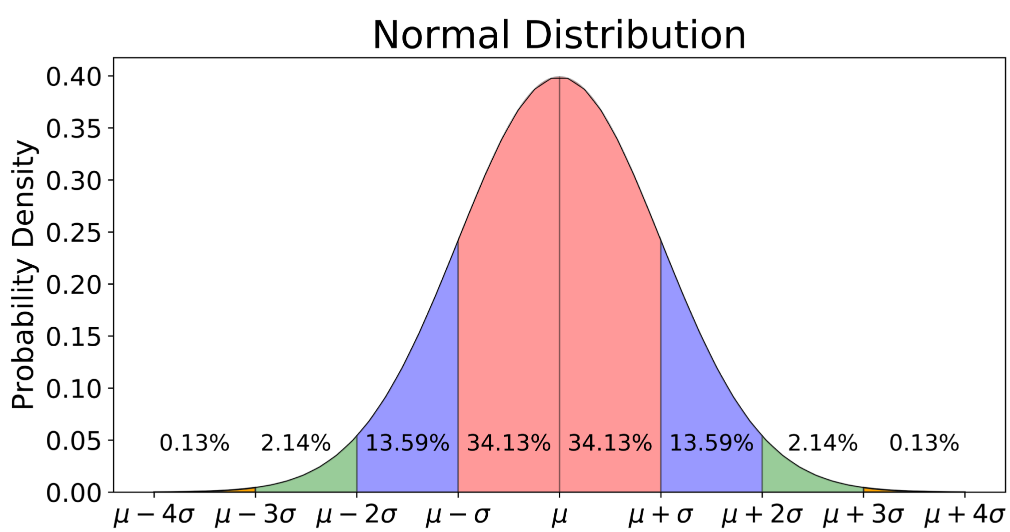
Worksheet 15: Final Exam Review III

1. What is the difference between a statistic and a parameter? Name examples of the symbols for each.

Solution: A statistic describes a sample, while a parameter describes a population. $\overline{x}$ is the sample mean, while μ is the population mean. Similarly, s is the sample standard deviation, while σ is the population standard deviation.

2. In the space below, sketch a Gaussian distribution and identify roughly the mean and standard deviation of the distribution.

Solution:



3. What do you get when you square the standard deviation?

Solution: The variance

4. How much of the area under a Gaussian curve is found within 1 standard deviation of the mean? How about within 2 standard devations? 3?

Solution: 68, 95, 99.7%

5. If you want to decrease the standard deviation of the mean by a factor of 5, by what factor should you increase the number samples?

Solution: 25

$$\sigma_n = \frac{\sigma}{\sqrt{n}}$$

6. What statistical test can you use to determine if there is a statistical difference between two standard deviations? How does it work?

Solution: The F-test:

$$F_{\text{calculated}} = \frac{s_1^2}{s_2^2}$$

where the larger standard deviation is always on top.

If the calculated value of F is greater than the tabulated one, the difference is significant. The null hypothesis in this case is that the two standard deviations are drawn from populations with the same standard deviation. In other words, if the calculated value is greater than the tabulated one, we reject the null.

7. What does a confidence interval represent? How can we construct one?

Solution: A confidence interval means that if we repeat n measurements many times to compute the mean and standard deviation, the 95% confidence interval would capture the true population mean in 95% of the sets of measurements. We can construct a confidence interval using the equation below.

$$\text{Confidence interval} = \overline{x} \pm \frac{ts}{\sqrt{n}}$$

8. What can a confidence interval tell us about our data?

Solution: If a confidence interval generated from our data captures an accepted value for something, it tells us that our data is acceptable to that confidence level.

9. What is the Student's t test used for? What are the three different cases?

Solution: This test can be used to compare the means from two sets of measurements for the same quantity. The null hypothesis states that the two means come from populations with the same population mean, and we reject the null if the calculated value of t is greater than the tabulated one. The different cases are listed below.

- Case 1: Comparing an average value and standard deviation generated by measuring a quantity several times to an accepted answer. In this case we just use a confidence interval generated as shown above.
- Case 2: We measure a quantity multiple times by two different methods, each with their own averages and standard deviations. This case splits depending on whether or not the standard deviations are statistically the same. If they are the same we use the pooled standard deviation, otherwise use them separately.
- Case 3: Sample A is measured once by method 1 and once by method 2; the two measurements do not give the same result. This is repeated for n samples. Do the two methods agree with each other statistically? This is the individual differences scenario and has its own long equation.

10. What is the difference between a one- and two-tailed t test?

Solution: A two-tailed test will reject the null if the t value is far enough above or below the mean, while a one-tailed test only looks in one direction.

11. Under what conditions is it acceptable to reject a potential outlier? What test have we learned in this class to test this?

Solution: If you have a clear explanation, like a blunder, for a potential outlier it is acceptable to reject it. Otherwise, you must use the Grubbs test to determine whether or not it can be considered an outlier.

$$G_{\text{calculated}} = \frac{|\text{questionable value} - \overline{x}|}{s}$$

If the calculated value is greater than the tabulated value, the value can be discarded.

12. How should you calculate the concentration of different components of a mixture when the individual spectra are well resolved? What about when they overlap substantially?

Solution: Simultaneous equations can be used when they are well resolved, but least-squares regression is needed when they overlap.