Sensorrettleiing utsett eksamen - ECON2130

Vår 2024

Oppgåve 1

a) Å føresette at Y_i er Poisson-fordelt inneber:

- Det forventa talet ulykker per arbeidsdag er lik λ_i.
- Kor mange ulykker som skjer innanfor eit gitt tidsrom er uavhengig av kor mange ulykker som skjer i alle andre tidsrom.
- To ulykker kan ikkje inntreffe nøyaktig samtidig.

Å føresette at Y_A og Y_B er uavhengige inneber at å vite at Y_A tek ein viss verdi, lærer oss ingenting om Y_B. Formelt:

$$P(Y_B = b|Y_A = a) = P(Y_B = b)$$

og

$$P(Y_B = b \cap Y_A = a) = P(Y_B = b) * P(Y_A = a)$$

b) Sidan Y_A og Y_B er uavhengige er $P(Y_A \leq 2|Y_B = 3) = P(Y_A \leq 2)$.

For hand:

$$P(Y_A \leq 2) = P(Y_A = 0) + P(Y_A = 1) + P(Y_A = 2)$$
$$P(Y_A \leq 2) = \frac{3^0e^{-3}}{0!} + \frac{3^1e^{-3}}{1!} + \frac{3^2e^{-3}}{2!} =\approx 0.423$$

I R: $ppois(2, 3)$

c) Sidan Y_A og Y_B er uavhengige har vi:

$$P(Y_A < 3 \cap Y_B < 3) = P(Y_A < 3) * P(Y_B < 3) = P(Y_A \leq 2) * P(Y_B \leq 2)$$

Vi finn $P(Y_B \leq 2)$ for hand som i a), eller ved å bruke $ppois(2, 4)$ i R.

$$P(Y_A < 3 \cap Y_B < 3) = 0.423 * 0.238 = 0.101$$

$$P(Y_A < 3 \cup Y_B < 3) = P(Y_A \leq 2 \cup Y_B \leq 2) = P(Y_A \leq 2) + P(Y_B \leq 2) - P(Y_A \leq 2 \cap Y_B \leq 2)$$

$$P(Y_A < 3 \cup Y_B < 3) = 0.423 + 0.238 - 0.101 = 0.56$$

d) Vi trekk 10 000 verdiar av Y_A og Y_B i R og finn kor stor del av gongene summen overstig 10:

$$mean(rpois(1e4, 3) + rpois(1e4, 4) > 10)$$

Vi finn eit simulert sannsyn lik 0.10.

I dette tilfellet er det relativt enkelt å finne teoretisk sannsyn ved å utnytte at $Y_A + Y_B$ er Poisson-fordelt med $\lambda_{A+B} = \lambda_A + \lambda_B$, men i oppgåveteksten står det eksplisitt at ein skal finne simulert sannsyn.

e) Viss Y_A er Poisson-fordelt med parameter $\lambda_A = 3$ på eit tidsrom på 8 timar er den fordelt med $\lambda'_A = 3 * 9/8$ på 9 timar. Vi finn svaret for hand som i a), eller ved å bruke R: $ppois(2, 3 * 9/8)$. Vi finn eit sannsyn lik 0.345.

Oppgåve 2

a) Å føresette at X_i er Binomisk fordelt inneber at X_i er summen av n_i uavhengige Bernoulli-forsøk:

- Kvart delforsøk har to relevante utfall
- Sannsynet p er den same i alle n_i delforsøk
- Delforsøka er uavhengige av kvarandre

b)

$$E[\hat{p}] = E\left[a * \frac{X_A}{n_A} + (1-a) * \frac{X_B}{n_b}\right]$$

$$E[\hat{p}] = \frac{a}{n_A} * E[X_A] + \frac{(1-a)}{n_B} * E[X_B]$$

$$E[\hat{p}] = \frac{a}{n_A} * n_A * p + \frac{(1-a)}{n_B} * n_B * p = p$$

$\hat{p}$ er forventningsrett.

$$Var[\hat{p}] = Var\left[a * \frac{X_A}{n_A} + (1-a) * \frac{X_B}{n_b}\right]$$

$$Var[\hat{p}] = \left(\frac{a}{n_A}\right)^2 * Var[X_A] + \left(\frac{(1-a)}{n_B}\right)^2 * Var[X_B]$$

$$Var[\hat{p}] = \left(\frac{a}{n_A}\right)^2 * n_A * p(1-p) + \left(\frac{(1-a)}{n_B}\right)^2 * n_B * p(1-p)$$

$$Var[\hat{p}] = p(1-p)\left[\frac{a^2}{n_A} + \frac{(1-a)^2}{n_B}\right]$$

c) For å minimere $Var[\hat{p}]$ minimerer vi:

$$\min_a \left[\frac{a^2}{n_A} + \frac{(1-a)^2}{n_B}\right]$$

Vi deriverer og set lik 0:

$$\frac{2a}{n_A} - \frac{2(1-a)}{n_B} = 0$$

$$a = \frac{n_A}{n_A + n_B}$$

Viss vi set inn for $n_A = 40$ og $n_B = 60$ får vi $a = 0.4$. Sidan uttrykket er konvekst veit vi at dette er minimum.

d)

$$\hat{p} = 0.4 * \frac{4}{40} + 0.6 * \frac{9}{60} = 0.13$$

$$SE[\hat{p}] = \sqrt{Var[\hat{p}]} = \sqrt{0.13(1-0.13)\left[\frac{0.4^2}{40} + \frac{0.6^2}{60}\right]} = 0.034$$

Vi finn $z = 1.96$ ved å bruke $qnorm(0.975)$ i R.

$$KI = [\hat{p} - z * SE[\hat{p}],\, \hat{p} + z * SE[\hat{p}]]$$

$$KI = [0.063,\, 0.197]$$

e) Vi set opp følgjande hypoteser:

$$H_0 : p_A = p_B,\ H_1 : p_A \neq p_B$$

Vi har estimata $p_A = 0.1$ og $p_B = 0.15$, og vi brukar teststatistikken:

$$Z = \frac{p_A - p_B}{\sqrt{\hat{p} * (1 - \hat{p}) * (\frac{1}{n_A} + \frac{1}{n_B})}}$$

Vi brukar same $\hat{p}$ som ovanfor i utrekninga av standardfeilen sidan $p_A = p_B$ under H_0. Teststatistikken er tilnærma normalfordelt og kritisk verdi er derfor $|z| = 1.96$.

$$z = \frac{0.1 - 0.15}{0.069} = -0.73$$

$|-0.73| < 1.96$ og vi forkastar derfor ikkje H_0.

f)

$$pverdi = 2 * \Phi(-|z|) = 2 * \Phi(-|-0.73|) = 0.47$$

P-verdien er ein stokastisk variabel. Den er ein funksjon av ein stokastisk variabel og har derfor eit utfallsrom og ei sannsynsfordeling over utfallsrommet.

Oppgåve 3

a) Vi brukar følgjande hypoteser:

$$H_0 : \mu = 15, \; H_1 : \mu \neq 15$$

$n = 25$ er under det vi har sagt er akseptabelt for å bruke sentralgrenseteoremet og vi kan derfor ikkje bruke at $Z = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ er tilnærma normalfordelt. Viss vi føreset at talet arbeidstimar er normalfordelt, kan vi bruke at:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

er t-fordelt med 24 fridomsgrader. Vi finn kritisk verdi ved å bruke $qt(0.995, 24)$ i R og finn at denne er lik 2.80.

$$T = \frac{19.7 - 15}{\sqrt{77.6/25}} = 2.67$$

Sidan $2.67 < 2.80$ forkastar vi ikkje H_0.

b) Vi kan nytte sentralgrenseteoremet når n blir stor, men kva som er stort nok er avhengig av den underliggande fordelinga. Vi har brukt ein tommelfingerregel at n må minimum vere 30, og frå det vi har brukt i dette faget kan vi då nytte sentralgrenseteoremet når $n = 60$. Då kan vi bruke at følgjande testestimator er tilnærma standard normalfordelt:

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

c) Hypotesene er dei same som i a), men kritisk verdi finn vi no ved $qnorm(0.995)$, og den er lik 2.58.

$$Z = \frac{19.7 - 15}{\sqrt{77.6/60}} = 4.13$$

Sidan $4.13 > 2.58$ forkastar vi no H_0.

d) Det er mange måtar å løyse denne oppgåva på og alle måtar som gjer dei same stega som i oppgåveteksten får full uttelling. Her er det vist med ein kombinasjon av '$function$' og '$replicate$'.

$$fun < -function()\{$$

$$x < -rnorm(25, 20, 9)$$

$$forkast < -t.test(x, mu = 15)\$p.value <= 0.01$$

$$return(forkast)\}$$

$$mean(replicate(1000, fun()))$$

Vi finn at vi forkastar den gale nullhypotesa omtrent halvparten av gongane. Det impliserer ein teststyrke på omtrent 0.5.