TILASTOMATEMATIIKKA

Harjoitus 5, kevät 2018

1. Tuottajan tutkimusten mukaan 50:ssä tuotteessa on keskimäärin 4 virheellistä ja niinpä hän hyvitti ostajia laskuttamalla 50:stä tuotteesta 46:n hinnan. Ostaja osti suuren erän (monta tuhatta) tuotetta ja mietti, voisiko vaatia enemmän hyvitystä viallisista kuin mitä tuottaja ilman vaatimuksia antaa. Ostaja rakensi seuraavan päätöksentekomallin: hän ottaa satunnaisen näytteen tuotteita ja määrittää siitä virheellisten osuuden. Osoittautui, että 150 tuotteen näytteessä oli 22 virheellistä.

   a) Määrää sopivat hypoteesit ja testaa ne luottamustasolla 95 %. Vaatiko ostaja lisähyvitystä?

   b) Määrää virheellisten osuuden 95 % yksisuuntainen luottamusväli.

   c) Laske näytteen p-arvo a)-kohdan hypoteeseilla.

   d) Mitä yhteistä on kohdilla a)-c)?

Ratkaisu:

a) Testataan virheellisten tuotteiden suhteellista osuutta p. Koska valmistajan mukaan 50 tuotteessa esiintyy keskimäärin 4 virheellistä, on nollahypoteesiksi syytä valita $H_0 : p = p_0 = \frac{4}{50}$. Koska suurempi määrä virheellisiä antaisi aiheen vaatia enemmän hyvitystä, on vastahypoteesiksi syytä valita $H_1 : p > \frac{4}{50}$.
Olkoon $X =$ "virheellisten lukumäärä otoksessa" $\sim \mathrm{Bin}(150, p)$, missä p on testattava suhteellinen osuus. Sopiva testimuuttuja on

$$Z = \frac{\frac{X}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1).$$

Luetaan taulukosta kynnysarvo r_0, jolla toteutuu ehto

$$P(Z \leq r_0) = 1 - \alpha$$
$$\Phi(r_0) = 0.95$$

Tämä arvo on $r_0 = 1.645$ ja hypoteesi H_0 hyväksytään, jos havaintoaineistosta laskettu testimuuttujan arvo $z < 1.645$.
Lasketaan testimuuttujan arvo

$$z = \frac{\frac{22}{150} - \frac{2}{25}}{\sqrt{\frac{\frac{2}{25} \cdot \frac{23}{25}}{150}}} \simeq 3.01 > 1.645$$

Voidaan siis tehdä johtopäätös, että nollahypoteesi H_0 on hylättävä ja vastahypoteesi H_1 on hyväksyttävä ts. virheellisten suhteellinen osuus ylittää ilmoitetun. Tämän otoksen perusteella ostajalla kannattaisi vaatia lisähyvitystä.

b) Luottamusväli saadaan a)-kohdan epäyhtälöstä $Z \leq 1.645$, joka on yhtäpitävää epäyhtälön

$$p \geq \frac{X}{150} - 1.645\sqrt{\frac{p(1-p)}{150}}$$

kanssa. Sijoittamalla epäyhtälön oikealle puolelle havainnoista saadut estimaatit $x = 22$ ja $\hat{p} = \frac{22}{150}$ saadaan luottamusväliksi

$$I_p = [0.10, \infty[,$$

jos käytetään luentorungossa ja luentokalvoissa esitettyä yksisuuntaisen luottamusvälin määritelmää. Koska suhteellinen osuus ei koskaan voi ylittää lukua 1, niin tässä tapauksessa järkevä yksisuuntainen luottamusväli olisi

$$I_p = [0.10, 1].$$

On huomattava, että käytetystä menetelmästä johtuen jälkimmäisen välin peittotodennäköisyys on alle 95%. Tälle ongelmalle on kirjallisuudessa esitetty erilaisia ratkaisuja, mutta jääkööt ne joillekin muille kursseille.
Koska luottamusväli ei sisällä annettua $p_0 = \frac{4}{50} = 0.08$, päädytään samaan johtopäätökseen kuin a)-kohdassa.

c) Näytteen p-arvo a)-kohdan hypoteeseilla tarkoittaa lukua p, jolle

$$P(Z \geq 3.01) = p,$$

missä 3.01 on testimuuttujan Z arvo otoksessa, joka laskettiin a)-kohdassa. Normaalijakauman taulukosta saadaan

$$p = 1 - \Phi(3.01) \simeq 0.0013.$$

Tämänkin avulla voidaan tehdä sama johtopäätös kuin edellisissä kohdissa.

d) Edellä jo tuli todettua, että samaan johtopäätökseen voi päätyä millä tahansa kohtien a)-c) tavoilla.

2. Eräs kalastustarvikevalmistaja väittää, että uuden kuitusiiman vetolujuus on keskimäärin 8.0 [kg] ja hajonta $\sigma = 0.5$ [kg]. Mitattiin 50 siiman vetolujuus ja keskimääräiseksi vetolujuudeksi saatiin 7.8 [kg]. Testaa riskitasolla 1%, voidaanko valmistajan ilmoittamaa vetolujuutta pitää oikeana.

Ratkaisu:
Olkoon X siiman vetolujuus, jolla on $E(X) = \mu$ ja $\mathrm{Var}(x) = \sigma^2 = 0,25$. Koska hajontaa pidetään tunnettuna, voidaan käyttää odotusarvon Z-testiä. Asetetaan hypoteesit
$H_0 : \mu = \mu_0 = 8,0$ valmistajan väite on totta
$H_1 : \mu \neq \mu_0$ vetolujuus poikkeaa ilmoitetusta
Testimuuttujaksi valitaan

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Luetaan normaalijakauman taulukosta kynnysarvo r_0, jolla toteutuu ehto

$$P(|Z| \leq r_0) = 1 - 0,01$$

Tämä arvo on $r_0 = 2,575$ ja hypoteesin H_0 hyväksymisalue on siis $|z| \leq 2,575$.
Lasketaan testimuuttujan arvo

$$|z| = |\frac{7,8 - 8,0}{\frac{0,5}{\sqrt{50}}}| = 2,83 > 2,575$$

Voidaan siis tehdä johtopäätös, että nollahypoteesi H_0 on hylättävä ja vastahypoteesi H_1 on hyväksyttävä ts. <u>vetolujuus poikkeaa ilmoitetusta</u>.

3. Testataan (polttoaineen) alkoholipitoisuuden vaikutusta bensiinin kulutukseen. Pitkän seurannan perusteella on todettu, että vanhan bensiinin, joka sisälsi vähemmän alkoholia, kulutus oli keskimäärin 8.6 litraa sadalla kilometrillä. Enemmän alkoholia sisältävällä 95E10-bensiinillä todettiin 10 toisistaan riippumattomaa kulutuslukemaa [l/100 km]

$$8.8;\ 8.6;\ 8.5;\ 8.7;\ 8.9;\ 8.4;\ 8.8;\ 9.0;\ 8.8;\ 9.1.$$

Voidaanko väittää 95E10-bensiinin kulutuksen olevan suurempaa aikaisempaan verrattuna? (Valitse riskitasoksi 5%.)

Ratkaisu:
Koska kulutuksen X hajontaa ei tunneta, arvioidaan se otoksesta laskemalla otoskeskihajonta ja käytetään odotusarvon T-testiä normaalijakautuneelle muuttujalle.
Asetetaan hypoteesit
$H_0 : \mu = \mu_0 = 8,6$ etanolia sisältävän polttoaineen kulutus on sama
$H_1 : \mu > \mu_0$ etanolia sisältävän polttoaineen kulutus on suurempi
Testimuuttujaksi valitaan

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

Luetaan t-jakauman taulukosta vapausasteella 9 yksisuuntaisen testin kynnysarvo $r_0 = 1,833$ merkitsevyystasolla 0,05 ja hypoteesi H_0 hyväksytään, jos havainnoista laskettu muuttujan T arvo $t \leq 1,833$.
Lasketaan havainnoista tunnusluvut $\overline{X} = 8,76$ ja $s = 0,2171$, joiden avulla saadaan testisuureelle arvo

$$t = \frac{8,76 - 8,6}{0,2171/\sqrt{10}} = 2,33 > 1,833$$

Voidaan siis tehdä johtopäätös, että nollahypoteesi H_0 on hylättävä ja vastahypoteesi H_1 on hyväksyttävä ts. kulutus on suurempaa kuin aikaisemmin.

4. Hankalasti vaihdettavan komponentin käyttöiältä vaadittiin, että sen odotettavissa oleva käyttöikä on ainakin 95 %:n varmuudella vähintään 1100 käyttötuntia. Järkevän hintatarjouksen tehneen valmistajan komponentteja testattiin valitsemalla tuotannosta satunnaisesti 15 komponenttia, jotka käytettiin loppuun aidoissa olosuhteissa. Havaitut komponenttien kestoiät tunteina olivat

$$998, 1217, 1093, 1124, 1270, 1220, 1018, 1096, 1150, 1111, 1311, 1005, 1190, 1075, 1060.$$

Aseta sopivat hypoteesit ja testaa ne riskitasolla 5 %. Kannattaako tämän otoksen perusteella hankkia komponentteja kyseiseltä valmistajalta?

Ratkaisu:
Testataan hypoteesit $H_0 : \mu \leq 1100$ vastaan $H_1 : \mu > 1100$. Tässä hypoteesien valinnassa kannattaa kiinnittää huomio siihen, että vastahypoteesi kantaa todistustaakkaa. Vastahypoteesiin päädytään vain, jos data **tukee vahvasti vastahypoteesia**. Koska komponentti on vaikeasti vaihdettava, haluamme olla vahvasti vakuuttuneita siitä, että käyttöikä on vähintään 1100 tuntia.
Koska hajonta σ on tuntematon, sopiva testimuuttuja on

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1},$$

missä $\overline{X}$ on käyttöikien aritmeettinen keskiarvo ja μ on (koko populaation) käyttöiän odotusarvo. Määrätään t-jakauman taulukosta vapausasteilla $n - 1 = 14$ kynnysarvo r_0 siten, että

$$P(T > r_0) = 0.05 \Rightarrow r_0 = 1.761.$$

Lasketaan testimuuttujan arvo otoksessa. Laskemalla saadaan $\overline{x} = 1129.2$ ja $s = 95.7901$, joten testimuuttujan arvo nollahypoteesin vallitessa on

$$t = \frac{1129.2 - 1100}{95.7901/\sqrt{15}} \simeq 1.18 \leq 1.761$$

joten H_0 hyväksytään. Otoksen perusteella käyttöikä on siis korkeintaan 1100 tuntia, joten tämän otoksen perusteella hankalasti vaihdettavaa komponenttia ei kannata hankkia kyseiseltä valmistajalta.

5. Haluttiin selvittää, hidastaako uusi seerumi leukemian etenemistä. Sitä varten valittiin 9 hiirtä, joilla tauti oli pitkälle kehittyneessä vaiheessa. Viisi hiirtä sai seerumihoitoa ja 4 ei. Hiirten elinajoiksi (vuosissa) hoidon aloittamisen jälkeen saatiin

Hoito	2.1	5.3	1.4	4.6	0.9
Ei hoitoa	1.6	0.5	2.8	3.1	

Voidaanko sanoa, että seerumihoito oli tehokasta merkitsevyystasolla $\alpha = 0.05$? Oletetaan, että elinajat ovat normaalijakautuneita.

Ratkaisu:
Tehdään odotusarvojen erotuksen testi. Merkitään
X = elinaika seerumihoidolla, $X \sim N(\mu_x, s_x^2)$
Y = elinaika ilman hoitoa, $Y \sim N(\mu_y, s_y^2)$
Asetetaan hypoteesit
$H_0 : \mu_x = \mu_y$ seerumihoito ei ole tehokasta
$H_1 : \mu_x > \mu_y$ seerumihoito on tehokasta
Testimuuttuja on

$$T = \frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{1}{n} + \frac{1}{m}} \cdot \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}} \sim t_{n+m-2}$$

Nollahypoteesi hyväksytään, jos $t \leq r_0 = 1,895$, missä r_0 löytyy t-jakauman taulukosta vapausasteella $n + m - 2 = 7$ ja merkitsevyystasolla 0,05.
Havainnoista saadaan $\overline{x} = 2,86$, $s_x = 1,9705$, $n = 5$ ja $\overline{y} = 2,0, s_y = 1,1916, m = 4$.
Testimuuttujan arvo saadaan sijoittamalla arvot yo. kaavaan ja käytetään nollahypoteesin mukaista arvoa $\mu_x - \mu_y = 0$.

$$t \simeq 0,762 < 1,895$$

Ei siis ole syytä hylätä nollahypoteesia ja todetaan, että <u>seerumihoito ei ole tehokasta</u>.