Ályktarnir fyrir flokkabreytur

Kennarar diffra 2024

Anna Helga Jónsdóttir

Helstu atriði:

1. Ályktanir um hlutfall þýðis
2. Ályktanir um hlutföll tveggja þýða
3. Ályktanir um hlutföll fleiri þýða
4. Tengslatöflur

Hvert erum við komin...

Mat á þýðishlutfalli

- Byrjum á að skoða ályktanir fyrir hlutfall þýðis.
- Við látum p lýsa þessu hlutfalli fyrir **allt þýðið** (sem við ekki þekkjum).
- Við látum $\hat{p}$ lýsa þessu hlutfalli fyrir **úrtakið okkar**.
- Við viljum nota mælingarnar okkar til að draga ályktanir um þýðishlutfallið p:
  - Með því að reikna öryggisbil fyrir p.
  - Með því að framkvæma tilgátupróf um p.

Bernoulli tilraun og tvíkostadreifing

Bernoulli tilraun

Sérhver tilraun í safni endurtekinna tilrauna flokkast sem **Bernoulli tilraun** ef eftirfarandi gildir:

1. Hver tilraun hefur aðeins tvær mögulegar útkomur.
2. Líkurnar á jákvæðri útkomu eru þær sömu í hverri tilraun fyrir sig.
3. Útkomurnar eru óháðar.

Fjöldi jákvæðra tilrauna úr n Bernoulli tilraunum fylgir **tvíkostadreifingu** með stikana n og p, skrifað $X \sim B(n, p)$, þar sem p eru líkurnar á jákvæðri útkomu.

Mat á þýðishlutfalli

Við metum þýðishlutfallið p, með úrtakshlutfallinu, þ.e.

$$\widehat{p} = \frac{x}{n}$$

þar sem x er fjöldi þeirra mælinga sem hljóta viðkomandi útkomu og n er stærð úrtaksins.

Normalnálgun

- Þegar ákveðnum skilyrðum er uppfyllt, líkist tvíkostadreifingin normaldreifingunni.
- Þá er hægt að nota aðferðir sem byggjast á eiginleikum normaldreifingarinnar til að draga ályktanir um slembistærðir sem í raun fylgja tvíkostadreifingu.
- Það köllum við að beita **normalnálgun**.

Hvenær má nota normalnálgun?

Séu $n\hat{p}$ og $n(1-\hat{p})$ stærri en 15 má nota normalnálgun til að draga ályktanir um hlutfall tvíkostadreifingar.

Öryggisbil

Öryggisbil fyrir hlutfall þýðis

Séu skilyrðin fyrir að nota normalnálgunina uppfyllt má reikna neðra öryggismark fyrir p með:

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

og efra öryggismarkið með

$$\widehat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

þar sem $\widehat{p} = \frac{x}{n}$ og $z_{1-\alpha/2}$ fæst með að fletta upp í töflu stöðluðu normaldreifingarinnar.

Núlltilgátan

- Tilgátuprófið í þessum hluta prófar núlltilgátuna hvort hlutfall þýðisins, p, sé jafnt einhverju ákveðnu gildi sem við köllum p_0.
- Núlltilgátuna ritum við $H_0 : p = p_0$.

Tilgátupróf fyrir hlutfall þýðis

Tilgátupróf fyrir hlutfall þýðis

Séu skilyrðin fyrir að nota normalnálgunina uppfyllt má nota eftirfarandi tilgátupróf:

Núlltilgátan er:

$$\mathrm{H}_0 : p = p_0$$

Prófstærðin er:

$$Z = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}}$$

þar sem X er fjöldi heppnaðra tilrauna og n er stærð úrtaksins.

Ef núlltilgátan er sönn fylgir prófstærðin stöðluðu normaldreifingunni, eða $Z \sim N(0, 1)$.

Gagntilgátur fyrir hlutfall þýðis

Gagntilgátur fyrir hlutfall þýðis

Gagntilgátan getur verið einhliða eða tvíhliða og má sjá þær ásamt höfnunarsvæðunum hér að neðan.

Gagntilgáta	Hafna H_0 ef:
$\text{H}_1 : p < p_0$	$Z < -z_{1-\alpha}$
$\text{H}_1 : p > p_0$	$Z > z_{1-\alpha}$
$\text{H}_1 : p \neq p_0$	$Z < -z_{1-\alpha/2}$ eða $Z > z_{1-\alpha/2}$

Hvert erum við komin...

Möt á hlutföllum í tveimur þýðum

- Við köllum hlutföllin í þýðunum tveimur p_1 og p_2 en hlutföll í tilsvarandi úrtökum $\hat{p}_1$ og $\hat{p}_2$.
- Við prófum $H_0 : p_1 = p_2$, en reiknum öryggisbil fyrir $p_1 - p_2$.
- Í þessu tilviki er ekki hægt að framkvæma próf sem byggir beint á tvíkostadreifingunni, heldur verðum við að nota normalnálgun.

Skilyrði fyrir normalnálgun

Beita má normalnálgun ef gera má ráð fyrir að $n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$ og $n_2(1-\hat{p}_2)$ séu öll stærri en 15

Öryggisbil fyrir hlutföll tveggja þýða

Öryggisbil fyrir hlutföll tveggja þýða

Séu skilyrðin fyrir að nota normalnálgunina uppfyllt má reikna neðra öryggismark fyrir **muninn á** p_1 og p_2 með:

$$\widehat{p_1} - \widehat{p_2} - z_{1-\alpha/2} \cdot \sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$

og efra öryggismarkið með

$$\widehat{p_1} - \widehat{p_2} + z_{1-\alpha/2} \cdot \sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$

þar sem $\widehat{p_1} = \frac{x_1}{n_1}$, $\widehat{p_2} = \frac{x_2}{n_2}$ og $z_{1-\alpha/2}$ fæst með að fletta upp í töflu stöðluðu normaldreifingarinnar.

Tilgátupróf fyrir hlutfall tveggja þýða

Séu skilyrðin fyrir að nota normalnálgunina uppfyllt má nota eftirfarandi tilgátupróf:

Núlltilgátan er:

$$\mathrm{H}_0 : p_1 = p_2$$

Prófstærðin er:

$$Z = \frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{\widehat{P}(1-\widehat{P})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad \text{þar sem} \quad \widehat{P} = \frac{X_1 + X_2}{n_1 + n_2}$$

Ef núlltilgátan er sönn fylgir prófstærðin stöðluðu normaldreifingunni, eða $Z \sim N(0,1)$.

Hvert erum við komin...

Töflur fyrir kí-kvaðrat próf

Töflur fyrir kí-kvaðrat próf

Þegar framkvæma á kí-kvaðrat próf er gott að búa til þrjár töflur:

- Tafla mældrar tíðni: Inniheldur tíðni sem við fáum úr rannsókninni, mæld tíðni táknuð með o.
- Tafla væntanlerar tíðni: Inniheldur væntanlega tíðni, táknuð með e. Gildin fást með því að margfalda samtalstölurnar úr Töflu mældrar tíðni úr þeim dálki og þeirri línu sem við erum stödd í og deila með heildarfjölda. Allar tölur í þessari töflu verða að vera hærri en 5 annars er ekki hægt að nota prófið.
- Tafla prófstærðar: Inniheldur framlag til prófstærðar reiknað með $\frac{(o-e)^2}{e}$. Að lokum eru allar tölurnar í töflunni lagðar saman til að fá gildið á prófstærðinni (sjá næstu glæru).

Kí-kvaðrat próf fyrir hlutföll

Kí-kvaðrat próf fyrir hlutföll

Tilgáturnar eru:

$$H_0 : p_1 = p_2 = ... = p_d$$

$$H_1 : \text{hlutföllin eru ekki öll jöfn}$$

Prófstærðin er:

$$\chi^2 = \sum_{i=1}^{l} \sum_{j=1}^{d} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

þar sem l er fjöldi lína, d er fjöldi dálka, o er mæld tíðni og e er væntanleg tíðni.

Sé núlltilgátan sönn fylgir prófstærðin χ^2-dreifingu með (l - 1) $\cdot$ (d - 1) fjölda frígráða. Hafna skal núlltilgátunni sé $\chi^2 > \chi^2_{1-\alpha,((l-1)\cdot(d-1))}$.

Hvert erum við komin...

Tengslatöflur

- Við vorum að sjá hvernig við getum borið saman hlutföll í mismunandi þýðum.
- Við viljum enn oftar bera saman tvær flokkabreytur þar sem gögnum er aflað úr sama þýðinu.
- Til þess eru notaðar svokallaðar tengslatöflur og prófin ganga út á að svara spurningunni hvort breyturnar tvær séu óháðar.
- Prófstærðin sem notast er við er sú sama og áður og eru allir útreikningar því eins.
- Tilgáturnar eru þó settar fram á annan máta.

Tengslatöflur

Tengslatöflur

- Gerum ráð fyrir að önnur breytan hafi l flokka, en hin d flokka.
- Tilgáturnar eru

$$
\begin{aligned}
H_0 &: \text{ Það er ekki samband á milli breytanna tveggja} \\
H_1 &: \text{ Það er samband á milli breytanna tveggja}
\end{aligned}
$$

- Prófstærðin er summa allra gildanna í töflu prófstærðar.
- Sé núlltilgátan sönn fylgir prófstærðin χ^2-dreifingu með (l - 1)·(d - 1) frígráður.
- Hafna skal núlltilgátunni sé $\chi^2 > \chi^2_{1-\alpha,((l-1)\cdot(d-1))}$.

Hlutföll eða tengslatafla?

Til að kanna hvort munur sé á milli sveitafélaga á tíðni fólks sem býr hjá foreldrum sínum eftir þrítugt voru 1500 Hafnfirðingar, 600 Garðbæingar og 550 Seltjarnarnesbúar spurðir hvort þeir búi í heimahúsum eða ekki. Niðurstöðurnar má sjá hér að neðan:

	Hafnarfjörður	Garðabær	Seltjarnarnes
Býr í foreldrahúsum	49	15	13
Býr ekki í foreldrahúsum	1451	585	537

Athugið að allar tölur í þessu dæmi eru hreinn uppspuni... :)

Kannið með viðeigandi tilgátuprófi hvort munur sé á hlutfalli fólks sem býr hjá foreldrum sínum eftir þrítugt eftir sveitarfélögum.

Hlutföll eða tengslatafla?

Til að kanna hvort munur sé á milli sveitafélaga á tíðni fólks sem býr hjá foreldrum sínum eftir þrítugt voru 2650 Hafnfirðingar, Garðbæingar og Seltjarnarnesbúar spurðir í hvaða sveitafélagi þeir búa og hvort þeir búi í heimahúsum eða ekki. Niðurstöðurnar má sjá hér að neðan:

	Hafnarfjörður	Garðabær	Seltjarnarnes
Býr í foreldrahúsum	49	15	13
Býr ekki í foreldrahúsum	1451	585	537

Athugið að allar tölur í þessu dæmi eru hreinn uppspuni... :)

Kannið með viðeigandi tilgátuprófi hvort samband sé á milli breytanna.