9. Intervallskattning

9.1 Tvåsidigt konfidensintervall för μ då σ känt

Ex.9.1: Datorutrusting kan vara mycket känslig för höga temperaturer. För att undersöka vilken temperatur som en viss komponent tål genomfördes en undersökning där 49 sådana komponenter utsattes för förhöjd temperatur. För var och en höjde man temperaturen tills komponenten gick sönder. Temperaturen då detta inträffade noterades. Av undersökningen framgick att genomsnittstemperaturen då en komponent gick sönder var 58.0°C. Standardavvikelsen, $\sigma = 14.0$°C, var känd från tidigare studier. Beräkna ett 99%-igt konfidensintervall för μ, den genomsnittliga temperaturen för hela populationen av komponenter.

Konfidensvallet för μ (ovan)

1) är centrerat i $\bar{\xi}$, symmetriskt m.a.p. $\bar{\xi}$

2) längden beror på
   - konfidensen man vill ha (desto större konfidens, desto längre intervall)
   - hur mycket $\bar{\xi}$ varierar (desto större varians, desto längre intervall)
   - stickprovsstorleken n (desto större n, desto kortare intervall)

3) kan användas endast om σ känd. Vanligen känner man inte till den och måste skatta den.

4) ξ är normalfördelad. Hur gör man om ξ inte är normalfördelad? Även då är det här konfidensintervallet en bra approximation om n är tillräckligt stort (på grund av centrala gränsvärdessatsen

9.2 Tvåsidigt konfidensintervall för μ då σ okänt

Om σ är känd, kan man använda statistikan

$$\frac{\bar{\xi} - \mu}{\sigma/\sqrt{n}}$$

för att hitta konfidensintervall för μ. Om σ inte är känd, måste man skatta den.

1) Hur kan man skatta σ? Genom att använda stickprovsstandardavvilkelsen

$$\hat{\sigma} = S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2}$$

2) Vad blir fördelningen av den ovanstående statistikan om man ersätter σ med S? Statistikan är inte längre normalfördelad men har en så-kallad Students T-fördelning.

Egenskaper av T-fördelningen

1) Fördelningen har en parameter γ, som är ett positivt heltal och som kallas antal frihetsgrader. Man betecknar T-fördelningen med γ frihetsgrader av T_γ.

2) Den är en kontinuerlig fördelning som är "klockformad" som normalfördelningen. Väntevärdet av T-fördelningen är 0.

3) Parametern γ påverkar variansen (formen) av fördelningen

   - när γ blir större, blir variansen mindre

   - när $\gamma \to \infty$, närmar T-fördelningen sig standardiserade normalfördelningen

Låt $\xi_1, \xi_2, ..., \xi_n$ vara ett stickprov från normalfördelningen med väntevärdet μ och variansen σ^2. Då är

$$\frac{\bar{\xi} - \mu}{S/\sqrt{n}} \sim T_{n-1}.$$

Statistikan ovan kan man använda för att hitta konfidensintervall för μ då variansen σ^2 inte är känd och antalet observationer är litet.

Ex: Svaveldioxid (sulfur dioxide) och kväveoxid (nitrogen oxide) får man när man använder fossilbränsle. Dessa föreningar kan resa långa avstånd och de kan förvandlas till syror innan de kommer ner som syrt regn. Nu har man 24 mätningar av SO_2-konsentrationen (enhet $\mu g/m^3$) från en skog i Bovaria, Tyskland, som har blivit skadad av syrt regn. Man vill hitta 95% konfidensintervall för medel-SO_2-konsentrati i denna skog.

Data:

52.7	43.9	41.7	71.5	47.6	55.1
62.2	56.5	33.4	61.8	54.3	50.0
45.3	63.4	53.9	65.5	66.6	70.0
52.4	38.6	46.1	44.4	60.7	56.4

Anta att om skogen inte är skadad är medelkonsentrationen $20 \mu g/m^3$). Har medelkonsentrationen ökat?

9.6 Jämförelse mellan två väntevärdena

9.6.1 Parvisa observationer

Ex: Man kan extrahera information om vädret från radarmätningar (radar returns) genom att använda en viss algoritm. Man vill studera skillnaden mellan vindhastigheten mätt på marken och mätt via Seasat satelliten genom att hitta ett 95% konfidensintervall för skillnaden. Vindhastigheten är mätt vid 12 tillfällen genom att använda de två metoderna.

Data:

Tidpunkt	Mark (ξ_M)	Satellit (ξ_S)
1	4.46	4.08
2	3.99	3.94
3	3.73	5.00
4	3.29	5.20
5	4.82	3.92
6	6.71	6.21
7	4.61	5.95
8	3.87	3.07
9	3.17	4.76
10	4.42	3.25
11	3.76	4.89
12	3.30	4.80

9.6.2 Två oberoende stickprov (varianserna lika)

Fördelningen för $\bar{\xi}_1 - \bar{\xi}_2$?

Låt $\bar{\xi}_1$ och $\bar{\xi}_2$ vara stickprovsmedelvärden av två oberoende stickprov av resp. storlekar n_1 och n_2 från $N(\mu_1, \sigma_1)$ och $N(\mu_2, \sigma_2)$. Då är

$$\bar{\xi}_1 \sim N(\mu_1, \sigma_1) \text{ och } \bar{\xi}_2 \sim N(\mu_2, \sigma_2)$$

och

$$\bar{\xi}_1 - \bar{\xi}_2 \sim N(\mu_1 - \mu_2, \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})$$

Nu antar man att varianserna av de två populationerna är lika, dvs att $\sigma_1^2 = \sigma_2^2 = \sigma^2$. $\bar{\xi}_1 - \bar{\xi}_2$ är en väntevärdesriktig skattning för $\mu_1 - \mu_2$ och

$$\bar{\xi}_1 - \bar{\xi}_2 \sim N(\mu_1 - \mu_2, \sqrt{\sigma^2(1/n_1 + 1/n_2)}.$$

Då är

$$\frac{\bar{\xi}_1 - \bar{\xi}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1).$$

Vanligen känner man inte σ^2 (eller σ) och man måste skatta den. Man kan skatta

$$\hat{\sigma}_1^2 = S_1^2 \text{ och } \hat{\sigma}_2^2 = S_2^2.$$

Nu är $\sigma_1^2 = \sigma_2^2 = \sigma^2$ och man har två skattningar för σ^2, nämligen S_1^2 och S_2^2.

Man vill ha bara en skattning men använda både två. Därför använder man

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

som en skattning för σ^2 (ett viktat medelvärde av S_1^2 och S_2^2).

Obs! Viktena är $n_1 - 1$ och $n_2 - 1$ (inte n_1 och n_2) därför att då blir

$$\frac{\bar{\xi}_1 - \bar{\xi}_2 - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$T_{n_1+n_2-2}$-fördelad.

Ex: Man jämför hur mycket diversarbetare var utsatta för radioaktivitet i 1973 ($n_1 = 16$, $\bar{x}_1 = 0.94$rem och $S_1^2 = 0.040$) och i 1979 ($n_2 = 16$, $\bar{x}_2 = 0.62$rem och $S_2^2 = 0.028$). Man har två oberoende stickprov, ett från 1973 och ett från 1979. Man antar att varianserna av de två populationerna är lika stora och vill hitta 95% konfidensintervall för $\mu_1 - \mu_2$.

Obs! Om varianserna inte kan antas vara lika, skattar man $Var(\bar{\xi}_1 - \bar{\xi}_2)$ med

$$\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$$

och använder teststatitikan

$$\frac{\bar{\xi}_1 - \bar{\xi}_2 - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}},$$

som också är T-fördelad, men antal frihetsgrader är inte samma som tidigare.