Matematisk statistik
Matematikcentrum
Lunds universitet

Tentamen: 2023–10–27 kl 8^{00}–13^{00}
FMSF30 & FMSF32
Matematisk statistik

- **Lösningsförslag**

---

1. Definiera händelserna S - *en person är sjuk*, samt T - *en person testar positivt.* Ur texten fås $P(S) = 0.001$, $P(T|S) = 0.99$ samt $P(T|S^c) = 0.005$, där S^c är komplementhändelsen till sjuk, dvs frisk.

   (a) Sannolikheten att en slumpmässig person, frisk eller sjuk, testar postitivt är enligt satsen om total sannolikhet (formelsamlingen) $P(T) = P(T|S) \cdot P(S) + P(T|S^c) \cdot (1 - P(S)) = \cdots = \underline{0.006}$, då (per definition) händelserna S och S^c är disjunkta och fyller hela utfallsrummet. (0.5)

   (b) Den omvända betingningen sjuk givet positiv, $P(S|T)$kan fås genom Bayes sats (formelsamling): $P(S|T) = \frac{P(T|S) \cdot P(S)}{P(T)} = \cdots = \underline{0.165}$, där $P(T)$ beräknats i (a). Sannolikheten för s.k. falskt positiv är alltså ganska stor, i genomsnitt är ca 5 av 6 som testar positivt inte sjuka. (0.5)

2. Efter Lottodragningen finns det 7 bollar (= nummer) som är rätt, och $35 - 7 = 22$ nummer som är fel, där ordningen inte är viktig. Låt ξ vara antal rätt en spelare har, där möjliga utfall är heltalen från 0 till 7. Eftersom bollarna dras utan återläggning är sannolikheterna att få rätt nummer i varje gissning beroende.

   (a) *Metod 1*: $p = P(\text{första numret rätt}) \cdot P(\text{andra numret rätt}) \times P(\text{tredje numret rätt}) \times \cdots \times P(\text{sjunde numret rätt}) = \frac{7}{35} \times \frac{6}{34} \times \frac{5}{33} \times \cdots \times \frac{1}{29} = \frac{7!}{\frac{35!}{28!}} = 0.000000149 = \underline{0.149 \cdot 10^{-6}}$
   Metod 2: Problemet kan ses som en urnmodell med kulor av två sorter, 'rätt' och 'fel'. Antal rätt följer då den hypergeometriska fördelningen, $\xi \in Hyp(N, n, q)$, där $N = 35, n = 7, q = \frac{7}{35}$. Formelsamlingen ger sannolikhetsfunktionen för $k = 7$ rätt som: (0.5)

$$p = P(\xi = 7) = \frac{\binom{7}{7}\binom{28}{0}}{\binom{35}{7}} = \frac{1 \cdot 1}{\frac{35!}{28! \cdot 7!}} = \frac{28! \cdot 7!}{35!} = \underline{0.149 \cdot 10^{-6}}$$

   (b) Händelsen att vinna på Lotto inträffar med sannolikhet p i varje försök. Spelaren gör n försök att vinna och försöken antas oberoende av varandra. Antal vinster på Lotto, η är därför binomialfördelat, $\eta \in Bin(2548, 0.149 \cdot 10^{-6})$. Sannolikheten för minst en vinst blir därför $P(\eta \geq 1) = 1 - P(\eta = 0) = [\text{formels.}] = 1 - \binom{2548}{0} p^0 \cdot (1-p)^{2548} = \underline{0.00038}$ (0.38 ‰), alternativt 0.0025 för $p = 10^{-6}$. Man kan konstatera att det är väldigt låg sannolikhet att vinna på Lotto, även för den trogne spelaren. (0.5)

Var god vänd!

3. Vi får

   (a) via standardisering $P\left(\frac{\overline{\xi}-\mu}{\sigma/\sqrt{n}} \leq \frac{\mu+1.65\sigma/\sqrt{n}-\mu}{\sigma/\sqrt{n}}\right) = \Phi(1.65) = 0.9505.$ (0.4)

   (b) Medelvärdet används som skattning av μ, med observation $\overline{x} = 106.975$. Eftersom σ är känd får konfidensintervallet en normalfördelningskvantil. Konfidensintervallet för förväntad forsforhalt μ blir således (formelsamling) $I_\mu = \left[\overline{x} - \lambda_\alpha \frac{s}{\sqrt{n}}\,,\ \infty\right] = \left[106.975 - 1.6449 \cdot \frac{16}{\sqrt{4}}\,,\ \infty\right] = [\,93.81\,,\ \infty\,]$ för $\alpha = 0.05$. Eftersom intervallet är ensidigt har hela α lagts i den nedre svansen. (0.5)

   (c) Eftersom konfidensintervallet I_μ innehåller $\mu = 100$ går det inte att påvisa att sjön skulle vara övergödd. (0.1)

4. Tabellen visar hur många syskon en person har, medan ξ räknar antal syskon inklusive personen som besvarat enkäten. Undersökningen tar dessutom inte hänsyn till familjer som inte har barn.

   (a) Utfallsrummet för antal barn i en familj kan antas vara $\Omega = \{1, 2, \ldots\}$, dvs minst ett barn, men utan övre gräns. (0.2)

   (b) Sannolikhetsfunktionen kan skattas som $P(\xi = k)^* = \frac{n_{k-1}}{n}$, för $k = 1, 2, \ldots$, där $n = 65$, vilket ger: (0.3)

k	1	2	3	4	5	6	7	8	9 och högre
$P(\xi = k)^*_{obs}$	0.03	0.38	0.25	0.18	0.06	0.05	0.02	0.03	0

   (c) $P(\text{hela familjen får plats}) = P(2 + \xi \leq 5) = P(\xi \leq 3) =$
   $P(\xi = 1) + P(\xi = 2) + P(\xi = 3) = 0.66.$ (0.2)

   (d) Förväntat antal syskon blir (formelsamling) $E(\xi) = \sum_{k=1}^{8} k \cdot P(\xi = k) = 3.22$ barn per familj. Som jämförelse kan nämnas att i Sverige föds i genomsnitt ca 2.1 barn per kvinna. Detta inkluderar även kvinnor som inte får barn, vilket inte kunnat undersökas här. (0.3)

5. Detta är ett problem av typen oberoende stickprov. Det ena stickprovet innehåller uppmätta densiteter av furu och betecknas $\xi_i \in N(\mu_1, \sigma)$ för $i = 1, \ldots, 10$. Det andra stickprovet är motsvarande för gran och betecknas $\eta_j \in N(\mu_2, \sigma)$ för $j = 1, \ldots, 8$.
   Vi misstänker att det finns skillnad i densitet mellan dessa träslag, och ställer upp nollhypotes $H_0 : \mu_1 - \mu_2 = 0$ (ingen skillnad) och mothypotes $H_1 : \mu_1 - \mu_2 \neq 0$ (pos. eller neg. skillnad). Som testvariabel använder vi (såsom i formelsamlingen) skattningen $\mu_1^* - \mu_2^* = \overline{\xi} - \overline{\eta}$, dvs skillnad mellan stickprovens medelvärden. Vi väljer teststrategin konfidensintervall (då denna finns i formelsamlingen) och börjar med att beräkna observerat medel och varians för respektive stickprov. Vi får:

$$\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 482.7, \quad s_1^2 = \frac{1}{9}\left(\sum_{i=1}^{10}(x_i^2) - 10 \cdot \overline{x}^2\right) = 815.5667, \quad \text{från tabellens första rad,}$$

$$\overline{y} = \frac{1}{8}\sum_{i=1}^{8} y_i = 447.625, \quad s_2^2 = \frac{1}{7}\left(\sum_{i=1}^{8}(y_i^2) - 8 \cdot \overline{y}^2\right) = 1291.9821, \quad \text{från tabellens andra rad,}$$

   vilket ger den poolade standardavvikelsen $s_p = \sqrt{\frac{9 \cdot s_1^2 + 7 \cdot s_2^2}{16}} = 32.0000$. Vi kan därmed beräkna konfidensintervallet för fallet okänd standardavvikelse (formelsamling) på nivån $\alpha = 0.05$:

$$\begin{aligned} I_{\mu_1-\mu_2} &= \overline{x} - \overline{y} \pm t_{0.025}(16) \cdot s_p\sqrt{\frac{1}{10} + \frac{1}{8}} = [\,t_{0.025}(16) = 2.12 \text{ enl. tabell}\,] \\ &= 35.075 \pm 32.179 = [\,2.89\,,\ 67.26\,]\,. \end{aligned}$$

Eftersom värdet under H_0, $\mu_1 - \mu_2 = 0$, inte ligger i konfidensintervallet kan vi förkasta H_0 på nivån $\alpha = 0.05$ och därmed med fog påstå att det finns skillnad i densitet mellan träslagen. (1.0)

6. Taklampornas livslängder är alla oberoende och $\xi \in Exp(\lambda)$, där $\lambda = \frac{1}{30\,000}$. Fördelningsfunktionen för exponentialfördelningen beräknas (med frekvensfunktionen från formelsamlingen)

$$F(x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_{t=0}^{x} = 1 - e^{-\lambda x}, \ x \geq 0.$$

   (a) Den kortaste livslängden är $\eta = \min(\xi_1, \ldots, \xi_n)$, vars fördelningsfunktion ges i formelsamlingen

   $$F_\eta(x) = 1 - (1 - F(x))^n = 1 - (1 - 1 + e^{-\lambda x})^{50} = 1 - e^{-50\lambda x}, \ x \geq 0,$$

   vilket innebär att den kortaste livslängden är också exponentialfördelad, $\eta \in Exp(\kappa)$, där $\kappa = \frac{50}{30\,000} = \frac{1}{600}$, dvs med förväntad livslängd 600 istället för 30 000 timmar. (0.7)

   (b) Vi vet från vaktmästarens observation att den kortaste livslängden var 120 timmar eller kortare, då vi inte vet exakt när den gick sönder. Sannolikheten för vaktmästarens observation är $P(\eta \leq 120) = F_\eta(120) = 1 - e^{-\frac{120}{600}} = 0.181$. Detta är inte något extremt utfall. Om vi hade formulerat ett test för om förväntad livslängd är 30 000 timmar hade 0.181 varit vårt p-värde och att förkasta en sådan nollhypotes hade inte varit trovärdigt. (0.3)

Lycka till!