Rettevejledning økonometri A juni 2015

Opgave 1

1. Poisson modellen kan anvendes fordi antallet af sygedage er indenfor et tidsrum, sygdage forekommer med et fast gennemsnit (intensitet) i tidsrummet, og der er uafhængighed af tiden siden sidste sygedag. Uafhængighed vil typisk være en skidt antagelse, da nogle sygdomme strækker sig over flere dage. Andre svar kan også accepteres. Lad X være antallet af sygedage om året for medarbejdere på Svanevej. $P(X > 6) = 1 - P(X \leq 6) = 0,238$.

2. Lad Y være en stokastisk variabel, der er lig 1, hvis ansat på Spurvevej og 0 ellers. $P(Y = 1) = 0,2$. Ved brug af loven om den totale sandsynlighed kan vi beregne sandsynligheden for at en tilfældigt valgt medarbejder har mere end 6 sygedage. $P(X > 6) = P(X > 6|Y = 1)P(Y = 1) + P(X > 6|Y = 0)P(Y = 0) = 0,005 \cdot 0,2 + 0,238 \cdot 0,8 = 0,191$.

3. Spørgsmålet kan omsættes til: Find $P(Y = 0|X > 6)$. Det kan gøres med Bayes regel. $P(Y = 0|X > 6) = \frac{P(X>6|Y=0)P(Y=0)}{P(X>6)} = \frac{0,190}{0,191} = 0,995$. Med 6 sygedage er der meget lille sandsynlighed for at medarbejderen er ansat på spurvevej og derfor er posterior sandsynlighed meget stor for at det er en medarbejder fra Svanevej, når vi ved medarbejderen har mere end 6 sygedage.

Opgave 2

1. Lad X være antal sygedage og Y være afstand. $P(X \geq 2, Y \leq 20) = \frac{99+53+62+34}{558} = 0,444$. De betingede fordelinger findes ved $P(X|Y) = \frac{P(X,Y)}{P(Y)}$ og findes i følgende tabel

	X (sygedage)		
	$0-1$	$2-4$	$5+$
$P(X\|0-10$ km$)$	$0,458$	$0,333$	$0,209$
$P(X\|10-20$ km$)$	$0,453$	$0,333$	$0,214$
$P(X\|20+$ km$)$	$0,451$	$0,333$	$0,216$

Der er afhængighed. Den er dog ret svag. Men jo længere der er til arbejdet desto større sandsynlighed for 5+ syge dage.

2. Den marginale fordeling er givet ved $P(X) = P(X, 0-10 \text{ km}) + P(X, 10-20 \text{ km}) + P(X, 20+ \text{ km})$ og findes i denne tabel:

X	$P(X)$
$0-1$	$0,455$
$2-4$	$0,333$
$5+$	$0,211$

Og de betinge fordelinger findes som før:

	X (sygedage)		
	$0-1$	$2-4$	$5+$
$P(X\rceil\text{Motorkøretøj})$	$0,444$	$0,333$	$0,222$
$P(X\|\text{Cykel})$	$0,5$	$0,333$	$0,167$

Dem som cykler har færre sygedage.

3. Denne opgave gælder det om at finde $P(Y,Z)$. Vi kender følgende: $P(X,Y), P(X,Z), P(X), P(Y), P(Z)$ og får oplyst betinget uafhængighed. Vi ved at $P(X,Y) = P(X,Y,Z) + P(X,Y,Z^C)$ fra loven om den totale sandsynlighed. Det første led på højre side af lighedstegnet kan omskrives ved mulitpikationsreglen $P(X,Y,Z) = P(Z)\cdot P(X|Z)\cdot P(Y|X,Z) = P(Z)\cdot P(X|Z)\cdot P(Y|Z) = P(X|Z)\cdot P(Y,Z)$. Det andet sidste lighedstegn følger af betinget uafhængighed. Vi får $P(X,Y) = P(X|Z)\cdot P(Y,Z) + P(X|Z^C)\cdot P(Y,Z^C)$. Vi ved også at $P(Y) = P(Y,Z) + P(Y,Z^C)$ fra loven om den totale sandsynlighed. Substitueres denne ind for $P(Y,Z^C)$ og løses for $P(Y,Z)$ får vi $P(Y,Z) = \frac{P(X,Y)-P(X|Z^C)P(Y)}{P(X|Z)-P(X|Z^C)}$. Den simultane fordeling af afstand og transportmiddel bliver:

	motor	cykel
0-10	$0,403$	$0,129$
10-20	$0,242$	$0,043$
20+	$0,161$	$0,021$

Opgave 3

1. Der er ingen systematisk afvigelse fra en pæn klokkeformet kurve. Så denne grafiske test understøtter at data kan beskrives med en normalfordeling.

$\frac{1}{n}\sum_{i=1}^{1500} x_i = \widehat{\mu} = \frac{731.904,2}{1500} = 487,9$
$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{1500}(x_i - \overline{x})^2 = \frac{10.151.618,6}{1499} = 6.772,3 = (82,3)^2$
$\widehat{\sigma} = 82,3 = s$

2. $P(X \leq 405) = P(\frac{X-487,9}{82,3} \leq \frac{405-487,9}{82,3}) = \Phi(-1,00729) = 0,157 = p(i)$

$H_0 :$ Data kan beskrives ved en normalfordeling $H_A :$ Data kan ikke beskrives ved en normalfordeling

$\sum_{i=1}^{5} \frac{(X(i)-1500p(i))^2}{1500p(i)} = 1,9$ som er χ^2 fordelt med frihedsgrader 2. Frihedsgrader er beregnet som antal kasser − antal estimerede parametre − 1.

p-værdi=Sss=$P(\chi^2 > 1,9) = 39\% >> 5\%$ hypotesen kan ikke afvises

3. $H_0 : \sigma = 100 \quad H_A : \sigma < 100$

$Q = \frac{(n-1)s^2}{\sigma^2} = \frac{(1500-1)*6.772,3}{10000} = 1.016,0$ som χ^2 fordelt med 1499 frihedsgrader.

$P(\chi^2 > 1016,0) = 0$ hvilket betyder at H_0 klar afvise. Konklusionen er dermed at Danmarks spredning er mindre end den europæiske.

4. $H_0 : \mu = 500 \quad H_A : \mu \neq 500$

$t = \sqrt{n}\frac{\overline{X}-\mu}{s} = \sqrt{1500}\frac{487,9-500}{82,3} = -5,7$ som er t-fordelt med 1499 frihedsgrade, hvilket I praksis er standardiseret normalfordelt. P-værdien bliver næsten nul og dermed kan H_0 afvises. Så Danmarks gennemsnit er forskellig fra det europæiske (i dette tilfælde er det under det europæiske)

5. Dette test kan udføres på forskellig måde. Antallet af drenge blandt de 1500 kan antages at være binomialfordelt. Så der skal testes om $p = 0,51$ mod alternativet at p er forskellig fra $0,51$. Nedenstående er udført et chi-i-anden test

	obs.	forv.	O-F	q-bidrag
drenge	732	765	-33	$1,42$
piger	768	735	33	$1,48$
i alt	1500	1500	0	$2,91$

Dermed er teststørrelsen på $2,91$ som er χ^2 fordelt med 1 frihedsgrad. P-værdien bliver $8,8\%$ så hypotesen opretholdes.

6. Der skal udføres et uafhængighedstest. De forventede værdier bliver

$553,4$	$178,6$	$732,0$
$580,6$	$187,4$	$768,0$
$1134,0$	$366,0$	$1500,0$

q-bidrag bliver

$0,17$	$0,52$
$0,16$	$0,49$

Og dermed bliver teststørrelsen $1,3$ som er χ^2 fordelt med 1 frihedsgrad. Ingen kan vi ikke afvise H_0 da p-værdien bliver 25%.