Empirical Distributions

```
clear

corg = load('organicmatter_one.txt');

plot(corg,zeros(1,length(corg)),'o')

nbin = round(sqrt(length(corg)))

vmin = min(corg) + 0.5*range(corg)/nbin;
vmax = max(corg) - 0.5*range(corg)/nbin;
vwth = range(corg)/nbin;
v = vmin : vwth : vmax;
```

```
v =
  Columns 1 through 4
    9.7383   10.3814   11.0245   11.6676
  Columns 5 through 8
   12.3107   12.9537   13.5968   14.2399
```

```
emin = min(corg);
emax = max(corg);
ewth = range(corg)/nbin;
e = emin : ewth : emax;
```

```
e =
  Columns 1 through 4
    9.4168   10.0598   10.7029   11.3460
  Columns 5 through 8
   11.9891   12.6322   13.2753   13.9184
  Column 9
   14.5615
```

```
for i = 1 : nbin
   if i < nbin
   corgb = corg(emin+(i-1)*ewth<=corg & emin+i*ewth>corg);
   n(i) = length(corgb);
   else
   corgb = corg(emin+(i-1)*ewth<=corg & emin+i*ewth>=corg);
   n(i) = length(corgb);
   end
end
```

```
n =
     2     2    10     7    14     9    12     4
```

```
h = histogram(corg);
```

```
h =
  Histogram with properties:
            Data: [60x1 double]
          Values: [2 6 13 20 15 4]
         NumBins: 6
        BinEdges: [9 10 11 12 13 14 15]
        BinWidth: 1
       BinLimits: [9 15]
    Normalization: 'count'

      FaceColor: 'auto'
      EdgeColor: [0 0 0]
```

```
cdfplot(corg)
```

```
mean(corg)
```

```
median(corg)
```

```
quantile(corg,[.25 .50 .75])
```

```
boxplot(corg)
```

```
range(corg)

var(corg)

std(corg)

kurtosis(corg)

skewness(corg)
```

Try to play the following two commands

randtool

disttool

The t-Test

$$\hat{t} = \frac{\left|\overline{a} - \overline{b}\right|}{\sqrt{\frac{n_a + n_b}{n_a n_b} \cdot \frac{(n_a - 1) \cdot {s_a}^2 + (n_b - 1) \cdot {s_b}^2}{n_a + n_b - 2}}}$$

The one-sample t-test is used to test the hypothesis that the mean of a Gaussian-distributed population has a value specified in the null hypothesis. The two-sample t-test is employed to test the hypothesis that the means of two Gaussian distributions are identical.

Two-sample test

```
clear

load('organicmatter_two.mat');

histogram(corg1,'FaceColor','b'), hold on
histogram(corg2,'FaceColor','r'), hold off

na = length(corg1); nb = length(corg2);
ma = mean(corg1); mb = mean(corg2);
sa = std(corg1); sb = std(corg2);

tcalc = abs((ma-mb))/sqrt(((na+nb)/(na*nb)) * ...
   (((na-1)*sa^2+(nb-1)*sb^2)/(na+nb-2)))

tcalc =
    0.7279

tcrit = tinv(1-0.05/2,na+nb-2)

tcrit =
    1.9803
```

```
[h,p,ci,stats] = ttest2(corg1,corg2,0.05)
```

```
h =
     0

p =
    0.4681

ci =
   -0.3028
    0.6547

stats =
    tstat: 0.7279
       df: 118
       sd: 1.3241
```

Test1:
Load ‘organicmatter_three.mat’ to do two-sample T-test

```
load('organicmatter_three.mat');
```

The F-Test

The two-sample *F*-test by Snedecor and Cochran (1989) compares the variances s_a^2 and s_b^2 of two distributions, where $s_a^2 > s_b^2$. An example is the comparison of the natural heterogeneity of two samples based on replicated measurements. The sample sizes n_a and n_b should be above 30. Both the sample and population distributions must be Gaussian. The appropriate test statistic with which to compare the variances is then

$$\hat{F} = \frac{s_a^2}{s_b^2}$$

```
clear

load('organicmatter_four.mat');
s1 = std(corg1)
s2 = std(corg2)

df1 = length(corg1) - 1
df2 = length(corg2) - 1

if s1 > s2
  slarger  = s1
  ssmaller = s2
else
  slarger  = s2
  ssmaller = s1
end
```

```
slarger =
    1.2550

ssmaller =
    1.2097
```

```
Fcalc = slarger^2 / ssmaller^2
Fcrit = finv(1-0.05/2,df1,df2)
```

```
Fcalc =
    1.0762

Fcrit =
    1.6741
```

```
[h,p,ci,stats] = vartest2(corg1,corg2,0.05)
```

```
h =
     0

p =
    0.7787

ci =
    0.6429
    1.8018

stats =
    fstat: 1.0762
      df1: 59
      df2: 59
```

Test2: Load ‘organicmatter_five.mat’ to do the two-sample F-test

```
load('organicmatter_five.mat');
```