Tentamen i matematisk statistik för KI den 19 dec 2012

Uppgift 1: Ett företag tillverkar säkerhetsutrustningar till bilar. Tillverkningen är förlagd till fyra olika länder, A, B C och D. I land A finns 40% av företagets produktion, i land B 25%, i land C 20% och i land D den resterande delen. Man vet sedan tidigare att 90% av utrustningarna som kommer från A har en livslängd som är minst 5 år. Motsvarande andelar för B, C och D är 95%, 97% och 99%. Hela företagets tillverkning samlas i ett stort centrallager innan några utrustningar säljs vidare till kund. När en bilfabrik köper företagets säkerhetsutrustningar så får den enheter från alla fyra länderna. Anta att man väljer ut en utrustning slumpmässigt från bilfabrikens lager.

a) Vad är sannolikheten att den utvalda utrustningen håller minst 5 år?
b) Om säkerhetsutrustningen håller minst 5 år, hur stor är sannolikheten att den kommer från land D?

(6 poäng)

Uppgift 2: Till en grossist levereras bl. a. en kartong med 50 jultomtar i porslin. Fyra av jultomtarna har blivit kantstötta i leveransen och måste returneras till fabriken. Ytterligare sex har bara blivit lite skavda så att den röda färgen har försvunnit på någon fläck. Dessa tomtar kan säljas till reducerat pris. Resterande fyrtio tomtar är oskadade. Anta att man slumpmässigt väljer ut fyra jultomtar från leveransen.

a) Vad är sannolikheten att man inte får någon tomte, som är kantstött?
b) Vad är sannolikheten att man får exakt en tomte som är kantstött och exakt en som är lite skavd?
c) Vad är sannolikheten att man får högst en tomte som är kantstött och högst en som är skavd?

(8 poäng)

Uppgift 3: Anta att en trollslända sätter sig på en mätsticka, som är graderad mellan 0 och 50 cm.

a) Vad är sannolikheten att den sitter närmare markeringen 0 än 50?
b) Vad är sannolikheten att trollsländan sitter minst tre gånger så långt från 0 som från 50?

(6 poäng)

Uppgift 4: När man kastar varpa siktar man på en pinne nedsatt i marken. Varpan (en speciell sorts sten) skall komma så nära pinnen som möjligt. Anta att man kastar varpan utefter en 7 meter lång linje med avsikten att komma så nära pinnen som möjligt. Pinnen identifieras med värdet 0 medan ξ är den faktiska landningspunkten för varpan. Frekvensfunktionen för landningspunkten kan skrivas enligt följande:

$$f(x) = \begin{cases} 0.75(1-x^2) & \text{för } -a \le x \le a \\ 0 & \text{för övrigt} \end{cases}$$

Kasta varpan 10 gånger. Beräkna variansen för antal gånger som varpan hamnar närmare pinnen än 0.5m.

(6 poäng)

Uppgift 5: Ett mejeri vill kontrollera densiteten på förpackad vispgrädde och väger därför 196 gräddförpackningar. Vikterna, ξ, (i gram) anses vara oberoende stokastiska variabler med väntevärdet $E(\xi)$ och standardavvikelsen 1.60 gram. Beräkna

$$P(|\bar{\xi} - E(\xi)| < 0.20)$$

där $\bar{\xi}$ är medelvärdet av de 196 gräddförpackningarnas vikter.

(6 poäng)

Uppgift 6: Anta att man genomförde ett faktorförsök där de 3 faktorerna A, B och C användes. I varje försökssituation användes 5 replikat. När ett försök hade pågått tillräckligt länge avbröts det och man användes sig då av de värden som framkommit dittills. Detta hände i försök 3, 5 och 6, vilket gör att dessa värden ser lite konstiga ut. Vid analysen användes ändå dessa värden på samma sätt som de övriga. Följande försöksresultat erhölls:

A	**B**	**C**	**Medel-värde**	**Standard-avvikelse**
-	-	-	1875	1541.2
+	-	-	1635	1333.4
-	+	-	150	0.0
+	+	-	165	32.6
-	-	+	3000	0.0
+	-	+	3000	0.0
-	+	+	1942	1346.8
+	+	+	2512	1090.3

Följande effekter beräknades

$\ell_A = 86.25$ $\ell_B = -1185.25$ $\ell_{AB} = 206.25$ $\ell_{ABC} = 78.75$

a) Använd ovanstående för att bestämma de resterande effekterna.
b) Beräkna ett 99%-igt referensintervall för effekterna.
c) Ange en matematisk modell för ovanstående situation där enbart signifikanta effekter ingår.

(6 poäng)

Uppgift 7: En importör av rundkornigt ris vill sälja sitt ris som högkvalitativt grötris. För att utmärka sig på marknaden vill importören klistra på en etikett med texten ”minst 95% hela riskorn” på paketen. Från ett slumpmässigt stickprov om 200 riskorn som valdes oberoende av varandra fann importören 184 hela riskorn. Hjälp importören avgöra om påståendet stämmer för hans ris.
a) Testa om importören på 5% signifikansnivå kan stå för sitt påstående.
b) Beräkna testets p-värde och kommentera resultatet med hänsyn till testets signifikansnivå.

(6 poäng)

Uppgift 8: En fabrik som tillverkar värmeljus experimenterar med en ny sorts paraffinolja för att ersätta stearinet i deras ljus. De har tre olika stearinblandningar i deras nuvarande sortiment och vill undersöka om de nya paraffinoljebaserade ljusen får lika bra brinntid.
Fabriken vet sedan tidigare att medelbrinntiden på deras värmeljus är normalfördelad och bestämmer sig för att utnyttja variansanalys för att se om de nya ljusen brinner lika länge som de gamla beprövade varianterna.
Utgå från följande påbörjade ANOVA-tabell och gör klart undersökningen. Beskriv hypoteserna bakom undersökningen och förklara testets utfall.

Variationskälla	**SS**	**df**	**MS**	**F_0**
Sammansättning	52052	3		
Okänd	114030			
Total	166082	9		

(6 poäng)

Lösningar till Matematisk statistik den 19/12-12

Uppgift 1: Låt A, B, C respektive D beteckna att säkerhetsutrustningen är tillverkad i land A, B, C respektive D. Låt vidare L betyda att utrustningen har en livslängd på minst 5 år.

$P(A) = 0.49 \quad P(B) = 0.25 \quad P(C) = 0.20 \quad P(D) = 0.15.$

$P(L|A) = 0.90 \quad P(L|B) = 0.95 \quad P(L|C) = 0.97 \quad P(L|D) = 0.99$

a) $P(L) = P(L|A)\cdot P(A) + P(L|B)\cdot P(B) + P(L|C)\cdot P(C) + P(L|D)\cdot P(D) =$

$= 0.90\cdot 0.40 + 0.95\cdot 0.25 + 0.97\cdot 0.20 + 0.99\cdot 0.15 = 0.94$

b) $P(D|L) = \text{(Bayes sats)} = \dfrac{P(L\,|\,D)\cdot P(D)}{P(L)} = \dfrac{0.99\cdot 0.15}{0.94} \approx 0.158$

Uppgift 2: N = 50 n = 4 Kantstötta = 4 Skavda = 6 Hela = 40

a) $P(\text{ingen kantstött}) = \dfrac{\binom{4}{0}\binom{46}{4}}{\binom{50}{4}} = \dfrac{163185}{230300} \approx 0.709$

b) $P(\text{en kantstött och en skavd}) = \dfrac{\binom{4}{1}\binom{6}{1}\binom{40}{2}}{\binom{50}{4}} = \dfrac{18720}{230300} \approx 0.08$

c) Det finns fyra fall som uppfyller villkoret att högst en är kantstött och högst en är skavd. Villkoret "högst en" innebär "noll eller en".
Låt K, S och H beteckna kantstött, skavd respektive hel jultomte. De fyra fallen kan de skrivas som

(K=0, S=0, H=4) och (K=0, S=1, H=3) och (K=1, S=0, H=3) och (K=1, S=1, H=2)

Motsvarande sannolikheter blir

P(K=0, S=0, H=4) + P(K=0, S=1, H=3) + P(K=1, S=0, H=3) + P(K=1, S=1, H=2) =

$$= \frac{\binom{4}{0}\binom{6}{0}\binom{40}{4}}{\binom{50}{4}} + \frac{\binom{4}{0}\binom{6}{1}\binom{40}{3}}{\binom{50}{4}} + \frac{\binom{4}{1}\binom{6}{0}\binom{40}{3}}{\binom{50}{4}} + \frac{\binom{4}{1}\binom{6}{1}\binom{40}{2}}{\binom{50}{4}} = \frac{208910}{230300} \approx 0.907$$

Uppgift 3: ξ = antal cm från 0:an på en 50 cm lång linjal. ξ är R(a, b) = R(0, 50)

a) P(närmare 0 än 50) = P(nedre halvan) = $P(\xi < 25) = \dfrac{25-0}{50-0} = 0.5$

b) P(minst 3 gånger så långt från 0 som från 50) = P(övre fjärdedelen av linjalen) =

$= P(\xi > \dfrac{3 \cdot 50}{4}) = 1 - P(\xi < 37.5) = 1 - \dfrac{37.5-0}{50-0} = 0.25$

Uppgift 4: ξ = landningspunkten för varpan räknat från pinnen

Bestäm variansen för antal av 10 kast som kommer närmare pinnen än 0.5m.

η = antal kast som kommer närmare pinnen än 0.5m

η är Bin(n, p) = Bin (10, p)   där p beräknas med hjälp av frekvensfunktionen f(x).

Beräkna värdet på a.

$$0.75\int_{-a}^{a}(1-x^2)dx = 0.75\left[x - \frac{x^3}{3}\right]_{-a}^{a} = 0.75[(a - \frac{a^3}{3}) - (-a - \frac{(-a)^3}{3})] = 1 \Rightarrow 3a - a^3 = 2$$

$\Rightarrow$ a = 1   Om a hade varit mindre än 0.5 så hade p blivit 1

$$p = 0.75\int_{-0.5}^{0.5}(1-x^2)dx = 0.75\left[x - \frac{x^3}{3}\right]_{-0.5}^{0.5} = 0.75[(0.5 - \frac{0.5^3}{3}) - (-0.5 - \frac{(-0.5)^3}{3})] =$$

$$= 0.75 \cdot (1 - \frac{0.25}{3}) = 0.6875$$

Alltså, η är Bin(10, 0.6875) $\Rightarrow$ $Var(\eta) = np(1-p) = 10 \cdot 0.6875\,(1 - 0.6875) \approx 2.148$

Uppgift 5: ξ = vikt $S(\xi) = 1.60$ gram   n = 196 förpackningar

$E(\bar{\xi}) = E(\xi)$ $Var(\bar{\xi}) = Var(\dfrac{Var(\xi)}{n} = \dfrac{1.6^2}{196}$

$$P(|\bar{\xi} - E(\xi)| < 0.20) = P(-0.20 < \bar{\xi} - E(\xi) < 0.20) = P(\bar{\xi} - E(\xi) < 0.20) - P(\bar{\xi} - E(\xi) < -0.20) =$$

$$= P(Z < \frac{0.20}{1.6/\sqrt{196}}) - (1 - P(Z < \frac{0.20}{1.6/\sqrt{196}}) = 2P(Z < 1.75) - 1 = 0.9198$$

Uppgift 6:

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) $s_p^2 = \dfrac{(5-1)1541.2^2 + (5-1)1332.4^2 + ... + (5-1)1090.3^2}{8 \cdot 4} = 223571.0716$

df = 32 Eftersom n > 30 $\Rightarrow$ använd normalfördelningsapproximation

Ett 99%-igt referensintervall

$0 \pm 2.575 \cdot \dfrac{2 \cdot \sqrt{223571.0716}}{\sqrt{8 \cdot 5}} \quad \Rightarrow \quad 0 \pm 385.02$

c) Faktorerna B och C samt samspelet BC ger signifikant effekter

$\hat{y} = \ell_M + \dfrac{\ell_B}{2} \cdot x_B + \dfrac{\ell_C}{2} \cdot x_C + \dfrac{\ell_{BC}}{2} \cdot x_{BC} \quad \Rightarrow$

$\hat{y} = 1784.875 - \dfrac{1185.25}{2} \cdot x_B + \dfrac{1657.25}{2} \cdot x_C + \dfrac{412.25}{2} \cdot x_{BC}$

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

Uppgift 7:

a) Om p(1-p)n > 10 kan vi använda normalfördelningstest. I vårt fall:
p(1-p)n = 184/200 * (16/200) * 200 = 14,72 > 10.
Vi genomför hypotestest och gör normalfördelningsapproximation av populationsproportionen.

Steg 1. Definiera hypoteserna.
Ett enkelsidigt test för att se om den sanna populationsproportionen hela riskorn är under 95%.
H_0: p = 0.95
H_1: p < 0.95

Steg 2. Bestäm signifikansnivå.
α = 0.01

Steg 3. Välj testvariabel.
Eftersom villkoret för normalfördelningsapproximation är uppfyllt använder vi följande testvariabel

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Steg 4. Beräkna testvariabelns värde utifrån stickprovet.

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{184}{200} - 0.95}{\sqrt{\frac{0.95(1-0.95)}{200}}} \cong -1.95$$

Steg 5. Jämför testvariabelns värde med kritiskt värde ur standardnormalfördelningen.
Från tabell får vi kritiskt Z-värde: $Z_{\alpha=0.05} = -1.645$
Vårt Z = -1.95 < $Z_\alpha = -1.645$ ➔ Vi förkastar H_0!
På 5% signifikansnivå kan inte importören med gott samvete påstå att det är minst 95% hela riskorn.

b) Testets p-värde är sannolikheten att vi, under antagandet om att H_0 stämmer, gör en observation som är ”minst lika extrem” som den vi precis gjort. Testets p-värde erhålls genom att slå upp sannolikheten för ett värde lika litet eller mindre än det värdet på vår testvariabel. Eftersom tabellen bara är given för P(Z<z) > 0.5 måste man slå upp komplementet och erhåller då:
p-värdet = P(Z<-1.95) = 1-P(Z<1.95) = 1 – 0.9744 = 0.0256
Eftersom p-värdet är mindre än nivån för vårt test förkastar vi H_0. Med ett p-värde på 2.56% hade vi inte kunnat förkasta testet om vi testat på signifikansnivån 1%.

Uppgift 8:

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)

Variationskälla	**SS**	**df**	**MS**	**F_0**
Sammansättning	52052	3	26026	1.369529
Okänd	114030	6	19005	
Total	166082	9		

Tabelluppslag av kritiskt F-värde med v_1=3 och v_2=6 frihetsgrader på signifikansnivån α=0.05 gav $F_{kritiskt}$ = 4.76
Eftersom F_0 < 4.76 kan vi EJ förkasta H_0, det verkar därför som om det inte finns någon skillnad i medelbrinntid för de olika ljusen.