Grundkurs i statistisk teori, del 2

Räkneövning 3 - Konfidensintervall, 10.04.2015

1. En fabrik tillverkar bultar. Längden (mm) av bultarna varierar på grund av ett slumpmässigt fel som antas vara normalfördelat kring 0 med standardavvikelsen $\sigma = 0.5$ vilket motsvarar precisionen för tillverkningsmetoden. Vi kan med andra ord betrakta längden av bultarna som en normalfördelad variabel med väntevärdet μ (= den avsedda längden) och variansen σ^2. Anta att vi tar ett stickprov på 80 bultar vars medellängd mäts till 49.94. Skapa ett 99%-konfidensintervall för den förväntade längden av bultarna.

   $X =$ " längden på en bult "

   $X \sim N(\mu, \sigma^2)$

   $\sigma = 0.5$ känd $\Rightarrow \; I_\mu = (\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = (49.80, 50.08)$

2. Låt X vara en normalfördelad stokastisk variabel med okänt väntevärde μ och okänd varians σ^2. Vi har följande stickprov av variabeln:

   42.38 41.80 43.01 42.55

   Skapa ett 95%-konfidensintervall för μ.

   $X \sim N(\mu, \sigma^2)$

   σ okänd $\Rightarrow \; I_\mu = (\bar{x} - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}) = (41.64, 43.23)$

3. Låt p beteckna sannolikheten att ett VR-tåg är mer än 10 minuter försenat vid ankomsten till sin ändstation. Man väljer slumpmässigt totalt $n = 80$ ankomster för olika tåg och undersöker hur mycket försenade de utvalda tågen är. Som resultat fås att 28 av tågen är mer än 10 minuter försenade. Skapa ett 95%-konfidensintervall för p.

   $X =$ " antal försenade tåg "

   $X \sim Bin(n, p)$ där $p =$ " snl att tåg är försenat "

   Vi använder normalapproximation $X \overset{\text{approx.}}{\sim} N(np, np(1-p))$ och skattar $\hat{p} = \frac{x}{n} \Rightarrow$

   $I_p = (\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) = (0.25, 0.45)$

4. En fysiker har gjort fem mätningar för att bestämma en fysikalisk konstant m. Mätningarna kan anses vara oberoende observationer av en normalfördelad stokastisk variabel med väntevärdet m. Anta att standardavvikelsen σ, som motsvarar mätmetodens precision, är känd. Fysikern fick ett 90%-konfidensintervall som $(7.02, 7.14)$ vilket hon tyckte var för brett. Hur många fler mätningar behövs för att få ett konfidensintervall som är hälften så brett.

   $X =$ " mätresultat "

   $X \sim N(m, \sigma^2)$

   σ känd $\Rightarrow \; I_m = (\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{5}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{5}}) \Rightarrow$

   Nuvarande intervallbredd: $2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{5}}$

Intervallbredd efter k tilläggsmätningar: $2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{5+k}}$

$$\frac{1}{2} \cdot 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{5}} = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{5+k}} \Leftrightarrow \ldots \Leftrightarrow k = 15$$

Det krävs 15 mätningar till för att halvera intervallbredden.

5. I ett miljöövervakningssystem studeras övergödning av vattendrag. I en viss å har man under en längre period gjort mätningar av fosforhalten (mg/l). Under perioden införde man ett nytt vattenreningssystem. För att undersöka vilken effekt det nya systemet har haft på fosforhalten beräknas årsmedelvärden av fosforhalten före och efter införandet av systemet.

Före	0.12	0.14	0.07	0.09	0.15	0.09	0.10	
Efter	0.09	0.03	0.07	0.09	0.07	0.08	0.11	0.07

Antag att fosforhalterna före och efter det nya systemet kan antas vara normalfördelade med den gemensamma variansen σ^2 och väntevärdena μ_F respektive μ_E. Skapa ett 95%-konfidensintervall för den genomsnittliga effekten ($= \mu_E - \mu_F$) utgående från datamaterialet.

$X_F =$ “ fosforhalt före nya systemet ”

$X_E =$ “ fosforhalt efter nya systemet ”

$X_F \sim N(\mu_F, \sigma^2)$

$X_E \sim N(\mu_E, \sigma^2)$

Gemensam okänd varians $\sigma^2 \Rightarrow$

$$I_{\mu_E - \mu_F} = (\bar{x}_E - \bar{x}_F - t_{\alpha/2}(f) \cdot s\sqrt{\frac{1}{n_E} + \frac{1}{n_F}}, \bar{x}_E - \bar{x}_F + t_{\alpha/2}(f) \cdot s\sqrt{\frac{1}{n_E} + \frac{1}{n_F}})$$

där

$$f = n_E + n_F - 2 \text{ samt } s^2 = \frac{(n_E - 1)s_E^2 + (n_F - 1)s_F^2}{n_E + n_F - 2}$$

$\Rightarrow \ I_{\mu_E - \mu_F} = (-0.061, -0.003)$