Statistics:

Statistics is a very broad subject, with applications in a vast number of different fields. In generally one can say that statistics is the methodology for collecting, analyzing, interpreting and drawing conclusions from information. Putting it in other words, statistics is the methodology which scientists and mathematicians have developed for interpreting and drawing conclusions from collected data.

What is meteorological Statistics?

Meteorological Statistics is concerned with the use of statistical methods in the atmospheric sciences, specifically, in the various specialties within meteorology and climatology. Students often resist statistics, and the subject is perceived by many to be the epitome of dullness. Before the available computers this negative view had some basis, at least with respect to applications of statistics involving the analysis of data. Performing hand calculations, even with the aid of a scientific pocket calculator, was, tedious, and time-consuming. Although the capacity of ordinary personal computers on today's desktops is comparable to the fastest mainframe computers of 30 years ago, many people have not noticed that the age of computational drudgery in statistics has long passed. In fact, some important and powerful statistical techniques were not even practical before the abundant availability of computing. Even when liberated from hand calculations, statistics is sometimes perceived as dull by people who do not appreciate its relevance to scientific problems.

Fundamentally, statistics is concerned with uncertainty. Evaluating uncertainty, as well as making inferences and forecasts in the face of uncertainty, are all parts of statistics. It should not be surprising, then, that statistics has many roles to play in the atmospheric sciences, since it is the uncertainty in atmospheric behavior that makes the atmosphere interesting. For example, many people are fascinated with weather forecasting. Weather forecasting remains interesting precisely because of the uncertainty that is intrinsic to the problem. If it were possible to make perfect forecasts even one day into the future (i.e., if there were no uncertainty involved), the practice of meteorology would be very dull.

Uncertainty about the Atmosphere

Uncertainty about the Atmosphere statistics is the notion of uncertainty. If atmospheric processes were constant, or periodic, describing them mathematically would be easy. Weather forecasting would also be easy, and meteorology would be boring. The atmosphere exhibits variations and fluctuations that are irregular. This uncertainty is the driving force behind the collection and analysis of the large data. It also implies that weather forecasts are inescapably uncertain. The weather forecaster predicting a particular temperature on the following day is not at all surprised (and perhaps is even pleased) if the subsequently observed temperature is different by a degree or two. In order to deal quantitatively with uncertainty it is necessary to employ the tools of probability, which is the mathematical language of uncertainty.

Before reviewing the basics of probability, it is worthwhile to examine why there is uncertainty about the atmosphere. After all, we now have large computer models that represent the physics of the atmosphere. Such models are used routinely for forecasting the future evolution of the atmosphere. These models are deterministic: They do not represent uncertainty. Once supplied with a particular initial atmospheric state (e.g., winds, temperatures, and humidities) and boundary forcings (e.g., solar radiation and, for some atmospheric models, such fields as sea-surface temperatures), each will produce a single particular result. Rerunning the model with the same inputs will not change that result. In principle, these atmospheric models could provide forecasts with no uncertainty, for two reasons. First, even though the models can be very impressive and give quite good approximations to atmospheric behavior, they are not complete and true representations of the governing physics. An important and essentially unavoidable cause of this problem is that some relevant physical processes operate on scales too small to be represented by these models. Even if all the relevant physics could be included in atmospheric models, however, we still could not escape the uncertainty because of what has come to be known as dynamical chaos. This is a problem "discovered" by an atmospheric scientist (Lorenz, 1963), and it has become the death knell for the dream of perfect (uncertainty-free) weather forecasts. Lorenz (1993) has also provided a very readable introduction to this subject. Simply and roughly put, the time evolution of a nonlinear, deterministic dynamical system (e.g., the equations of atmospheric motion, or of the atmosphere itself) is very sensitive to the initial conditions of the system. If two realizations of such a system are started from two only very slightly different initial conditions, the two solutions will eventually diverge markedly.

For the case of weather forecasts, imagine that one of these systems is the real atmosphere and the other is a perfect mathematical model of the physics governing the atmosphere. Since the atmosphere is always incompletely observed, it will never be possible to start the mathematical model in exactly the same state as the real system. So even if the model is perfect, it will still be impossible to calculate what the atmosphere will do indefinitely far into the future. Therefore, deterministic forecasts of future atmospheric behavior will always be uncertain, and probabilistic methods will always be needed to adequately describe that behavior. Even if the atmosphere is not fundamentally a random system, for many practical purposes it might as well be. This thought has been put very nicely by Zeng e t al. (1993), who observe "Just as relativity eliminated the Newtonian illusion of absolute space and time, and as quantum theory eliminated the Newtonian and Einsteinian dream of a controllable measurement process, chaos eliminates the Laplacian fantasy of long-term deterministic predictability." Finally, it is worth noting that randomness is not a state of "unpredictability," or "no information," as is sometimes thought. Rather, random means "not precisely predictable." For example, the amount of precipitation that will occur tomorrow where you live is a random quantity, not known to you today. However, a simple statistical analysis of climatological precipitation records at your location would yield relative frequencies of precipitation amounts that would provide substantially more information about tomorrow's precipitation at your location than I have as I sit writing this sentence. A still less uncertain idea of tomorrow's rain might be available to you in the form of a weather forecast. Reducing uncertainty about random meteorological events is the purpose of weather forecasts. Furthermore, statistical methods allow estimation of the precision of predictions, which can itself be valuable information.

Descriptive and Inferential Statistics

It is convenient, although somewhat arbitrary, to divide statistics into two broad areas: descriptive statistics and inferential statistics. Both are relevant to the atmospheric sciences. Descriptive statistics relates to the organization and summarization of data. The atmospheric sciences are awash with data. Worldwide, operational surface and upper-air observations are routinely taken at thousands of locations in support of weather forecasting activities. These are supplemented with aircraft, radar, profiler, and--increasingly--satellite data. Observations of the atmosphere specifically for research purposes are less widespread, but often involve very dense sampling in time and space. In addition, models of the atmosphere consisting of numerical integration of equations describing atmospheric dynamics produce yet more numerical output for both operational and research purposes. As a consequence of these activities, one is often confronted with extremely large batches of numbers that, we hope, contain information about natural phenomena of interest. It can be a nontrivial task just to make some preliminary sense of such data sets. It is typically necessary to organize the raw data, and to choose and implement appropriate summary representations. When the individual data values are too numerous to be grasped individually, a summary that nevertheless portrays important aspects of their variationsma statistical model--can be invaluable in understanding the data. It may be worth emphasizing that it is not the purpose of descriptive data analyses to play with numbers. Rather, these analyses are undertaken because it is known, suspected, or hoped that the data contain information about a natural phenomenon of interest. Inferential statistics is traditionally understood to consist of methods and procedures used to draw conclusions regarding underlying processes that generate the data. Pedder (1987) express this point prosaically, when they state that statistics is "the art of persuading the world to yield information about itself." There is a kernel of truth here: Our physical understanding of atmospheric phenomena comes in part through statistical manipulation and analysis of data. In the context of the atmospheric sciences it is probably sensible to interpret inferential statistics a bit more broadly as well, and include statistical weather forecasting. By now this important field has a long tradition, and it is an integral part of operational .

Organizing and Graphing Data.

This chapter explains how to organize and display data using tables and graphs. We will learn how to prepare frequency distribution tables for data; how to construct bar graphs, histogram and polygons for data.

Raw data

Raw data is the data recorded in the sequence in which they are collected and before they are processed or ranked.

Frequency Distributions

Frequency distribution for data lists all categories and number of elements that belong to each of the categories.

Relative Frequency and percentage distributions.

Relative frequency of category is obtained by dividing the frequency of the category by the sum of all frequencies. A relative frequency distribution lists the relative frequencies for all categories.

$$\text{Relative frequency of a category} = \frac{\text{Frequency of that category}}{\text{Sum of all frequencies}}.$$

The percentage for a category is obtained by multiplying the relative frequency of that category by 100. A percentage distribution lists the percentages for all categories.

$$\text{percentage} = (\text{Relative frequency}) \cdot 100$$

Cumulative Frequency.

The cumulative frequency of data class is the number of data elements in that class and all previous classes (can be done ascending or descending).

Definition

Class Boundary.

The class boundary is given by the midpoint of the upper limit of one class and the lower limit of the next class.

The difference between the two boundaries of a class gives the class width. The class width is also called the class size.

$$\text{class width} = \text{upper boundary} - \text{lower boundary}.$$

The class midpoint or mark is obtained by dividing the sum of the two limits (or the two boundaries) of a class by 2.

lower class limit.
is the smallest data value that can be included in the class.

upper class limit.
is the largest data value that can be included in the class.

The list shows the average temperature (°K) for 30 cities, make a cumulative frequency table of the data.

30 , 35 , 38 , 40 , 42 , 34 , 39 , 41 , 34 , 37
35 , 39 , 41 , 42 , 40 , 37 , 36 , 34 , 38 , 40
41 , 39 , 40 , 36 , 30 , 32 , 33 , 35 , 38 , 39 .

classe width is 2 .

solution :

Sort the data from smallest to largest.

30 , 30 , 32 , 33 , 34 , 34 , 34 , 35 , 35 , 35
36 , 36 , 37 , 37 , 38 , 38 , 38 , 39 , 39 , 39
39 , 40 , 40 , 40 , 40 , 41 , 41 , 41 , 42 , 42 .

Class width is so, 30 – 32
32 – 34
⋮
42 – 44

classes	tally	frequency	Cumulative frequency
30 – 32	||	2	2
32 – 34	||	2	4
34 – 36	~~|||||~~ |	6	10
36 – 38	||||	4	14
38 – 40	~~|||||~~ ||	7	21
40 – 42	~~|||||~~ ||	7	28
42 – 44	||	2	30

$\Sigma f_i = 30$

from the table find the frequency for each class?

42, 62, 46, 54, 41, 37, 54, 44, 32, 45
47, 50, 58, 49, 51, 42, 46, 37, 42, 39
54, 39, 51, 58, 47, 64, 43, 48, 49, 48
49, 61, 41, 40, 58, 49, 59, 57, 57, 34
56, 38, 45, 52, 46, 40, 63, 41, 51, 41.

size of class is 5.

Classes	Tally	frequency.
30-35	||	2
35-40	~~||||~~ |	6
40-45	~~||||~~ ~~||||~~ ||	12
45-50	~~||||~~ ~~||||~~ ||||	14
50-55	~~||||~~ |	6
55-60	~~||||~~ |	6
60-65	||||	4
		50.

The size of class can be determine by the formula

$$\text{number of class intervals} = \frac{\text{Range}}{1 + 3.322 \log N} = \frac{64-32}{6.64}$$

$$= 4.8 = 5$$

where N is number of observation.