Just because statistical software—including *R* and *Python*—generates

output by default does not mean that all the output is useful or

relevant. You can see that the preceding standard deviations are not

that useful; on their face they suggest that numerous values might

be negative, when negative revenue is not feasible. This data consists

of a small set of relatively high values (page views with conversions)

and a huge number of 0-values (page views with no

conversion). It is difficult to sum up the variability of such data

with a single number, though the mean absolute deviation from the

mean (7.68 for A and 8.15 for B) is more reasonable than the standard

deviation. **O**Blinding in studies

A *blind study* is one in which the subjects are unaware of whether

they are getting treatment A or treatment B. Awareness of receiving

a particular treatment can affect response. A *double-blind* study is

one in which the investigators and facilitators (e.g., doctors and

nurses in a medical study) also are unaware which subjects are getting

which treatment. Blinding is not possible when the nature of

the treatment is transparent—for example, cognitive therapy from

a computer versus a psychologist. **B**

Getting Permission

In scientific and medical research involving human subjects, it is

typically necessary to get their permission, as well as obtain the

approval of an institutional review board. Experiments in business

that are done as a part of ongoing operations almost never do this.

In most cases (e.g., pricing experiments, or experiments about

which headline to show or which offer should be made), this practice

is widely accepted. Facebook, however, ran afoul of this general

acceptance in 2014 when it experimented with the emotional tone

in users’ newsfeeds. Facebook used sentiment analysis to classify

newsfeed posts as positive or negative, and then altered the positive/

negative balance in what it showed users. Some randomly

selected users experienced more positive posts, while others experienced

more negative posts. Facebook found that the users who

experienced a more positive newsfeed were more likely to post positively

themselves, and vice versa. The magnitude of the effect was

small, however, and Facebook faced much criticism for conducting

the experiment without users’ knowledge. Some users speculated

that Facebook might have pushed some extremely depressed users

over the edge if they got the negative version of their feed. **O**

False Discovery Rate

The term *false discovery rate* was originally used to describe the rate

at which a given set of hypothesis tests would falsely identify a significant

effect. It became particularly useful with the advent of

genomic research, in which massive numbers of statistical tests

might be conducted as part of a gene sequencing project. In these

cases, the term applies to the testing protocol, and a single false

“discovery” refers to the outcome of a hypothesis test (e.g., between

two samples). Researchers sought to set the parameters of the testing

process to control the false discovery rate at a specified level.

The term has also been used for classification in data mining; it is

the misclassification rate within the class 1 predictions. Or, put

another way, it is the probability that a “discovery” (labeling a

record as a “1”) is false. Here we typically are dealing with the case

where 0s are abundant and 1s are interesting and rare (see Chapter

5 and “The Rare Class Problem” on page 223). **B**

Decomposition of Variance

Observed values in a data set can be considered sums of different

components. For any observed data value within a data set, we can

break it down into the grand average, the treatment effect, and the

residual error. We call this a “decomposition of variance”:

1. Start with grand average (173.75 for web page stickiness data).

2. Add treatment effect, which might be negative (independent

variable = web page).

3. Add residual error, which might be negative.

Thus the decomposition of the variance for the top-left value in the

A/B/C/D test table is as follows:

1. Start with grand average: 173.75.

2. Add treatment (group) effect: –1.75 (172 – 173.75).

3. Add residual: –8 (164 – 172).

4. Equals: 164.  **B**