Chem 454 Statistics Problem Set – mod. Feb 4, 2019

Note for 2/4/19, problems 2 and 4 have been modified. They are in red for emphasis.

1] What is the 95% confidence interval for 6 measurements whose average is 4.22 and with a standard deviation of 0.07? ¹

2] An analysis of chromium in a steel sample was repeated 5 times. The results are listed below. Using Grub’s test which of the five results may be removed in the determination of the mean? ²

<table>
<thead>
<tr>
<th>Run #</th>
<th>Cr mass %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.22</td>
</tr>
<tr>
<td>2</td>
<td>8.55</td>
</tr>
<tr>
<td>3</td>
<td>6.02</td>
</tr>
<tr>
<td>4</td>
<td>5.98</td>
</tr>
<tr>
<td>5</td>
<td>5.43</td>
</tr>
</tbody>
</table>

3] What is the relative population in pph (%) that lies below the value of 88.7 for a Gaussian distribution whose mean is 64.4 and with a standard deviation of 17.3? ³

4] Sketch and label a diagram that illustrates the concepts of
   a) linear range
   b) sensitivity
   c) background
   d) dynamic range

part B

Problem 5-A p 112 Harris 8th Ed.

Absorbance measurements were made on a blank and a low concentration sample. The data are summarized below.
A) What is the absorbance detection limit of method?

B) A calibration curve gives a slope of 2.24e4 1/M. What is the LOD?

C) What is the LOQ?

5] What is the 95% confidence limit for the following set of data?  

4.88, 5.13, 4.92, 5.22, 5.07, 4.72

6] What is the 90% confidence interval for the following measurements?  

10.22 11.03 10.71 10.47 10.83

7] A blood sample was sent to two different labs for cholesterol analysis. The results are:

Lab 1 $\bar{x} = 221$ mg/dL $s = 11$ $n = 10$

Lab 2 $\bar{x} = 233$ mg/dL $s = 14$ $n = 10$

<table>
<thead>
<tr>
<th>blank</th>
<th>low conc</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0006</td>
<td>0.0047</td>
</tr>
<tr>
<td>0.0012</td>
<td>0.0054</td>
</tr>
<tr>
<td>0.0022</td>
<td>0.0062</td>
</tr>
<tr>
<td>0.0005</td>
<td>0.0060</td>
</tr>
<tr>
<td>0.0016</td>
<td>0.0046</td>
</tr>
<tr>
<td>0.0008</td>
<td>0.0056</td>
</tr>
<tr>
<td>0.0017</td>
<td>0.0052</td>
</tr>
<tr>
<td>0.001</td>
<td>0.0044</td>
</tr>
<tr>
<td>0.0011</td>
<td>0.0058</td>
</tr>
</tbody>
</table>

std dev 0.000556 0.000644

avg 0.001189 0.005322
Are the two standard deviations different significantly different at the 95% confidence limit? 7

8] You carefully followed an analytical procedure with n = 6 and found a mean of 6.37 mM with s = 0.37. Meanwhile, Joe Cutcorners used a modified procedure with n = 4, \( \bar{x} = 6.87 \) mM with s = 0.22. Assuming that the standard deviations are not statistically different from each other, does Joe’s method have a systematic error, i.e. statistically different at the 95% confidence limit? 8

9] The analysis of Mn (m/m) was conducted on a Martian rock sample. The following values were obtained:

\[
4.77\% \quad 4.82\% \quad 5.22\% \quad 4.92\% \quad 5.82\% \quad 4.99\%
\]

The average is 5.09 and s = 0.357, Using valid statistics, which if any of the values can be rejected? 9

10] The Rope-A-Dope fishing line company guarantees that their “Jaws-Max” nylon line will haul in at least an 80 lbs gilled monster. Their chief statistician, Myron Knumbers has 200 samples of the Jaws-Max line tested and finds that the mean weight for line breakage is 120 lbs with a standard deviation of 60 lbs. What are the chances that the hooked 80 pounder will get away if you were using Jaws-Max and end up being another fishing story? 10

11] Linear Range is an expression of 11

\[
\begin{array}{llllll}
\text{a) signal to} & \text{b) analyte} & \text{c) analyte} & \text{d) accuracy and} & \text{e) precision of} \\
\text{noise ratio} & \text{concentration} & \text{detection limits} & \text{precision of} & \text{repeated} \\
& \text{range over} & \text{of method} & \text{method or} & \text{experiment} \\
& \text{which the c \( \propto \)} & & \text{technique} & \text{results} \\
& \text{signal} & & & \\
\end{array}
\]

12] What is the 95% confidence interval for 5 measurements whose average is 3.44 and with a standard deviation of 0.04? 12

\[
\begin{array}{llllll}
\text{a) } \pm 0.04 & \text{b) } \pm 0.4 & \text{c) } \pm 0.05 & \text{d) } \pm 0.1 & \text{e) } \pm 0.06 \\
\end{array}
\]
13] What is the relative population that lies above the value of 55.1 for a Gaussian distribution whose mean is 33.8 and with a standard deviation of 11.8?  

a) 0.50%  
b) 3.6%  
c) 1.8%  
d) 46%  
e) 0.18%

14] Which of the following values may be discarded with 90% confidence?  

9.11  
8.89  
9.01  
9.77  
9.05

15] Two methods of analyses were compared. Method A had a mean of 23.2 with a standard deviation of 4.4. Method B had a mean of 24.1 with a standard deviation of 4.8. Both sets of measurements were done 6 times. What is the F ratio and are the standard deviations significantly different from each other at the 95% confidence level?  

a) 1.19, no  
b) 1.19, yes  
c) 0.840, no  
d) 0.840, yes  
e) 1.09, no

16] Standard deviation can be best described as a measure of  

a) detection  
b) accuracy  
c) sensitivity  
d) linearity  
e) precision

17] The method of least squares fits a line (L) to a set of x,y data by  

a) maximizing  
b) minimizing  
c) minimizing  
d) maximizing  
e) minimizing

\[ \Sigma (x_i - x_L) \]  
\[ \Sigma (x_i - x_L)^2 \]  
\[ \Sigma (y_i - y_L) \]  
\[ \Sigma (y_i - y_L)^2 \]  
\[ \Sigma (y_i - y_L) \]

18] Standard Deviation is a measure of
a) accuracy  
b) how close the mean is to the true result  
c) the mean relative to the true result  
d) precision  
e) precision and accuracy

19] Two sets of measurements were made by different technicians. The first has a mean of 55.6 ppm with a standard deviation of 7.3 ppm over 7 measurements. The second had $\overline{x} = 62.1$ ppm with $s = 8.5$ ppm over 6 measurements. Are the two standard deviations significantly different from each other?  

20] A final quantity, D is calculated by the ratio of $D = H/G$. If H was measured 6 times with a mean of 987.2 grams and a standard deviation of 11.9 grams and G had $\overline{x} = 554.2$ liters with a standard deviation of 32.7 liters over 10 measurements. What is the absolute uncertainty of D?  

21] Calculate the limit of detection of Method A given the calibration curve below. Also note that the curve has 9 data points each replicated 5 times. The data point at the lowest concentration has a standard deviation of 0.12 signal units. 

![Calibration Curve](image.png)

22] Replicate runs of an analysis gave 5 values of 1.77, 1.45, 1.91, 1.85 and 1.82. Can any of these values be discarded with 95% statistical confidence?  

23] Replicate runs of an analysis gave 5 values of 9.88, 8.92, 9.62, 9.33 and 9.27. What is the 95% confidence interval of this set of data?  

24] When is it appropriate to calculate $s_{pooled}$ for two sets of data?  
   a) When the 2 standard deviations are statistically the same  
   b) When $t$-calculated $> t$-test  
   c) When the 2 standard deviations are statistically different  
   d) When $F$-calculated $= F$-table
25] Two different methods of Fe analysis were compared to an NIST standard containing 6.50% Fe by mass. The results follow:

<table>
<thead>
<tr>
<th>Method</th>
<th>%Fe</th>
<th>±</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>6.33%</td>
<td>± 0.23%</td>
</tr>
<tr>
<td>Method 2</td>
<td>6.55%</td>
<td>± 0.45%</td>
</tr>
</tbody>
</table>

Which of the following statements is true?

a) Method 1 is less precise and less accurate
b) Method 1 is more precise and less accurate
c) Method 2 is less precise and less accurate
d) Method 2 is more precise and less accurate
e) Method 2 is more precise and more accurate

26] Trace analysis were conducted on a sample 25 times. The average concentration was found to be 10.0 ppb with a standard deviation of 5.0 ppb. What is the chance that a single analysis will yield a result that is twice this average?

27] An analysis for lead in groundwater was conducted. What is the correct terms for the lead and the water?

a) Lead is the sample and the groundwater is the analyte
b) Both the lead groundwater are the analytes
c) Both the lead groundwater are the samples
d) Lead is the matrix and the groundwater is the analyte
e) Lead is the analyte and the groundwater is the matrix

Answers

1 $\mu = \bar{x} \pm t \frac{s}{\sqrt{n}} = 4.22 \pm 2.776 \left(\frac{0.07}{6}\right)^{1/2} = 0.08 \; \mu = 4.22 \pm 0.08$

2 $G = \frac{|value - \bar{x}|}{s}$

$n = 5$, avg = 6.24, $s = 1.34$  most likely value to consider is 8.55

$G = \frac{(8.55 - 6.24)}{1.34} = 1.72$
Use table

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>99.9</th>
<th>99.5</th>
<th>99</th>
<th>97.5</th>
<th>95</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.155</td>
<td>1.155</td>
<td>1.155</td>
<td>1.155</td>
<td>1.153</td>
<td>1.148</td>
</tr>
<tr>
<td>4</td>
<td>1.496</td>
<td>1.496</td>
<td>1.482</td>
<td>1.481</td>
<td>1.463</td>
<td>1.425</td>
</tr>
<tr>
<td>5</td>
<td>1.780</td>
<td>1.764</td>
<td>1.749</td>
<td>1.715</td>
<td>1.672</td>
<td>1.602</td>
</tr>
<tr>
<td>6</td>
<td>2.011</td>
<td>1.973</td>
<td>1.944</td>
<td>1.887</td>
<td>1.822</td>
<td>1.729</td>
</tr>
<tr>
<td>7</td>
<td>2.201</td>
<td>2.139</td>
<td>2.067</td>
<td>2.020</td>
<td>1.938</td>
<td>1.828</td>
</tr>
<tr>
<td>8</td>
<td>2.358</td>
<td>2.274</td>
<td>2.221</td>
<td>2.126</td>
<td>2.032</td>
<td>1.909</td>
</tr>
<tr>
<td>9</td>
<td>2.492</td>
<td>2.387</td>
<td>2.323</td>
<td>2.215</td>
<td>2.110</td>
<td>1.977</td>
</tr>
<tr>
<td>10</td>
<td>2.606</td>
<td>2.482</td>
<td>2.410</td>
<td>2.290</td>
<td>2.176</td>
<td>2.036</td>
</tr>
<tr>
<td>11</td>
<td>2.705</td>
<td>2.564</td>
<td>2.465</td>
<td>2.355</td>
<td>2.234</td>
<td>2.088</td>
</tr>
<tr>
<td>12</td>
<td>2.791</td>
<td>2.636</td>
<td>2.550</td>
<td>2.412</td>
<td>2.285</td>
<td>2.134</td>
</tr>
<tr>
<td>13</td>
<td>2.867</td>
<td>2.699</td>
<td>2.607</td>
<td>2.462</td>
<td>2.331</td>
<td>2.175</td>
</tr>
<tr>
<td>14</td>
<td>2.935</td>
<td>2.755</td>
<td>2.659</td>
<td>2.507</td>
<td>2.371</td>
<td>2.213</td>
</tr>
<tr>
<td>15</td>
<td>2.997</td>
<td>2.806</td>
<td>2.705</td>
<td>2.549</td>
<td>2.409</td>
<td>2.247</td>
</tr>
</tbody>
</table>

At 95% CL table is 1.672

G-calc > G-table we can discard 8.55

\[^3\] Z = (64.4 – 88.7)/17.3 = 1.40 use table 4-1 area = 0.4192
Above = 0.5000 – 0.4192 = 0.0808 or 8.08%
Therefore the population below is 91.92%

\[^4\] see also notes and your book
Part b)

A) What is the absorbance detection limit of method? Want less than 1% chance for false positive or negative for LOD.

Signal detection limit would be:
\[ y_{DL} = y_{blank} + 3s \]
\[ y_{DL} = 0.001189 + 3 \times 0.000644 = 0.0031 \text{ absorbance units} \]

**B) A calibration curve gives a slope of 2.24e4 1/M. What is the LOD?**

For a straight line:

\[ y_{DL} = m \cdot c + y_{\text{blank}} \]

And we have

\[ y_{DL} = y_{\text{blank}} + 3s \]

solving for \( c \) gives the LOD

\[ c = \frac{3s}{m} \]

\[ c_{\text{LOD}} = \frac{3 \times (0.000644)}{2.24 \times 10^4 (1/M)} \]

\[ c_{\text{LOD}} = 8.6 \times 10^{-8} M \]

**C) What is the LOQ?**

\[ c_{\text{LOQ}} = 10s/m = \frac{10 \times (0.000644)}{2.24 \times 10^4 (1/M)} = 2.9 \times 10^{-7} M \]

\(^5\text{Avg} = 4.99 \quad \text{s.d.} = 0.184\)

\[ \mu = x\text{-bar} \pm ts/n^{1/2} = 4.99 \pm 2.571(0.184)/(6)^{1/2} \]

\[ \mu = 4.99 \pm 0.19 \]
x-bar = 10.65

\[ s = \sqrt{\frac{(10.22-10.65)^2 + (11.03-10.65)^2 + (10.71-10.65)^2 + (10.47-10.65)^2 + (10.83-10.65)^2}{5-1}} = 0.32 \]

\[ \mu = x-bar \pm \frac{ts}{\sqrt{n}} = 10.65 \pm 2.132(0.32)5^{1/2} = 10.65 \pm 0.31 \]

\[ F = \frac{14^2}{11^2} = 1.62 \quad F-\text{Table} = 3.18 \text{ so they are not different from each other} \]

\[ S_{\text{pooled}} = \sqrt{0.37^2 \times 5 + 0.22^2 \times 3/6+4} = 0.33 \]

\[ t = \frac{6.87-6.37}{0.332} \sqrt{\frac{6*4}{6+4}} = 2.41 \quad t_{\text{Table}@95\%} = 2.306 \text{ so they are different from each other.} \]

\[ G-\text{calc} = \frac{5.82 - 5.09}{0.357} = 2.045 \quad \text{for n = 6 @ 95\% CL} \quad G-\text{Table} = 1.822 \]

\[ G-\text{calc} > G-\text{Table} \quad 5.82 \text{ may be discarded} \]

\[ z = \frac{80 - 120}{60} = 0.667 \quad z \approx 0.7 \quad \text{area} = 0.258 \]

\[ \text{Area from 0 to 80 is} \quad 0.500 - 0.258 = 0.242 \quad \approx 24\% \]

\[ b \]

\[ c \pm 2.776 \frac{(0.04)}{(5)^{1/2}} \]

\[ z = \frac{33.8-55.1}{11.8} \approx 1.80 \quad \text{area} = 0.4641 \]

\[ \text{above} = 0.5000 - 0.04641 \approx 3.6\% \]

\[ 9.77-9.11/9.77-8.89 = 0.75 \quad Q_{\text{table}} = 0.64 \text{ for n = 5 so 9.77 can be discarded} \]

\[ a \]

\[ e \]

\[ c \]

\[ d) \text{ precision} \]

\[ \text{Use F-test} \quad s_1 = 8.5 \text{ and } s_2 = 7.3 \quad F = \frac{8.5^2/7.3^2}{1.35579} \quad df_1 = 6-1 = 5 \quad df_2 = 7-1 = 6 \]

\[ F-\text{table} 4.39 \quad F < F-\text{table} \text{ so std. dev. Are not statistically different.} \]
\[ \sigma(\%) = \sqrt{\sigma(\%)_1^2 + \sigma(\%)_2^2 + \sigma(\%)_3^2 + \ldots} \]

\[ \sigma_1(\%) = \frac{11.9}{987.2} \times 100 = 1.21\% \quad \sigma_2(\%) = \frac{32.7}{554.2} \times 100 = 5.90\% \]
\[ \sigma_i(\%) = (1.21^2 + 5.90^2)^{1/2} = 6.02\% \quad D = \frac{987.2}{554.2} \times \frac{L}{g} = 1.781 \text{ g/L} \]

6.02 % of 1.781 g/L = 0.107 g/L

21 LOD = 3s/m = 3(0.12 signal units) / 2 signal/conc = 0.18 conc. units

22 1.77, 1.45, 1.91, 1.85 and 1.82 mean = 1.76 \quad s = 0.181 

Use Grubbs Test \[ G = 1.76 - 1.45 / 0.181 = 1.713 \quad G\text{-Table for n = 5 is 1.672} \]

G > G\text{-table}, 1.45 can be discarded.

23 9.88, 8.92, 9.62, 9.33 and 9.27 mean = 9.404 \quad s = 0.3643 \quad d.f. = 4 \quad t = 2.776

\[ \mu = \bar{x} \pm \frac{t \sigma}{\sqrt{n}} = \bar{x} \pm 2.776(0.36)/\sqrt{5} = 0.4469 = 0.45 \]

24 a) When the 2 standard deviations are statistically the same

25 B) Method 1 is more precise and less accurate

26 \[ Z = \frac{X - \mu}{S} = \frac{20 - 10}{5} = 2 \] look up 2 on z-table.

Area = 0.4773 \quad Area above 2 = 0.5000 - 0.4773 = 0.0227 or 2.27%

27 d) Lead is the analyte and the groundwater is the matrix