Introduction to Quantitative Research Monday 17 Nov - Monday 24 Nov 2014

Pertti Jokivuori, Ph.D Lecturer

# Wikipedia:

In the social sciences, **quantitative research** refers to the **systematic empirical investigation** of social phenomena via **statistical** techniques.

The process of **measurement** is central to quantitative research

*🡪* it provides the fundamental connection between empirical observation and statistical expression of *quantitative* relationships.



Quantitative research is generally made using scientific methods, which can include:

* The generation of models, theories and hypotheses
* The development of instruments and methods for measurement
* Experimental control and manipulation of variables
* Collection of empirical data
* Modeling and analysis of data
* Evaluation of results

Sociology can be broadly defined as the study of people and institutions within society. Sociologists from every field utilize similar data-collection techniques when conducting research.

These research techniques include group observations, census/sample taking, interviewing, distributing surveys, and so on. Each of these methods will produce results that can be analyzed either qualitatively or quantitatively.

### http://www.greenbook.org/Content/DataStar/tick-box.jpg

### What is Quantitative Research?

The word "quantitative" expresses data that is subject to **measurements** and **numbers**. Quantitative methods of research in sociology employ data that is **objective** and **statistical**. This type of research generates information with trends and correlations.

Quantitative information is similar to scientific-style data; where objects are analyzed in exact descriptions, often utilizing logic and reasoning.

### http://courses.tlt.psu.edu/course/bio12/module03/population%20sample.png

### Significant Differences between Qualitative and Quantitative

* When employing quantitative research methods, the **researcher usually already knows what is expected** from the results. On the other hand, researchers that rely on qualitative techniques cannot expect results similar to their hypothesis.
* Qualitative information is subjective and open to interpretation. **Quantitative information is more objective**; expressed in exact descriptions such as statistics, percentages and numbers.
* Qualitative research methods portray what the social world has in store for people of various backgrounds and living situations. This type of research communicates results through detailed observation and description.
* Conversely; **quantitative research is a reflection of people and situations from a specific point of view**, such as the rate of divorce amongst parents of suicidal teenagers, organisational commitment amongst female and male employees.

*Quantitative research consists of those studies in which the data concerned can be analysed in terms of numbers ... Research can also be qualitative, that is, it can describe events, persons and so forth scientifically without the use of numerical data ... Quantitative research is based more directly on its original plans and its results are more readily analysed and interpreted. Qualitative research is more open and responsive to its subject. Both types of research are valid and useful. They are not mutually exclusive. It is possible for a single investigation to use both methods. (Best and Khan, 1989: 89-90) 🡪 MIXED METHOD*

*Quantitative research is empirical research where the data* ***are in the form of numbers****.*

*Qualitative research is empirical research where the data* ***are not in the form of numbers****. (Punch, 1998: 4)*



***STRENGTHS of Quantitative research:***

· Precision and transparency - through quantitative and reliable measurement

· Control - through sampling and design

· Ability to produce causality statements, through the use of controlled experiments

· Statistical techniques allow for sophisticated analyses

· Replicable

# *Quantitative research is qualitatative*!!!

Social scientist is not interested in numbers or coefficients in itself (%-distributions, alfa, eta and beta values or correlation coeffficients), but he/she is interested in social phenomena (connections between variables refer/represent connections between social phenomena)

🡪 Social reality is a research subject that a researcher is interested in!



***Key elements of Quantitative research****:*

* ***Observational units*** *(human being (survey), country, organisation, voluntary association…)*
* ***Variables*** *(quality)*
* ***Values*** *(all qualities have numeric symbols)*

*♦ Quantitative research means that all variables (qualities) have been expressed in numerical way.*

Observational matrix:



SPSS = **Statistical** **Package** for the **Social** **Sciences**

**Analyse** 🡪 descriptive statistics: frequencies, crosstabs, compare means, anova, general linear model, correlate, regression, classify (discriminant), data reduction (factor), scale (reliability analysis).

**Transformations** 🡪 recode ja Transform > compute

**Data** 🡪 split file and select cases

**Graphs** 🡪 bar, line, pie, scatter, boxplot

Three Levels of Quantitative Analysis:

•**Descriptive** level of analysis (frequencies, graphs)

• The **association** (or independence/dependence) between variables (variables-oriented study)

• **Explanatory** analysis (elaboration, multivariate models)

1. Analysing a single variable: Tables, distributions and graphs

2. Analysing two variables: Cross tabulations

3. Analysing two variables: Compare means, One-way ANOVA and linear relationships between variables (correlation)

4. Generalising beyond a sample: Statistical significance

5. Multivariate methods

1. Analysing a single variable: Tables, distributions and graphs

What kind of contracts Finnish employees have?

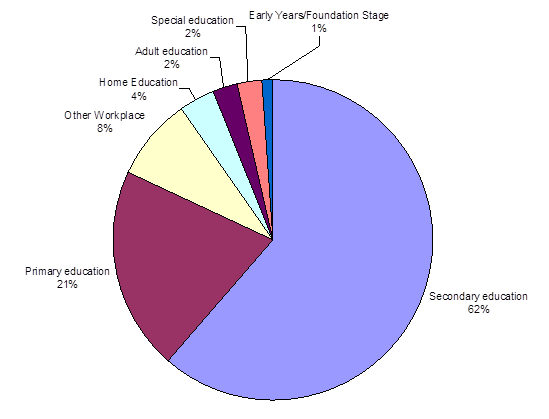
****

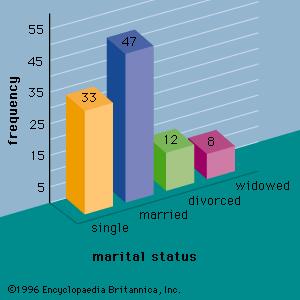
**Permanent =** employment contract made for the time being

****

Graphs 🡪 bar (X9)

Examples of descriptive analysis:

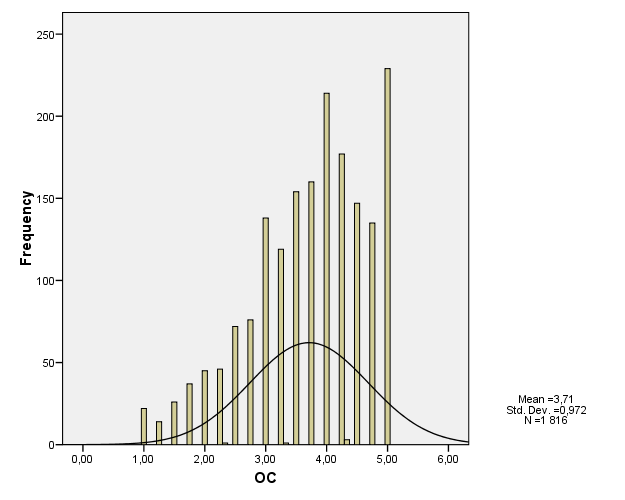




|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| [c1] How happy would you say you are? (0-10) | 1877 | 0 | 10 | 7,96 | 1,430 |
| Valid N (listwise) | 1877 |  |  |  |  |

Distribution of a Statement:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Belonging to this organisation has a great deal  of personal meaning for me | Totally agree | Partially agree | Not agree, not disagree or can’t say | Partially  disagree | Totally  Disagree |
|  | 25% | 38% | 22% | 11% | 4% |



2. Analysing two variables: Cross tabulations

**Type of contract by Gender:**

****

****

**Statistical tests are used in quantitative analyses to show that is it possible to generalize results in data (differences between groups, connections between variables) to cover the whole population too.**

**Chi-square (independence-test): association between proportions**

**Significance levels:**

p< .05 statistically almost significant (950/1000) 🡪 5 % likelihood that a result is accurate just in sample

p< .01 statistically significant (990/1000) 🡪 1% likelihood that a result is accurate just in sample

p< .001 statistically highly significant (999/1000) 🡪 0.1 % likelihood that a result is accurate just in sample

Example 2. Type of contract by Generations:

****

****

Example 3 (3 varibles). Type of contract by Generations by Gender Crosstabulation:

****

****

Quantitative Research - Operationalization and Sum variables

|  |  |
| --- | --- |
| [Charles T. Schmidt, Jr. Labor Research Center](http://www.uri.edu/research/lrc/) | [Professor Richard W. Scholl](http://www.uri.edu/research/lrc/scholl/index.html) |
| Dilbert.com Organizational Commitment | |

What is organisational commitment (OC)?

Organizational commitment is one of those concepts that is used in a number of different ways. In most cases, we use the term to refer to a type of employee, that is, an employee with high organizational commitment. In this case, we generally refer to three observable patterns of behavior. When we see an employee exhibiting these patterns of behavior, we attribute these patterns to something we call commitment and we say that individual is committed. More recently, these behaviors have been termed organizational citizenship behaviors. The three components of commitment are:

**Social identification** with the organization's goals and/or mission manifested in pride in and defense of the organization.

**Long-term membership** in the organization and intention to remain with the organization, often termed loyalty

High levels of **extra role behavior**, that is, behavior beyond required performance- Often referred to as pro-social behavior.

Sum variables

Operationazation of the concept

Example: **organisational commitment**

All statements must be formulated colloquially!

**The statements indicating organisational commitment:**

* I would be very happy to spend the rest of my career in this organisation
* I would rather be working somewhere else than in my current organisation (reversed)
* I don’t feel emotionally attached to this organisation (reversed)
* Belonging to this organisation has a great deal of personal meaning

for me

\* Organisational commitment-indicator is measured within four items/statements. Each of those statements has alternatives for answer on a 5-point Likert-scale. Organisational commitment items refer to employees’ emotional attachment to, identification with and involvement in an organisation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Commitment items | Totally agree  % | Partially agree  % | Not agree, not disagree or can’t say  % | Partially  disagree  % | Totally  disagree  % |
| Organisational commitment I would be very happy to spend the rest of my career in this organisation  I would rather be working somewhere else than in my current organisation (reversed)  I don’t feel emotionally attached to this organisation  (reversed)  Belonging to this organisation has a great deal of personal meaning for me | 32  6  4  25 | 29  11  12  38 | 21  27  17  22 | 10  27  33  11 | 8  29  36  4 |

Steps:

1. Recode variables (Transforms 🡪 Recode into different variables)
2. Compute variable (Transform 🡪 Compute)

Sum variable 🡪 all items that indicates same factor/ in one indicator (totally disagree=1…totally agree=5)

OC=Mean(OC1,OC2,OC3,OC4)

Cronbach's alpha (internal consistency reliability) of the scale:

* Cronbach's **alfa** is a measure of consistency of the indicator/scale; the alfa is the fraction of this total variance that can be explained by a single trait that people have to a different amount.  This unique trait is the skill that the test is designed to measure.  The alfa ranges from 0 to 1, and higher is better (Alfa should be at least 0.5/0.6).

Analyses 🡪 scale 🡪 reliability analyses 🡪 items: OC1, OC2, OC3 and OC4

α = .86 🡪 OK

Decriptive statistics of OC-indicator:

****

Histogram and normal curve:



What does the mean of OC and the normal curve of OC tell about the commitment rate of Finnish employees?



**Analysing two variables: Compare means (one-way anova) and linear relationships between variables (correlation)**

How OC varies in generations?

### One-way anova

****

****

****



**Correlation:**

Correlation means linear relationship between two variables. Variables should be so-called continual variables (min --------------------- max)

* Is there a positive correlation (dual commitment) or negative correlation (cognitive dissonance) between organizational commitment (OC) and trade union commitment (TC)? How organisational commitment is related to union commitment (dual commitment or cognitive dissonance)?

**Union Commitment:**

* My union and I have approximately the same values (X31\_1)
* The decisions made by my union usually reflect my own opinions (x31\_2)
* My union’s problems are my problems (X31\_3)
* My union means a great deal to me personally (X31\_4)

Trade union commitment-indicator is measured within four items. Each of those questions has alternatives for answer on a 5-point scale. Union commitment items refer to employees’ emotional attachment to, identification with and involvement in an trade union.

First: Recode -> into different variables

1->5

2->4

3->3

4->2

5->1

9->3

TC1, TC2, TC3 and TC4

Then, Transform -> compute

TC= mean(TC1,TC2,TC3,TC4)

****

Is there a positive correlation (dual commitment) or negative correlation (cognitive dissonance) between organizational commitment (OC) and trade union commitment (TC)?

****

**Scatter, correlation diagram:**

****

Elaboration

Explanatory analysis (elaboration, multivariate models)

**Partial correlation** is the correlation of two variables while controlling for a third or more other variables. The technique is commonly used in "causal" modeling of small models.

Example:

****

We know that there are positive correlations between age and organisational commitment and between age and trade union commitment:

****

(syntymävuosi = the year of birth which is vice versa age)

Now, we have to study is the connection between OC and UC true or so-called spurious connection (or false connection) which is produced by age, because the older employees are more committed to their organisation and trade union.

The reliability of **dual commitment** could be judged by the probability that the connection between OC and UC is spurious, resulting from the age of employees rather than from a true relationship between variables.

Partial correlation is a method used to describe the relationship between two variables whilst taking away the effects of another variable, or several other variables, on this relationship.

Partial correlation is a correlation between two variables that remains after controlling for (e.g., partialling out) one or more other variables.

Analyse 🡪 correlation partial (OC and UC into the list of variables, and x3 into the list of controlling for.

PARTIAL CORR

/VARIABLES= OC TC BY x3

/SIGNIFICANCE=TWOTAIL

/MISSING=LISTWISE.



Thus, we can see that correlation between OC and UC decreased when it is analysed by using age as control variable. BUT, the age explains just a small part of connection, and it remains still at the level of statistically highly significance.

So, we’ll add more controlling variables… for instance: gender (x2), type of contract (x9)

PARTIAL CORR

/VARIABLES= OC TC BY x3 x2 x9

/SIGNIFICANCE=TWOTAIL

/MISSING=LISTWISE.



So, despite the fact of number of the added controlling variables the correlation between OC and UC remains at the level statistically highly significance (r=0.17\*\*\*) which means that it is true connection! Only a small part of original correlation is explained by age.

Multivariate analyses: Linear Regression

General Purpose of Multiple Regression Analyses

The general purpose of multiple regression is to learn more about the relationship between several independent or predictor variables and one dependent variable.

**Unique Prediction and Partial Correlation.** Note that in this equation, the regression coefficients (or *B* coefficients) represent the *independent* contributions of each independent variable to the prediction of the dependent variable. Another way to express this fact is to say that, for example, variable *X1* is correlated with the *Y* variable, after controlling for all other independent variables.

This type of correlation is also referred to as a *partial correlation* (which was our topic last lecture).

**Partial correlation coefficient**: Value that measures the relationship between the independent variable and the dependent variable when the other independent variables in the equation are held constant.

Perhaps the following **example** will clarify this issue:

One would probably find a significant negative correlation between hair length and height in the population (i.e., short people have longer hair). At first this may seem odd; however, if we were to add the variable *Gender* into the multiple regression equation, this correlation would probably disappear. This is because women, on the average, have longer hair than men; they also are shorter on the average than men. Thus, after we remove this gender difference by entering *Gender* into the equation, the relationship between hair length and height disappears because hair length does *not* make any unique contribution to the prediction of height, above and beyond what it shares in the prediction with variable *Gender*. Put another way, after controlling for the variable *Gender*, the partial correlation between hair length and height is zero.

The example. Predictors of organisational commitment

Dependent variable = OC (organisational commitment)

Independent variables/predictors

x2 (Gender; female/male)

x3 (age/year of birth 34-82)

x9 (type of contract; permanent/temporary)

x15a (changes in salary; 🡪weaken)

x15d (changes in work interesting; 🡪weaken)

x16b (uncertainty of employment🡪weaken)

x21 (climate of communications🡪weaken)

x25d (conflicts between personnel groups 🡪weaken)

**Scales:**

**15a; 1=became better … 5 =weakened**

**15d; 1=became better … 5 =weakened**

**16a; threat of dismissal 1=probable … 4 =unlikely**

**21; 1=good … 3=bad**

**25d; 1=lots of conflicts … 4=not at all**

**Analyze 🡪 Regression…Linear**

**Dependent = OC**

Independent variables

x2 (Gender)

x3 (age/year of birth)

x9 (type of contract)

x15a (changes in salary)

x15d (changes in work interesting)

x16b (uncertainty of employment)

x21 (climate of communications)

x25d (conflicts between personnel groups)

**Method: Enter and then click OK.**

Output:

****

In Model Summary we can see that all variables predict together 28 % of organisational commitment (R Square 0,283).

****

The model predicts organisational commitment with statistically highly significance (p<.001).

****

The main results are in table of coefficients. Standardized **Beta-coefficients** show the net effect in OC which is associated with one unit change in X. We can see that the strongest predictors of organisational commitment are age (year of birth), changes in work interesting and climate of communications. Instead, gender, type of contract and uncertainty of employment have no effect to OC.

* + Regression analysis tries to produce the best **additive** combination of independent variables that produces the best the **linear** relationship between the **observed** Y values and the Y values **predicted** by the resulting regression equation.
  + The **multiple correlation coefficient, R**, is equal to the **product moment correlation, *r***, that would be calculated if the observed values were correlated with the values computed from the regression equation.
  + Similarly, the **multiple R-squared** is equal to the **proportion of variance** in the dependent variable Y that is explained by the additive combination of effects of the independent variables, X1 to Xn.