Aalto-universitetet

Gripenberg, Metsalo

MS-A0509 Grundkurs i sannolikhetskalkyl och statistik
Mellanförhör 2, 17.2.2014

Skriv ditt namn, nummer och övriga uppgifter på varje papper!
Du får använda en räknare och Ilkka Mellins tabeller och "formelsamling".

1. Ett företag tillverkar plaströr som skall ha längden 3 m. Vid ett tillfälle mättes längden på 10 rör och resultaten var x_j, $j = 1, \ldots, 10$ med $\overline{x} = \frac{1}{n}\sum_{j=1}^{10} x_j = 3.035$. Dessutom räknade man ut ett konfidensintervall med konfidensgraden 95% under antagandet att längderna är oberoende och normalfördelade med samma fördelning och resultatet blev $[2.958, 3.112]$

(a) Vad var stickprovsvariansen?
(b) Vad innebär dethär konfidensintervallet?

Lösning: (a) Eftersom $T = \frac{\overline{X}-\mathrm{E}(X)}{\sqrt{\frac{S^2}{n}}}$ har fördelningen $\mathrm{t}(10-1)$ så är $\Pr(-t_{0.025} \leq T \leq t_{0.025}) = 0.95$ där $t_{0.025} = F^{-1}_{\mathrm{t}(9)}(0.975) = 2.2622$. Detta leder till att det konfidensintervall vi räknar ut blir

$$\left[\overline{x} - 2.622 \cdot \sqrt{\frac{s^2}{10}}, \overline{x} + 2.622 \cdot \sqrt{\frac{s^2}{10}}\right]$$

vilket innebär att

$$2 \cdot 2.622 \cdot \sqrt{\frac{s^2}{10}} = 3.112 - 2.958 = 0.154$$

och detta ger

$$s^2 = 10 \cdot \left(\frac{0.154}{2 \cdot 2.622}\right)^2 = 0.0116.$$

(b) Om vi tar många stickprov så kommer väntevärdet av längden på ett rör att i ungefar 95% av fallen ligga i det konfidensintervall vi räknar ut. Alternativt, sannolikheten för att vi får ett konfidensintervall i vilket väntevärdet ligger är 0.95 men väntevärdet ligger antingen i intervallet $[2.958, 3.112]$ eller också inte och och det är inte förnuftigt att tala om sannolikheter om man inte förklarar att man med sannolikhet menar något slag av övertygelse.

---

2. En arbetsgrupp som planerat en omläggning av busslinjerna i en stad hävdar att minst 60% av stadens invånare kommer att understöda förändringarna. När 70 personer blev tillfrågade svarade 33 att de understödde förändringarna. Vad kan man säga om arbetsgruppens påstående?

Lösning: Låt p vara sannolikheten att en slumpmässigt vald person stöder omläggningen. Som nollhypotes väljer vi $\mathrm{H}_0 : p \geq 0.6$. Om nu X_j, $i = 1, \ldots, n$ är ett stickprov av fördelningen $\mathrm{Bernoulli}(p)$ så är $\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim_a \mathrm{N}(0,1)$. Testvariabeln är

$$Z = \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}},$$

och dess värde, då $p = 0.6$ är $z = -2.1958$ Eftersom nollhypotesen är $p \geq 0.6$ har den ett ensidigt alternativ och p-värdet blir $F^{-1}_{\mathrm{N}(0,1)}(-2.1958) = 0.014$ vilket betyder att nollhypotesen kan förkastas på signifikansnivån 5% men inte 1%. Ett annat sätt att komma till samma resultat är att räkna ut de kritiska värdena $-z_{0.05} = -1.6449$ och $-z_{0.01} = -2.3263$ och konstatera att $z < -1.6449$ men $z > -2.3263$.

Om man räknar med binomialfördelningen blir p-värdet 0.019909 vilket ger samma slutsats.

---

3. Slumpvariabeln X antas ha fördelningen $\mathrm{t}(10)$ och för att undersöka detta togs ett stickprov med storleken 100 av denna slumpvariabel och värdena ordnades i 5 grupper beroende på i vilket av intervallen $(-\infty, -0.88)$, $[-0.88, -0.26)$, $(-0.26, 0.26)$, $[0.26, 0.88)$ och $[0.88, \infty)$ de låg. Antalen värden i de olika intervallen blev följande:

$(-\infty, -0.88)$	$[-0.88, -0.26)$	$(-0.26, 0.26)$	$[0.26, 0.88)$	$[0.88, \infty)$
15	27	28	14	16

Om man nu räknar väntevärdet av hur många av dessa 100 värden som skulle ligga i dessa intervall om fördelning verkligen var $\mathrm{t}(10)$ så får man följande värden (avrundade till heltal):

$(-\infty, -0.88)$	$[-0.88, -0.26)$	$(-0.26, 0.26)$	$[0.26, 0.88)$	$[0.88, \infty)$
20	20	20	20	20

Testa nollhypotesen att fördelningen verkligen är $\mathrm{t}(10)$ på signifikansnivån 5%.

Lösning: Vi använder χ^2-testet och värdet av testvariabeln är

$$c = \sum_{j=1}^{5} \frac{(o_j - e_j)^2}{e_j} = \frac{(15-20)^2}{20} + \frac{(27-20)^2}{20} + \frac{(28-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(16-20)^2}{20} = 9.5$$

Eftersom det finns 5 grupper blir antalet frihetsgrader 4 (det värde som inte kommer i en grupp kommer i en annan) och p-värdet blir $1 - F_{\chi^2(4)}(10.5) \approx 0.00497 < 0.05$ så nollhypotesen förkastas. Ett annat sätt att komma till samma slutsats är att konstatera att om $C \sim \chi^2(4)$ så gäller $\Pr(C \geq 9.49) = 0.05$ och eftersom $9.5 > 9.49$ så förkastas nollhypotesen.

---

4. Vi har ett observerat stickprov (x_j, y_j), $j = 1, \ldots, 17$ och har räknat ut medelvärdena $\overline{x} = 2$ och $\overline{y} = 4$, stickprovsvarianserna $s_x^2 = 5$ och $s_y^2 = 6$, och stickprovskovariansen $s_{xy} = 3$. Räkna ut estimat b_0 och b_1 för regressionskoefficienterna β_0 och β_1 i regressionsekvationen $Y = \beta_0 + \beta_1 x + \varepsilon$ och testa nollhypotesen $\beta_1 = 0.25$ på signifikansnivån 1%.

Lösning: Regressionskoefficienterna blir

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{3}{5} = 0.6 \quad \text{och} \quad b_0 = \overline{y} - b_1\overline{x} = 4 - 0.6 \cdot 2 = 2.8.$$

Korrelationskoefficienten $r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} = \frac{3}{\sqrt{30}}$ så att restvariansen s^2 blir

$$s^2 = \frac{(n-1)s_y^2(1-r_{xy}^2)}{n-2} = \frac{16 \cdot 6}{15} \cdot \left(1 - \frac{9}{30}\right) = 4.48$$

Värdet av testvariabeln blir

$$t = \frac{b_1 - \beta_1}{\sqrt{\frac{s^2}{(n-1)*s_x^2}}} = \frac{0.6 - 0.25}{\sqrt{\frac{4.48}{16 \cdot 5}}} = 1.4790.$$

Eftersom testvariabeln är $\mathrm{t}(17-2)$ fördelad blir p-värdet $2(1-F_{\mathrm{t}(15)}(1.4790)) = 0.16 >> 0.01$ (nollhypotesen har ett dubbelsidigt alternativ!) och nollhypotesen kan inte förkastas på signifikansnivån 1%. Ett annat sätt att komma till samma resultat är att konstatera att $t_{0.005}(15) = 2.9467$ (0.005 för alternativet är dubbelsidigt) och eftersom värdet av testvariabeln ligger i intervallet $[-2.9467, 2.9467]$ förkastas nollhypotesen inte.
