

Statistik och sannolikhetslära, TNIU66

Tentamen, 24 augusti 2019, Svar och lösningsskisser

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

1. Medelvärde fås som $(29 \cdot (-1) + 14 \cdot 0 + 2 \cdot 1 + 4 \cdot 2 + 1 \cdot 3)/50 = -0,32$. Median är mittenvärdet, och eftersom mer än hälften av talen är -1 är det också medianvärdet.
2. (a) $\Pr(X = 0) = 0,424$, $\Pr(X = 3) = 0,0408$, $\Pr(X = 25) = 0$. Det sista resultatet inses ur förutsättningarna för binomialfördelningen, i detta fall att det är 21 delförsök.
(b) $\Pr(Y = 0) = 0,438$, $\Pr(Y = 3) = 0,0426$, $\Pr(Y = 25) = 3,56 \cdot 10^{-28}$
(c) 1,74%, 4,46%, samt odefinierat (eller oändligt stort). Att approximationen ändå anses acceptabel beror på att värdet från Poissonfördelningen är så pass litet att det för alla praktiska ändamål kan anses vara noll det också.
3. Låt X vara slumpvariabeln som beskriver volymen ketchup (och minns att sannolikheter även är andelar).
(a) Givet i uppgiften är $X \sim N(\mu = 500; \sigma = 15)$. Då blir $\Pr(X < 490) = \Pr(Z < -0.67) = 0,25143 \approx 25\%$
(b) Låt nu $X \sim N(\mu; \sigma = 15)$, μ okänt. Vi söker värdet på μ så att $\Pr(X > 500) = 0,99$. Då blir $0,01 = \Pr(X < 500) = \Pr(Z < (500 - \mu)/15) = \Pr(Z < z_{0,01})$, dvs $\mu = 500 - z_{0,01}15 \approx 534$ gram (med $z_{0,01} = -2,326$ ur tabell).
4. Stickprovets storlek är $n = 5414$ (antalet tillfrågade är irrelevant sålänge bortfallet inte är systematiskt) medan populationens storlek är ovidkommande (sålänge den är substantiellt större än stickprovet). Eftersom $np(1-p) = 826,5 > 5$ går den vanliga normalapproximationen bra att använda. Testvariabeln blir i både (a) och (b) $z = (p - \pi_0)/\sqrt{\pi_0(1 - \pi_0)/n} = -1,8465$, med $\pi_0 = 19,8\%$, valresultatet.
(a) Låt $H_0 : \pi = \pi_0$ och $H_a : \pi \neq \pi_0$. Kritiskt område är $z < z_{\alpha/2} = -1,96$ och $z > z_{1-\alpha/2} = 1,96$. Här är $z = -1,8465$ och ligger alltså *inte* i det kritiska området. Slutsatsen är att nollhypotesen inte kan förkastas, dvs att opinionsundersökningen inte ger någon grund för att säga att väljarstödet har *ändrats*.
(b) Låt $H_0 : \pi \geq \pi_0$ och $H_a : \pi < \pi_0$. Kritiskt område är $z < z_\alpha = -1,645$. Här är $z = -1,8465$ och ligger alltså i det kritiska området, vilket gör att nollhypotesen kan förkastas. Slutsatsen är att mothypotesen är statistiskt bevisad på signifikansnivån 5%, dvs att opinionsundersökningen visar att väljarstödet har *minskat*.

Det kan tyckas paradoxalt att vi inte kan säga om stödet har *ändrats*, men att vi kan säga att det har *minskat*. Orsaken är att vi i det senare fallet har formulerat frågan mer specifikt, vilket gör att det inte krävs lika mycket av undersökningen för att påvisa det vi önskar. Att göra som vi gör här är dock ett visst missbruk av hypotesprövningsmetodiken, att göra undersökningen om *minskningen* är statistiskt säkerställd *efter* att vi sett att stickprovsandelen är mindre än valresultatet, är att använda samma data två gånger. Det är ett vanligt förfarande, men egentligen otillåtet.

5. (a) Falskt. p -värdet erhålls alltid *efter* ett experiment, medan styrkan kan räknas ut förväg.
 (b) Falskt. Det är aldrig sannolikheten för en hypotes som beräknas.
 (c) Falskt. Det är aldrig nollhypotesen som bevisas.
 (d) Falskt. Ett test kan aldrig visa att nollhypotesen är sann (som mest att den inte kan förkastas).
 (e) Sant.
 (f) Sant.

6. (a) Regressionslinjen är $y = 4,21 - 0,263x$. Med $x = 2,0\%$ blir $y = 3,7\%$.
 (b) Konfidensintervall för genomsnittsnivå ges av formeln på sidan 269 i Wahlin (sidan 279 i Løvås). Alla storheter utom $\bar{x} = 6,35\%$ finns i utdatasammanfattningen ($\bar{x}$ får tas fram med miniräknare). Notera att $\sum_i (x_i - \bar{x})^2 = s/SE(b_1)$. Intervallet blir $[1,6\%, 5,8\%]$.
 (c) Visst är korrelationskoefficienten negativ, samma belopp som Multipel-R anger fast med minustecken. Det här är en defekt hos Dataanalysverktyget i Excel, som jag inte har sett påtalas någonstans.

Hela regressionsanalysen här står på skakig grund. Man kan se att konfidensintervallet för lutningsparametern b_1 omfattar värdet noll, dvs det är tveksamt om det visar på något samband mellan arbetslöshet och inflation överhuvud taget.