Linköpings universitet
Institutionen för teknik och naturvetenskap
Michael Hörnquist

Statistik och sannolikhetslära, TNIU66

Tentamen, 21 augusti 2021, Svar och lösningsskisser

1. Total årsinkomst före Gates inflyttning är $16\,747 \cdot 270\,100 = 4\,523\,364\,700 = 4,523$ miljarder kr. Gates årsinkomst skattas till $12 \cdot 35 \cdot 10^9 = 420$ miljarder kronor. Ny medellön blir alltså $(420 + 4,5)/(1 + 16747) = 25,3$ miljoner kr. Medianlönen påverkas (i stort sett) inte av extremvärden, varför den kvarstår på 247 237 kr.

2. (a) $13/52 = 25,0\%$. (b) $(13 - 1)/(52 - 1) = 12/51 \approx 23,5\%$ (Kan även lösas med betingad sannolikhet, men det tillför ingenting.) (c) Låt A vara händelsen ”översta kortet en hjärter” och B händelsen ”näst översta kortet en hjärter”. Då blir enligt multiplikationssatsen för beroende händelser $\Pr(A \cap B) = \Pr(A)\Pr(B|A) = (12/51) \cdot (12 - 1)/(51 - 1) = 132/2550 \approx 5,2\%$.

3. (a) Vi har slumpvariabel $X \sim N(\mu = 100, \sigma = 15)$. $\Pr(X > 120) = \Pr(Z > (120 - 100)/15) = 1 - \Pr(Z < 1,33) = 0,09176$. (b) Nu har vi X_1, X_2, X_3 oberoende slumpvariabler fördelade på samma sätt som X ovan. Då blir $Y = (X_1 + X_2 + X_3)/3 \sim N(\mu_Y = \mu, \sigma_Y = \sigma/\sqrt{3}) = N(\mu_Y = 100, \sigma_Y = 8,66)$. Då blir $\Pr(Y > 120) = \Pr(Z > (120 - 100)/8,66) = 1 - \Pr(Z < 2,31) = 0,01044$. Jämför gärna denna uppgift med uppgift 3 på tentan i juni 2021.

4.
   (a) Falskt. Större stickprov ger *kortare* konfidensintervall.
   (b) Falskt. Konfidenintervallet för medellängd gäller hur säkra vi är på det skattade medelvärdet, inte på hur hela populationen ser ut.
   (c) Falskt. Bara konstigt, en typisk konfidensnivå är 95%, den kan inte bli dubbelt så stor.
   (d) Sant. Självklart blir vi säkrare (får större konfidens) på att intervallet täcker det sanna värdet om vi gör det större.
   (e) Falskt. Jämför (d) ovan.
   (f) Sant. Se exempelvis t-tabellen i Appendix B i Wahlin.

5. Det är alltid alternativhypotesen som statistiskt bevisas, så här måste H_a vara ”medelrestiden *mindre än* 30 minuter”. Nollhypotesen H_0 blir då ”medelrestiden är (mer än) 30 minuter”. Vi antar att restiden för varje enskild dag är oberoende av de andra tiderna och att tiderna är dragna från en normalfördelning. Testvaribeln blir $t = (\bar{x} - \mu_0)/(s/\sqrt{n}) = (27,0 - 30,0)/(3,16\ldots/\sqrt{5}) \approx -2,12$. Detta ska jämföras med det kritiska värdet från t-fördelningen, $t_{5-1;0,10} = -1,53$. Eftersom t är *mindre än* det kritiska värdet betyder det att testvaribeln hamnar i det kritiska området och att H_0 kan förkastas. Alternativhypotesen H_a är därmed statistiskt bevisad (på signfikansnivån 10%) och Margareta kan hävda att hennes medelrestid till jobbet är mindre än en halvtimme.

6.
   (a) Korrelationskoefficienten $r =$ ”Multipel-R” $= 0,99$
   (b) Förklaringsgraden $r^2 =$ ”R-kvadrat” $= 0,98$
   (c) $\hat{y}_x = -2373,4 + 0,3412x$ med värden från ”Konstant” och ”X-variabel 1”.
   (d) $0,303 \leq \beta_1 \leq 0,379$ från uttrycket $b_1 \pm t_{n-1;1-\alpha/2}\mathrm{SE}(\beta_1)$.