

Part 2:

Discussions and exercises

4-5.Sept.2025

SDC- course at Statistics Iceland and University of Iceland

Ex: Well-being at work

Table 2.9 shows part of the results of a well-being at work survey conducted in company X. The staff of all four departments (A, B, C and D) answered to the following question

How satisfied are you with your immediate superior?

- a) *completely satisfied (+)*
- b) *rather satisfied (+)*
- c) *rather dissatisfied (-)*
- d) *completely dissatisfied (-).*

Dept.	Satisfaction with the immediate superior		Staff total
	"+"	"-"	
A	0	10	10
B	2	13	15
C	1	1	2
D	12	1	13
Total	15	25	40

Table 2.9: Distribution of answers to the question: "How satisfied are you with your immediate superior?"

Well-being at work –contin.

Table 2.9 was published in company X's staff newsletter.

Let us assume that everybody except the superiors themselves answered the question, i.e. there are no missing answers. Let us also assume that everybody knows in which department everybody works.

Discuss:

1. Is it possible to disclose answers of individual employees based on Table 2.9?
2. If the answer to previous question is “yes”, which would be a better way to present the results so that individual answers could not be disclosed?
3. From SDC point of view which is the best and the worst distribution of answers in the two categories (“+” / “-“)?

Well-being at work –contin.

1. Yes:

- Everybody working in department A are (rather or completely) dissatisfied
- The two persons working in department C know if the other one is satisfied or not.
- In department D the one that is dissatisfied knows that everybody else is satisfied.
- Only the superior of department A knows which of his/her employees are satisfied and which are not (since everybody is dissatisfied). The superiors of other departments cannot single out the satisfied and dissatisfied employees of their department.

Well-being at work –contin.

Dept.	Satisfaction with the immediate superior		Staff total
	"+"	"-"	
A	0	10	10
B	2	13	15
C	1	1	2
D	12	1	13
Total	15	25	40

-satisfaction varies a lot between different departments and thus it makes sense to publish some department level results. For this reasons it also might not be sensible to group answers from different departments, for example A+B and C+D.

A minimum frequency rule should be used so that if a department has less than 3 employees, the results are not published.



Dept.	Is more than half of the employees satisfied (+) or not satisfied (-) with their immediate superior?	Staff total
A	-	10
B	-	15
C	<i>(too few observations -> no information published)</i>	2
D	+	13

Well-being at work –contin.

3.

The best distribution: at least two (+) answers and two (-) answers and all the answers are almost evenly distributed in (+) and (-).

The worst distribution: all answers in more sensitive category (-).

Ex: Research and development expenditure – contin.

⊕	Company size (number of staff)	Total R&D expenditure (EUR million)	Company's own funding	External funding	Number of companies
	0-9	155.5	91.8	63.7	4
⊕	10-49	394.7	258.1	136.6	7
	50-99	272.9	234.4	38.5	1
	100-249	192.2	146.3	45.9	4
	250-499	478.5	373.1	105.4	3
	500+	3189.1	2949.7	239.4	1
	Total	4682.9	4053.4	629.5	20

Table 2.10: Research and development expenditure for companies in area A

Ex: Research and development expenditure

Discuss:

1. Does Table 2.10 pose a disclosure risk if published?
2. Which sensitivity rules could be applied to estimate the disclosure risk of the companies? Let us assume that for estimating the risk company level microdata would be available for you.
3. Which SDC methods could be applied to lower the disclosure risk?

Ex: Research and development expenditure – contin.

1.

- If we assume that all companies in area A are present in the Table 2.10 at least the R&D expenditure of the largest company (by staff size) is disclosed.

Probably the one company with 50-99 staff members could also be identified and its expenditure disclosed assuming information about company size class is somewhere available.

- If Table 2.10 is based on a sample of companies, i.e. all companies in area A are not present in the table, the identification of companies can be more difficult.

If there are (more than one) other companies in area A with 500 or more staff members we cannot know which one of these big companies have R&D expenditures, i.e. which one appears in Table 2.10. We would need additional information to evaluate the risk.

Ex: Research and development expenditure – contin.

2.

- Threshold rule could be applied within every size class.
- Dominance rule or $p\%$ -rule only make sense if the R&D expenditure can be considered “identifying” information, i.e. companies are able to some extent estimate the size of R&D expenditures of other companies and compare it to their own expenditure. If no one has any information or no one can make any reasonable guess which company or companies have the highest R&D expenditures the use of dominance rule or $p\%$ -rule is questionable.
- On the other hand, if one company is strongly dominating the R&D expenditures in a certain size class this company knowing its own expenditure can disclose that other companies do not invest much or anything at all in research and development.

Ex: Research and development expenditure – contin.

3.

For example, redesigning the table (combining size categories) or cell suppression could be used.

Removing the number of companies can be used also as an SDC method but this lowers the identification risk only if the same information (i.e. number of companies in area A) is not available elsewhere.