Lösningsförslag TMSB18 Matematisk statistik IL 10081
Tid: 12.00-17.00
Telefon: 0707-463397,
Examinator: F Abrahamsson

TEKNISKA HÖGSKOLAN
HÖGSKOLAN I JÖNKÖPING

1. Om ett visst företags inkomster en månad är fördelade enligt $N(7000, 300)$ och samma företags utgifter denna månad är $N(6700, 400)$ (i tusentals kronor), beräkna sannolikheten att företaget går med vinst denna månad. (2p)

   Lösning: Om ξ är inkomsterna och η utgifterna så är $\xi - \eta \in N(300, 500)$ och vi söker sannolikheten
$$P(\xi - \eta > 0) = \Phi\left(\frac{300 - 0}{500}\right) = \Phi(0.6) \approx 0.73.$$

2. Givet en datamängd $x_1, \ldots, x_n$ så är ett vanligt spridningsmått (stickprovets) standardavvikelse s där $s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2$. Beskriv två andra spridningsmått. (2p)

   Lösning: Variationsbredden $R = x_{max} - x_{min}$ samt kvartilavståndet $Q = Q_3 - Q_1$ (där Q_1, Q_3 är 1:a resp. 3:e kvartilen).

3. I en studie av tendenser till inavel bland en speciell flugart testade man 121 flugor genetiskt, och beräknade en sk. inavelskoefficient. Medelvärdet blev $\overline{x} = 0.044$ med en beräknad standardavvikelse $s = 0.884$. Vi antar att testet ger ett observerat stickprov från en normalfördelning $N(\mu, \sigma)$, där μ är den verkliga inavelskoefficienten, som skall vara lika med 0 om ingen inavel sker.

   (a) Genomför ett relevant hypotestest på 5% signifikansnivå för att avgöra om inavel sker eller inte. (2p)

   (b) Beräkna ett lämpligt konfidensintervall för att besvara samma fråga som i (a) med en konfidensgrad som motsvarar den signifikansnivå som användes i (a). (2p)

   Lösning: (a) Vi genomför ett tvåsidigt t-test (σ ej känd) där
$$\begin{array}{ll} H_0: & \mu = 0 \\ H_1: & \mu \neq 0 \end{array}$$
   på signifikansnivån $\alpha = 0.05$. Testvariabeln är
$$T_0 = \frac{\overline{X} - 0}{S/\sqrt{121}}.$$
   Det observerade värdet blir med våra data $t_0 = .55$. Eftersom $t_{\alpha/2,n-1} = t_{0.025,120} = 1.98$ så är $t_0 < t_{\alpha/2,n-1}$ och vi kan då inte förkasta H_0 på 5% signifikansnivå.

   (b) Eftersom konfidensintervallet skall motsvara hypotestestet ovan med signifikansnivån 5% så måste vi beräkna ett tvåsidigt konfidensintervall för μ med konfidensgrad 95%. Enligt teorin ges detta av
$$I_\mu = \overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = 0.044 \pm 1.98\frac{0.884}{\sqrt{121}} \approx [-0.11, 0.21].$$
   Eftersom detta intervall innehåller 0 så kan vi inte förkasta H_0 på 5% signifikansnivå.

4. I en burk finns 5 mynt; 2 st enkronor, 2 st femkronor och 1 st 10-krona. En person drar 2 mynt, ett i taget utan återläggning.

   (a) Beräkna det förväntade värdet av det första draget samt det andra draget. (2p)

   (b) Vad skulle du högst betala för att få dra 2 mynt ur burken? Motivera ditt svar. (1p)

   Lösning: (a) Låt ξ_1, ξ_2 vara värdet av första resp. andra draget. Då är
$$E(\xi_1) = 1 \cdot \frac{2}{5} + 5 \cdot \frac{2}{5} + 10 \cdot \frac{1}{5} = \frac{22}{5}.$$
   För att beräkna $E(\xi_2)$ så låter vi A_k resp. B_k för $k = 1, 5, 10$ beteckna händelserna att vi drar ett mynt värt k kronor i första resp. andra draget. Då är $P(A_1) = P(A_5) = \frac{2}{5}$ och $P(A_{10}) = \frac{1}{5}$. Lagen om total sannolikhet ger att
$$P(B_1) = P(B_1|A_1)P(A_1) + P(B_1|A_1^c)P(A_1^c) = \frac{1}{4} \cdot \frac{2}{5} + \frac{2}{4} \cdot \frac{3}{5} = \frac{8}{20} = \frac{2}{5}$$

och på samma sätt fås att $P(B_5) = \frac{2}{5}$ och därmed att $P(B_{10}) = \frac{1}{5}$. Alltså är

$$E(\xi_2) = 1 \cdot P(B_1) + 5 \cdot P(B_5) + 10 \cdot P(B_{10}) = \frac{22}{5}.$$

(b) Det förväntade värdet av 2 dragna mynt är $E(\xi_1 + \xi_2) = E(\xi_1) + E(\xi_2) = \frac{44}{5}$ vilket är det belopp som man maximalt skulle betala (om man kan lite grundläggande sannolikhetslära).

5. Ann har gjort en serie mätningar som antas vara observationer av en stokastisk variabel fördelad enligt $N(\mu, 0.8)$ där μ är den konstant som Ann vill mäta.

   (a) Ann beräknar ett 95% konfidensintervall för μ till $[8.76, 8.94]$. Hur många mätningar hade Ann gjort? (2p)

   (b) Om Ann gjorde 250 mätningar och fick fram konfidensintervallet $[8.76, 8.94]$, vilken konfidensgrad har Ann använt då? Ge ett exakt svar. (2p)

Lösning: (a) Eftersom intervallets bredd ges av

$$8.94 - 8.76 = 2 \cdot z_{0.025} \frac{\sigma}{\sqrt{n}} = 2 \cdot 1.96 \frac{0.8}{\sqrt{n}}$$

så fås sambandet

$$0.18 = \frac{3.136}{\sqrt{n}} \Leftrightarrow n = \left(\frac{3.136}{0.18}\right) \approx 304.$$

(b) Samma resonemang som i (a) ger sambandet

$$0.18 = 2 \cdot z_{\alpha/2} \frac{0.8}{\sqrt{250}} \Leftrightarrow z_{\alpha/2} = \frac{0.18\sqrt{250}}{2 \cdot 0.8} \approx 1.78\,.$$

Enligt definitionen av kvantil så innebär det att

$$P(Z > 1.78) = \alpha/2 \Leftrightarrow \alpha = 2 \cdot (1 - \Phi(1.78)) \approx 0.075$$

och konfidensgraden är därmed $1 - \alpha = 0.925$.

6. En konsumentorganisation gör ett oberoende test av livslängden av en viss typ av glödlampor. Resultatet från 30 testade lampor redovisas nedan. (2p)

Livslängd (år)	0-1	1-2	2-3	3-4	4-5	5-6	6-7
Observerad frekvens	11	8	6	3	1	0	1

Konsumentorganisationen misstänker att medianlivslängden är mindre än 3 år. Om vi inte kan anta att livslängderna är normalfördelade, genomför ett lämpligt hypotest på 1% signifikansnivå för att se om det går att styrka konsumentorganisationens misstankar.

Lösning: Vi genomför ett teckentest;

$$\begin{aligned} H_0 &: \tilde{\mu} = 3 \\ H_1 &: \tilde{\mu} < 3 \end{aligned}$$

på 1% signifikansnivå. Om R^+ är antalet positiva differenser $x_i - 3$ så är det observerade resultatet $R^+_{\text{obs}} = r^+ = 5$. Eftersom

$$\begin{aligned} P-\text{värdet} &= P(R^+ \leq 5 | R^+ \in \text{Bin}(30, \frac{1}{2})) \\ &= \begin{pmatrix} 30 \\ 0 \end{pmatrix} (\frac{1}{2})^0 (\frac{1}{2})^{30} + \cdots + \begin{pmatrix} 30 \\ 5 \end{pmatrix} (\frac{1}{2})^5 (\frac{1}{2})^{25} \\ &\approx 0.00016 < 0.01 \end{aligned}$$

så kan vi förkasta H_0 på 1% signifikansnivå.

7. En student gör en tenta där varje fråga besvaras med antingen 'sant' eller 'falskt'. Om studenten inte kan en fråga så gissar studenten ett svar. Vid rättningen av tentan justeras resultatet med formeln (3p)

$$Y = X_r - \alpha X_f$$

där X_r är antalet korrekta svar och X_f är antalet felaktiga svar från studenten. Man vill välja faktorn α så att det förväntade värdet av det justerade resultatet återspeglar studentens faktiska kunskaper, dvs. $E(Y)$ skall vara antalet frågor som studenten kunde svaret på. Hur skall faktorn α väljas?

Lösning: Om tentan består av N frågor, låt N_k vara antalet frågor som studenten kan och N_g antalet frågor som studenten gissar svaret på. Då är $X_r + X_f = N$ och vi vill att $E(Y) = N_k$. Då är

$$E(X_r) = N_k + N_g\frac{1}{2} = N_k + (N - N_k)\frac{1}{2} = \frac{1}{2}(N + N_k)$$

och

$$E(X_f) = \frac{1}{2}N_g = \frac{1}{2}(N - N_k).$$

Alltså får vi villkoret

$$\begin{aligned} E(Y) = N_k \quad &\Leftrightarrow \quad E(X_r) - \alpha E(X_f) = N_k \Leftrightarrow \frac{1}{2}(N + N_k) - \alpha\frac{1}{2}(N - N_k) = N_k \\ &\Leftrightarrow \quad \alpha\frac{1}{2}(N - N_k) = \frac{1}{2}(N - N_k) \Leftrightarrow \alpha = 1 \end{aligned}$$

8. Ett grossistföretag använder beställningspunktssystem för att styra ett större antal artiklar med medelhög efterfrågan både vad gäller volym och uttagsfrekvens. Inom logistikavdelningen är man dock inte överens om hur prognostiseringen ska läggas upp. Som underlag för diskussionen har en representativ produkt valts ut och den ska nu användas för en konsekvensanalys av hur en prognos (skattning av osäkerheten) bör beräknas om den ska utgöra underlag för säkerhetslagerdimensionering. Utgångspunkten är att SERV1 används för dimensionering av säkerhetslager och att en servicegrad på 95% används, vilket ger $k = 1,64$. Ledtiden för den aktuella produkten är fyra veckor (dvs en prognosperiod). Det är viktigt att motivera tydligt och förklara ev. skattningar som används i analysen nedan.

	Jan	Feb	Mars	April	Maj	Juni	Juli	Aug	Sept
Försäljningshistorik	100	107	111	122	126	122	134	137	140

Tabell 1: Försäljningshistorik för en artikel över de senaste nio månaderna

(a) Beräkna säkerhetslagret för artikeln, vars försäljningshistorik återfinns i Tabell 1, utan tidsseriedekomposition och med användning av hela tidsserien. (1p)

(b) Beräkna säkerhetslagret, för artikeln i Tabell 1, med tidsseriedekomposition. Utgå från en additiv efterfrågemodell som har komponenterna basnivå, trend och slump. Ange också hur många procent mindre/större säkerhetslager som behövs vid samma servicenivå som i deluppgift (a). (1.5p)

(c) Förklara vilken av de båda beräkningarna av säkerhetslager som bör användas och vad skillnaden innebär i termer av servicenivå kopplat till kapitalbindning. (0.5p)

9. Produktion i funktionell verkstad med flera olika produktionsgrupper (PG) är ofta svårplanerad då många olika artiklar tillverkas i låga volymer vilket gör att de planerade operationstiderna är osäkra (för återkommande artiklar i högre volymer kan data över operationstider samlas in för att få bättre underlag). För att hantera osäkerheten i tiderna så används köer vid de olika PG för att säkerställa att det hela tiden finns något att jobba med. Vid låg beläggningsgrad är köerna ett mindre problem då det finns gott om ledig kapacitet. När beläggningsgraden ökar så blir dock utmaningen större. Eftersom operationstiderna är osäkra så blir flödet svårt att styra i detalj och ur en PGs perspektiv kan därför inflödet av

arbete modelleras som att de skapas av en stokastisk process, vilken här antas vara exponentialfördelad. På samma sätt är operationsledtiden i den aktuella PG osäker. Vanligtvis brukar ledtider modelleras med t.ex. en Erlang-k fördelning av högre ordning men i det här fallet så kan operationstiden antas vara exponentialfördelad (vilken är en Erlang-fördelning av lägsta ordningen).

(a) Beräkna den förväntade kölängden vid PG som funktion av förväntad beläggningsgrad. Anta att den förväntade ankomstintensiteten är λ och den förväntade serviceintensiteten (operationstiden) är μ. Rita även en graf över den förväntade kölängden som funktion av beläggningsgrad. (1.5p)

(b) För att bestämma den planerade kölängden (dvs den kö som man ger plats för i planeringen) kan den förväntade kölängden beräknad i deluppgift (a) användas. Förklara vilka problem som kan uppstå om den förväntade kölängden används och motivera varför en längre eller kortare kötid (välj det du tycker verkar mest lämplig) bör användas. (0.5p)

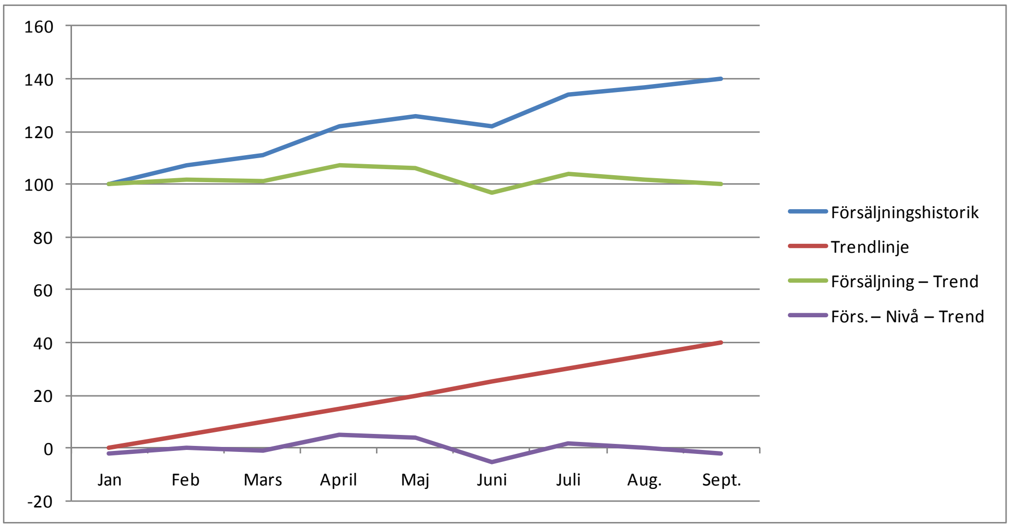
Lösning 8

Lösningsförslag

a) Säkerhetslager (SS) baserat på brist under ordercykel definieras som $SS = k \cdot \sigma_L$ (där k = 1,64 är givet i uppgiften). Ledtiden för artiklarna i uppgiften är lika lång som prognosperioden (L = 1) vilket innebär att den standardavvikelse som beräknas från historiken kan användas direkt i säkerhetslagerdimensioneringen: $SS = 1{,}64 \cdot \sigma$.

Standardavvikelsen för data kan beräknas som: $s = \sqrt{\frac{\sum_{i=Jan}^{Sept}(x_i - \bar{x})^2}{9-1}}$ där x_i tas från Tabell 1, vilken också är underlag för beräkning av medelvärdet $\bar{x} = 122{,}11$. Detta ger $s = \sqrt{\frac{1539}{8}} = 13{,}87$ st vilket ger ett säkerhetslager $SS = 1{,}64 \cdot 13{,}87 = 22{,}75$ st.

b) Nivå- och trendkomponenten kan beräknas med statistiska metoder som t.ex. regressionsanalys men de teknikerna har inte behandlats i kursen. I brist på dessa får man göra en skattning av trend och nivå genom inspektion av data. En bra start är som alltid att plotta data som i Figur 1.



Figur 1. Graf över försäljningsdata och dekompositionsdata

De data som ligger till grund för Figur 1 finns i Tabell 2. Enligt uppgiften ska en efterfrågemodell antas som är: $E_t = B + T \cdot t + \text{Slump}$. I det här fallet kan man ganska enkelt se att det finns en trend i materialet och ett exempel på skattning kan man få genom att lägga en linjal över grafen och ta ut en trend. I det här fallet är det inte orimligt att utgå från den sista data som är 140 st och dra en rät linje till första data som är 100 st. Det ger en trend på T = (140–100)/(9–1) = 5 st/månad vilken är inlagd i rad 2 i Tabell 2. Trendkomponenten kan då dras bort ifrån försäljning vilket ger rad 3. Baserat på rad 3 kan basnivån beräknas som medelvärdet av rad 3 vilket ger B = 102,11 st. Slutligen dras basnivån bort och det är den slumpmässiga komponenten S som återstår på rad 4.

	Jan	***Feb***	***Mars***	***April***	***Maj***	***Juni***	***Juli***	***Aug.***	***Sept.***

Försäljningshistorik	100	107	111	122	126	122	134	137	140
Trendlinje	0	5	10	15	20	25	30	35	40
Försäljning – Trend	100	102	101	107	106	97	104	102	100
Förs. – Nivå – Trend	-2,1	-0,1	-1,1	4,9	3,9	-5,1	1,9	-0,1	-2,1

Tabell 1. Försäljningsdata och dekompositionsdata

Standardavvikelsen för den rensade tidsserien på rad 4 kan nu skattas med $s = \sqrt{\frac{79}{8}} = 3{,}14$ st vilket ger ett säkerhetslager $SS = 1{,}64 \cdot 3{,}14 = 5{,}15$ st.
Den procentuella förändringen är $\frac{5{,}15-22{,}75}{22{,}75} = -77\%$. Kostnaden för kapitalbindning i säkerhetslager blir alltså 77% lägre vid en dekompositionsanalys.

c) Teknikerna med dekomposition (MD) och utan dekomposition (UD) utgår från två olika scenarios. I UD är ansatsen att ingen trend finns och därmed bör trenden inte heller finnas med i själva styrningsmodellen utan där får säkerhetslagret försöka ta den rollen. Modellen är dock inte bra och det finns en hel del frågetecken kring hur nivån, som planeringen använder, ska beräknas då det sker en efterfrågeökning (trend) som inte beaktas/bekräftas enligt den prognosmodell som används. Servicenivån kommer därför att försämras med tiden eftersom efterfrågeökningen inte beaktas. Om istället MD används så bör styrlogiken utgå ifrån att det finns ökande behov att ta hänsyn till (t.ex. genom att höja beställningspunkten över tiden om ett beställningspunktssystem används). Säkerhetslagret behöver endast täcka upp avvikelserna från denna trend och den varierar inte så mycket över planeringshorisonten. I det här fallet kommer kapitalbindningen att bli relativt låg och servicenivån ungefär densamma över hela planeringshorisonten. Alternativet MD är således genomgående bättre än UD vad gäller kapitalbindning i kombination med att säkerställa en förväntad servicenivå.

Poängfördelning enligt rättningsmall

a) 0,5 poäng för beräkning av standardavvikelse och 0,5 poäng för beräkning av säkerhetslager.

b) 1 poäng för beräkning av standardavvikelse med dekomposition och 0,5 poäng för jämförelse av säkehetslager dimensionerat med respektive utan dekomposition.

c) 0,5 poäng för diskussion och tolkning.

Referenser

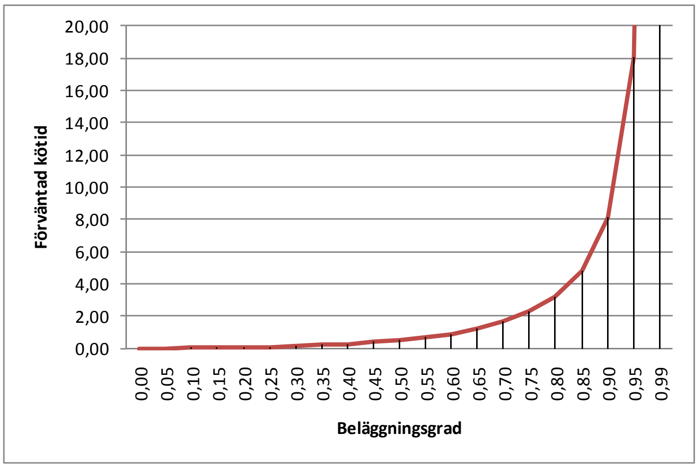
a-c) Presentationsunderlag: Ti1 – Prognostisering och
Presentationsunderlag: Ti2 – Säkerhetslagerdimensionering

Lösning 9

Lösningsförslag

a) ”Förväntad” beläggningsgraden (FBG) definieras som: $\mathrm{FBG} = \frac{\mathrm{E[Kapacitetsbehov]}}{\mathrm{E[Nettokapacitet]}}$. Ett mått på ”förväntat” kapacitetsbehov (FKB) kan definieras baserat på ankomstintensiteten: $\mathrm{FKB} = \frac{1}{\mathrm{E[Tid\ mellan\ nya\ order]}} = \lambda$ och på motsvarande sätt kan ”förväntad” nettokapaciteten (FNK) definieras baserat på serviceintensiteten: $\mathrm{FNK} = \frac{1}{\mathrm{E[Tid\ mellan\ färdiga\ order]}} = \mu$

Beläggningsgraden kan således skrivas: $FBG = \frac{\lambda}{\mu} = \rho$. För den här typen av M/M/1-köer kan förväntad tid i produktionsgruppen beräknas till $E[\text{Tid i PG}] = \frac{\rho}{1-\rho}$ och förväntad kötid blir då (om servicetiden, dvs operationstiden, dras bort) $E[\text{Kötid i PG}] = \frac{\rho^2}{1-\rho}$ där ρ är beläggningsgraden, se även Figur 2.



Figur 2. Förväntad kötid som funktion av beläggningsgrad.

b) Den förväntade kötiden utgör ett ”medelvärde” för kölängden men i praktiken kan kön under perioder vara mycket längre. För att inte riskera alltför många störningar (förseningar) i planen är det därför vanligt att använda planerade kötider som avsevärt överstiger den förväntade kötiden. Tyvärr innebär detta som konsekvens långa ledtider men det är ett pris som man ofta väljer att ”betala” för att säkerställa utnyttjandegraden.

Poängfördelning enligt rättningsmall

a) 1 poäng för modellering av FBG och FKB och 0,5 poäng för modellering av beläggningsgrad och plottning.

b) 0,5 poäng för att visa förståelse för att den verkliga kötiden under perioder avsevärt kan överstiga den förväntade kötiden.

Referenser

a,b) Presentationsunderlag: Ti6 – Beläggningsanalys 2