Linköpings universitet
Institutionen för teknik och naturvetenskap
Michael Hörnquist

Statistik och sannolikhetslära, TNIU66

Tentamen, 8 juni 2020, Svar och lösningsskisser

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

1. (a) Alla löner ökar med samma faktor, det innebär att också medelvärdet, medianen, typvärdet och standardavvikelsen ökar med den faktorn, 1,025. Vi får (avrundat till heltal) medelvärdet=28 667·1,025=29 384:-, medianen=29 000·1,025=29 725 (den 14:e högsta/lägsta lönen), typvärdet=32 000·1,025=32 800:- och standardavvikelsen=3702·1,025=3795:-. (b) När alla löner ökas med samma värde, 800:-, så förflyttas sig alla staplar uppåt längs x-axeln lika mycket, vilket innebär att spridningen inte ändras. Det innebär att medelvärdet=29 384+800=30 184:-, medianen=29 725+800=30 525:-, typvärdet=32 800+800=33 600:- och standardavvikelsen=3795:- oförändrad.

2. Här handlar det om multiplikationsprincipen (Wahlin), också känd som produktregeln (Løvås). För var och en av de två stavningarna av s-ljudet, ”s” eller ”z”, finns två stavningar av i-judet, ”i” eller ”ii”. Det betyder att om det bara funnes dessa varianter skulle vi ha två *gånger* två, dvs fyra, varianter av stavning av vårt namn. Av en ren tillfällighet, kära Lisa, råkar du få samma svar när du ”plussar” antalet varianter här. Men nu har vi för var och en av dessa fyra varianter också två varianter av a-ljudet, ”a” och ”ah”, vilket innebär att totala antalet varianter blir fyra *gånger* två, dvs åtta varianter av stavning.

   Nåväl, säger Lisa, du har säkert rätt, men det är ändå alltid jag som är ”esset” av oss två.

3. (a) Tärningens utfall kan beskrivas med en diskret slumpvariabel X med en likformig sannolikhetsfördelning. Väntevärdet blir då $\mu = E(X) = (1 + 60)/2 = 30,5$.

   (b) För en diskret slumpvariabel gäller $\sigma = \sqrt{\sum_i (x_i - \mu)^2 \mathrm{Pr}(X = x_i)}$. Med X som likformig diskret slumpvariabel med värden mellan 1 och 60 blir det $\sigma = \sqrt{\sum_{i=1}^{60} (i - 30,5)^2/60}$.

   (c) $\mathrm{Pr}(\mu - \sigma \leq X \leq \mu + \sigma) = \mathrm{Pr}(13,18 \leq X \leq 47,82) = \mathrm{Pr}(14 \leq X \leq 47) = 34/60 \approx 0,567$.

   (d) $\mathrm{Pr}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = \mathrm{Pr}(-4,14 \leq X \leq 65,14) = \mathrm{Pr}(1 \leq X \leq 60) = 1$.

   (e) Utfallen är heltal mellan 1 och 60, så $\mathrm{Pr}(X = \mu) = 0$.

   (f) Utfallsrummet vid två kast med en D60 är alla kombinationer av utfall, (1,1), (1,2),...,(60,60), dvs totalt $60 \cdot 60 = 3600$ stycken. Gynnsamma utfall är (1,60),(2,59),...,(60,1), dvs totalt 60 stycken. Det ger sannolikheten att få väntevärdet som utfall är $60/3600 = 1,67\%$.

4. (a) Låt X vara en slumpvariabel som beskriver mängden läsk som fylls i *en* flaska, $X \sim N(\mu; \sigma = 1,1)$ (allt i cl om inget annat anges). Det är μ som är sökt. Givet är att $\Pr(X > 35) = 0,001$. Vi får $0,001 = \Pr(X > 35) = 1 - \Pr(Z < (35 - \mu)/1,1))$ och ur normalfördelningstabell (Appendix B Wahlin) fås $(35 - \mu)/1,1 = 3,09$, vilket ger $\mu = 31,6$ cl.

   (b) Låt Y vara slumpvariabel för den totala mängd läsk som fylls på flaska under en dag. Eftersom påfyllningen av varje enskild flaska är oberoende av varandra har vi då $Y \sim N(\mu_Y = 238 \cdot 31,6; \sigma_Y = \sqrt{238} \cdot 1,1)$. Sökt är en volym läsk V så att $\Pr(Y \leq V) = 0,95$. Vi har $0,95 = \Pr(Y \leq V) = \Pr(Z \leq (V - 7520,8)/16,97)$, och ur normalfördelningstabell (Appendix B Wahlin) fås $(V - 7520,8)/16,97 = 1,645$, vilket ger $V = 7548,7$ cl ≈ 76 liter. [Du som räknar med väntevärde 33 cl för flaskorna bör få 79 liter.]

5. Med $p = 0,25$ och $n = 100$ blir $np(1 - p) = 18,75 > 5$ och därmed kan den vanliga testvariabeln $z = (p - \pi_0)/\sqrt{\pi_0(1 - \pi_0)/n} = 1,25$ användas. Detta jämförs med z-värde på någon signifikansnivå, exempelvis 1%; 5% eller 10%. Motsvarande kvantiler från standardnormalfördelningen är $z_{0,99} = 2,326$, $z_{0,95} = 1,645$ och $z_{0,90} = 1,282$. I inget av fallen kan nollhypotesen H_0 förkastas.

   (a) Konsumentverket verkar kunna sin hypotesprövning, hypoteserna som ställdes upp skulle i princip kunna ha besvarat deras oro och visat att lotterna som såldes i kiosken hade en högre andel vinster än förväntat. Nu var emellertid andelen vinster, 25%, alltför liten för att en sådan slutsats skulle kunna dras, ens på signifikansnivån $\alpha = 10\%$. Alltså, slutsatsen för Konsumentverket är att det inte är bevisat att påståendet är sant (men inte heller att det är falskt). Ingen slutsats, med andra ord. Trist, men så kan hypotestest lätt sluta.

   (b) Lotteriinspektionen är mer illa ute. Där borde man ha reagerat direkt på försöksuppställningen, då den inte ens i princip kan leda till önskad slutsats. Det går aldrig att statistiskt bevisa en nollhypotes, bara att notera att den inte går att förskasta. Och det är precis det som hände här, det är inte bevisat att andelen vinstlotter är högre än förväntat, men det innebär inte att allt är som det ska. Framför allt innebär det inte att man har bevisat att $\pi = 0,20$. Slutsatsen för Lotteriinspektionen är att deras försök att påvisa att påståendet är *falskt* inte ens i princip hade varit möjligt med de här hypoteserna.

6. (a) Falskt. Riktningskoefficienten i regressesionsanalys visar brantheten, korrelationskoefficientens tecken visar i sig om det är positiv elller negativ lutning, men dess belopp har ingenting med brantheten att skaffa, se (b) nedan.

   (b) Sant. Det är exakt vad korrelationskoefficienten uppskattar, där värdet ± 1 anger att punkterna ligger exakt på en rät linje, se även (a) ovan för tecknet.

   (c) Falskt. Kan omöjligt vara sant, då korrelationskoefficienten även kan vara negativ. Kvadraten på korrelationskoefficienten är det som brukas kallas ”förklaringsgrad” och mäter andel av variationen som förklaras av den räta linjen.

   (d) Falskt. Konfidensintervall för medelvärden, andelar, etc., dvs för endimensionella datamängder har den angivna egenskapen, men för tvådimensionella datamängder är det inte lika enkla samband. Se exempelvis Wahlin sid. 269 (Løvås sid. 279).

   (e) Sant. Andragradspolynom har tre anpassningsparametrar, där två av dem återfinns även för en rät linje. Det innebär att det aldrig kan bli sämre anpassning med ett andragradspolynom.

   (f) Sant. Formeln på sid. 256 i Wahlin (sid. 274 i Løvås) går alltid att räkna ut, oavsett hur dålig anpassningen är (om någon har svarat att det inte går för det fall att alla x_i-värden är lika, så är det också ok – det är sant, men ett totalt urartat fall som aldrig lär inträffa i praktiken).