

# Output checking

Liesbeth Steenhof  
Coordinator output checking Microdata Services

# Why do we check output?

- CBS-law: ensure privacy persons / companies
- data which researchers like to use outside the secure environment (output) is checked by CBS for not violating the CBS-law (disclosure risk)
- Safe output is shared responsibility of CBS and researchers
- Quality of the statistical analyses is not responsibility of CBS

# Disclosure

2 components:  
Identification AND attribution

Recognizing a statistical unit (person, company)  
AND  
Finding out more about this unit

# Disclosure: example

1 male person aged 40 living in a district in The Hague.

Disclosure or not?

NOT

Same table with mean income per age.

Disclosure or not?

DISCLOSURE

# To make it easier: Output guidelines

These are guidelines and no rules

If the researcher wants to deviate from the guidelines, the researcher should explain why there is no disclosure risk

# Output guidelines

- NO microdata
- At least 10 units per cell
- At least 10 degrees of freedom for modeled output
- Group disclosure: no cell should contain more than 90 % of the total number of units in its row or column
- Dominance: largest contributor to a cell should not contribute more than 50% of the total amount in the cell

# At least 10 units per cel

Housholds with children by number of children and pets

Number of children	Number of Pets			
	0	1	2+	Total
1-3	32	64	84	180
4-10	28	48	28	104
11-20	10	16	10	36
>20	1	0	0	1
Total	71	128	122	321

# Group disclosure: 90% at most per cell

Percentage persons with social security benefit by postcode

Postcode	Social security benefit	No social security benefit
....		
1011	95%	5%
1012	35%	65%
....		



# Dominance: contributing 50% at most

Investments of companies (mln euros) by region and activity

	Region 1	Region 2	Region 3	Total
Activity 1	78	12	9	99
Activity 2	2	367	12	381
Activity 3	2	28	389	419
Total	82	407	410	899

# Dominance: contributing 50% at most

Total number of contributors per cell

	Region 1	Region 2	Region 3	Total
Activity 1	14	10	12	36
Activity 2	13	15	11	39
Activity 3	10	21	23	54
Total	37	46	46	129

Percentual contribution of biggest contributor

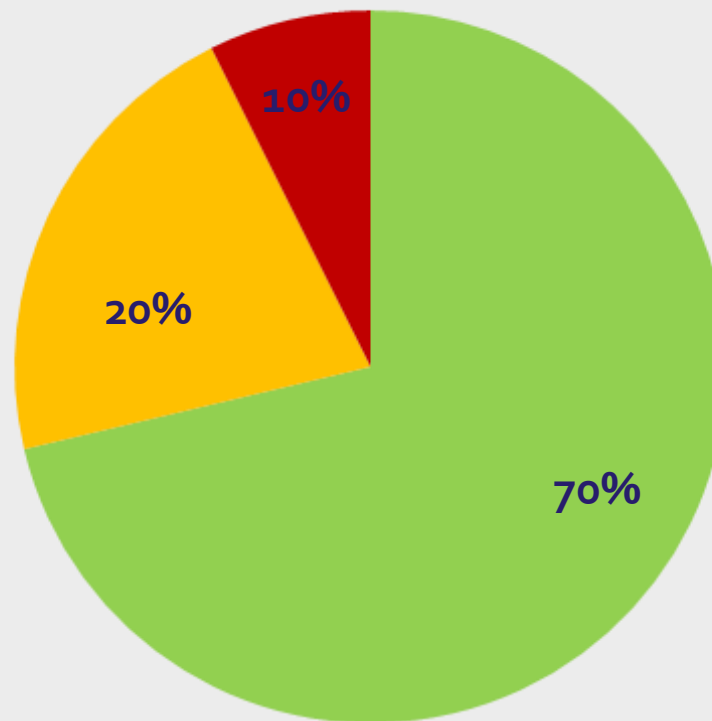
	Region 1	Region 2	Region 3	Total
Activity 1	44%	33%	22%	
Activity 2	25%	46%	42%	
Activity 3	30%	36%	80%	
Total				

# How do we check output?

- Researchers put output in their output folder
- Once a day (mornings) output is collected and distributed
- 17 checkers
- Changing pairs of checkers
- Output checking is not the main activity of the checkers
- If necessary we consult a theme expert or methodologist
- On average 3 to 4 outputs per day
- Checking within 3 working days
- 200 euro per output

# Results of output checking

In 2016 almost 1000 outputs



■ released ■ adjusted ■ stopped

# Changes in output checking?

- Checking afterwards
- Output light

# Questions?

