Linköpings universitet
Institutionen för teknik och naturvetenskap
Michael Hörnquist

Statistik och sannolikhetslära, TNIU66

Tentamen, 22 augusti 2020, Svar och lösningsskisser

Nedanstående är svar, med viss diskussion och antydningar till hur uppgifterna *kan* lösas.

1. (a) Sant. Om det är ett jämt antal värden i mängden är medianen fortfarande väldefinierad som medelvärdet av de två mittersta värdena (när de är ordnade efter storlek).
   (b) Falskt. Med ett jämt antal värden i mängden är medianen medelvärdet av de två mittersta, och det behöver inte vara ett heltal. Exempelvis mängden $\{1, 2, 3, 4\}$.
   (c) Sant, men det kan finnas mer än ett typvärde i en mängd. Om alla element är olika stora blir alla förekommande värden typvärden, varvid begreppet blir helt meningslöst. Därför godkänns även svaret ”Falskt” på denna uppgift med en god motivering.
   (d) Sant, typvärdet är ju det vanligaste värdet.
   (e) Sant. Finns ingenting i formeln som hindrar uträkningen.
   (f) Sant, även om ”de flesta” standardavvikelser inte är heltal. Enklaste exemplet på när det är ett heltal är förstås när alla element är lika, $\{1, 1, 1, 1\}$. Mindre banalt exempel på stickprovs-standardavvikelse är $\{1, 2, 3\}$.

2. Låt X vara slumpvariabeln som anger antal extrema regn under ett enskilt år, och Y slumpvariabeln motsvarande för tio år. Med intensiten $\lambda = 0,25$ extrema regn per år blir $X \sim \mathrm{Po}(\lambda \cdot t) = \mathrm{Po}(0,25)$ då $t = 1$ år och $Y \sim \mathrm{Po}(\lambda \cdot t) = \mathrm{Po}(2,5)$ då $t = 10$ år.
   (a) $\Pr(X = 0) = (0,25^0/0!)e^{-0,25} = 0,7788 \approx 78\%$
   (b) $\Pr(X > 1) = 1 - \Pr(X = 0) - \Pr(X = 1) = 1 - (0,25^0/0!)e^{-0,25} - (0,25^1/1!)e^{-0,25} = 0,0265 \approx 2,6\%$
   (c) $\Pr(Y = 0) = (2,5^0/0!)e^{-2,5} = 0,0821 \approx 8,2\%$

3. Min gissning på medelvärde och standardavvikelse för en medlem i Logistisksektionen är $\mu = 65$ kg respektive $\sigma = 8$ kg. Det betyder att 68% av studenterna väger mellan 57 och 73 kg och att 95% av dem väger mellan 49 och 81 kg. Med $X_i \sim N(\mu = 65; \sigma = 8)$ som vikten hos *en* student, blir summan av alla 14 studenter $X = \sum_{i=1}^{14} X_i \sim N(14 \cdot 65; \sqrt{14} \cdot 8) = N(910; 30)$. Den sökta sannolikheten blir därmed $\Pr(X > 1000) = 1 - \Pr(Z < (1000 - 910)/30) \approx 0,13\%$.

4. Här handlar det om konfidensintervall för andelar. Punktskattning $p = 32/40 = 0,8$ (32 av 40 kuber har en hållfasthet som överstiger 30 MPa. Eftersom $np(1 - p) = 40 \cdot 0,8 \cdot 0,2 = 6,4 > 5$ kan den vanliga approximationen användas. Med 95% konfidensnivå bildar vi ett nedåt begränsat konfidensintervall på vanligt sätt, $\pi > p - z_{1-\alpha}\sqrt{p(1 - p)/n}$, vilket med $z_{1-\alpha} = z_{0,95} = 1,645$ blir $\approx 70\%$. Slutsats: Mätprotokollet ger inget stöd för påståendet att *minst 80% av testkuberna lever upp till kravet 30 MPa.*

5. (a) Med den enkelsidiga alternativhypotesen $H_a : \pi > 0,20$ och en stickprovsstorlek $n = 500$ får vi kravet på andelen vinstlotter p ur sambandet $(p - \pi_0)/\sqrt{\pi_0(1-\pi_0)/n} > z_{1-\alpha}$, dvs $p > \pi_0 + z_{1-\alpha}\sqrt{\pi_0(1-\pi_0)/n} = 22,9\%$ då $z_{1-\alpha} = z_{0,95} = 1,645$ och $\pi_0 = 20\%$.

   (b) Att visa att ett visst värde är det sanna värdet är inte görligt, men det går att ringa in det med ett konfidensintervall. Lotteriinspektionen skulle kunna skapa ett litet konfidensintervall baserat på *många* lotter, och se om värdet 20% täcks av intervallet. De skulle också kunna ställa upp en hypotesprövning för att se om den sanna andelen vinstlotter är *mindre* än något givet tal bara lite större än 20%, exempelvis 21%. Då bevisar de i och för sig inte att den sanna andelen är 20%, men att den åtminstone inte är större än 21% (i princip, det gäller ju också att utfallet blir rätt).

6. (a) Temperatur: $\hat{y}_{28} = b_0 + b_1 x^* = 7,1088 + 1,2697 \cdot 28 = 42,6604 \approx 43$ kg Tryck: $\hat{y}_{1020} = b_0 + b_1 x^* = -282.8605 + 0,3169 \cdot 1020 = 40,3775 \approx 40$ kg.

   (b) Eftersom det är förutsägelse för en enskild dag används prognosintervall (prediktionsintervall enligt Løvås). Enkelsidigt uppåt begränsat ges övre gränsen av uttrycket $b_0 + b_1 x^* + t_{n-2;1-\alpha} s \sqrt{1 + (1/n) + (x^* - \bar{x})^2 / \sum_{i=1}^{n}(x_i - \bar{x})^2}$. Värdet på $\sum_{i=1}^{n}(x_i - \bar{x})^2$ kan beräknas direkt ur data givet i tabell 1, men kan även fås som $(n-1)s_x^2$, där s_x är stickprovsstandardavvikelsen för x-värdena (Wahlin) eller som $(s/\mathrm{SE}(\beta_1))^2$ (Løvås). Med $t_{n-2;1-\alpha} = t_{5;0.95} = 2,015$ ur tabell blir mängden glass för förutsägelsen baserad på temperatur ≈ 57 kg och baserad på lufttryck $48,9616 \approx 49$ kg.

   (c) Förklaringsgraden läses av direkt i databladen som ”R-kvadrat”, vilken är 60,5% för temperatur som förklaringsvariabel och 85,5% för tryck.