Tilastollisten mallien peruskurssi – harjoitustyö

Alustetaan 800:n kokoinen otos elinolo-tilastosta:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

## Regressiomalli

Luodaan sirontakuvio:

Kuva, joka sisältää kohteen diagrammi

Kuvaus luotu automaattisesti

Tarkasteltaessa asunnon pinta-alan suhdetta muihin, vaikuttaa siltä, että ainoastaan kuluttajayksiköiden lukumäärällä on yhteys siihen.

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisestiKorrelaatiot:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

Pearsonin korrelaatiokertoimien mukaan kaikki selittäjät ovat tilastollisesti merkitseviä, mutta ainoastaan kuluttajayksiköiden lukumäärän korrelaatio asunnon pinta-alaan on edes jokseenkin suuri.

Myös Spearmanin korrelaatioista ainoastaan kuluttajayksiköiden lukumäärän voidaan sanoa korreloivan asunnon pinta-alan kanssa, vaikka myös alueella asumisaika on tilastollisesti merkitsevä selittäjä. Voidaan sanoa, että alueella asumisajan ja asunnon pinta-alan välillä on heikkoa ei-lineaarista riippuvuutta.

Suoraviivauksien tarkastelu:

Tilastollisesti merkitseville selittäjille Spearmanin korrelaatiot ovat hieman suurempia, joten ei ole syytä ajatella yhteyksien olevan suoraviivaisia.

Yksinkertainen regressiomalli:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

Mallin selitysaste: 0.28, eli asunnon pinta-ala seuraa kuluttajayksiköiden lukumäärää 28 %:n tarkkuudella

p<0.001, joten malli on merkitsevä

Jäännöstarkastelu:

Kuva, joka sisältää kohteen kaavio

Kuvaus luotu automaattisestiKuva, joka sisältää kohteen kaavio

Kuvaus luotu automaattisesti

Jäännösten hajonta vaikuttaa hieman laskevan. Jakauma on myös positiivisesti vino, eli huomataan että yhteys ei ole suoraviivainen.

# Toistomittausmalli

Normaalijakaumatestit kaikille:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

Kaikille p-arvo<0.001, joten kaikki poikkeaa normaalijakaumasta merkittävästi.

Dataa on kuitenkin huomattavan paljon, joten voidaan suorittaa toistettujen mittausten varianssianalyysi.

Normaalijakaumatestit naisille erikseen:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

Normaalijakaumatestit miehille erikseen:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

Toistettujen mittausten varianssianalyysi:

Kaikki:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

F-testin p-arvo<0.001, eli on tilastollisesti merkitseviä eroja mielipiteiden keskiarvot eroavat toisistaan.

Naiset:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

F-testin p-arvo<0.001, eli on tilastollisesti merkitseviä eroja mielipiteiden keskiarvot eroavat toisistaan.

Miehet:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

F-testin p-arvo<0.001, eli on tilastollisesti merkitseviä eroja mielipiteiden keskiarvot eroavat toisistaan.

# Kategoristen vastemuuttujien mallitus

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

sukupuoli p=0.890, ei ole merkitsevä selittäjä

ikä p<0.001, on merkitsevä selittäjä

Merkitsevien selittäjien (ainoastaan ikä) malli:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **selittäjä** | **p** | **OR** | **95 % luottamusvälin alaraja OR:lle** | **95 % luottamusvälin yläraja OR:lle** |
| ikä | <0.001 | 0.960 | 0.948 | 0.972 |



Mallin selitysaste (Nagelkerke): 0.101

# Monimuuttujamenetelmät

Pääkomponenttianalyysi:

Kuva, joka sisältää kohteen teksti

Kuvaus luotu automaattisesti

Valitaan kolme pääkomponenttia, joista lataukset:

Kuva, joka sisältää kohteen teksti, piha-, kyltti

Kuvaus luotu automaattisesti