

D: 0.58

Uppgift 3

De oberoende stokastiska variablerna X och Y är båda sådana att de är för-första-gången-fördelade med $p = \frac{1}{2}$, dvs. $X \in \text{ffg}(\frac{1}{2})$ och $Y \in \text{ffg}(\frac{1}{2})$. Låt $Z = \max(X, Y)$ (dvs. Z är det största värdet av X och Y). Bestäm $\mathrm{P}(Z > 2)$.

A: $\frac{1}{16}$

B: $\frac{7}{16}$

C: $\frac{9}{16}$

D: $\frac{15}{16}$

Uppgift 4

Låt X och Y vara stokastiska variabler så att $D(X) = 2$, $D(Y) = 2$ och $C(X, Y) = -2$. Bestäm $C(X + 2Y, 2X - Y)$.

A: -10

B: -6

C: 6

D: 10

Uppgift 5

En butik tar emot en leverans av 14 bakelser, varav 5 har hamnat på golvet vid uppackningen men ändå lagts fram till försäljning tillsammans med de 9 andra. Anta att du köper 4 bakelser. Vad är sannolikheten att exakt två av dem har varit på golvet vid uppackningen?

A: 0.3012

B: 0.3288

C: 0.3596

D: 0.3704

Uppgift 6

Låt X och Y vara oberoende normalfördelade stokastiska variabler där vi har väntevärdena $E(X) = -2$ och $E(Y) = 4$. Standardavvikelserna är $D(X) = D(Y) = 1$. Bestäm konstanten a så att $P(2X + Y < a) = 0.05$.

A: -4.38

B: -3.68

C: 3.68

D: 4.38

Uppgift 7

Antag att X_1 och X_2 är oberoende stokastiska variabler, sådana att $X_1 \in \text{Bin}(n, p)$ och $X_2 \in \text{Bin}(2n, p)$. Två skattningar av p har föreslagits: $p^*_{\text{obs}} = \frac{x_1+x_2}{3n}$ och $\hat{p}_{\text{obs}} = \frac{x_1}{2n} + \frac{x_2}{4n}$. Vilket av nedanstående påståenden är sant?

A: p^*_{obs} är den effektivaste skattningen av p.

B: $\hat{p}_{\text{obs}}$ är den effektivaste skattningen av p.

C: Bägge skattningarna är lika effektiva.

D: Man kan inte avgöra vilken av skattningarna som är effektivast, eftersom minst en av dem inte är väntevärdesriktig.

Uppgift 8

Antag att följande normalfördelade stickprov erhålls

$$-17.1 \quad 1.24 \quad 21.7 \quad 10.9.$$

Ange det nedåt begränsade ensidiga konfidensintervallet för väntevärdet med konfidensgrad 95%.

A: -34.5

B: -9.36

C: -0.652

D: -15.2

Uppgift 9

Givet följande oberoende observationer från en $\mathrm{Po}(\mu)$-fördelad stokastisk variabel

$$391 \quad 392 \quad 388,$$

beräkna den övre gränsen för det dubbelsidiga approximativa 99% konfidensintervallet för μ.

A: 416.9

B: 412.7

C: 441.2

D: 419.7

Uppgift 10

Antag att vi vill testa hypotesen $H_0 : \theta = 3.7$ mot alternativet $H_1 : \theta > 0$ med hjälp av ett stickprov med $n = 7$ observationer från en $\mathrm{N}(\theta, 2)$-fördelning. Hypotestestet är

Förkasta H_0 om $\overline{x} > 4.67$.

Vad är signifikansnivån?

A: 1%

B: 2.5%

C: 5%

D: 10%

Uppgift 11

Vi antar att $X \in \text{Bin}(18, p)$ och testar nollhypotesen $H_0 : p = 0.25$ mot mothypotesen $H_1 : p = 0.50$ och får utfallet $x = 9$.

A: Vi får p-värdet 0.019 och kan därför förkasta $H_0 : p = 0.25$ på risknivån 5%

B: Vi får p-värdet 0.019 och kan därför inte förkasta $H_0 : p = 0.25$ på risknivån 5%

C: Vi får p-värdet 0.30 och kan därför förkasta $H_0 : p = 0.25$ på risknivån 5%

D: Vi får p-värdet 0.30 och kan därför inte förkasta $H_0 : p = 0.25$ på risknivån 5%

Uppgift 12

Låt X och Y vara oberoende stokastiska variabler så att X har fördelningen $\text{Exp}(2\lambda)$ medan Y har fördelningen $\text{Exp}(3\lambda)$. Bestäm MK-skattningen för λ givet observationerna $x = 1.59$ och $y = 0.54$.

A: 0.939

B: 0.417

C: 0.466

D: 0.370

Del II

Uppgift 13

Låt X, Y vara likformigt fördelade över mängden

$$D = \{(x, y) \in \mathbb{R}^2 : |x| + |y| \leq 1\},$$

dvs. över rektangeln med hörn $(1, 0)$, $(0, 1)$, $(-1, 0)$ och $(0, -1)$. Är X och Y oberoende? Motivera ditt svar. (10 p)

Uppgift 14

Du pendlar till skolan antingen med bil eller cykel. Färdvägen är 2 kilometer och din medelhastighet med bilen och cykel är 40 km/h respektive 30 km/h. På vägen finns det dock 20 trafikljus placerade som alternerar mellan rött i 40 sekunder och grönt i 20 sekunder. Som bekant måste både bilar och cyklar stanna när trafikljuset visar rött.
Antag att du ankommer till varje trafikljus slumpmässigt under dess 60-sekunderscykel (dvs. det finns ingen "grön våg"). En dag tar du bilen medan en annan dag tar du cykeln till skolan. Vad är sannolikheten att du kommer fram snabbare den dagen du tar cykeln? Motivera eventuella approximationer och antaganden. (10 p)

Uppgift 15

Livslängder för två olika kullagertyper mättes och man erhöll följande data (i driftdagar):

Typ A	120	125	121	123	122
Typ B	200	203	201	207	205

Man antog att livslängderna var utfall av oberoende normalfördelade stokastiska variabler som är $\mathrm{N}(\mu_\mathrm{A}, \sigma)$ och $\mathrm{N}(\mu_\mathrm{B}, \sigma)$ för kullager av typ A respektive B.

(a) Bestäm ett konfidensintervall för $\mu_\mathrm{A} - \mu_\mathrm{B}$ med konfidensgrad 95%. (4 p)

(b) Eftersom kullager av typ B var dubbelt så dyra som kullager av typ A så ville man ha ett konfidensintervall för $2\mu_\mathrm{A} - \mu_\mathrm{B}$ med konfidensgrad 95%. Beräkna detta intervall. (6 p)

Uppgift 16

Kryptoanalysen studerar hur krypton kan forceras (dvs. ”knäckas”) utan tillgång till en kodnyckel. Detta innebär ofta att olika nycklar testas genom att dekryptera datan med en viss nyckel och analysera den dekrypterade texten. Om den dekrypterade texten ser rimlig ut kan man konstatera att nyckeln är korrekt.
Ett sätt att analysera den dekrypterade texten är att studera dess bokstavsfrekvens. I svenskan kan man bland annat konstatera att bokstäverna ’a’ (inklusive ’å’ och ’ä’), ’e’, ’n’, ’t’ och ’o’ (inklusive ’ö’) har följande frekvenser

Bokstav	’a’/’å’/’ä’	’e’	’n’	’t’	’o’/’ö’
Frekvens	12.52%	10.15%	8.54%	7.70%	5.79%

Givet att den krypterade texten har exakt 70 bokstäver (där endast tecknena ’a’–’z’ används), ta fram ett test som kan särskilja svenskspråkig text från annan text med 5% felrisk. Verifiera sedan ditt test på följande två dekrypterade textfragment:

```
korparenstorsvartkrakfagelsomaterfinnsoverstorredelenavnorrahalvklotet

pvknmzdnewemslamfsjtpkdldoqyxqgcmmguoisuiyojhlfpdysvceisndyjgrodypqlto
```

(10 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1910/SF1925 TILLÄMPAD STATISTIK, TORSDAG 9 JANUARI 2025 KL 8.00–13.00.

Del I, Svar.

1. D

2. A

3. B

4. B

5. C

6. B

7. A

8. D

9. D

10. D

11. A

12. D

Del I, Lösningsförslag.

Uppgift 1

$$\begin{aligned}
P(A \cup C|B) &= \frac{P([A \cup C] \cap B)}{P(B)} = [\text{A och C är disjunkta}] = \frac{P(A \cap B) + P(B \cap C)}{P(B)} \\
&= [\text{B och C är oberoende}] = \frac{P(A|B)P(B) + P(B)P(C)}{P(B)} \\
&= P(A|B) + P(C) = 0.6 + 0.2 = 0.8
\end{aligned}$$

Uppgift 2

Först noterar vi att $V(Y) = E(Y^2) - E^2(Y)$. Detta ger då

$$E(Y^2) = E(X^4) = \int_{-\infty}^{\infty} x^4 f_X(x)dx = \int_0^1 x^4 \cdot 2xdx = \left[\frac{x^6}{3}\right]_0^1 = \frac{1}{3}$$

$$E(Y) = E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^1 x^2 \cdot 2xdx = \left[\frac{x^4}{2}\right]_0^1 = \frac{1}{2}$$

$$\Rightarrow V(Y) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} = 0.083$$

Uppgift 3

$$\begin{aligned}
P(Z > 2) &= P(\max(X,Y) > 2) = 1 - P(\max(X,Y) \le 2) = 1 - P(X \le 2 \cap Y \le 2) \\
&= [\text{X och Y är oberoende}] = 1 - P(X \le 2)P(Y \le 2) = 1 - \left(\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}\right)^2 \\
&= 1 - \left(\frac{3}{4}\right)^2 = 1 - \frac{9}{16} = \frac{7}{16}
\end{aligned}$$

Uppgift 4

$$\begin{aligned}
C(X + 2Y, 2X - Y) &= C(X, 2X) + C(X, -Y) + C(2Y, 2X) + C(2Y, -Y) \\
&= 2C(X,X) - C(X,Y) + 4C(Y,X) - 2C(Y,Y) \\
&= 2V(X) + 3C(X,Y) - 2V(Y) = 2 \cdot 4 - 3 \cdot 2 - 2 \cdot 4 = -6
\end{aligned}$$

Uppgift 5

Här har vi hypergeometrisk fördelning eftersom vi här har situationen dragning utan återläggning utan hänsyn till ordning. Alltså blir sannolikheten för 2 bakelser som legat på golvet $\frac{\binom{5}{2}\binom{9}{2}}{\binom{14}{4}} = 0.3596$.

Uppgift 6

Låt $Z = 2X + Y$. Då är Z en linjärkombination av två oberoende normalfördelade stokastiska variabler och är således också normalfördelad.

$$E(Z) = E(2X + Y) = 2 \cdot (-2) + 4 = 0$$

$$V(Z) = V(2X + Y) = 4V(X) + V(Y) = 4 \cdot 1 + 1 = 5$$

$$\text{Alltså gäller att } Z \in N(0, \sqrt{5})$$

$$\text{Vi har nu } \mathrm{P}(Z < a) = 0.05\text{. Vi gör om till } N(0,1) \Rightarrow \mathrm{P}\left(\frac{Z}{\sqrt{5}} < \frac{a}{\sqrt{5}}\right) = 0.05$$

$$W = \frac{Z}{\sqrt{5}} \in N(0,1)\text{, och vi har } \mathrm{P}\left(W < \frac{a}{\sqrt{5}}\right) = 0.05$$

$$\text{Symmetri ger nu att } \mathrm{P}\left(W > -\frac{a}{\sqrt{5}}\right) = 0.05 \Rightarrow -\frac{a}{\sqrt{5}} = \lambda_{0.05} = 1.6449 \Rightarrow a = -3.68$$

.

Uppgift 7

$$E(p^*) = E\left(\frac{X_1 + X_2}{3n}\right) = \frac{np + 2np}{3n} = p$$

$$E(\hat{p}) = E\left(\frac{X_1}{2n} + \frac{X_2}{4n}\right) = \frac{np}{2n} + \frac{2np}{4n} = p$$

Båda skattningarna är således väntevärdesriktiga.

$$V(p^*) = V\left(\frac{X_1 + X_2}{3n}\right) = \frac{np(1-p) + 2np(1-p)}{9n^2} = \frac{p(1-p)}{3n}$$

$$V(\hat{p}) = V\left(\frac{X_1}{2n} + \frac{X_2}{4n}\right) = \frac{np(1-p)}{4n^2} + \frac{2np(1-p)}{16n^2} = \frac{3p(1-p)}{8n}$$

$$V(p^*) < V(\hat{p})\text{, så } p^*_{\text{obs}} \text{ är den effektivaste skattningen av } p.$$

Uppgift 8

För att bestämma konfidensintervallet tar vi först fram stickprovsmedelvärdet

$$\bar{x} = \frac{-17.1 + 1.24 + 21.7 + 10.9}{4} = 4.185$$

och stickprovsstandardavvikelsen

$$s \approx 16.47.$$

Vi får då

$$\bar{x} - t_{0.025}(3) \cdot \frac{s}{\sqrt{n}} = 4.185 - 2.353 \cdot \frac{16.47}{2} \approx -15.2$$

Svaret är D.

Uppgift 9

Från datat har vi

$$\mu^*_{\text{obs}} = \bar{x} \approx 390.33.$$

Standardavvikelsen för $\bar{X}$ ges av

$$D\left(\bar{X}\right) = \frac{D\left(X_1 + X_2 + X_3\right)}{3} = \frac{\sqrt{3}}{3} D\left(X_1\right) = \frac{\sqrt{\mu}}{\sqrt{3}} = \sqrt{\frac{\mu}{3}}$$

eftersom X_1, X_2, X_3 är oberoende och standardavvikelsen för X_1 är $\sqrt{\mu}$.
Vi får alltså medelfelet

$$d(\mu^*) = D\left(\bar{X}\right)\Big|_{\mu=\mu^*_{\text{obs}}} \approx 11.407.$$

Eftersom μ troligvist är stort så approximerar vi pivotvariabeln med en normalfördelning, vilket ger ett approximativt 99%-igt konfidensintervall på formen

$$390.33 \pm \lambda_{0.005} \cdot 11.407 \approx [360.95, 419.71].$$

Svaret är alltså D.

Uppgift 10

Under nollhypotesen är $\bar{X}$ fördelad enligt $\mathrm{N}(3.7, 2/\sqrt{7})$, dvs. sannolikheten att förkasta är

$$\begin{aligned}\mathrm{P}\left(\bar{X} > 4.67\right) = 1 - \mathrm{P}\left(\bar{X} \leq 4.67\right) &= 1 - \mathrm{P}\left(\frac{\bar{X} - 3.7}{2/\sqrt{7}} \leq \frac{4.67 - 3.7}{2/\sqrt{7}}\right) \\ &= 1 - \Phi\left(\frac{4.67 - 3.7}{2/\sqrt{7}}\right) \approx 1 - \Phi\left(1.28\right) \approx 0.1003.\end{aligned}$$

Med andra ord är svaret 10%, dvs. D.

Uppgift 11

P-värdet är sannolikheten att förkasta H_0 om H_0 är sann givet att man fått det utfall man har fått eller ett utfall ännu mer extremt. Alltså söker vi $\mathrm{P}\left(X \geq 9\right)$ där $X \in \mathrm{Bin}(18, 0.25)$. I tabell 6 ser man då att om $X \in \mathrm{Bin}(18, 0.25)$ så är $\mathrm{P}\left(X \leq 8\right) = 0.98065$.
Då är $\mathrm{P}\left(X \geq 9\right) = 1 - 0.98065 \approx 0.019$. Eftersom p-värdet är mindre än 0.05 så kan vi förkasta H_0 på risknivån 5%.

Uppgift 12

Eftersom väntevärdet för en exponentialfördelad stokastisk variabel med intensitet λ är lika med $1/\lambda$ får vi följande att $E[X] = \frac{1}{2\lambda}$ och $E[Y] = \frac{1}{3\lambda}$.
Vi bildar då kvadratsumman

$$Q(\lambda) = \left(x - \frac{1}{2\lambda}\right)^2 + \left(y - \frac{1}{3\lambda}\right)^2,$$

vars derivata ges av

$$\frac{dQ(\lambda)}{d\lambda} = 2\left(x - \frac{1}{2\lambda}\right)\frac{1}{2\lambda^2} + 2\left(y - \frac{1}{3\lambda}\right)\frac{1}{3\lambda^2}.$$

Sätter vi detta till noll och förenklar får vi

$$\left(x - \frac{1}{2\lambda}\right)\frac{1}{2} + \left(y - \frac{1}{3\lambda}\right)^2\frac{1}{3} = 0,$$

vilket har lösningen

$$\lambda = \frac{13}{18x + 12y}.$$

Om vi beräknar andraderivatan ser vi att detta är ett globalt minimum.
MK-skattningen för λ ges då av:

$$\lambda^* = \frac{13}{18 \cdot 1.59 + 12 \cdot 0.54} \approx 0.370.$$

Svaret är alltså D.

Del II, Lösningsförslag:

Uppgift 13

Eftersom D är en kvadrat med sida $\sqrt{2}$ har den area 2. Vi får alltså täthetsfunktionen

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2}, & |x|+|y| \leq 1, \\ 0, & \text{annars.} \end{cases}$$

Marginalfördelningen för X blir alltså

$$f_X(x) = \int_{y=-\infty}^{+\infty} f_{X,Y}(x,y)dy = \int_{-1+|x|}^{1-|x|} \frac{1}{2}dy = \frac{1}{2}(1-|x|-(-1+|x|)) = \frac{2-2|x|}{2} = 1-|x|,$$

för $|x| \leq 1$ och noll annars. På samma sätt har vi $f_Y(y) = 1-|y|$ för $|y| \leq 1$ och noll annars. Nu ser vi att X och Y inte är oberoende eftersom

$$f_{X,Y}(x,y) \neq f_X(x)f_Y(y) = (1-|x|)(1-|y|).$$

Bland annat ser vi att vänster- och högerled skiljer sig i punkten $(x,y) = (0,0)$ där vi får $\frac{1}{2} \neq 1$.

Uppgift 14

Utan trafikljus tar det 180 sekunder (3 minuter) för bilen och 240 sekunder (4 minuter) för cykeln. Tidpunkten X du anländer till ett trafikljus är fördelad enligt $\mathrm{U}(0,60)$ under dess cykel på 60 sekunder. Antag att den visar rött under de första 40 sekunderna och grönt under de resterande 20 sekunderna.
Tiden du måste vänta, Y, är då

$$Y = \begin{cases} 40-X, & X \leq 40, \\ 0, & \text{annars.} \end{cases}$$

Vi kan beräkna $E(Y)$ och $D(Y)$ genom

$$E(Y) = \int_0^{40} (40-x)\frac{1}{60}dx = 13.33$$

$$E\left(Y^2\right) = \int_0^{40} (40-x)^2\frac{1}{60}dx = 355.55$$

$$D(Y) = \sqrt{E\left(Y^2\right) - E(Y)^2} \approx 13.34$$

Den totala tiden du väntar vid ett trafikljus (oberoende om du är i bil eller på cykel) är då summan av 20 stycken oberoende kopior av Y. Vi antar att 20 är tillräckligt stort och approximerar därför summan då vi har oberoende och likafördelade termer. Den totala väntetiden är alltså approximativt fördelad enligt

$$\mathrm{N}(20 \cdot 13.33, \sqrt{20} \cdot 13.34) \approx \mathrm{N}(266.6, 59.65)$$

Skillnaden mellan de totala färdtiderna är nu appromativt fördelad enligt

$$\mathrm{N}(240+266.6-(180+266.6), \sqrt{2}59.65) \approx \mathrm{N}(60, 84.36).$$

Sannolikheten att denna är mindre än noll ges av

$$1 - \Phi(60/84.36) \approx 1 - \Phi(0.71) \approx 1 - 0.7611 = 0.2389.$$

Uppgift 15

(a) Två oberoende stickprov med $N(\mu_A, \sigma)$- respektive $N(\mu_B, \sigma)$-fördelade observationer. Ett 95%-igt konfidensintervall för $\mu_A - \mu_B$ blir med t-metoden (FS 12.2)

$$\bar{x} - \bar{y} \pm t_{0.025}(5+5-2)s\sqrt{\frac{1}{5} + \frac{1}{5}}$$

där $\bar{x} = (120 + 125 + \cdots + 122)/5 = \frac{611}{5} = 122.2$ och $\bar{y} = (200 + 203 + \cdots + 205)/5 = \frac{1016}{5} = 203.2$. Vidare får vi

$$s_x^2 = \frac{1}{5-1}\left(\sum_{i=1}^{5} x_i^2 - 5 \cdot (\bar{x})^2\right) = \frac{1}{4}\left(74679 - \frac{611^2}{5}\right) = \frac{14.8}{4} = 3.7$$

$$s_y^2 = \frac{1}{5-1}\left(\sum_{i=1}^{5} y_i^2 - 5 \cdot (\bar{y})^2\right) = \frac{1}{4}\left(206484 - \frac{1016^2}{5}\right) = \frac{32.8}{4} = 8.2$$

som ger

$$s = \sqrt{\frac{(5-1)s_x^2 + (5-1)s_y^2}{5+5-2}} = 2.4393$$

och vi får intervallet till $122.2 - 203.2 \pm 2.31 \cdot 2.4393\sqrt{2/5} = -81.0 \pm 3.6 = \underline{(-84.6, -77.4)}$.

(b) $2\mu_A - \mu_B$ skattas med $2\bar{x} - \bar{y}$. Vidare gäller att $V(2\bar{X} - \bar{Y}) = \frac{4\sigma^2}{5} + \frac{\sigma^2}{5} = \sigma^2$. Alltså $2\bar{X} - \bar{Y}$ är $N(2\mu_A - \mu_B, \sigma)$. σ okänd och skattas p.s.s. som ovan. Ett 95%-igt konfidensintervall för $2m_A - m_B$ blir med t-metoden (FS 12.2)

$$2\bar{x} - \bar{y} \pm t_{0.025}(5+5-2)s = 244.4 - 203.2 \pm 2.31 \cdot 2.4393 = 41.2 \pm 5.6 = \underline{(35.6, 46.8)}$$

Uppgift 16

Vi bildar ett χ^2-test där nollhypotesen är att bokstäverna har den givna fördelningen med $n = 70$ observationer. Problemet här är att den minsta gruppen ('o'/'ö') är för liten ($0.0579 \cdot 70 = 4.053 < 5$), så vi måste slå ihop den med en annan grupp, förslagsvis den näst minsta gruppen ('t'). Tillsammans med en grupp för övriga bokstäver får vi då tabellen

Bokstav	'a'/'å'/'ä'	'e'	'n'	't'/'o'/'ö'	andra
p_i	0.1252	0.1015	0.0854	0.1349	0.5530
np_i	8.764	7.105	5.978	9.443	38.71

Om vi låter x_1 vara antalet 'a', x_2 vara antalet 'e', x_3 vara antalet 'n', x_4 vara antalet 't' eller 'o' samt x_5 vara antalet andra bokstäver i texten kan vi bilda kvadratsumman

$$Q = \frac{x_1^2}{8.764} + \frac{x_2^2}{7.105} + \frac{x_3^2}{5.978} + \frac{x_4^2}{9.443} + \frac{x_5^2}{38.71} - 70$$

ock förkasta nollhypotesen om Q överstiger $\chi^2_{0.05}(5-1) = \chi^2_{0.05}(4) \approx 9.49$.
För de tre exempeltexterna får vi $Q = 1.87$, $Q = 12.5$ respektive $Q = 11.8$, så allt verkar stämma.