Tilastomatematiikka, Harjoitus 5

Tehtävä 1.

Maaperän arseenipitoisuutta mitattiin kahdella eri tavalla. Tavalla 1 mitaten saatiin n1 = 10 mittauksen otoshajonnaksi s1 = 0, 8 ppm ja tavalla 2 mitaten n2 = 12 mittauksen otoshajonnaksi s2 = 1, 2 ppm. Mittausten tulosten voidaan olettaa olevan normaalijakautuneita. Testaa riskitasolla α = 0, 05 onko mittaustapojen variansseissa eroa vai ovatko varianssit yhtäsuuret. a) luottamusrajaa b) testisuuretta c) p-arvoa.

```
n1 = 10;
s1 = 0.8;
n2 = 12;
s2 = 1.2;
a = 0.05;
va1 = n1-1
```

```
va1 = 9
```

```
va2 = n2-1
```

```
va2 = 11
```

Hypoteesit ovat siis H0: var1 = var2 ja H1: var1 =/= var2

a) Ratkaisu:

Jos varianssit ovat yhtäsuuret, niiden suhde on 1. Jos 1 on luottamusvälillä, pysytään nollahypoteesissa.

Käytetään F-jakautunutta muuttujaa: $F = \frac{\sigma_2^2}{\sigma_1^2} \times \frac{S_1^2}{S_2^2}$

Luottamusrajat saadaan kaavoilla $\frac{\hat{\sigma_1}}{\sigma_{2L}} = \frac{s_1}{s_2} \cdot \frac{1}{\sqrt{f_{2;\alpha/2}}}$ ja $\frac{\hat{\sigma_1}}{\sigma_{2U}} = \frac{s_1}{s_2} \cdot \frac{1}{\sqrt{f_{1;\alpha/2}}}$

```
f1 = finv(a/2, va1, va2)
```

```
f1 = 0.2556
```

```
f2 = finv(1-a/2, va1, va2)
```

```
f2 = 3.5879
```

```
alaraja = (s1/s2)*(1/sqrt(f2))
```

```
alaraja = 0.3520
```

```
ylaraja = (s1/s2)*(1/sqrt(f1))
```

```
ylaraja = 1.3186
```

a) Vastaus:

Pysytään nollahypoteesissa koska 1 osuu luottamusvälille.

b) Ratkaisu:

Tutkittava testisuure on $f = \frac{s_1^2}{s_2^2}$

```
f = (s1^2)/(s2^2)
```

```
f = 0.4444
```

Tutkitaan osuuko se arvojen $f_{1,\alpha/2}$ ja $f_{2,\alpha/2}$ välille.

```
f1 = finv(a/2, va1, va2)
```

```
f1 = 0.2556
```

```
f2 = finv(1-a/2, va1, va2)
```

```
f2 = 3.5879
```

b) Vastaus:

Koska osuu, pysytään nollahypoteesissa.

c) Ratkaisu:

Etsitään P-arvo kaavalla $2 \cdot min\{P(F \leq f), 1 - P(F \leq f)\}$, jossa muutuja f saadaan kaavalla $f = \frac{s_1^2}{s_2^2}$

```
f = (s1^2)/(s2^2)
```

```
f = 0.4444
```

```
p_arvo = 2*min(fcdf(f, va1, va2), 1-fcdf(f, va1, va2))
```

```
p_arvo = 0.2336
```

c) Vastaus:

Koska p_arvo = 0.2336 on isompi kuin riskitaso a = 0.05 ei hylätä nollahypoteesia.

Tehtävä 2.

Edellisten harjoitusten viimeisessä tehtävässä todettiin, että otoskoon ollessa suuri otosvarianssi S noudattaa likimain normaalijakaumaa odotusarvolla σ ja varianssilla σ^2 / 2n . Tämän jälkeen testattiin hypoteesiparia H0 : σ = 10 ja H1 : σ > 10 otoksella, jossa otoskoko oli 100 ja realisoitunut otoshajonta oli s = 13, 03. Testaus suoritettiin käyttämällä sekä tilanteesta S ~ N(σ, σ^2/2n) standardoimalla saatavaa muuttujaa että tavallista varianssin testauksessa käytettävää χ 2 -jakautunutta muuttujaa. Laske p-arvot molemmilla tavoilla.

Ratkaisu:

Otetaan tehtävän muuttujat talteen ja merkitään myös sigmaksi 10, koska se on nollahypoteesi.

```
n = 100;
S = 13.03;
sigma = 10;
va = n-1
```

```
va = 99
```

Koska H1 on sigma > 10, saadaan p-arvo todennäköisyydestä $P(A \geq a)$

Lasketaan ensiksi p-arvo standardoidulle Z muuttujalle

Standardoidaan muuttuja S. Kaavalla $Z = \frac{S-\sigma}{\sigma/\sqrt{2n}}$ saadaan muuttuja Z.

```
Z = (S-sigma)/(sigma/sqrt(2*n))
```

```
Z = 4.2851
```

```
p_arvo_std = normcdf(Z, 0, 1, 'upper')
```

```
p_arvo_std = 9.1342e-06
```

Ja sitten X^2 jakautunutta muuttujaa

Lasketaan ensin kyseinen muuttuja kaavalla $X^2 = \frac{(n-1)S^2}{\sigma^2}$

```
X2 = ((n-1)*S^2)/(sigma^2)
```

```
X2 = 168.0831
```

```
p_arvo_x2 = chi2cdf(X2, va, 'upper')
```

```
p_arvo_x2 = 1.8213e-05
```

Vastaus:

Standardoimalla saadaan p-arvo 9.1342e-06 ja X^2 muuttujalla 1.8213e-05

Tehtävä 3.

Suomessa eri veriryhmät esiintyvät Punaisen ristin veripalvelun mukaan väestössä seuraavasti:

Veriryhmä O A B AB

% väestöstä 33 42 17 8

Yhdysvalloissa aikanaan tehdyssä tutkimuksessa saatiin 200 ihmisen jakautumisesta veriryhmiin seuraavat tulokset

Veriryhmä O A B AB

fi 89 84 20 7

Testaa riskitasolla α = 0, 01 onko Yhdysvalloissa jakautuminen eri veriryhmiin sama kuin Suomessa? 'Vastaus': h ≈ 18, 84.

Ratkaisu:

```
s_O = 33;
s_A = 42;
s_B = 17;
s_AB = 8;
fin = [33, 42, 17, 8];

y_O = 89;
y_A = 84;
y_B = 20;
y_AB = 7;
usa = [89, 84, 20, 7];

n = 200;
a = 0.01;
```

Nollahypoteesiksi H0 otetaan että Yhdysvaltojen veriryhmien jakauma on sama kuin Suomessa. Vaihtoehtoinen H1 on että näin ei ole, eli joku (tai useampi) em. todennäköisyyksistä ei ole sama kuin Suomessa.

Aloitetaan laskemalla X^2 jakautunut muuttuja $H = \sum_{i=1}^{k} \frac{(F_i - np_i)^2}{np_i}$, jossa k on eri frekvenssiluokkien lkm (tässä 4), F_i toteunut frekvenssi ja np_i nollahypoteesin mukainen frekvenssi.

```
np = n*fin/(sum(fin));
F = usa;
h = sum( ((F-np).^2)./np )
```

```
h = 18.8424
```

Lasketaan p-arvo X^2 jakauman loppuhännästä

```
k = length(fin)
```

```
k = 4
```

```
va = k-1
```

```
va = 3
```

```
p_arvo = chi2cdf(h, va, 'upper')
```

```
p_arvo = 2.9470e-04
```

```
p_arvo < a
```

```
ans =
     1
```

Vastaus:

Koska p-arvo on pienempi kuin riskitaso, hylätään nollahypoteesi ja jatketaan vaihtoehtoisellahypoteesilla H1.

Tehtävä 4.

Kolmea eri yskänlääkettä V, R ja T testattiin kutakin 50 lääketieteen opiskelijalla. Näin saatiin alla olevat tulokset. Eroavatko yskänlääkkeet toisistaan vaikutuksen puolesta? 1 Käytä p-arvoa.

Vaikutus V R T Σ

Olematon 11 13 09 33

Lievä 32 28 27 87

Selvä 07 09 14 30

Σ 50 50 50 150

'Vastaus': h ≈ 3, 81.

Ratkaisu:

Tarkastellaan kahta eri populaatiota, lääke ja vaikutus. Molemmat ovat jakautuneet 3 tapaukseen.

Otetaan kontingenssitaulukko talteen.

```
F = [11, 13, 9; 32, 28, 27; 7, 9, 14];
f = [33, 87, 30]';
g = [50, 50, 50];
k = length(f);
l = length(g);
n = 150;
```

Lasketaan testisuure h kaavalla $H = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{(f_{i,j}-f_i g_j/n)^2}{f_i g_j/n}$.

```
h = 0;
for i=1:k
    for j=1:l
        h = h + ( ( F(i,j)-f(i)*g(j)/n )^2 ) / ( f(i)*g(j)/n );
    end
end
h
```

```
h = 3.8100
```

Muuttuja h noudattaa X^2 jakaumaa vapausastein (k-1)(l-1)

```
va = (k-1)*(l-1)
```

```
va = 4
```

```
p_arvo = chi2cdf(h, va, 'upper')
```

```
p_arvo = 0.4323
```

Vastaus:

P-arvoksi saatiin 0.4323. P-arvon ollessa näin suuri, ei nollahypoteesia H0 voida hylätä.