

**TENTAMEN:** Matematisk statistik för K (TMA072 och TMA073)

Torsdagen den 28 maj 2009, kl 14:00-18:00

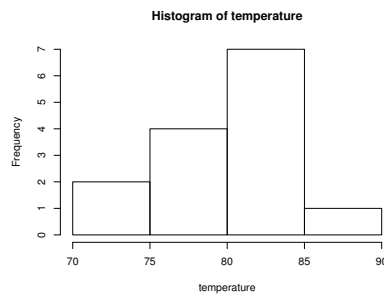
**Lärare och jour:** Aila Särkkä, telefon 772 3542

**Jour:** Tommy Norberg, telefon 0730 79 42 09

**Hjälpmedel:** Formelsamling, tabeller (även BETA, Physics Handbook, skoltabeller, t.ex. TEFYMA), valfri miniräknare.

- 1) Ett varningssystem består av två komponenter. Var och en av dessa komponenter kan ge signal om en olycka inträffar. Tyvärr kan hela systemet förstöras vid vissa typer av olyckor. Erfarenheter har visat att komponent  $A$  fungerar med sannolikheten 0.95 medan komponent  $B$  fungerar med sannolikheten 0.93. Sannolikheten att båda kommer att fungera vid en olycka är 0.90.
  - a) Är händelserna "A fungerar" och "B fungerar" oberoende?
  - b) Antag att komponent  $B$  fungerar. Hur stor är sannolikheten att även  $A$  fungerar?
  - c) Beräkna sannolikheten att olyckor där komponent  $A$  har blivit förstörd även förstör komponent  $B$ ? (4p)
- 2) Enligt en meteorolog är den genomsnittliga (medel/median) dagliga högsta temperaturen i juli i Pittsburgh  $83^\circ$  Fahrenheit. Nedan hittar du den högsta temperaturen under 14 slumpmässigt valda dagar i juli och ett histogram av värdena.

74, 79, 81, 90, 79, 81, 83, 81, 74, 78, 76, 84, 82, 85



- a) Använd ett lämpligt parametriskt test ( $Z$ -test eller  $T$ -test) och signifikansnivån 0.01 för att se om meteorologen har rätt. (Medelvärdet av observationerna ovan är  $80.5^\circ$  och standardavvikelsen  $4.36^\circ$ .) Tolka resultatet.
- b) Samma som i a) men använd teckentestet. Tolka resultatet.
- c) Vilka antaganden har du gjort i a) och b)?
- d) Vilket test skulle du rekommendera? Varför? (5p)

3) Antag att  $X$  har täthetsfunktionen

$$f(x) = \begin{cases} a + bx & \text{för } |x| < 1 \\ 0 & \text{annars} \end{cases}$$

Vilka värden antar  $a$  och  $b$  om väntevärdet av  $X$  är  $\frac{1}{5}$ ? (4p)

4) En studie genomfördes där man ville skatta den genomsnittliga vikten för en viss typ av får. Vikten antas vara en normalfördelad variabel med standardavvikelsen 5 (kg). Efter att ha vägt 100 får så erhöll man konfidensintervallet

$$92.6 \leq \mu \leq 95.2.$$

för vikten (enheten kg).

a) Vilken konfidensgrad (i procent) har intervallet?

b) Hur många observationer borde man ha undersökt om man vill ha samma intervallängd men en 95%-ig konfidensgrad? (4p)

5) En grupp sjuksköterskor vill studera om antalet mödravårdsbesök (före födelsen) påverkar födelsevikten av barnet så att mödrar med ett stort antal besök skulle föda tyngre barn än de med ett litet antal besök. Mödrarna (tagna slumpmässigt från födelseregistret) delades i två grupper: 14 mödrar hade haft 5 eller färre mödravårdsbesök före födelsen och 14 mödrar 6 eller fler. Låt  $X$  (5 eller färre) och  $Y$  (6 eller fler) vara de respektive födelsevikterna hos bebisarna i dessa två grupper av mödrar.

a) Formulera noll- och mothypotesen för att hjälpa sjuksköterkorna med sin frågeställning. Motivera ditt val av mothypotes.

b) Testa hypotesen i a) på signifikansnivån 5% genom att använda viktorna (i uns) i tabellen nedan:

$x$	49	108	110	82	93	114	134	114	96	52	101	114	120	116
$y$	133	108	93	119	119	98	106	131	87	153	116	129	97	110

( $\bar{x} = 100.21$ ,  $s_X = 24.59$ ,  $\bar{y} = 114.21$ ,  $s_Y = 18.15$ ,  $\bar{d} = -14$  och  $s_D = 38.50$ , där  $d = x - y$ )

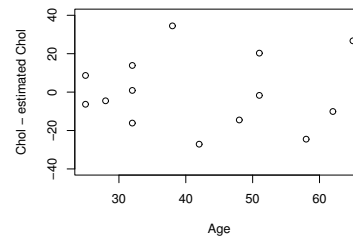
c) Skulle du få samma resultat om du hade haft en annan mothypotes (ensidig om du använde tvåsidig eller tvåsidig om du använde ensidig)?

d) Vilka antaganden har du gjort i b)? Varför? (4p)

- 6) En läkare är intresserad av hur patienters kolesterolvärde påverkas av åldern. Hon väljer slumpmässigt 14 av sina manliga patienter och får följande data:

Åldern (Age)	Kolesterol (Chol)
25	180
25	195
28	186
32	180
32	210
32	197
38	239
42	183
48	204
51	221
51	243
58	208
62	228
65	269

- a) Skatta regressionslinjen för att beskriva det linjära sambandet mellan kolesterolvärde och åldern. Tolka!  
 b) Baserat på den skattade regressionslinjen, vad kan du säga om kolesterolvärde vid 40-års ålder och vid 80-års ålder?



- c) Beskriva och tolka figuren ovan.  
 d) Vilka antaganden har du gjort i a)?

( $\sum c_i = 2943$ ,  $\sum a_i = 589$ ,  $\sum c_i^2 = 628055$ ,  $\sum a_i^2 = 27253$ ,  $\sum c_i a_i = 127276$ , där  $c$  är kolesterolvärdet och  $a$  åldern och summan alltid är över alla 14 värden) (5p)

- 7) Man vill göra inferens om ett okänt väntevärde  $\mu$ . För detta har man konstruerat ett  $T$ -test och använt den nedanstående MatLab-koden. De observerade värdena är i vektor  $x$ .

```
>> length(x)
ans =
    100
>> mean(x)
ans =
    1.0980
>> std(x)
ans =
    0.9213
>> [h p ci]=ttest(x,0,0.05,'both')
h =
     1
p =
 7.6804e-021
ci =
    0.9152    1.2808
>>
```

- Ge noll- och mothypotesen (som motsvarar MatLab-koden).
- Vilken signifikansnivå har man använt och vad är slutsatsen av testet? Förklara.
- Ge (beräkna) ett 99% konfidensintervall för  $\mu$ . (4p)

**Lycka till!**