



Python-harjoitustyö

Outi Laurila, 1903248

HARJOITUSTYÖ
Joulukuu 2021

Tietojenkäsittely
Ohjelmistotuotanto

TIIVISTELMÄ

Tampereen ammattikorkeakoulu
Tietojenkäsittely
Ohjelmistotuotanto

LAURILA, OUTI
Python-Harjoitustyö

Harjoitustyö 18 sivua
Joulukuu 2021

Harjoitustyö käsittelee golfpalvelujen hintoja, golfkenttien sijainteja, kenttien kokonaispituuksia sekä par-3-, par-4- ja par-5-väylien määriä. Työssä tuotetaan ja kaumat golfautojen vuokrasta ja harjoituspallojen hinnasta sekä tehdä niistä ristiintaulukointi.

Golfauton vuokra vaihteli 10 ja 50 euron välillä ja pallokorin hinta 2 ja 5 euron välillä. Hintojen välillä ei havaittu yhteyttä.

Kohderyhmän golfkentät sijaitsevat kahdeksan maakunnan alueella, suurin osa Pirkanmaalla. Hämeenlinnan ja Porin kaupunkien alueilla on viisi kenttää, Naantalissa neljä, Heinolassa ja Tampereella kolme. Kuudessa kunnassa on kaksi kenttää ja yhden kentän kuntia on 18.

Koko aineistossa par-3-väylien osuus oli 29,8 %, par-4-väylien 50,4 % ja par-5-väylien osuus 19,8 %. Kenttien kokonaismatkojen tyyppiarvo eli moodi kauimmaiselta tiiltä mitattuna on 5789 metriä. Väylien määrä selittää 94 % kenttien pituuden vaihtelusta. Myös par-3-väylien määrän ja kentän kokonaispituuden välillä on riippuvuutta.

Asiasanat: golf, hinnat, par-3-väylät, par-4-väylät, par-5-väylät, pituudet

SISÄLLYS

1	JOHDANTO	4
2	GOLFAUTON VUOKRA JA HARJOITUSPALLOJEN HINTA	5
3	GOLFKENTTIEN SIJAINNIT	8
4	GOLFKENTTIEN KOKONAISMATKAT	11
5	PAR-LUVUT	14
5.1	Par-3-, par-4- ja par-5-väylien määrät	14
5.2	Par-3-väylien määrä ja kentän kokonaispituus	16
6	YHTEENVETO	18

1 JOHDANTO

Tämän harjoitustyön aihe on sama kuin SPSS-harjoitustyössäni eli se liittyy golfiin. Työn tarkoitus on tuottaa jakaumat golfautojen vuokrasta ja harjoituspallojen hinnasta sekä tehdä niistä ristiintaulukointi.

Kohderyhmän kenttien etäisyydet Tampereelta kuvataan pylväskaaviossa. Eri maakunnissa olevien kenttien määrät (frekvenssit) lasketaan. Kenttien määrät eri kunnissa luetellaan.

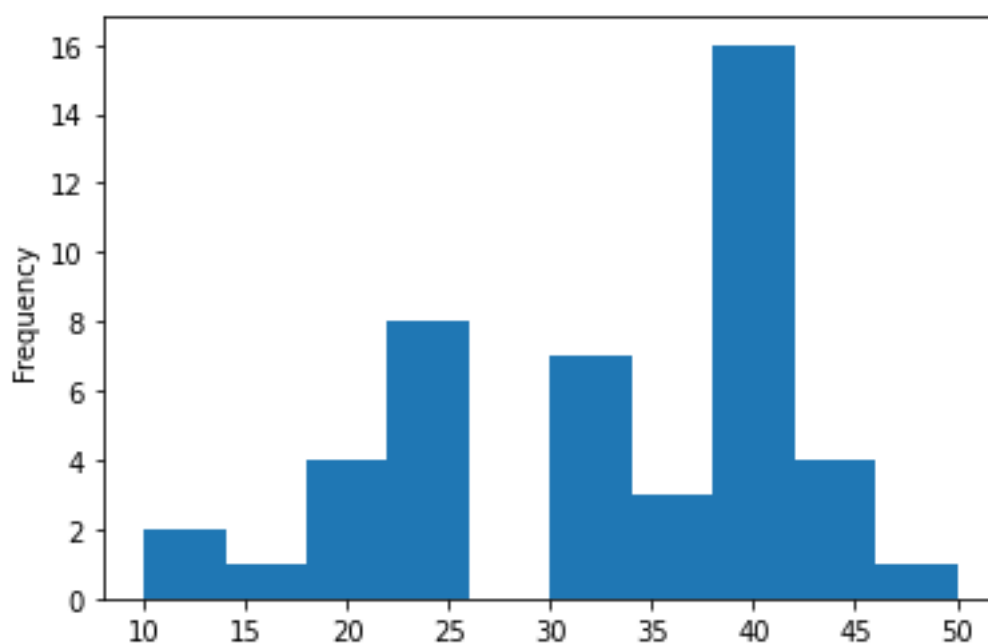
Golfkenttien väylien määrästä ja kokonaispituuksista tehdään regressioanalyysi. Par-3-, par-4- ja par-5-väylien määrästä lasketaan absoluuttinen ja prosentuaalinen frekvenssijakauma ja tehdään kaavio.

Par-3-väylien lukumääriä ja kentän kokonaispituutta tutkitaan khiin neliö -testin ja kontingenssianalyysin avulla.

Python-koodit esitetään harjoitustyössä kehystettyinä.

2 GOLFAUTON VUOKRA JA HARJOITUSPALLOJEN HINTA

Golfauton vuokraaminen pelikierroksen ajaksi maksaa kohderyhmän kentillä 10 – 50 euroa. Kuviossa 1 esitetään vuokrahintojen jakauma.



KUVIO 1. Golfauton vuokrahinnan jakauma

```
# -*- coding: latin-1 -*-
```

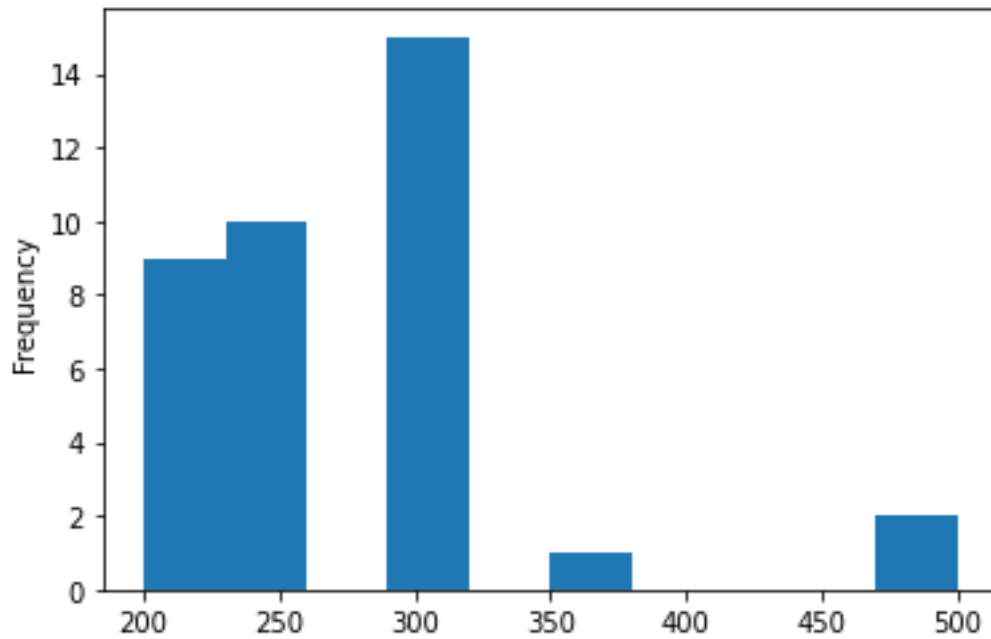
```
import pandas as pd
import matplotlib.pyplot as plt
```

```
data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analytiikka/CSV_LaurilaOuti.csv")
```

```
data["AutonHinta"].plot(kind = 'hist')
```

```
plt.show()
```

Harjoittelualueella käytettävien lainapallojen hinta/kori vaihtelee kahden ja viiden euron välillä. Kuviossa 2 on pallokorin hinnan jakauma.



KUVIO 2. Pallokorin hinnan jakauma (senttejä)

```
# -*- coding: latin-1 -*-

import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analyt-
iikka/CSV_LaurilaOuti.csv")

data["Rangekori_c"].plot(kind = 'hist')

plt.show()
```

TAULUKKO 1. Pallokorin hinnan ja golfauton vuokran ristiintaulukointi

row_0	200.0	250.0	300.0	350.0	500.0
col_0					
10.0	0	2	0	0	0
15.0	1	0	0	0	0
20.0	1	1	1	0	0
22.0	0	0	1	0	0
25.0	2	0	0	0	0
30.0	1	2	2	0	0
32.0	0	0	2	0	0
35.0	1	1	1	0	0
40.0	1	4	6	0	2
45.0	2	0	1	1	0
50.0	0	0	1	0	0

```
# -*- coding: latin-1 -*-
```

```
import pandas as pd
```

```
data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analytiikka/CSV_LaurilaOuti.csv")
```

```
x = data['Rangekori_c'].values
```

```
y = data['AutonHinta'].values
```

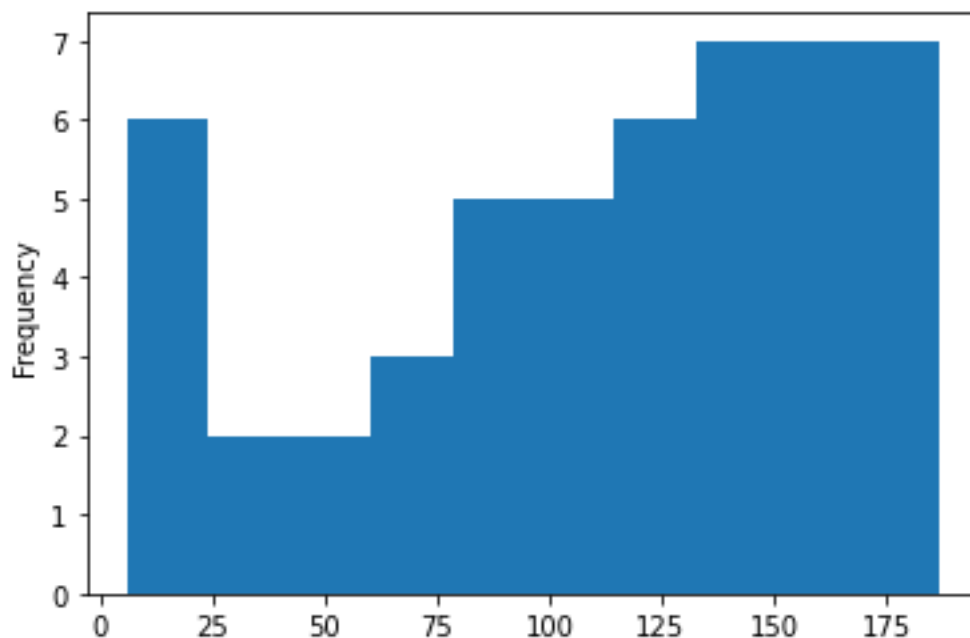
```
x = pd.crosstab([x], [y]).T
```

```
print(x)
```

Taulukon 1 silmämääräinen tarkastelu ei osoita selvää riippuvuutta. Pallokorien hintojen ei oletettukaan olevan yhteydessä golfautojen vuokrahintaan.

3 GOLFKENTTIEN SIJAINNIT

Kohderyhmän golfkentät sijaitsevat Tampereella tai korkeintaan noin 190 kilometrin etäisyydellä. Kuviossa 3 on etäisyyden jakauma.



KUVIO 3. Golfkentän etäisyys Tampereen Keskustorilta, jakauma

```
# -*- coding: latin-1 -*-

import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analytiikka/CSV_LaurilaOuti.csv")

data["EtäisyysTre"].plot(kind = 'hist')

plt.show()
```

Kohderyhmän golfkenttien sijainnin jakauma maakunnittain:

Etelä-Pohjanmaa :	4
Kanta-Häme :	7
Keski-Suomi :	3
Pirkanmaa :	13
Päijät-Häme :	7

Satakunta : 7
 Uusimaa : 2
 Varsinais-Suomi : 7

Mode: Pirkanmaa

Yleisin muuttujan arvo (moodi) on Pirkanmaa, jossa on 13 kenttää.

```
# -*- coding: utf-8 -*-

import pandas as pd
import statistics as st

data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analyt-
iikka/CSV_LaurilaOuti.csv")

my_list = data['Maakunta'].values
my_list.sort()
freq = {}
for item in my_list:
    if (item in freq):
        freq[item] += 1
    else:
        freq[item] = 1

for key, value in freq.items():
    print("\n", key, "\t", value)

most_often = st.mode(my_list)
print("Mode: ", most_often)
```

Monen kunnan alueella on useita golfkenttiä:

Asikkala : 1
 Hattula : 1
 Heinola : 3
 Hollola : 1
 Hyvinkää : 2
 Hämeenlinna : 5
 Ikaalinen : 2
 Jyväskylä : 1

Jämsä : 1
Kauhajoki : 1
Kuortane : 1
Lahti : 2
Lempäälä : 2
Loimaa : 1
Muurame : 1
Naantali : 4
Nakkila : 1
Nokia : 2
Parainen : 1
Pirkkala : 1
Pori : 5
Rauma : 1
Sastamala : 2
Seinäjoki : 1
Tampere : 3
Turku : 1
Virrat : 1
Ypäjä : 1
Ähtäri : 1

```
import pandas as pd

data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analy-
tiikka/CSV_LaurilaOuti.csv")

my_list = data['Kunta'].values
my_list.sort()
freq = {}
for item in my_list:
    if (item in freq):
        freq[item] += 1
    else:
        freq[item] = 1

for key, value in freq.items():
    print("\n", key, ": \t", value)
```

4 GOLFKENTTIEN KOKONAISMATKAT

Kenttien kokonaismatkat kauimmaiselta tiiltä mitattuna on lueteltu seuraavassa:

```
[ 607 680 834 1074 1079 1465 1470 2091 2337 2370 2386 2524 2548 2582
2663 2793 2797 2956 3097 4936 5249 5411 5716 5751 5760 5818 5819 5820
5840 5851 5936 5960 6023 6026 6032 6046 6055 6129 6134 6139 6160 6188
6257 6305 6340 6369 6411 6495 6624 6685]
```

Kenttien keskiarvopituus on 4492 metriä, tyyppiarvo 6129 metriä ja mediaani 5789 metriä. Keskihajonta eli muuttujan arvojen keskimääräinen poikkeama keskiarvosta on 2042 metriä. Variaatiokerroin (0.455) ilmoittaa vertailukelpoisen keskihajonnan.

Mean: 4492	Median: 5789.0
Mode: 6129	Standard deviation: 2041.6358637132137
Variation: 0.4545048672558356	

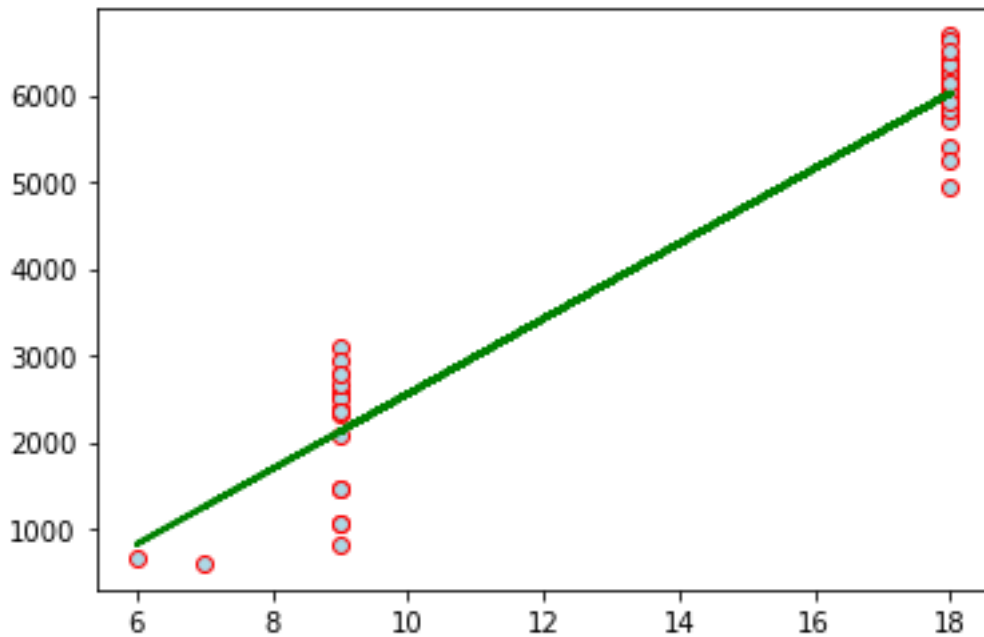
```
# -*- coding: utf-8 -*-
import statistics as st
import pandas as pd

data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analyt-
iikka/CSV_LaurilaOuti.csv")
x = data['PisinKokonaismatka'].values
x.sort()

print("\n", x, "\n")

average = st.mean(x)
middle_of = st.median(x)
most Often = st.mode(x)
deviation = st.stdev(x)
variation = st.stdev(x)/st.mean(x)

print("\nMean: ", average, "\tMedian: ", middle_of)
print("Mode: ", most Often, "\tStandard deviation: ", deviation)
print("Variation: ", variation)
```



KUVIO 4. Pisteparvikuvaaja ja trendiviiva: väylien määrä (x) ja kentän kokonaispituus metreinä kauimmaiselta tiiltä mitattuna (y)

Coefficient of determination, R-square is 0.9367334233013649

Intercept, constant is [-1770.86190439]

Slope, coefficient is [[432.57057351]]

Kentän väylien määrän vaikutusta kentän kokonaispituuteen tutkittiin regressioanalyysillä (kuvio 4). Regressioanalyysin mallin selitysaste R-square kertoo, että väylien määrä selittää noin 94 prosenttia kentän kokonaispituudesta. Intercept-arvo tarkoittaa, että trendiviivan y-arvo on noin -1771 kohdassa $x = 0$. Trendiviivan kulmakerroin on noin 433.

```
# -*- coding: utf-8 -*-
```

```
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.linear_model import LinearRegression
```

```
data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analyt-
iikka/CSV_LaurilaOuti.csv")
```

```
x = data['Vaylia'].values
x = x.reshape(-1,1) # x pystyvektoriksi
```

```
y = data['PisinKokonaismatka'].values
y = y.reshape(-1,1) # y pystyvektoriksi
```

```
model = LinearRegression()
model.fit(x, y) # muodostetaan lineaarinen regressiomalli
```

```
y_predicted = model.predict(x) # trendiviivan mukainen y-arvo
R_square = model.score(x, y)

print("")
print("Coefficient of determination, R-square is ", R_square)
print("Intercept, constant is ", model.intercept_)
print("Slope, coefficient is ", model.coef_)
print("")

plt.plot(x, y, 'ro', markerfacecolor = "lightblue")
plt.plot(x, y_predicted, 'g-', linewidth = 2)
plt.show()
```

5 PAR-LUVUT

5.1 Par-3-, par-4- ja par-5-väylien määrät

Yleensä golfkentillä on sekä par-3, par-4 että par-5 väyliä. Lyhyillä kentillä voi kuitenkin olla pelkkiä par-3-väyliä. Väylien jakaumat koko aineistossa olivat seuraavanlaiset:

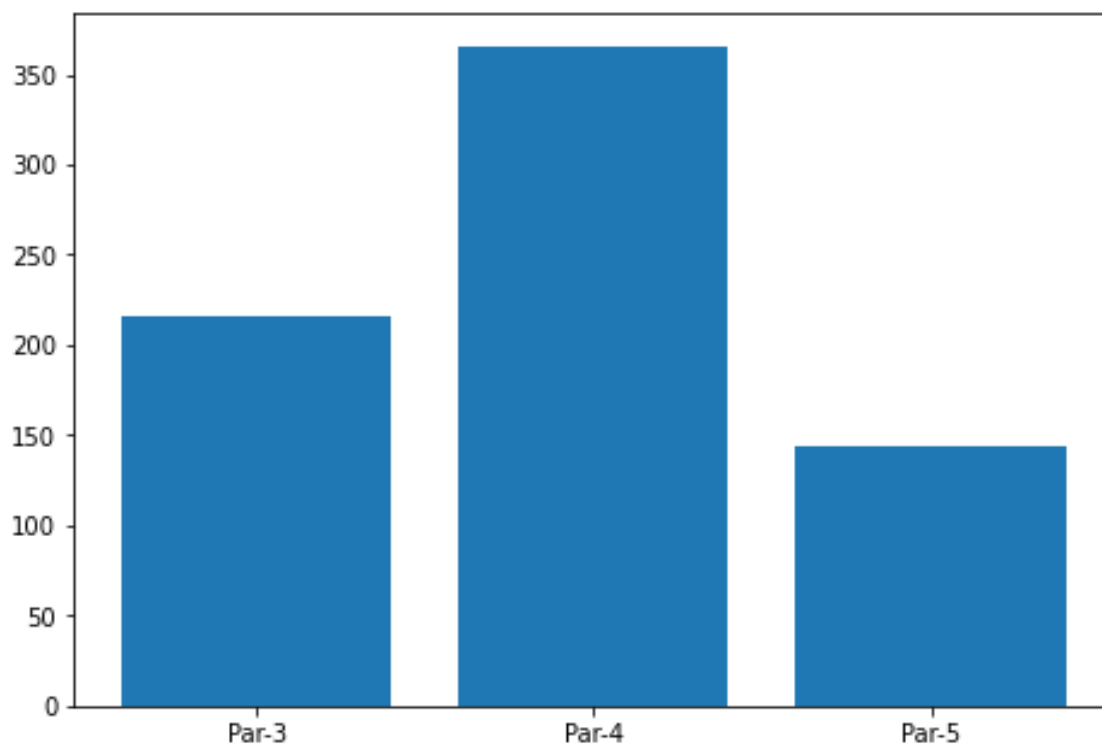
Par-3 väyliä: 216 (29.8 %)

Par-4 väyliä: 366 (50.4 %)

Par-5 väyliä: 144 (19.8 %)

Väyliä kaikkiaan: 726

Jakauma esitetään myös kuviossa 5.



KUVIO 5. Eri par-lukemaisten väylien jakauma

```

# -*- coding: utf-8 -*-
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analyt-
iikka/CSV_LaurilaOuti.csv")

par3 = data['Par3'].values
par4 = data['Par4'].values
par5 = data['Par5'].values

sum_par3 = 0
sum_par4 = 0
sum_par5 = 0

for i in range(len(par3)):
    sum_par3 += par3[i]

for i in range(len(par4)):
    sum_par4 += par4[i]

for i in range(len(par5)):
    sum_par5 += par5[i]

vaylat = sum_par5 + sum_par4 + sum_par3

par3proc = round(sum_par3/vaylat*100, 1)
par4proc = round(sum_par4/vaylat*100, 1)
par5proc = round(sum_par5/vaylat*100, 1)

print("Par-3 väyliä: ", sum_par3, '(',par3proc, '%) ')
print("Par-4 väyliä: ", sum_par4, '(',par4proc, '%) ')
print("Par-5 väyliä: ", sum_par5, '(',par5proc, '%) ')
print("Väyliä kaikkiaan: ", vaylat )

fig = plt.figure()

ax = fig.add_axes([0,0,1,1])

pars = ['Par-3', 'Par-4', 'Par-5']
sums = [sum_par3, sum_par4, sum_par5]

ax.bar(pars,sums)

plt.show()

```

5.2 Par-3-väylien määrä ja kentän kokonaispituus

Par-3-väylien lukumääriä ja kentän kokonaispituutta lähimmältä tiiltä mitattuna tutkittiin khiin neliö -testin ja kontingenssianalyysin avulla:

===Chi-Square Test Value===

291.3140401580271

===Degrees of Freedom===

Df = 49

===Significance===

Sig = 3.594616501074363e-36

Please, give me your value of risk: e.g. 0.001, 0.01 or 0.05: > 0.001

Result: These variables depend on each other (reject H0)

Riskiarvolla 0.001 saadaan tulokseksi, että muuttujilla on riippuvuutta: df=49; $X^2(2) = 291.31$; Sig=3.59 e-36.

```
# -*- coding: utf-8 -*-
import pandas as pd
import numpy as np
from scipy import stats

dice = pd.read_csv("C:/Users/Outi/OneDrive - TUNI.fi/Data-analyt-
iikka/CSV_LaurilaOuti.csv")

PAR3 = dice['Par3'].values
meters = dice['LyhinKokonaismatka'].values

dice = np.array([PAR3, meters])

stats.chi2_contingency(dice)

chi2_stat, sig, df, expected = stats.chi2_contingency(dice)
print("")
print("===Chi-Square Test Value===")
print(chi2_stat)
print("")
```



```
print("===Degrees of Freedom===")
print("Df = ", df)
print("")
print("===Significance===")
print("Sig = ", sig)
print("")

#Interpret sig value
critical = input("Please, give me your value of risk: e.g. 0.001, 0.01 or 0.05: > ")
critical = float(critical)

if sig <= critical:
    print("Result: These variables depend on each other (reject H0)")
else:
    print("Result: These variables are independent (H0 holds true)")
```

6 YHTEENVETO

Golfautojen vuokrahinnat vaihtelivat mielestäni melko paljon. Alle 18 väylän kentillä hinta on luonnollisesti edullisempi kuin täysmittaisilla kentillä, koska auton käyttöaika on lyhempi. Kallein hinta on kuitenkin jopa viisinkertainen halvimpaan nähden.

Pallokorien hinnoissa oli vaihtelua 2 ja 5 euron välillä. Yleisin hinta oli 2,5 e, mutta kahdella kentällä hinta oli 5 euroa. Ristiintaulukoinnissa ei havaittu riippuvuutta pallokorin ja auton vuokrahinnan välillä.

Kohderyhmän golfkentät sijaitsevat kahdeksan maakunnan alueella, suurin osa Pirkanmaalla. Hämeenlinnan ja Porin kaupunkien alueilla on viisi kenttää, Naantalissa neljä, Heinolassa ja Tampereella kolme. Kuudessa kunnassa on kaksi kenttää ja yhden kentän kuntia on 18.

Kenttien kokonaismatkat kauimmaiselta tiiltä mitattuna ovat keskimäärin 4492 metriä. Koska jakauma ei ole symmetrinen (pienimpiä arvoja on vähemmän kuin suurimpia), mediaani on parempi tunnusluku. Sen arvo on 5789 metriä. Keskihajonta on jopa 2041 metriä, kun taas lyhimmat kentät ovat alle 1000 metriä.

Regressioanalyysin avulla saatiin selville, että väylien määrä selittää 94 prosenttia kentän kokonaispituudesta. Tämä oli yllättävän suuri prosentti, sillä esimerkiksi yhdeksän väylän kenttien pituudet vaihtelevat alle tuhannesta yli kolmeen-tuhanteen metriin.

Aineiston 50 golfkentällä laskettiin olevan 726 väylää. Par-3-väyliä on 29,8 prosenttia, par-4-väyliä 50,4 prosenttia ja par-5-väyliä 19,8 prosenttia. Khiin neliö -testin ja kontingenssianalyysin avulla selvitettiin, että 0.001 riskiarvolla par-3-väylien määrän ja kentän kokonaispituuden (lähimmältä tiiltä mitattu) välillä on riippuvuutta. Tämä on ymmärrettävää, koska par-3 väylät ovat lyhyitä.