

Oplossing Examen Statistiek

14 juni 2024

Vincent Van Schependom

Vraag 1

- Leg gedetailleerd uit wat de mediaantest is en geef de hypothesen die we aan de hand hiervan kunnen testen.
- Leg uit wat de determinatiecoëfficiënt is en illustreer grafisch.
- Veronderstel dat $X \sim N(\mu, \sigma^2)$. Geef twee verschillende schatters voor de variantie σ^2 en toon aan dat de onvertekende schatter niet de minimale variantie heeft.

Vraag 2

Laad de dataset met het commando `load(N:\\SRebry\\people.RData)`. We maken gebruik van onderstaande variabelen:

<code>leis_time</code>	Tijd besteed door de respondent aan ontspanning in uur per week
<code>hh_nchild</code>	Aantal leden van het gezin jonger dan 18 jaar
<code>ind_income</code>	Persoonlijk netto inkomen van de respondent in euro per maand
<code>ind_gender</code>	Geslacht van de respondent

- Is de proportie van werkende vrouwen verschillend naar gelang er in het gezin kinderen jonger dan 18 jaar zijn?
- Gegeven het lineair model `lm(log10(leis_time)~log10(ind_income)+children)`. Hierbij is `children` een binaire dummy-variabele die aangeeft of er kinderen zijn in het gezin (`TRUE`) of niet (`FALSE`). Je hoeft geen aanpassingen te maken aan het model.
 - Noteer de vergelijking(en) van het gegeven model.
 - Bespreek de betekenis van de coëfficiënten in dit model (ongeacht hun significantie).
 - Bespreek de inferentie van de individuele coëfficiënten.

Vraag 3

Gustav heeft 14 winkels en verkoopt onder andere kippenboutjes. Hij wil weten of dit effectief rendeert. Dat is het geval is het aantal verkochte kippenboutjes minstens 1000 is. Als hij komende week minder dan 950 kippenboutjes verkoopt, zal hij ze schrappen uit zijn catalogus. Uit voortdurende analyses van de verkoop weet Gustav dat de standaarddeviatie van de verkochte kippenboutjes 100 bedraagt.

- Formuleer gepaste hypothesen, teststatistiek en criteria. Welke voorwaarden zijn er? Ga ervan uit dat aan deze voorwaarden voldaan is.

Oplossing

We doen een test voor één gemiddelde met gekende variantie. De nul- en alternatieve hypothese zijn:

$$\begin{cases} H_0 : \mu \geq \mu_0 = 1000 \Rightarrow \text{neem } \mu = 1000 \\ H_1 : \mu < \mu_0 = 1000 \end{cases}$$

Stel

- X = het aantal verkochte kippenboutjes voor 1 winkel in 1 week tijd
- \bar{X} = het gemiddeld aantal verkochte kippenboutjes voor 14 winkels in 1 week tijd

Nu geldt wegens de Centrale Limietstelling dat $\bar{X} \sim N\left(\mu, \frac{100^2}{14}\right)$, onder voorwaarde dat X zelf niet te scheef (en liefst ook normaal) verdeeld is.

Onder H_0 kennen we bijgevolg de verdeling van de teststatistiek Z :

$$Z = \frac{\bar{X} - 1000}{\frac{100}{\sqrt{14}}} \sim N(0, 1)$$

- Wat is de kans dat Gustav kippenboutjes schrapt van zijn catalogus, als de verkoop weldegelijk rendeert?

Oplossing

$$\begin{aligned} \alpha &= P(\text{type I fout}) \\ &= P(H_0 \text{ verwerpen} \mid H_0 \text{ waar}) \\ &= P\left(\bar{X} < 950 \mid \frac{\bar{X} - 1000}{100/\sqrt{14}} \sim N(0, 1)\right) \\ &= P\left(Z < \frac{950 - 1000}{100/\sqrt{14}} \mid Z \sim N(0, 1)\right) \\ &= \text{pnorm}(950, 1000, 100/\text{sqrt}(14)) \\ &= 0.03 \end{aligned}$$

- Wat is het onderscheidingsvermogen van de test? Hoe groot moet het verschil zijn tussen de werkelijke gemiddelde verkoop en de beoogde verkoop om met 95% zekerheid kippenboutjes te schrappen uit de catalogus?

Oplossing

$$\begin{aligned}
 1 - \beta|_{\mu_1} &= 1 - P(\text{type II fout}) \\
 &= 1 - P(H_0 \text{ niet verwerpen} \mid H_0 \text{ niet waar}) \\
 &= 1 - P(\bar{X} > 950 \mid \mu = \mu_1 \neq \mu_0) \\
 &= 1 - \left[1 - P\left(\frac{\bar{X} - 1000}{100/\sqrt{14}} < \frac{950 - 1000}{100/\sqrt{14}} \mid \frac{\bar{X} - \mu_1}{100/\sqrt{14}} \sim N(0, 1)\right) \right] \\
 &= P\left(\frac{\bar{X} - 1000}{100/\sqrt{14}} + \frac{1000 - \mu_1}{100/\sqrt{14}} < \frac{950 - 1000}{100/\sqrt{14}} + \frac{1000 - \mu_1}{100/\sqrt{14}} \mid \frac{\bar{X} - \mu_1}{100/\sqrt{14}} \sim N(0, 1)\right) \\
 &= P\left(\frac{\bar{X} - \mu_1}{100/\sqrt{14}} < \frac{950 - \mu_1}{100/\sqrt{14}} \mid \frac{\bar{X} - \mu_1}{100/\sqrt{14}} \sim N(0, 1)\right) \\
 &= \Phi\left(\frac{950 - \mu_1}{100/\sqrt{14}}\right)
 \end{aligned}$$

We zoeken μ_1 zodat het onderscheidingsvermogen 95% is:

$$\begin{aligned}
 1 - \beta|_{\mu_1} &= 0.95 \\
 \iff \Phi\left(\frac{950 - \mu_1}{100/\sqrt{14}}\right) &= 0.95 \\
 \iff \frac{950 - \mu_1}{100/\sqrt{14}} &= \Phi^{-1}(0.95) \\
 \iff 950 - \mu_1 &= \left(\frac{100}{\sqrt{14}}\right) \Phi^{-1}(0.95) \\
 \iff \mu_1 &= 950 - \left(\frac{100}{\sqrt{14}}\right) \Phi^{-1}(0.95) \\
 \iff \mu_1 &= 950 - (100/\text{sqrt}(14))*\text{qnorm}(0.95) \\
 \iff \mu_1 &= 906
 \end{aligned}$$

- Hoe groot moet de steekproef zijn zodat Gustav met 95% zekerheid kan zeggen dat hij niet genoeg verkoopt, als de werkelijke verkoop 940 bedraagt?

Oplossing

$$\begin{aligned}
 \Phi\left(\frac{950 - 940}{100/\sqrt{n}}\right) &= 0.95 \iff \frac{10}{100/\sqrt{n}} = \Phi^{-1}(0.95) \\
 &\iff \sqrt{n} = 10 \cdot \Phi^{-1}(0.95) \\
 &\iff n = (10 * \text{qnorm}(0.95)) ** 2 \\
 &\iff n = 270
 \end{aligned}$$