

# 6

## Measurement Error

*Jesus: I am come to bear witness unto the truth.  
Pilate: What is truth?*

### 1. INTRODUCTION

In an ideal world, if the same thing is measured several times, the same result would be obtained each time. In practice, there are differences. Each result is thrown off by chance error, and the error changes from measurement to measurement. One of the earliest scientists to deal with this problem was Tycho Brahé (1546–1601), the Danish astronomer. But it was probably noticed first in the market place, as merchants weighed out spices and measured off lengths of silk.

There are several questions about chance errors. Where do they come from? How big are they likely to be? How much is likely to cancel out in the average? The first question has a short answer: in most cases, nobody knows. The second question will be dealt with later in this chapter, and the third will be answered in part VII.

### 2. CHANCE ERROR

This section will discuss chance errors in precision weighing done at the National Bureau of Standards.<sup>1</sup> First, a brief explanation of standard weights. Stores weigh merchandise on scales. The scales are checked periodically by county

weights-and-measures officials, using county standard weights. The county standards too must be *calibrated* (checked against external standards) periodically. This is done at the state level. And state standards are calibrated against national standards, by the National Bureau of Standards in Washington, D.C.

This chain of comparisons ends at the International Prototype Kilogram (for short, The Kilogram), a platinum-iridium weight held at the International Bureau of Weights and Measures near Paris. By international treaty—The Treaty of the Meter, 1875—“one kilogram” was defined to be the weight of this object under standard conditions.<sup>2</sup> All other weights are determined relative to The Kilogram. For instance, something weighs a pound if it weighs just a bit less than half as much as The Kilogram. More precisely,

$$\text{The Pound} = 0.4539237 \text{ of The Kilogram.}$$

To say that a package of butter weighs a pound means that it has been connected by some long and complicated series of comparisons to The Kilogram in Paris, and weighs 0.4539237 times as much.

Each country that signed the Treaty of the Meter got a national prototype kilogram, whose exact weight had been determined as accurately as possible relative to The Kilogram. These prototypes were distributed by lot, and the United States got Kilogram #20. The values of all the U.S. national standards are determined relative to K<sub>20</sub>.

In the U.S., accuracy in weighing at the supermarket ultimately depends on the accuracy of the calibration work done at the Bureau. One basic issue is reproducibility: if a measurement is repeated, how much will it change? The Bureau gets at this issue by making repeated measurements on some of their own weights. We will discuss the results for one such weight, called NB 10 because it is owned by the National Bureau and its nominal value is 10 grams—the weight of two nickels. (A package of butter has a “nominal” weight of 1 pound; the exact weight will be a little different—chance error in butter; similarly, the people who manufactured NB 10 tried to make it weigh 10 grams, and missed by a little.)

NB 10 was acquired by the Bureau around 1940, and they’ve weighed it many times since then. We are going to look at 100 of these weighings. These measurements were made in the same room, on the same apparatus, by the same technicians. Every effort was made to follow the same procedure each time. All the factors known to affect the results, like air pressure or temperature, were kept as constant as possible.

The first five weighings in the series were

9.999591 grams  
 9.999600 grams  
 9.999594 grams  
 9.999601 grams  
 9.999598 grams

At first glance, these numbers all seem to be the same. But look more closely. It is only the first 4 digits that are solid, at 9.999. The last 3 digits are shaky, they change from measurement to measurement. This is chance error at work.<sup>3</sup>

NB 10 does weigh a bit less than 10 grams. Instead of writing out the 9.999 each time, the Bureau just reports the amount by which NB10 fell short of 10 grams. For the first weighing, this was

0.000409 grams.

The 0's are distracting, so the Bureau works not in grams but in micrograms: a *microgram* is the millionth part of a gram. In these units, the first five measurements on NB 10 are easier to read. They are

409      400      406      399      402.

All 100 measurements are shown in table 1. Look down the table. You can see that the results run around 400 micrograms, but some are more, some are less. The smallest is 375 micrograms (#94); the largest is 437 micrograms (#86). And there is a lot of variability in between. To keep things in perspective, one microgram is the weight of a large speck of dust; 400 micrograms is the weight of a grain or two of salt. This really is precision weighing!

Even so, the different measurements can't all be right. The exact amount by which NB 10 falls short of 10 grams is very unlikely to equal the first number

Table 1. One hundred measurements on NB 10. Almer and Jones, National Bureau of Standards. Units are micrograms below 10 grams.

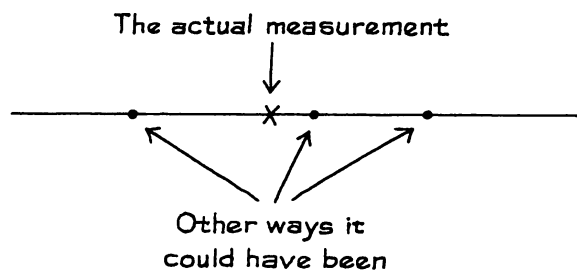
<i>No.</i>	<i>Result</i>	<i>No.</i>	<i>Result</i>	<i>No.</i>	<i>Result</i>	<i>No.</i>	<i>Result</i>
1	409	26	397	51	404	76	404
2	400	27	407	52	406	77	401
3	406	28	401	53	407	78	404
4	399	29	399	54	405	79	408
5	402	30	401	55	411	80	406
6	406	31	403	56	410	81	408
7	401	32	400	57	410	82	406
8	403	33	410	58	410	83	401
9	401	34	401	59	401	84	412
10	403	35	407	60	402	85	393
11	398	36	423	61	404	86	437
12	403	37	406	62	405	87	418
13	407	38	406	63	392	88	415
14	402	39	402	64	407	89	404
15	401	40	405	65	406	90	401
16	399	41	405	66	404	91	401
17	400	42	409	67	403	92	407
18	401	43	399	68	408	93	412
19	405	44	402	69	404	94	375
20	402	45	407	70	407	95	409
21	408	46	406	71	412	96	406
22	399	47	413	72	406	97	398
23	399	48	409	73	409	98	406
24	402	49	404	74	400	99	403
25	399	50	402	75	408	100	404

in the table, or the second, or any of them. Despite the effort of making these 100 measurements, the exact weight of NB 10 remains unknown and perhaps unknowable.

Why does the Bureau bother to weigh the same weight over and over again? One of the objectives is quality control. If the measurements on NB 10 jump from 400 micrograms below 10 grams to 500 micrograms above 10 grams, something has gone wrong and needs to be fixed. (For this reason, NB 10 is called a *check weight*; it is used to check the weighing process.)

To see another use for repeated measurements, imagine that a scientific laboratory sends a nominal 10-gram weight off to the Bureau for calibration. One measurement can't be the last word, because of chance error. The lab will want to know how big this chance error is likely to be. There is a direct way to find out: send the same weight back for a second weighing. If the two results differ by a few micrograms, the chance error in each one is only likely to be a few micrograms in size. On the other hand, if the two results differ by several hundred micrograms, each measurement is likely to be off by several hundred micrograms. The repeated weighings on NB 10 save everybody the bother of sending in weights more than once. There is no need to ask for replicate calibrations because the Bureau has already done the work.

No matter how carefully it was made, a measurement could have come out a bit differently. If the measurement is repeated, it will come out a bit differently. By how much? The best way to answer this question is to replicate the measurement.



The SD of the 100 measurements in table 1 is just over 6 micrograms. The SD tells you that each measurement on NB 10 was thrown off by a chance error something like 6 micrograms in size. Chance errors around 2 or 5 or 10 micrograms in size were fairly common. Chance errors around 50 or 100 micrograms must have been extremely rare. The conclusion: in calibrating other 10-gram weights by the same process, the chance errors should be something like 6 micrograms in size.

The SD of a series of repeated measurements estimates the likely size of the chance error in a single measurement.

There is an equation which helps explain the idea:

$$\text{individual measurement} = \text{exact value} + \text{chance error}.$$

The chance error throws each individual measurement off the exact value by an amount which changes from measurement to measurement. The variability in repeated measurements reflects the variability in the chance errors, and both are gauged by the SD of the data. Mathematically, the SD of the chance errors must equal the SD of the measurements: adding the exact value is just a change of scale (pp. 92–93).

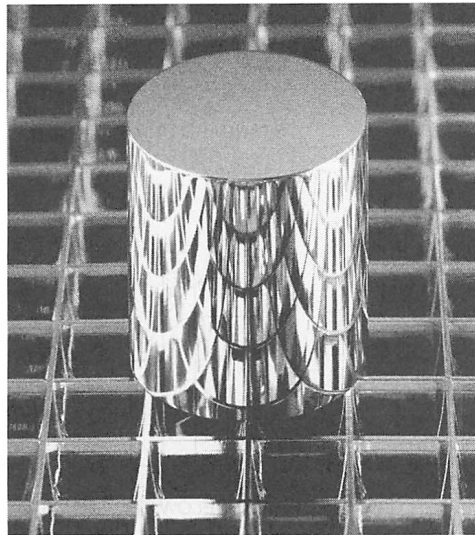
To go at this more slowly, the average of all 100 measurements reported in table 1 was 405 micrograms below 10 grams. This is very likely to be close to the exact weight of NB 10. The first measurement in table 1 differed from the average by 4 micrograms:

$$409 - 405 = 4.$$

This measurement must have differed from the exact weight by nearly 4 micrograms. The chance error was nearly 4 micrograms. The second measurement was below average by 5 micrograms; the chance error must have been around  $-5$  micrograms. The typical deviation from average was around 6 micrograms in size, because the SD was 6 micrograms. Therefore, the typical chance error must have been something like 6 micrograms in size.

Of course, the average of all 100 measurements (405 micrograms below 10 grams) is itself only an estimate for the exact weight of NB10. This estimate too must be off by some infinitesimal chance error. Chapter 24 will explain how to figure the likely size of the chance error in this sort of average.

Figure 1. The U.S. national prototype kilogram, K<sub>20</sub>.



Source: National Institute of Science and Technology.