[MTTTP1] TILASTOTIETEEN JOHDANTOKURSSI, Kevät 2019
https://coursepages.uta.fi/mtttp1/kevat-2019/

HARJOITUS 5 viikko 15

Joitain ratkaisuja

1. $H_0: \pi = 10$, $H_1: \pi \neq 10$. Kun H_0 on tosi, niin

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(100-\pi_0)/n}} \overset{likimain}{\sim} N(0,1).$$ Lasketaan testisuureelle arvo. Saadaan

$$z = \frac{9-10}{\sqrt{10(100-10)/200}} = -0.47.$$ Normaalijakauman taulukosta luentomonisteen liitteestä 2 saadaan $z_{0,05} = 1{,}6449$. Koska $-1{,}6449 < -0{,}47 < 1{,}6449$, niin H_0 hyväksytään kaksisuuntaisessa testissä 10 %:n ristitasolla tarkasteltuna. Voidaan siis päätellä, että 10 % pelaajista saa voiton. Pienin riskitaso, jolla nollahypoteesisi voidaan hylätä (p-arvo), on $P(Z > 0{,}47 \text{ tai } Z < -0{,}47) > 0{,}10$.

2. $H_0: \pi = 4$, $H_1: \pi > 4$. Kun H_0 on tosi, niin

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(100-\pi_0)/n}} \overset{likimain}{\sim} N(0,1).$$ Nyt $z = \frac{5{,}6-4}{\sqrt{4(100-4)/500}} = 1{,}83.$

Normaalijakauman taulukosta luentomonisteen liitteestä 2 saadaan $z_{0,05} = 1{,}6449$ ja $z_{0,025} = 1{,}96$. Havaittu arvo on tällä välillä, joten 5 %:n riskitasolla nollahypoteesi hylätään, päätellään virheellisiä olevan yli 4 %, mutta 2,5 %:n riskitasolla hyväksytään. Pienin riskitaso, jolla nollahypoteesisi voidaan hylätä (p-arvo), on $0{,}025 < P(Z > 1{,}83) < 0{,}05$.

3. $H_0: \mu = 1000$, $H_1: \mu \neq 1000$. Kun H_0 on tosi, niin $t = \frac{\overline{X} - 1000}{s/\sqrt{n}} \sim t(n-1)$.

Aineistosta laskettu $t = \frac{1002-1000}{3{,}4/\sqrt{20}} = 2{,}63$. Onko saatu t-arvo harvinaisten vai tavanomaisten arvojen joukkoon kuuluva? Verrataan laskettua t-arvoa taulukkoarvoon. Nyt $\alpha = 0{,}01$, $\alpha/2 = 0{,}005$ ja $t_{0,005;19} = 2{,}861$. Koska $-2{,}861 < 2{,}63 < 2{,}861$, niin H_0 hyväksytään 1 % riskitasolla tarkasteltuna. Päätellään koneen toimivan toivotulla tavalla. Jos riskitaso olisi 2 %, niin nollahypoteesi tulisi hylättyä, koska $t_{0,02/2;19} = 2{,}539$. Tällöin pääteltäisiin, että kone ei tuota keskimäärin kilon pusseja. Siis p-arvo on suurempi kuin 1 %, tarkemmin $0{,}01 <$ p-arvo $< 0{,}02$ (kaksisuuntaisessa testissä).

4. $H_0: \mu = 32{,}0$, $H_1: \mu \neq 32{,}0$. Kun H_0 on tosi, niin $t = \dfrac{\overline{X} - 1000}{s/\sqrt{n}} \sim t(n-1)$.

Aineistosta laskettu $t = \dfrac{35{,}0 - 32{,}0}{6{,}8/\sqrt{30}} = 2{,}42$. Onko saatu t-arvo harvinaisten vai tavanomaisten arvojen joukkoon kuuluva? Verrataan laskettua t-arvo taulukkoarvoon. Nyt $\alpha = 0{,}05$, $\alpha/2 = 0{,}025$ ja $t_{0,025;29} = 2{,}045$. Koska $2{,}045 < 2{,}42$, niin H_0 hylätään 5 % riskitasolla tarkasteltuna. Päätellään, että ei kyseisen kirjailijan tuotantoa. Jos riskitaso olisi 2 %, niin nollahypoteesi tulisi hyväksyttyä, koska $t_{0,02/2;29} = 2{,}462 > 2{,}42$. Siis $0{,}02 < p < 0{,}05$ (kaksisuuntaisessa testissä).

5.
H_0 : Miesten ja naisten keskimääräinen päivässä saama kolesterolimäärä on sama (odotusarvot samoja)
H_1 : eivät samoja

Testisuureen arvo on 4,665 (katsotaan 1. riviltä, koska voidaan olettaa populaatioiden varianssit yhtä suuriksi, F-testiin liittyvä p = 0,083) ja H_0 hylätään, koska pienin riskitaso, jolla H_0 voidaan hylätä, on < 0,001 (kolmella desimaalilla 0,000). Päätellään, että keskimääräiset kolesterolimäärät eivät samoja.

6.
H_0: Yötyötä ja päivätyötä tekevien (kaikki työntekijät) poissaolojen keskiarvot (odotusarvot) samoja
H_1: eivät samoja

Testisuureen arvo on 2,570 (katsotaan 1. riviltä, koska voidaan olettaa populaatioiden varianssit yhtä suuriksi, F-testiin liittyvä p = 0,173) ja H_0 hylätään 5 %:n riskitasolla, koska pienin riskitaso, jolla H_0 voidaan hylätä, on 0,019 < 0,05. Päätellään, että keskimääräiset poissaolomäärät eivät samoja. Jos tarkastellaan 1 %:n riskitasolla, niin H_0 hyväksytään.

7.

H_0 : ei riippuvuutta

H_1 : on riippuvuutta.

Koska $p = 0{,}344 > 0{,}05$, niin nollahypoteesi hyväksytään eli ei riippuvuutta.

8. H_0 : ei riippuvuutta

H_1 : on riippuvuutta.

Koska $p = 0{,}0142$ eli pienin riskitaso, jolla H_0 voidaan hylätä, niin H_0 hylätään tätä suuremmilla riskitasoilla, mutta hyväksytään pienemmillä. Esim. 1 % riskitasolla H_0 hyväksytään, mutta 1,5 % riskitasolla hylätään, jolloin päätellään riippuvuutta olevan

9. Tutkitaan lineaarisen riippuvuuden olemassaoloa. Testattu hypoteesi

H_0 : Muuttujien välinen korrelaatiokerroin populaatiossa nolla, merkitään $\rho = 0$

H_1 : $\rho \neq 0$.

Pienin riskitaso, jolla H_0voidaan hylätä on tuloksista kolmen desimaalin tarkkuudella 0,000; voidaan sanoa siis, että $p < 0{,}001$. H_0 hylätään. Päätellään, että populaatiossa muuttujien välillä lineaarista riippuvuutta.