

Kortfattade lösningar till tentamen i LMA135 - Industriell matematik och statistik, fredagen den 13 mars 2009.

1. Låt X vara vikten hos ett paket.

$$P(X < 735) = P\left(Z < \frac{735 - 750}{7,3}\right) = P(Z < -2,05) = P(Z > 2,05) = 2,0\%.$$

Z är standardnormalfördelad.

2. Hypoteser

$$H_0: \mu = 0,13$$

$$H_1: \mu < 0,13$$

Eftersom stickprovet är stort kan vi använda testvariabeln

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

som är approximativt standardnormalfördelad om nollhypotesen är sann. Observerat värde på testvariabeln är

$$z_{\text{obs}} = \frac{0,11 - 0,13}{0,03/\sqrt{50}} = -4,71.$$

Detta värde finns inte i vår normalfördelningstabell, men vi kan se att p-värdet < 0,1 %. Vi har trestjärnig signifikans och nollhypotesen kan förkastas. Konkurrenten har mycket starka bevis för att fröinnhålllet är lägre än det angivna.

3. Hypoteser

H_0 : Åldersgrupp och åsikt om paketen är oberoende

H_1 : Åldersgrupp och åsikt om paketen är inte oberoende

Vi beräknar förväntade frekvenser under förutsättning att nollhypotesen är sann.

Ålder (år)/	Paket			Summa
	1	2	3	
20-40	5 ($\frac{11 \cdot 30}{60} = 5,5$)	15 (10,0)	10 (14,5)	30
40-60	6 (5,5)	5 (10,0)	19 (14,5)	30
Summa	11	20	29	60

För att kunna använda ett χ^2 -test bör helst alla förväntade frekvenser vara större 5 eller så får högst 20 % vara under 5 men inte under 1. Detta är uppfyllt här. Testvariabeln

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

är då χ^2 -fördelad med 2 (antal rader -1 multiplicerat med antal kolumner -1) frihetsgrader om nollhypotesen är sann. Observerat värde på testvariabeln är

$$\chi^2_{\text{obs}} = \frac{(5 - 5,5)^2}{5,5} + \frac{(15 - 10)^2}{10} + \frac{(10 - 14,5)^2}{14,5} + \frac{(6 - 5,5)^2}{5,5} + \frac{(5 - 10)^2}{10} + \frac{(19 - 14,5)^2}{14,5} = 7,88.$$

Om vi jämför med värden från en χ^2 -tabell med 2 frihetsgrader ser vi att $\chi^2_{\text{obs}} > 5,99$ men $< 9,21$ vilket ger att p-värdet är $< 5 \%$ men $> 1 \%$. Alltså har vi enstjärningssignifikans och nollhypotesen kan förkastas. Det är signifikant skillnad i åsikt mellan de två åldersgrupperna.

4. Typvärdet är det vanligaste värdet, d.v.s. 0, medianen är 2 eftersom 50 % av observationerna är mindre än 2, första kvartilen är 0 eftersom 25 % av observationerna är lägre än 0, tredje kvartilen är 3 eftersom 75 % av observationerna är lägre än 3. Medelvärdet är

$$\bar{x} = \frac{\sum_{i=1}^9 f_i \cdot x_i}{n} = \frac{799 \cdot 0 + 469 \cdot 1 + 657 \cdot 3 + \dots + 10 \cdot 8}{799 + 469 + \dots + 10} = \frac{4805}{2736} = 1,76$$

och standardavvikelsen är

$$\begin{aligned} s &= \sqrt{\frac{\sum_{i=1}^9 f_i \cdot x_i^2 - \frac{(\sum_{i=1}^9 f_i \cdot x_i)^2}{n}}{n - 1}} \\ &= \sqrt{\frac{799 \cdot 0^2 + 469 \cdot 1^2 + 657 \cdot 3^2 + \dots + 10 \cdot 8^2 - \frac{(4805)^2}{2736}}{2735}} \\ &= \sqrt{\frac{15369 - \frac{(4805)^2}{2736}}{2735}} = 1,59. \end{aligned}$$

5. Se sidorna 21-24 i boken.

6. Ett 95 % konfidsintervall för den sanna andelen rojalister, π , ges av

$$p \pm z \sqrt{\frac{p(1-p)}{n}} = 0,75 \pm 1,96 \sqrt{\frac{0,75 \cdot 0,25}{1000}} = 0,75 \pm 0,0268.$$

Alternativt $0,723 < \pi < 0,777(95\%)$. Observera att $np(1-p) = 187,5 > 5$. Bredden på konfidsintervallet ges av

$$2 \cdot 1,96 \sqrt{\frac{0,75 \cdot 0,25}{1000}}.$$

Om bredden ska minska med 20 % och vi antar att stickprovsandelen rojalister i det nya stickprovet också är 75 %, får vi,

$$2 \cdot 1,96 \sqrt{\frac{0,75 \cdot 0,25}{n}} < 0,8 \cdot 2 \cdot 1,96 \sqrt{\frac{0,75 \cdot 0,25}{1000}}.$$

Lös ut n ,

$$n > \frac{1000}{0,8^2} = 1562,5.$$

Alltså bör man fråga 1563 personer. Det går även bra att anta att stickprovsandelen är okänd i det nya stickprovet och använda det värsta fallet 0,5 istället.

7. Skattning av koefficienterna,

$$b = \frac{\sum_{i=1}^{15} x_i y_i - \frac{\sum_{i=1}^{15} x_i \sum_{i=1}^{15} y_i}{n}}{\sum_{i=1}^{15} x_i^2 - \frac{(\sum_{i=1}^{15} x_i)^2}{n}} = \frac{6725 - \frac{399,8 \cdot 249,5}{15}}{10850 - \frac{399,8^2}{15}} = 0,387$$

och

$$a = \bar{y} - b\bar{x} = \frac{249,5}{15} - 0,387 \cdot \frac{399,8}{15} = 6,33.$$

Tolkningen av b är att antal knarrande ökar i genomsnitt med 0,387 stycken per grad som temperaturen ökar. Tolkningen av a är att vid 0 °C är antal knarranden i genomsnitt 6,33 per sekund om vi nu kan lita på modellen vid den temperaturen. Det är inte säkert att sambandet är linjärt ner till 0. Förklaringsvärdet är

$$\begin{aligned} r^2 &= \frac{\left(\sum_{i=1}^{15} x_i y_i - \frac{\sum_{i=1}^{15} x_i \sum_{i=1}^{15} y_i}{n}\right)^2}{\left(\sum_{i=1}^{15} x_i^2 - \frac{(\sum_{i=1}^{15} x_i)^2}{n}\right) \left(\sum_{i=1}^{15} y_i^2 - \frac{(\sum_{i=1}^{15} y_i)^2}{n}\right)} = \frac{\left(6725 - \frac{399,8 \cdot 249,5}{15}\right)^2}{\left(10850 - \frac{399,8^2}{15}\right) \left(4192 - \frac{249,5^2}{15}\right)} \\ &= 0,6905. \end{aligned}$$

Alltså kan 69 % av variationen i knarrande förklaras av sambandet med temperaturen.

8. Testvariabeln är (observera $n\pi_0(1 - \pi_0) = 25 > 5$)

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}.$$

Testet ska förkasta nollhypotesen för stora värden på z , närmare bestämt om z är större än 1,64, vilket ger att testet förkastar nollhypotesen om den observerade stickprovsandelen

$$p > \pi_0 + 1,64 \sqrt{\frac{\pi_0(1 - \pi_0)}{n}} = 0,582.$$

Om nollhypotesen är sann så är stickprovsandelen, p , approximativt normalfördelad med medelvärde 0,60 och standardavvikelse $\sqrt{0,60 \cdot 0,40/100}$, se tipset. Styrkan är

$$\begin{aligned} P(p > 0,582 \text{ om } H_1 \text{ är sann med } \pi = 0,60) &= P\left(Z > \frac{0,582 - 0,60}{\sqrt{\frac{0,60 \cdot 0,40}{100}}}\right) \\ &= P(Z > -0,37) = P(Z < 0,37) = 1 - P(Z \geq 0,37) = 100\% - 35,6\% \\ &= 64,4\%. \end{aligned}$$