Chapter 8 – Manipulation through Statistics?

Do statistics lie?

The only statistics you can trust are those you falsified yourself.
(pretended to be from) Sir Winston Churchill, British Prime Minister (1874 - 1965)

Counterfeit or truth?

Does that mean that statistics are just in general faked? Wouldn't we rather agree that statistical data can be interpreted in different ways? The data can be put together in a certain way so that the results can be seen differently. Sometimes we can assume that we can have great faith in numbers. Normally the credibility of a statistic is not challenged. If a moderator says that the consumer price index (CPI) (basis 1995 = 100) for alcoholic beverages and tobacco products is in 2000 was as high as 107.5 (in all German households; Federal Statistical Office Germany, Mai 31, 2001), then the audience will usually accept that number. They believe in that number! The audience will not think that the number may be 108.5 or 106.5 instead of 107.5. As a moderator you can use the belief in numbers to your advantage. Namely:

- the number is said
- the number is believed

Your statement continues

- and this statement is also believed (because the number is already defined as 'true')

The source of the statistical data <u>can,</u> but does not have to, be mentioned. If you have a source that is generally accepted such as a credible news magazine then the desired effect of the data is considerable. From this, it follows that, in order to rely (or even to emphasize) on our statement, we include statistical data. Thereby we notice:

- one, two or a maximum of three statistical numbers related to the same information
- numbers with many of digits (if possible round up or down)
  - instead of 3,724,624, approximately 3.7 million is better
  - instead of 0.8769, approximately 0.88 better
- include the word 'statistic' or 'survey'
  - "according to a statistic …" or

- "statistically seen …"
- References only if the source is credible
- statement may sound funny to the audience
  - "X % of all Germans sleep naked."

Analysis of the collected data

In statistics the following terms are used:

- Midpoint
  - The midpoint is the sum of all values divided by the number of single values (of the interviewees).
- Median
  - The median halves the collection of data after the number of single values (of the interviewees) into half-half (50 % to 50 %).
- Mode
  - The mode is the nominal value that most frequently appears.

Example:

Question: "Which grade (German high school grading system) would you give the TV show ABC?"

Grade	1	2	3	4	5	6	Sum
Quantity	5	4	7	19	20	5	60
	5 x 1	4 x 2	7 x 3	19 x 4	20 x 5	5 x 6	
Value	5	8	21	76	100	30	240

n = number of interviewees (single value) who gave that grade.

From this table we can analyze the following:

- Midpoint
  - 240 divided by 60 equals 4.0
- Median
  - The same amounts of single value are on the right and left side of the middle. Therefore we get 4.5.
- Mode
  - The most mentioned value is 5.0.

We see that depending on the used term, we get a number in between 4.0 and 5.0. How we can manipulate with this data is demonstrated in the next chapter.

Real manipulation through statistics?

In the following examples we assume real manipulation. If you, as a lecturer want to manipulate like this, is your personal decision.

More important is that you can realize when a conversation partner tries to manipulate you and therefore you can react.

We refer in the following example to numbers which have been listed in the previous chapter.

Result	Statement
1	"Statistically seen there is a midpoint of 4.0 for the evaluated TV show."
2a	"Most evaluations got a grade of 5.0."
2b	"Most of the interviewees (35 out of 60) gave the TV show a good grade (in between 1.0 and 4.0)."
3a	"Half of the interviewees gave the TV show a grade in between 1 and 4.5."
3b	"Half of the interviewees gave the TV show a grade in between 4.5 and 6."
4a	"At least 26.6 % (16 out of 60) of the interviewees gave the TV show a grade in between 1 to 3."
4b	Only ¼ (16 out of 60) of the interviewees gave the TV show a grade in between 1 to 3."
5a	"41.6 % (25 out of 60) of the interviewees gave the TV show a grade in between 5 to 6."
5b	"Almost half of the interviewees (25 out of 60) gave the TV show a grade in between 5 and 6."
6a	"Only 5 persons gave the TV show a grade of 6."
6b	"Almost 10 % (exactly 8.33 %) gave the TV show a grade of 1."
6c	"At least every twelfth person gave the TV show a grade of 1."

Do not forget: The same data was used.

Practice tip: look for any statistical data. Use the numbers on one hand to strengthen a statement and on the other hand for a negative interpretation of the subject.

Smart display of the results?

Due to an (fictional) survey the following result occurred:

in favor	15
neutral	55
against	30

How can you present the result (without lying)?

- "Just 15 % are in favor."
- "70 % are not against it."
- "30 % are against it."
- "85 % are not in favor."
- "Twice as much is against it than in favor."

Independent of the data, every statistical number can be used positively as well as negatively.

(Fictional) example:

- Ten percent of all people living in Cologne agree with solution A.

Interpretation and display:

- "Only 10 % of all people living in cologne agree with solution A. That is clearly the minority of people living in Cologne."
- "After all, 100,000 people living in Cologne agree with solution A. Just imagine that outrageous crowd on the street."

Another example:

The following data resulted in a (fictional) survey about the subject. "How did you like the first-run movie?"

good	65 %
bad	35 %

Obviously the majority of the people questioned did like the movie.

Result: notably more good than bad.

Let's see if the results possibly change if we give three alternatives instead of two (good vs. bad) three:

outstanding	30 %
good	35 %
bad	35 %

Result: the same amount of people that said good, also said bad.

Or if we even give six different possibilities to answer:

outstanding	30 %
good	20 %
rather good	15 %
rather bad	5 %
bad	25 %
very bad	5 %

Result: more bad than good.

That means: depending on how we collect the data (only two possibilities to answer or six possibilities) it may end up in clearly different (or even opposite) results.

Depending on the direction you want to manipulate the data you can change the options. Obviously, data can be manipulated so far that you can actually get the opposite statement of what originally was correct.

Just for clarification: the reader should not be encouraged to manipulate but he should be able to recognize when and how manipulation is tried to influence him and that he knows how to react to it.

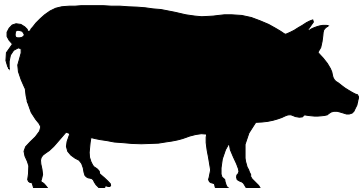
Are 2 times 2 equal 2 to the power of 2?

In year X, the average European eats 100 pork cutlets. In year Y, he eats twice as much, i.e., 200 pork cutlets. Therefore we have the following pictographic display:

year X

100

year Y

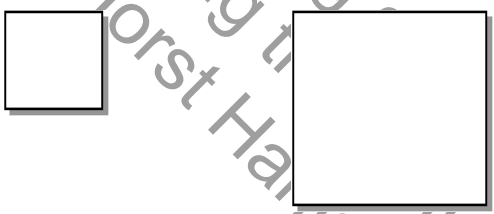


200

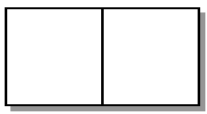
Is it right? NO – it is wrong! And why? Because the second pig is not displayed in the right proportion to the first one.

The pig on the right side is not displayed twice as big as the one on the left side. It is more likely four times as big. Because it is doubled in length and height. Therefore 2 times 2 are not 2 to the power of 2.

For clarification let's demonstrate it with squares.

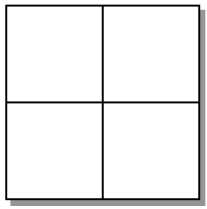


To double, looks like this:



Or so:

But not like that:



This pictographic display of our pig is more likely correct:

In reality numbers are rarely exactly 100 or 200. They are most likely 97 or 213 so that possible manipulation is even more difficult to recognize.

In a three dimensional sample the third dimension is added. Therefore the size does not increase four times. It expands exponentially at the same time, namely 8 times.