Aalto-universitetet

MS-A0509 Grundkurs i sannolikhetskalkyl och statistik Gripenberg
Övning 5
Vecka 6, 2–6.2.2015

I1. Under åren 2007 och 2008 uppmättes vid samma tidpunkter under året följande vattentemperaturer i Ule träsk :

2007	5.3	8.5	11.8	17.9	15.2	16.1	15.8	17.2	18.9	22.1	18.2	13.9	11.2
2008	5.5	4.1	13.1	11.1	16.1	15.8	13.0	18.1	17.4	14.4	15.6	13.1	11.0

Finns det någon signifikant skillnad mellan temperaturerna dessa år. Använd signifikansnivån 0.05. Du kan inte anta att temperaturerna under ett år är observationer av samma slumpvariabel, men däremot kan det vara en god idé att räkna skillnaderna mellan mätvärdena de olika åren vid samma tidpunkt.

Lösning: Om vi räknar skillnaden mellan temperaturerna 2008 och 2007 får vi följande resultat:

0.2	-4.4	1.3	-6.8	0.9	-0.3	-2.8	0.9	-1.5	-7.7	-2.6	-0.8	-0.2

Som nollhypotes tar vi att väntevärdet av skillnaderna i temperaturerna är 0 och att dessa skillnader är normalfördelade. Medelvärdet och stickprovsvariansen är $\overline{x} = -1.8308$ och $s^2 = 8.5240$. Som testvariabel använder vi

$$W = \frac{\overline{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}$$

och eftersom nollhypotesen är $\mu_0 = 0$ så får testvariabeln värdet

$$\frac{-1.8308 - 0}{\sqrt{\frac{8.524}{13}}} = -2.2609.$$

Eftersom alternativet till nollhypotesen är tvåsidigt blir p-värdet

$$p = 2F_{\mathrm{t}(13-1)}(-2.2609) = 0.043$$

och eftersom $0.043 < 0.05$ kan vi förkasta nollhypotesen på signifikansnivån 0.05.

I2. Antag att vi går tillväga på följande (felaktiga) sätt. Vi tar ett observerat stickprov $x_1, x_2, \ldots, x_{15}$ av en $\mathrm{N}(\mu, \sigma^2)$ fördelad slumpvariabel och räknar ut medelvärdet $\overline{x}$ och stickprovsvariansen s^2. Ifall $\overline{x} > \mu_0$ så tar vi som nollhypotes $\mathrm{H}_0 : \mu \leq \mu_0$ och i andra fall tar vi som nollhypotes $\mathrm{H}_0 : \mu \geq \mu_0$, (och felet består i att vi låter nollhypotesen bero på det observerade stickprovet). Sedan testar vi nollhypotesen på normalt sätt på signifikansnivån 0.01. Vad är sannolikheten att vi förkastar nollhypotesen om $\mu = \mu_0$? Vad är denna sannolikhet om stickprovets storlek är n och signifikansnivån α istället för 0.01?

Ledning: Antag först att $\overline{x} > \mu_0$. Bestäm det kritiska värdet för testet, dvs. ett tal A så att om testvariabeln $w = \frac{\overline{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} > A$ så förkastas nollhypotesen $\mu \leq \mu_0$ på signifikansnivån 0.01. Om sedan $\overline{x} < \mu_0$ så blir motsvarande kritiska värde på grund av symmetrin $-A$, dvs. nollyhoptesen (som nu är $\mu \geq \mu_0$) förkastas på signifikansnivån 0.01 om testvariabeln $w < -A$. Bestäm sedan sannolikheten för att testvariabeln w får ett värde som ligger i mängden $(-\infty, -A) \cup (A, \infty)$ så att nollhypotesen förkastas eftersom $w > A$ innebär att $\overline{x} > \mu_0$ och $w < -A$ innebär att $\overline{x} < \mu_0$.

Lösning: Det observerade värdet av testvariabeln är

$$w = \frac{\overline{x} - \mu_0}{\sqrt{\frac{s^2}{n}}}.$$

Ifall $w > t_{0.01}(15-1) = -F^{-1}_{\mathrm{t}(15-1)}(0.01) = 2.6245$ så är $\overline{x} > \mu_0$ och nollhypotesen är $\mathrm{H}_0 : \mu \leq \mu_0$ så att den förkastas eftersom

$$p = 1 - F_{\mathrm{t}(14)}(w) < 1 - F_{\mathrm{t}(14)}(2.6245) = 0.01.$$

På motsvarande sätt ser vi att om $t < -2.6245$ så är $\overline{x} < \mu_0$ och nollhypotesen är $\mathrm{H}_0 : \mu \geq \mu_0$ och den förkastas eftersom

$$p = F_{\mathrm{t}(14)}(w) < F_{\mathrm{t}(14)}(-2.6245) = 0.01.$$

Om däremot $|w| \leq 2.6245$ så förkastas inte nollhypotesten oberoende av vad den blir. Men

$$\Pr(|w| > 2.6245) = 2 \cdot 0.01 = 0.02,$$

dvs. sannolikheten för att nollhypotesen förkastas är 0.02 och inte 0.01 som var meningen.

Om signifikansnivån är α och stickprovets storlek är n så är enda skillnaden att talet 2.6245 byts ut mot $t_{\alpha,n-1} = -F^{-1}_{\mathrm{t}(n-1)}(\alpha)$ och nollhypotesen förkastas med sannolikheten 2α istället för med sannolikheten α.

I3. Vi kastar en tärning 180 gånger och får n_k gånger resulatet k enligt följande tabell:

Resultat	1	2	3	4	5	6
n_k	41	22	20	36	36	25

Är tärningen felfri, dvs. är alla resultat lika sannolika? Testa med χ^2-testet på signifikansnivån 0.01.

Lösning: Nollhypotesen är att alla resultat förekommer med sannolikheten $\frac{1}{6}$. Då är $np_k = \dfrac{180}{6} = 30$ för alla k och testvariabeln $C = \sum_{k=1}^{m} \dfrac{(n_k - np_k)^2}{np_k}$ får värdet

$$c = \frac{(41-30)^2}{30} + \frac{(22-30)^2}{30} + \frac{(20-30)^2}{30} + \frac{(36-30)^2}{30} + \frac{(36-30)^2}{30} + \frac{(25-30)^2}{30} = 12.733.$$

Eftersom testvariabeln är approximativt $\chi^2(6-1)$-fördelad så blir p-värdet $p = \Pr(C \geq 12.733) = 1 - F_{\chi^2(5)}(12.733) = 0.026 > 0.01$ och vi förkastar inte nollhypotesen på signifikansnivån 0.01.

I4. I samband med en kvalitetskontroll av en viss produkt skall vi testa nollhypotesen $\mathrm{H}_0 : p \leq 0.06$ på signifikansnivån 0.05 på ett stickprov med storleken 500 av en Bernoulli(p)-fördelad slumpvariabel (som får värdet 1 om kvalitetskraven på produkten inte uppfylls, annars 0) genom att använda normalapproximation. Vad är sannolikheten att nollhypotesen inte förkastas fast värdet av p i verkligheten är 0.08. Använd också för denna del normalapproximation.

Ledning: Bestäm talet A så att nollhypotesen förkastas om minst A exemplar i stickprovet inte uppfyller kvalitetskraven och bestäm sedan sannolikheten för att färre än A stycken av produkterna i stickprovet inte uppfyller kvalitetskraven.

Svar: 0.37

Lösning: Enligt (extremfallet $p = 0.06$ av) nollhypotesen har antalet bristfälliga produkter Y fördelningen $\mathrm{Bin}(400, 0.06)$ och eftersom vi använder normalapproximation förkastar vi nollahypotesen på signifikansnivån 0.05 om testvariabeln $Z = \frac{Y - \mathrm{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}$ får sitt värde i intervallet $[z_{0.05}, \infty)$ där $z_{0.05} = F^{-1}_{\mathrm{N}(0,1)}(0.95) = 1.6449$ dvs. $Z \geq 1.6449$. Eftersom $\mathrm{E}(Y) = 0.06 \cdot 500$ och $\mathrm{Var}(Y) = 0.06 \cdot 0.94 \cdot 500$ så förkastas nollhypotesen då

$$Y \geq 0.06 \cdot 500 + \sqrt{0.06 \cdot 0.94 \cdot 600} \cdot 1.6449 = 38.7,$$

så att vi får $A = 39$.

I nästa steg skall vi beräkna sannolikheten för att om $U \sim \mathrm{Bin}(500, 0.08)$ så är $U \leq 38$. Med normalapproximation får vi

$$\begin{aligned}\Pr(U \leq 38) &= \Pr\left(\frac{U - \mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \leq \frac{38 - \mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}}\right) \\ &= \Pr\left(\frac{U - \mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \leq \frac{38 - 0.08 \cdot 500}{\sqrt{0.08 \cdot 0.92 \cdot 500}}\right) = \Pr\left(\frac{U - \mathrm{E}(U)}{\sqrt{\mathrm{Var}(U)}} \leq -0.32969\right) \approx 0.37.\end{aligned}$$

I5. Antag att $X \sim \mathrm{N}(\mu, \sigma^2)$. Vad är sannolikheten att nollhypotesen $H_0 : \sigma^2 \leq 3$ förkastas på signikansnivån 0.01 om i verkligheten $\sigma^2 = 5$ (dvs. man drar rätt slutsats) och hypotesen testas med ett stickprov med storleken 60?

Ledning: Bestäm det kritiska värdet A så att nollhypotesen förkastas om $s^2 > A$ (då $n = 60$) och beräkna sedan sannolikheten för att stickprovsvariansen $S^2 > A$ om $X \sim \mathrm{N}(\mu, 5)$ och $n = 60$. Kom ihåg vad man vet om fördelningen av stickprovsvariansen for normalfördelade slumpvariabler!

Svar: 0.72

Lösning: Ifall $X_j \sim \mathrm{N}(\mu, \sigma^2)$, $j = 1, \ldots, n$ så gäller som bekant $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

Dethär innebär att om $\sigma^2 = 3$ (dvs. extremfallet av nollhypotesen) så är

$$\Pr\left(\frac{(60-1)S^2}{3} > F^{-1}_{\chi^2(59)}(1 - 0.01)\right) = 0.01,$$

så att den kritiska gränsen A, så att nollhypotesen förkastas då $s^2 > A$, blir

$$A = \frac{3 \cdot F^{-1}_{\chi^2(59)}(0.99)}{59} = \frac{3 \cdot 87.166}{59} = 4.4322.$$

I nästa stegt skall vi räkna sannolikheten för att $S^2 > 4.4322$ om variansen i verkligheten är $\sigma^2 = 5$, dvs.

$$\begin{aligned}\Pr(S^2 > 4.4322) &= \Pr\left(\frac{(60-1)S^2}{5} > \frac{(60-1)\cdot 4.4322}{5}\right)\\ &= \Pr\left(\frac{(n-1)S^2}{\sigma^2} > 52.3\right) = 1 - F_{\chi^2(59)}(52.3) = 0.72.\end{aligned}$$
