CHAPTER 1

Measurement

1.1 What Is Biostatistics?

Biostatistics is the discipline concerned with the treatment and analysis of numerical data derived from biological, biomedical, and health-related studies. The discipline encompasses a broad range of activities, including the design of research, collection and organization of data, summarization of results, and interpretation of findings. One of its important functions is to weigh evidence and help draw conclusions based on numerical data. In all these functions, biostatistics is the *servant of the sciences.*[a]

Biostatistics is *more* than just a compilation of computational techniques. It is *not* merely pushing numbers through formulas or computers, but rather it is a way to *detect* patterns and *judge* responses. The statistician is both a data detective and judge.[b] The data detective uncovers patterns and clues, while the data judge decides whether the evidence can be trusted. Goals of biostatistics include[c]:

- Improvement of the intellectual content of the data
- Organization of data into understandable forms
- Reliance on tests of experience as a standard of validity

---

[a] Neyman, J. (1955). Statistics—servant of all sciences. *Science*, *122*, 401–406.
[b] Tukey, J. W. (1969). Analyzing data: sanctification or detective work? *American Psychologist*, *24*, 83–91.
[c] Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics*, *33*(1), 1–67, esp. p. 5.

1.2 Organization of Data

Observations, Variables, Values

Measurement is how we get our data. More formally, *measurement* is "the assigning of numbers or codes according to prior-set rules."[d] Measurement may entail either positioning observations along a numerical continuum (e.g., determining a person's age) or classifying observations into categories (e.g., determining whether an individual is seropositive or seronegative for HIV antibodies).

The term **observation** refers to the unit upon which measurements are made. Observations may correspond to individual people or specimens. They may also correspond to aggregates upon which measurements are made. For example, we can measure the smoking habits of an individual (in terms of "pack-years" for instance) or we can measure the smoking habits of a region (e.g., per capita cigarette consumption). In the former case, the *unit of observation* is a person; in the later, the *unit of observation* is a region.

Data are often collected with the aid of a **data collection form**, with data on individual data forms usually corresponding to observations. **Figure 1.1** depicts four such observations. Each field on the form corresponds to a **variable**. We enter **values** into these fields. For example, the *value* of the fourth *variable* of the first *observation* in **Figure 1.1** is "45."

Do not confuse *variables* with *values*. The *variable* is the generic thing being measured. The value is a number or code that has been realized.

Observations are the unit upon which measurements are made.
Variables are the characteristic being measured.
Values are realized measurements.

Data Table

Once data are collected, they are organized to form a **data table**. Typically, each row in a data table contains an observation, each column contains a variable, and each cell contains a value.

---

[d] Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

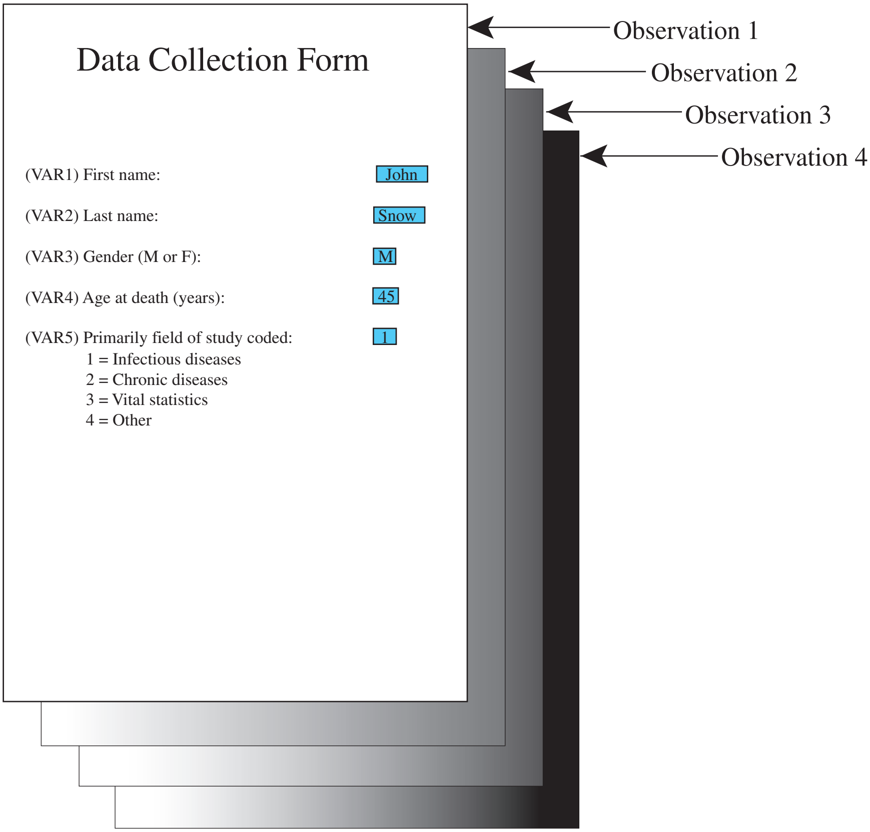

FIGURE 1.1 Four *observations* with five *variables* each.

Data table
Observations → rows
Variables → columns
Values → table cells

Table 1.1 corresponds to data collected with the form depicted in **Figure 1.1**. This data table has four *observations*, five *variables*, and 20 *values*. For example, the value of VAR1 for the first observation is "John." As another example, the value of VAR4 for the second observation is "75."

Table 1.1 Data table for data collected with the forms in Figure 1.1.

VAR1	VAR2	VAR3	VAR4	VAR5
John	Snow	M	45	1
William	Farr	M	75	3
Joseph	Goldberger	M	54	2
Janet	Lane-Claypon	F	90	2

Table 1.2 is a data table composed of three variables: country of origin (COUNTRY), per capita cigarette consumption (CIG1930), and lung cancer mortality (LUNGCA). The unit of observation in this data set is a country, not an individual person. Data of this type are said to be *ecological*.[e] This data table has 11 *observations*, three *variables*, and 33 *values*.

Exercises

1.1 ***Value, variable, observation.*** In **Table 1.2**, what is the value of the LUNGCA variable for the seventh observation? What is the value of the COUNTRY variable for the eleventh observation?

1.2 ***Value, variable, observation (cont.).*** What is the value of the CIG1930 variable for observation 3 in **Table 1.2**?

Table 1.2 Per capita cigarette consumption in 1930 (CIG1930) and lung cancer cases per 100,000 in 1950 (LUNGCA) in 11 countries.

COUNTRY	CIG1930	LUNGCA
USA	1300	20
Great Britain	1100	46
Finland	1100	35
Switzerland	510	25
Canada	500	15
Holland	490	24
Australia	480	18
Denmark	380	17
Sweden	300	11
Norway	250	9
Iceland	230	6

Source: Doll, R. (1955). Etiology of lung cancer. *Advances Cancer Research*, *3*, 1–50.

---

[e] The term *ecological* in this context should not be confused with its biological use.

1.3 ***Value, variable, observation (cont.).*** In the form depicted in **Figure 1.1**, what does VAR3 measure?

1.4 ***Value, variable, observation (cont.).*** In **Table 1.1**, what is the value of VAR4 for observation 3?

1.3 Types of Measurements

There are different ways to classify variables and measurements. We consider three types of measurements: categorical, ordinal, and quantitative.[f] As we go from categorical to ordinal to quantitative, each scale will take on the assumptions of the prior type and adds a further restriction.

- Categorical measurements place observations into unordered categories.
- Ordinal measurements place observations into categories that can be put into rank order.
- Quantitative measurements impose equal spacing between ordered intervals.

Additional explanation follows.

Categorical measurements place observations into classes or groups. Examples of categorical variables are SEX (male or female), BLOOD_TYPE (A, B, AB, or O), and DISEASE_STATUS (case or noncase). Categorical measurements may occur naturally (e.g., diseased/not diseased) or can be created by grouping quantitative measurements into classes (e.g., classifying blood pressure as normotensive or hypertensive). Categorical variables are also called *nominal variables* (nominal means "named"), *attribute variables*, and *qualitative variables*.

Ordinal measurements assign observations into categories that can be put into rank order. An example of an ordinal variable is STAGE_OF_CANCER classified stage I, stage II, or stage III. Another example is OPINION ranked on a 5-point scale (e.g., 5 = "strongly agree," 4 = "agree," and so on). Although ordinal scales place observation into order, the "distance" (difference) between ranks is not uniform. For example, the difference between stage I cancer and stage II cancer is not necessarily the same as the difference between stage II and stage III. Ordinal variables serve merely as a ranking, and do not truly quantify differences.

Quantitative measurements position observations along a numeric scale. Examples of quantitative measures are chronological AGE (years), body WEIGHT (pounds), systolic BLOOD_PRESSURE (mmHg), and serum GLUCOSE (mmol/L). Some

---

[f] Distinctions between measurement scales often get blurred in practice. See Velleman, P. F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician*, *47*, 65–72.

statistical sources use terms such as *ratio/interval measurement*, *numeric variable*, *scale variable*, and *continuous variable* to refer to quantitative measurements.

Illustrative Example: *Weight change and coronary heart disease.*[g] A group of 115,818 women between 30 and 55 years of age were recruited to be in a study. Individuals were free of coronary heart disease at the time of recruitment. Body weight of subjects was determined as of 1976. Let us call this variable WT_1976. Weight was also determined as of age 18. Let's call this variable WT_18. From these variables, the investigators calculated weight change for individuals (WT_CHNG = WT_1976 − WT_18). Adult height in meters was determined (HT) and was used to calculate body mass index according to the formula: BMI = weight in kilograms ÷ (height in meters)^2. BMI was determined as of age 18 (BMI_18) and at the time of recruitment in 1976 (BMI_1976). All of these variables are quantitative.

BMI was classified into *quintiles*. This procedure divides a quantitative measurement into five ordered categories with an equal number of individuals in each group. The lowest 20% of the values are put into the first quintile, the next 20% are put into the next quintile, and so on. The quintile cutoff points for BMI at age 18 were: $<$19.1, 19.1–20.3, 20.4–21.5, 21.6–23.2, and $\geq$23.3. Let us put this information into a new variable called BMI_18_GRP encoded 1, 2, 3, 4, 5 for each of the quintiles. This is an ordinal variable.

The study followed individuals over time and monitored whether they experienced adverse coronary events. A new variable (let us call it CORONARY) would then be used to record this new information. CORONARY is a categorical variable with two possible values: either the person did or did not experiences an adverse coronary event. During the first 14 years of follow-up, there were 1292 such events. ■

Exercises

1.5 ***Measurement scale.*** Classify each variable depicted in Figure 1.1 as either quantitative, ordinal, or categorical.

1.6 ***Measurement scale (cont.).*** Classify each variable in Table 1.2 as quantitative, ordinal, or categorical.

---

[g] Willett, W. C., Manson, J. E., Stampfer, M. J., Colditz, G. A., Rosner, B., Speizer, F. E., et al. (1995). Weight, weight change, and coronary heart disease in women. Risk within the "normal" weight range. *JAMA*, *273*, 461–465.

1.4 Data Quality

Meaningful Measurements

How reliable is a single blood pressure measurement? What does an opinion score *really* signify? How is cause of death determined on death certificates? Responsible statisticians familiarize themselves with the measurements they use in their research. This requires a critical mind and, often, some research. It may also require consultation with a subject matter specialist. We must always do our best to understand the variables we are analyzing.

In our good intentions to be statistical, we might be tempted to collect data that is several steps removed from what we really want to know. This is often a bad idea.

> A drunken individual is searching for his keys under a street lamp at night. A passerby asks the drunk what he is doing. The drunken man slurs that he is looking for his keys. After helping the man unsuccessfully search for the keys under the streetlamp, the Good Samaritan inquires whether the drunk is sure the keys were lost under the street lamp. "No," replies the drunk, "I lost them over there." "Then why are you looking here?" asks the helpful Samaritan. "Because the light is here," says the drunk.

Beware of looking for statistical relationships in data that are far from the information that is actually required.

Here is a story you may be less familiar with. This story comes from the unorthodox scientist Richard Feynman who calls pseudoscientific work **Cargo Cult science.**[h] This story is based on an actual occurrence in a South Seas island people during World War II. During the war, the inhabitants of the island saw airplanes land with goods and materials. With the end of the war, the cargo airplanes ceased and so did deliveries. Since the inhabitants wanted the deliveries to continue, they arranged to imitate things they saw when cargo arrived. Runways lights were constructed (in the form of fires), a wooden hut with bamboo sticks to imitate antennas was built for a "controller" who wore two wooden pieces on his head to emulate headphones, and so on. With the Cargo Cult in place, the island inhabitants awaited airplanes to land. The form was right on the surface, but of course things no longer functioned as they had hoped. Airplanes full of cargo failed to bring goods and services to the island inhabitants. "Cargo Cult" has come to mean a pseudoscientific method that follows precepts and forms, but is missing in the honest, self-critical assessments that is essential to scientific investigation.

---

[h] Feynman, R. P. (1999). Cargo Cult science: Some remarks on science, pseudoscience, and learning how not to fool yourself. In *The Pleasure of Finding Things Out* (pp. 205–216). Cambridge, MA: Perseus.

These two stories are meant to remind us that sophisticated numerical analyses cannot compensate for poor-quality data. Statisticians have a saying for this: "**Garbage in, garbage out,**" or **GIGO**, for short.

GIGO stands for "garbage in, garbage out."

Objectivity (the intent to a measure things as they are without shaping them to conform to a preconceived worldview) is an important part of measurement accuracy. Objectivity requires a suspension of judgment; it requires us to look at *all* the facts, not just the facts that please us. Consider how subtle word choices may influence responses. Suppose I ask you to remember the word "jam." I can influence the way you interpret the word by preceding it with the word "traffic" or "strawberry." If I influence your interpretation in the direction of traffic jam, you are less likely to recognize the word subsequently if it is accompanied by the word "grape."[i] This effect will occur even when you are warned not to contextualize the word. The point is that we do not interpret words in a vacuum. When collecting information, nothing should be taken for granted.

Two Types of Inaccuracies: Imprecision and Bias

There are two distinct types of measurement errors: imprecision and bias. **Imprecision** is the random inability to get the same result upon repetition (irreproducibility). **Bias** is a systematic deviation from the truth. When something is *un*biased, it is said to be **valid**.

Figure 1.2 depicts how imprecision and bias may play out in practice. This figure considers repeated glucose measurement in a single serum sample. The true glucose level in the sample is 100 milligrams per deciliter. Measurements have been taken with four different instruments.

- Instrument A is precise and unbiased.
- Instrument B is precise and has a positive bias.

---

[i] Example based on Baddeley cited in Gourevitch, P. (1999, 14 June). The Memory Thief. *The New Yorker*, 48–68.

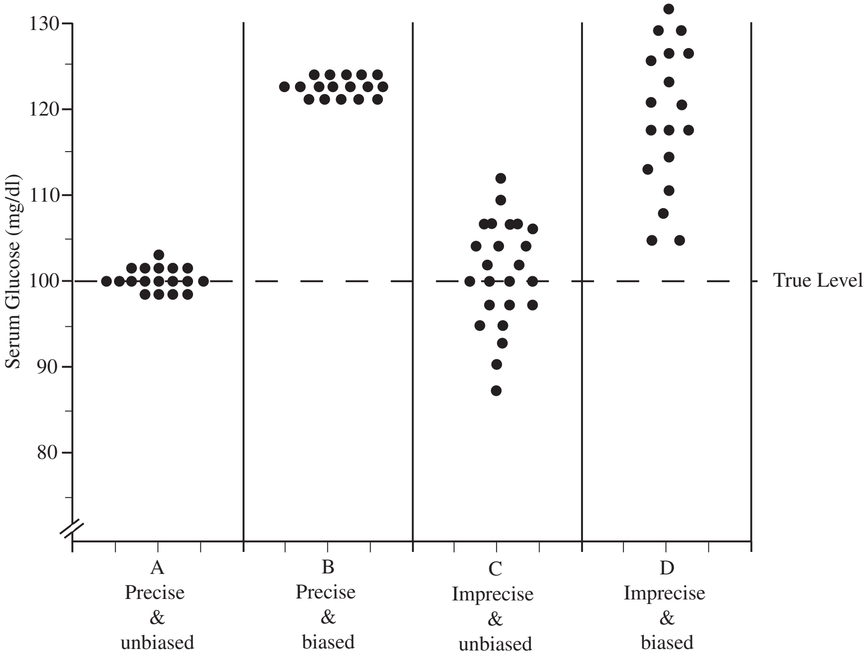

FIGURE 1.2 Repeated glucose measurements on a single sample.

- Instrument C is imprecise and unbiased.
- Instrument D is imprecise and has a positive bias.

In practice, it is easier to quantify imprecision than bias. This fact can be made clear by an analogy. Imagine an archer shooting at a target. A brave investigator is sitting behind the target at a safe distance. Because the investigator is behind the target, he cannot see the location of the actual bull's-eye. He can, however, see where the arrow pokes out of the back of the target (**Figure 1.3**). This is analogous to looking at the results of a study—we see where the arrows stick out but do not actually know the location of the bulls-eye.

Figure 1.4 shows exit sites of arrows from two different archers. From this we can tell that Archer B is more **precise** that Archer A (values spaced tightly). We cannot, however, determine which Archer's aim centers in on the bull's-eye. Characterization of precision is straightforward—measure the scatter in the results. Characterization of validity, however, requires additional information.

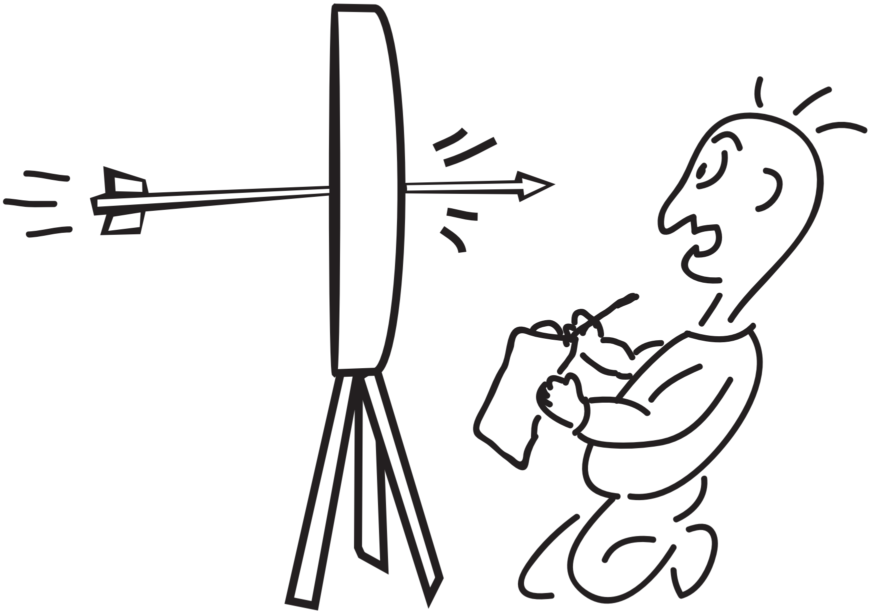

FIGURE 1.3 A brave investigator sits behind the target to see what he can see.

FIGURE 1.4 The investigator sees exit sites of arrows but cannot see the bull's-eye.

Vocabulary

Bias
Cargo cult science
Categorical measurements
Data table
Garbage in, garbage out (GIGO)
Imprecision
Measurement
Objectivity
Observation
Ordinal measurements
Precise
Quantitative measurements
Valid
Values
Variable

Exercises

1.7 ***Duration of hospitalization.*** **Table 1.3** contains data from an investigation that studied antibiotic use in hospitals.

(a) Classify each variable as quantitative, ordinal, or categorical.
(b) What is the value of the DUR variable for observation 4?
(c) What is the value of the AGE variable for observation 24?

1.8 ***Clustering of adverse events.*** An investigation was prompted when the U. S. Food and Drug Administration received a report of an increased frequency of an adverse drug-related event after a hospital switched from the innovator company's product to a generic product. To address this issue, a team of investigators completed chart reviews of patients who had received the drugs in question. **Table 1.4** lists data for the first 25 patients in the study.

(a) Classify each variable in the table as either quantitative, ordinal, or categorical.
(b) What is the value of the AGE variable for observation 4?
(c) What is value of the DIAG (diagnosis) variable for observation 2?

1.9 ***Dietary histories.*** Prospective studies on nutrition often require subjects to keep detailed daily dietary logs. In contrast, retrospective studies often rely on recall. Which method—dietary logs or retrospective recall—do you believe is more likely to achieve accurate results? Explain your response.

1.10 ***Variable types.*** Classify each of the measurements listed here as quantitative, ordinal, or categorical.

(a) Response to treatment coded as 1= no response, 2 = minor improvement, 3 = major improvement, 4 = complete recovery
(b) Annual income (pretax dollars)
(c) Body temperature (degrees Celsius)

Table 1.3 Twenty-five observations derived from hospital discharge summaries.

ID	DUR	AGE	SEX	TEMP	WBC	AB	CULT	SERV
1	5	30	2	99.0	8	2	2	1
2	10	73	2	98.0	5	2	1	1
3	6	40	2	99.0	12	2	2	2
4	11	47	2	98.2	4	2	2	2
5	5	25	2	98.5	11	2	2	2
6	14	82	1	96.8	6	1	2	2
7	30	60	1	99.5	8	1	1	1
8	11	56	2	98.6	7	2	2	1
9	17	43	2	98.0	7	2	2	1
10	3	50	1	98.0	12	2	1	2
11	9	59	2	97.6	7	2	1	1
12	3	4	1	97.8	3	2	2	2
13	8	22	2	99.5	11	1	2	2
14	8	33	2	98.4	14	1	1	2
15	5	20	2	98.4	11	2	1	2
16	5	32	1	99.0	9	2	2	2
17	7	36	1	99.2	6	1	2	2
18	4	69	1	98.0	6	2	2	2
19	3	47	1	97.0	5	1	2	1
20	7	22	1	98.2	6	2	2	2
21	9	11	1	98.2	10	2	2	2
22	11	19	1	98.6	14	1	2	2
23	11	67	2	97.6	4	2	2	1
24	9	43	2	98.6	5	2	2	2
25	4	41	2	98.0	5	2	2	1

Here's a codebook for the data:

Variable	Description
DUR	Duration of hospitalization (days)
AGE	Age (years)
SEX	1 = male, 2 = female
TEMP	Body temperature (degrees Fahrenheit)
WBC	White blood cells per 100ml
AB	Antibiotic use: 1 = yes, 2 = no
CULT	Blood culture take: 1 = yes, 2 = no
SERV	Service: 1 = medical, 2 = surgical

Sources: Townsend, T. R., Shapiro, M., Rosner, B., & Kass, E. H. (1979). Use of antimicrobial drugs in general hospitals. I. Description of population and definition of methods. *Journal of Infectious Disease*, *139*(6), 688–697 and Rosner, B. (1990). *Fundamentals of Biostatistics* (Third ed.). Belmont, CA: Duxbury Press, p. 36.

(d) Area of a parcel of land (acres)

(e) Population density (people per acre)

(f) Political party affiliation coded 1 = Democrat, 2 = Republican, 3 = Independent, 4 = Other

Table 1.4 First 25 observations from a study of cerebellar toxicity.

I	AGE	SEX	MANUF	DIAG	STAGE	TOX	DOSE	SCR	WEIGHT	GENERIC
1	50	1	J	1	1	1	36.0	0.8	66	1
2	21	1	J	1	2	2	29.0	1.1	68	1
3	35	1	J	2	2	2	16.2	0.7	97	1
4	49	2	S	1	1	2	29.0	0.8	83	2
5	38	1	J	2	2	1	16.2	1.4	97	1
6	42	1	S	2	2	2	18.0	1.0	82	2
7	17	1	J	1	2	2	17.4	1.0	64	1
8	20	1	S	2	2	2	17.4	1.0	73	2
9	49	2	J	1	1	2	37.2	0.7	103	1
10	41	2	J	1	2	2	18.6	0.9	58	1
11	20	1	S	2	2	2	18.0	1.1	113	2
12	55	1	S	1	1	2	36.0	0.8	87	2
13	44	2	J	1	1	1	22.4	1.2	59	1
14	23	1	S	2	2	2	39.6	0.8	83	2
15	64	2	S	1	1	2	30.0	0.9	69	2
16	65	1	S	1	1	1	23.2	1.7	106	2
17	23	2	S	1	2	2	16.8	0.9	66	2
18	44	1	S	1	2	2	17.4	1.0	84	2
19	29	2	S	2	1	2	18.0	0.7	56	2
20	32	1	S	1	2	2	18.0	1.0	84	2
21	18	1	S	2	2	2	17.4	0.9	70	2
22	22	1	S	1	1	1	26.1	1.7	69	2
23	43	2	J	2	2	2	18.0	0.8	63	1
24	39	2	S	1	2	2	18.0	0.9	55	2
25	38	2	J	1	1	1	16.0	1.0	112	1

Here's a codebook for the data:

Variable	**Description**
AGE	Age (years)
SEX	1 = male; 2 = female
MANUF	Manufacturer of the drug: Smith or Jones
DIAG	Underling diagnosis: 1 = leukemia; 2 = lymphoma
STAGE	Stage of disease: 1 = relapse; 2 = remission
TOX	Did cerebellar toxicity occur?: 1 = yes; 2 = no
DOSE	Dose of drug (grams /meters2)
SCR	Serum creatinine (mg/dl)
WEIGHT	Body weight (kg)

Source: Jolson, H. M., Bosco, L., Bufton, M. G., Gerstman, B. B., Rinsler, S. S., Williams, E., et al. (1992). Clustering of adverse drug events: analysis of risk factors for cerebellar toxicity with high-dose cytarabine. *JNCI, 84*, 500–505.

1.11 ***Variable types 2.*** Here is more practice in classifying variables as quantitative, ordinal, or categorical.

(a) White blood cells per deciliter of whole blood

(b) Leukemia rates in geographic regions (cases per 100,000 people)

(c) Presence of type II diabetes mellitus (yes or no)

(d) Body weight (kilograms)
(e) Low-density lipoprotein level (mg/dl)
(f) Grade in a course coded: A, B, C, D, or F
(g) Religious identity coded 1 = Protestant, 2 = Catholic, 3 = Muslim, 4 = Jewish, 5 = Atheist, 6 = Buddhist, 7 = Hindu, 8 = Other
(h) Blood cholesterol level classified as either 1 = hypercholesterolemic, 2 = borderline hypercholesterolemic, 3 = normocholesterolemic
(i) Course credit (pass or fail)
(j) Ambient temperature (degrees Fahrenheit)
(k) Type of life insurance policy: 1 = none, 2 = term, 3 = endowment, 4 = straight life, 5 = other
(l) Satisfaction: 1 = very satisfied, 2 = satisfied, 3 = neutral, 4 = unsatisfied, 5 = very unsatisfied
(m) Movie review rating: 1 star, 1½ stars, 2 stars, 2½ stars, 3 stars, 3½ stars, 4 stars
(n) Treatment group: 1 = active treatment, 2 = placebo