

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

Måndagen den 8 mars 2003, kl 14:15-18:15

**Lärare:** Aila Särkkä

**Jour:** Aila Särkkä

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

1) Anta att man har två händelser  $A$  och  $B$  och att  $P(A) = 0.4$  och  $P(B) = 0.2$ .

a) Anta att  $A$  och  $B$  är oberoende. Räkna  $P(A \cap B)$ ,  $P(A|B)$  och  $P(A \cup B)$ .

b) Anta att  $A$  och  $B$  är beroende och att  $P(A|B) = 0.6$ . Räkna  $P(A \cup B)$  och  $P(B|A)$ .

(4p)

2) a) Vilka två egenskaper bör en bra punktskattning ha?

b) Har medelvärdet dessa egenskaper då det används för att skatta väntevärdet? Förklara.

(3p)

3) Är pojkar genomsnittligt bättre i matte än flickor? För att svara på frågan jämför man poäng i ett mattetest av pojkar och flickor på åttonde klass. Testpoäng var följande:

	Pojkar	Flickor
Stickprovsstorlek	32	34
Medelvärde	48.9	48.4
Standardavvikelse	12.96	11.85

a) Testa om de genomsnittliga mattetestpoängen för pojkar och flickor är olika.

b) Använd ett 90% konfidensintervall för att skatta den sanna skillnaden mellan pojkar och flickor. Stöttar konfidensintervallet resultatet av testet som du gjorde i del a)?

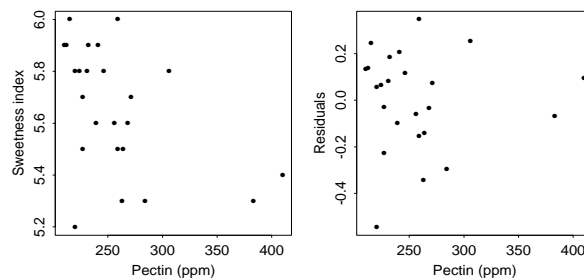
c) Vad måste du anta om fördelningarna av populationerna för att kunna garantera giltigheten av testet i a) och konfidensintervallet i b)?

(5p)

- 4) Man studerar förorenad jordmån. Man har 72 prov av jordmånen (400g var) som man har torkat och analyserat för cyanid. Medelcyanidnivån i stickprovet är  $\bar{x} = 116\text{mg/kg}$  och standardavvikelsen  $s = 80\text{mg/kg}$ .
- Testa hypotesen att den sanna cyanidnivån i jordmånen är högre än  $100\text{mg/kg}$ . Använd signifikansnivån  $0.10$ .
  - Skulle du göra samma slutsats än i del a) om signifikansnivån var  $0.05$ ? Eller  $0.01$ ? Varför kan olika signifikansnivåer leda till olika beslut?

(5p)

- 5) Kvaliteten av apelsinjuice från en viss tillverkare övervakas hela tiden. Det finns många sensoriska och kemikaliska komponenter som påverkar smaken av apelsinjuice. En av dem är ett så kallat söthetsindex. Man skulle vilja veta om det finns ett samband mellan söthetsindexet och mängden av vattenlöslig pektin (i ppm) i apelsinjuice. Tillverkaren vill använda enkel linjär regression för att prediktera söthetsindexet ( $Y$ ) från mängden av pektin ( $X$ ).
- Från ett stickprov av storlek 24 har man räknat att  $\sum x_i = 6167$ ,  $\sum y_i = 135.8$ ,  $\sum x_i^2 = 1641115$ ,  $\sum y_i^2 = 769.72$  och  $\sum x_i y_i = 34764.5$ . Skatta regressionslinjen.
  - Förklara figurerna nedan. Är den linjära regressionsmodellen en bra modell enligt dem?



(4p)

6) Anta att alkalinitet av vattenprov tagna från Han River i Seoul, Korea, är normalfördelad med väntevärdet 50mg/l och standardavvikelsen 3.2mg/l.

a) Vad är sannolikheten att ett vattenprov från flodet har alkalinitet som är högre än 55mg/l?

b) Anta att vi har fem (oberoende) vattenprov från flodet. Vad är sannolikheten att minst en av dem har alkalinitet som är högre än 55mg/l?

(5p)

7) Anta att man har ett stickprov av storlek 20 från en normalfördelning med väntevärde  $\mu$  och varians  $\sigma^2 = 4$ . Man vill testa  $H_0 : \mu = 10$  mot  $H_1 : \mu \neq 10$  på signifikansnivån 0.05. Teststatistikan man använder är den vanliga, dvs.  $(\bar{X} - \mu)\sqrt{n}/\sigma$ . Anta att det sanna väntevärdet är 9. Vad är sannolikheten att testet inte kan upptäcka detta? (4p)

**Lycka till!**