Linköpings universitet
Institutionen för teknik och naturvetenskap
Michael Hörnquist

Statistik och sannolikhetslära, TNIU66

Tentamen, 4 juni 2018, Svar och lösningsskisser

Nedanstående är svar, med viss diskussion och antydningar till hur uppgifterna *kan* lösas. Det är *inte* nödvändigtvis fullständiga lösningar.

1. Direkt räkning ger att den vanliga stickprovssstandardavvikelsen i första fallet blir $s = 1$ och i andra fallet blir $s = \sqrt{2}$. Peterstandardavvikelen blir i bägge fallen noll (0), så ett svar till Peter skulle kunna vara: ”Käre Peter, din ’Peterstandardavvikelse’ blir exakt noll för två mängder som dels är olika, dels verkligen har spridningar i sina värden. Jag skulle gissa att den blir noll för alla mängder, men det har jag ännu inte visat.”

2. (a) Likformig fördelning, dvs $p = 1/30 = 3,3\%$.
   (b) Komplementhändelser och oberoende ger $P(\text{minst en}) = 1 - P(\text{ingen}) = 1 - P(\text{inte första}) \cdots P(\text{inte sjätte}) = 1 - (1-p)^6 = 18,4\%$.
   (c) Låt den stokastiska variabeln X vara antalet gånger av sex Eva behöver redovisa, dvs antalet ”lyckade” försök av sex möjliga, med samma sannolikhet p att ”lyckas” i vart och ett av dem och oberoende. Då blir $X \in \text{Bin}(6, p)$. och $P(X = 2) = \binom{6}{2} p^2 (1-p)^4 = 1,5\%$.
   (d) Eftersom varje omgång är oberoende av de andra ges sökt sannolikhet på samma sätt som i (b), fast för fem omgångar. $P(\text{minst en av fem}) = 1 - (29/30)^5 = 15,6\%$.

3. (a) Direkt avläsning ur grafen vid den vertikala linjen för 5 år ger medel $\mu = 19,6$ kg och standardavvikelse $\sigma = 2,4$ kg (behöver inte ges rätt på decimalen, kan vara svårt att avläsa).
   (b) Andelar är sannolikheter, så med X som stokastisk variabel för vikten hos en godtycklig femårig pojke i Sverige söker vi $P(X < 15)$ då $X \in \text{N}(19,6; 2,4)$. Vi får $P(X < 15) = P((X-\mu)/\sigma < (15-\mu)/\sigma) = P(Z < -1,92) = 2,7\%$, där $Z \in \text{N}(0, 1)$ och tabell D3 i Løvås har nyttjats.
   (c) Här söker vi gränsen k så att $P(X > k) = 0,01$, dvs en procent av pojkarna ska väga mer än k kg. Vi får $0,01 = P(X > k) = P((X-\mu)/\sigma > (k-\mu)/\sigma) = P(Z > z_{0,01})$, och avläser ur tabell D4 $z_{0,01} = 2,326$. Alltså $(k-\mu)/\sigma = z_{0,01} \Leftrightarrow k = \mu + z_{0,01}\sigma = 25,2$ kg.

4. (a) Vikter tillhör sådant som brukar vara normalfördelat, vilket vi antar att även dessa är. Eftersom standardavvikelsen är känd från mätbladet, $\sigma = 5$ g, kan vi använda normalfördelningen för skattningen av konfidensintervall för medelvärde, $[\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \ \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}]$. Med $\alpha = 0,01 (\Leftrightarrow 1 - \alpha = 0,99)$ fås ur tabell D4 $z_{0,01/2} = 2,576$. Stickprovets storlek är $n = 5$ och dess medelvikt $\bar{x} = 2383,8$ kg. Konfidensintervallet blir $[2378,0; 2389,6]$ kg.
   (b) Nu är inte standardavvikelsen att betrakta som känd längre, utan skattas med stickprovsstandardavvikelsen (samma formel som i uppgift 1). Vi får $s = 1,3$ g. Vi har därför ett konfidensintervall för medelvärde som ges av $[\bar{x} - t_{\alpha/2}s/\sqrt{n}, \ \bar{x} + t_{\alpha/2}s/\sqrt{n}]$. Ur tabell D5 och $n - 1 = 4$ frihetsgrader fås $t_{0,01/2} = 4,604$. Konfidensintervallet blir $[2381,1; 2386,5]$ kg.

5. (a) Falskt. Signifikansnivån är sannolikheten att fälla en oskyldig, dvs att förkasta en sann nollhypotes.
   (b) Sant. Se (a).
   (c) Sant. Styrkefunktionen, γ, ger alltid sannolikheten att förkasta nollhypotesen, och om man är i området att nollhypotesen är falsk blir godtagandefelet (sannolikheten att *inte* förskasta en falsk nollhypotes, β) lika med $1-\gamma$.
   (d) Falskt. Totalt nonsens.
   (e) Sant. Notera att p-värdet är sannolikheten för ett visst utfall (det utfall man fick, ifall man hade beräknat det innan experimentet gjordes), och att det beror av vilken nollhypotes som har valts.
   (f) Falskt. Det är aldrig sannolikheten att en hypotes är sann eller falsk som beräknas. Det vore i och för sig en mycket relevant storhet för ett försök, men den klarar vi inte av att bestämma på ett rimligt sätt.

6. (a) Korrelstionskoefficienten i Excels utdatasammanfattning kallas för "Multipel-R" och avläses till 80%.
   (b) Här efterfrågas ett prediktionsintervall för en framtida observation, dvs vi behöver använda sambandet

$$\left[\hat{\alpha}+\hat{\beta}x-s\,t_{\alpha/2}\sqrt{1+\frac{1}{n}+\left(\frac{x-\bar{x}}{s/\mathrm{SE}(\bar{\beta})}\right)^2},\ \hat{\alpha}+\hat{\beta}x+s\,t_{\alpha/2}\sqrt{1+\frac{1}{n}+\left(\frac{x-\bar{x}}{s/\mathrm{SE}(\bar{\beta})}\right)^2}\right].$$

   Från utdatasammanfattningen kan vi utläsa det mesta som behövs, $\hat{\alpha}=13,3$, $\hat{\beta}=19,8$, $n=12$, $s=2,259$ och $\mathrm{SE}(\hat{\beta})=4,7$. Konfidensnivån var given till 95%, dvs $\alpha=0,05$, vilket med $n-2=10$ frihetsgrader ger ur tabell D5 $t_{0,05/2}=2,228$. Önskad koncentration uppgavs till $x=0,50$. Till sist behövs medelvärdet av samtliga observation av koncentrationer, vilket får räknas ut ur de tolv givna värdena till $\bar{x}=0,487$. Allt detta sammantaget ger konfidensintervallet för värmeutvecklingen i Joule per gram till $[17,96;\ 28,44]$. Nu efterfrågades inte Joule per gram, dvs för ett gram, utan total värmeutveckling för 10 kg = 10 000 gram. Multiplicera därför intervallets ändpunkter med 10 för att erhålla svaret [180, 284] kJ.