

# Teema 10: Regressio- ja varianssianalyysi

**Regressioanalyysi** lienee  $t$ -testin ohella maailman eniten käytetty tilastollinen menetelmä. Sitä sivuttiin jo alustavasti Teemassa 4.

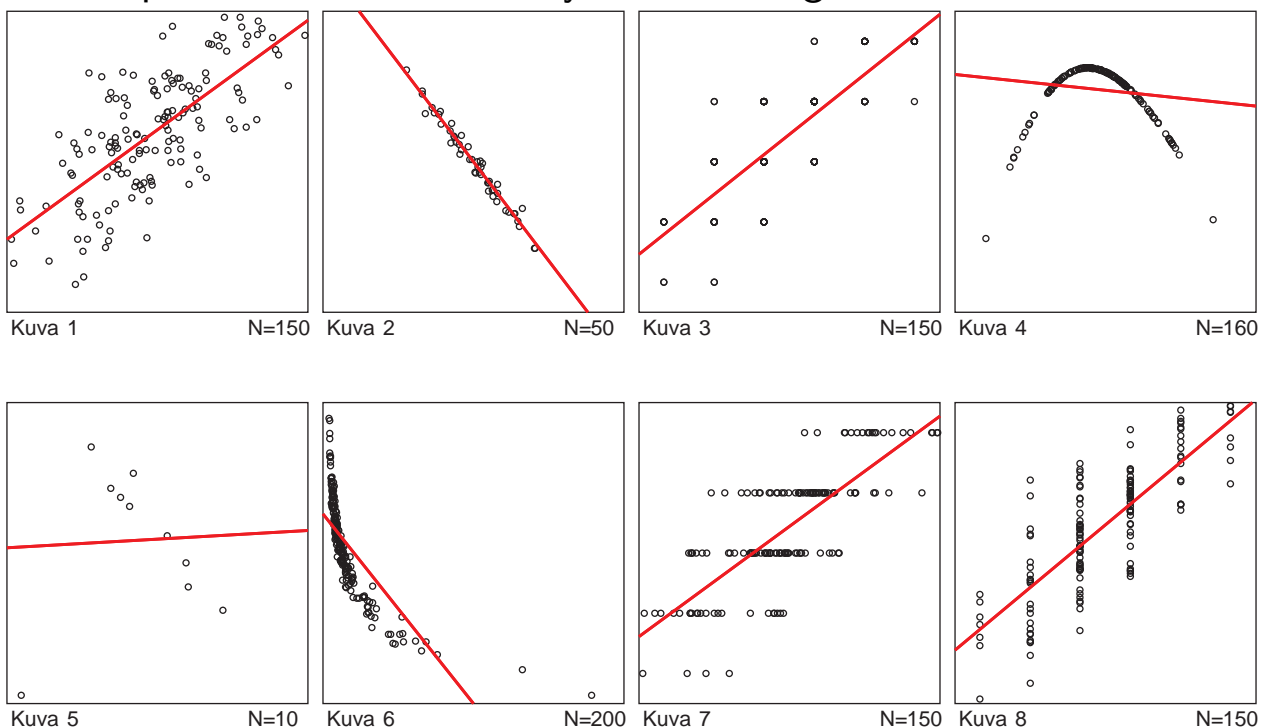
**Varianssianalyysi** liittyy useallakin tavalla regressioanalyysiin, ja on toisaalta  $t$ -testin yleistys silloin kun vertaillaan useampaa kuin kahta ryhmää. Menetelmät kietoutuvat siis vahvasti toisiinsa.

**Aivan lopuksi** täydennetään aiempia ristiintaulukoita koskevia tarkasteluja  $\chi^2$ -riippumattomuustestillä (luetaan "khi-toiseen").

Navigation icons: back, forward, search, etc.

## Esimerkkejä hajontakuvista simuloiduilla aineistoilla

Mieleepalautus Teemasta 4, nyt mukana regressiosuorat:



Navigation icons: back, forward, search, etc.

A set of navigation icons typically found in Beamer presentations, including symbols for back, forward, search, and other slide controls.

# Regressioanalyysin perusteet (jatkoa)

Regressioanalyysia käytetään paljon—valitettavasti myös väärin. Elleivät menetelmän oletukset päde (edes jollain tavalla), eivät tulokset tai johtopäätöksetkään voi olla kovin ihmeellisiä. Sama pätee tietenkin kaikkiin tilastollisiin menetelmiin.

Regressiomallin tärkeimmät oletukset koskevat mallivirhettä  $\varepsilon$ . Siitä oletetaan yleensä, että  $\varepsilon \sim N(0, \sigma^2)$ , mikä tarkoittaa että mallivirhe tuo regressiomalliin vain satunnaista, samansuuruista vaihtelua nollan molemmin puolin. Sen oletetaan siis noudattavan normaalijakaumaa, mikä on monissa sovelluksissa luonteva oletus.

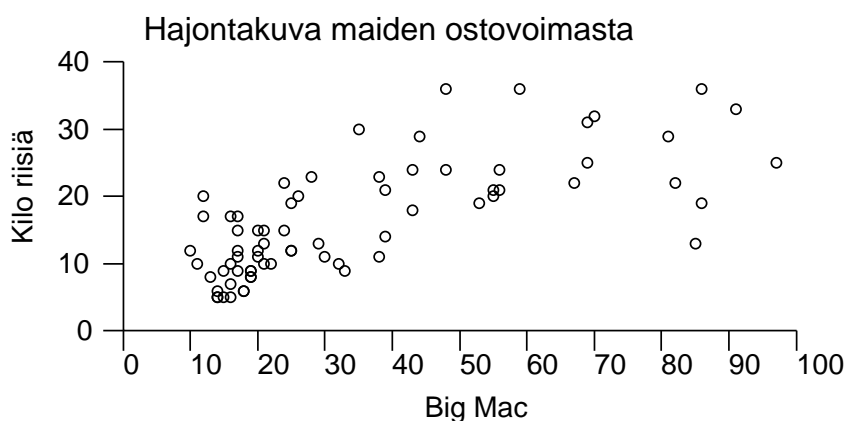
Regressioanalyysiin liittyy monenlaisia testaustilanteita, joissa käytetään mm.  $t$ -testiä. Jos normaalijakaumaoletus ei päde, testeistä tehdyt johtopäätökset menettävät luotettavuuttaan.

Mallivirhe on satunnaismuuttuja eikä sellaisenaan havaittavissa, mutta oletusten voimassaoloa tutkitaan mallin jäännöstermin eli **residuaalin** ( $y$ :n ja  $\hat{y}$ :n erotuksen) avulla.

Navigation icons: back, forward, search, etc.

## Esimerkki: *Prices and Earnings*, BigMac (Survo)

Tietoja 70 maasta ja kaupungista vuodelta 2006 (ks. Teema 4)



Hajontakuvia voi toki piirtää kumminpäin hyvänsä. Tutkitaan nyt tilannetta toisinpäin ja selitetään BigMac-indeksiä riisi-indeksillä:

```
Linear regression analysis: Data BIGMAC06, Regressand BigMac      N=70
Variable  Regr.coeff.      Std.dev.      t      p
Rice      2.002298         0.237603     8.427   0.000
constant  2.204914             4.367544     0.505   0.308
Variance of regressand BigMac=557.2730849 df=69
Residual variance=276.6006701 df=68
R=0.7147  R^2=0.5108
```

Navigation icons: back, forward, search, etc.

## Esimerkki jatkuu: tulkitaan tulostusta

Tulosteesta nähdään, että BigMac-indeksistä 51 % selittyy riisi-indeksillä (Rice), loppu satunnaisvaihtelulla:

Variable	Regr.coeff.	Std.dev.	t	p
Rice	2.002298	0.237603	8.427	0.000
constant	2.204914	4.367544	0.505	0.308

Residual variance=276.6006701 df=68  
R=0.7147 R^2=0.5108

Rice-muuttujan regressiokertoimen estimaatti on  $\hat{\beta}_1 = 2$ , eli kun Rice kasvaa yhden yksikön, BigMac kasvaa kaksi yksikköä. Tämä relaatio on nähtävissä myös kuvasta (vaikka se onkin toisinpäin). Yksikkönähän molemmissa muuttujissa on minuutti (työaika).

Kertoimen  $\hat{\beta}_1$  estimoitu hajonta (eli keskivirhe) on 0.237. Kerroin jaettuna keskivirheellään on  $t$ -testisuure (vrt. Teema 9) 8.427.

Hypoteesi  $H_0 : \beta_1 = 0$  kaatuu selvästi (vapausasteita on  $df=68$ ), joten kerroin poikkeaa nolasta (tilastollisesti merkitsevästi).

Vakiotermin kerroin on  $\hat{\beta}_0 = 2.2$ . Sillä on mallissa oma tehtävänsä, vaikkei se olekaan tilastollisesti merkitsevä ( $p=0.308$ ).

Navigation icons: back, forward, search, etc.

## Esimerkki jatkuu: lisätään toinen selittäjä

Lisätään malliin toiseksi selittäjäksi ruokakorin keskimääräistä hintaa kuvaava Basket (kotitalouden ruokamenot/kk, US\$):

Variable	Regr.coeff.	Std.dev.	t	p
Rice	1.322144	0.269971	4.897	0.000
Basket	-0.074029	0.017947	-4.125	0.000
constant	42.84683	10.54302	4.064	0.000

Residual variance=223.8777762 df=67  
R=0.7810 R^2=0.6099

Selitysaste paranee ja on nyt 61 %. Molemmat selittäjät ovat merkitseviä, mutta Rice menettää painoarvoaan ( $\hat{\beta}_1 = 1.32$ ). Vapausasteet vähenevät, kun estimoidaan yksi parametri lisää ( $\beta_2$ ).

Basket:in estimoitu kerroin  $\hat{\beta}_2 = -0.07$  on lukuna pieni, koska muuttujan arvot ovat vastaavasti suurehkoja lukuja.

Regressiomalleissa muuttujat voivat olla eri mittayksiköissä ja eri suuruusluokkaa.

Vakiotermikin nousee tilastollisesti merkitseväksi, vaikkei sillä edelleenkään ole sen ihmeellisempää merkitystä.

Navigation icons: back, forward, search, etc.



## Esimerkki jatkuu: lisätään kolmas selittäjä

Lisätään malliin vielä lomapäivien määrää kuvaava Vacdays, jota on tarkasteltu aiemmissa harjoituksissa:

Variable	Regr.coeff.	Std.dev.	t	p
Rice	1.216663	0.279237	4.357	0.000
Basket	-0.077837	0.018052	-4.312	0.000
Vacdays	-0.450943	0.331290	-1.361	0.178
constant	55.08178	13.76280	4.002	0.000

Residual variance=221.0640444 df=66  
R=0.7878 R^2=0.6206

Selitysaste paranee edelleen, mutta se ei ole ihme, sillä näin käy vaikka malliin lisättäisiin *mitä tahansa selittäjiä*. Selitysaste ei sellaisenaan olekaan riittävä peruste hyvälle mallille.

Huomaa, että Rice menettää taas painoarvoaan ja vapausasteet vähenevät yhdellä, kun  $\beta_3$  estimoidaan aineistosta.

Vacdays-selittäjän ottamiselle malliin ei kuitenkaan ole perusteluja, sillä sen estimoitu kerroin  $\hat{\beta}_3 = -0.45$  ei ole merkitsevä ( $p=0.178$ ). Hypoteesi  $H_0 : \beta_3 = 0$  jää voimaan, ja näin ollen Vacdays voidaan pudottaa pois ja palata aiempaan, suppeampaan malliin.

Navigation icons: back, forward, search, etc.

## Esimerkki jatkuu: vaihdetaan kolmas selittäjä

Otetaan Vacdays'in tilalle malliin muuttuja Bread, joka kuvaa sitä, kuinka kauan pitää työskennellä voidakseen ostaa kilon leipää:

Variable	Regr.coeff.	Std.dev.	t	p
Rice	0.802165	0.206413	3.886	0.000
Basket	-0.062346	0.013083	-4.766	0.000
Bread	0.826217	0.105148	7.858	0.000
constant	28.61589	7.844347	3.648	0.000

Residual variance=117.4217887 df=66  
R=0.8936 R^2=0.7985

Nyt selitysaste nousee selvästi, miltei 80 %:iin. Kaikki selittäjät ovat merkitseviä. On aika katsastaa mallia hieman syvemmältä.

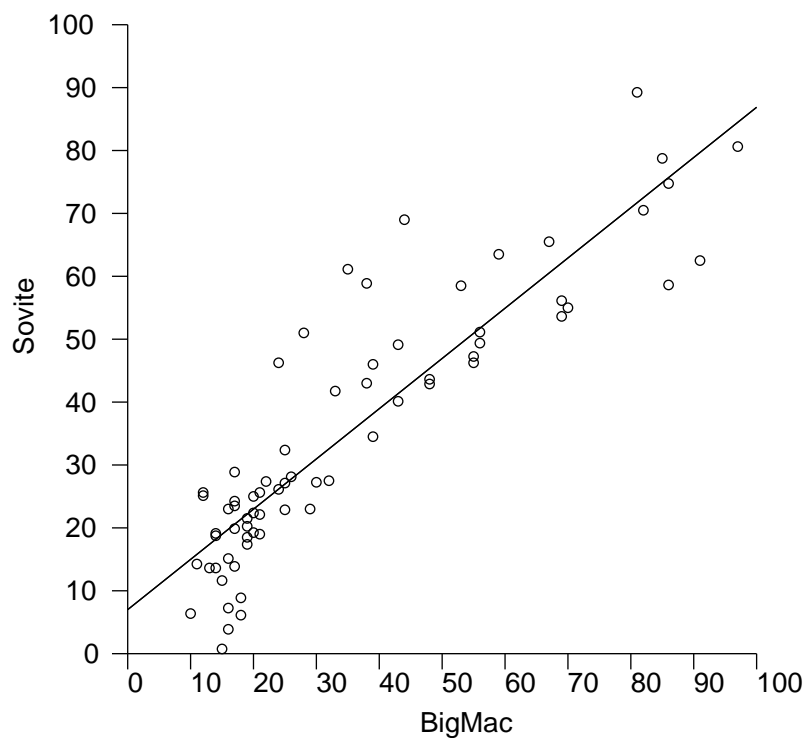
Piirretään muutama **diagnostiikkakuva**, jotta voidaan tutkia miten regressiomallin oletukset toteutuvat tämän mallin osalta:

- ▶ selitettävä muuttuja vastaan mallin sovite
- ▶ sovite vastaan residuaali (ns. *jäännösvaihteludiagrammi*)
- ▶ residuaalin normalisuus (ns. *todennäköisyyspaperikuva*)
- ▶ residuaalin normalisuus (histogrammi, normaalijakauma)

Navigation icons: back, forward, search, etc.



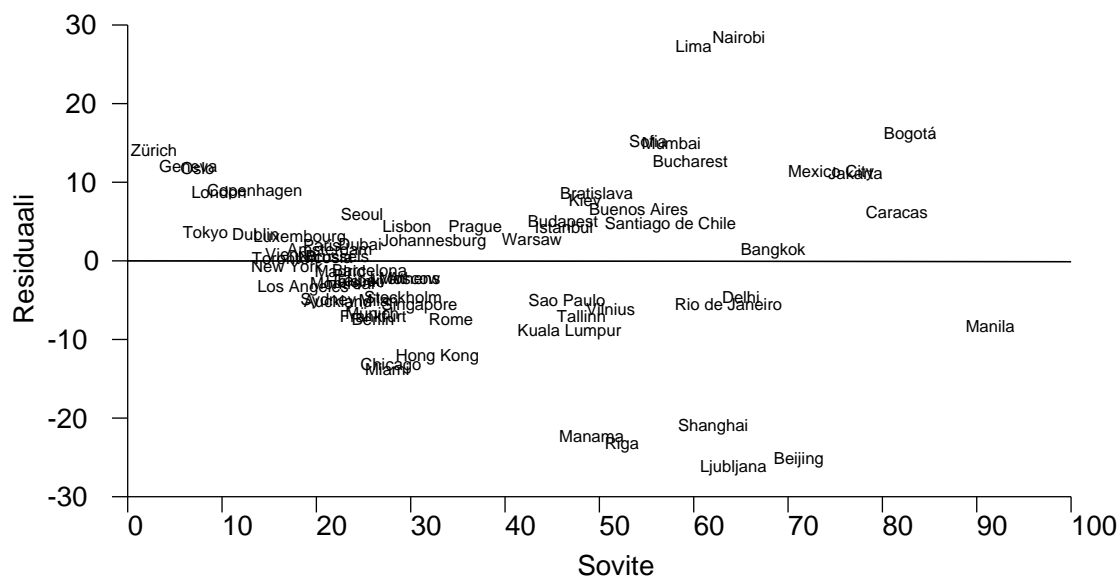
# Diagnostiikkakuva 1: selitettävä ja sovite



Mitä paremmin pisteet asettuvat suoralle, sitä lähempänä mallin antama sovite on selitettävää muuttujaa.

Navigation icons: back, forward, search, etc.

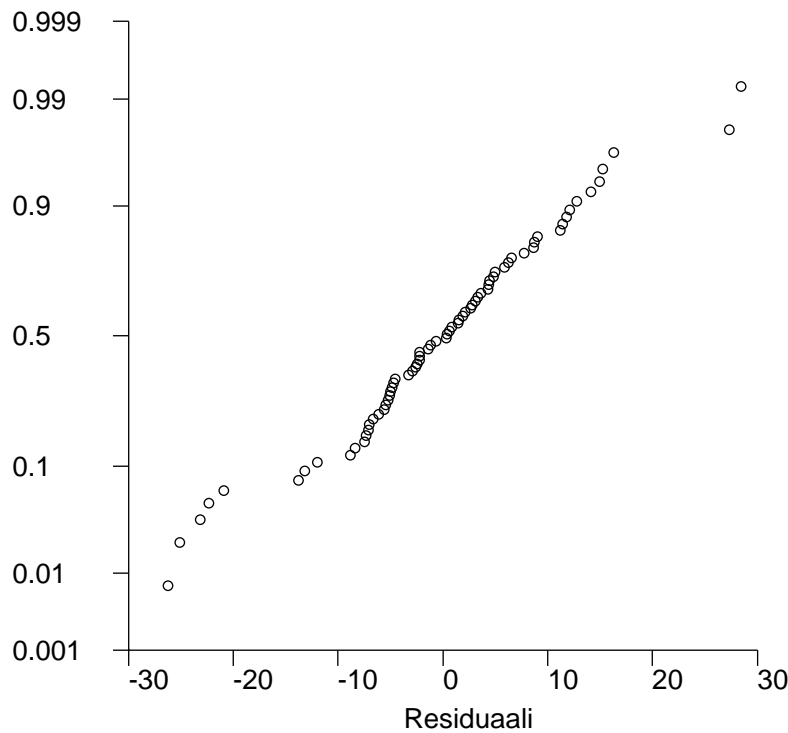
# Diagnostiikkakuva 2: sovite ja residuaali



Tässä ei saisi näkyä mitään systemaattista, mutta nyt ilmenee että mallissa on vaikeuksia. Aineiston heterogeenisuus tuo ongelmia. **Jäännösvaihtelu** ei ole satunnaista. Malli on lievästi harhainen.

Navigation icons: back, forward, search, etc.

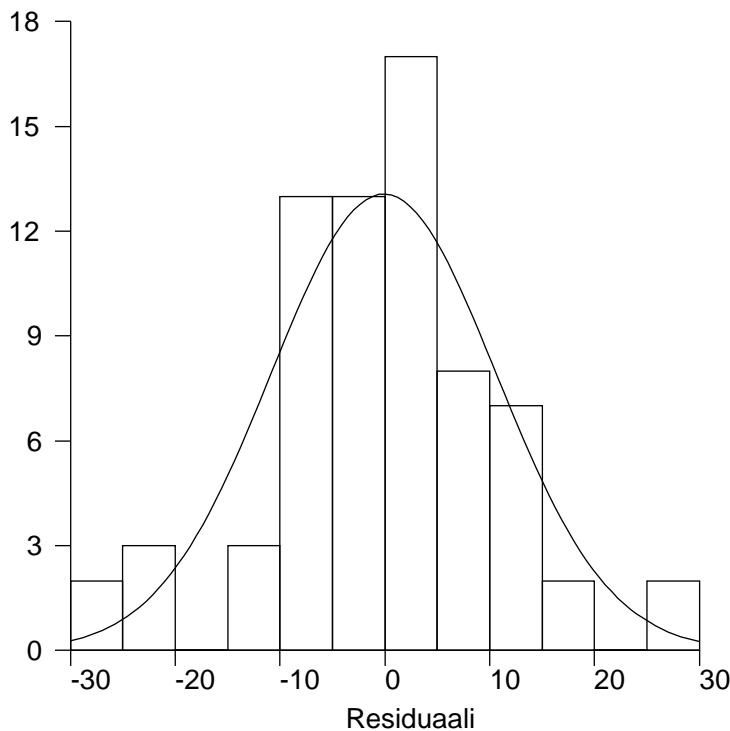
## Diagnostiikkakuva 3: residuaali todennäköisyyspaperilla



Todennäköisyyspaperikuvassa pystyakselilla ovat normaalijakauman kertymäfunktion arvot. Mitä paremmin pisteet kuvautuvat suoraksi, sitä paremmin empiirinen jakauma vastaa normaalijakaumaa.

Navigation icons: back, forward, search, etc.

## Diagnostiikkakuva 4: residuaalin histogrammi



Teema 9:stä tutussa  $\chi^2$ -yhteensopivuustestissä  $p=0.37$ , joten residuaaleja voidaan pitää normaalisti jakautuneina. Oletus mallivirheen normaalisuudesta jää siis voimaan.

Navigation icons: back, forward, search, etc.

# Esimerkki päättyy: johtopäätökset

Näin rakennettu malli on kohtuullisen hyvä, mutta lievä harhaisuus olisi hyvä saada korjattua, joko paremmilla selittäjävalinnoilla tai mahdollisilla regressiodiagnostisilla keinoilla kuten muunnoksilla.

Malliin liittyvä normaalijakaumaoletus näyttäisi pitävän paikkansa, joten  $t$ -testeihin perustuva päättely on perusteltua.

Sisällölliset tulkinnot on tässä jätetty vähemmälle huomiolle, ja saatuun malliin onkin syytä suhtautua enemmänkin teknisenä esimerkkinä regressiomallin laatimisesta.

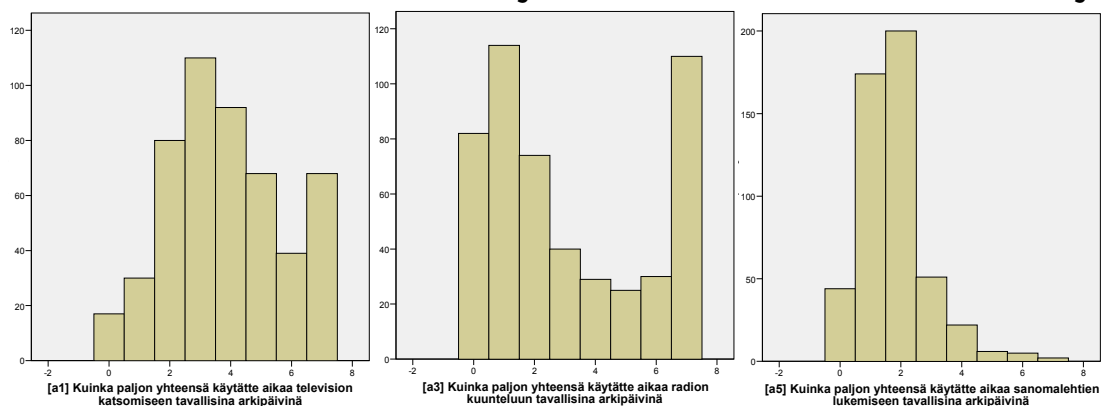
Huomaa jälleen, miten kurssilla aiemmin opitut asiat ja käsitteet (jakaumat, parametrit, estimointi, testit,  $p$ -arvot jne.) sekä erilaiset merkintätavat tulivat luontevasti käyttöön.

Tästä olisi hyvä jatkaa pidemmällekin, mutta sitä ei tämän kurssin puitteissa tehdä. Siirrytään sen sijaan toiseen esimerkkiin.

Navigation icons: back, forward, search, etc.

## Esimerkki: ESS ja mediaseuranta (SPSS)

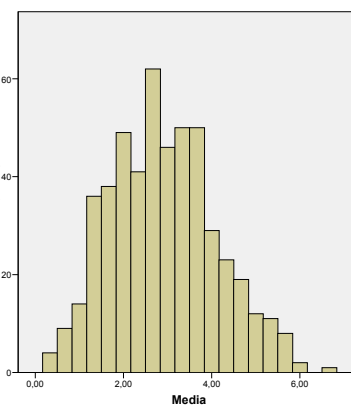
Tarkastellaan ESS-aineistoa ja siitä valittua kolmea muuttujaa:



Muodostetaan summamuuttuja Media:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
[a1] Kuinka paljon yhteensä käytätte aikaa television katsomiseen tavallisina arkipäivinä	504	0	7	3.85	1.890
[a3] Kuinka paljon yhteensä käytätte aikaa radion kuunteluun tavallisina arkipäivinä	504	0	7	3.12	2.607
[a5] Kuinka paljon yhteensä käytätte aikaa sanomalehtien lukemiseen tavallisina arkipäivinä	504	0	7	1.76	1.132
Media	504	.33	6.67	2.9101	1.19450
Valid N (listwise)	504				



Navigation icons: back, forward, search, etc.



## Esimerkki jatkuu: ESS ja mediaseuranta

Laaditaan regressiomalli, jossa selitetään mediaseurannan aktiivisuutta vastaajan sukupuolella ja iällä. Ikää käytetään mallissa jatkuvana muuttujana (2002 – syntymävuosi), ja sukupuoli on koodattu 1=mies, 2=nainen. Tulostus alkaa yhteenvedolla:

Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.327 <sup>a</sup>	.107	.103	1.13127

a. Predictors: (Constant), Sukupuoli (1=mies, 2=nainen), Ikä (vuosina)

b. Dependent Variable: Media

Selitysaste ei ole kovin kaksinen (n. 11 %). Samaa luokkaa on myös korjattu selitysaste, joka laajemmissa malleissa rankaisee liioista selittäjistä ja on siksi hieman parempi mallin hyvyyden mitta.

Selitysasteen neliöjuurta merkitään usein (myös edellä olevissa Survon tulosteissa) symbolilla R ja kutsutaan *yhteiskorrelaatiokertoimeksi*. Nimi tulee siitä, että kyseessä on  $y$ :n ja  $\hat{y}$ :n korrelaatiokerroin. Jos selittäjiä on vain yksi, R on sama kuin  $y$ :n ja  $x$ :n tavallinen korrelaatiokerroin.

Navigation icons: back, forward, search, etc.

## Esimerkki jatkuu: ESS ja mediaseuranta

Seuraava tuloslaatikko on otsikoitu lyhenteellä ANOVA, joka johtuu **varianssianalyysia** tarkoittavista sanoista ANalysis Of VAriance.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	76.538	2	38.269	29.903	.000 <sup>a</sup>
	Residual	641.162	501	1.280		
	Total	717.700	503			

Varianssianalyysi liittyy läheisesti regressioanalyysiin (nämä menetelmät ovat oikeastaan saman *lineaarisen mallin* ilmentymiä). Tässä kohtaa on kysymys siitä, miten regressioanalyysi on yleisesti onnistunut: paljonko  $y$ -muuttujan (Media) vaihtelusta selittyy mallin avulla ja paljonko jää selittämättä.

Nämä tiedot esitetään perinteisesti tässä näkyvänä *varianssitauluna*, joka koostuu kumpaakin osuutta kuvaavista neliösummista, vapausasteista ja näiden osamäärinä saatavista varianssitermeistä (Mean Square). Varianssien suhteena saadaan  $F$ -testisuure, joka testaa (*todella skeptistä!*) hypoteesia

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0,$$

siis että kaikkien selittäjien (paitsi vakion) regressiokertoimet olisivat nollia.

Testin  $p$ -arvon perusteella *ainakin yksi* kertoimista poikkeaa nolasta, ts. mallin selittäjävalinta ei ole aivan pielessä (vaikka selitysaste onkin melko vähäinen).

Navigation icons: back, forward, search, etc.



## Esimerkki jatkuu: ESS ja mediaseuranta

Kiinnostavin osa tulostuksesta on yleensä seuraava laatikko:

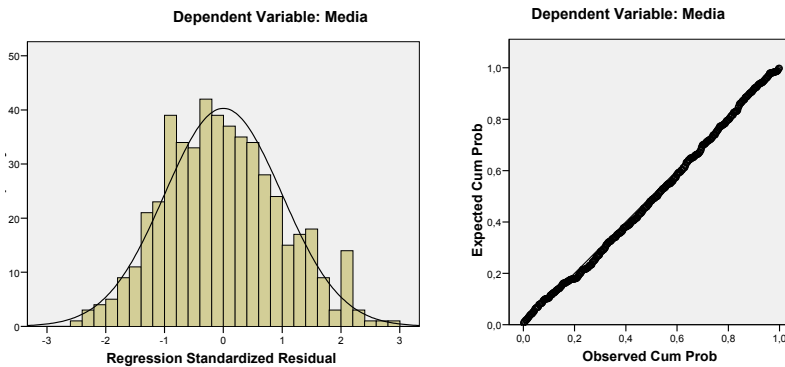
Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.270	.203		11.159	.000
	Ikä (vuosina)	.020	.003	.317	7.518	.000
	Sukupuoli (1=mies, 2=nainen)	-.190	.101	-.079	-1.878	.061

a. Dependent Variable: Media

Analyysia täydentävät kaksi diagnostista kuvaa (vrt. Survon kuvat):

Normal P-P Plot of Regression Standardized Residual



Tulkintoihin palataan tarkemmin harjoituksissa.

Navigation icons: back, forward, search, etc.

## Esimerkki jatkuu: ESS ja mediaseuranta

Äskeisen kaltaisia tilanteita on yhteiskuntatieteissä tyypillistä tutkia myös varianssianalyysilla. Luokitellaan seuraavassa ikä (karkeasti) ryhmiin 15-34 -vuotiaat, 35-54 -vuotiaat ja yli 54-vuotiaat.

Jos ikäryhmiä olisi vain kaksi, niitä voitaisiin vertailla *t*-testillä. Sen yleistys useamman ryhmän vertailuun tunnetaan nimellä **yksisuuntainen varianssianalyysi**. Tutkittavana muuttujana on edelleen mediaseurannan aktiivisuus.

Aluksi saadaan ryhmiä kuvaavia tunnuslukuja ja luottamusvälejä:

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
15-34	158	2.5865	1.07499	.08552	2.4176	2.7554
35-54	176	2.6913	1.13865	.08583	2.5219	2.8607
55-	170	3.4373	1.18463	.09086	3.2579	3.6166
Total	504	2.9101	1.19450	.05321	2.8055	3.0146

Navigation icons: back, forward, search, etc.

## Esimerkki jatkuu: ESS ja mediaseuranta

Seuraavaksi tulee varianssitaulu, joka muistuttaa hyvin paljon regressioanalyysin yhteydessä nähtyä vastaavaa esitystä. Tässä kysymys on ryhmien välisen ja ryhmien sisäisen vaihtelun suhteesta. Kiinnostuksen kohteena on nimenomaan ryhmien välinen vaihtelu, ts. miten suuria eroja ikäryhmien välillä on havaittavissa.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	72.214	2	36.107	28.025	.000
Within Groups	645.486	501	1.288		
Total	717.700	503			

F-testin perusteella eroja on selvästi havaittavissa. Tarkemmin erot paljastuvat vasta **monivertailutesteillä**:

(I) ikäryhm	(J) ikäryhm	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
15-34	35-54	-.10479	.12440	.677	-.3972	.1876
	55-	-.85076*	.12543	.000	-1.1456	-.5559
35-54	15-34	.10479	.12440	.677	-.1876	.3972
	55-	-.74597*	.12206	.000	-1.0329	-.4590
55-	15-34	.85076*	.12543	.000	.5559	1.1456
	35-54	.74597*	.12206	.000	.4590	1.0329

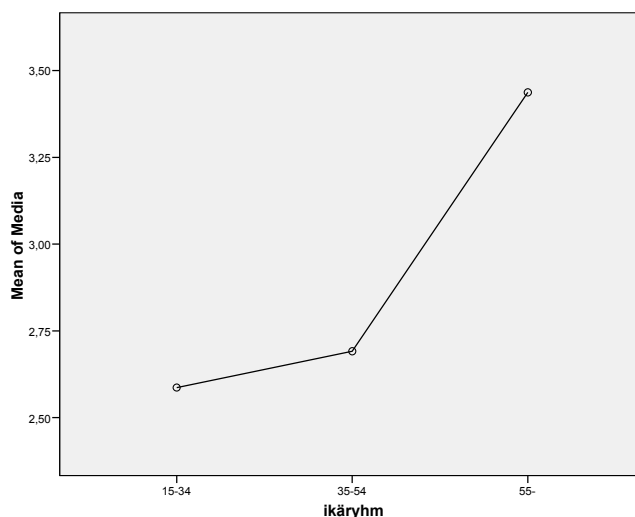
\*. The mean difference is significant at the .05 level.

Navigation icons: back, forward, search, etc.

## Esimerkki päättyy: ESS ja mediaseuranta

Varianssianalyysin (joskus hieman harhaanjohtavalta tuntuva) nimi tulee siis y-muuttujan varianssin hajottamisesta erilaisiin osiin, mutta mielenkiinto keskittyy **ryhmien keskiarvojen vertailuun**.

Tuloksia on usein tapana visualisoida **keskiarvoprofiileilla**:



Varianssianalyysilläkin voidaan tutkia useita muuttujia sekä niiden **yhdysvaikutuksia**, mutta näihin asioihin ei tässä perehdytä.

Navigation icons: back, forward, search, etc.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ≡ ▶ ≡ ▶ ↺ 🔍 ↻

$\chi^2$ -testi: odotetut ja havaitut frekvenssit

Diskreetin jakauman yhteensopivuustesteissä (nopat, lotto) verrattiin **havaittuja frekvenssejä** pistetodennäköisyyksien mukaisesti **odotettuihin frekvensseihin**. Samoin tehdään nyt.

Odotetut frekvenssit vastaavat  $H_0$ :n mukaista riippumatonta tilannetta, joten ne seuraavat aivan suoraan reunajakaumista.

Esimerkiksi taulukon ensimmäisen solun arvo saadaan laskemalla  $\frac{228 \cdot 228}{403} = 128.99$ . Ohessa luvut on pyöristetty:

	Ikä (vuotta)				
Tiedekunta	17–22	23–27	28–32	33–46	yht.
Valtiotieteellinen	129.0	67.3	18.7	13.0	<b>228</b>
Matem.–luonnontiet.	52.0	27.2	7.5	5.3	92
Muut tiedekunnat	47.0	24.5	6.8	4.7	83
yhteensä	<b>228</b>	119	33	23	<b>403</b>



## $\chi^2$ -testi: kontribuutiot päättelyn tukena

Odotettuja ja havaittuja frekvenssejä verrataan toisiinsa laskemalla niiden *neliöidyt erotukset* suhteessa odotettuihin frekvensseihin. Jokaista taulukon solua kohti saadaan kyseistä eroa kuvaava luku. Näitä lukuja kutsutaan toisinaan  $\chi^2$ -kontribuutioiksi:

	Ikä (vuotta)				
Tiedekunta	17–22	23–27	28–32	33–46	yht.
Valtiotieteellinen	0.07	0.42	0.29	0.00	0.78
Matem.–luonnontiet.	0.49	0.30	0.29	0.11	1.18
Muut tiedekunnat	0.09	0.25	<b>2.12</b>	0.11	<b>2.58</b>
yhteensä	0.65	0.97	<b>2.70</b>	0.22	4.54

Korostetut luvut kertovat suurimmat erot (vrt. taulukot).

Mukana ovat myös  $\chi^2$ -kontribuutioiden rivi- ja sarakesummat. Kokonaissumma 4.54 on  $\chi^2$ -testisuure, jonka **vapausasteet** (*degrees of freedom*) ovat  $df = (\text{rivien lkm} - 1) \times (\text{sarakkeiden lkm} - 1)$ , kun summarivejä ei oteta mukaan, siis  $(3 - 1)(4 - 1) = 6$ . Testin p-arvo saadaan  $\chi^2$ -jakauman kertymäfunktioista tai taulukosta.



## $\chi^2$ -testi: kevään 2008 vastaava taulukko

**Kevään 2008 kurssin osallistujien ikäjakauma tiedekunnittain**  
(lähde: WebOodi/KV)

	Ikä (vuotta)				
Tiedekunta	18–22	23–27	28–32	33–62	yht.
Valtiotieteellinen	86	43	6	12	147
Matem.–luonnontiet.	46	33	20	7	106
Muut tiedekunnat	35	28	5	4	72
yhteensä	167	104	31	23	325

Tässä tilanteessa testi johtaa p-arvoon 0.003, joten  $H_0$  hylätään. Ikäjakaumissa **on** eroja tiedekunnittain. Mitä nuo erot ovat?



## $\chi^2$ -testi: odotetut ja havaitut frekvenssit

Tehdään vastaavat tarkastelut kuin edellä:

Odotetut frekvenssit vastaavat  $H_0$ :n mukaista riippumatonta tilannetta, joten ne seuraavat aivan suoraan reunajakaumista. Esimerkiksi taulukon ensimmäisen solun arvo saadaan laskemalla  $\frac{167 \cdot 147}{325} = 75.5$ .

	Ikä (vuotta)				
Tiedekunta	18–22	23–27	28–32	33–62	yht.
Valtiotieteellinen	75.5	47.0	14.0	10.4	<b>147</b>
Matem.–luonnontiet.	54.5	33.9	10.1	7.5	106
Muut tiedekunnat	37.0	23.0	6.9	5.1	72
yhteensä	<b>167</b>	104	31	23	<b>325</b>



# $\chi^2$ -testi: kontribuutiot päättelyn tukena

Lasketaan  $\chi^2$ -kontribuutiot ja -testisuure:

$$\frac{(86-75.5)^2}{75.5} + \frac{(43-47)^2}{47.0} + \dots + \frac{(4-5.1)^2}{5.1} = 1.45 + 0.35 + \dots + 0.24 = 19.6$$

Kerätään luvut taulukkoon (vastaavasti kuin edellä):

Tiedekunta	Ikä (vuotta)				yht.
	18–22	23–27	28–32	33–62	
Valtiotieteellinen	1.45	0.35	<b>4.59</b>	0.25	<b>6.63</b>
Matem.–luonnontiet.	1.32	0.02	<b>9.67</b>	0.03	<b>11.05</b>
Muut tiedekunnat	0.11	1.07	0.51	0.24	1.92
yhteensä	2.87	1.44	<b>14.77</b>	0.51	19.60

Korostetut luvut kertovat, miksi  $H_0$  hylätään (vrt. taulukot).

Navigation icons: back, forward, search, etc.

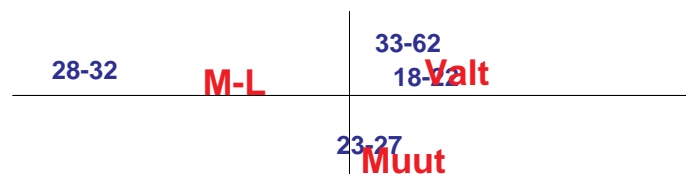
## $\chi^2$ -testin visualisointi: korrespondenssianalyysi

Taulukoiden visualisointiin erikoistunut monimuuttujamenetelmä on nimeltään **korrespondenssianalyysi**. Tarkastellaan äskeisiä taulukoita vielä sen avulla:

### Kevät 2008:

Correspondence analysis on data JK08K: Rows=3 Columns=4

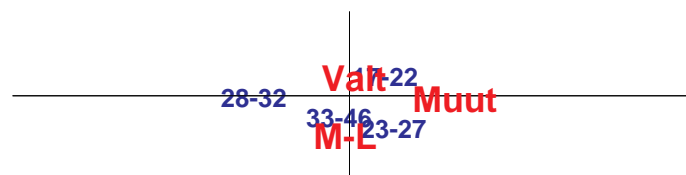
	Canonical correlation	Eigen-value	Chi^2	Cumulative percentage
1	0.2311	0.0534	17.3646554	88.61
2	0.0829	0.0069	2.23264005	100.00
		0.0603	19.5973 (df=6 P=0.00326527)	



### Syksy 2008:

Correspondence analysis on data JK08S: Rows=3 Columns=4

	Canonical correlation	Eigen-value	Chi^2	Cumulative percentage
1	0.0899	0.0081	3.25890317	71.82
2	0.0563	0.0032	1.27855915	100.00
		0.0113	4.53746 (df=6 P=0.604347)	



Lisää aiheesta: *Kyselytutkimuksen mittarit ja menetelmät* (luku 7).

Navigation icons: back, forward, search, etc.