Goals for the day:

1. Words: Weibull model
2. R: dweibull, pweibull, qweibull, rweibull, curve
3. Big idea: the Weibull density provides a model for many populations

The Weibull density function is given by the function

\[
f(x) = \frac{\alpha}{\beta} \left( \frac{x}{\beta} \right)^{\alpha-1} e^{-\left( \frac{x}{\beta} \right)^\alpha} \quad 0 < x < \infty
\]

This density has two parameters: \( \alpha \) and \( \beta \) which are called the shape and scale parameters. Both \( \alpha \) and \( \beta \) are positive real numbers. (Alert: The textbook has a different form for the density. The difference is that the shape parameter of the book’s density is the reciprocal of the one that R and most other sources use.)

Shape

This density is often used to model a variety of phenomena since it can model distributions that are reasonably symmetric as well as those that are quite skew.

Sketch the Weibull density function for several values of \( \alpha \) and \( \beta \), Remember that \( \alpha \) is a “shape” parameter and \( \beta \) is a scale parameter. So you might fix \( \beta \) and vary \( \alpha \) to see what kinds of shapes are possible. For values of \( \alpha \), make sure that you use 1 and positive numbers that are less than 1 and greater than 1. See if you can make choices of \( \alpha \) and \( \beta \) that lead to skew left, symmetric, and skew right distributions.

You can use curve to sketch a graph of a function of \( x \). So here is a sketch of the Weibull density function with \( \alpha = 2 \) and \( \beta = 10 \). Notice that we don’t know what the graphics “window” should be so we use qweibull to determine where most of the area is.

Use R efficiently so that you don’t have to do a lot of typing. For example, use the source window.

```r
> alpha=2
> beta=10
> high=qweibull(.99,shape=alpha,scale=beta)
> curve(dweibull(x,alpha,beta),from=0,to=high,ylab='density')
```
Fitting a Weibull distribution to data

A certain type of glass is manufactured. A sample of 31 sheets of glass are tested for tensile strength and the load (in mPa) at which they failed was recorded. The data are in the dataframe `windowstrength`. (This is in the M241 package.) The variable is named `ksi` under the mistaken impression that those are the units! Draw a density histogram of the variable and describe the distribution below:

We are going to investigate using a Weibull model to describe the population from which this sample is taken. We will use the model with $\alpha = 4.6$ and $\beta = 33.7$.

According to the model

- What percentage of the population will have breaking strength less than 50 kPa?
- What percentage of the population will have breaking strength less than 15 kPa?
- What percentage of the population will have breaking strength greater than 30 kPa?
- What is the first quartile of the breaking strengths of the population?
- What is the third quartile of the breaking strengths of the population?

Compare the last two answers above with the values for this particular sample.

It would be nice to superimpose the density function that we are using to model the population on the histogram of the sample values. We use the R function `lines` to do this. The function lines takes two arguments: a vector of $x$ values and a corresponding vector of $y$ values and it plots the points $(x,y)$ and draws a curve connecting them.

```r
> hist(windowstrength$ksi, freq=F, main='')
> x=seq(15,50,.01)
> y=dweibull(x, 4.6, 33.7)
> lines(x,y)
```
Fitting models

How do we know what are good values to choose for $\alpha$ and $\beta$ when modeling a population? For example, why choose 4.6 and 33.7 in the previous example? No choice of $\alpha$ and $\beta$ will fit the data perfectly and just from the picture one could make reasonable arguments for other values.

What would be a good way to use the histogram and the superimposed curve to measure the goodness of the fit of the model to the data?

From your understanding of the $\alpha$ and $\beta$ parameters, how might you modify one or the other to fit these data better? Try a different choice and superimpose it on a density histogram. Record your $\alpha$ and $\beta$ here:

$\alpha$:  
$\beta$: 

The parameters for the `windowstrength` model were found using an extremely powerful function `fitdistr` which can be found in the `MASS` package. Load that package. We fit both a normal and weibull to the data. (The parenthetical values below the estimates are estimates of the standard deviation of the parameter.)

```r
> fitdistr(windowstrength$ksi, 'weibull')
shape   scale   
4.63539 33.67424  
( 0.62922) ( 1.38288)   
> fitdistr(windowstrength$ksi, 'normal')
mean     sd      
30.8114  7.1354  
( 1.2816) ( 0.9062)   
```

Draw a histogram of the data and superimpose both the weibull and normal models that are fitted above. Which model fits the data better? What is your evidence? Besides the graphical evidence, look at predictions that each model makes about the data.