Linköpings universitet
Institutionen för teknik och naturvetenskap
Michael Hörnquist

# Statistik och sannolikhetslära, TNIU66

## Tentamen, 25 augusti 2018, Svar och lösningsskisser

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

1. Medelvärdet fås som summan av alla värden multiplicerade med dess respektive andel, dvs $2 \cdot 0,15 + 3 \cdot 0,05 + 4 \cdot 0,45 + 5 \cdot 0,05 + 6 \cdot 0,20 + 7 \cdot 0,10 = 4,4$. Medianvärdet fås som det heltal som ligger närmast den punkt som delar stapelareorna i två lika stora delar, dvs median$= 4$.

2. (a) Att Eva behöver redovisa vid första tillfället ges av likformig fördelning, dvs $p = 1/30 = 3,3\%$, och att inte behöva redovisa är komplementet $1 - p$. Oberoende händelser, sex stycken, gör att sannolikheten att inte behöva redovisa någon gång ges av $(1-p)^6 = 81,6\%$. Alternativt, låt $X$ vara den stokastiska variabel som anger hur många gånger hon får redovisa. Då blir $X \in \text{Bin}(6, p)$ och den sökta sannolikheten ges som $P(X = 0)$.

   (b) Med samma $X$ som i (a), $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1,5\%$ (komplementhändelsen används för att minska antalet beräkningar, notera också att $P(X = 1) = \binom{6}{1} p^1 (1-p)^{6-1}$).

   (c) Fem oberoende försök med sannolikheten att "lyckas" som $p$ i vart och ett av dem, ger sökt sannolikhet som $p^5 = 4,1 \cdot 10^{-8}$.

   (d) Eftersom varje omgång är oberoende av de andra är sökt sannolikhet $p = 3,3\%$.

3. (a) Med $X$ som den stokastiska variabel som ger tidsåtgången blir $X \in \text{N}(\mu, \sigma)$, med $\mu = 19,6$ timmar och $\sigma = 2,4$ timmar. Andelar är sannolikheter, så $P(X < 15) = P((X-\mu)/\sigma < (15-\mu)/\sigma) = P(Z < -1,92) = 2,7\%$, där $Z \in \text{N}(0,1)$ och tabell D3 i Løvås har nyttjats.

   (b) Här söker vi gränsen $k$ så att $P(X > k) = 0,01$, dvs en procent av alla projekt ska vara längre än $k$ timmar. 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 $(k-\mu)/\sigma = z_{0,01} \Leftrightarrow k = \mu + z_{0,01}\sigma = 25,2$ timmar.

4. Stickprovets storlek är $n = 2176$ medan populationens storlek är ovidkommande (sålänge den är substantiellt större än stickprovet).

   (a) Skattad andel $\hat{p} = 0,252$. Eftersom $n\hat{p}(1-\hat{p}) = 410 > 5$ har vi en approximativ normalfördelning och kan använda Regel 6.11 i Løvås. 95% konfidensintervall ger $1 - \alpha = 0,95 \Leftrightarrow \alpha = 0,05$ och tabell D4 $z_{0,05/2} = 1,960$. Konfidensintervallets gränser är $\hat{p} \pm z_{\alpha/2}\sqrt{(\hat{p}(1-\hat{p}))/n}$, vilket med insatta värden blir $[23,4\%,\ 27,0\%]$.

   (b) Skattad andel $\hat{p} = 0,048$. Eftersom $n\hat{p}(1-\hat{p}) = 99 > 5$ har vi en approximativ normalfördelning och konfidensintervallets undre gräns ges av $\hat{p} - z_\alpha\sqrt{(\hat{p}(1-\hat{p}))/n}$. Tabell D4 ger $z_{0,05} = 1,645$, och allt insatt i formeln ger den undre gränsen $4,0\%$. Alltså, med 95% sannolikhet erhåller partiet mer än 4,0% av rösterna och kommer därmed att finnas kvar i riksdagen.

5. Det som ska visas är att medellivslängden är mer än tre år. Låt $\mu_0 = 3$ år $= 26298$ timmar och $\mu$ den sanna (men okända) medellivslängden för lamporna. Sätt därför hypoteserna $H_0 : \mu \leq \mu_0$ och $H_1 : \mu > \mu_0$, så att det blir mothypotesen som är den som vi önskar visa. Med $\bar{x} = 26390$ timmar som uppmätt medellivslängd blir testvariabeln $T = (\bar{x} - \mu_0)/(s/\sqrt{n}) = 1,84$, där $s = 500$ timmar och $n = 100$ har använts. Valet av testvariabel baseras på att antalet mätningar är över 30 och att standardavvikelsen $\sigma$ är okänd (den har skattats som stickprovsstandardavvikelse ur de 100 mätningarna). Vi väljer signifikansnivå till exempelvis $\alpha = 5\%$, varvid gränsen för förkastelseområden fås ur tabell D5 som $t_{0,05} = 1,660$ med 99 frihetsgrader (approximerat ur tabellen för 100 frihetsgrader). Eftersom $T$ ligger i förkastelseområdet, $T > t_{0,05}$ förkastar vi $H_0$ och säger att mothypotesen $H_1$ är statistiskt bevisad på aktuell signifikansnivå. Alltså kan Rania känna sig (hyfsat) trygg i att den sanna medellivslängden hos lamporna är minst tre år. (Med ett mindre $\alpha$, exempelvis 1%, fås slutsatsen att $H_0$ inte kan förkastas, dvs försöket ger ingen information med den säkerheten.)

6. (a) Sant.

   (b) Sant.

   (c) Falskt. Värdet på $\hat{\beta}$ ger ingen information om hur pass bra modellen beskriver datamängden.

   (d) Sant. Men om du erhåller det värdet för en datamängd ur verkligheten bör du noga fundera över om den kan vara tillfuskad.

   (e) Falskt. Se (c).

   (f) Sant. Se (d).