

6. Muuttujamuunnokset

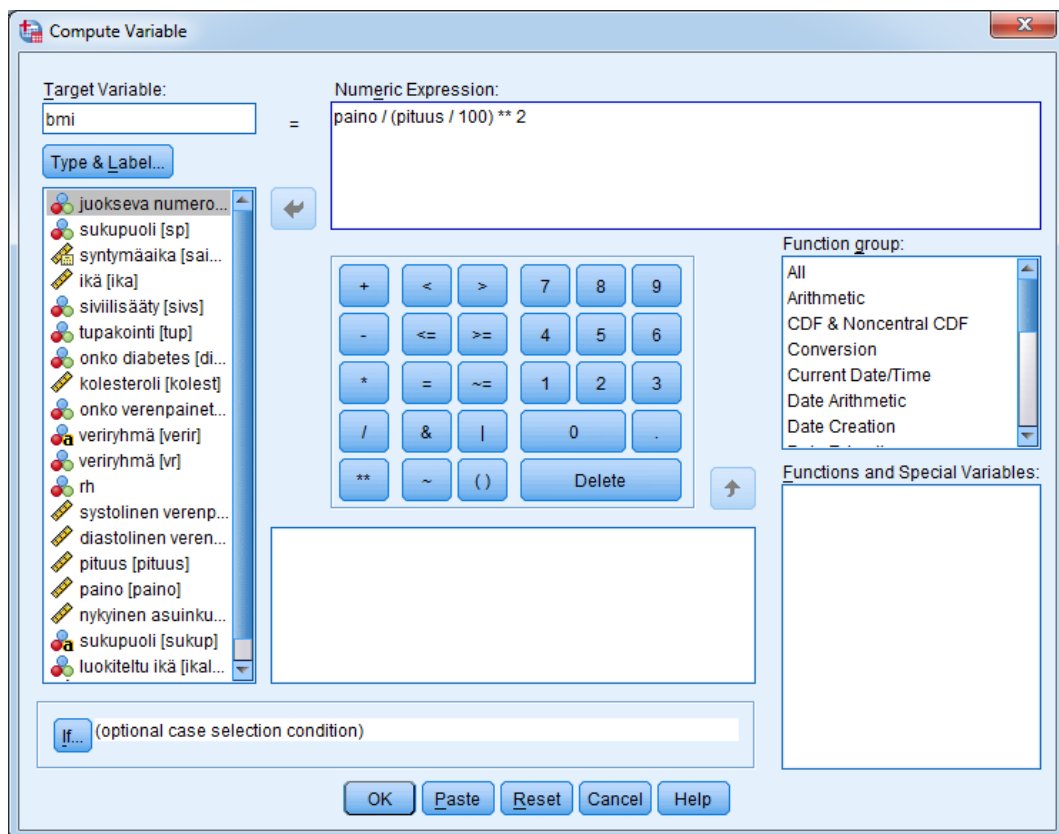
6.1. Compute-valinta

Avaa esimerkkiaineisto *sydant.sav*. Lasketaan uusi painoindeksiä kuvaava muuttuja **bmi** jo olemassa olevien muuttujien, painon ja pituuden, funktiona. Aineistossa on jo muuttuja **bindex**, joka on samalla tavalla laskettu painoindeksi. Muuttuja **bmi** tehdäänkin tässä tapauksessa vain harjoituksen vuoksi. Valitse

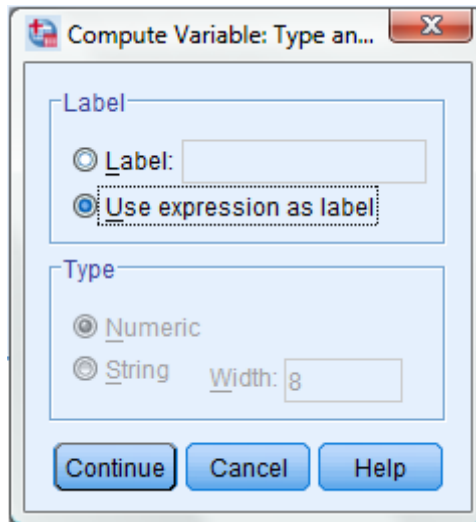
Transform → Compute Variable...

Compute Variable -ikkunassa (kuva 6.1.1) uuden muuttujan tyyppi ja kommenttiteksti määritellään **Type & Label...**-painikkeella. Voit avata *Type&Label* -ikkunan kirjoitettua uuden muuttujan nimen. Oletusarvoisesti uusi muuttuja on numeerinen. Kommentti voi olla omaisi tai funktio, jolla uusi muuttuja lasketaan. Uuden muuttujan laskemiseksi käytettävä funktio kirjoitetaan *Numeric Expression* -kohtaan. Apuna voidaan käyttää valmista muuttujalistaa, laskinnäppäimiä (taulukko 6.1.1) ja funktiolistaa (taulukko 6.2.2). Voit myös kirjoittaa kaavan suoraan näppäimistöllä, mutta ole tällöin huolellinen, että kirjoitat muuttujien nimet oikein.

Kirjoita *Compute Variable* -ikkunassa uuden muuttujan nimeksi *Target Variable* -kenttään **bmi** ja napsauta **Type & Label...**. Valitse kommentiksi *Use expression as label* (kuva 6.1.2) ja napsauta **Continue**. Kirjoita muuttujalistaa ja laskinnäppäimiä apuna käyttäen *Numeric Expression* -kohtaan laskukaavaksi $\text{paino} / (\text{pituus} / 100) ** 2$. Muuttujien nimet saat siirrettyä laskukaavaan kaksois-napsauttamalla muuttujan nimeä. Napsauta **OK**-painiketta, jolloin uusi muuttuja siirtyy aineistoikkunaan viimeiseksi muuttujaksi.



Kuva 6.1.1: Uuden muuttujan laskeminen *Compute Variable* -ikkunassa.



Kuva 6.1.2: Uuden muuttujan selitetekstin lisääminen Type&Label -ikkunassa.

Taulukko 6.1.1: Laskinpainikkeiden toiminta

Painike	Selite	Painike	Selite	Painike	Selite
+	yhteenlasku	<	pienempi kuin	&	ja
-	vähennyslasku	>	suurempi kuin		tai
*	kertolasku	<=	pienempi tai yhtä suuri kuin	~	ei
/	jakolasku	>=	suurempi tai yhtä suuri kuin	()	sulkeet
**	potenssiin korotus	=	yhtä suuri kuin		
		^	erisuuri kuin		

Kohdassa *Function group* SPSS:n funktiot on järjestetty ryhmittäin. Kun jokin funktioryhmistä valitaan, ryhmän funktiot näkyvät kohdassa *Functions and Special Variables*. Funktiosta saadaan lyhyt kuvaus *Compute*-ikkunan keskellä olevaan kenttään, kun sen valitsee hiirellä. Taulukossa 6.1.2 on esimerkkejä muutamista yleisesti käytetyistä funktioista.

Esimerkki.

Kolmen muuttujan keskiarvo voidaan laskea sekä lausekkeella

$$(\text{muuttuja1} + \text{muuttuja2} + \text{muuttuja3}) / 3$$

että funktiolla

$$\text{MEAN}(\text{muuttuja1}, \text{muuttuja2}, \text{muuttuja3}).$$

Jos yhdessäkin käsiteltävässä muuttujassa on puuttuva tieto, lauseke antaa uuden muuttujan arvoksi puuttuvan tiedon, kun taas funktio laskee keskiarvon niistä muuttujista, joista tieto on olemassa. Myös summan laskemisessa voidaan käyttää lauseketta tai funktiota SUM. Puuttuvaa tietoa käsitellään samalla tavoin kuin aiemmin.

Taulukko 6.1.2: Esimerkkejä funktioista

Funktio	Selite	Toiminta
ABS	itseisarvo	$ABS(-3,6) = 3,6$
SQRT	neliöjuuri	$SQRT(25) = 5$
LG10	10-kantainen logaritmi	$LG10(10) = 1$
LN	luonnollinen logaritmi	$LN(2,718282) = 1$
SUM	argumenttien (esim. muuttujien) summa	
MEAN	argumenttien keskiarvo	
MIN	argumenttien pienin arvo	
MAX	argumenttien suurin arvo	

6.1.1. Uuden muuttujan muodostaminen ehtolausekkeita käyttäen

Muodostetaan edelleen aineistossa *sydant.sav* riskiryhmämuuttuja **riskir**, joka saa arvon

- 0: ei riskiä, kun henkilö ei tupakoi tai on lopettanut tupakoinnin eikä hänellä ole verenpainetautia (muuttuja **tup** = 0 tai 2 ja **verpt** = 0)
 1: lievä riski, kun henkilö joko tupakoi tai hänellä on verenpainetauti (**tup** = 1 tai **verpt** = 1)
 2: vahva riski, kun henkilö sekä tupakoi että hänellä on verenpainetauti (**tup** = 1 ja **verpt** = 1)

Useasta ehdosta koostuvaa muuttujaa voidaan lähteä rakentamaan kahdella tavalla.

Tapa 1:

Valitse

Transform → Compute Variable

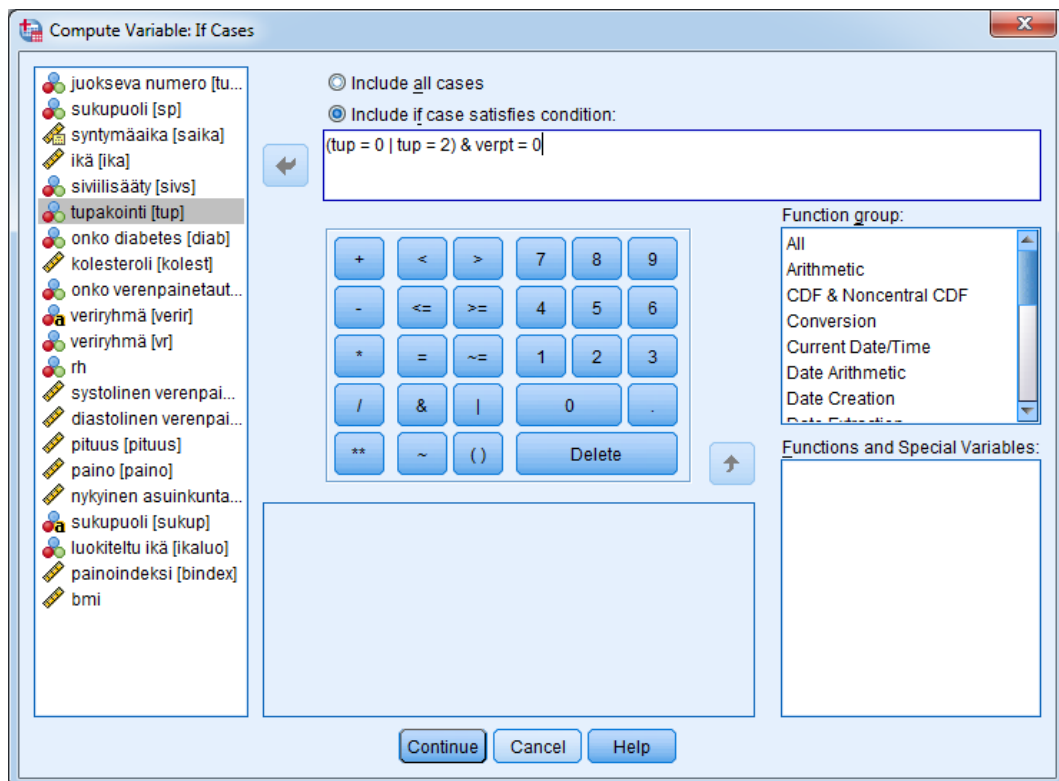
Anna uuden muuttujan nimeksi **riskir** ja kirjoita *Numeric expression* -kohtaan arvoksi 0. Napsauta -painiketta ja valitse *Include if case satisfies condition*. Kirjoita kenttään ehto (**tup=0|tup=2**) & **verpt=0** (kuva 6.1.3). Napsauta painiketta  ja vielä *Compute Variable* -ikkunassa . Aineistoikkunaan ilmestyy muuttuja **riskir**, jolla annetun ehdon täyttävillä havainnoilla on arvona 0, muilla puuttuva tieto.

Valitse nyt uudelleen

Transform → Compute Variable

Muuta *Numeric Expression* -kohdassa nollan paikalle luku 1 ja napsauta -painiketta. Kirjoita nyt *If case satisfies condition* -ehdoksi **tup=1 | verpt=1**. Napsauta  ja vielä . Vastaa kysymykseen *Change existing variable?* napsauttamalla -painiketta. Aineistoikkunaan muuttujan **riskir** arvoksi tulee yksi aina kun **tup=1** tai **verpt=1** tai molemmat ovat ykkösiä. Koska tilanne, jossa molemmat muuttujat saavat arvon 1, halutaan erottaa vahvempaa riskiä kuvaavaksi arvoksi 2, on muunnosoperaatio tehtävä vielä kerran. Toimitaan kuten edellä, mutta *Numeric Expression* -kohtaan laitetaan luku 2 ja *If case satisfies condition* -ehdoksi kirjoitetaan **tup=1 & verpt = 1**.

Huomaa, että peräkkäisiä *If case satisfies condition* -ehtoja annettaessa järjestys on tärkeä, koska myöhemmin annettu ehdollinen toimenpide toteutetaan riippumatta siitä, mitä aiemmin on tehty. Jos edellä tehdyssä **riskir**-muuttujan tapauksessa olisi ensin annettu ja-ehto (arvo 2) ja sen jälkeen tai-ehto (arvo 1), olisi muuttuja saanut vain arvoja 0 ja 1.



Kuva 6.1.3: Ehtojen asettaminen *If Cases* -ikkunassa

Tapa 2:

SPSS osaa luoda 0/1-arvoisen muuttujan, kun lausekkeeksi kirjoitetaan ehto. Tehdään esimerkkinä *Compute*-valinnan kautta uusi muuttuja **r1**. Uuden muuttujan nimeksi (*Target variable*) annetaan **r1** ja laskentakaavaksi **tup=1 | verpt=1**. Muuttuja **r1** saa arvon 0, jos ehto ei ole tosi ja arvon 1, jos ehto on tosi.

Kyseessä olevassa useamman ehdon esimerkissä ehdot ovat sisäkkäisiä eli havainto voi toteuttaa useampia ehtoja ja jälkimmäisen ehdon täytyessä ovat myös edelliset ehdot tosia. Ehtolausekkeita voidaan laskea yhteen. Muuttujan **riskir** tapauksessa oikeaan lopputulokseen päästään kirjoittamalla *Compute*-valinnassa muuttujan nimeksi **riskir** ja laskentakaavaksi **(tup=1 | verpt=1) + (tup=1 & verpt=1)**. Jos kumpikaan ehto ei ole tosi, summaksi (ja muuttujan **riskir** arvoksi) tulee 0.

Ehdot voivat olla myös toisensa poissulkevia eli havainto voi toteuttaa korkeintaan yhden ehdon. Tällaisessa tapauksessa ehtolausekkeet lasketaan yhteen ja kunkin ehdon eteen laitetaan kertoimeksi se arvo, joka muuttujalle ehdon täytyessä halutaan. Laskentakaava annetaan siis muodossa **1*(ehto1) + 2*(ehto2) + 3*(ehto3)**... Kukin ehtolauseke saa arvon 0 tai 1 ja toteutuneen ehdon edessä oleva kerroin tulee muuttujan arvoksi. Jos mikään ehdoista ei ole tosi, muuttujan arvoksi tulee 0.

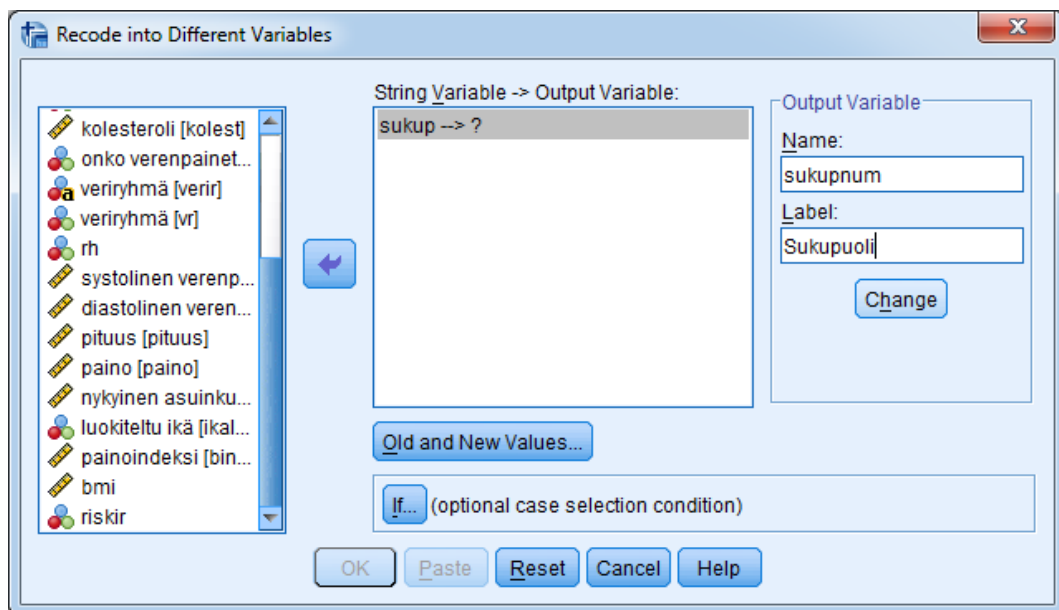
6.2. Recode-valinta

Muuttujamuunnoksia voidaan tehdä myös koodaamalla olemassa oleva muuttuja uudelleen joko samaan tai uuteen muuttujaan. Useimmiten koodaaminen kannattaa tehdä uuteen muuttujaan, koska tällöin alkuperäinen muuttuja säilyy ennallaan.

Koodataan merkkietomuuttuja **sukup** (arvoina *m* ja *n*) numeeriseksi muuttujaksi **sukupnum** siten, että *m* korvataan luvulla 1 ja *n* korvataan luvulla 2. Valitse

Transform → Recode into Different Variables...

Valitse **sukup** muuttujalistasta kohtaan *Input Variable → Output Variable*, jolloin kentän otsikoksi tulee *String Variable → Output Variable* (kuva 6.2.1). Anna *Output Variable* -kohdassa nimeksi **sukupnum**, selitteeksi Sukupuoli ja napsauta **Change**-painiketta.

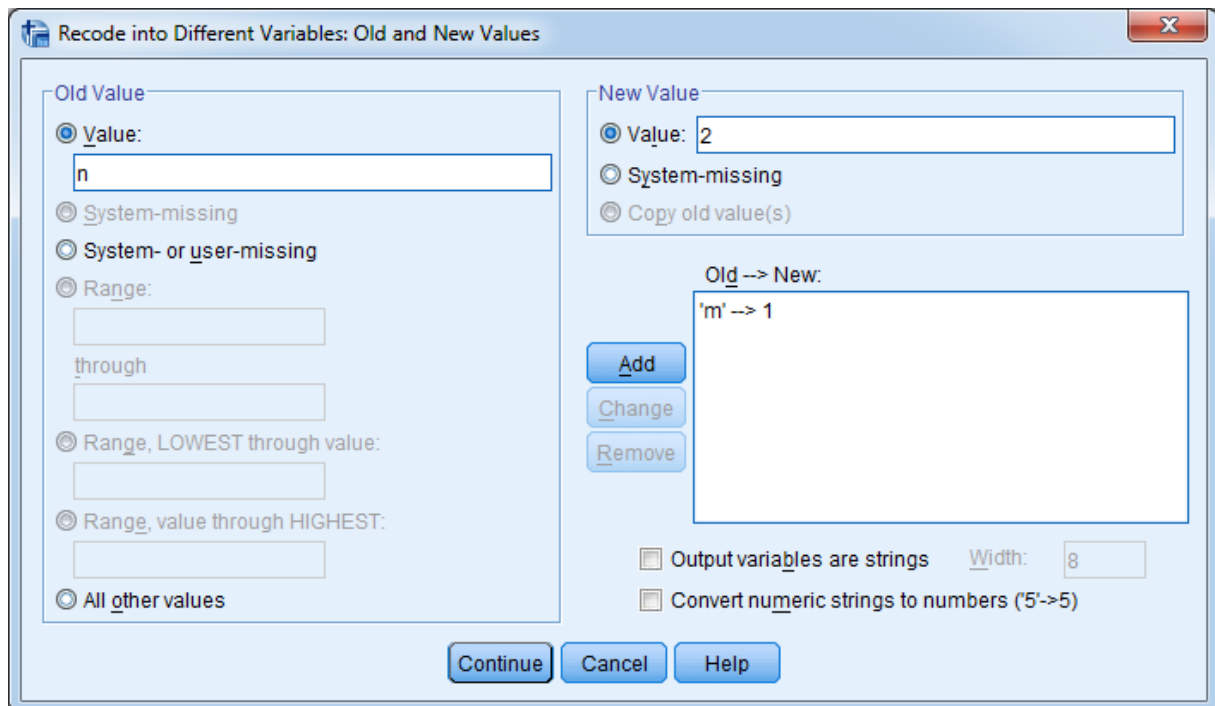


Kuva 6.2.1: Uudelleenkoodattavan muuttujan valitseminen Recode -ikkunassa.

Napsauta **Old and New Values...**-painiketta. Kirjoita vanhaksi arvoksi *Old Value* -kohtaan *m*, uudeksi arvoksi *New Value* -kohtaan 1 ja napsauta **Add**-painiketta (kuva 6.2.2). Määrittele samoin vanhalle arvolle *n* uudeksi arvoksi 2 ja napsauta **Add**-painiketta. Lopuksi napsauta **Continue** ja vielä *Recode into Different Variables* -ikkunassa **OK**, jolloin uusi muuttuja siirtyy aineisto-ikkunaan.

Old Value -kohdassa voidaan koodata uudelleen useita erilaisia arvoja taulukon 6.2.1 mukaan.

Uuden muuttujan arvoksi voidaan määritellä haluttu arvo (*Value*), puuttuva tieto (*System-missing*) tai vanhan muuttujan arvon kopio (*Copy old value(s)*). Jos uusi muuttuja on merkkietoa, valitse *Output variables are strings*. Valinnalla *Convert numeric strings to numbers* voit muuttaa vanhan muuttujan merkkietodoksi määritellyt numeroarvot numeerisiksi.



Kuva 6.2.2: Uuteen muuttujaan koodaaminen Old and New Values -ikkunassa.

Taulukko 6.2.1: Uudelleen koodattavia arvoja.

Valinta	Selite
Value	Yksittäinen arvo
System-missing	Systeemin tunnistamat puuttuvat tiedot
System- or user-missing	Systeemin tunnistamat tai käyttäjän määrittelemät puuttuvat tiedot
Range __ through __	Väli alarajasta ylärajaan
Range, LOWEST through value: __	Väli pienimmästä arvosta ylärajaan
Range, value through HIGHEST: __	Väli ylärajasta suurimpaan arvoon
All other values	Kaikki muut arvot

Luokitellaan vielä muuttuja **ika** kaksiluokkaiseksi. Valitse

Transform → Recode into Different Variables

Tyhjennä aluksi *Recode into Different Variables* -ikkuna napsauttamalla **Reset** -painiketta. Valitse sitten **ika** muuttujalistasta. Anna uuden muuttujan nimeksi **ikaluo**, kommentiksi *Luokiteltu ikä* ja napsauta **Change** -painiketta.

Napsauta **Old and New Values...** -painiketta. Valitse *Old value* -kohdassa *Range, LOWEST through value* ja kirjoita ylärajaksi 59. Anna uudeksi arvoksi 1 ja napsauta **Add**. Vastaavasti valitse *Range, value through HIGHEST*, syötä alarajaksi 60, uudeksi arvoksi 2, ja napsauta **Add**, **Continue** ja **OK**.

Huom! Kun saman SPSS-istunnon aikana tehdään useita uusia muuttujia, kannattaa vanhat määrittelyt tarvittaessa poistaa napsauttamalla *Recode into Different Variables* -ikkunan **Reset** -painiketta.