Lösningar till Matematisk statistik den 14/12-11

Uppgift 1: Det finns 4 olika situationer som kan uppstå när man kastar tärningen:

(udda, udda) (udda, jämn) (jämn, udda) (jämn, jämn)

Vart och ett av fallen inträffar med sannolikheten $\frac{1}{4}$

<u>Situation 1: (udda, udda)</u> I denna situation flyttas en kula från uran U_1 efter både första och andra kastet. Det betyder att kulorna som flyttas båda måste vara svarta.

P(svart efter kast 1 ∩ svart efter kast 2) = $\frac{3}{7} \cdot \frac{2}{6} = \frac{1}{7}$

<u>Situation 2: (udda, jämn)</u> I denna situation flyttas en kula från uran U_1 efter första kastet och en kula från urna U_2 efter andra kastet. Nu kan 2 situationer uppkomma: (svart, svart) eller (vit, vit)

P(svart efter kast 1 ∩ svart efter kast 2) + P(vit efter kast 1 ∩ vit efter kast 2) =

$$= \frac{3}{7} \cdot \frac{7}{11} + \frac{4}{7} \cdot \frac{5}{11} = \frac{41}{77}$$

<u>Situation 3: (jämn, udda)</u> I denna situation flyttas en kula från uran U_2 efter första kastet och en kula från urna U_1 efter andra kastet. Även här kan 2 situationer uppkomma: (svart, svart) eller (vit, vit)

P(svart efter kast 1 ∩ svart efter kast 2) + P(vit efter kast 1 ∩ vit efter kast 2) =

$$= \frac{6}{10} \cdot \frac{4}{8} + \frac{4}{10} \cdot \frac{5}{8} = \frac{11}{20}$$

<u>Situation 1: (jämn, jämn)</u> I denna situation flyttas en kula från uran U_2 efter både första och andra kastet. Det betyder att kulorna som flyttas båda måste vara svarta.

P(svart efter kast 1 ∩ svart efter kast 2) = $\frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}$

P(4 vita i urna U_2 efter 2 kast) = $\frac{1}{4}\left(\frac{1}{7} + \frac{41}{77} + \frac{11}{20} + \frac{1}{3} \right) = \frac{7201}{18480} \approx 0.3897$

Uppgift 2 ξ = antal lådor med repiga CD-skivor i urvalet

$\xi = \text{Hyp}(N, Np, n) = \text{Hyp}(10, 2, 4)$

a) $$P(\xi > 0) = 1 - P(\xi = 0) = 1 - \frac{\binom{2}{0}\binom{8}{4}}{\binom{10}{4}} = 1 - \frac{8\cdot7\cdot6\cdot5}{10\cdot9\cdot8\cdot7} = 1 - \frac{1}{3} = \frac{2}{3}$$

b) η = antal tillfällen man finner repiga CD-skivor i urvalet

$\eta = \text{Bin}(n, p) = \text{Bin}(5, \frac{2}{3})$

$$P(\eta > 1) = 1 - P(\eta = 0) = 1 - \binom{5}{0}\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^5 \approx 1 - 0.00412 = 0.99588$$

Uppgift 3:

a) $$\int_{-\infty}^{\infty} f(x)\,dx = 1 \;\Rightarrow\; -c\int_{-1}^{0} x\,dx + c\int_{0}^{\infty} e^{-6x}dx = -c\left[\frac{x^2}{2}\right]_{-1}^{0} + c\left[\frac{e^{-6x}}{-6}\right]_{0}^{\infty} =$$

$$= -c\left(0 - \frac{1}{2}\right) + c\left(0 - (-\frac{1}{6})\right) = c\frac{4}{6} \;\Leftrightarrow\; c = 1.5$$

b)

<u>$x \le -1$</u> :

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = 0$$

<u>$-1 < x < 0$</u> :

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-1} 0\,dt + 1.5\int_{0}^{x}(-t)\,dt = 1.5\left[-\frac{t^2}{2}\right]_{-1}^{x} = 1.5\,(-\frac{x^2}{2} + \frac{1}{2}) = 0.75(-x^2 + 1)$$

<u>$0 \le x$</u> :

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-1} 0\,dt + 1.5\int_{-1}^{0}(-t)\,dt + 1.5\int_{0}^{x} e^{-6t}\,dt =$$

$$= 0 + 1.5\left[-\frac{t^2}{2}\right]_{-1}^{0} + 1.5\left[\frac{e^{-6t}}{-6}\right]_{0}^{x} = 1.5(0 + \frac{1}{2}) + 1.5(\frac{e^{-6x}}{-6} + \frac{1}{6}) = -0.25e^{-6x} + 1$$

fortsättning uppgift 3b på nästa sida

fortsättning uppgift 3b

Fördelningsfunktionen:

$$F(x) = \begin{cases} 0 & \text{för } x \le -1 \\ 0.75(-x^2+1) & \text{för } -1 < x < 0 \\ -0.25e^{-6x}+1 & \text{för } x \ge 0 \end{cases}$$

c) $P(0.5 < \xi) + P(\xi < -0.5) = 1 - P(\xi < 0.5) + P(\xi < -0.5) =$

$$= 1 - \left[-0.25e^{-6x}+1\right]_{x=0.5} + \left[0.75(-x^2+1)\right]_{x=-0.5} = 0.25\, e^{-3} + 0.5625 \approx 0.5749$$

Uppgift 4: ξ_i = vinsten efter ett spel

$\xi = x$	$P(\xi = x)$
-10	0.5
10	0.5

$E(\xi) = -10 \cdot 0.5 + 10 \cdot 0.5 = 0$ $\quad$ $Var(\xi) = (-10)^2 \cdot 0.5 + 10^2 \cdot 0.5 = 100$

η = total vinst efter 1000 spel $\quad$ där $\quad$ $\eta = \xi_1 + \xi_2 + \ldots.. + \xi_{1000}$

$E(\eta) = 1000\, E(\xi) = 0$ $\quad$ $Var(\eta) = 1000\, Var(\xi) = 1000 \cdot 100 = 100\,000$

$n \ge 30$. använd normalapproximation (centrala gränsvärdessatsen)

$$P(\eta > 100) = 1 - P(\eta < 100) = 1 - P(Z < \frac{100-0}{\sqrt{100000}}) = 1 - P(Z < 0.32) \approx$$

$\approx 1 - 0.6255 = 0.3745$

Uppgift 5:

a) $\ell_C = \dfrac{3000+3000+1942+2512}{4} - \dfrac{1875+1635+150+165}{4} = 1657.25$

$\ell_{AC} = \dfrac{1875+150+3000+2512}{4} - \dfrac{1635+165+3000+1942}{4} = 198.75$

$\ell_{BC} = \dfrac{1875+1635+1942+2512}{4} - \dfrac{150+165+3000+3000}{4} = 412.25$

b) σ är okänd $\Rightarrow$ använd t-fördelningen med $5 \cdot 8 - 8 = 32$ df

Eftersom N = 40 > 30 kan man välja om man vill göra beräkningen exakt eller om man vill använda centrala gränsvärdessatsen och använda normalfördelningen.

$$s_p^2 = \frac{1}{5 \cdot 8 - 8}\left[(5-1) \cdot 1541.2^2 + (5-1) \cdot 1333.4^2 + (5-1) \cdot 0^2 + (5-1) \cdot 32.6^2 + \right.$$

$$\left. + (5-1) \cdot 0^2 + (5-1) \cdot 0^2 + (5-1) \cdot 1346.8^2 + (5-1) \cdot 1090.3^2 \right] = 894617.5113$$

$$0 \pm t \cdot \frac{2 \cdot s_p}{\sqrt{N}} \quad \Rightarrow \quad 0 \pm 2.75 \cdot 2\sqrt{\frac{894617.5113}{40}} \quad \Rightarrow \quad 0 \pm 822.529$$

c) Faktorerna B och C ger signifikanta effekter

$$\hat{y} = \ell_M + \frac{\ell_B}{2} \cdot x_B + \frac{\ell_C}{2} \cdot x_C$$

$$\Rightarrow \quad \hat{y} = 1784.875 - \frac{1185.25}{2} \cdot x_B + \frac{1657.25}{2} \cdot x_C$$

där $\ell_M = \dfrac{1875 + 1635 + 150 + 165 + 3000 + 3000 + 1942 + 2512}{8} = 1784.875$

Uppgift 6:

a) 5 faktorer, A, B, C, D och E
b) 8 försök
c) 2 generatorer, t.ex. D = AB och E = AC
d) 3 definierande relationer. Av ovanstående generatorer erhålls
$I_1 = ABD$
$I_2 = ACE$
$I_3 = I_1 \cdot I_2 = \cancel{A}BD \cdot \cancel{A}CE = BCDE$

Uppgift 7: Vi gör ett hypotestest

Steg 1 H0: $\mu = 28$
H1: $\mu \neq 28$

Steg 2 Signifikansnivå $\alpha = 0.01$

Steg 3 Vi har normalfördelad data med okänd varians, och färre än 30 observationer -- vi måste göra ett t-test!

Testvariabel: $t = \frac{\bar{x}-\mu}{s_x/\sqrt{n}}$

Fortsättning uppgift 7 på nästa sida

Fortsättning uppgift 7

Steg 4

$$t = \frac{27.42 - 28}{2.39/\sqrt{24}} = -1.1936$$

Steg 5 Vi jämför vårt t = -1.1936 med de kritiska värdena från tabell (dubbelsidigt): 23 frihetsgrader och nivån α= 0.01 ger kritiska värden tc = ±2.81.
Vårt t > -2.81 och t < 2.81 så vi hamnar i acceptansområdet. Vi kan EJ förkasta H0 och vi kan därmed ej säga om köttbullarnas diameter skiljer från 28 mm.

Uppgift 8:

Variationskälla	SS	df	MS	F_0	p-värde (behövs ej!)
Leverantör	9.93636667	3	3.31212222	3.067442554	0.0910586519
Okänd	8.63813333	8	1.07976667		
Total	18.5745	11			

Hypoteser:
H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (alla medelvärden lika)
H_1: $\mu_i \neq \mu_j$ för $i \neq j$ (alla medelvärden inte lika, räcker att minst ett skiljer från resten)

Tabelluppslag av kritiskt F-värde med v_1=3 och v_2=8 frihetsgrader på signifikansnivån α=0.05 gav $F_{kritiskt}$ = 4.066181

Eftersom F_0 < 4.066 kan vi EJ förkasta H_0, det verkar därför som om tillverkarnas saltlösningar har samma medelsaltkoncentration.

Notera att testets p-värde (som inte behöver beräknas i uppgiften) visar att vi hade förkastat H0 på signifikansnivån 0.1, men som sagt inte på 0.05.

Vanligt IV-dropp har en saltkoncentration på 9 mg/mL. Kul att veta är också att den egna produktionens värden egentligen kommer från en normalfördelning med medelvärde 10.5 och en större standardavvikelse än datan från övriga leverantörer. För att finna detta behövs fler mätpunkter!

Lösningar till Matematisk statistik den 14/12-11

Uppgift 1: Det finns 4 olika situationer som kan uppstå när man kastar tärningen:

(udda, udda) (udda, jämn) (jämn, udda) (jämn, jämn)

Vart och ett av fallen inträffar med sannolikheten $\frac{1}{4}$

<u>Situation 1: (udda, udda)</u> I denna situation flyttas en kula från uran U_1 efter både första och andra kastet. Det betyder att kulorna som flyttas båda måste vara svarta.

$$P(\text{svart efter kast 1} \cap \text{svart efter kast 2}) = \frac{3}{7} \cdot \frac{2}{6} = \frac{1}{7}$$

<u>Situation 2: (udda, jämn)</u> I denna situation flyttas en kula från uran U_1 efter första kastet och en kula från urna U_2 efter andra kastet. Nu kan 2 situationer uppkomma: (svart, svart) eller (vit, vit)

$$P(\text{svart efter kast 1} \cap \text{svart efter kast 2}) + P(\text{vit efter kast 1} \cap \text{vit efter kast 2}) =$$

$$= \frac{3}{7} \cdot \frac{7}{11} + \frac{4}{7} \cdot \frac{5}{11} = \frac{41}{77}$$

<u>Situation 3: (jämn, udda)</u> I denna situation flyttas en kula från uran U_2 efter första kastet och en kula från urna U_1 efter andra kastet. Även här kan 2 situationer uppkomma: (svart, svart) eller (vit, vit)

$$P(\text{svart efter kast 1} \cap \text{svart efter kast 2}) + P(\text{vit efter kast 1} \cap \text{vit efter kast 2}) =$$

$$= \frac{6}{10} \cdot \frac{4}{8} + \frac{4}{10} \cdot \frac{5}{8} = \frac{11}{20}$$

<u>Situation 1: (jämn, jämn)</u> I denna situation flyttas en kula från uran U_2 efter både första och andra kastet. Det betyder att kulorna som flyttas båda måste vara svarta.

$$P(\text{svart efter kast 1} \cap \text{svart efter kast 2}) = \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}$$

$$P(\text{4 vita i urna } U_2 \text{ efter 2 kast}) = \frac{1}{4}\left(\frac{1}{7} + \frac{41}{77} + \frac{11}{20} + \frac{1}{3} \right) = \frac{7201}{18480} \approx 0.3897$$

Uppgift 2 ξ = antal lådor med repiga CD-skivor i urvalet

$\xi = \text{Hyp}(N, Np, n) = \text{Hyp}(10, 2, 4)$

a) $P(\xi > 0) = 1 - P(\xi = 0) = 1 - \dfrac{\binom{2}{0}\binom{8}{4}}{\binom{10}{4}} = 1 - \dfrac{8\cdot7\cdot6\cdot5}{10\cdot9\cdot8\cdot7} = 1 - \dfrac{1}{3} = \dfrac{2}{3}$

b) η = antal tillfällen man finner repiga CD-skivor i urvalet

$\eta = \text{Bin}(n, p) = \text{Bin}(5, \dfrac{2}{3})$

$P(\eta > 1) = 1 - P(\eta = 0) = 1 - \binom{5}{0}\left(\dfrac{2}{3}\right)^0\left(\dfrac{1}{3}\right)^5 \approx 1 - 0.00412 = 0.99588$

Uppgift 3:

a) $\int\limits_{-\infty}^{\infty} f(x)dx = 1 \;\Rightarrow\; -c\int\limits_{-1}^{0} x\,dx + c\int\limits_{0}^{\infty} e^{-6x}dx = -c\left[\dfrac{x^2}{2}\right]_{-1}^{0} + c\left[\dfrac{e^{-6x}}{-6}\right]_{0}^{\infty} =$

$= -c\left(0 - \dfrac{1}{2}\right) + c\left(0 - (-\dfrac{1}{6})\right) = c\dfrac{4}{6} \;\Leftrightarrow\; c = 1.5$

b)

<u>$x \le -1$</u> :

$F(x) = \int\limits_{-\infty}^{x} f(t)dt = 0$

<u>$-1 < x < 0$</u> :

$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{-1} 0\,dt + 1.5\int\limits_{0}^{x}(-t)dt = 1.5\left[-\dfrac{t^2}{2}\right]_{-1}^{x} = 1.5\,(-\dfrac{x^2}{2} + \dfrac{1}{2}) = 0.75(-x^2 + 1)$

<u>$0 \le x$</u> :

$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{-1} 0\,dt + 1.5\int\limits_{-1}^{0}(-t)dt + 1.5\int\limits_{0}^{x} e^{-6t}\,dt =$

$= 0 + 1.5\left[-\dfrac{t^2}{2}\right]_{-1}^{0} + 1.5\left[\dfrac{e^{-6t}}{-6}\right]_{0}^{x} = 1.5(0 + \dfrac{1}{2}) + 1.5(\dfrac{e^{-6x}}{-6} + \dfrac{1}{6}) = -0.25e^{-6x} + 1$

fortsättning uppgift 3b på nästa sida

fortsättning uppgift 3b

Fördelningsfunktionen:

$$F(x) = \begin{cases} 0 & \text{för } x \le -1 \\ 0.75(-x^2+1) & \text{för } -1 < x < 0 \\ -0.25e^{-6x}+1 & \text{för } x \ge 0 \end{cases}$$

c) $P(0.5 < \xi) + P(\xi < -0.5) = 1 - P(\xi < 0.5) + P(\xi < -0.5) =$

$$= 1 - \left[-0.25e^{-6x}+1\right]_{x=0.5} + \left[0.75(-x^2+1)\right]_{x=-0.5} = 0.25\,e^{-3} + 0.5625 \approx 0.5749$$

Uppgift 4: ξ_i = vinsten efter ett spel

$\xi = x$	$P(\xi = x)$
-10	0.5
10	0.5

$E(\xi) = -10\cdot 0.5 + 10\cdot 0.5 = 0$ $\quad$ $Var(\xi) = (-10)^2\cdot 0.5 + 10^2\cdot 0.5 = 100$

η = total vinst efter 1000 spel där $\eta = \xi_1 + \xi_2 + \ldots + \xi_{1000}$

$E(\eta) = 1000\,E(\xi) = 0$ $\quad$ $Var(\eta) = 1000\,Var(\xi) = 1000\cdot 100 = 100\,000$

n ≥ 30. använd normalapproximation (centrala gränsvärdessatsen)

$$P(\eta > 100) = 1 - P(\eta < 100) = 1 - P\left(Z < \frac{100-0}{\sqrt{100000}}\right) = 1 - P(Z < 0.32) \approx$$

$\approx 1 - 0.6255 = 0.3745$

Uppgift 5:

a) $$\ell_C = \frac{3000+3000+1942+2512}{4} - \frac{1875+1635+150+165}{4} = 1657.25$$

$$\ell_{AC} = \frac{1875+150+3000+2512}{4} - \frac{1635+165+3000+1942}{4} = 198.75$$

$$\ell_{BC} = \frac{1875+1635+1942+2512}{4} - \frac{150+165+3000+3000}{4} = 412.25$$

b) σ är okänd $\Rightarrow$ använd t-fördelningen med $5 \cdot 8 - 8 = 32$ df

Eftersom N = 40 > 30 kan man välja om man vill göra beräkningen exakt eller om man vill använda centrala gränsvärdessatsen och använda normalfördelningen.

$$s_p^2 = \frac{1}{5 \cdot 8 - 8}\left[(5-1)\cdot 1541.2^2 + (5-1)\cdot 1333.4^2 + (5-1)\cdot 0^2 + (5-1)\cdot 32.6^2 + \right.$$

$$\left. + (5-1)\cdot 0^2 + (5-1)\cdot 0^2 + (5-1)\cdot 1346.8^2 + (5-1)\cdot 1090.3^2\right] = 894617.5113$$

$$0 \pm t \cdot \frac{2 \cdot s_p}{\sqrt{N}} \quad \Rightarrow \quad 0 \pm 2.\,75 \cdot 2\sqrt{\frac{894617.5113}{40}} \quad \Rightarrow \quad 0 \pm 822.529$$

c) Faktorerna B och C ger signifikanta effekter

$$\hat{y} = \ell_M + \frac{\ell_B}{2} \cdot x_B + \frac{\ell_C}{2} \cdot x_C$$

$$\Rightarrow \quad \hat{y} = 1784.875 - \frac{1185.25}{2} \cdot x_B + \frac{1657.25}{2} \cdot x_C$$

där $\ell_M = \dfrac{1875 + 1635 + 150 + 165 + 3000 + 3000 + 1942 + 2512}{8} = 1784.875$

Uppgift 6:

a) 5 faktorer, A, B, C, D och E
b) 8 försök
c) 2 generatorer, t.ex. D = AB och E = AC
d) 3 definierande relationer. Av ovanstående generatorer erhålls
$I_1 = ABD$
$I_2 = ACE$
$I_3 = I_1 \cdot I_2 = \cancel{A}BD \cdot \cancel{A}CE = BCDE$

Uppgift 7: Vi gör ett hypotestest

Steg 1 H0: $\mu = 28$
H1: $\mu \neq 28$

Steg 2 Signifikansnivå $\alpha = 0.01$

Steg 3 Vi har normalfördelad data med okänd varians, och färre än 30 observationer -- vi måste göra ett t-test!

Testvariabel: $t = \frac{\bar{x}-\mu}{s_x/\sqrt{n}}$

Fortsättning uppgift 7 på nästa sida

Fortsättning uppgift 7

Steg 4

$$t = \frac{27.42 - 28}{2.39/\sqrt{24}} = -1.1936$$

Steg 5 Vi jämför vårt t = -1.1936 med de kritiska värdena från tabell (dubbelsidigt): 23 frihetsgrader och nivån $\alpha = 0.01$ ger kritiska värden tc = ±2.81.
Vårt t > -2.81 och t < 2.81 så vi hamnar i acceptansområdet. Vi kan EJ förkasta H0 och vi kan därmed ej säga om köttbullarnas diameter skiljer från 28 mm.

Uppgift 8:

Variationskälla	SS	df	MS	F_0	p-värde (behövs ej!)
Leverantör	9.93636667	3	3.31212222	3.067442554	0.0910586519
Okänd	8.63813333	8	1.07976667		
Total	18.5745	11			

Hypoteser:
H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (alla medelvärden lika)
H_1: $\mu_i \neq \mu_j$ för $i \neq j$ (alla medelvärden inte lika, räcker att minst ett skiljer från resten)

Tabelluppslag av kritiskt F-värde med $v_1=3$ och $v_2=8$ frihetsgrader på signifikansnivån $\alpha=0.05$ gav $F_{kritiskt} = 4.066181$

Eftersom $F_0 < 4.066$ kan vi EJ förkasta H_0, det verkar därför som om tillverkarnas saltlösningar har samma medelsaltkoncentration.

Notera att testets p-värde (som inte behöver beräknas i uppgiften) visar att vi hade förkastat H0 på signifikansnivån 0.1, men som sagt inte på 0.05.

Vanligt IV-dropp har en saltkoncentration på 9 mg/mL. Kul att veta är också att den egna produktionens värden egentligen kommer från en normalfördelning med medelvärde 10.5 och en större standardavvikelse än datan från övriga leverantörer. För att finna detta behövs fler mätpunkter!