

TENTAMEN: Matematisk statistik för K (TMA073 och TMA072)

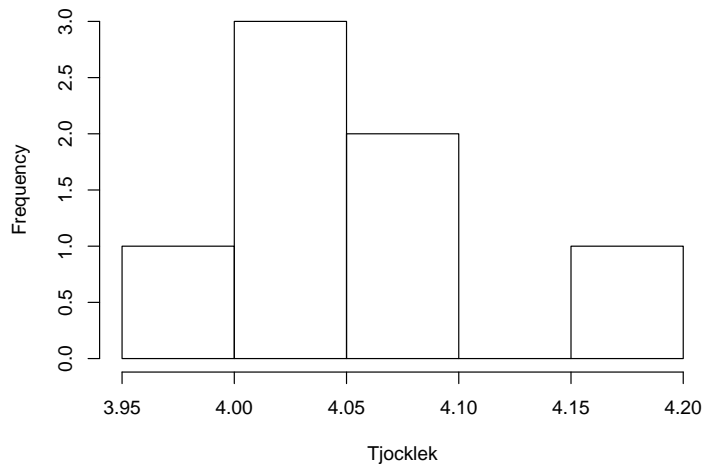
Fredagen den 27 maj 2011, kl 14:00-18:00

Lärare och jour: Aila Särkkä, telefon 772 3542

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

- 1) I en låda finns det 10 komponenter av vilka 3 är felaktiga. Man tar ut två komponenter på måfå ur lådan. Låt A vara händelsen att den första komponenten man tar är felaktig, och B händelsen att den andra man tar är felaktig. Hitta
 - a) $P(A)$
 - b) $P(B|A)$
 - c) $P(B)$
 - d) Är A och B oberoende händelser? Varför/varför inte? (4p)
- 2)
 - a) För vilka värden på konstanten c är $f(x) = cx(2-x)$, $0 \leq x \leq 1$ (och $f(x) = 0$ annars) en täthetsfunktion?
 - b) Låt X vara en stokastisk variabel med täthetsfunktionen i a). Hitta fördelningsfunktionen för X .
 - c) Hitta väntevärdet för X .
 - d) Hitta sannolikheten $P(0.5 < X \leq 1.2)$. (4p)
- 3) I ett stickprov av 100 stålkablar är medelvärdet av brytningsstyrkan 50kN. Från en tidigare studie vet man att standardavvikelsen av brytningsstyrkan är 2kN.
 - a) Hitta ett 95% konfidensintervall för den genomsnittliga brytningsstyrkan för den här typen av stålkabel.
 - b) En ingenjör har hittat konfidensintervallet $[49.7, 50.3]$ för den genomsnittliga brytningsstyrkan. Vilken konfidensgrad har hon använt?
 - c) Vilka antaganden måste man göra i a) och b)? (4p)
- 4) En ingenjörsförening utförde en studie för att undersöka medianåldern av doktorander som disputerar inom ingenjörsvetenskaper. Enligt föreningen är medianåldern högre än 30 år. Man valde 20 doktorander på måfå för studien. Bland dem var det 4 som hade doktorerat innan de hade fyllt 30.
 - a) Testa på signifikansnivå 0.01 om föreningen verkar ha det rätt?
 - b) Vilka antaganden har du gjort i a)?
 - c) Vad är den minsta signifikansnivån för att kunna förkasta nollhypotesen? (4p)

- 5) Enligt reglerna borde vägg tjockleken av 2-liters polykarbonatflaskor vara genomsnittligt 4.0 mils. Man tar ett stickprov som innehåller 7 flaskor och mäter vägg tjockleken av varje flaska. Resultaten blir: 4.065, 3.967, 4.028, 4.008, 4.195, 4.057 och 4.010. Man kan räkna från datavärdena att medelvärdet är 4.05 och stickprovsvariansen 0.0053.



Baserat på datan ovan vill man testa om den genomsnittliga vägg tjockleken skilljer sig från 4.0 mils.

- Ge noll- och mothypoteserna.
- Förklara de praktiska konsekvenserna av att göra ett fel av typ I och typ II när man gör ett hypotestest baserat på a).
- Utför ett hypotes- eller signifikanstest för att testa hypotesparet i a). Använd signifikansnivå 0.05. Kan man säga att den genomsnittliga vägg tjockleken skilljer sig från 4.0 mils?
- Vilka antaganden har du gjort i c)? Verkar de rimliga?
- Vilket fel (typ I eller typ II) kan du ha gjort i c)? Varför? (5p)

- 6) Man har utvecklat en ny metod för att tillverka en viss kemikalie och man hoppas på att den nya metoden skulle vara billigare än den som numera används (gamla metoden). Man tillverkade 100l kemikalie genom att använda var och en av metoderna 6 gånger och räknade kostnaden för varje körning. Resultaten (i dollar):

| | | | | | | |
|----------------------|----|----|----|----|----|----|
| Ny metod (X) | 51 | 52 | 55 | 53 | 54 | 53 |
| Gammal metod (Y) | 50 | 54 | 59 | 56 | 50 | 58 |

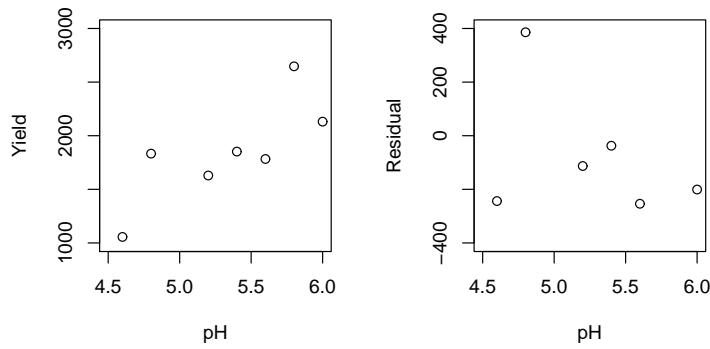
Från data kan man räkna att $\bar{x} = 53$, $s_X = 1.41$, $\bar{y} = 54.5$, $s_Y = 3.89$, $\bar{d} = -1.5$ (där $d_i = x_i - y_i$) och $s_D = 3.39$ (där s_D är standardavvikelsen av skillnaderna).

- Ge noll- och mothypoteserna.
- Vilket test kan man använda för att testa hypotesparet i a)? Varför?
- Utför testet i b) på signifikansnivå 0.10. Stödjer data påståendet att den nya metoden är billigare än den gamla?
- Ge antaganden du har gjort i c).
- Anta att man vet att den sanna skillnaden av de genomsnittliga priserna (ny metod - gammal metod) är 2 dollar. Ge en formel (i form av en sannolikhet) för teststyrkan genom att använda teststatistikan i c) och den sanna skillnaden. (5p)

- 7) En jordbruksforskare planerade alfalfa på flera jordtyper, som var identiska utom att det fanns olika pH värden i jordmånen i de olika ställena. Nedan hittar du viktena (i "pounds per acre") av skörden för varje yta.

| pH (x) | 4.6 | 4.8 | 5.2 | 5.4 | 5.6 | 5.8 | 6.0 |
|---------------|------|------|------|------|------|------|------|
| Skörd (y) | 1056 | 1833 | 1629 | 1852 | 1783 | 2647 | 2131 |

Vi har att $\sum x_i = 37.4$, $\sum x_i^2 = 201.4$, $\sum y_i = 12931$, $\sum y_i^2 = 25285429$ och $\sum x_i y_i = 70251$, där summan är alltid från 1 till 7.



- Skatta regressionslinjen för att beskriva det linjära sambandet mellan pH och vikten av skörden (dvs. att hur vikten av skörden beror på pH). Tolka resultatet!
- Hur definierar man "Residual" på bilden ovan till höger?
- Är den skattade linjära regressionsmodellen bra? Varför/varför inte?
- Om man ökar pH med 0.1, hur mycket skulle du förutsäga att vikten av skörden skulle minska/öka? (4p)

Lycka till!