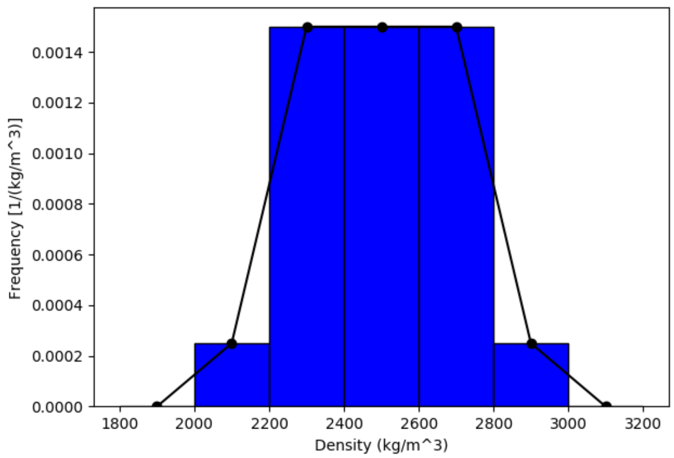
The following table lists repeated measurements of the density of glass particles.

a. Plot a frequency distribution of the density values in a plot with the x-axis ranging from [1800, 3200] kg/m^3 with seven total bins (each bin size is 200 kg/m^3).
b. Determine the sample mean of the distribution.
c. Determine the true mean of the particle density within a confidence interval of 95%.
d. What fraction of the density measurements lie within the range [2200, 2800] kg/m^3?

Measurement #	Density [kg/m^3]
1	2694
2	2516
3	2628
4	2831
5	2342
6	2505
7	2612
8	2531
9	2452
10	2380
11	2657
12	2335
13	2668
14	2516
15	2701
16	2222
17	2003
18	2565
19	2222
20	2316

SOLUTION:

Following is a frequency distribution plot of the data. Refer to the python code at the end of this document for how it was generated.



Note that the area under the frequency distribution curve is equal to one.

The sample mean of the measurements, *m*, is,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{i=N} x_i, \tag{1}$$

where $N = 20$ and x_i is measurement number *i*. Using the given data,

$$\boxed{\bar{x} = 2484.8\ \mathrm{kg/m^3}}.$$

The sample standard deviation of the measurements, *s*, is,

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{i=N}(x_i - \bar{x})^2}. \tag{2}$$

Using the given data,

$$\boxed{s = 201.9\ \mathrm{kg/m^3}}.$$

The standard error of the sample means, *SEM*, is,

$$SEM = \frac{s}{\sqrt{N}}. \tag{3}$$

Using the given data,

$$\boxed{SEM = 45.2\ \mathrm{kg/m^3}}.$$

The true mean, μ, will lie within the range,

$$\mu = \bar{x} \pm t_{95\%} SEM, \tag{3}$$

where the value for $t_{95\%}$ is found from a Student's *t* distribution at a 95% confidence interval to be 2.093 for $N = 20$ ($N - 1 = 19$ degrees of freedom). Thus,

$$\boxed{\mu = 2484.8 \pm 94.5\ \mathrm{kg/m^3}}. \tag{4}$$

The fraction of density measurements in the range [2200, 2800] is,

$$\mathrm{fraction}(x_i, x_f) = \sum_{x_i}^{x_f - \Delta x_{f-1}} f(x_i, x_i + \Delta x_i)\Delta x_i \tag{5}$$

$$\mathrm{fraction}(2200, 2800) = f(x_{2200}, x_{2400})(200\ \mathrm{kg/m^3}) + f(x_{2400}, x_{2600})(200\ \mathrm{kg/m^3}) + f(x_{2600}, x_{2800})(200\ \mathrm{kg/m^3}) \tag{6}$$

$$\mathrm{fraction}(2200, 2800) = [f(x_{2200}, x_{2400}) + f(x_{2400}, x_{2600}) + f(x_{2600}, x_{2800})](200\ \mathrm{kg/m^3}) \tag{7}$$

$$\mathrm{fraction}(2200, 2800) = \left[0.001500\frac{\mathrm{m^3}}{\mathrm{kg}} + 0.001500\frac{\mathrm{m^3}}{\mathrm{kg}} + 0.001500\frac{\mathrm{m^3}}{\mathrm{kg}}\right](200\ \mathrm{kg/m^3}) \tag{8}$$

$$\mathrm{fraction}(2200, 2800) = 0.9. \tag{9}$$

Thus, 90% of the measurements lie in the range [2200, 2800] $\mathrm{kg/m^3}$.

```
# uncertainty_10.py

import scipy.stats as stats
import numpy as np
import pylab as plt

# Put the data into an array.  Normally we would read this data from
# an input file.
my_data = np.array([2694, 2516, 2628, 2831, 2342, 2505, 2612, 2531, 2452, 2380, 2657, 2335, 2668, 2516, 2701,
2222, 2003, 2565, 2222, 2316])

# Report some statistics about the data.
N = len(my_data)  # number of samples
sample_mean = np.mean(my_data)  # sample mean
sample_stdev = np.std(my_data, ddof=1)      # sample standard deviation;
                                            # divisor is N-1 since we
                                            # don't know the entire
                                            # population
sem = stats.sem(my_data) # std error of the sample mean
CI = 0.05 # 95% confidence interval (alpha = 0.05)
t = stats.t.ppf(1-CI/2, N-1)  # compute t-factor for the specified confidence interval

# Print the data
print("# of data entries =", N)
print("sample mean (kg/m^3) = %.1f" % sample_mean)
print("sample standard deviation (kg/m^3) = %.1f" % sample_stdev)
print("standard error of the sample means (kg/m^3) = %.1f" % sem)
print("t_95 for %d" % N, "samples = %.3f" % t)
print("true mean (kg/m^3) = %.1f" % sample_mean, " +/- %.1f" % (t*sem))

# Generate the frequency distribution data.  Set the bin edges.
bin_list = np.linspace(1800, 3200, num=8)
#Nbins = 6  # number of bins to use in the frequency plot
counts, bin_edges = np.histogram(my_data, bins=bin_list, density=True)

# Determine the bin centers.
bin_centers = np.empty([len(bin_edges)-1])
for i in range(len(bin_edges)-1):
    bin_centers[i] = (bin_edges[i]+bin_edges[i+1])/2

# Print the bin edges, the bin centers, and the counts.
print("[lower_bin_value, upper_bin_value)\tbin_center\tfrequency [1/(kg/m^3)]")
for i in range(len(bin_edges)-1):
    print("[%.1f," % bin_edges[i], "%.1f)" % bin_edges[i+1], "\t%.1f" % bin_centers[i], "\t%3f" % counts[i])

# Plot the frequency distribution.  Plotting it two ways: once showing
# the bin sizes with a bar chart and once showing the center of the
# bins with a scatter plot.
plt.figure(1)
plt.hist(my_data, bins=bin_list, density=True, color="blue", edgecolor="black")
plt.plot(bin_centers, counts, color='black', marker='o', linestyle='solid')
plt.ylabel('Frequency [1/(kg/m^3)]')
plt.xlabel('Density (kg/m^3)')
plt.show()
```

Running the program gives the following output:

# of data entries = 20
sample mean (kg/m^3) = 2484.8
sample standard deviation (kg/m^3) = 201.9
standard error of the sample means (kg/m^3) = 45.2
t_95 for 20 samples = 2.093
true mean (kg/m^3) = 2484.8 +/- 94.5

[lower_bin_value, upper_bin_value)	bin_center	frequency [1/(kg/m^3)]
[1800.0, 2000.0)	1900.0	0.000000
[2000.0, 2200.0)	2100.0	0.000250
[2200.0, 2400.0)	2300.0	0.001500
[2400.0, 2600.0)	2500.0	0.001500
[2600.0, 2800.0)	2700.0	0.001500
[2800.0, 3000.0)	2900.0	0.000250
[3000.0, 3200.0)	3100.0	0.000000