

# Extrauppgifter - Statistik

## Uppgifter

1. Den stokastiska variabeln  $Y \sim t_{10}$ . Bestäm  $c$  så att  $P(-c < Y < c) = 0.95$ .

2. Vid tillverkning av en viss sorts färg tillsätts färgpigmentet med hjälp av en doseringsapparat, men man är inte säker på om den har tillräcklig precision. Om man ställer in apparaten på dosen  $\mu$  blir den verkliga dosen  $X = \mu + \epsilon$ , där  $\epsilon \sim N(0, \sigma^2)$ . Man vill att minst 99% av burkarna ska ha en acceptabel pigmentdos, dvs en dos i intervallet  $(\mu - 0.2, \mu + 0.2)$ , i det långa loppet.

Man genomförde 51 provdoseringar med  $\mu = 10$  och fick stickprovsstandardavvikelsen  $s = 0.0629$ . Uppfyller apparaten precisionskravet vid  $\mu = 10$ ? Besvara frågan med hjälp av ett lämpligt 95 % konfidenstervall.

3. För att jämföra effekterna hos tre olika blodtryckssänkande mediciner behandlades tre grupper om vardera 10 patienter med de olika medicinerna. Efter en månad mättes sänkningen av blodtrycket.

	Medelvärde	Stickprovsstandardavvikelse
Medicin 1	17.3	6.19
Medicin 2	21.1	7.26
Medicin 3	10.8	5.23

Modell: Tre oberoende stickprov från  $N(\mu_i, \sigma^2)$ ,  $i = 1, 2, 3$ .

Konstruera ett konfidensintervall för  $\sigma$  av typen  $(0, a)$  med konfidensgrad 0.95.

4. Betjäningstiderna för ett kösystem är exponentialfördelade med väntevärde  $\mu$ . Man har observerat 80 oberoende betjäningstider och fått medelvärdet  $\bar{x} = 4.5$  minuter.
  - (a) Konstruera ett konfidensintervall för  $\mu$  med konfidensgraden approximativt 0.95.
  - (b) Låt  $p$  vara sannolikheten att en betjäningstid är mer än tio minuter. Konstruera ett konfidensintervall för  $p$  med konfidensgraden approximativt 0.95.
  
5. På misstänkta rattfyllerister gör man tre bestämningar av alkoholhalten i blodet. Resultaten  $x_1$ ,  $x_2$  och  $x_3$  antas utgöra ett slumpmässigt stickprov från  $N(\mu, 0.05^2)$  där  $\mu$  är den verkliga alkoholhalten i blodet. Om  $\mu > 0.2$  har personen gjort sig skyldig till rattonykterhet. Låt oss anta att den domstol som skall döma tar hänsyn till osäkerheten i mätningarna genom att beräkna det aritmetiska medelvärdet  $\bar{x}$  av de tre analysresultaten och därefter förklara personen skyldig om

$$\bar{x} > 0.2 + \lambda_{0.01} \cdot 0.05 / \sqrt{3}$$

men oskyldig annars. Med statistisk terminologi kan man säga att domstolen prövar  $H_0 : \mu = 0.2$  mot  $H_1 : \mu > 0.2$  på nivån 0.01. Vilka av följande påståenden ger en någorlunda korrekt beskrivning av vad som kommer hända i det långa loppet?

- (a) högst 1% av alla frikända är skyldiga
- (b) högst 1% av alla oskyldiga blir dömda
- (c) högst 1% av alla skyldiga blir frikända
- (d) högst 1% av alla dömda är oskyldiga

6. Med en viss mätutrustning registreras den radioaktiva bakgrundsstrålningen på en ort. Det är rimligt att anta att antalet registrerade partiklar under  $t$  minuter är  $Po(\lambda t)$  där  $\lambda = 5$  ( $\text{min}^{-1}$ ). Efter ett radioaktivt utsläpp misstänker man att strålningen har ökat. Hur länge behöver man mäta strålningen om man vill pröva

$$H_0 : \lambda = 5 \quad \text{mot} \quad H_1 : \lambda > 5$$

på nivån 0.01 med ett test som ger utslag med säkerheten 0.99 om strålningen ökat 50%. Lämplig approximation får utnyttjas.

7. Man har mätt en viss föroreningshalt i jordprover tagna dels i närheten av en industri och dels i ett "rent område" med motsvarande jordmån. Det förutsätts att en mätning ger ett lognormalfördelat värde och därför är nedanstående medelvärden beräknade för de logaritmerade mätvärdena.

	antal obs.	medelvärde	standardavvikelse
Nära industrin	8	1.8	0.49
Rent område	9	1.1	0.46

Undersök om föroreningshalten kan anses vara högre i det förorenade området.

8. Ett företag tillverkar reläer vid fyra olika fabriker; A, B, C och D. För att undersöka om fabriker har samma kvalitet på sina reläer undersökte man 200 enheter från var och en av fabriker och räknade antalet defekta och korrekta enheter.

	A	B	C	D
Defekta	6	4	4	16
Korrekta	194	196	196	184

Det förefaller rimligt att anta att A, B och C har samma defektsannolikhet  $p_1$  medan D har en annan defektsannolikhet  $p_2$ . Konstruera ett 95 % konfidensintervall för  $p_1 - p_2$ .

9. Genotyperna  $AA$ ,  $Aa$  och  $aa$  för en viss allel förekommer med sannolikhet  $\theta^2$ ,  $2\theta(1 - \theta)$  och  $(1 - \theta)^2$  i en population. Man har på måfå valt ut ett antal individer och fått följande resultat:

Typ	$AA$	$Aa$	$aa$
Antal	34	50	19

ML-skatta andelen  $A$ -gener i populationen, dvs  $\theta$ .

10. Använd ML-skattningen av  $\theta$  från uppgift 9 och utför ett  $\chi^2$ -test för att pröva om den angivna modellen är rimlig.
11. En kaffeleverantör påstår att mängden kaffe i ett 500g-paket kan betraktas som en stokastisk variabel  $X \sim N(500, 10^2)$ . Vid en undersökning av 100 paket fick man följande resultat:

Vikt	< 495 g	495 – 504 g	> 505g
Antal paket	39	45	16

Kaffemängder i olika paket kan anses vara oberoende stokastiska variabler.

Pröva med hjälp av ett  $\chi^2$ -test på nivån 0.01 hypotesen att kaffemängderna är normalfördelade som leverantören påstår.

12. I följande tabell är  $y_i$  observerade värden på en stokastisk variabel  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ , där  $\epsilon_1, \dots, \epsilon_n$  är oberoende och  $N(0, \sigma^2)$ .

$x_i$	1	2	3	4	5	6	7	8	9	10
$y_i$	11.5	9.8	7.0	5.5	3.9	2.0	0.4	-1.4	-3.1	-5.0

- (a) Beräkna punktskattningarna av  $\beta_0$ ,  $\beta_1$ , och  $\sigma$ .
- (b) Rita in värdena tillsammans med den beräknade regressionslinjen i ett koordinatsystem.