

Veckoblad 6

Kapitel 9 i Matematisk statistik, Blomqvist U.

(Läs: 9.1 – 9.5, 9.8)

Nya begrepp: Konfidensgrad, hur längd av konfidensintervall beror av konfidensgrad, fördelningens varians och stickprovsstorlek

Övningar att räkna: 9.1-9.13, 9.35, 9.36, 9.39, 9.40

Snabbrepetition:

1. Ett observerat konfidensintervall ex.vis $23.4 < \mu < 28.6$ (t.ex. ett 95%-igt) är en lögn eller en sanning. 95% anger att metoden fungerar i 95% av fallen.
2. Konfidensintervall för μ vid normalfördelning.
 - Om σ är känd använder vi normalfördelningsfraktilen. Detta gör vi även vid stora stickprovsstorlekar då σ är okänd.
 - Om σ är okänd använder vi t-fördelningen (vid små stickprovsstorlekar). Om σ är okänd använder vi det skattade värdet av σ , som vi kallar s . Det observerade värdet av denna är stickprovsstandardavvikelsen

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}}$$

- Om vi ej har normalfördelning kan vi beräkna konfidensintervall för μ vid stora stickprovsstorlekar genom att utnyttja att $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ är appr. $N(0,1)$
3. Om vi ex.vis vill ha ett nedåt begränsat konfidensintervall för σ kan det vara lämpligt att först sätta upp sannolikheten $P\left(\frac{(n-1)s^2}{\sigma^2} < \chi^2\right) = F(x)$ där F är fördelningsfunktionen för $\chi^2(n-1)$ -fördelning. Observera vilket håll olikheten går. Gör man sedan om olikheten får man ett nedåt begränsat intervall $\sigma^2 > \frac{(n-1)s^2}{\chi^2}$
 4. Vid ett konfidensintervall för μ gäller att:
 - Intervallets längd blir mindre om n ökar.
 - Intervallets längd blir mindre om vi väljer lägre konfidensgrad.
 - Intervallets längd blir mindre om σ varit mindre.

Gammalt tentamenstal (data/elektro 110319):

Vid en undersökning av alkohols inverkan på reaktionstiden på 6 slumpmässigt utvalda personer fick man följande resultat (tid i sekunder):

Person	före alkohol	efter alkohol
A	0.15	0.55
B	0.10	0.60
C	0.10	1.00
D	0.25	0.55
E	0.25	0.50
F	0.10	0.35

Anta att reaktionstiderna före och efter alkohol är normalfördelade.

- Beräkna ett 95%-igt konfidensintervall för den genomsnittliga skillnaden mellan reaktionstiderna före och efter alkohol.
- Tyder denna undersökning på att alkohol påverkar reaktionstiden? Du måste motivera ditt svar för att få poäng.
- Beräkna ett 95%-igt ensidigt uppåt begränsat konfidensintervall för motsvarande standardavvikelse.

(6 poäng)

Kapitel 10 i Matematisk statistik, Blomqvist U.

(Läs: hela kapitel 10)

Nya begrepp: Enkel linjär regression, icke-linjär regression, multipel linjär regression, minsta-kvadrat-metoden, intercept, riktningskoefficient, korrelationskoefficienten, determinationskoefficienten, normalekvationerna

Övningar att räkna: 10.1, 10.5, 10.7-10.10, 10.13, 10.16

Snabbrepetition:

- En rät linje, $y = a + bx$, kan beräknas med hjälp av ekvationerna

$$a = \bar{y} - b\bar{x} \quad \text{och} \quad b = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}$$

x väljs om möjligt som den oberoende variabeln och y som den beroende.

- a kallas *interceptet* och b kallas *riktningskoefficienten*.
- Den metod man brukar använda för att anpassa en rät linje kallas för *minsta-kvadratmetoden*. Konfidensintervall för μ vid normalfördelning.

8. För att ta reda på om denna räta linje beskriver vårt undersökta material bra så kan vi beräkna Pearsons produkt-moment-korrelationskoefficient, r , vanligtvis bara kallad *korrelationskoefficienten*. Man kan visa att $-1 \leq r \leq 1$.

9. Korrelationskoefficienten beräknas med hjälp av formeln

$$r = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right) \left(\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n} \right)}}$$

10. Observera att ett värde på r som ligger nära $+1$ eller -1 inte automatiskt innebär att det finns ett orsak-verkan samband.

11. Ett annat mått som ofta beräknas är *determinationskoefficienten*, R^2 , även kallad *förklaringsgraden*. När man har två variabler, x och y , så är $R^2 = r^2$, d.v.s. förklaringsgraden är korrelationskoefficienten i kvadrat.

12. Ett icke-linjärt samband $y = a \cdot b^x$ kan beräknas genom att ekvationen logaritmeras, $\ln y = \ln a + x \cdot \ln b$.

13. En multipel regressionsmodell $y = a + b_1 x_1 + b_2 x_2$ kan uppskattas med hjälp av de s.k. normalekvationerna

$$\begin{cases} a \cdot n + b_1 \sum_{i=1}^n x_{1i} + b_2 \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i} x_{2i} = \sum_{i=1}^n x_{1i} y_i \\ a \sum_{i=1}^n x_{2i} + b_1 \sum_{i=1}^n x_{1i} x_{2i} + b_2 \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n x_{2i} y_i \end{cases}$$

Gammalt tentamenstal (data/elektro 991022):

En undersökning om studenters studievevanor genomfördes. Bland annat frågades om ålder, kön och antal timmar/vecka utöver schemalagd tid som man ägnade åt sina studier. Från denna undersökning har fem studenter valts ut och deras svar finns i tabellen på nästa sida.

Student nr	ålder	antal timmar	kön
1	19	11	Man
2	20	22	Kvinna
3	18	14	Kvinna
4	19	5	Man
5	22	19	Man

- a) Beräkna a , b_1 , och b_2 i en multipel regressionsmodell, $y = a + b_1x_1 + b_2x_2$, där y = antal timmar, x_1 = ålder och x_2 = kön. Vid beräkningarna skall variabeln kön kodas med Man=0 och Kvinna=1.
- b) Tolka a , b_1 och b_2 i termer av antal timmar, ålder och kön. (Observera att om du bara förklarar koefficienterna med hjälp av termerna y , x_1 och x_2 så erhålles inga poäng på b-uppgiften).