STATISTICS OF PALEOMAGNETIC DATA

The need for statistical analysis of paleomagnetic data has become apparent from the preceding chapters. For instance, we require a method for determining a mean direction from a set of observed directions**.** This method should provide some measure of uncertainty in the mean direction. Additionally, we need methods for testing the significance of field tests of paleomagnetic stability. Basic statistical methods for analysis of directional data are introduced in this chapter. It is sometimes said that statistical analyses are used by scientists in the same manner that a drunk uses a light pole: more for support than for illumination. Although this might be true, statistical analysis is fundamental to any paleomagnetic investigation. An appreciation of the basic statistical methods is required to understand paleomagnetism.

Most of the statistical methods used in paleomagnetism have direct analogies to “planar” statistics. We begin by reviewing the basic properties of the normal distribution (Gaussian probability density function). This distribution is used for statistical analysis of a wide variety of observations and will be familiar to many readers. Statistical analysis of directional data are developed by analogy with the normal distribution. Although the reader might not follow all aspects of the mathematical formalism, this is no cause for alarm. Graphical displays of functions and examples of statistical analysis will provide the more important intuitive appreciation for the statistics.

THE NORMAL DISTRIBUTION

Any statistical method for determining a mean (and confidence limit) from a set of observations is based on a *probability density function*. This function describes the distribution of observations for a hypothetical, infinite set of observations called a *population*. The *Gaussian probability density function* (*normal distribution*) has the familiar bell-shaped form shown in Figure 6.1. The meaning of the probability density function $f(z)$ is that the proportion of observations within an interval of width dz centered on z is $f(z)\,dz$.

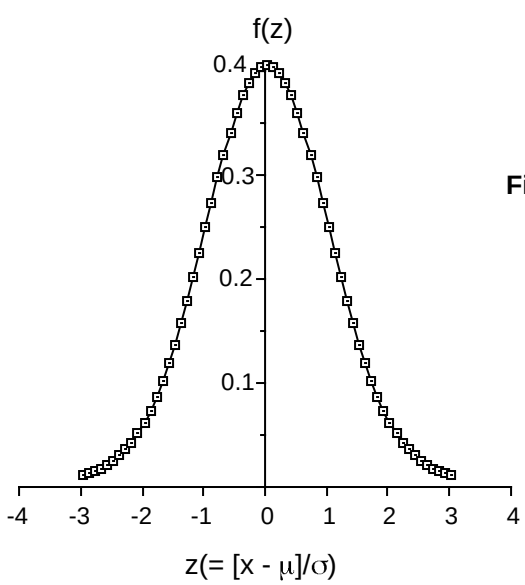

Figure 6.1 The Gaussian probability density function (normal distribution, Equation (6.1)). The proportion of observations within an interval dz centered on z is $f(z)dz$; x = measured quantity; μ = true mean; σ= standard deviation.

The normal distribution is given by

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \tag{6.1}$$

where

$$z = \frac{(x-\mu)}{\sigma}$$

x is the variable measured, μ is the *true mean*, and σ is the *standard deviation*. The parameter μ determines the value of x about which the distribution is centered, while σ determines the width of the distribution about the true mean. By performing the required integrals (computing area under curve $f(z)$), it can be shown that 68% of the readings in a normal distribution are within σ of μ, while 95% are within 2σ of μ.

The usual situation is that one has made a finite number of measurements of a variable x. In the literature of statistics, this set of measurements is referred to as a *sample*. By using the methods of Gaussian statistics, one is supposing that the observed sample has been drawn from a population of observations that is normally distributed. The true mean and standard deviation of the population are, of course, unknown. But the following methods allow estimation of these quantities from the observed sample.

The best estimate of the true mean (μ) is given by the mean, m, of the measured values:

$$m = \frac{\sum_{i=1}^{n} x_i}{n} \tag{6.2}$$

where n is the number of measurements, and x_i is an individual measurement.

The *variance* of the sample is

$$\mathrm{var}(x) = \frac{\sum_{i=1}^{n} (x_i - m)^2}{(n-1)} = s^2 \tag{6.3}$$

The estim*ated standard deviation* of the sample is s and provides the best estimate of the standard deviation (σ) of the population from which the sample was drawn. The *estimated standard error of the mean,* Δm, is given by

$$\Delta m = \frac{s}{\sqrt{n}} \tag{6.4}$$

Some intuitive understanding of the effects of sampling errors can be gotten by the following theoretical results. For multiple samples drawn from the same normal distribution, 68% of the sample means will be within $\sigma / \sqrt{n}$ of μ and 95% of sample means will be within $2\sigma / \sqrt{n}$ of μ. So the sample means are themselves normally distributed about the true mean with standard deviation $\sigma / \sqrt{n}$.

The estimated standard error of the mean, Δm, provides a *confidence limit* for the calculated mean. Of all the possible samples that can be drawn from a particular normal distribution, 95% have means, m, within $2\Delta m$ of μ. (Only 5% of possible samples have means that lie farther than $2\Delta m$ from μ.) Thus the 95% confidence limit on the calculated mean, m, is $2\Delta m$, and we are 95% certain that the true mean of the population from which the sample was drawn lies within $2\Delta m$ of m.

It should be appreciated and emphasized that the estimated standard deviation, s, does not fundamentally depend upon the number of observations, n. However, the estimated standard error of the mean, Δm, does depend on n and decreases as $1/\sqrt{n}$. Because we imagine each sample as having been drawn from

a normal distribution with a definite true mean and standard deviation, it follows that our best estimate of the standard deviation does not depend on the number of observations in the sample. However, it is also reasonable that a larger sample will provide a more precise estimation of the true mean, and this is reflected in the smaller confidence limit with increasing *n*.

THE FISHER DISTRIBUTION

A probability density function applicable to paleomagnetic directions was developed by the British statistician R. A. Fisher and is known as the *Fisher distribution*. Each direction is given unit weight and is represented by a point on a sphere of unit radius. The Fisher distribution function $P_{dA}(\theta)$ gives the probability per unit angular area of finding a direction within an angular area, *dA*, centered at an angle θ from the true mean. The angular area, *dA*, is expressed in steredians, with the total angular area of a sphere being 4π steredians. Directions are distributed according to the probability density function

$$P_{dA}(\theta) = \frac{\kappa}{4\pi \sinh(\kappa)} \exp(\kappa \cos\theta) \tag{6.5}$$

where θ is the angle from true mean direction (= 0 at true mean), and κ is the *precision parameter*. The notation $P_{dA}(\theta)$ is used to emphasize that this is a probability per unit angular area.

The distribution of directions is azimuthally symmetric about the true mean. κ is a measure of the concentration of the distribution about the true mean direction. κ is 0 for a distribution of directions that is uniform over the sphere and approaches ∞ for directions concentrated at a point. $P_{dA}(\theta)$ is shown in Figure 6.2a for κ = 5, 10, and 50. As expected from the definition, the Fisher distribution is maximum at the true mean (θ = 0), and, for higher κ, the distribution is more strongly concentrated towards the true mean.

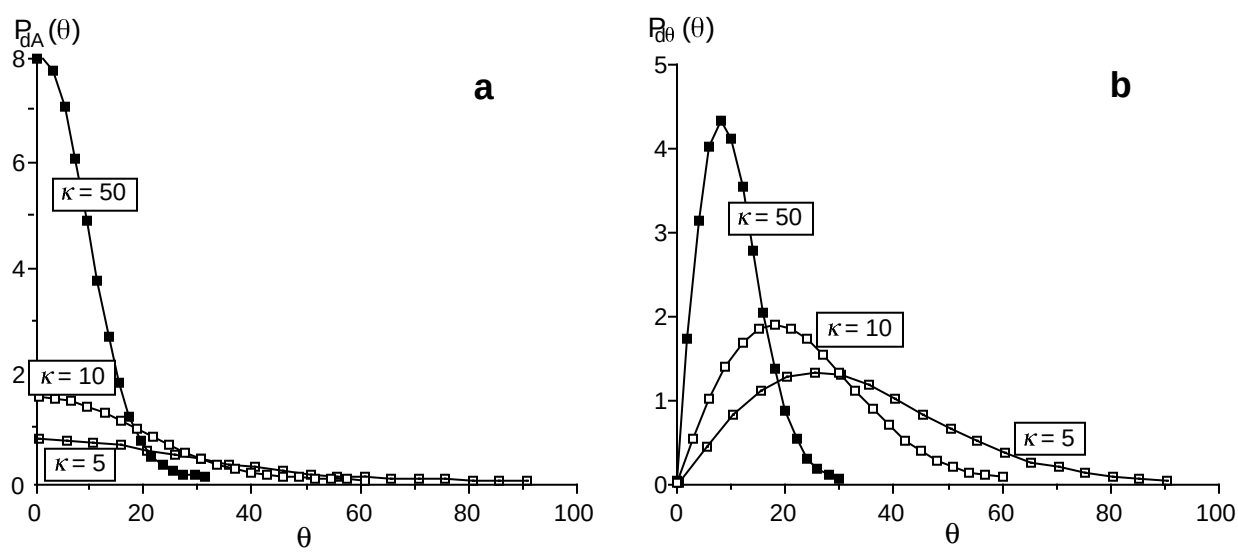

Figure 6.2 The Fisher distribution. (**a**) $P_{dA}(\theta)$ is shown for κ = 50, κ = 10, and κ = 5. $P_{dA}(\theta)$ is the probability *per unit angular area* of finding a direction within an angular area, *dA*, centered at an angle θ from the true mean; $P_{dA}(\theta)$ is given by Equation (6.5); κ = precision parameter. (**b**) $P_{d\theta}(\theta)$ is shown for κ = 50, κ = 10, and κ = 5. $P_{d\theta}(\theta)$ is the probability of finding a direction within a band of width $d\theta$ between θ and $\theta + d\theta$. $P_{d\theta}(\theta)$ is given by Equation (6.8).

If ξ is taken as the azimuthal angle about the true mean direction, the probability of a direction within an angular area, *dA*, can be expressed as

$$P_{dA}(\theta) dA = P_{dA}(\theta) \sin(\theta)\, d\theta \;\; d\xi \tag{6.6}$$

The sin (θ) term arises because the area of a band of width $d\theta$ varies as sin (θ). It should be understood that the Fisher distribution is normalized so that

$$\int_{\xi=0}^{2\pi}\int_{\theta=0}^{\pi} P_{dA}(\theta)\sin(\theta)\,d\theta\,d\xi = 1.0 \tag{6.7}$$

Equation (6.7) simply indicates that the probability of finding a direction somewhere on the unit sphere must be 1.0. The probability $P_{d\theta}(\theta)$ of finding a direction in a band of width $d\theta$ between θ and $\theta + d\theta$ is given by

$$P_{d\theta}(\theta) = \int_{\xi=0}^{2\pi} P_{dA}(\theta)dA = 2\pi\, P_{dA}(\theta)\sin(\theta)\,d\theta$$

$$= \frac{\kappa}{2\sinh(\kappa)}\exp(\kappa\cos\theta)\sin\theta\,d\theta \tag{6.8}$$

This probability (for κ = 5, 10, and 50) is shown in Figure 6.2b, where the effect of the sin (θ) term is apparent.

The angle from the true mean within which a chosen percentage of directions lie can also be calculated from the Fisher distribution. The angle within which 50% of directions lie is

$$\theta_{50} = \frac{67.5°}{\sqrt{\kappa}} \tag{6.9}$$

and is analogous to the *interquartile* of the normal distribution. The angle analogous to the standard deviation of the normal distribution is

$$\theta_{63} = \frac{81°}{\sqrt{\kappa}} \tag{6.10}$$

This angle is often called the *angular standard deviation.* But notice that only 63% of directions lie within θ_{63} of the true mean direction, while 68% of observations in a normal distribution lie within σ of μ. The final critical angle of interest is that containing 95% of directions and given by

$$\theta_{95} = \frac{140°}{\sqrt{\kappa}} \tag{6.11}$$

Computing a mean direction

The above equations apply to a population of directions that are distributed according to the Fisher probability density function. But we commonly have only a small sample of directions (e.g., a data set of ten directions) for which we must calculate (1) a mean direction, (2) a statistic indicating the amount of scatter of the directions (analogous to the estimated standard deviation in Gaussian statistics), and (3) a confidence limit for the calculated mean direction (analogous to the estimated standard error of the mean). By employing the Fisher distribution, the following calculation scheme can provide the desired quantities.

The mean of a set of directions is found simply by vector addition (Figure 6.3). To compute the mean direction from a set of N unit vectors, the direction cosines of the individual vectors are first determined by

$$l_i = \cos I_i \cos D_i \qquad m_i = \cos I_i \sin D_i \qquad n_i = \sin I_i \tag{6.12}$$

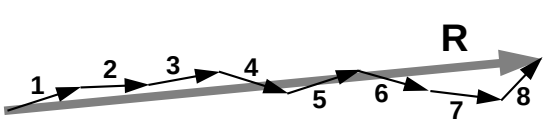

Figure 6.3 Vector addition of eight unit vectors to yield resultant vector ***R***.

where D_i is the declination of the i vector; I_i is the inclination of the i vector; and l_i, m_i, and n_i are the direction cosines of the i vector with respect to north, east, and down directions. The direction cosines, l, m, and n, of the mean direction are given by

$$l = \frac{\sum_{i=1}^{N} l_i}{R} \qquad m = \frac{\sum_{i=1}^{N} m_i}{R} \qquad n = \frac{\sum_{i=1}^{N} n_i}{R} \tag{6.13}$$

where $\boldsymbol{R}$ is the resultant vector with length R given by

$$R^2 = \left(\sum_{i=1}^{N} l_i\right)^2 + \left(\sum_{i=1}^{N} m_i\right)^2 + \left(\sum_{i=1}^{N} n_i\right)^2 \tag{6.14}$$

The relationship of $\boldsymbol{R}$ to the N individual unit vectors is shown in Figure 6.3. R is always $\leq N$ and approaches N only when the vectors are tightly clustered. From the mean direction cosines given by Equations (6.13) and (6.14), the declination and inclination of the mean direction can be computed by

$$D_m = \tan^{-1}\left(\frac{m}{l}\right) \quad \text{and} \quad I_m = \sin^{-1}(n) \tag{6.15}$$

Dispersion estimates

Having calculated the mean direction, the next objective is to determine a statistic that can provide a measure of the dispersion of the population of directions from which the sample data set was drawn. One measure of the dispersion of a population of directions is the precision parameter, κ. From a finite sample set of directions, κ is unknown, but a best estimate of κ can be calculated by

$$k = \frac{N-1}{N-R} \tag{6.16}$$

Examination of Figure 6.3 provides intuitive insight into Equation (6.16). It can readily be seen that k increases as R approaches N for a tightly clustered set of directions.

By direct analogy with Gaussian statistics (Equation (6.3)), the angular variance of a sample set of directions is

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N} \Delta_i^2 \tag{6.17}$$

where Δ_i is the angle between the i direction and the calculated mean direction. The *estimated angular standard deviation* (often called *angular dispersion*) is simply s. As expected from Equation (6.10), s can be approximated by

$$s \approx \frac{81^\circ}{\sqrt{k}} \tag{6.18}$$

Another statistic, δ, which is often used as a measure of angular dispersion (and is often called the angular standard deviation) is given by

$$\delta = \cos^{-1}\left(\frac{R}{N}\right) \tag{6.19}$$

The advantages of using δ for an estimated angular standard deviation are ease of calculation and the intuitive appeal (e.g., Figure 6.3) that δ decreases as R approaches N and the set of directions becomes more tightly clustered. In practice (at least for reasonable values of $N \geq 10$),

$$s \approx \delta \approx \frac{81°}{\sqrt{k}} \tag{6.20}$$

Although s from Equation (6.17) is the rigorously correct estimator of angular standard deviation, all of the above techniques will yield essentially the same result.

In analyzing paleomagnetic directions, it is common to report the statistic k as a measure of *within-site* scatter of directions (from multiple samples of a site). When an analysis is made of *between-site* dispersion of directions (dispersion of mean directions from one site to another), one of the above measures of angular dispersion is usually reported.

A confidence limit

We need a method for determining a confidence limit for the calculated mean direction. This confidence limit is analogous to the estimated standard error of the mean Δm of Gaussian statistics. For Fisher statistics, the confidence limit is expressed as an angular radius from the calculated mean direction. A probability level must be indicated for the confidence limit to be fully defined.

For a directional data set with N directions, the angle $\alpha_{(1-p)}$ within which the unknown true mean lies at confidence level $(1 - p)$ is given by

$$\cos\alpha_{(1-p)} = 1 - \frac{N-R}{R}\left\{\left(\frac{1}{p}\right)^{\frac{1}{N-1}} - 1\right\} \tag{6.21}$$

The usual choice of probability level $(1 - p)$ is 0.95 (= 95%), and the confidence limit is usually denoted as α_{95}. Two convenient approximations (reasonably accurate for both $k \geq 10$ and $N \geq 10$) are

$$\alpha_{63} \approx \frac{81°}{\sqrt{kN}} \quad \text{and} \quad \alpha_{95} \approx \frac{140°}{\sqrt{kN}} \tag{6.22}$$

The α_{63} is analogous to the estimated standard error of the mean, while α_{95} is analogous to two estimated standard errors of the mean.

When we calculate the mean direction, a dispersion estimate, and a confidence limit, we are supposing that the observed data came from random sampling of a population of directions accurately described by the Fisher distribution. But we do not know the true mean of that Fisherian population, nor do we know its precision parameter κ. We can only estimate these unknown parameters. The calculated mean direction of the directional data set is the best estimate of the true mean direction, while k is the best estimate of κ. The confidence limit α_{95} is a measure of the precision with which the true mean direction has been estimated. One is 95% certain that the unknown true mean direction lies within α_{95} of the calculated mean. The obvious corollary is that there is a 5% chance that the true mean lies more than α_{95} from the calculated mean.

Some illustrations

Having buried the reader in mathematical formulations, we present the following illustrations to develop some intuitive appreciation for the statistical quantities. One essential concept is the distinction between statistical quantities calculated from a directional data set and the unknown parameters of the sampled population.

The six synthetic directional data sets illustrated in Figure 6.4 were generated and analyzed in the following manner:

1. A population of directions distributed according to the Fisher probability density distribution was generated by computer. The true mean direction of this Fisherian population was $I = +90°$ (directly downward) and the precision parameter was $\kappa = 20$.
2. This Fisher distribution was randomly sampled 20 times to produce a "synthetic" directional data set with $N = 20$. A total of six such data sets were produced, each being an independent random sampling of the same population of directions. These six data sets are shown on the equal-area projections of Figure 6.4.
3. For each synthetic data set, the following quantities were calculated: (a) mean direction (D_m, I_m), (b) k, and (c) the confidence limit α_{95}. These quantities are also illustrated for each data set in Figure 6.4.

There are several important observations to be taken from this example. Note that the calculated mean direction is never exactly the true mean direction ($I = +90°$). The calculated mean inclination I_m varies from 85.7° to 88.8°, and at least one calculated mean declination falls within each of the four quadrants of the equal-area projection. The calculated mean direction thus randomly dances about the true mean direction and varies from the true mean by between 1.2° and 4.3°.

The calculated k statistic varies considerably from one synthetic data set to another with a range of 17.3 to 27.2 that contains the known precision parameter $\kappa = 20$. The variation of k and differences in angular variance of the data sets are simply due to the vagaries of random sampling. (Techniques for determining confidence limits for k do exist. When applied to these data sets, none of the k values is, in fact, significantly removed from the known value $\kappa = 20$ at 95% confidence. See Suggested Readings for these techniques.)

The confidence limit α_{95} varies from 6.0° to 7.5° and is shown by the stippled oval surrounding the calculated mean direction. For these six directional data sets, none has a calculated mean that is more than α_{95} from the true mean. However, if 100 such synthetic data sets had been analyzed, on average five data sets would have a calculated mean direction removed from the true mean direction by more than the calculated confidence limit α_{95}. That is, the true mean direction would lie <u>outside</u> the circle of 95% confidence, on average, in 5% of the cases.

It is also important to appreciate which statistical quantities are fundamentally dependent upon the number of observations N. Neither the k value (Equation (6.16)) nor the estimated angular deviation s or δ (Equation (6.18) or (6.19)) is fundamentally dependent upon N. These statistical quantities are estimates of the intrinsic dispersion of directions in the Fisherian population from which the data set was sampled. Because that dispersion is not affected by the number of times the population is sampled, the calculated statistics estimating that dispersion should not depend fundamentally on the number of observations N.

However, the confidence limit α_{95} should depend on N; the more individual measurements there are in our sample, the greater must be the precision in estimating the true mean direction. This increased precision should be reflected by a decrease in α_{95} with increasing N. Indeed Equation (6.22) indicates that α_{95} depends approximately on $1/\sqrt{N}$.

Figure 6.5 illustrates these dependences of calculated statistics on number of directions in a data set. The following procedure was used to construct this diagram:

1. A synthetic data set of $N = 30$ was randomly sampled from a Fisherian population of directions with angular standard deviation $\theta_{63} = 15°$ ($\kappa = 29.2$).
2. Starting with the first four directions in the synthetic data set, a subset of $N = 4$ was used to estimate κ and θ_{63} by calculating k and s from Equations (6.16) and (6.20), respectively. In addition, α_{95} (using Equation (6.21)) was calculated. Resulting s and α_{95} values are plotted at $N = 4$ in Figure 6.5.

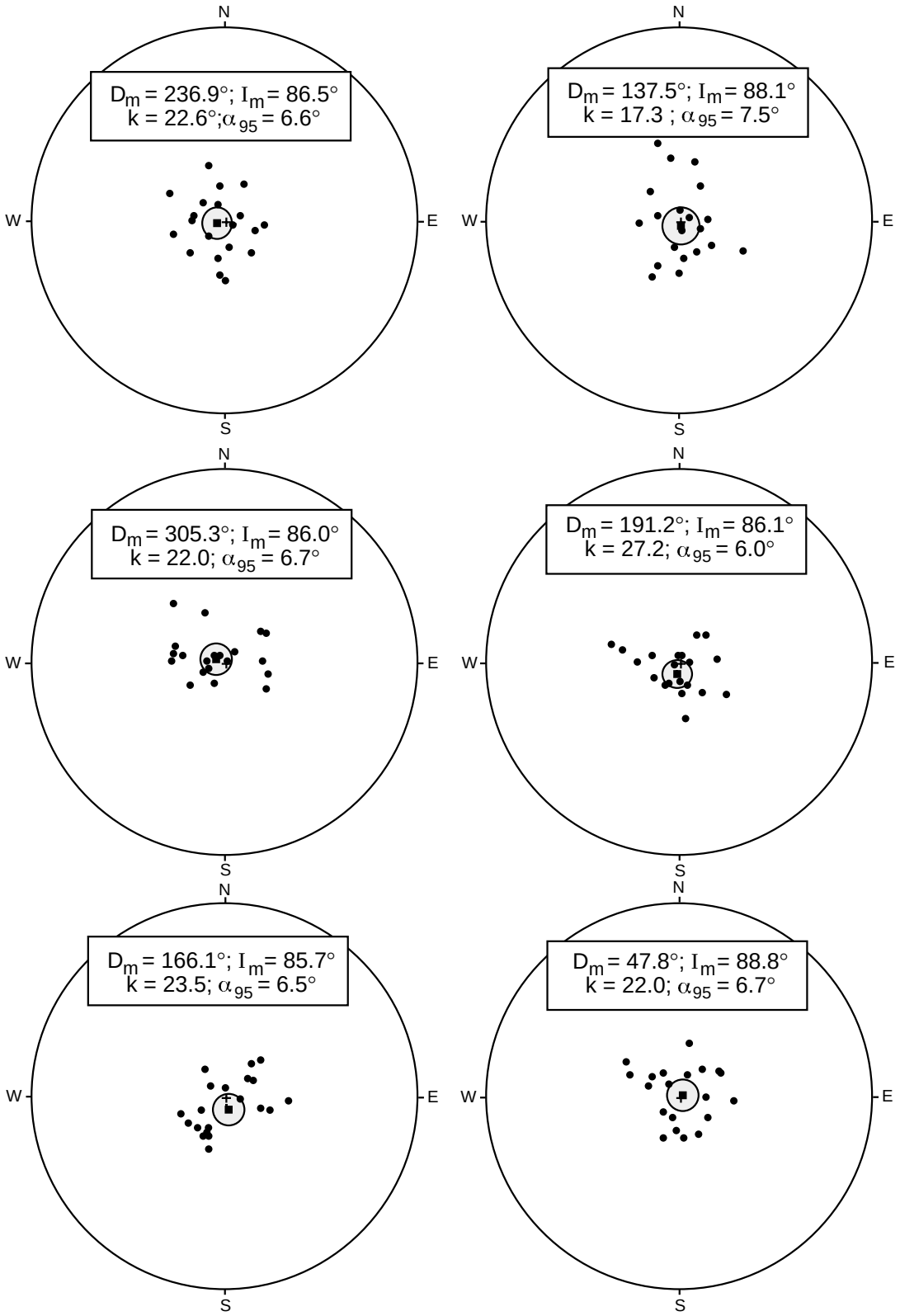

Figure 6.4 Equal-area projections of six synthetic directional data sets, mean directions, and statistical parameters. The data sets were randomly selected from a Fisherian population with true mean direction I = +90° and precision parameter κ = 20; individual directions are shown by solid circles; mean directions are shown by solid squares with surrounding stippled α_{95} confidence limits.

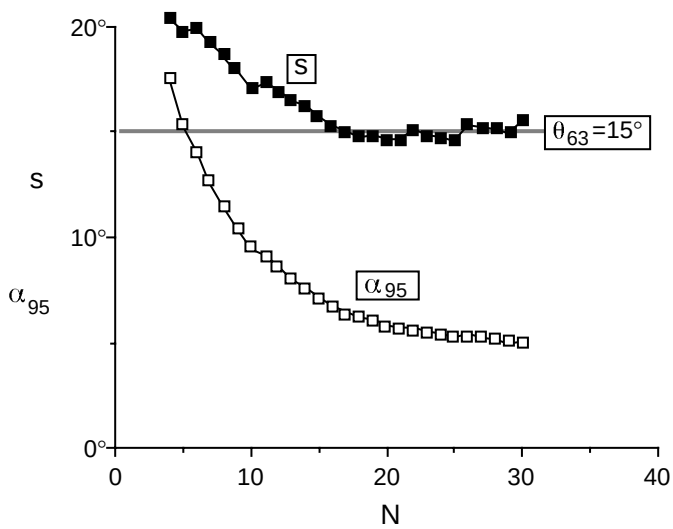

Figure 6.5 Dependence of estimated angular standard deviation, *s*, and confidence limit, α_{95}, on the number of directions in a data set. An increasing number of directions were selected from a Fisherian population of directions with angular standard deviation θ_{63} = 15° (κ = 29.2) shown by the stippled line.

3. For each succeeding value of *N* in Figure 6.5, the next direction from the *N* = 30 synthetic data set was added to the previous subset of directions, continuing until the full *N* = 30 synthetic data set was utilized.

The effects of increasing *N* are readily apparent in Figure 6.5. Although not fundamentally dependent upon *N*, in practice the estimated angular standard deviation, *s*, systematically overestimates the angular standard deviation θ_{63} for values of *N* < 10. (If uncertainties in the calculated values of *s* are considered, it is found that these errors become quite large for *N* < 10.) For *N* > 10, the calculated value of *s* approaches the known angular standard deviation θ_{63} = 15°. As expected, the calculated confidence limit α_{95} decreases approximately as $1/\sqrt{N}$, showing a dramatic decrease in the range $4 \le N \le 10$ and more gradual decrease for *N* > 10.

Another example of the effects of increasing *N* on the calculated statistical quantities is provided in Figure 6.6. The following procedure was used:

1. Two independent synthetic directional data sets of *N* = 50 were randomly selected from a Fisherian population of directions with angular standard deviation θ_{63} = 15°. The true mean direction is vertically down (*I* = +90°).
2. Two subsets of these *N* = 50 data sets were then produced by selecting the first five directions, to yield two sets of *N* = 5, then the first ten directions, to yield two sets of *N* = 10.
3. The mean of each of the six data sets was calculated along with the statistics *k*, *s*, and α_{95} as described in the example above.

The resulting data sets are illustrated in the equal-area projections of Figure 6.6. The results are arranged in two columns: the left-hand column resulting from the first *N* = 50 synthetic data set and the right-hand column resulting from the second *N* = 50 data set. As expected, the calculated mean direction provides a "better" estimation of the true mean as the number of directions, *N*, increases. This effect is most dramatic when the results for *N* = 5 are compared with those for *N* = 10. Notice that the mean directions calculated from the two *N* = 5 data sets are ~15° apart. For the *N* = 10 and *N* = 50 data sets, the calculated mean directions quite closely approximate the true mean direction, and the α_{95} continues to decrease.

Non-Fisherian distributions

The Fisher distribution is azimuthally symmetric about the true mean direction. Occasionally, in analysis of paleomagnetic data, a set of directions that is strongly elliptical in shape is encountered. A statistical method allowing treatment of such data is sometimes required. The *Bingham distribution* (see Suggested Read-

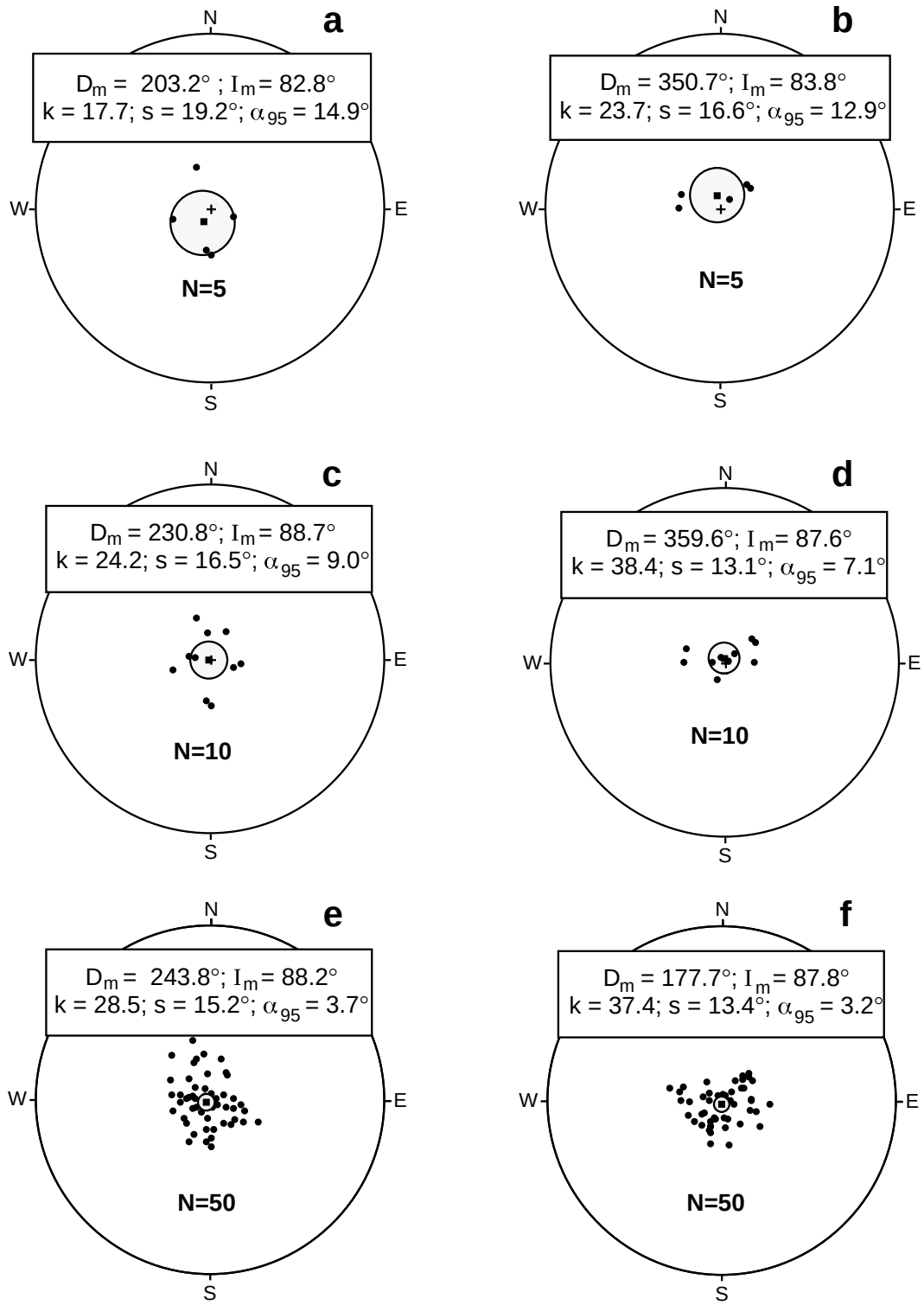

Figure 6.6 Equal-area projections showing mean directions and statistical quantities calculated from increasing numbers of directions drawn from two synthetic directional data sets. The Fisherian population had angular standard deviation θ_{63} = 15° and true mean direction I = +90°; results from one data set are shown in parts (**a**), (**c**), and (**e**) and for the other data set in parts (**b**), (**d**), and (**f**); individual directions are shown by solid circles; mean directions are shown by solid squares with surrounding stippled α_{95} confidence limits.

ings) allows for azimuthal asymmetry and is appropriate for such analyses. Some researchers prefer the Bingham distribution to the Fisher distribution for statistical analysis of all paleomagnetic data. However, the Fisher distribution remains the basis of most statistical treatments in paleomagnetism because (1) Fisher statistics provides fairly straightforward techniques for determining confidence limits, whereas the Bingham distribution does not, and (2) significance tests based on the Fisher distribution are fairly simple and have intuitive appeal, whereas significance tests based on the Bingham distribution are more complex.

SITE-MEAN DIRECTIONS

There are several levels of paleomagnetic data analysis at which mean directions must be calculated:

1. If more than one specimen was prepared from a sample, then ChRM directions for the multiple specimens must be averaged.
2. A site-mean ChRM direction is then calculated from the sample ChRM directions.
3. Generally, a paleomagnetic investigation involves numerous sites within a particular rock unit. These site-mean directions must be averaged to yield either the average ChRM direction or a paleomagnetic pole position from the rock unit.

Straightforward application of the Fisher statistical procedures (Equations (6.12)–(6.15)) is used to calculate both sample-mean directions and site-mean directions. For site-mean directions, R, k and α_{95} are often listed in a table of data. Each site-mean direction ideally provides a record of the geomagnetic field direction at a single point in time. The desired result is that site-mean directions are precisely determined. But it is important to gain an appreciation for the range of results that are actually observed.

Figure 6.7 illustrates examples of sample and site-mean ChRM directions grading from “fantastic” to “poor.” The site-mean result shown in Figure 6.7a is from a single lava flow containing essentially no secondary components of NRM. The ChRM direction for each sample was revealed over a large range of peak AF demagnetization fields. Anchored line-fits from principal component analysis (p.c.a.) were extraordinarily well defined (MAD angles ~1°). For the nine samples collected from this site, the sample ChRM directions are so tightly grouped that they cannot be resolved on the equal-area plot of Figure 6.7a! The site-mean direction has k = 2389 and α_{95} = 1.1°. Such precisely determined site-mean directions are uncommon and generally observed only in very fresh volcanic rocks. Paleomagnetists dream about rocks like this but do not often find them.

In Figure 6.7b, a more typical “good” result from a basalt flow is shown. Minor secondary NRM components (probably lightning-induced IRM) were removed during AF demagnetization to reveal a ChRM direction for each of the seven samples. These sample ChRM directions are reasonably well clustered and yield a site-mean direction with k = 134 and α_{95} = 4.6°. Site-mean directions with $k \approx 100$ and $\alpha_{95} \approx 5°$ would be considered good quality paleomagnetic results and are typical of fresh volcanic rocks. Well-behaved intrusive igneous rocks and red sediments also can yield paleomagnetic data of similar quality.

The clustering of sample ChRM directions shown in Figure 6.7c is only “fair.” These results are from a single bed of Mesozoic red siltstone. Substantial secondary VRM was present in samples from this site, and thermal demagnetization into the 600° to 660°C range was required to isolate the ChRM. Anchored lines (from p.c.a.) fit to four progressive thermal demagnetization results for each sample within the 600° to 660°C range had average MAD ≈ 10°. When plotted on a vector component diagram, the progressive thermal demagnetization data are similar to those of Figure 5.7b. Even with this detailed analysis, the sample ChRM directions are not particularly well clustered. The resulting site-mean direction has k = 42.5 and α_{95} = 11.9°. This site-mean direction was considered acceptable for inclusion in the set of site means used to calculate a paleomagnetic pole. However, this site-mean result was one of the least precise of the 23 site-mean directions considered acceptable.

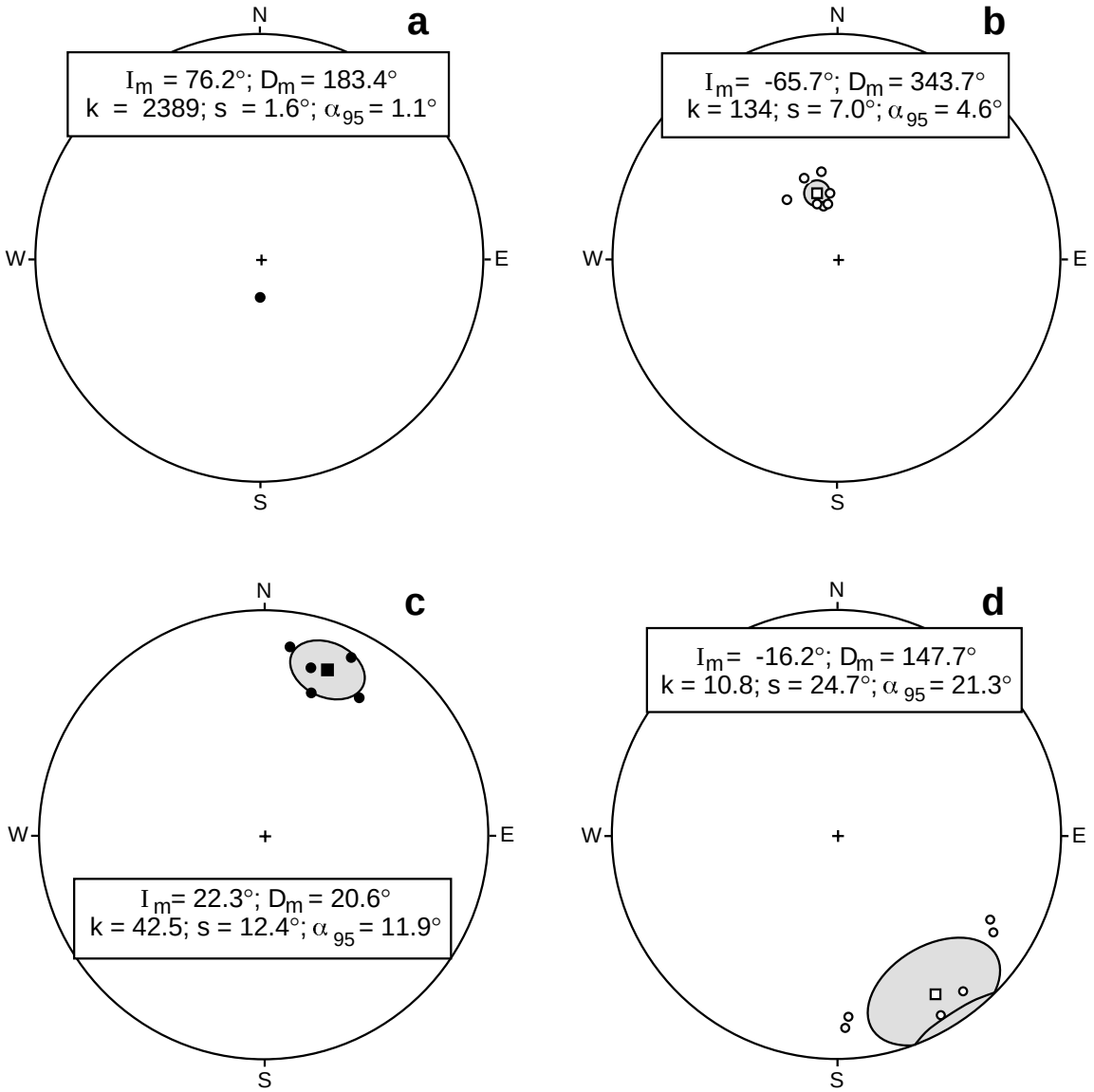

Figure 6.7 Equal-area projections showing examples of sample and site-mean ChRM directions. Sample ChRM directions are shown by circles; site-mean directions are shown by squares with surrounding stippled α_{95} confidence limits; directions in the lower hemisphere are shown by solid symbols; directions in the upper hemisphere are shown by open symbols. (**a**) Unusually well-determined site-mean direction from a single Late Cretaceous lava flow in southern Chile. (**b**) More typical "good" site-mean direction from a Late Cretaceous basalt flow in southern Argentina. (**c**) Site-mean direction determined with "fair" precision from a bed of red siltstone in the Early Jurassic Moenave Formation of northern Arizona. (**d**) A "poor"-quality site-mean direction from a bed of the Late Triassic Chinle Formation in eastern New Mexico.

In Figure 6.7d, "poor"-quality results obtained from a site in Mesozoic red sediment are shown. Despite thermal demagnetization at numerous temperatures and analysis of progressive demagnetization data using p.c.a., the ChRM directions for samples from this site are scattered. The site-mean direction is correspondingly poorly determined. Most paleomagnetists would regard the results from this site as unacceptable for inclusion in a set of site means from which a paleomagnetic pole might be determined. However, these results might still be useful for determination of polarity of ChRM.

Although no firm criteria exist for acceptability of paleomagnetic data, within-site $k > 30$ and $\alpha_{95} < 15°$ would generally be regarded as minimally acceptable site-mean results from which a paleomagnetic pole could be determined. The above examples illustrate that precisely determined site-mean directions (minimal *within-site dispersion*) are desired. The situation for dispersion of site-mean directions (*between-site dispersion*) is considerably more complex. Let's defer consideration of this subject until techniques for calculation of paleomagnetic poles are presented in the next chapter.

SIGNIFICANCE TESTS

From examples of field tests of paleomagnetic stability given in Chapter 5, it is evident that techniques for quantitative evaluation of those tests are required. We must be able to give quantitative answers to such questions as the following: (1) Are two paleomagnetic directions significantly different from one another? (2) Does a set of site-mean directions pass the bedding-tilt test, as evidenced by significantly improved clustering of directions following structural correction? Quantitative evaluations of these questions require *statistical significance tests*.

There are two fundamental principles of statistical significance tests that are important to the proper interpretation:

1. Tests are generally made by comparing an observed sample with a *null hypothesis*. For example, in comparing two mean paleomagnetic directions, the null hypothesis is that the two mean directions are separate samples from the same population of directions. (This is the same as saying that the samples were not, in fact, drawn from different populations with distinct true mean directions.) Significance tests do not <u>prove</u> a null hypothesis but only show that observed differences between the sample and the null hypothesis are unlikely to have occurred because of sampling errors. In other words, there is <u>probably</u> a real difference between the sample and the null hypothesis, indicating that the null hypothesis is <u>probably</u> incorrect.

2. Any significance test must be applied by using a *level of significance*. This is the probability level at which the differences between a set of observations and the null hypothesis may have occurred by chance. A commonly used significance level is 5%. In Gaussian statistics, when testing an observed sample mean against a hypothetical population mean μ (the null hypothesis), there is only a 5% chance that μ is more than $2\Delta m$ from the mean, m, of the sample. If m differs from μ by more than $2\Delta m$, m is said to be "statistically significant from μ at the 5% level of significance," using proper statistical terminology. However, the corollary of the actual significance test is often what is reported by statements such as "m is distinct from μ at the 95% confidence level." The context usually makes the intended meaning clear, but be careful to practice safe statistics.

An important sidelight to this discussion of level of significance is that too much emphasis is often put on the 5% level of significance as a magic number. Remember that we are often performing significance tests on data sets with a small number of observations. Failure of a significance test at the 5% level of significance means only that the observed differences between sample and null hypothesis cannot be shown to have a probability of chance occurrence that is $\leq 5\%$. This <u>does not</u> mean that the observed differences are unimportant. Indeed the observed differences might be significant at a marginally higher level of significance (for instance, 10%) and might be important to the objective of the paleomagnetic investigation.

Significance tests for use in paleomagnetism were developed in the 1950s by Watson and Irving (see Suggested Readings). These versions of the significance tests are fairly simple, and an intuitive appreciation of the tests can be developed through a few examples. Because of their simplicity and intuitive appeal, we investigate these "traditional" significance tests in the development below. However, many of these tests have been revised by McFadden and colleagues (see Suggested Readings) using advances in statistical sampling theory. These revisions are technically superior to the traditional significance tests and are gener-

ally employed in modern paleomagnetic literature. However, they are more complex and less intuitive than the traditional tests.

There are two important points regarding the traditional versions of the significance tests as opposed to the revised versions:

1. Results of these versions of the significance tests differ only when the result is close to the critical value (at a specified significance level). If a result using the traditional version of the appropriate significance test just misses a critical value for being significant at the 5% significance level, it is worthwhile reformulating the test using the revised approach.
2. The revised significance tests are generally more "lenient" than the traditional tests. Results that are significant using the traditional tests will also be significant using the revised test. But some results that were not significant at the 5% significance level according to the traditional test might, in fact, be significant using the revised test.

Comparing directions

A very simple form of significance test is used to determine whether the mean of a directional data set is distinguishable from a known direction. The two directions are distinguishable at the 5% significance level if the known direction falls outside the α_{95} confidence limit of the mean direction. If the known direction is within α_{95} of the calculated mean, the two directions are not distinguishable at the 5% significance level. This test is often used to compare a site-mean direction with the present geomagnetic field or geocentric axial dipole field direction at the sampling locality.

Comparison of two mean directions is more complicated. If the confidence limits surrounding two mean directions do not overlap, the directions are distinct at that level of confidence. For example, if α_{95} circles surrounding two mean directions do not overlap, those directions are distinct at the 5% significance level. Another way of stating this result is that, with 95% probability, the directional data sets yielding these mean directions were selected from different populations with distinct true mean directions. In the case that one or both of the mean directions falls within the α_{95} circle of the other mean direction, the mean directions are not distinct at the 5% significance level.

For intermediate cases in which neither mean direction is contained within the α_{95} circle of the other mean but the α_{95} circles overlap, a further test of significance is required. In this test, the null hypothesis is that the two directional data sets are samplings of the same population and the difference between the means is due to sampling errors.

Consider two directional data sets: one has N_1 directions (described by unit vectors) yielding a resultant vector of length R_1; the other has N_2 directions yielding resultant R_2. The statistic

$$F = (N-2)\frac{(R_1 + R_2 - R)}{(N - R_1 - R_2)} \tag{6.23}$$

must be determined, where

$$N = N_1 + N_2$$

and R is the resultant of all N individual directions. This *F statistic* is compared with tabulated values for 2 and $2(N-2)$ degrees of freedom. If the observed F statistic exceeds the tabulated value at the chosen significance level, then these two mean directions are different at that level of significance.

The tabulated *F-distribution* indicates how different two sample mean directions can be (at a chosen probability level) because of sampling errors. If the calculated mean directions are very different but the individual directional data sets are well grouped, intuition tells us that these mean directions are distinct. The mathematics described above should confirm this intuitive result. With two well-grouped directional data sets with very different means, $(R_1 + R_2) >> R$, R_1 approaches N_1, and R_2 approaches N_2, so that

($R_1 + R_2$) approaches N. With these conditions, the F statistic given by Equation (6.23) will be large and will easily exceed the tabulated value. So this simple intuitive examination of Equation (6.23) yields a sensible result.

Comparison of mean directions is useful for examining the independence of site-mean directions in stratigraphic superposition. Implications of independence of site means will be discussed in the next chapter. Comparison of mean directions is also used in the reversals test for paleomagnetic stability. The mean of the normal-polarity sites is compared with the antipode of the mean of reversed-polarity sites. It is important to realize that this comparison really tests for failure of the reversals test because the null hypothesis is that the two means were selected from the same population. If the mean of normal-polarity sites is distinct from the antipode of the mean of reversed-polarity sites, then there is only a 5% chance that the two directions were samples of the same population (with one true mean direction). Such a result would constitute failure of the reversals test. The desired result ("passage of the reversals test") is that the two means are not distinct at the 5% significance level.

In the illustration of the reversals test shown in Figure 5.16, the mean of the normal-polarity sites is $I_m = 51.7°$, $D_m = 345.2°$, $\alpha_{95} = 5.4°$. The mean of the reversed-polarity sites is $I_m = –51.0°$, $D_m = 163.0°$, $\alpha_{95} = 3.6°$. When the antipode of the reversed-polarity mean is compared with the normal-polarity mean, these means are less than 2° from one another, and each is contained within the α_{95} circle of the other. These directions are not distinct at the 5% significance level, and the site means pass the reversals test.

Test of randomness

When widely scattered directions are observed, the question arises whether the observed directions could have resulted from sampling a random population of directions. (A random population is uniformly distributed over the sphere, has no mean direction, and has $\kappa = 0$.) Even for a directional data set selected from a random population, the observed data set (sample) will rarely have $k = 0$; sampling errors will yield finite R and finite k. But for a given number of directions, N, there is a critical value of R (= R_0) that is unlikely to result from an unusual sampling of a random population. If the 5% significance level is chosen and the observed R exceeds R_0, then there is only a 5% chance that the observed directions resulted from sampling a random population. The corollary is that, with 95% probability, the directional data set resulted from sampling of a nonrandom population with $\kappa > 0$.

The test for randomness is often used in magnetostratigraphic investigations in which site-mean polarity of ChRM is the fundamental information sought. To ensure that a mean ChRM observed at a site is not simply the result of sampling from a random population, the randomness test is applied. For $N = 3$, the critical $R_0 = 2.62$, and $R > 2.62$ is required for 95% probability that the observed mean direction did not result from selection from a random population. In this application, $R > R_0$ is obviously the desired result.

In applying the test for randomness to the conglomerate test for paleomagnetic stability, the desired result is that the ChRM directions observed in clasts of a conglomerate are consistent with selection from a random population. For the conglomerate test shown in Figure 5.14, $N = 7$ and $R = 1.52$. But for $N = 7$, $R_0 = 4.18$ for 5% significance level. Because $R < R_0$, the test for randomness indicates that the observed set of directions could indeed have been selected from a random population. This result constitutes "passage of the conglomerate test."

Comparison of precision (the fold test)

In the fold test (or bedding-tilt test), one examines the clustering of directions before and after performing structural corrections. If the clustering improves on structural correction, the conclusion is that the ChRM was acquired prior to folding and therefore "passes the fold test." The appropriate significance test determines whether the improvement in clustering is statistically significant.

Consider two directional data sets, one with N_1 directions and k_1, and one with N_2 directions and k_2. If we assume (null hypothesis) that these two data sets are samples of populations with the same κ, the ratio k_1 / k_2 is expected to vary because of sampling errors according to

$$\frac{k_1}{k_2} = \frac{\mathrm{var}\left[2(N_2 - 1)\right]}{\mathrm{var}\left[2(N_1 - 1)\right]} \tag{6.24}$$

where var[2(N_2 – 1)] and var[2(N_1 – 1)] are variances with 2(N_2 – 1) and 2(N_1 – 1) degrees of freedom. This ratio should follow the *F*-distribution if the assumption of common κ is correct. Fundamentally, one expects this ratio to be near 1.0 if the two samples were, in fact, selections from populations with common κ. The *F*-distribution tables indicate how far removed from 1.0 the ratio may be before the deviation is significant at a chosen probability level. If the observed ratio in Equation (6.24) is far removed from 1.0, then it is highly unlikely that the two data sets are samples of populations with the same κ. In that case, the conclusion is that the difference in the *k* values is significant and the two data sets were most likely sampled from populations with different κ.

As applied to the fold test, one examines the ratio of *k* after tectonic correction (k_a) to *k* before tectonic correction (k_b). The significance test for comparison of precisions determines whether k_a / k_b is significantly removed from 1.0. If k_a / k_b exceeds the value of the *F*-distribution for the 5% significance level, there is less than a 5% chance that the observed increase in *k* resulting from the tectonic correction is due only to sampling errors. There is 95% probability that the increase in *k* is meaningful and the data set after tectonic correction is a sample of a population with κ larger than the population sampled before tectonic correction. Such a result constitutes a "statistically significant passage of the fold test."

As an example, consider the illustration of the bedding-tilt test shown in Figure 5.12. For the multiple collecting locations in the Nikolai Greenstone, $N = 5$, $k_b = 5.17$, $k_a = 21.51$, and $k_a / k_b = 4.16$. The degrees of freedom are $2(N - 1) = 8$ and the *F*-distribution value $F_{8,8}$ for 5% significance level is 3.44. With ratio $k_a / k_b > F_{8,8}$, the improvement in clustering produced by applying tectonic correction is significant at the 5% level. The bedding-tilt test is thus significant at the 5% significance level, implying that the ChRM was acquired prior to folding.

In examining the possibility of synfolding magnetization, the significance test is applied during a stepwise application of tectonic corrections. Results are usually reported as (1) percent unfolding producing the maximum *k* value and (2) range of unfolding percentage surrounding that producing maximum *k* over which the change in *k* is <u>not</u> significant at the 5% level.

These statistical significance tests are often crucial features of paleomagnetic investigations. Although specific cases can be complex, the background provided above should allow the reader to understand essential elements of the significance tests that are commonly used in paleomagnetism.

SUGGESTED READINGS

INTRODUCTIONS TO STATISTICAL METHODS APPLIED TO DIRECTIONAL DATA:

R. A. Fisher, Dispersion on a sphere, *Proc. Roy. Soc. London*, v. A217, 295–305, 1953.
The classic paper introducing the Fisher distribution.

E. Irving, *Paleomagnetism and Its Applications to Geological and Geophysical Problems*, John Wiley and Sons, New York, 399 pp., 1964.
Chapter 4 contains an excellent introduction to statistical methods in paleomagnetism.

D. H. Tarling, *Palaeomagnetism*: *Principles and Applications in Geology, Geophysics and Archaeology*, Chapman and Hall, 379 pp. 1983.
Chapter 6 presents a discussion of statistical methods.

G. S. Watson, *Statistics on Spheres*, Univ. Arkansas Lecture Notes in the Mathematical Sciences, Wiley, New York, 238 pp., 1983.

N. I. Fisher, T. Lewis, and B. J. J. Embleton, *Statistical Analysis of Spherical Data*, Cambridge, London, 329 pp., 1987.

More advanced texts on statistical analysis of directional data.

SIGNIFICANCE TESTS:

G. S. Watson, Analysis of dispersion on a sphere, *Monthly Notices Geophys. J. Roy. Astron. Soc.*, v. 7, 153–159, 1956.

G. S. Watson and E. Irving, Statistical methods in rock magnetism, *Monthly Notices Geophys. J. Roy. Astron. Soc.*, v. 7, 289–300, 1957.

G. S. Watson, A test for randomness of directions, *Monthly Notices Geophys. J. Roy. Astron. Soc.*, v. 7, 160–161, 1956.

M. W. McElhinny, Statistical significance of the fold test in palaeomagnetism, *Geophys. J. Roy. Astron. Soc.*, v. 8, 338–340, 1964.

The traditional approaches to statistical significance tests applied to paleomagnetism are introduced in these articles.

P. L. McFadden and F. J. Lowes, The discrimination of mean directions drawn from Fisher distributions, *Geophys. J. Roy. Astron. Soc.*, v. 67, 19–33, 1981.

P. L. McFadden and D. L. Jones, The fold test in palaeomagnetism, *Geophys. J. Roy. Astron. Soc.*, v. 67, 53–58, 1981.

Revised treatments of the significance tests.

SOME ADVANCED TOPICS:

T. C. Onstott, Application of the Bingham distribution function in paleomagnetic studies, *J. Geophys. Res.*, v. 85, 1500–1510, 1980.

T. Lewis and N. I. Fisher, Graphical methods for investigating the fit of a Fisher distribution for spherical data, *Geophys. J. Roy. Astron. Soc.*, v. 69, 1–13, 1982.

P. L. McFadden, The best estimate of Fisher's precision parameter *k*, *Geophys. J. Roy. Astron. Soc.*, v. 60, 397–407, 1980.

P. L. McFadden and A. B. Reid, Analysis of palaeomagnetic inclination data, *Geophys. J. Roy. Astron. Soc.*, v. 69, 307–319, 1982.

P. L. McFadden, Determination of the angle in a Fisher distribution which will be exceeded with a given probability, *Geophys. J. Roy. Astron. Soc.*, v. 60, 391–396, 1980.

PROBLEMS

6.1 The rigorous expression for α_{95} is Equation (6.21). A reasonable approximation can be obtained from Equation (6.22). Consider a directional data set with $N = 9$ and $R = 8.6800$. Investigate the accuracy of the approximation given by Equation (6.22) by determining α_{95} for this data set, using both Equation (6.21) and Equation (6.22).

6.2 Consider the table of ChRM directions given below from which a reversals test can be evaluated. Use Equation (6.22) to estimate α_{95} for the mean of the normal-polarity sites and for the mean of the reversed-polarity sites. Then use an equal-area projection to evaluate the reversals test (a simple comparison of the mean directions will suffice in this case).

	N	I_m (°)	D_m (°)	R
Normal-polarity sites:	16	–46.8	26.6	15.4755
Reversed-polarity sites:	12	48.1	215.0	11.4836

6.3 A common response to inspection of Figures 6.2a and 6.2b is that the numbers on the probability axes are too large: "How can $P_{dA}(\theta) \approx 8$ for $\theta = 0°$ and $\kappa = 50$?" But remember that $P_{dA}(\theta)$ is a probability *per unit angular area* of finding a direction within an angular area dA centered at angle θ from the true mean direction (at $\theta = 0°$). To prove that the probabilities shown in Figures 6.2a and 6.2b are not too large but instead are intuitively reasonable, do the following calculation:

a. Determine the angular area, A (in steredians), of a spherical cap that is centered on $\theta = 0°$ and extends to $\theta = 5°$ (the angular radius is 5°). To do this calculation, recall that the angular area of a spherical cap centered on $\theta = 0°$ is given by

$$A = \int_{\xi=0}^{\xi=2\pi} \int_{\theta} \sin\theta \, d\theta d\xi = 2\pi \int_{\theta} \sin\theta \, d\theta$$

where the integral is over the range of θ (0° to 5° in this case).

b. By inspection of Figure 6.2a, you can see that $P_{dA}(\theta)$ does not change dramatically between θ = 0° and θ = 5° (even for κ = 50). So the probability of finding a direction within a spherical cap centered on θ = 0° with angular area A is approximately given by $P_{dA}(0°)A$. Use the value of A determined above and the plot of $P_{dA}(\theta)$ in Figure 6.2a to calculate the approximate probability of finding a direction within a spherical cap centered on θ = 0° and extending to θ = 5° for a population of directions with κ = 50. Does your numerical result make intuitive sense?