Tentamensskrivning i Matematisk statistik för **D3**
Lärare: Dan Mattsson, tfn 772 5349
Hjälpmedel: Utdelad formelsamling med tabeller (även BETA, Physics Handbook, skoltabeller, till exempel TEFYMA). Valfri räknedosa utan kommunikationsmöjlighet med andra räknedosor samt med *tömda* minnen. Inga egna anteckningar eller lärobok.

---

1. Att en exponentialfördelad stokastisk variabel X är minneslös formuleras matematiskt
$$\mathsf{P}(X > s+t | X > s) = \mathsf{P}(X > t).$$
Förklara med ord innebörden av föregående påstående och visa sedan likheten. (3p)

2. Vad är $\overline{X}$ och S^2? Bevisa att den första är en väntevärdesriktig skattning av väntevärdet, och att den andra är en väntevärdesriktig skattning av variansen. (3p)

3. En brandvarnare innehåller två batterier; en för rökdetekteringsenheten och en för signaleringen. Livslängderna för batterierna antas vara oberoende och exponentialfördelade med parametrar λ_1 och λ_2 respektive. Låt T vara tiden som brandvarnaren fungerar, dvs båda batterierna är hela, och bestäm fördelningen för T. Vad är sannolikheten att brandvarnaren fungerar efter 8 år om $\lambda_1 = 1/7$ [år^{-1}] och $\lambda_2 = 1/10$ [år^{-1}]? (3p)

4. En tulltjänsteman gör stickprovsundersökning bland väskor. I ett parti om 50 väskor väljer han på måfå ut fem stycken.
   (a) Om partiet innehåller 3 väskor med illegala substanser, bestäm fördelningen för X, antalet sådana väskor bland de utvalda. (1p)
   (b) Bestäm $\mathsf{P}(X = 0)$. (1p)
   (c) Är stickprovsstorleken tillräckligt stor för att med 90% sannolikhet hitta åtminstone en sådan väska, om två av tio väskor innehåller illegala substanser? (1p)

5. Låt $X_1, \ldots, X_n$ vara ett stickprov på X, där $X \sim \mathrm{N}(0, \sigma^2)$. Tag fram moment- och maximi-metodens skattningar av σ^2 och visa att båda är väntevärdesriktiga. (3p)

6. I en modell för en population är hälften av alla hushåll enpersonshushåll, varav hälften är ensamboende kvinnor. I flerpersonshushåll antas en kvinna svara i telefonen i 60% av fallen.

   En anställd vid en telemarketingfirma misstänks fuska genom att fabricera gjorda telefonsamtal. Av 500 rapporterade telefonsamtal uppges 252 blivit besvarade av män och 248 av kvinnor. Verkar detta rimligt? (3p)

7. Två butiker konkurrerar om kaffekunder. Butik A tar 40kr/kg, butik B 50kr/kg. Kunder kommer till butik A med en intensitet av 13 i timmen, och till butik B med 9 i timmen. Butik A har öppet i åtta timmar medan butik B har öppet i nio.

   Om varje kund köper ett kilo kaffe, bestäm butikernas förväntade kaffevinster och approximera sannolikheten att butik B tjänar mer än A under en dag. (3p)

Vänd!

8. Två rollspelare utkämpar en strid med hjälp av en fyrsidig tärning. Vid 400 kast erhölls följande resultat

Resultat	1	2	3	4
Antal	84	113	116	87

Är tärningen verkligen rättvis? (3p)

9. Temperaturmätningar på en reglerad process (etanol/vatten-destillation) ger stickprovet $X_1, \ldots, X_n$ på X som kan antas vara normalfördelad. För $n = 20$ observationer erhölls

$$\overline{x} = 78.12°\text{C} \qquad s^2 = 2.348.$$

Låt X_{n+1} vara resultatet vid en ännu ej gjord temperaturmätning.

(a) Vad är fördelningen för $X_{n+1} - \overline{X}$?

(b) Vad är fördelningen för

$$\frac{X_{n+1} - \overline{X}}{s\sqrt{\frac{1}{n} + 1}} \; ?$$

(c) Använd resultatet ovan för att bestämma ett 95% intervall för *mätresultatet* X_{n+1}.

(3p)

10. Vid mätningar av resistansen hos en förstärkare antar man att mätfelen är multiplikativa i responsen Y. Betrakta därför modellen

$$Y(x) = a \cdot \mathrm{e}^{bx+\epsilon}, \qquad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Baserat på $n = 3$ observationer

y	2.1	5.3	10.2
x	1	2	3

skatta a och b. (3p)

1. En stokastisk variabel X är minneslös om

$$\mathsf{P}\left(X > t + s | X > s\right) = \mathsf{P}\left(X > t\right),$$

vilket med ord innebär att den återstående livslängden efter en fix tid s har samma fördelning som för en ny komponent.

$$\mathsf{P}\left(X > t + s | X > s\right) = \frac{\mathsf{P}\left(X > t + s\right)}{\mathsf{P}\left(X > s\right)} = \frac{\mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda s}} = \mathrm{e}^{-\lambda t} = \mathsf{P}\left(X > t\right).$$

2. $\overline{X}$ betecknar medelvärdet för ett stickprov $X_1, \ldots, X_n$ på X,

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

och S^2 är stickprovsvariansen

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

$$\mathsf{E}\left[\overline{X}\right] = \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left[X_i\right] = \mathsf{E}\left[X\right].$$

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

3. Låt X_1 och X_2 vara batteriernas livslängder. Då är

$$\mathsf{P}\left(T > t\right) = \mathsf{P}\left(X_1 > t,\ X_2 > t\right) = \{\text{ober.}\} = \mathsf{P}\left(X_1 > t\right)\mathsf{P}\left(X_2 > t\right) = \mathrm{e}^{-\lambda_1 t}\mathrm{e}^{-\lambda_2 t},$$

det vill säga

$$F_T(t) \;=\; \mathsf{P}\left(T \leq t\right) \;=\; 1 - \mathrm{e}^{-(\lambda_1+\lambda_2)t},$$

och T är exponentialfördelad med parameter $\lambda = \lambda_1 + \lambda_2 = 17/70$. Vi får sedan att $\mathsf{P}\left(T > 8\right) = \mathrm{e}^{-\lambda 8} \approx 14\%$.

4. Låt X vara antalet illegala väskor bland de $n = 5$ utvalda.

   (a) Om partiet består av N väskor varav r är illegala så är

$$\mathsf{P}\left(X = k\right) = \binom{r}{k}\binom{N-r}{n-k}\Big/\binom{N}{k}, \qquad \max(0, n + r - N) \leq k \leq \min(n, r).$$

   Här får vi att, $N = 50$, $r = 3$,

$$\mathsf{P}\left(X = k\right) = \frac{\binom{3}{k}\binom{47}{5-k}}{\binom{50}{5}}, \qquad 0 \leq k \leq 3.$$

(b) $\mathsf{P}(X=0) = \binom{47}{5}/\binom{50}{5} = 0.7240$.

(c) Med $r = 10$ fås $\mathsf{P}(X=0) = \binom{40}{5}/\binom{50}{5} = 0.3106$, dvs med sannolikhet $68.9\% < 90\%$ kommer vi att hitta åtminstone en illegal väska.

5. Låt $X_1, \ldots, X_n$ vara ett stickprov på $X \sim \mathrm{N}(0, \sigma^2)$. Eftersom $\mathsf{E}[X] = 0$ och $\mathsf{E}\left[X^2\right] = \mathrm{Var}(X) + \mathsf{E}[X]^2 = \sigma^2$, så blir momentskattningen av σ^2,

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Skattningen är väntevärdesriktig ty

$$\mathsf{E}\left[\widehat{\sigma}^2\right] = \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left[X_i^2\right] = \frac{1}{n}\sum_{i=1}^{n}(\sigma^2+0) = \sigma^2.$$

Likelihoodfunktionen för stickprovet givet de observerade värdena är

$$L(\sigma) = \prod_{i=1}^{n} f_{X_i}(x_i) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-0)^2\right\}.$$

Logaritmering och derivering ger

$$\frac{d}{d\sigma^2}\log L = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} x_i^2 = 0.$$

Förenkling och omarrangering ger slutligen

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad\Rightarrow\quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Detta är samma skattning som momentskattningen så den är också väntevärdesriktig.

6. Låt p_0 vara andelen gånger män som svarar i telefon vid ett slumpvist utvalt hushåll enligt vår modell. Då är $p_0 = 1/4 + 1/2 \cdot 4/10 = 0.45$. Bland $n = 500$ telefonsamtal, låt X vara antalet gånger en man svarade. Antag att $X \sim \mathrm{Bin}(n, p)$, och ställ upp hypoteserna

$$H_0\colon p = p_0 \quad \text{och} \quad H_1\colon p \neq p_0,$$

som vi vill testa på nivå $\alpha = 5\%$. Vi skattar p med $\widehat{p} = X/n$, och får det observerade värdet $\widehat{p} = 0.5040$. Vi bildar statistikan

$$T = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

som under H_0 är approximativt $\mathrm{N}(0,1)$ och har det observerade värdet $t = 2.4271$. Vi förkastar H_0 för stora värden på $|T|$ och ser att $|t|$ motsvarar p-värdet 0.0152. Slutsatsen är att vi förkastar H_0 på nivå 5%. Om vi tror på modellen så ljuger förmodligen den anställde undersökaren.

7. Låt X_A och X_B vara antalet kunder som kommer till butik A resp. B under en dag. Antag att de är oberoende och Poissonfördelade med intensiteter $\lambda_A = 13$ och $\lambda_B = 9$. Då är $\mathsf{E}[X_A] = \mathrm{Var}(X_A) = 8\lambda_A = 104$ och $\mathsf{E}[X_B] = \mathrm{Var}(X_B) = 9\lambda_B = 81$. Bilda skillnaden i kaffevinst som $Y = 40X_A - 50X_B$ som har $\mathsf{E}[Y] = 110$ och $\mathrm{Var}(Y) = 368900$. Vi approximerar fördelningen för Y med normalfördelningen.

$$\mathsf{P}(Y < 0) = \mathsf{P}\left(\frac{Y - 110}{\sqrt{368900}} < \frac{0 - 110}{\sqrt{368900}}\right) \approx \Phi(-0.1811) = 0.43.$$

8. Om tärningen är rättvis så skall $p_i = \mathsf{P}(\text{Tärning visar sida } i)$ vara $1/4$. Vi vill testa hypoteserna

$$H_0 : p_1 = \cdots = p_4 = \frac{1}{4} \quad \text{mot} \quad H_1 : p_i \neq \frac{1}{4}, \text{ något } i.$$

på nivå $\alpha = 0.05$. Bland $n = 400$ kast förväntar vi oss att $e_i = np_i = 100$ skall visa sida i. Vi jämför de observerade frekvenserna O_i med de förväntade med hjälp av statistikan

$$Q = \sum_{i=1}^{4} \frac{(O_i - e_i)^2}{e_i}$$

som under H_0 är approximativt χ^2-fördelad med 3 frihetsgrader. Vi förkastar H_0 för stora värden på Q, och ur tabell får vi att $\mathsf{P}(Q > 7.815) = 0.05$. Vi observerar utfallet $q = 8.5 > 7.815$ och förkastar H_0 på nivå 5%. (p-värde 0.0367)

9. $X_1, \ldots, X_{n+1}$ är ett stickprov på $X \sim \mathrm{N}(\mu, \sigma^2)$. Då är $X_{n+1} - \overline{X}$ normalfördelad med $\mathsf{E}\left[X_{n+1} - \overline{X}\right] = \mu - \mu = 0$ och varians $\mathrm{Var}\left(X_{n+1} - \overline{X}\right) = \sigma^2 + \sigma^2/n$.
Eftersom $(n-1)S^2/\sigma^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2$ är χ^2_{n-1} så är

$$T = \frac{X_{n+1} - \overline{X}}{S\sqrt{\frac{1}{n} + 1}} = \frac{X_{n+1} - \overline{X}}{\sigma\sqrt{\frac{1}{n} + 1}} \Bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2}\frac{1}{n-1}} \sim t_{n-1}\text{-fördelad.}$$

(En kvot mellan oberoende $\mathrm{N}(0,1)$ och roten av $\chi^2_{n-1}/(n-1)$.)
Sålunda kan vi ur tabell t_{n-1} bestämma $t_{\alpha/2}$ så att $\mathsf{P}\left(|T| \leq t_{\alpha/2}\right) = 1 - \alpha$. För $n = 20$ och $\alpha = 5\%$ får vi $t_{\alpha/2} = 2.0930$. Dvs, med sannolikhet 95% är

$$-t_{\alpha/2} \leq \frac{X_{n+1} - \overline{X}}{S\sqrt{\frac{1}{n} + 1}} \leq t_{\alpha/2}$$

eller, omformulerat,

$$\overline{X} - t_{\alpha/2}S\sqrt{\frac{1}{n} + 1} \leq X_{n+1} \leq \overline{X} + t_{\alpha/2}S\sqrt{\frac{1}{n} + 1}.$$

Med värden får vi

$$74.83 \leq X_{n+1} \leq 81.41 \quad (95\%).$$

10. Låt $Y(x) = a \cdot \mathrm{e}^{bx+\epsilon}$. Låt $Z(x) = \log Y(x) = \log(a) + bx + \epsilon$. Gör linjär regression på Z med data

$$\sum x_i = 6.0 \quad \sum z_i = 4.732 \quad \sum x_i^2 = 14.0 \quad \sum x_i z_i = 11.045 \quad n = 3$$

och de observerade skattningarna blir

$$\widehat{b} = \frac{n\sum x_i z_i - (\sum x_i)(\sum z_i)}{n\sum x_i^2 - (\sum x_i)^2} = 0.79, \quad \widehat{\log a} = \overline{z} - \widehat{b}\overline{x} = -0.0031 \quad \Rightarrow \quad \widehat{a} = 0.997.$$

Tentamensskrivning i Matematisk statistik för **D3**
Lärare: Dan Mattsson, tfn 772 5349
Hjälpmedel: Utdelad formelsamling med tabeller (även BETA, Physics Handbook, skoltabeller, till exempel TEFYMA). Valfri räknedosa utan kommunikationsmöjlighet med andra räknedosor samt med *tömda* minnen. Inga egna anteckningar eller lärobok.

---

1. Vi har 15 observationer på X med medelvärde $\overline{x} = 20.1340$ och standardavvikelse $s_X = 0.9550$. Motsvarande 16 mätningar på Y ger $\overline{y} = 20.9318$ och $s_Y = 0.9921$.

   Testa hypotesen H_0: $\mu_X = \mu_Y$ mot H_1: $\mu_X < \mu_Y$ på signifikansnivå 5%. Vilken sats motiverar att du kan göra detta test? (3p)

2. Dan är intresserad av att lägga ett bud på en ministereo på en auktionsfirma (Svensk teleauktion). Eftersom han har data från tidigare auktioner bestämmer han sig för att göra regression på säljpriserna för att få en skattning på vad som är ett rimligt bud. Ansätt modellen:

$$(\text{Försäljningspris}) \;=\; \alpha + \beta \cdot (\text{utropspris}) + (\text{fel})$$

   där felen är oberoende $\sim \mathrm{N}(0, \sigma^2)$. Skatta parametrarna α och β med nedstående data och bestäm förväntat försäljningspris på ministereon vars utropspris är 1250.

Produkt	Säljpris y_i [tKr]	Utropspris x_i [tKr]
Sanyo Widescreen-TV	8.005	3.000
Sanyo Video	2.512	1.125
Sanyo MT2,TV	2.050	1.000
Sony Walkman WM-FX571	0.800	0.375
Sony Bioljudanläggning DP-150	3.000	1.250
Sony MiniDisc Walkman	2.775	1.000
Sony DVD-spelare	4.500	1.875
Sony 29 tum TV	8.621	3.875
Sony Digital 8 Kamera	7.610	3.750
Sony Videobandspelare	2.500	1.250

   Observationerna kan sammanfattas i

$$\sum x_i = 18.5 \quad \sum x_i^2 = 48.125 \quad \sum y_i = 42.373$$
$$\sum y_i^2 = 250.667 \quad \sum x_i y_i = 109.222.$$

   (3p)

3. Följande är 5 observationer på en exponentialfördelad stokastisk variabel X med intensitet λ. Tag fram maximimetodens skattning av λ och beräkna $\widehat{\lambda}$ för dessa observationer. (3p)

   3.8708 2.8297 1.1520 14.1356 2.5303

4. Man vet att 50% av alla datorchips som tillverkas är defekta. Kontrollverksamhet garanterar att endast 5% av de chips som säljs (legalt) är defekta. Dessvärre stjäls chips innan kontroll sker. Om 1% av alla datorchips på marknaden är stulna (och därför okontrollerade) finn

Vänd!

(a) Sannolikheten att ett på måfå valt chip är defekt. (1p)

(b) Sannolikheten att ett defekt chip även är stulet. (2p)

5. Dan behöver hjälp med att tvätta tavlorna på rasten och plockar därför ut tre personer, Kajsa, Ola och Sharon, på första raden före lektionen. Dock hann inte Dan så långt som han tänkt så bara två personer behövdes när rasten kom, och Dan valde ut den person som slapp hjälpa till genom att i smyg kasta en tresidig tärning. Innan resultatet blev känt, tänkte Kajsa ut en briljant plan och frågade Dan om namnet på *en av de andra* utvalda personerna. Någon av Ola och Sharon måste ju stanna, och deras namn har lika stor chans att nämnas. Om Ola nämns så har Kajsa och Sharon båda sannolikhet 1/2 att få ledigt. På samma sätt om Sharons namn nämns har Kajsa sannolikhet 1/2 att slippa taveljobbet. Så, bara genom att fråga ökade Kajsa sin sannolikhet att få ledigt från 1/3 till 1/2. Är det inte en briljant plan?

6. Baserat på följande 10 observationer, ta fram ett symmetriskt 95% konfidensintervall för väntevärdet μ och ett 90% uppåt begränsat konfidensintervall för standardavvikelsen σ.
(3p)

8.89 9.02 8.91 9.49 8.87 9.39 7.00 8.18 8.04 6.60

7. På en fråga om de har röstat i år svarade 834 personer av 1006 att de hade gjort det. Låt p vara den sanna andelen som röstat och testa hypotesen H_0: $p = 0.80$ mot H_1: $p \neq 0.80$ på signifikansnivå 5%. Bestäm p-värdet. Kan vi förkasta på 1%? (3p)

8. Pepparkakor har vikter $X \sim \mathrm{N}(\mu, \sigma^2)$ där $\mu = 5$ och $\sigma = 0.9$ [gram]. Man vill att ett paket med n pepparkakor skall väga åtminstone 300 gram. Bestäm det minsta n så att

$$\mathsf{P}\,(\text{paketets vikt} \geq 300)$$

är åtminstone 0.90. (3p)

9. I en vattenledning sker läckor enligt en Poissonprocess med i genomsnitt 10 läckor under 45 år. Vad är sannolikheten att det under 10 år, finns åtminstone 7 *år* utan läckor? (3p)

10. En rökare har två tändsticksaskar med från början n stycken tändstickor i varje ask. Den ena asken förvarar han i vänster ficka och den andra i höger. Varje gång han vill ha en tändsticka väljer han en ask på måfå. Vad är fördelningen för antalet tändstickor i andra asken när han precis tagit den sista tändstickan ur den ask han valde? (3p)

1. Vi skattar $\mu_X - \mu_Y$ med $\overline{X} - \overline{Y}$. Om X_i och Y_i har samma varians så är enligt centrala gränsvärdessatsen

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \overset{\text{approx}}{\sim} t_{n+m-2}\text{-fördelad}$$

Vi skattar σ med S där

$$S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} \quad \Rightarrow \quad s^2 = 0.9494 \quad \Rightarrow \quad s = 0.9744.$$

Vi förkastar H_0 för små värden på T och vi bestämmer en undre gräns t_α så att $\mathsf{P}(T \le t_\alpha) = 0.05$, och ur t_{29}-tabeller får vi att $t_\alpha = -1.6991$.

Dvs vi förkastar H_0 om vi observerar $T < -1.6991$. Här är $t = -2.28$ och vi förkastar H_0 på nivån 5%.

2. Med formelsamlingens beteckningar får vi

$$\overline{x} = 1.8500 \quad S_{xx} = \sum x_i^2 - n\overline{x}^2 = 13.9 \quad \overline{y} = 4.2373$$

och

$$S_{xy} = \sum x_i y_i - n\overline{x} \cdot \overline{y} = 30.8323$$

vilket ger skattningarna

$$\widehat{\beta} = \frac{S_{xy}}{S_{xx}} = 2.2182 \quad \widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x} = 0.1337,$$

vilket resulterar i skattningen $\mathsf{E}\left[\widehat{\alpha} + \widehat{\beta} \cdot 1.250\right] = 2.906$ [tKr].

3. För exponentialfördelningen är $f_X(x) = \lambda e^{-\lambda x}$, $x \ge 0$, vilket ger trolighetsfunktionen

$$L(\lambda) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.$$

Logaritmering och derivering ger i maximeringspunkten $0 = \frac{d}{d\lambda} \log L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$, och eftersom $\frac{d^2}{d\lambda^2} \log L(\lambda) = -\frac{n}{\lambda^2} < 0$ är trolighetsfunktionen maximal för $\widehat{\lambda} = n/\sum X_i = 1/\overline{X}$. Vi beräknar $\widehat{\lambda} = 1/4.9037 = 0.2039$.

4. Låt $A = \{\text{chip stulet}\}$ och $B = \{\text{chip defekt}\}$. Då är $\mathsf{P}(B|A) = 0.50$, $\mathsf{P}(B|A^c) = 0.05$ och $\mathsf{P}(A) = 0.01$.

   (a) $\mathsf{P}(B) = \mathsf{P}(B|A)\,\mathsf{P}(A) + \mathsf{P}(B|A^c)\,\mathsf{P}(A^c) = 0.0545.$

   (b) $\mathsf{P}(A|B) = \dfrac{\mathsf{P}(B|A)\,\mathsf{P}(A)}{\mathsf{P}(B)} = 0.0917.$

5. Nej, det är ingen briljant plan. Om till exempel Olas namn nämns, har inte Sharon och Kajsa samma sannolikhet att få ledigt. Betrakta följande tabell:

Sannolikhet	Ola	Kajsa	Sharon	Dan säger:
1/3	✗			"Sharon"
1/6		✗		"Sharon"
1/6		✗		"Ola"
1/3			✗	"Ola"

där kryssen visar resultatet av tärningskastet. Så

$$\mathsf{P}\left(\text{Kajsa ledig}|\text{"Ola"}\right) = \frac{1/6}{1/6+1/3} = \frac{1}{3}.$$

och

$$\mathsf{P}\left(\text{Kajsa ledig}\right) = \frac{1}{3}$$

som alla insett.

6. Vi beräknar $\overline{x} = 8.439$ och $s^2 = 0.9596$. Enligt centrala gränsvärdessatsen är $(\overline{X} - \mu)\sqrt{n}/S \sim t_9$-fördelad, och $(n-1)S^2/\sigma^2 \sim \chi^2_9$. Sålunda gäller

$$\mathsf{P}\left(\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right| \leq 2.2622\right) = 0.95 \qquad \mathsf{P}\left(\frac{(n-1)S^2}{\sigma^2} \geq 4.1682\right) = 0.90.$$

Vi får de observerade intervallen

$$\begin{aligned} \mu &= \overline{x} \pm 2.2622\, s/\sqrt{n} \\ &= 8.439 \pm 0.701 \qquad (95\%) \end{aligned} \qquad \begin{aligned} \sigma^2 &\leq (n-1)s^2/4.1682 \\ \sigma^2 &\leq 2.072 \qquad (90\%) \\ \sigma &\leq 1.439 \qquad (90\%) \end{aligned}$$

7. Vi skattar p med $\widehat{p} = 834/1006 = 0.829$. Enligt CGS och Slutsky är under $H_0 : p = p_0 = 0.80$

$$\mathsf{P}\left(\left|\frac{\widehat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}\right| \geq 1.96 \;\middle|\; H_0\right) = 0.95.$$

Det vill säga vi förkastar H_0 om $|\widehat{p} - p_0| \geq 1.96\sqrt{p_0(1-p_0)/n} = 0.0207$. Vi observerar $\widehat{p} - p_0 = 0.0290$ så vi förkastar H_0. Slutligen är

$$\text{p-värdet} = 2\left(1 - \Phi\left(\frac{\widehat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}\right)\right) = 2(1-\Phi(2.3016)) \approx 2.14\%,$$

så vi kan inte förkasta H_0 på 1%.

8. Låt $S = X_1 + \cdots + X_n$ vara paketets vikt. Då är $S \sim \mathrm{N}(n\mu, n\sigma^2)$ och relationen $\mathsf{P}(S \geq 300) = 1 - \Phi((300 - n\mu)/(\sqrt{n}\sigma)) = 0.90$ ger att $(300 - n\mu)/(\sqrt{n}\sigma) = z_\alpha = -1.2816$. Då kan vi skriva

$$n + \frac{z_\alpha \sigma}{\mu}\sqrt{n} - \frac{300}{\mu} = 0 \;\Rightarrow\; \sqrt{n} = -\frac{z_\alpha\sigma}{2\mu} \pm \sqrt{\left(\frac{z_\alpha\sigma}{2\mu}\right)^2 + \frac{300}{\mu}} = 7.86.$$

Då är $n = 61.8$, det vill säga man skall ta 62 kakor eller fler.

9. Enligt uppgift är antalet läckor $\sim \mathrm{Po}(\lambda T)$ där $\lambda = 10/45$. Under ett år är antalet läckor $X \sim \mathrm{Po}(\lambda \cdot 1)$ och $p = \mathsf{P}(X = 0) = \mathrm{e}^{-\lambda} \approx 0.8007$ är sannolikheten att det under ett år inte sker några läckor. Under 10 år är antalet år utan läckor $Y \sim \mathrm{Bin}(10, p)$ och $\mathsf{P}(Y \geq 7) \approx 1 - 0.12 = 88\%$.

10. Låt oss säga att tändstickorna i vänster ask tar slut först, och beteckna med X antalet tändstickor tagna ur höger ask. Då motsvarar händelsen $X = 0$ att vi enbart plockat tändstickor ur vänster ask, och $\mathsf{P}(X = 0) = (0.5)^n$. Händelsen $X = 1$ blir att vi plockat n stycken tändstickor ur vänster ask och 1 tändsticka ur höger ask, totalt $n + 1$ tändstickor där den sista kom från den vänstra. Tändstickan från den högra kan då väljas som 1:a, 2:a,…,n:te tändsticka, dvs på $\binom{n}{1}$ sätt, och

$$\mathsf{P}(X = 1) = \binom{n}{1}(0.5)^{n+1}.$$

Generellt, $X = k$ blir att vi plockat $n + k$ tändstickor, där de k stycken från höger ask kan väljas bland de $n + k - 1$ första tändstickorna, dvs

$$\mathsf{P}(X = k) = \binom{n+k-1}{k}(0.5)^{n+k},\ k = 0, 1, \ldots, n-1.$$

Eftersom problemet är symmetriskt i höger och vänster ask så är

$$\mathsf{P}(k \text{ stickor kvar i andra asken}) = 2 \cdot \mathsf{P}(X = n-k) = \binom{2n-k-1}{n-k}(0.5)^{2n-k-1},$$

för $k = 1, 2, \ldots, n$.

Tentamensskrivning i Matematisk statistik för **D3**
Lärare: Dan Mattsson, tfn 772 5349
Hjälpmedel: Utdelad formelsamling med tabeller (även BETA, Physics Handbook, skoltabeller, till exempel TEFYMA). Valfri räknedosa utan kommunikationsmöjlighet med andra räknedosor samt med *tömda* minnen. Inga egna anteckningar eller lärobok.

---

1. Konstruera två rektanglar på följande sätt: Låt U_1, U_2 och U_3 vara oberoende likformigt fördelade på intervallet $[0, a]$ för någon konstant a. Den första rektangeln A har sidlängder U_1 och U_2, medan rektangeln B är en kvadrat med sidlängd U_3. Bestäm de förväntade areorna av rektanglarna A och B. Kommentarer?

2. Jag erbjuder dig att delta i följande spel. Vi singlar två tior och en enkrona. Om två eller fler mynt visar krona (kungens profil) så vinner jag, i annat fall du. Segraren får de mynt som blev krona. Vad blir våra förväntade vinster? Är spelet rättvist? (3p)

3. Följande är 20 observationer på passningstiden i timmar för en mobiltelefon med specialbatteri. Passningstiden kan antas vara $X \sim \exp(\lambda)$.

167	201	169	59	303	35	130	55	1	173
53	73	65	829	72	10	83	42	979	39

   Sätt upp ett 95% symmetriskt konfidensintervall för λ. Är påståendet att ett batteri i genomsnitt räcker i 300 timmar, rimligt? (3p)

4. Vid oljeborrning i Kuwait sipprar naturgas ut ur oljekällorna. För varje fat råolja får man en viss mängd naturgas [m^3], $X \sim \mathrm{N}(\mu, \sigma^2)$. För tjugo mätningar av utsläppt gasmängd per fat olja, fick man följande resultat:

$$\sum_{i=1}^{20} x_i = 11790 \quad \sum_{i=1}^{20} x_i^2 = 7503700.$$

   Tag fram ett 95% konfidensintervall för μ, den förväntade mängden utsläppt naturgas per fat upptappad råolja. (3p)

5. Antalet förolyckade i trafiken under perioden 1970–1996 ser ut enligt följande (källa SCB):

År	Antal döda
1970	1307
1980	848
1985	808
1990	772
1993	632
1994	589
1995	572
1996	537

   Ansätt modellen:

$$\text{(antalet dödsolyckor)} = \alpha + \beta \cdot \text{år} + \text{fel},$$

   där felen oberoende och $\sim \mathrm{N}(0, \sigma^2)$. Skatta α och β och därur antalet dödsolyckor under 1999. Vid vilket år är nollvisionen (inga dödsolyckor) uppnådd enligt modellen? (3p)

Vänd!

6. Åtta reservkraftsaggregat skall leverera ström, men när man mätte utspänningen erhöll man följande resultat:

$$3.1 \quad 4.1 \quad 3.4 \quad 3.7 \quad 3.3 \quad 3.2 \quad 4.0 \quad 4.0 \quad [\text{kV}]$$

Antag att spänningarna kan ses som utfall av oberoende stokastiska variabler $\sim \mathrm{N}(\mu, \sigma^2)$. Voltmetern som användes har dock mätfel som kan ses som utfall av $\mathrm{N}(0, \sigma^2)$-fördelade stokastiska variabler, $\sigma = 0.1$. Observera att vi har två källor till variation; batteriernas individuella variation och spänningsmätarens mätfel. Tag fram ett 95% konfidensintervall för σ, standardavvikelsen för ett kraftaggregats utspänning. (3p)

7. Vid en studie av läsbarheten för ingenjörers rapporter använde man sig av ett mått, kallat "index of confusion", utformat så att låga värden återspeglar god läsbarhet. Man jämförde sedan 13 stycken publicerade rapporter med 12 stycken refuserade (ej publicerade) och fick följande resultat:

Publicerade, X				Ej publicerade, Y			
1.79	1.87	1.62	1.96	2.39	2.56	2.36	2.49
1.75	1.74	2.06		2.62	2.51	2.29	2.33
1.69	1.67	1.94		2.58	2.41	2.86	1.94
1.33	1.70	1.65					
$\overline{x} = 1.7515,\ s_x = 0.1844$				$\overline{y} = 2.4450,\ s_y = 0.2224$			

Antag att X och Y har samma varians σ^2 och testa hypotesen H_0: $\mathsf{E}[X] = \mathsf{E}[Y]$ mot hypotesen H_1: $\mathsf{E}[X] < \mathsf{E}[Y]$ på nivå 5%. Vad är p-värdet? Om det inte finns i tabellerna, svara med "p-värdet $< \ldots$". (3p)

8. Bromsklossar tillverkas för att hålla åtminstone 12000 mil. Låt X vara livslängden för en godtycklig bromskloss. Då har X frekvensfunktionen

$$f_X(x) = (1/\theta^2)x\mathrm{e}^{-x/\theta}, \quad x > 0, \quad \theta > 0.$$

Bestäm maximiskattningen av θ och räkna ut den baserat på följande tio observationer ([Mm]): (3p)

166	355	140	564	424
1190	55.9	297	461	654

9. Vid övergången till år 2002 blev larmcentralen vid ett till Danmark närliggande kärnkraftverk knäpp och registrerade endast det första larmet som kom in, och därefter enbart varannat larm; alltså, larm två registreras ej. Antag att antalet larm är poissonfördelat med intensitet 2 larm/dag. Låt X vara antalet larm som registreras under årets första två dagar, och bestäm sannolikheterna $\mathsf{P}(X = k)$ för $k = 0, 1, 2, 3$. Är X poissonfördelad? (3p)

10. Två personer, A och B, står och skjuter prick på en måltavla. Båda har samma sannolikhet att träffa tavlan och de träffar oberoende av varandra. Om de skjuter varannat skott och slutar skjuta när de tillsammans träffat tavlan två gånger, vad är sannolikheten att det var samma person som sköt de båda skotten? (3p)

1. En likformigt $[0, a]$ fördelad stokastisk variabel U har väntevärde $a/2$ och varians $a^2/12$. Alltså

$$\mathsf{E}\left[\text{area } A\right] = \mathsf{E}\left[U_1 U_2\right] = \text{ober.} = \mathsf{E}\left[U_1\right]\mathsf{E}\left[U_2\right] = \frac{a^2}{4},$$

medan

$$\mathsf{E}\left[\text{area } B\right] = \mathsf{E}\left[U_3^2\right] = \mathrm{Var}\left(U_3\right) + \mathsf{E}\left[U_3\right]^2 = \frac{a^2}{12} + \frac{a^2}{4} = \frac{a^2}{3}.$$

Kvadraten har i förväntan större area.

2. Följande tabell anger alla möjliga utfall, där T =en tia visar krona, t =en tia visar klave, K =enkronan visar krona och k =enkronan visar klave.

Utfall	Slh.	Vinnare	Utdelning
ttk	1/8	du	0
ttK	1/8	du	1
tTk	2/8	du	10
tTK	2/8	jag	11
TTk	1/8	jag	20
TTK	1/8	jag	21

Min förväntade vinst är alltså $(21 + 20 + 2 \cdot 11)/8 = 5.38$, medan din är $(2 \cdot 10 + 1 + 0)/8 = 2.63$. Vilket man inte kan kalla rättvist.

3. Låt $S = X_1 + \cdots + X_n$. Då är $S \sim \Gamma(n, \lambda)$ och $\lambda S \sim \Gamma(n, 1)$. Ur $\Gamma(20, 1)$-tabell får vi att $\mathsf{P}\left(12.217 \leq \lambda S \leq 29.671\right) = 0.95$, och med 95% sannolikhet kommer

$$\frac{12.217}{S} \leq \lambda \leq \frac{29.671}{S}, \qquad \text{dvs} \qquad 0.00345 \leq \lambda \leq 0.00839.$$

Ett värde på $\lambda = 1/300$ är osannolikt, och vi skulle förkasta en sådan hypotes till förmån för $\lambda \neq 1/300$ på signifikansnivå 5%.

4. Vi får direkt att $\overline{x} = 11790/20 = 589.5$ och $s^2 = (7503700 - 11790^2/20)/19 = 29131$. Eftersom $(\overline{X} - \mu)/(S/\sqrt{n}) \sim t_{19}$ kan vi ur tabeller bestämma $t_{\alpha/2} = 2.0930$ så att

$$-t_{\alpha/2} \leq \frac{\overline{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2} \text{ eller } \overline{X} - t_{\alpha/2} S \frac{1}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2} S \frac{1}{\sqrt{n}},$$

med 95% sannolikhet. Med insatta värden får vi det observerade konfidensintervallet

$$509.6 \leq \mu \leq 669.4.$$

5. Ur data får vi att

$$\begin{aligned} \overline{x} &= 1988 \quad & S_{xx} &= 574.875 \\ \overline{y} &= 758.1 \quad & S_{xy} &= -15438 \end{aligned}$$

varur vi beräknar

$$\widehat{\beta} = S_{xy}/S_{xx} = -26.8543 \quad \widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x} = 54141.$$

Under 1999 kan vi enligt modellen förvänta oss $\widehat{\alpha} + \widehat{\beta} \cdot 1999 = 459$ olyckor med dödlig utgång, och ekvationssystemet $0 = \widehat{\alpha} + \widehat{\beta} x$ ger oss att år $x = 2016$ förväntas vi ha 0 stycken sådana olyckor.

6. Eftersom mätningarna gjordes med mätfel $\sim \mathrm{N}(0, \sigma^2)$ är våra mätdata observationer $\sim \mathrm{N}(\mu, \sigma'^2)$, där $\sigma' = \sqrt{\sigma^2 + 0.1^2}$, och det är σ'^2 vi skattar med $s^2 = 0.16$. Nu är $7S^2/\sigma'^2 \sim \chi^2_9$ så ur tabeller kan vi bestämma $q_1 = 1.6899$ och q_2 =16.0128 så att vi får ett 95% konfidensintervall för σ'^2 som

$$0.07 = \frac{7s^2}{q_2} \leq \sigma'^2 \leq \frac{7s^2}{q_1} = 0.66, \quad (95\%)$$

det vill säga motsvarande för σ blir

$$0.24 = \sqrt{\frac{7s^2}{q_2} - 0.1^2} \leq \sigma \leq \sqrt{\frac{7s^2}{q_1} - 0.1^2} = 0.81, \quad (95\%).$$

7. Först tar vi fram den bästa skattningen av den gemensamma variansen σ^2 som

$$S^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2} \quad \Rightarrow \quad s^2 = 0.0414,$$

och enligt centrala gränsvärdessatsen är med $\mu_X = \mathsf{E}[X]$ och $\mu_Y = \mathsf{E}[Y]$

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S\sqrt{1/n_X + 1/n_Y}} \overset{\text{approx}}{\sim} t_{n_X+n_Y-2}\text{-fördelad.}$$

Vi kan således under H_0 bestämma t_α så att $\mathsf{P}(T \leq t_\alpha) = 0.05$, det vill säga $t_\alpha = -1.7139$. Vi förkastar H_0 till förmån för H_1 om vi observerar att $T < -1.7139$. Eftersom vi observerar $t = -8.51$ så förkastar vi H_0 på nivån 5%. (p-värdet= $7.2 \cdot 10^{-9}$.)

8. Låt $X_1, \ldots, X_n$ vara vårt stickprov med observationer $x_1, \ldots, x_n$. Då är trolighetsfunktionen

$$L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i) = \frac{1}{\theta^{2n}} \left(\prod_{i=1}^{n} x_i \right) \mathrm{e}^{-\sum x_i/\theta}.$$

Logaritmering ger

$$\log L(\theta) = -2n \log \theta + \sum_{i=1}^{n} \log x_i - \frac{1}{\theta} \sum_{i=1}^{n} x_i.$$

Deriveras detta uttryck med avseende på θ och derivatan sätts till noll fås följande uttryck efter multiplikation med θ^2/n.

$$-2\theta + \overline{x} = 0,$$

det vill säga vår skattning är $\widehat{\theta} = \overline{X}/2$, och med insatta värden $\widehat{\theta} = 215$. De som känner igen en Γ-fördelning inser att skattningen är väntevärdesriktig.

9. Låt Y vara antalet larm som kommer in under de två första dagarna, $Y = \mathrm{Po}(4)$. Då är

$$\begin{aligned}
\mathsf{P}(X = 0) &= \mathsf{P}(Y = 0) = 0.018 \\
\mathsf{P}(X = 1) &= \mathsf{P}(Y = 1) + \mathsf{P}(Y = 2) = 0.0733 + 0.1465 = 0.2198 \\
\mathsf{P}(X = 2) &= \mathsf{P}(Y = 3) + \mathsf{P}(Y = 4) = 0.3907 \\
\mathsf{P}(X = 3) &= \mathsf{P}(Y = 5) + \mathsf{P}(Y = 6) = 0.2605.
\end{aligned}$$

X är inte poissonfördelad eftersom tiden mellan larm är summan av två exponentialfördelade stokastiska variabler, det vill säga, mellanankomsttiden är Γ-fördelad.

10. Låt p vara sannolikheten att någon träffar tavlan. Efter att första skottet träffat tavlan är händelsen att samma person träffar tavlan igen, händelsen för ett udda antal missar följt av en träff, dvs sannolikheten blir:

$$\sum_{k=0}^{\infty} p(1-p)^{1+2k} = (1-p)p\sum_{k=0}^{\infty}((1-p)^2)^k = \frac{(1-p)p}{1-(1-p)^2} = \frac{1-p}{2-p}.$$

Observera att $p = 1$ ger sannolikheten 0 (dvs båda lyckas vid första försöket) och $p \to 0$ gör att sannolikheten $\to 1/2$.