Matematisk statistik LKT325
Lösningar Tentamen 20180109

- **Uppgift 1, 3+3 poäng**

  (a)
  $$E(\xi) = \int_0^2 x(\frac{3}{2}x - \frac{3}{4}x^2)dx = \ldots = \boxed{1}.$$
  $$E(\xi^2) = \int_0^2 x^2(\frac{3}{2}x - \frac{3}{4}x^2)dx = \ldots = \frac{6}{5}.$$
  $$\sigma = \sqrt{E(\xi^2) - E(\xi)^2} = \sqrt{\frac{6}{5} - 1} = \boxed{\frac{1}{\sqrt{5}} \approx 0.4472}.$$

  (b)
  $$E(\eta) = 700E(\xi) = 700.$$
  $$S(\eta) = 700S(\xi) = \frac{700}{\sqrt{5}}.$$

  Låt T vara utgifterna under ett år. Enligt centrala gränsvärdessatsen är T approximativt $N(365 \times 700, \frac{\sqrt{365 \times 700}}{\sqrt{5}}) \approx N(255500, 5980.8)$-fördelad. Nu fås att

  $$P(T \leq 245000) = P(\frac{T - 255500}{5980.8} \leq \frac{245000 - 255500}{5980.8})$$
  $$\approx \Phi(-1.76) = 1 - \Phi(1.76) \approx 1 - 0.9608 = \boxed{0.0392.}$$

- **Uppgift 2, 2+2+2 poäng**

  Låt ξ vara antalet gröna bollar som dras och η antalet blå bollar som dras. Det gäller att både ξ och η är $Hyp(N = 14, n = 5, p = 1/2)$. Däremot är de förstås inte oberoende.

  (a)
  $$P(A) = P(\xi = 5) = \frac{\binom{7}{5}\binom{7}{0}}{\binom{14}{5}} = \ldots = \boxed{\frac{3}{286} \approx 0.010490}.$$

  (b) Vi beräknar först $P(B)$. Komplementet till B är att $\eta = 5$ eller $\xi = 5$ vilket ger

  $$P(B) = 1 - P(\xi = 5) - P(\eta = 5) = 1 - 2P(A) = 0.979021.$$

  Ovan användes att η och ξ har samma fördelning. Eftersom A och B är disjunkta fås att

  $$P(A \cup B) = P(A) + P(B) = 0.010490 + 0.979021 \approx \boxed{0.9895}.$$

(c) *Lösning via resonemang*: Att betinga på B^c betyder att vi betingar på att vi fick antingen bara gula bollar, eller bara gröna bollar. Eftersom det finns lika många gula som gröna bollar måste alltså

$$P(A|B^c) = \boxed{\frac{1}{2}}.$$

Alternativ lösning: Från deluppgift b vet vi att $P(B^c) = 2P(A)$. Dessutom gäller att $A \cap B^c = A$. Alltså fås att

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{P(A)}{2P(A)} = \boxed{\frac{1}{2}}.$$

- **Uppgift 3, 4 poäng** Vi får att $\bar{x} = 34.7$ och $s = 3.90299$. Antal frihetsgrader är $4 - 1 = 3$. Från t-tabellen får vi med $\alpha = 0.05$ att $t(3) = 3.18$ (se rad 3, kolumn 0.05). Intervallet blir alltså

$$34.7 \pm \frac{3.18 \times 3.90299}{\sqrt{4}} = \boxed{34.7 \pm 6.2058} = \boxed{[28.4942, 40.9058]}.$$

- **Uppgift 4, 7 poäng**

Låt A beteckna händelsen att man går till sjö A, och låt B beteckna händelsen att man går till sjö B. Låt ξ beteckna antalet fångade fiskar. Låt C beteckna händelsen $\{\xi \geq 3\}$. Vi söker $P(A|C)$, och vi kommer att använda Bayes sats. Om A inträffar blir ξ en Po(5)-fördelad stokastisk variabel och om B inträffar blir ξ en $Po(4)$-fördelad stokastisk variabel. Vi beräknar först $P(A)$ och $P(B)$. Om η betecknar tiden man väntar på bussen blir

$$P(A) = P(\eta \leq 9) = \int_0^9 \frac{1}{12} dx = 0.75.$$

Eftersom han går till B om han inte går till A blir

$$P(B) = 1 - P(A) = 0.25.$$

Enligt ovan har vi att

$$\begin{aligned} P(C|A) = P(\xi \geq 3|A) = 1 - P(\xi \leq 2|A) &= 1 - \left(\frac{e^{-5} \times 5^0}{0!} + \frac{e^{-5} \times 5^1}{1!} + \frac{e^{-5} \times 5^2}{2!} \right) \\ &= 1 - 0.124652 = 0.875348. \end{aligned}$$

På liknande sätt fås att

$$P(C|B) = 1 - \left(\frac{e^{-4} \times 4^0}{0!} + \frac{e^{-4} \times 4^1}{1!} + \frac{e^{-4} \times 4^2}{2!} \right) = 0.761897.$$

Nu ger Bayes sats att

$$\begin{aligned} P(A|C) &= \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|B)P(B)} \\ &\approx \frac{0.875348 \times 0.75}{0.875348 \times 0.75 + 0.761897 \times 0.25} \approx \boxed{0.7751}. \end{aligned} \tag{1}$$

- **Uppgift 5, 3+3+1 poäng**

  (a)

  $l_A = (54{+}52{+}56{+}55{+}53{+}51{+}57{+}59{-}42{-}43{-}41{-}44{-}41{-}44{-}43{-}48)/8 = 11.375$

  $l_{AB} = (42{+}52{+}41{+}55{+}41{+}51{+}43{+}59{-}54{-}43{-}56{-}44{-}53{-}44{-}57{-}48)/8 = -1.875$

  (b) Från generatorerna får vi att $I_1 = ABCDE$, $I_2 = ABCF$ och $I_3 = BCDG$. Sedan får vi att $I_4 = I_1I_2 = ABCDEABCF = DEF$, $I_5 = I_1I_3 = ABCDEBCDG = AEG$, $I_6 = I_2I_3 = ABCFBCDG = ADFG$ och $I_7 = I_1I_2I_3 = ABCDEABCFBCDG = BCEFG$. Det kortaste ordet har längd 3, och upplösningen är därför 3 (III). Vi ser till exempel att $E = EI_4 = EDEF = DF$, så E kommer att blandas ihop med (bland annat) två-faktorsamspelet DF i detta reducerade faktorförsök.

  (c) Antalet tre-faktorsamspel ges av $\binom{8}{3} = \ldots = 56$.

- **Uppgift 6, 4+5 poäng**

  (a) Sannolikheten att en given lampa lyser efter 105 timmar ges av

  $$\int_{105}^{\infty} \frac{1}{100} e^{-x/100} = \ldots = e^{-1.05}.$$

  Låt η vara antalet lysande lampor efter 105 timmar. Då är η, eftersom lamporna antas oberoende, $Bin(n = 100, p = e^{-1.05})$. Eftersom $np(1-p) = 22.7481 > 10$ så blir η approximativt $N(np, \sqrt{np(1-p)}) = N(34.99, 4.7694)$. Vi får att

  $$P(\eta > 40) = 1 - P(\eta \leq 40) = 1 - P\left(\frac{\eta - 34.99}{4.7694} \leq \frac{40 - 34.99}{4.7694}\right)$$
  $$\approx 1 - \Phi(1.05) = 1 - 0.8531 = \boxed{0.1469.}$$

  (b) Vi skall beräkna $F(x) = P(\xi \leq x)$ för alla x. Om $x < 0$ blir förstås $F(x) = 0$. Så antag nu att $x \geq 0$. Låt η_x vara antalet lampor som gått sönder vid tiden x. Observera att

  $$F(x) = P(\xi \leq x) = P(\eta_x \geq 2).$$

  Sannolikheten att en given lampa gått sönder vid tiden x ges av

  $$\int_0^x \frac{1}{100} e^{-t/100} = 1 - e^{-x/100}.$$

  Eftersom lamporna är oberoende av varandra är η_x $Bin(n = 5, p = 1 - e^{-x/100})$-fördelad. Alltså blir,

  $$P(\eta_x \geq 2) = 1 - P(\eta_x < 2) = 1 - P(\eta_x = 0) - P(\eta_x = 1)$$
  $$= 1 - \binom{5}{0}(1 - e^{-x/100})^0 (e^{-x/100})^5 - \binom{5}{1}(1 - e^{-x/100})^1 (e^{-x/100})^4$$
  $$= \ldots = \boxed{1 + 4e^{-x/20} - 5e^{-x/25}}.$$

Variationskälla	SS	df	MS	F
Mellan grupper	184.17	3	61.3886	2.5
Inom grupper	983.52	40	24.588	
Totalt	1167.69	43		

- **Uppgift 7, 5 poäng**

  Lösning: Den färdiga ANOVA-tabellen bör se ut som ovan.

  Vi använder de fem stegen.

Steg 1. Låt μ_1, μ_2, μ_3 och μ_4 beteckna genomsnittet för hur hårt han skjuter med de olika klubborna. Våra hypoteser blir $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ och $H_1 : \exists i, j : \mu_i \neq \mu_j$.

Steg 2. Vi väljer signifikansnivå $\alpha = 0.05$. (Man är fri att välja signifikansnivå men det är bara 0.05 som finns i tabellen så om man väljer någon annan nivå får man problem.)

Steg 3. Testvariabeln vi vill använda är

$$F = \frac{MSB}{MSE} = \frac{(SSB)/(k-1)}{(SSE)/(n-k)}$$

som är F-fördelad med 3 och 40 frihetsgrader.

Steg 4. Vi vill ta fram vårt observerade värde på F. För att göra det så fyller vi i ANOVA-tabellen succesivt. Vi vet att $SST = SSB + SSE$ så vi vet att $SST = 184.17 + 983.52 = 1167.69$. Eftersom det är fyra grupper och tio observationer i varje grupp blir frihetsgraderna 3 och 40. Vi vet också att $MSB = SSB/df = 61.3886$ och att $MSE = SSE/df = 24.588$. Tillslut får vi att $F = MSB/MSE = 2.5$.

Steg 5. Vi tittar i F-fördelningstabellen med 3 och 40 frihetsgrader och ser att det kritiska värdet är 2.84. Eftersom vårt observerade värde är mindre än det kritiska värdet så förkastar vi inte H_0 på signifikansnivå $\alpha = 0.05$.

Slutsats: Vi kan inte utesluta att han i genomsnitt skjuter lika hårt med alla klubborna. Han bör således inte lägga för stor vikt vid skotthastigheten när han bestämmer sig för vilken klubba han ska spela med.

- **Uppgift 8, 4+1+1 poäng**

  Lösning:

  (a) Låt μ vara det riktiga medelvärdet för antalet russin i paketen.

  Steg 1 $H_0 : \mu \leq 200$, $H_1 : \mu > 200$.

  Steg 2 Bestäm signifikansnivå, t ex $\alpha = 0.05$.

  Steg 3 Vi får ingen information om variansen, alltså måste vi göra ett t-test. Testvariabeln vi använder är således

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

Steg 4 I vårt fall är $\mu_0 = 200$, $\bar{x} = 206.4$, $s = 14.5$ och $n = 12$. Vårt värde på testvariabeln blir då tillslut $T = 1.5290$.

Steg 5 Det kritiska värdena tas fram genom att titta i t-tabellen med 11 frihetsgrader ($\alpha = 0.05$): $t_c = 1.80$. Eftersom vår observerade testvariabel $T = 1.5290$ är mindre än de kritiska värdet så kan vi inte förkasta H_0 på signifikansnivå $\alpha = 0.05$.

Slutsats: Vi kan inte utesluta att antalet russin i paketen i genomsnitt är mindre än 200. Vi behöver alltså inte vidta några åtgärder. (Observera att p-värdet är 0.0767 så om man väljer $\alpha = 0.1$ så ska man förkasta H_0 och dra motsvarande slutsats.)

(b) Det viktiga här är att påpeka att stickprovet är just ett stickprov och inte nödvändigtvis representativt för hela populationen. Om det riktiga medelvärdet är exakt 200 så är det 50% chans att stickprovsmedelvärdet är större än 200 (förutsatt att antalet russin i paketen är normalfördelat).

(c) Det enklaste argumentet är att normalfördelningen är kontinuerlig men antalet russin i ett paket måste vara ett heltal. Dessutom kan normalfördelningen anta negativa värden men det kan inte finnas negativt antal russin i ett paket. Man kan godta bägge argumenten.