

Quantitative Business Research Methods – Day 5

Spring 2023
Sami Kajalo



Data Analysis with SPSS Software: Correlations and Regression Analysis

Product moment correlation

- The product moment correlation, r , summarises the strength of association between two metric (interval or ratio scaled) variables, say X and Y .
- It is an index used to determine whether a linear or straight-line relationship exists between X and Y .
- As it was originally proposed by Karl Pearson, it is also known as the *Pearson correlation coefficient*. It is also referred to as *simple correlation*, *bivariate correlation* or merely the *correlation coefficient*.

Correlation analysis: Example

Source: Samaha, S., Palmatier, R., & Dant, R. (2011). Poisoning Relationships: Perceived Unfairness in Channels of Distribution. *Journal of Marketing*, 75(3), 99-117.

TABLE 2
Descriptive Statistics and Correlations

Constructs	M	SD	Extracted Variance	Correlations											
				1	2	3	4	5	6	7	8	9	10	11	12
1. Channel member cooperation	4.14	.60	.77	.93											
2. Channel member flexibility	3.78	.73	.64	.51**	.84										
3. Seller opportunism	2.23	.75	.70	-.45**	-.42**	.90									
4. Perceived unfairness	2.90	.88	.72	-.47**	-.42**	.26**	.89								
5. Channel member conflict	2.10	.77	.50	-.37**	-.35**	.31**	.35**	.83							
6. Contract utilization	1.97	.80	.67	-.28**	-.25**	.49**	.08*	.22**	.89						
7. Channel member performance	3.99	.69	.84	.33**	.32**	-.16**	-.24**	-.16**	-.06*	.94					
8. Total interdependence	6.65	1.33	N.A.	.17**	.13**	-.04	-.09**	-.06	.02	.16**	N.A.				
9. Interdependence asymmetry	.97	.82	N.A.	-.02	.00	.02	.03	.01	-.01	.05	-.18**	N.A.			
10. Relationship age	6.46	6.19	N.A.	.00	-.01	.00	-.01	.05	-.04	-.05	-.04	-.03	N.A.		
11. Competitive intensity	29.65	42.30	N.A.	.01	.03	-.03	.03	-.07*	-.05	-.01	.02	.03	-.04	N.A.	
12. Firm size	1.94	2.50	N.A.	.00	.00	.00	.03	.03	-.02	-.03	-.08**	.02	.00	.04	N.A.

* $p < .05$.

** $p < .01$.

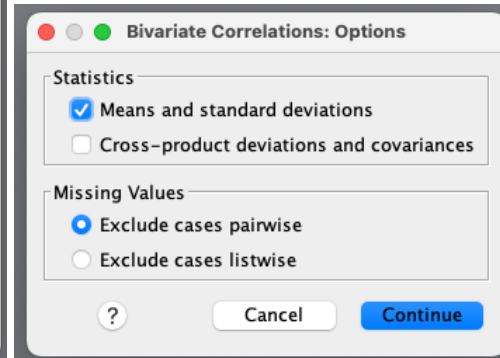
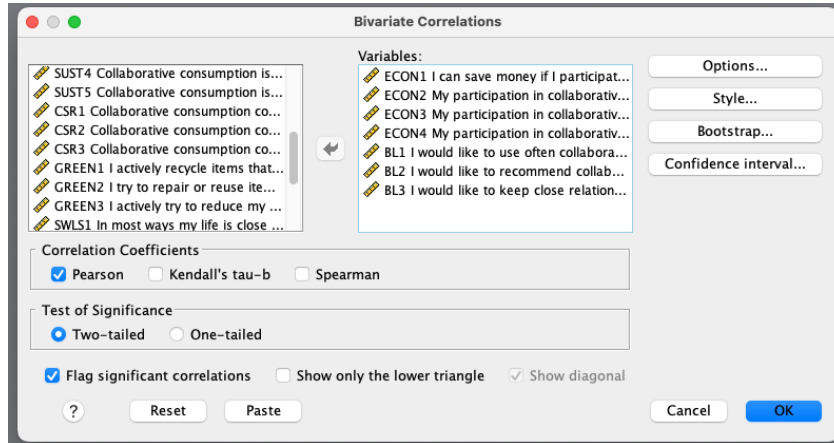
Notes: Composite reliabilities are presented along main diagonal; N.A. = not applicable. The results reported in this table are pooled results across Years 1, 2, and 3.

Often correlations are reported with other statistics. Here means (M) and standard deviations (SD) are important basic statistics. Statistical significance of correlations is here marked with stars. Extracted variances and composite reliabilities (along main diagonal) are advanced statistics used in structural equation modeling (SPSS is not able to provide them).

Performing correlation analysis with SPSS

- Analyze-Correlate-Bivariate

You can get means and standard deviation from the options (often you want to present these and the correlations together)



Pairwise - the respondent is dropped only on analyses involving variables that have missing values. This is usually better choice.

Performing correlation analysis with SPSS

Correlations

		ECON1 I can save money if I participate in collaborative consumption.	ECON2 My participation in collaborative consumption benefits me financially.	ECON3 My participation in collaborative consumption can improve my economic situation.	ECON4 My participation in collaborative consumption saves me time.	BL1 I would like to use often collaborative consumption companies' services	BL2 I would like to recommend collaborative consumption companies to others	BL3 I would like to keep close relationship collaborative consumption companies
ECON1 I can save money if I participate in collaborative consumption.	Pearson Correlation	1	.872**	.698**	.468**	.415**	.451**	.398**
	Sig. (2-tailed)		<.001	<.001	<.001	<.001	<.001	<.001
	N	200	200	200	200	200	200	200
ECON2 My participation in collaborative consumption benefits me financially.	Pearson Correlation	.872**	1	.742**	.465**	.456**	.489**	.384**
	Sig. (2-tailed)	<.001		<.001	<.001	<.001	<.001	<.001
	N	200	200	200	200	200	200	200
ECON3 My participation in collaborative consumption can improve my economic situation.	Pearson Correlation	.698**	.742**	1	.573**	.443**	.393**	.331**
	Sig. (2-tailed)	<.001	<.001		<.001	<.001	<.001	<.001
	N	200	200	200	200	200	200	200
ECON4 My participation in collaborative consumption saves me time.	Pearson Correlation	.468**	.465**	.573**	1	.383**	.424**	.327**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001	<.001	<.001
	N	200	200	200	200	200	200	200
BL1 I would like to use often collaborative consumption companies' services	Pearson Correlation	.415**	.456**	.443**	.383**	1	.732**	.704**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001		<.001	<.001
	N	200	200	200	200	200	200	200
BL2 I would like to recommend collaborative consumption companies to others	Pearson Correlation	.451**	.489**	.393**	.424**	.732**	1	.746**
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001		<.001
	N	200	200	200	200	200	200	200
BL3 I would like to keep close relationship collaborative consumption companies	Pearson Correlation	.398**	.384**	.331**	.327**	.704**	.746**	1
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001	
	N	200	200	200	200	200	200	200

** . Correlation is significant at the 0.01 level (2-tailed).

Interpretation: $r = .732$ (high positive correlation). It is statistically significant at the 0.01 level.

Regression analysis

- Regression analysis examines associative relationships between a metric-dependent variable and one or more independent variables in the following ways:
 - Determine whether the independent variables explain a significant variation in the dependent variable: whether a relationship exists.
 - Determine how much of the variation in the dependent variable can be explained by the independent variables: strength of the relationship.
 - Determine the structure or form of the relationship: the mathematical equation relating the independent and dependent variables.
- Predict the values of the dependent variable.
- Control for other independent variables when evaluating the contributions of a specific variable or set of variables.
- Regression analysis is concerned with the nature and degree of association between variables and does not imply or assume any causality.

Statistics associated with bivariate regression analysis (1 of 3)

- **Bivariate regression model.** The basic regression equation is $Y_i = \beta_0 + \beta_1 X_i + e_i$, where Y = dependent or criterion variable, X = independent or predictor variable, β_0 = intercept of the line, β_1 = slope of the line, and e_i is the error term associated with the i th observation.
- **Coefficient of determination.** The strength of association is measured by the coefficient of determination, r^2 . It varies between 0 and 1 and signifies the proportion of the total variation in Y that is accounted for by the variation in X .
- **Estimated or predicted value.** The estimated or predicted value of Y_i is $\hat{Y}_i = a + bx$, where $\hat{Y}_i$ is the predicted value of Y_i , and a and b are estimators of β_0 and β_1 , respectively.

Statistics associated with bivariate regression analysis (2 of 3)

- **Regression coefficient.** The estimated parameter, b , is usually referred to as the non-standardised regression coefficient.
- **Scattergram.** A scatter diagram, or scattergram, is a plot of the values of two variables for all the cases or observations.
- **Standard error of estimate.** This statistic, SEE , is the standard deviation of the actual Y values from the predicted $\hat{y}$ values.
- **Standard error.** The standard deviation of b , SE_b , is called the standard error.

Statistics associated with bivariate regression analysis (3 of 3)

- **Standardised regression coefficient.** Also termed the beta coefficient or beta weight, this is the slope obtained by the regression of Y on X when the data are standardised.
- **Sum of squared errors.** The distances of all the points from the regression line are squared and added together to arrive at the sum of squared errors, which is a measure of total error, $\sum e_j^2$.
- **t statistic.** A t statistic with $n - 2$ degrees of freedom can be used to test the null hypothesis that no linear relationship exists between X and Y , or:
 $H_0: \beta_1 = 0$, where $t = \frac{b}{SE_b}$

Statistics associated with multiple regression

- **Adjusted R^2 .** R^2 , coefficient of multiple determination, is adjusted for the number of independent variables and the sample size to account for the diminishing returns. After the first few variables, the additional independent variables do not make much contribution.
- **Coefficient of multiple determination.** The strength of association in multiple regression is measured by the square of the multiple correlation coefficient, R^2 , which is also called the coefficient of multiple determination.
- **F test.** The F test is used to test the null hypothesis that the coefficient of multiple determination in the population, R_{pop}^2 , is zero. This is equivalent to testing the null hypothesis. The test statistic has an F distribution with k and $(n - k - 1)$ df .

Regression analysis: Example

Profiling Victims of Investment Fraud: Mindsets and Risky Behaviors

MARGUERITE DELIEMA
DOUG SHADEL
KARLA PAK

Millions of Americans are targeted by investment scams, resulting in billions of dollars lost each year. Previous research indicates that investment fraud victims are more likely to be male, white, and married, and to have higher socioeconomic status compared to the general US population, but little research examines what behaviors and mindsets differentiate them from other investors. A telephone survey was administered to 214 investment fraud victims and 813 general investors recruited using random digit dialing. Based on the opportunity model of predatory victimization, the aim was to identify differences in investment behaviors and psychological mindsets that may affect exposure to investment scams and make individuals more attractive and susceptible targets. In addition to being older and male, victims were more materialistic than general investors and were more frequent stock traders, and purchased more investments sold through unsolicited calls, emails, television advertisements, or “free lunch” seminars, but were less likely to invest based on a social network member’s recommendation. As more retirees begin to take on managing their retirement assets, many may be tempted by unreasonable investment returns promised by unscrupulous brokers. Findings point to specific areas where investor education is needed to counteract poor investment decision-making and risky mindsets.

Keywords: opportunity model of predatory victimization, materialism, investment scam, remote investing

Regression analysis: Example

Logistic Regression Analysis. Dependent variable: person has been (or has not been) a victim of investment fraud.

JOURNAL OF CONSUMER RESEARCH

7

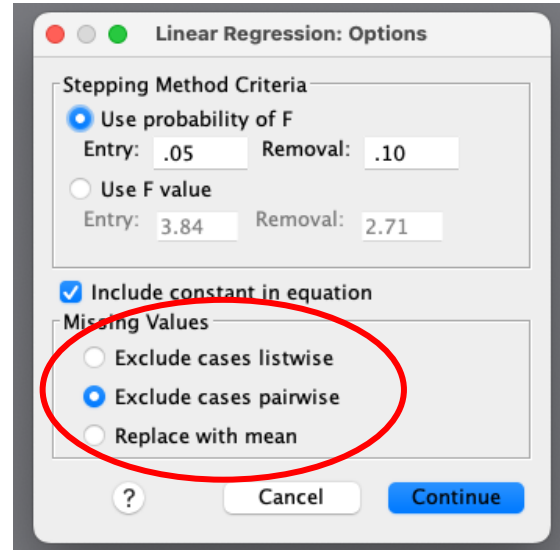
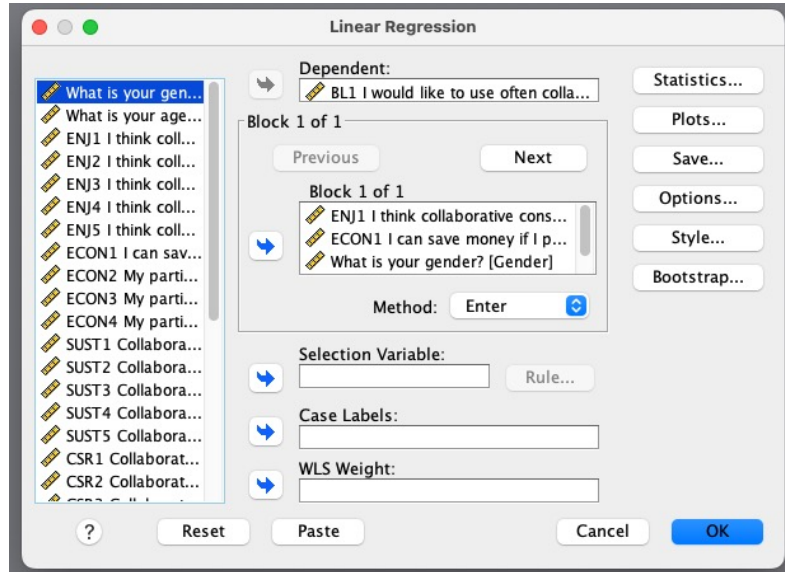
TABLE 2

FACTORS ASSOCIATED WITH INVESTMENT FRAUD VICTIMIZATION (KNOWN VICTIM = 1; N = 1,027).

	Model 1: Sociodemographic characteristics			Model 2: Behaviors			Model 3: Mindsets		
	Odds ratio (95% confidence interval) <i>p</i> -value								
<i>Sociodemographic characteristics</i>									
Age	1.04	(1.03, 1.06)	.000	1.04	(1.02, 1.06)	.000	1.04	(1.02, 1.06)	.000
Sex (male=1)	5.50	(3.76, 8.04)	.000	3.64	(2.33, 5.68)	.000	3.89	(2.31, 6.55)	.000
Non-Hispanic white	1.38	(0.81, 2.35)	.233	1.39	(0.68, 2.86)	.370	1.51	(0.69, 3.29)	.306
Married/partnered	1.22	(0.85, 1.80)	.318	1.13	(0.72, 1.78)	.605	1.28	(0.76, 2.17)	.359
High school or less (ref)	—	—	—	—	—	—	—	—	—
Some college	0.78	(0.46, 1.35)	.379	0.67	(0.35, 1.26)	.209	0.67	(0.34, 1.34)	.261
College or more	1.05	(0.65, 1.71)	.839	0.82	(0.46, 1.45)	.483	0.89	(0.48, 1.66)	.717
Income ≤ \$35,000 (ref)	—	—	—	—	—	—	—	—	—
Income \$35,000–74,999	0.87	(0.49, 1.55)	.631	0.83	(0.41, 1.67)	.596	0.67	(0.31, 1.44)	.302
Income \$75,000–124,999	1.33	(0.71, 2.49)	.381	1.27	(0.60, 2.69)	.539	1.00	(0.44, 2.30)	.997
Income ≥ \$125,000	1.10	(0.54, 2.21)	.800	0.72	(0.31, 1.69)	.450	0.56	(0.22, 1.42)	.222
<i>Exposure to offenders/risk behaviors</i>									
Frequency of solicitations				1.10	(1.02, 1.19)	.021	1.08	(0.99, 1.18)	.075
Trading frequency				1.56	(1.36, 1.77)	.000	1.63	(1.38, 1.94)	.000
Remote investments				1.64	(1.23, 2.19)	.001	1.51	(1.08, 2.10)	.015
Invested based on recommendation				0.70	(0.51, 0.97)	.032	0.68	(0.47, 0.98)	.036
<i>Target suitability/investment mindset</i>									
Materialism							2.25	(1.01, 5.02)	.047
Interest in new investment opportunities							0.62	(0.29, 1.35)	.229
Favorable attitudes toward unregulated investments							1.27	(1.01, 1.59)	.043

Performing regression analysis with SPSS

- Analyze-Regression-Linear



Pairwise - the respondent is dropped only on analyses involving variables that have missing values. This is usually better choice.

Performing regression analysis with SPSS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.652 ^a	.425	.413	1.05265

a. Predictors: (Constant), SUST1 Collaborative consumption helps save natural resources. , What is your gender?, ECON1 I can save money if I participate in collaborative consumption. , ENJ1 I think collaborative consumption is enjoyable.

R Square (Coefficient of multiple determination) and Adjusted R Square: the model explains 41.3%-42.5% of the variance. Both statistics are reported. Adjusted R Square takes into account the number of independent variables and is lower than R Square.

Performing regression analysis with SPSS

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	159.848	4	39.962	36.065	<.001 ^b
	Residual	216.072	195	1.108		
	Total	375.920	199			

a. Dependent Variable: BL1 I would like to use often collaborative consumption companies' services

b. Predictors: (Constant), SUST1 Collaborative consumption helps save natural resources. , What is your gender?, ECON1 I can save money if I participate in collaborative consumption. , ENJ1 I think collaborative consumption is enjoyable.

F-test reports the statistical significance of the model. Both F and Sig. are reported. This model is significant as Sig.<0.05.

Performing regression analysis with SPSS

B = regression coefficients.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.480	.469		1.022	.308
	ENJ1 I think collaborative consumption is enjoyable.	.511	.075	.458	6.839	<.001
	ECON1 I can save money if I participate in collaborative consumption.	.080	.070	.075	1.137	.257
	What is your gender?	.024	.151	.009	.161	.872
	SUST1 Collaborative consumption helps save natural resources.	.256	.072	.234	3.566	<.001

a. Dependent Variable: BL1 I would like to use often collaborative consumption companies' services

t and Sig. show whether the variable is significant (<.05 means significant). Here, ENJ1 and SUST1 are significant and explain BL1. From Beta column we see that ENJ1 has bigger impact than SUST1.

Structural Equation Modeling

- Structural Equation Modeling (SEM) is a family of statistical techniques which incorporates and integrates factor analysis and path analysis (=regression analysis).
- Use of SEM software for a model in which each variable has only one indicator is a type of path analysis (=regression analysis).
- Use of SEM software for a model in which each variable has multiple indicators but there are no direct effects connecting the variables is a type of factor analysis.
- Usually SEM refers to a hybrid model with both multiple indicators for each variable (called latent variables or factors), and paths connecting the latent variables.

Structural Equation Modeling: key concepts and statistics

- Confirmatory factor analysis (CFA). A technique used to estimate the measurement model. It seeks to confirm if the number of factors (or constructs) and the loadings of observed (indicator) variables on them conform to what is expected on the basis of theory.
- Construct In SEM, a construct is a latent or unobservable concept that can be defined conceptually but that cannot be measured directly or without error. Also called a factor, a construct is measured by multiple indicators or observed variables.
- Measurement model The first of two models estimated in SEM. It represents the theory that specifies the observed variables for each construct and permits the assessment of construct validity.
- Structural model The second of two models estimated in SEM. It represents the theory that specifies how the constructs are related to each other, often with multiple dependence relationships.

Research Ethics

The importance of research ethics

- Research ethics is not a 'step' in the research process. It is an integral part of the research process that should be applied throughout.
- If participants feel that they or their views are being misused or misrepresented, they either will not take part in future studies or may do so without honesty or full engagement in the issues being researched.
- Ethics codes are a core part of professional practice and supported by key market research associations.

The Ethics Codes of Market Research Society (MRS)

1. Researchers shall ensure that participation in their activities is based on voluntary informed consent
2. Researchers shall be straightforward and honest in all their professional and business relationships
3. Researchers shall be transparent as to the subject and purpose of data collection
4. Researchers shall respect the confidentiality of information collected in their professional activities
5. Researchers shall respect the rights and well-being of all individuals
6. Researchers shall ensure that participants are not harmed or adversely affected by their professional activities
7. Researchers shall balance the needs of individuals, clients, and their professional activities
8. Researchers shall exercise independent professional judgement in the design, conduct and reporting of their professional activities
9. Researchers shall ensure that their professional activities are conducted by persons with appropriate training, qualifications and experience
10. Researchers shall protect the reputation and integrity of the profession

The Ethics Codes of the European Society for Opinion and Marketing Research (ESOMAR)

1. Market researchers shall conform to all relevant national and international laws
2. Market researchers shall behave ethically and shall not do anything which might damage the reputation of market research
3. Market researchers shall take special care when carrying out research among children and young people
4. Respondents' cooperation is voluntary and must be based on adequate, and not misleading, information about the general purpose and nature of the project when their agreement to participate is being obtained and all such statements shall be honoured
5. The rights of respondents as private individuals shall be respected by market researchers and they shall not be harmed or adversely affected as the direct result of cooperating in a market research project
6. Market researchers shall never allow personal data they collect in a market research project to be used for any purpose other than market research
7. Market researchers shall ensure that projects and activities are designed, carried out, reported and documented accurately, transparently and objectively
8. Market researchers shall conform to the accepted principles of fair competition

Informed consent (person's voluntary well-informed decision to share information)

- Informed consent is at the core of marketing research ethics.
- However, with the growing use of secondary data it is becoming harder to ensure that informed consent is being gathered for all the data sources we analyse.
- For example, can we be sure that we have gathered informed consent when using data gathered with social media APIs? (API=application programming interface; process where multiple applications share data).
- Some researchers have argued that with new technology platforms, a lot of the time we are dealing with uninformed consent.

Anonymity

- Anonymisation is rooted in the defining principle of research ethics: that participants in research should not be harmed as a result of participation.
- Data collected during a research process could, if gathered in the wrong hands, cause harm to respondents by making public information that was not designed to be. Anonymisation protects respondents against this outcome.
- **Always secure the anonymity of the respondents!**

Quantitative or Qualitative Master's Thesis

Putting the pieces together

- Searching literature & doing literature review -> ability to learn anything
- Mendeley, Dropbox, VPN, Evernote etc. -> getting organized
- Webropol & SPSS -> data collection & analysis
- .. and now we put the pieces together for Master's thesis
- First we have to know whether we want to use qualitative or quantitative data in the master's thesis.
- Next we discuss how to choose the topic for master's thesis.



Quantitative or qualitative Master's thesis?

- On the beginning of the course we studied what quantitative master's thesis are like.
- As similar excellent examples of qualitative master's thesis I recommend checking the following two (both available at Aalto library's website):
 - Kerkelä, Iina: Access-Based Consumption of Clothes: Ownership, Attachment, and Identity
 - Helaniemi, Markus: Pursuit of happiness: qualitative study of Finnish craft beer hobbyists' consumer identity

Choosing the Topic – Impactful and Interesting Research

Impactful and interesting research

- Very ambitious target: prove something generally believed wrong.
- “What matters is whether the research runs counter to what is currently assumed as true. That’s interesting. That’s impactful.” (Tellis 2017)
- Unfortunately, it is not easy to find some relevant phenomena and then prove that what has been found out about is wrong.
- Enough ambitious target: add knowledge to some field.
 - Studies about mechanism how things have impact on each other (often mediation).
 - Studies about how some things strengthen or weaken impacts (often moderation).
 - Sometimes (not very recommended): Studies about different industries, countries, customer groups etc.

Theory + Context/Data + Method = Excellent research

- Providing scientific contribution can be simple and easy. We just need to combine following elements:
 - Theory. These are often general views that explain how the world works. For example, Unified Theory of Acceptance, and Use of Technology (UTAUT). Theory is related to theoretical concepts, such as Brand Loyalty.
 - Context and data. For example Sharing Economy (Uber, AirBnB).
 - Method. For example a survey among consumers. And analysis by advanced methods, for exaple Structural Equation Modeling.
- If we combine these things in an innovative way we are able to create new knowledge and scientific contribution.

**Thank you for your
participation in the course!**

Please answer the course feedback survey.

All feedback and ideas are greatly appreciated.

sami.kajalo@aalto.fi