

Lösningsförslag 081106

Uppgift 1

A

Trädslag: kvalitativ, nominalskala (diskret)

Diameter: kvantitativ, kvotskala, kontinuerlig

Höjd: kvantitativ, kvotskala, kontinuerlig

Ålder: kvantitativ, kvotskala, kontinuerlig

B

Trädslag: Frekvenstabell, cirkeldiagram, stapeldiagram, typvärde

Diameter, höjd och ålder: Klassindeldad frekvenstabell, histogram, boxplot, dotplot, medelvärde, standardavvikelse, median

C

Trädslag vs diameter: Medelvärden för respektive trädslag, stapeldiagram som visar medelvärden

Diameter vs höjd: Spridningsdiagram, korrelationskoefficient

D

Förklaringsgraden är korrelationskoefficienten i kvadrat, dvs $0,817^2 = 0,6675$

66,7 % av den totala variationen (variationen i fastighetspriset) förklaras av variationen i virkesförrådet

E

Priset på fastigheten ökar med i genomsnitt 225.000:- per kubikmeter skog

En fastighet utan virkesförråd har sålts för i genomsnitt 53.000 kronor. Man bör vara försiktig med att tolka konstanten då det inte framgår för vilka värden som modellen tagits fram. Rimligen borde en skogsfastighet ha ett virkesförråd, men det är inte säkert.

F

$Y = 53 + 225 * 12.000 = 2700053$ (1000 kronor)

Uppgift 2

A = Händelsen att lögnedektorn indikerar lögn

B = Händelsen att en person ljuger

Givna uppgifter enligt den inledande texten:

$$\Pr(\text{icke-A} | \text{icke-B}) = 0,84$$

$$\Pr(A | B) = 0,8$$

$$\Pr(B) = 0,01$$

$$\Pr(A \text{ och } B) = \Pr(A | B) * \Pr(B) = 0,8 * 0,01 = 0,008$$

$$\Pr(\text{icke-A} | \text{icke-B}) = \Pr(\text{icke-A} | \text{icke-B}) * \Pr(\text{icke-B}) = 0,84 * 0,99 = 0,8316$$

	A	Icke-A	
B	0,008	0,002	0,01
Icke-B	0,1584	0,8316	0,99
	0,1664	0,8336	1,00

A

$$\Pr(B | A) = \Pr(A \text{ och } B) / \Pr(A) = 0,008 / 0,1664 = 0,048$$

B

$$\Pr(\text{icke-B} | A) = 1 - \Pr(B | A) = 1 - 0,048 = 0,952$$

C

$$E[X] = \sum x * \Pr(X=x) = 0 * 0,1 + 1 * 0,3 + 2 * 0,3 + 3 * 0,2 + 4 * 0,1 = 1,9$$

$$E[X^2] = \sum x^2 * \Pr(X=x) = 0 * 0 * 0,1 + 1 * 1 * 0,3 + 2^2 * 0,3 + 3^2 * 0,2 + 4^2 * 0,1 = 0,3 + 1,2 + 1,8 + 1,6 = 4,9$$

$$\text{Var}[X] = E[X^2] - E[X]^2 = 4,9 - 1,9^2 = 1,29 \rightarrow s = 1,14$$

D

$$E[Y] = 5 + 3 * E[X] = 5 + 3 * 1,9 = 10,7$$

$$s_Y = |3| * s_X = 3 * 1,14 = 3,42$$

E

X är $N(\mu=0,71; \sigma=0,05)$

$$\Pr(X < 0,7) = \Pr(Z < 0,7 - 0,71 / 0,05) = \Pr(Z < -0,2) = 42\%$$

F

Y är $N(\mu=4,26; \sigma=0,12)$

$$\Pr(Y < 4,2) = \Pr(Z < 4,2 - 4,26 / 0,12) = \Pr(Z < -0,5) = 31\%$$

Uppgift 3

A

p = Andel burkar som inte håller måttet

$$\hat{p} = \frac{n_A}{n} = \frac{2}{200} = 0,01$$

Kravet för att få approximera med normalfördelningen: $np(1-p) \geq 9$

$np(1-p) \geq 9 = 200 \cdot 0,01 \cdot 0,99 = 1,98$ (ej tillräckligt stort för normalapproximation egentligen)

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0,01 \pm 1,96 \sqrt{\frac{0,01(1-0,01)}{200}} = 0,01 \pm 0,0138$$

Med 95% säkerhet ligger andelen burkar som måste kasseras inom intervallet 0-2,4%, med reservation för att intervallet inte är tillförlitligt pga otillåten normalapproximation

B

För att studera effekten av medicinen har man studerat två olika variabler. Dels andelen som dött i kranskärslssjukdom och dels andelen som haft ej dödlig hjärtinfarkt

P = Andelen som dött i kranskärslssjukdom (Variabel 1)

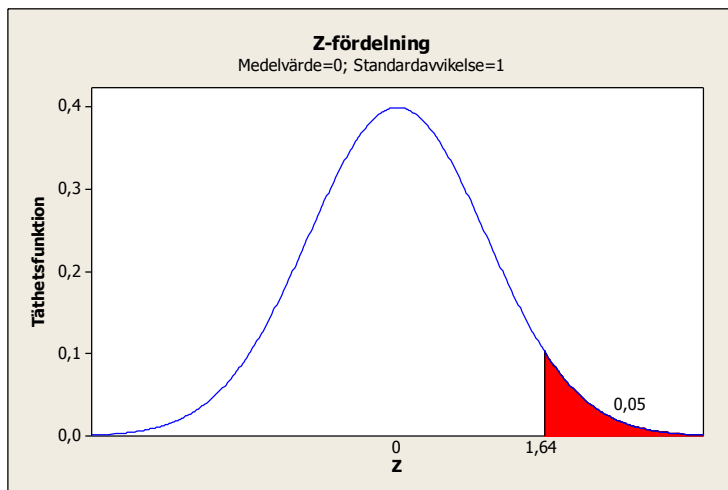
$$\hat{p}_{\text{Kontrollgrupp}} = 0,0014$$

$$\hat{p}_{\text{Placebogrupp}} = 0,0158$$

H₀: Lika stora populationsandelar = $p_{\text{Placebo}} - p_{\text{Kontroll}} = 0$

H₁: Populationsandelen är lägre för behandlingsgruppen = $p_{\text{Placebo}} - p_{\text{Kontroll}} > 0$

Vi väljer signifikansnivån α till 5% vilket ger följande kritiska område:



$$\text{Tesvariabel: } Z = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_0(1 - \hat{p}_0)\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \quad \text{där } \hat{p}_0 = \frac{n_X \hat{p}_X + n_Y \hat{p}_Y}{n_X + n_Y}$$

Vi ska förkasta nollhypotesen på 5%-nivån då $Z_{\text{Observerat}} > 1,645$

$$\hat{p}_0 = \frac{n_{\text{Placebo}} \hat{p}_{\text{Placebo}} + n_{\text{Kontroll}} \hat{p}_{\text{Kontroll}}}{n_{\text{Placebo}} + n_{\text{Kontroll}}} = \frac{0,0158 \cdot 3128 + 0,0014 \cdot 3302}{3128 + 3302} = 0,01354$$

$$Z = \frac{\hat{p}_{\text{Placebo}} - \hat{p}_{\text{Kontroll}}}{\sqrt{\hat{p}_0(1 - \hat{p}_0)\left(\frac{1}{n_{\text{Placebo}}} + \frac{1}{n_{\text{Kontroll}}}\right)}} = \frac{0,0158 - 0,0014}{\sqrt{0,01354 \cdot (1 - 0,01354)\left(\frac{1}{3128} + \frac{1}{3302}\right)}} = 1,5259$$

$Z_{\text{Observerat värde}} (=1,5259) < Z_{\text{Kritiskt värde}} (=1,645)$. Vi kan inte förkasta nollhypotesen på 5 % -nivån

Vi kan inte säga att andelen som dött i kranskärslssjukdom är mindre för de som fått medicin jämfört med de som fått placebobehandling.

P = Andelen som haft ej dödlig hjärtinfarkt (Variabel 2)

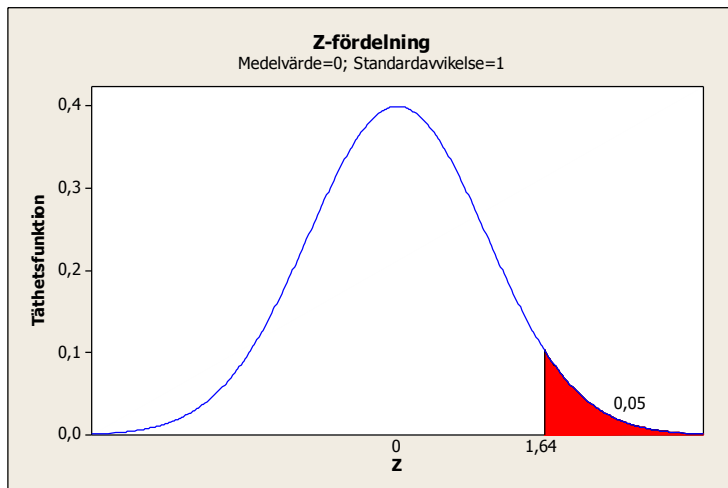
$$\hat{p}_{\text{Kontrollgrupp}} = 0,062$$

$$\hat{p}_{\text{Placebogrupp}} = 0,043$$

H₀: Lika stora populationsandelar = $p_{\text{Placebo}} - p_{\text{Kontroll}} = 0$

H₁: Populationsandelen är lägre för behandlingsgruppen = $p_{\text{Placebo}} - p_{\text{Kontroll}} > 0$

Vi väljer signifikansnivån α till 5% vilket ger följande kritiska område:



Tesvariabel:
$$Z = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_0(1 - \hat{p}_0)\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \text{ där } \hat{p}_0 = \frac{n_X\hat{p}_X + n_Y\hat{p}_Y}{n_X + n_Y}$$

Vi ska förkasta nollhypotesen på 5 % -nivån då $Z_{\text{Observerat}} > 1,645$

$$\hat{p}_0 = \frac{n_{\text{Placebo}} \hat{p}_{\text{Placebo}} + n_{\text{Kontroll}} \hat{p}_{\text{Kontroll}}}{n_{\text{Placebo}} + n_{\text{Kontroll}}} = \frac{0,062 \cdot 3128 + 0,043 \cdot 3302}{3128 + 3302} = 0,0522$$

$$Z = \frac{\hat{p}_{\text{Placebo}} - \hat{p}_{\text{Kontroll}}}{\sqrt{\hat{p}_0(1 - \hat{p}_0)\left(\frac{1}{n_{\text{Placebo}}} + \frac{1}{n_{\text{Kontroll}}}\right)}} = \frac{0,062 - 0,043}{\sqrt{0,0522 \cdot (1 - 0,0522)\left(\frac{1}{3128} + \frac{1}{3302}\right)}} = 3,4235$$

$Z_{\text{Observerat värde}} (=3,4235) > Z_{\text{Kritiskt värde}} (=1,645)$. Vi förkastar nollhypotesen på 5 % -nivån

Andelen som haft ej dödliga hjärtinfarkter är mindre för de som fått medicin jämfört med de som fått placebo-behandling.

C

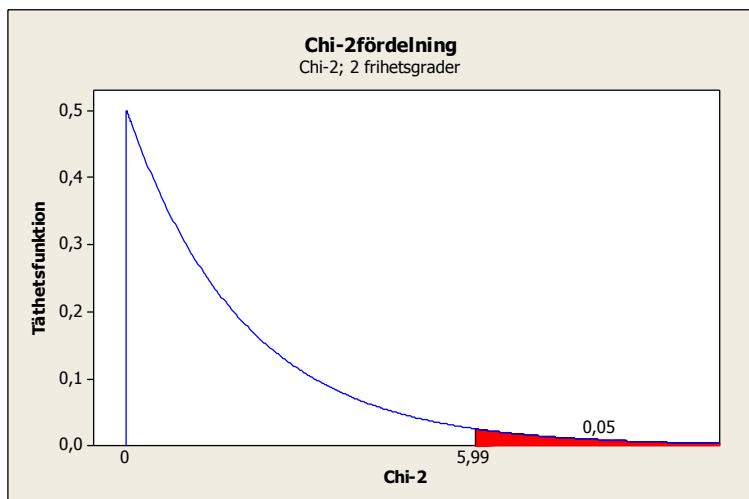
Chitest – fördelningstest

Vi ska jämföra dessa båda värderingsmän och undersöka ifall deras fördelning när det gäller klassificeringen är lika. Vi betraktar det observerade resultatet som två oberoende stickprov.

H_0 : Fördelningarna för de båda avsyningsmännen avseende klassificeringen är lika

H_1 : Fördelningarna för de båda avsyningsmännen avseende klassificeringen är **inte** lika

Vi väljer signifikansnivån till 5 %, vilket ger det kritiska värdet 5,99



	Olle	Jocke	Summa
Klass A	180	200	380
Klass B	250	220	470
Klass C	160	180	340
Summa	590	600	1190

Om nollhypotesen är sann så ska fördelningarna vara lika och de relativa frekvenserna för de olika klasserna vara:

Klass A = $280/1190$ dvs 31,93 %, Klass B = $470/1190$ dvs 39,50 %, Klass C = $340/1190$ dvs 28,57 %

	C1	C2	Total
1	180 188,40	200 191,60	380
2	250 233,03	220 236,97	470
3	160 168,57	180 171,43	340
Total	590	600	1190

Chi-Sq = 4,060; DF = 2; P-Value = 0,131

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \text{ är } \chi^2\text{-fördelad med 2 (antal kategorier-1) frihetsgrader.}$$

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(180-188,40)^2}{188,40} + \frac{(250-233,03)^2}{233,03} + \frac{(160-168,57)^2}{168,57} + \frac{(200-191,60)^2}{191,60} + \frac{(220-236,97)^2}{236,97} + \frac{(180-171,43)^2}{171,43} = 4,06$$

Vi kan inte förkasta nollhypotesen på 5 % nivå, då vi inte hamnar i det kritiska området. Vi kan inte säga att de båda värderingsmännen gör olika kvalitetsvärderingar.

Uppgift 4

A

För att beräkna erforderlig stickprovsstorlek kan följande formel användas:

$$n > \frac{z^2 \sigma^2}{L^2} \text{ där } L \text{ är felmarginalen}$$

$$n > \frac{1,96^2 8^2}{5^2} = 9,83 \Rightarrow \text{Stickprovet måste bestå av minst 10 observationer}$$

B

Man brukar dela in felen i två typer, slumpmässiga fel och systematiska fel.

Det slumpmässiga felet är kopplat till urvalet och storleken på detta fel kan vi uppskatta då vi genomför slumpmässiga urval.

Det systematiska felet kan vi inte bestämma hur stort det är. Det systematiska felet ger en snedvridning av resultatet och detta fel kan delas in i följande feltyper:

Bortfall: Inträffar då personer som ingår i undersökning väljer att inte delta.

Täckningsfel: Inträffar då målpopulation och rampopulation inte sammanfaller. Undertäckning innebär att ramen inte innehåller alla element som ingår i populationen. Övertäckning innebär att ramen innehåller element som inte ingår i populationen.

Bearbetningsfel: Kan uppstå då man vid manuell inmatning av data av misstag anger felaktiga värden.

Fel kopplade till mätinstrument: Tolkningsfel, prestigebias etc