

TENTAMEN: Matematisk statistik för K (TMA072)

Onsdagen den 24 maj 2006, kl 14:00-18:00

Lärare: Aila Särkkä

Jour: Aila Särkkä, telefon 772 35 42

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

1) Antag att man har två händelser A och B och att $P(A) = 0.4$ och $P(B|A) = 0.3$.

a) Vad betyder $B|A$?

b) Är A och B disjunkta händelser? Varför?

Låt $P(B) = 0.5$ (i c) och d)).

c) Beräkna $P(A \cup B)$.

d) är A och B oberoende? (4p)

2) Definiera

a) p -värdet

b) signifikansnivån av testet (2p)

3) IQ-värdena för 600 sökande till en viss högskola är normalfördelade med väntevärdet 115 och standardavvikelsen 12. Om högskolan kräver att IQ är minst 95, hur många av dessa studenter kommer inte in? (3p)

4) Datamängden nedan beskriver hur många timmar en laddbar häckklippare fungerar innan den behövs att ladda om: 1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2 and 1.7. Testa hypotesen på signifikansnivån 5% att denna häckklippare fungerar genomsnittligt (använd väntevärdet eller medianen) 1.8 timmar innan den behövs att ladda om

a) genom att använda ett vanligt Z - eller T -test. Ge antaganden du har gjort.

b) genom att använda ett ickeparametriskt test. Ge antaganden du har gjort.

c) Jämför resultaten i a) och b). Varför tror du att resultaten är lika/olika?

(5p)

- 5) Man har länge använt kedjor i brännugnar i cementanläggningar för att minska värmeförbrukning. Man studerar om kedja har samma effekt när billigare material med hög svavel- och klorinnehåll används. Man använder två oberoende stickprov:

Utan kedjor	Med kedjor
$n_1 = 16$	$n_2 = 14$
$\bar{x}_1 = 6150\text{kJ/kg}$	$\bar{x}_2 = 5250\text{kJ/kg}$
$s_1 = 80\text{kJ/kg}$	$s_2 = 75\text{kJ/kg}$

Är det bra att använda kedjor. Förklara varför/varför inte genom att

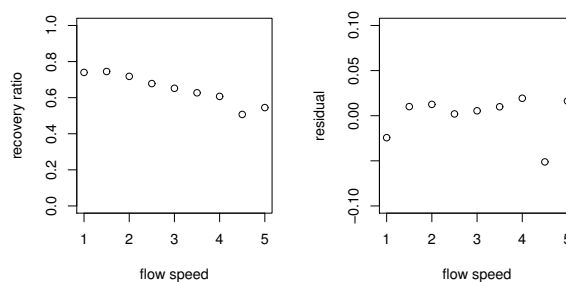
- hitta ett 99% konfidensintervall för $\mu_1 - \mu_2$, dvs. differensen i värmeförbrukningen i brännugnar med och utan kedjor.
 - använda ett lämpligt hypotestest. Du kan använda signifikansnivån 0.01.
 - Vilka antaganden har du gjort i a) och b)? (5p)
- 6) Forskare experimenterar med en ny förening som används för att binda teflon till stål. Föreningar som används numera behöver genomsnittligt 3 minuter att torka. Forskarna tror att den nya föreningen torkar fortare.
- Bestäm nollhypotesen och mothypotesen som behövs för att testa om den nya föreningen torkar snabbare än den nyvarande.
 - Vilka är de praktiska konsekvenserna om man gör ett fel av typ I/typ II?
 - Enligt en tidigare studie kan standardavvikelsen av torkningstiden antas vara 0.5. Antag att den nya produkten är värd att marknadsföra om man kan visa att den genomsnittliga torkningstiden är 2.5 minuter eller mindre. Hur stort stickprov behöver man för att upptäcka att torkningstiden är 2.5 minuter med sannolikhet 0.95 om signifikansnivån α är 0.05?
 - Efter att ha gjort experimentet har man fått följande data: 1.4, 2.1, 2.8, 0.9, 2.4, 1.7, 3.7, 2.7, 2.6, 1.9, 2.8, 2.8, 2.2, 2.2, 3.4, 1.9. Medelvärdet av mätningarna är 2.34 och standardavvikelsen 0.71. Testa nollhypotesen i a) på signifikansnivån 0.05. Skulle du rekommendera att marknadsföra denna produkt?
 - Vilka antaganden har du gjort i c) och d)? (8p)

- 7) Man vill studera återhämtning av värme som man förlorar till omgivningen i form av avgaser från masugnar. Man har fixa värden av strömhastigheten (m/s) och man mäter återhämtningskvoten:

Strömhastigheten (x_i)	Återhämtningskvoten (y_i)
1.0	.740
1.5	.745
2.0	.718
2.5	.678
3.0	.652
3.5	.627
4.0	.607
4.5	.507
5.0	.545

Man har $\sum x_i = 27$, $\sum y_i = 5.819$, $\sum x_i^2 = 96$, $\sum y_i^2 = 3.819$ och $\sum x_i y_i = 16.574$. I samtliga summor går i från 1 till 9.

- a) Hitta regressionslinjen $\mu_{Y|x} = \beta_0 + \beta_1 x$.
b) Förklara bilderna nedan. Verkar en enkel linjär regressionsmodell vara



en lämplig model i detta fall?

- c) Hurdana problem kan man upptäcka genom att titta på residualplottar generellt? (3p)

Lycka till!