Linköpings universitet
Institutionen för teknik och naturvetenskap
Michael Hörnquist

Statistik och sannolikhetslära, TNIU66

Tentamen, 15 mars 2018, Svar och lösningsskisser

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

1. Direkt räkning utifrån sambandet för medelvärde $\sum$ antal $\cdot$ värde/(totalt antal) ger $(3 \cdot (-2) + 0 \cdot (-1) + 6 \cdot 0 + 4 \cdot 1 + 1 \cdot 2 + 4 \cdot 3 + 1 \cdot 4 + 2 \cdot 5 + 1 \cdot 6)/(3 + 0 + 6 + 4 + 1 + 4 + 1 + 2 + 1) = 32/22 \approx 1,45$. Med 22 element i mängden fås medianen som medelvärdet av talen på position 11 och 12 om de ordnas i storleksordning. Eftersom båda elementen har värdet 1, blir det också medianens värde.

2. (a) Produktregeln ger direkt för Margareta $5 \cdot 8 \cdot 3 = 120$ olika kombinationer
   (b) Eftersom ordningen på fyllningarna är oviktig, men det ska vara två olika, så finns $\binom{8}{2}$ fyllningskombinationer. Totalt antal kombinationer för Eva blir därmed $5 \cdot 28 \cdot 3 = 420$.
   (c) Nu är även ordningen väsentlig, varför antalet fyllningskombinationer blir $8 \cdot 7 = 56$. Totalt antal kombinationer för Thomas blir $5 \cdot 56 \cdot 3 = 840$.

3. (a) Sant. X är hur många försök som lyckas utav 10, och kan inte bli större än 10.
   (b) Falskt. X är hur många händelser som inträffar (exempelvis kunder som kommer) om det sker tio händelser i snitt.
   (c) Falskt. Täthetsfunktionen f för normalfördelningen har $f(x) > 0$ för alla x, så alla intergraler blir positiva. Gäller i *modellen*, som alltså har en brist.
   (d) Falskt. Se (c)
   (e) Falskt. Någon sådan regel finns inte för exponentialfördelningar.
   (f) Falskt. Väntevärden kan adderas, men väntevärdet för exempelvis X här är $E(X) = 1/5$.

4. (a) Vikter tillhör sådant som brukar vara normalfördelat, vilket vi antar att även dessa är. Eftersom standardavvikelsen är beräknad från ett stickprov var den inte känd i förväg, och vi har därför ett konfidensintervall för medelvärde som ges av $[\bar{X} - t_{\alpha/2}S/\sqrt{n}, \bar{X} + t_{\alpha/2}S/\sqrt{n}]$. Med $\bar{X} = 57,5$, $S = 5,0$, $\alpha = 0,05$, $n = 20$ och $t_{0,05/2}(19) = 2,093$ fås $[57,5 - 2,093 \cdot 5,0/\sqrt{20}, 57,5 + 2,093 \cdot 5,0/\sqrt{20}]$, dvs intervallet mellan 55,2 gram och 59,8 gram innehåller medelvikten med 95% säkerhet.
   (b) Intervallängden är omvänt proportionell mot kvadratroten ur antalet mätningar (det står $\sqrt{n}$ i nämnaren i uttrycket för intervallängd $L = 2t_{\alpha/2}S/\sqrt{n}$). Det innebär att ett halverat konfidensintervall kräver fyra gånger fler mätningar, vilket här innebär att man måste väga 80 kulor. Den som vill vara ännu mer noggrann noterar även att värdet på $t_{\alpha/2}$ minskar med ett större stickprov, och att 80 därmed är en övre begränsning på hur stort stickprov som krävs.
   (c) Om Fritz väger stickprovet i klump tapppar han information om variationen, och allt han får fram är en punktskattning av medelvikten som 1,15 kg/20 = 57,5 gram. Han har inte en aning om hur pass säker han kan vara på det värdet.

5. Låt p beteckna den sanna men okända andelen patienter som blir friska av FÄSOID:s medicin. Eftersom vi (=de) vill bevisa att deras medicin är bäst är alternativhypotesen, den som ska bevisas, $H_1 : p > p_0 = 65\%$. Nollhypotesen är $H_0 : p \leq p_0 = 65\%$, där vi räknar i värsta-fall scenario för FÄSOID, vilket är gränsen p_0. Hypotestestning är konservativ till nollhypotesens fördel.

   Stickprovets storlek är $n = 128$, vilket med 90 botade patienter ger punktskattningen $\hat{p} = 90/128 = 70,3\%$ och $n\hat{p}(1 - \hat{p}) = 26,7 > 5$. Normalapproximationsformlerna kan därmed användas.

   Vi bildar $Z = (\hat{p} - p_0)/\sqrt{p_0(1 - p_0)/n}$, beräknar $z = 1,260$ och jämför med normalfördelningstabellens värde $z_{0,05} = 1,645$. Eftersom $z \leq z_{0,05}$ följer att nollhypotesen *inte* kan förkastas.

   FÄSOID har inget statistiskt belägg för att deras medicin skulle vara bättre än den bästa som idag finns på marknaden. Givetvis kan de heller inte säga motsatsen, utan slutsatsen är att det inte går att dra någon slutsats av den gjorda undersökningen.

6. (a) Korrelstionskoefficienten i Excels datablad kallas för ”Multipel-R” och avläses till 60%.

   (b) Vi ska göra ett ensidigt hypotestest för att se om den sanna riktningskoefficienten β kan anses statistiskt säkert vara större än noll. Vi sätter då alternativhypotesen $H_1 : \beta > 0$. Nollhypotesen är $H_0 : \beta \leq 0$, där vi räknar i värsta-fall scenario vilket är gränsen $\beta = 0$.

   Vi bildar uttrycket $T = \hat{\beta}/\mathrm{SE}(\hat{\beta}) \in t(n - 2)$ och bestämmer dess värde till $1,988$ Eftersom $n = 9$ och signifikansnivån $\alpha = 5\%$ slår vi upp i tabell värdet $t_{0,05}(7) = 1,895$. Noterar $T > t_{0,05}$.

   Slutsatsen är att på den valda signifikansnivån är det statistiskt bevisat att riktningskoefficienten är större än noll, dvs att det finns ett positivt samband mellan nedlagd tid på studier och resultat på den aktuella kursen.

   Kommentar: Frågan är komplex, och rimligen bör man studera mer än medelvärden på nedlagd tid. Likaså bör de låga svarsnivåerna på kursvärderingar där nedlagd tid skattas leda till försiktighet i tolkningen av resultatet, samt att det i tentamensresultaten även ingår de studenter som läste kursen ett tidigare år men då inte fick godkänt resultat och nu skriver om. Nivån på tentamen kan också variera mellan åren, även om avsikten är att den skall ligga fast. Data är hämtade från kursen TNSL01 vid LiU.