Much geological data contain information about orientation in a plane where the orientation itself and not the length of the feature is the important part. Examples abound: Strikes and dips of bedding planes, fault surfaces, joints are some, augmented by glacial striations, sole marks, lineaments in Landsat photographs, and much more. First we must distinguish between directional and oriented data. Directional data can take on unique values in the entire 0-360° range, like drumlins, while oriented data consist of 2-headed vectors: there is a 180° ambiguity inherent in the data. Examples of the oriented data include fault traces and lineaments on maps. Such data require special care in the analysis.

Directional data can be displayed in circular diagrams. One can either plot each direction as a unit vector, or count the number of vectors within a given sector and draw a polar histogram of the distribution.

Unfortunately, the sector diagram as described here is biased in the way it presents the data, but is still the most commonly used type of display for directional data. Consider the area of a sector of width $\Delta \alpha$

$$A = \frac{\pi r^2 \Delta \alpha}{360} \propto r^2$$

We see that the area is proportional to the radius squared, while for a conventional histogram the column area is proportional to height, not height squared. Therefore, we should let the sector
radius be proportional to the square root of the frequency or count, so that the final rose diagram may be have correct area. If this is not done, the larger counts in some sectors will completely swamp smaller sectors due to the \( r^2 \) effect. Thus, small but significant trends may not be detected or simply considered noise.

Before we can statistically analyze orientation data the 180° ambiguity must be accounted for. The simplest way to do this is to double the angles and analyze these angles instead. E.g., if two fault traces are reported as having strikes of 45° and 225°, which means the strikes have the same orientation, we double the angles and find 90° and 450° - 360° = 90°. Statistics derived from the doubled angles can then be divided by two to recover the original orientations.

The dominant direction or mean direction can be found by computing the vector resultant or vector sum of the unit vectors that represent the various directions in the data. Since the coordinates of these vectors are

\[
X_i = \cos \theta_i, \\
Y_i = \sin \theta_i
\]  

we find the coordinates of the resultant \( R \) as

\[
X_r = \sum_{i=1}^{n} \cos \theta_i, \\
Y_r = \sum_{i=1}^{n} \sin \theta_i
\]  

(6.3)

The mean direction \( \bar{\theta} \) is then simply given by

\[
\bar{\theta} = \tan^{-1} \left( \frac{Y_r}{X_r} \right) = \tan^{-1} \left( \frac{\sum_{i=1}^{n} \sin \theta_i}{\sum_{i=1}^{n} \cos \theta_i} \right)
\]  

(6.4)

The magnitude of the resultant should be normalized by the number of vectors before it can be used further. This gives us the normalized length

\[
\bar{R} = \frac{R}{n} = \frac{\sqrt{X_r^2 + Y_r^2}}{n}
\]  

(6.5)

or mean resultant length, which will range from 0 to 1 and is a measure of dispersion analogous to the variance. Since it increases for focused distribution we often convert it to circular variance, i.e.

\[
s_\circ^2 = 1 - \bar{R} = \frac{(n - \bar{R})}{n}
\]  

(6.6)

In order to test various statistical hypothesis about circularly distributed data we need a probability distribution that we can compare to our data. A traditional distribution which has been used extensively is the von Mises distribution, given by

\[
p(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp \left( \kappa \cos (\theta - \mu) \right)
\]  

(6.7)

where \( \kappa \) is a measure of the concentration of the distribution about the mean direction \( \mu \), and \( I_0 \) is the modified Bessel function of the first kind and order zero. (It is there to normalize the cumulative distribution over 360° to unity). \( \kappa \) is obviously related to \( R \) (or \( s_\circ^2 \)) and can be
derived if we assume that the data are a sample from a population having a von Mises distribution. Tables exist (e.g., Appendix 2.12) that relate $\kappa$ and $R$.

**Testing for Randomness**

The simplest test on circular data is to check whether the directional observations are random, that is, there is no preferred direction. In terms of the von Mises distribution (6.7), this means that $\kappa$ must be zero. Hence, the test becomes

$$H_0: \kappa = 0 \quad \text{vs.} \quad H_1: \kappa > 0$$

If the data comes from a circular distribution ($\kappa = 0$), we expect a small $R$. Lord Rayleigh devised this test which gives critical values for $R$ depending on $\nu = n - 1$ and the chosen level of significance. Let us look at an example. We have measured the strikes of fault planes in an area of the sea floor imaged by a side-scan sonar device and found the following orientations

$\Theta$: 110°, 300°, 310°, 135°, 320°, 141°, 145°, 330°, 335°, 280°, 160°, 170°  \quad n=13$

Since these are oriented features, we double the angles first:

$\Theta_2$: 220°, 240°, 260°, 270°, 276°, 280°, 282°, 290°, 300°, 310°, 200°, 320°, 340°$

Summing up the sines and cosines, we find

$$X_r = 0.0998 \quad \text{and} \quad Y_r = -0.796$$

which gives us

$$\theta_2 = 277° \quad R_2 = 0.802$$

We convert back to original angles by dividing $\theta_2$ by 2:

$$\theta_2 = 138.5°$$

The null hypothesis is not affected by doubling the angles since we are checking if $\kappa = 0$. At the $\alpha = 0.05$ level of significance we find critical $R_{12,0.05} = 0.475$ which is greatly exceeded by the observed $R = 0.802$. We conclude that the fault strikes are not randomly oriented.

**Test for a specific trend**

There are instances when we would like to test whether an observed trend equals some specified trend. The determination of critical values for such a scenario is difficult and involves using complex charts. As an alternative, we instead find the confidence angle $\Delta \theta$ around the mean direction of the sample and see if this angle is large enough to contain the specified trend we want to compare the data to. The approximate standard error in $\theta$ is given (in radians) by

$$s_c \approx \sqrt{nR\kappa}$$

(6.8)

If we assume that the estimation errors are normally distributed, we may use the $z$ values for a given confidence level to construct the angular interval:

$$\theta \pm z_{\alpha} s_c$$
which will contain the true mean direction \( \mu \) \( \alpha \% \) of the time. In our strike data case, it is believed that the faults are tensional features due to tectonic forces operating in the N30°E direction. We would then theoretically expect the faults to trend at 90° to these tensile stresses or in the N120°E direction. From Appendix 2.12 we find \( \kappa = 2.897 \), which gives us

\[
se = \frac{1}{\sqrt{13 \cdot 0.802 \cdot 2.897}} = 10.4^\circ
\]

Now, this is the standard deviation about the doubled angle direction 277°. We divide by 2 to find the uncertainty associated with the original angles, \( s_e = 5.2^\circ \). The 95% confidence interval corresponds to \( z_{95\%} = 1.96 \) so the interval around \( \bar{\theta} \) becomes

\[
\theta = 138.5^\circ \pm 10.2^\circ
\]

Since the predicted direction of 120° is outside this interval we conclude that the observed mean direction deviates from that predicted from the orientation of current tectonic forces. The orientation of stresses may have changed since the formation of the faults.

### Testing if two mean directions are equal

Another common situation where we may want to apply a statistical test is to find out if two sample mean directions are equal at some prescribed level of confidence. For example, we may have more side-scan sonar data for an area farther away, and we would like to test whether the mean direction at the second location is the same as what we just found. We can carry out this test by comparing the vector resultant of the two groups to that produced by pooling. The idea is that if the mean directions are different, then the pooled resultant should be shorter than the sum of the resultants. This can be checked using an F-test. Unfortunately, the nature of the F-statistic changes somewhat with the dispersion parameter \( \kappa \). For large \( \kappa (>10) \), we compute

\[
F_{1, n-2} = \frac{(n-2)(R_1 + R_2 - R_p)}{(n - R_1 - R_2)}
\]

with \( n \) the number of total observations, \( R_1 \) and \( R_2 \) the resultants for each data set, and \( R_p \) the resultant from the combined data. \( \kappa \) is estimated from \( R_p \), the mean resultant. If \( 2 < \kappa < 10 \), a more accurate statistic is

\[
F_{1, n-2} = \left(1 + \frac{3}{8\kappa}\right) \frac{(n-2)(R_1 + R_2 - R_p)}{(n - R_1 - R_2)}
\]

while for smaller \( \kappa \) special tables must be used. We will use this test to see whether the strikes of the faults in our second area are the same as what we found in the first area (\( \Theta = 138.5^\circ \); double angle = 277°). The new data are:

\[
\theta = 91^\circ, 280^\circ, 111^\circ, 115^\circ, 118^\circ, 300^\circ, 122^\circ, 126^\circ, 130^\circ, 80^\circ, 320^\circ, 149^\circ \quad (n = 12)
\]

which we double to get

\[
\theta_2 = 182^\circ, 200^\circ, 222^\circ, 230^\circ, 236^\circ, 240^\circ, 240^\circ, 244^\circ, 252^\circ, 260^\circ, 160^\circ, 280^\circ, 298^\circ
\]

We find \( \theta_2 = 234.5^\circ \) and \( R_2 = 0.805 \). The test becomes \( H_0: \theta_1 = \theta_2 \) versus \( H_1: \theta_1 \neq \theta_2 \) at \( a = 0.05 \). For the combined data set, with \( n = 25 \), we get

\[
\theta_p = 256.7^\circ \quad R_p = 0.749
\]
We find $\kappa$ from the appendix to be $\kappa = 2.36$. Then, $F$ becomes

$$F_{1,23,0.05} = \left(1 + \frac{3}{8(2.36)}\right) \frac{23[(13)(0.802) + (12)(0.805) - (25)(0.749)]}{25 - (13)(0.802) - (12)(0.805)} = 7.42$$

The $F$ table gives critical $F$ value for 1 and 23 degrees of freedom as 4.28. Therefore, we must reject the null hypothesis and conclude that the fracture directions in the two areas appear to have differing orientations at the 95% confidence level.

**Robust Directions**

The robustness idea is also applicable to directional data since such data is not free of outliers either. Directional outliers cause less harm than the general outliers encountered earlier since the data are forced to be periodic. It is clear from Figure 6–2 that an outlier causes most damage when it is pointing 90° away from the trend of the bulk of the data.

![Fig. 6-2. The effect of outliers on mean direction.](image)

We may find the circular analog of the median by finding the $\theta$ that minimizes the sum

$$\text{minimize } \sum_{i=1}^{n} d\left(\theta_i, \tilde{\theta}\right)$$

where $d(\theta_i, \theta)$ is the arc distance between $\theta_i$ and $\theta$ measured along the perimeter of the unit circle. Since this distance is proportional to $|\theta_i - \theta|$ it is equivalent of minimizing

$$\text{minimize } \sum_{i=1}^{n} |\theta_i - \tilde{\theta}|$$

which we know gives the median of the $\theta_i$ set. Similarly, a LMS mode estimate can be found by determining the midpoint of the shortest arc containing $n/2 + 1$ points. Again, this would involve sorting the directions first. For example, if we had the directions

$$\theta_i: 75^\circ, 85^\circ, 90^\circ, 98^\circ, 170^\circ$$

The shortest arc over 5/2+1=3 points is between 85° and 98°, giving the mode estimate $\theta = 91.5^\circ$, which seems to indicate that 170° is an outlier with respect to the rest of the data.
Data with length and direction

In the analysis so far we have only considered the direction of a feature and not its length. However, in many cases, like fracture data, the features will have very different lengths. Analysis of such data will give both a 1 km long and a 100 km long fault the same weight. We can account for this bias by weighting the directions by the length of the feature. By keeping track of the length of faults (and not their numbers) per sector we obtain a rose diagram that reflects the proportions of the various fracture directions. The rationale here is that large and/or long faults may be more representative of the tectonic stresses than a few short fractures. The rose diagram may then be normalized by the total length of the fractures to give proportions in percent.

Staying with fault strikes, it is clear that in many regions the faults are not entirely straight lines but may actually curve along its strike. Such fractures must be approximated by shorter straight line segments. It then becomes obvious that we must weight the pieces by their length, otherwise the directional frequencies would depend on the number of pieces used. Fault traces must therefore be digitized prior to analysis on a computer. A fault trace may look like the one in Fig. 6-3.

![Fig. 6-3. A digitized fault trace.](image)

Depending on the angular width of the polar histogram, the digitized fault may end up in two bins corresponding to the angles $\alpha_1$ and $\alpha_2$. The digitizing process will introduce errors in the digitized points. To see how, consider the line segment in the figure (Fig. 6-4) below.

![Fig. 6-4. The uncertainty in the location of the digitized points introduces error in both the length and angle of a line segment.](image)
We may assume that the exact position of each point is in error, here shown by the $1\sigma$ estimate (circles) for each point. There are two things of interest here. First, the length of the segment, $d$, will have an uncertainty since it basically is a difference between two uncertain values. We found this to be

\[ d = \bar{d} \pm \sqrt{2}\sigma, \]

(6.12)

Furthermore, we see that the angle or direction $\alpha$ may be in error by $\pm\Delta\alpha$ given by

\[ \Delta\alpha = \tan^{-1}\left(\frac{2\sigma}{d}\right) \]

(6.13)

One should not digitize so frequently that $\Delta\alpha$ exceeds the histogram interval. If this interval is $10'$, you are best served by making the average digitizing interval $d > 10\sigma_r$. (Alternatively, one can filter the digitized track so that it is smooth before binning the segments.)

The length of the chains that have a direction within the width of a bin is simply computed by adding up all the individual lengths. Note that since internal nodes are shared by adjacent line segments the uncertainty of the length is independent of the number of line segments and only depends on (6.12). However, the sum of the lengths of the chains will have an uncertainty that is cumulative since the chains are not connected. This method will provide a frequency distribution with error bars in which directions with many small segments have higher uncertainty than directions with a few long faults.

**Spherical Data Distributions**

Much geological data involve not only directions in the plane but spatial directions as well. This introduces another degree of complexity, but because 3-D vector data are very common in the earth sciences we need to look at such data in more detail. We find that much of the measurements used in structural geology, such as strike and dip of fault planes, can be expressed as a normal vector to the plane. Other examples include vectoral measurements of the geomagnetic field, palaeomagnetic measurements, stress directions, and determinations of crystallographic axes for petrofabric studies.

It is common to require that these vectors have unit length so that their endpoints all lie on the surface of a sphere with unit radius - hence the name spherical distributions. Similar to the case in the plane, there will be data that only contain orientation (i.e., axes) rather than direction.

![Fig. 6–5. The relation between the Cartesian and spherical coordinate systems.](Image)
We need to use a 3-D Cartesian coordinate system to describe the unit vectors. Thus, any vector \( \mathbf{V} \) is uniquely determined by the triplet \((x,y,z)\). We could also use spherical angles \( \theta \) (colatitude) and \( \phi \) (longitude) to specify the vector direction. We can relate the Cartesian coordinates and the spherical angles as follows:

\[
\begin{align*}
X &= \sin(\theta) \cos(\phi) \\
Y &= \sin(\theta) \sin(\phi) \\
Z &= \cos(\theta)
\end{align*}
\]  

\( (6.14) \)

Fig. 6–6. Local, right-handed coordinate system shows the convention used in structural geology.

However, geological measurements like strike and dip are more common than Cartesian coordinates and spherical angles, and follow their own convention. We define a new local coordinate system in which \( X \) points toward north, \( Y \) points east, and \( Z \) points vertically down (in order to maintain a right-handed coordinate system). In such a system, fault plane dips are expressed as positive angles. For the fault plane in Fig. 6-6 we find that the angle \( A \) is the azimuth of the strike of the plane, and \( D \) is the dip, measured positive down. The slip-vector \( \mathbf{OP} \) is then given by its components

\[
\begin{align*}
X &= -\sin(A) \cos(D) \\
Y &= \cos(A) \cos(D) \\
Z &= \sin(D)
\end{align*}
\]  

\( (6.15) \)

Once we have converted our \((A, D)\) data to \((x,y,z)\) we can compute such quantities as mean direction and spherical variance, which are simple extensions of the 2-D or directional analogs. The length of the resultant vector is simply
\[ R = \sqrt{\left(\sum x_i\right)^2 + \left(\sum y_i\right)^2 + \left(\sum z_i\right)^2} \]  
\[ (6.16) \]

which is usually normalized as \( R = R/n \). The coordinates \( X, Y, \) and \( Z \) of the mean vector are then

\[
\tilde{X} = \frac{\sum x_i}{R}, \\
\tilde{Y} = \frac{\sum y_i}{R}, \\
\tilde{Z} = \frac{\sum z_i}{R}
\]
\[ (6.17) \]

so that the mean slip-vector is given by

\[
\tilde{D} = \sin^{-1} \tilde{Z}, \\
\tilde{A} = \tan^{-1}\left(\frac{-\tilde{X}}{\tilde{Y}}\right)
\]
\[ (6.18) \]

If all the vectors are close together, the resultant \( R \) will approach \( n \), and as the vectors are more randomly distributed \( R \) will approach 0. We can use \( R \) (or \( R/n \)) as the basis for the spherical variance \( s^2_s \):

\[
s^2_s = \frac{n}{n-R} = 1 - \frac{R}{n} \]
\[ (6.19) \]

We can perform simple tests in a similar manner to the ones we carried out for 2-D directional data. As was the case for 2-D, we now require a spherical probability distribution to which we may compare our statistics. In response to the need for 3-D statistical analysis, in particular for palaeomagnetic studies, the famous statistician R. Fisher developed a theoretical distribution for spherical data. His probability function has since been called the Fisher distribution and is

\[
P = \frac{\kappa}{4\pi \sinh(\kappa)} \exp\left(\kappa \cos\psi\right)
\]
\[ (6.20) \]

where \( \psi \) is the spherical angular distance between the mean direction \( \mu \) and a point’s direction and \( \kappa \) is the precision parameter, similar to what we found for the von Mises distribution, and \( \sinh \) is the hyperbolic sine (needed to normalize volume to unity). Fisher showed that one can estimate \( \kappa \) from the samples provided \( n > 7 \) and \( \kappa > 3 \). The estimate is then

\[
k = \kappa = \frac{n-2}{n-R}
\]
\[ (6.21) \]

Testing a spherical distribution for randomness follows the approach used for directional data: We must first evaluate \( R \), the mean resultant. Then, we state the null hypothesis

\[
H_0: \kappa = 0 \quad H_1: \kappa > 0
\]

Again, this test is converted to look for critical values of \( R \) given the prescribed level of confidence \( \alpha \). Tables for critical \( R \) for selected values of \( \alpha \) and \( n \) exist.
Often, we will be interested in testing whether the observed mean direction equals some prescribed direction, given the uncertainties due to random errors. As for circular data, such tests are best performed by constructing the $\alpha$ confidence region around the mean direction. This statistic is based on the Fisher distribution and gives the spherical radius of a cone of confidence around the mean direction. As usual, this radius, $\delta$, is a function of the confidence level $\alpha$ and $R$. We find

$$\delta_{1-\alpha} = \cos^{-1}\left\{1 - \frac{n-R}{R}\left[\left(\frac{1}{\alpha}\right)^{\frac{1}{n-1}} - 1\right]\right\}$$

(6.22)

This is quite a messy expression, but it simplifies considerably if we assume (actually know) that $\kappa > 7$ and use $\alpha = 0.05$. Then,

$$\delta_{95} \approx \frac{140}{\sqrt{\kappa n}}$$

(6.23)

We then have that there is a 95% probability that the true mean lies within the cone of confidence with angular radius $\delta_{95}$. Let us examine a set of palaeomagnetic measurements. We have been given 6 measurements of declination and inclination:

<table>
<thead>
<tr>
<th>Dec</th>
<th>115°</th>
<th>130°</th>
<th>111°</th>
<th>120°</th>
<th>118°</th>
<th>125°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inc</td>
<td>43°</td>
<td>49°</td>
<td>51°</td>
<td>39°</td>
<td>55°</td>
<td>45°</td>
</tr>
</tbody>
</table>

We convert these angles to $x, y, z$ using $x = \cos D \sin I$, $y = \sin D \cos I$, and $z = \sin I$ and find

<table>
<thead>
<tr>
<th>$x$</th>
<th>-0.31</th>
<th>-0.42</th>
<th>-0.23</th>
<th>-0.38</th>
<th>-0.27</th>
<th>-0.41</th>
<th>-0.337</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>0.66</td>
<td>0.5</td>
<td>0.59</td>
<td>0.66</td>
<td>0.51</td>
<td>0.58</td>
<td>0.583</td>
</tr>
<tr>
<td>$z$</td>
<td>0.68</td>
<td>0.75</td>
<td>0.75</td>
<td>0.64</td>
<td>0.82</td>
<td>0.71</td>
<td>0.725</td>
</tr>
</tbody>
</table>

Computing the mean direction and resultant gives

$$\tilde{I} = \sin^{-1}\tilde{Z} = 46.5^\circ$$

$$\tilde{D} = \tan^{-1}\tilde{Y}/\tilde{X} = 120^\circ$$

$$\tilde{R} = \frac{1}{6}5.87 = 0.978$$

$$\kappa = \frac{6-2}{6-5.87} = 30.77$$

It is clear that our distribution has a preferred direction since $\kappa$ is so large. To see what the radius of the 95% confidence cone is we use (6.23) and find

$$\delta_{95} \approx \frac{140}{\sqrt{30.77} \cdot 6} = 10.3^\circ$$

This means that the true population direction probably (@ 95%) lies within a spherical cone of radius 10° centered on the observed mean direction. We also obtained some measurements from a nearby site:

<table>
<thead>
<tr>
<th>Dec</th>
<th>63°</th>
<th>70°</th>
<th>55°</th>
<th>65°</th>
<th>50°</th>
<th>45°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inc</td>
<td>25°</td>
<td>23°</td>
<td>30°</td>
<td>23°</td>
<td>28°</td>
<td>30°</td>
</tr>
</tbody>
</table>
which converts to

\[
\begin{array}{cccccccc}
  x & 0.38 & 0.31 & 0.50 & 0.39 & 0.57 & 0.61 & 0.46 \\
  y & 0.81 & 0.86 & 0.71 & 0.83 & 0.68 & 0.61 & 0.75 \\
  z & 0.42 & 0.39 & 0.50 & 0.39 & 0.47 & 0.50 & 0.445 \\
\end{array}
\]

which gives the mean resultant as

\[
\overline{R}_2 = 0.986 \quad \hat{I} = 26.4^\circ \quad \hat{D} = 58.5^\circ
\]

We would like to compare these two mean directions to see if they are similar at the 95% confidence level, i.e.

\[
H_0: I_1 = I_2, D_1 = D_2
\]

The best fit for this equality is exactly the same as the one we used for 2-D directional data. We use an F-test to check if the resultant from the pooled data is significantly different from the linear sum of the individual resultants:

\[
F = \frac{(n_1 + n_2 - 2)(R_1 + R_2 - R_p)}{(n_1 + n_2 - R_1 - R_2)}
\]

We need to find the resultant for the combined data set of 12 points. It is given by

\[
R_p = \frac{6}{12} \left( \overline{R}_1 + \overline{R}_2 \right) = 0.89
\]

The F statistic becomes

\[
F = \frac{10(5.87 + 5.92 - 10.68)}{12 - 5.87 - 5.92} = 52.9
\]

The observed F value by far exceeds the critical \( F_{0.05, 1, 10} = 4.96 \) and we must reject the idea that the two directions are the same.