Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1900/SF1912 SANNOLIKHETSTEORI OCH STATISTIK,
ONSDAG 23 OKTOBER 2019 KL 8.00–13.00.

*Examinator:* Camilla Landén, 08-790 6197.

*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. På denna del skall endast svar anges, antingen i form av ett numeriskt värde med tre värdesiffrors noggrannhet eller i form av val av ett av de möjliga svarsalternativen. Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1-3, utan får tillgodoräkna sig dessa tre uppgifter. Studenter som är godkända på datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift. Detta gäller vid ordinarie tentamen och vid första omtentamen. Gränsen för godkänt är preliminärt 9 poäng. Möjlighet att komplettera ges för tentander med, preliminärt, 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 skall resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar skall förklaras och definieras och numeriska svar skall anges med minst två värdesiffrors noggrannhet. Studenter som är godkända på datorlaborationen får 3 bonuspoäng på del II vid ordinarie tentamenstillfället och det första omtentamenstillfället.

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

Låt $X$ och $Y$ vara stokastiska variabler, sådana att $V(X) = 4$, $V(Y) = 9$, och $C(X, Y) = 2$. Beräkna $V(2X - Y)$.

### Uppgift 2

I ett stort bostadsområde är $\frac{1}{6}$ av lägenheterna ettor, $\frac{1}{3}$ tvåor, $\frac{2}{5}$ treor och $\frac{1}{10}$ fyror. Under en 10-årsperiod anses risken för vattenskada för en etta vara 10% , för en tvåa 5%, för en trea 8 % och för en fyra 10%. Vad är då sannolikheten att en slumpmässigt vald lägenhet drabbas av en vattenskada under denna 10-årsperiod?

A: 6.08%

B: 7.53%

C: 8.36%

D: 9.02%

### Uppgift 3

En stokastisk variabel $X$ har täthetsfunktionen

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \dfrac{x}{10}, & 0 \leq x \leq 2 \\ \dfrac{2}{10}, & 2 < x \leq 6 \\ 0, & x > 6 \end{cases}$$

Bestäm $E(X^2)$.

A: 12.0

B: 12.8

C: 13.6

D: 14.3

### Uppgift 4

Låt $X$ och $Y$ vara oberoende stokastiska variabler sådana att $X \in Bin(2, 0.5)$ och $Y \in Hyp(8, 2, 0.5)$. Bestäm $P(X + Y > 1)$.

A: 0.7500

B: 0.2500

C: 0.3036

D: 0.6964

### Uppgift 5

För händelserna $A$, $B$ och $C$ är följande sannolikheter givna:

$$P(A^*|C) = 0.5, \qquad P(A \cup C) = 0.6, \qquad P(C) = 0.2$$

samt

$$P(A \cap B) = 0.25, \qquad P(A \cap B \cap C) = 0.05, \qquad P(A \cap C|B) = 0.125.$$

Vilket av följande påståenden är korrekt?

A: Händelsen $A$ är oberoende av både händelsen $B$ och händelsen $C$.

B: Händelsen $A$ är varken oberoende av händelsen $B$ eller händelsen $C$.

C: Händelsen $A$ är oberoende av händelsen $B$, men inte av händelsen $C$.

D: Händelsen $A$ är oberoende av händelsen $C$, men inte av händelsen $B$.

### Uppgift 6

Antag att $X$ och $Y$ är två oberoende normalfördelade stokastiska variabler med väntevärdena $E(X) = E(Y) = 0$ och standardavvikelserna $D(X) = D(Y) = 2$. Beräkna $P(|X - Y| < 3)$.

### Uppgift 7

Antag att vi har observationer $x_1, \ldots, x_n$ av oberoende stokastiska variabler $X_1, \ldots, X_n$ sådana att $E(X_i) = \mu$ och $V(X_i) = \sigma^2$, $i = 1, \ldots, n$. Skatta $\mu$ med

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

och $\sigma^2$ med

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

Vilken approximativ fördelning har stickprovsvariabeln $\mu^*$ då $n \geq 50$? Andra parametern avser standardavvikelsen i påståendena nedan.

A: $N(\bar{x}, s)$

B: $N(\bar{x}, s/\sqrt{n})$

C: $N(\mu, \sigma/\sqrt{n})$

D: $N(\mu, \sigma)$

## Uppgift 8

Följande konfidensintervall för skillnaden mellan två stickprovs väntevärden är givet:

$$I_{\mu_x-\mu_y} = (149.34, 172.66).$$

Antag att stickproven kommer från $X_i \in N(\mu_x, \sigma)$, respektive $Y_i \in N(\mu_y, \sigma)$ samt att alla dessa stokastiska variabler är oberoende. Eftersom $\sigma$ är okänd har man skattat $\sigma$. Man har haft 10 observationer i varje stickprov och ur dessa har man fått varsin skattning av $\sigma^2$: $s_x = 10.0$ och $s_y = 8.0$ Ange konfidensgraden för det tvåsidiga konfidensintervallet $I_{\mu_x-\mu_y}$ ovan.

A: 90%

B: 95 %

C: 99%

D: 99.9%

## Uppgift 9

Följande konfidensintervall med konfidensgrad 99% för skillnaden mellan två stickprovs väntevärden är givet

$$I_{\mu_x-\mu_y} = (0.12, 0.32).$$

Antag att stickproven kommer från $X_i \in N(\mu_x, \sigma_x)$, respektive $Y_i \in N(\mu_y, \sigma_y)$ samt att dessa alla stokastiska variabler är oberoende. Antag vidare att $\sigma_x$ och $\sigma_y$ är kända. Man önskar testa nollhypotesen $H_0 : \mu_x = \mu_y$ mot $H_1 : \mu_x \neq \mu_y$. Vilken slutsats kan man dra med hjälp av det givna konfidensintervallet?

A: $H_0$ kan varken förkastas på risknivån 1% eller risknivån 5%

B: $H_0$ kan både förkastas på risknivån 1% och risknivån 5%

C: $H_0$ kan förkastas på risknivån 1% , men inte på risknivån 5%

D: $H_0$ kan inte förkastas på risknivån 1% , men på risknivån 5% krävs ytterligare utredning.

## Uppgift 10

Vid ett homogenitetstest av 7 serier med 4 kategorier vardera finner man att Q=33.2.
Bestäm $P$-värdet (eller observerad signifikansnivå).

A: 1.58 %

B: 4.40 %

C: 9.99 %

D: 0.58 %

## Uppgift 11

En stokastisk variabel $X$ har täthetsfunktionen

$$f_X(x) = \begin{cases} \dfrac{3x^2}{\theta} e^{-x^3/\theta}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Vi har två oberoende observationer från ovanstående fördelning: $x_1 = 1.65$ och $x_2 = 6.13$.

Bestäm maximum-likelihood-skattningen av $\theta$ utifrån detta.

A: 117

B: 20.1

C: 3.89

D: 737

## Uppgift 12

Vid enkel linjär regression gäller följande samband

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

där man observerat paren $(x_i, y_i)$, $i = 1, \dots, n$ och $\varepsilon_i$ betecknar de slumpmässiga felen. Vilket av följande påståenden är INTE sant?

A: Felen antas vara normalfördelade.

B: Observationerna av förklaringsvaribeln, $x_i$, antas vara normalfördelade.

C: Observationerna av responsvariabeln, $y_i$, antas vara normalfördelade givet värdena på $x_i$.

D: Felen antas ha samma varians.

# Del II

### Uppgift 13

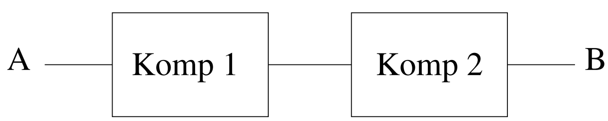
I september förekom en artikel i SvD där det uppgavs att det finns 661 adelsätter i Sverige och att en eller två adelsätter utslocknar per år. Därför kan adeln alltså sluta att existera om 300 år sägs det i artikeln. Låt oss för enkelhetens skull anta att antalet adelsätter som utslocknar under ett år är Poissonfördelat med väntevärdet 2. Vi antar också att variablerna för olika år är oberoende av varandra.

a) Beräkna sannolikheten för att alla adelsätter i Sverige ska ha utslocknat inom 300 år under förutsättning att antalet adelsätter som utslocknar under ett år är Poissonfördelat med väntevärdet 2. Lämpliga och välmotiverade approximationer är tillåtna. (3 p)

b) Vilket värde ska väntevärdet ovan ha om sannolikheten för att alla adelsätter i Sverige ska ha utslocknat inom 300 år ska vara 95%? Lämpliga och välmotiverade approximationer är tillåtna. (7 p)

### Uppgift 14

a) I ett system är två komponenter seriekopplade enligt figuren. Systemet fungerar om både komponent 1 och 2 fungerar.




Antag att livslängderna $T_1$ och $T_2$ för komponent 1, respektive 2 är obeoroende stokastiska variabler bägge med fördelningsfunktion $F(t) = t/100$ för $0 \leq t \leq 100$ och 0 för övrigt (komponenten får aldrig användas mer än 100 timmar av säkerhetsskäl). Beräkna väntevärde och varians för systemets livslängd. (5 p)

b) Antag att man i en fabrik har 57 hela system av den typ som behandlats i a)-delen av denna uppgift. Fabriken behöver ett helt system för att kunna hålla igång produktionen, så när ett system går sönder sätts nästa igång. Antag att ett arbetsår består av 250*8 = 2000 timmar. Vad är sannolikheten att de 57 systemen kommer att räcka minst ett helt arbetsår? Lämpliga och väl motiverade approximationer är tillåtna. (5 p)

### Uppgift 15

Körsträckan för bildäck med en ny gummiblandning testades för sexton däck och följande resultat ( i 100 mil) erhölls:

| | | | |
|---|---|---|---|
| 60.6 | 59.8 | 59.6 | 60.3 |
| 59.8 | 60.2 | 60.3 | 50.0 |
| 60.5 | 60.3 | 60.0 | 59.9 |
| 69.9 | 60.1 | 60.2 | 60.1 |

Pröva hypotesen att medelkörsträckan är 6000 mil på signifikansnivån 0.05. Det får antas att observationerna ovan är utfall av oberoende normalfördelade variabler. (10 p)

### Uppgift 16

En kemist analyserar $n = 10$ olika prov med två olika mätmetoder $A$ och $B$ och erhåller då $n$ par av värden $(x_i, y_i)$, $i = 1, \ldots, n$. Antag att $x_i$:na är observationer av $X_i \in N(\mu_i, \sigma_1)$ och $y_i$:na är observationer av $Y_i \in N(\mu_i + \Delta, \sigma_2)$ och att alla stokastiska variabler är oberoende. Parametern $\sigma_1$ antas känd, men parametrarna $\mu_1, \ldots, \mu_n$ och $\sigma_2$ är okända. Tag fram ett konfidensintervall för $\sigma_2$ med konfidensgrad 95% med hjälp av följande data.

| Prov | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Metod $A$ | 1.08 | 0.75 | 2.81 | 4.85 | 5.90 | 6.94 | 7.00 | 8.39 | 8.91 | 10.25 |
| Metod $B$ | 1.68 | 2.28 | 4.11 | 5.99 | 7.05 | 7.53 | 7.97 | 10.48 | 10.03 | 11.01 |

(10 p)

**Lycka till!**

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1900/SF1912 SANNOLIKHETSTEORI OCH STATISTIK, ONSDAG 23 OKTOBER 2019 KL 8.00–13.00.

## Del I

Rätt rad:

1. 17
2. B
3. D
4. D
5. D
6. 0.711
7. C
8. C
9. B
10. A
11. A
12. B

Kortfattade lösningar:

### Uppgift 1

$$\begin{aligned}
V(2X-Y) &= C(2X-Y,2X-Y) = C(2X,2X)+C(2X,-Y)+C(-Y,2X)+C(-Y,-Y)\\
&= 2\cdot 2C(X,X)+2\cdot(-1)C(X,Y)+(-1)\cdot(2)C(X,Y)+(-1)\cdot(-1)C(Y,Y)\\
&= 4V(X)-4C(X,Y)+V(Y) = 16-8+9 = 17
\end{aligned}$$

### Uppgift 2

Här har vi lagen om total sannolikhet.
Låt $A, B, C, D$ beteckna händelserna att en lägenhet har ett, två, tre resp. fyra rum.

Låt $V$ beteckna händelsen att en vattenskada inträffat i en slumpmässigt vald lägenhet under 10-årsperioden. Då blir

$$\begin{aligned} P(V) &= P(V|A)P(A) + P(V|B)P(B) + P(V|C)P(C) + P(V|D)P(D) \\ &= 0.10 \cdot \frac{1}{6} + 0.05 \cdot \frac{1}{3} + 0.08 \cdot \frac{2}{5} + 0.10 \cdot \frac{1}{10} = 0.0753 \end{aligned}$$

Svar: Sannolikheten är 7.53% att en slumpvis vald lägenhet har drabbats av en vattenskada under 10-årsperioden.

### Uppgift 3

$$E(X^2) = \int_0^2 x^2 \cdot \frac{x}{10}dx + \int_2^6 x^2 \cdot \frac{1}{5}dx = \left[\frac{1}{40}x^4\right]_0^2 + \left[\frac{1}{15}x^3\right]_2^6 = \frac{16}{40} + \frac{216-8}{15} = 14.3$$

### Uppgift 4

$$\begin{aligned} P(X+Y>1) &= 1 - P(X+Y \le 1) \\ &= 1 - [P(X+Y=0) + P(X+Y=1)] = \{\text{oberoende}\} \\ &= 1 - [P(X=0)P(Y=0) + P(X=0)P(Y=1) + P(X=1)P(Y=0)] \\ &= 1 - \left[\binom{2}{0}0.5^2 \cdot \frac{\binom{4}{0}\binom{4}{2}}{\binom{8}{2}} + \binom{2}{0}0.5^2 \cdot \frac{\binom{4}{1}\binom{4}{1}}{\binom{8}{2}} + \binom{2}{1}0.5 \cdot 0.5 \cdot \frac{\binom{4}{2}\binom{4}{0}}{\binom{8}{2}}\right] \\ &= 1 - \frac{1}{4}\cdot\frac{3}{14} - \frac{1}{4}\cdot\frac{8}{14} - \frac{1}{2}\cdot\frac{3}{14} \approx 1 - 0.3036 = 0.6964 \end{aligned}$$

### Uppgift 5

$P(A|C) = 1 - P(A^*|C) = 0.5$. $P(A \cap C) = P(A|C)P(C) = 0.5 \cdot 0.2 = 0.1$.
Unionsformeln $\Rightarrow P(A \cup C) = P(A) + P(C) - P(A \cap C) \Rightarrow 0.6 = P(A) + 0.2 - 0.1 \Rightarrow P(A) = 0.5$.
Således gäller att $P(A)P(C) = 0.1 = P(A \cap C)$. Alltså är $A$ och $C$ oberoende.
Vi har alltså att $P(A) = 0.5$ och att $P(A \cap B) = 0.25$ $P(B)$ fås från att $P(A \cap C|B)P(B) = P(A \cap C \cap B) \Rightarrow P(B) = \frac{0.05}{0.125} = 0.4$ D.v.s. $P(A)P(B) = 0.5 \cdot 0.4 = 0.2 \neq P(A \cap B) = 0.25$ D.v.s. $A$ och $B$ är ej ober.

### Uppgift 6

Sätt $Z = X - Y$. Detta är en linjärkombination av två oberoende normalfördelade stokastiska variabler och är således också den normalfördelad. För parametrarna har vi att $E(Z) = E(X) - E(Y) = 0 - 0 = 0$ och $V(Z) = V(X) + V(Y) = 4 + 4 = 8$, dvs $Z \in N(0, \sqrt{8})$

$$\begin{aligned} P(|Z| < 3) &= P(-3 < Z < 3) = P(\frac{-3-0}{\sqrt{8}} < \frac{Z-0}{\sqrt{8}} < \frac{3-0}{\sqrt{8}}) \\ &= \Phi(\frac{3}{\sqrt{8}}) - \Phi(\frac{-3}{\sqrt{8}}) = \Phi(\frac{3}{\sqrt{8}}) - [1 - \Phi(\frac{3}{\sqrt{8}})] \\ &= 2 \cdot \Phi(1.06) - 1 = 2 \cdot 0.8554 - 1 = 0.711 \end{aligned}$$

**Uppgift 7**

Vi har att

$$E[\mu^*] = E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E\left[X_i\right]}_{=\mu} = \mu.$$

samt att

$$\begin{aligned} V(\mu^*) &= V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{V\left(X_i\right)}_{=\sigma^2} = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

Då $n \geq 50$ följer det av centrala gränsvärdessatsen att summan är approximativt normalfördelad och därmed blir även medelvärdet det eftersom linjärkombinationer av oberoende normalfördelade stokastiska variabler är normalfördelade.

**Uppgift 8**

Vi har givet $I_{\mu_x-\mu_y} = (149.41, 172.66)$. Detta är ett konfidensintervall för skillnaden mellan två stickprovs väntevärden där vi antar att stickproven har samma okända varians. M.h.a. §12.2 och § 11.2 fås konfidensintervallet till

$$I_{\mu_x-\mu_y} = \bar{x} - \bar{y} \pm s\cdot\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}\cdot t_{\frac{\alpha}{2}}(n_x+n_y-2)$$

där

$$s^2 = \frac{(n_x-1)\cdot s_x^2 + (n_y-1)\cdot s_y^2}{n_x+n_y-2}$$

Intervallet $I_{\mu_x-\mu_y} = (149.34, 172.66)$ kan också skrivas $I_{\mu_x-\mu_y} = 161 \pm 11.66$. Viket ger att

$$s\cdot\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}\cdot t_{\frac{\alpha}{2}}(n_x+n_y-2) = 11.66$$

Nu har vi med insatta värden

$$\sqrt{\frac{s_x^2+s_y^2}{2}}\sqrt{\frac{1}{10}+\frac{1}{10}}\cdot t_{\frac{\alpha}{2}}(18) = 11.66 \Rightarrow t_{\frac{\alpha}{2}}(18) = 2.88$$

Om man tittar i tab 3 så ser man att $\frac{\alpha}{2} = 0.005$. Alltså har vi ett konfidensintervall med graden 99%.

**Uppgift 9**

Eftersom $0 \notin I_{\mu_x-\mu_y}$ så kan vi förkasta $H_0$ på risknivån 1% D.v.s. risken är mindre än 1% att vi förkastar $H_0$ om den är sann. Då är risken även mindre än 5% att vi förkastar $H_0$ om den är sann och således är alltså B rätt svarsalternativ.

## Uppgift 10

Vi har här att $Q \sim \chi^2$-fördelad med $(4-1)(7-1) = 18$ frihetsgrader. D.v.s. $Q \sim \chi^2(18)$. $\chi^2_{0.025}(18) = 31.5 < 33.2 < 34.8 < \chi^2_{0.01}(18)$ Alltså ligger p-värdet mellan 1% och 2.5%. Således måste A vara rätt svar.

## Uppgift 11

Likelihoodfunktionen ges av

$$L(\theta) = f_X(x_1;\theta) f_X(x_2;\theta) = \frac{3x_1^2}{\theta} e^{-x_1^3/\theta} \frac{3x_2^2}{\theta} e^{-x_2^3/\theta} = \frac{9x_1^2 x_2^2}{\theta^2} e^{-(x_1^3+x_2^3)/\theta}$$

ML-skattningen ges av det värde $\theta^*_{ML}$ som maximerar $L(\theta)$. Logaritmering ger

$$\ln L(\theta) = \ln(9x_1^2 x_2^2) - 2\ln(\theta) - (x_1^3 + x_2^3)/\theta$$

Derivering m.a.p. $\theta$ ger

$$-2 \cdot \frac{1}{\theta} + \frac{x_1^3 + x_2^3}{\theta^2}$$

Denna derivata satt $= 0$ ger ML-skattningen $\underline{\theta^*_{ML} = \frac{x_1^3+x_2^3}{2}}$. Med de erhållna värdena insatta så fås $\theta^*_{ML} = \frac{1.65^3+6.13^3}{2} = \underline{117}$

## Uppgift 12

I enkel linjär regression skall $x_i$ kunna uppfattas som givna värden och det är alltså påstående B som är felaktigt.

# Del II

### Uppgift 13

a) Anta att $X$ är antalet adelsätter som utslocknar under 300 år. Vi kan se det som att $X$ är en Poissonfördelad stokastisk variabel med utslocknande intensiten $\lambda = 2/\text{år}$. Då gäller att $X \in Po(\lambda \cdot t) = Po(2 \cdot 300) = Po(600)$. Eftersom $Po(\mu) \sim N(\mu, \sqrt{\mu})$ om $\mu \geq 15$ enligt §15 i F.S. så gäller att $X \sim N(600, \sqrt{600})$ Nu löser vi $P(X \geq 661) =$ [Gör om till N(0,1)]

$$= P(\frac{X - 600}{\sqrt{600}} \geq \frac{661 - 600}{\sqrt{600}}) = 1 - \Phi(\frac{61}{\sqrt{600}}) \approx 1 - \Phi(2.49) = 1 - 0.99361 \approx 0.0064$$

b) Även här kan vi med samma argument som i a-uppgiften anta att $X \sim N(\mu, \sqrt{\mu})$ och nu ska vi lösa ut $\mu$ ur ekvationen

$$P(\frac{X - 300 \cdot \mu}{\sqrt{300 \cdot \mu}} \geq \frac{661 - 300 \cdot \mu}{\sqrt{300 \cdot \mu}}) = 0.95$$

Sätt

$$Y = \frac{X - 300 \cdot \mu}{\sqrt{300 \cdot \mu}}$$

Då får vi

$$P(Y \geq \frac{661 - 300 \cdot \mu}{\sqrt{300 \cdot \mu}}) = 0.95$$

Symmetri ger då att

$$P(Y \geq \frac{300 \cdot \mu - 661}{\sqrt{300 \cdot \mu}}) = 0.05$$

Dvs

$$\frac{300 \cdot \mu - 661}{\sqrt{300 \cdot \mu}} = \lambda_{0.05} = 1.6449$$

$$300 \cdot \mu - 661 = 1.6449 \cdot \sqrt{300 \cdot \mu}$$

Sätt $\sqrt{\mu} = n \Rightarrow$

$$300 \cdot n^2 - 661 = 1.6449 \cdot \sqrt{300} \cdot n$$

$$n^2 - \frac{1.6449}{\sqrt{300}} \cdot n - \frac{661}{300} = 0$$

$$n = \frac{1}{2} \cdot \frac{1.6449}{\sqrt{300}} \pm \sqrt{\left(\frac{1}{2} \cdot \frac{1.6449}{\sqrt{300}}\right)^2 + \frac{661}{300}}$$

Eftersom $n = \sqrt{\mu}$ måste vara positiv behåller vi bara den positiva lösningen $n = 1.532606417 \Rightarrow \mu = n^2 = 2.348882429$

Svar: $\underline{\mu \approx 2.35}$

**Uppgift 14**

a) Låt $T$=systemets livslängd. Då gäller att $T = \min\{T_1, T_2\}$ och

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(\min\{T_1, T_2\} \leq t) = 1 - P(\min\{T_1, T_2\} > t) \\ &= 1 - P(T_1 > t, T_2 > t) = \{\text{oberoende}\} = 1 - P(T_1 > t)P(T_2 > t) \\ &= 1 - [1 - P(T_1 \leq t)]\,[1 - P(T_2 \leq t)] \\ &= 1 - \left[1 - \frac{t}{100}\right]^2 \end{aligned}$$

Derivera så fås att

$$f_T(t) = \frac{d}{dt}F_T(t) = \frac{d}{dt}\left(1 - \left[1 - \frac{x}{100}\right]^2\right) = \frac{2}{100}\left[1 - \frac{t}{100}\right].$$

För väntevärdet fås

$$\begin{aligned} E(T) &= \int_{-\infty}^{\infty} t f_T(t)dt = \int_0^{100} t\frac{2}{100}\left[1 - \frac{t}{100}\right] \\ &= \frac{2}{100}\left[\frac{t^2}{2} - \frac{1}{100}\frac{t^3}{3}\right]_0^{100} = \frac{100}{3} \end{aligned}$$

och för variansen

$$\begin{aligned} E(T^2) &= \int_{-\infty}^{\infty} t^2 f_T(t)dt = \int_0^{100} t^2\frac{2}{100}\left[1 - \frac{t}{100}\right] \\ &= \frac{2}{100}\left[\frac{t^3}{3} - \frac{1}{100}\frac{t^4}{4}\right]_0^{100} = \frac{100^2}{6} \\ V(T) &= E[T^2] - (E[T])^2 = \frac{100^2}{6} - \left(\frac{100}{3}\right)^2 = \frac{100^2}{18} \end{aligned}$$

b) Låt $U$ beteckna den sammanlagda livslängden för de 57 systemen, dvs

$$U = \sum_{i=1}^{57} T_i$$

där $T_i$, $i = 1, \ldots, 57$ nu betecknar livslängden för system $i$ (inte livslängderna för komponenterna 1 och 2). Då 57 är förhållandevis stort är $U$ alltså en summa av många oberoende likafördelade stokastiska variabler och enligt centrala gränsvärdessatsen approximativt normalfördelad. För parametrarna får vi $E[U] = 57 \cdot 100/3$, respektive $D(U) = \sqrt{57} \cdot 100/\sqrt{18}$ (eller $V(U) = 57 \cdot 100^2/18$).

Nu gäller att

$$P(U \geq 2000) = \{\text{kontinuerlig s.v.}\} = 1 - P(U \leq 2000)$$

$$1 - P\left(\frac{U - 1900}{\sqrt{57} \cdot 100/\sqrt{18}} \leq \frac{2000 - 1900}{\sqrt{57} \cdot 100/\sqrt{18}}\right) \approx 1 - \Phi(0.5620) \approx \underline{0.2871}$$

## Uppgift 15

Det antas att $x_1, ...., x_{16}$ är utfall av de oberoende stokastiska variablerna $X_1, ...., X_{16}$, som är Normalfördelade $N(\mu, \sigma)$.
För att testa nollhypotesen $H_0 : \mu = 60.0$ mot mothypotesen $H_1 : \mu \neq 60.0$ bildas -eftersom $\sigma$ är okänd- det tvåsidiga konfidensintervallet

$$I_\mu = \bar{x} \pm \frac{s}{\sqrt{n}} \cdot t_{\frac{\alpha}{2}}(n-1)$$

.

$$\bar{x} = \frac{1}{16} \sum_{i=1}^{16} x_i = 60.10$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{16} (x_i - \bar{x})^2} = 3.6430756$$

Eftersom signifikansnivån är 0.05 fås $t_{\alpha/2}(n-1) = t_{0.025}(16-1) = 2.13$
Detta ger konfidensintervallet $I_\mu = 60.10 \pm 1.9399 = [58.16, 62.04]$
Eftersom 60.0 ligger i $I_\mu$ kan vi inte förkasta $H_0$ på risknivån 5%
Svar:<u>Vi kan inte förkasta $H_0$ på risknivån 5%</u>

## Uppgift 16

Här har vi typsituationen för stickprov i par. Normalt kan man då inte skatta $\sigma_1$ eller $\sigma_2$, men i och med att $\sigma_1$ kommer vi kunna skatta $\sigma_2$. Börja med att bilda differenserna $z_i = y_i - x_i$:

| Prov | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Metod $A$, $x_i$ | 1.08 | 0.75 | 2.81 | 4.85 | 5.90 | 6.94 | 7.00 | 8.39 | 8.91 | 10.25 |
| Metod $B$, $y_i$ | 1.68 | 2.28 | 4.11 | 5.99 | 7.05 | 7.53 | 7.97 | 10.48 | 10.03 | 11.01 |
| $z_i = y_i - x_i$ | 0.60 | 1.53 | 1.30 | 1.14 | 1.15 | 0.59 | 0.97 | 2.09 | 1.12 | 0.76 |

Då gäller att $z_i$ är observationer av oberoende stokastiska variabler $Z_i \in N(\Delta, \sigma)$, där $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ ($\sigma$ betecknar standardavvikelsen). Med hjälp av observationerna $z_i$ kan vi på vanligt sätt göra ett konfidensintervall för $\sigma^2$

$$I_{\sigma^2} = \left( \frac{s^2(n-1)}{\chi^2_{\alpha/2}}, \frac{s^2(n-1)}{\chi^2_{1-\alpha/2}} \right)$$

där

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{z})^2 = 0.4529^2$$

Eftersom

$$\sigma_2^2 = \sigma^2 - \sigma_1^2$$

fås ett konfidensintervall för $\sigma_2$ som

$$I_{\sigma_2} = \left( \sqrt{\frac{s^2(n-1)}{\chi^2_{\alpha/2}} - \sigma_1^2}, \sqrt{\frac{s^2(n-1)}{\chi^2_{1-\alpha/2}} - \sigma_1^2} \right)$$

Med insatta siffror blir det

$$I_{\sigma_2} = \left( \sqrt{\frac{0.4529^2(10-1)}{19.0228} - \sigma_1^2}, \sqrt{\frac{0.4529^2(10-1)}{2.7004} - \sigma_1^2} \right)$$

alltså

$$\underline{I_{\sigma_2} = \left( \sqrt{0.0970 - \sigma_1^2}, \sqrt{0.6835 - \sigma_1^2} \right)}$$