November 20, 2014

# BioStatistics Problem Set 2

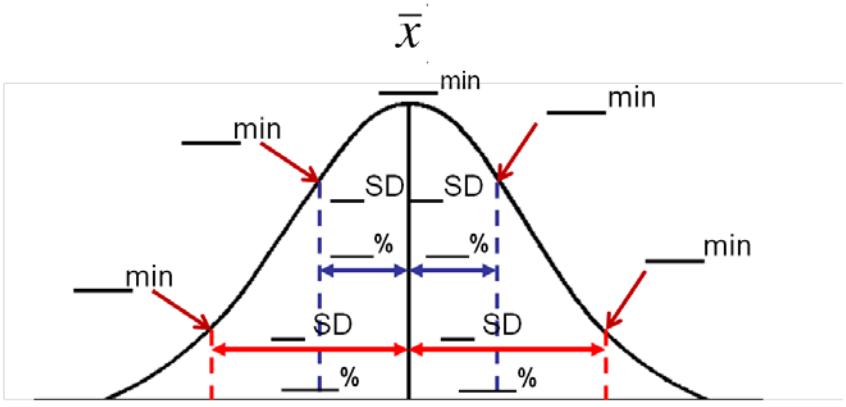
1. Describe what a "normal distribution" of data means. What does the graph of data that is normally distributed typically look like?

2. You are interested in measuring the amount of viral production from the new 2014 influenza strain. DNA constructs from this influenza strain are introduced into an appropriate host cell line and after three days, you measure the amount of viral proteins released in the cell culture media. To get as accurate a measurement as possible for virus production, you repeat the experiment 5 times, recording viral protein levels to generate the data table below. Calculate the sample mean ($\bar{x}$) and standard deviation ($s$) for the data you collected:

| Experiment # | Viral protein production (µg protein/mL media) |
|---|---|
| 1 | 12.4 |
| 2 | 18.0 |
| 3 | 11.3 |
| 4 | 15.2 |
| 5 | 23.5 |

   Create a bar graph plotting ($\bar{x}$) and use the standard deviation for + and - error bars. Be sure to label all axis' and units correctly.

3. Fill the values in the distribution curve below using your analysis of the above data.




4. In the data from problem 2, what is the sample variance ($s^2$)? What are the degrees of freedom (df)?
5. Explain the difference between the standard deviation ($s$) and the standard error of the mean (SEM). What happens to each value as your sample size ($n$) increases?

6. A local farm would like to compare the growth rate of tomatoes grown outside in natural sunlight versus 24-hour artificial light indoors. They are hoping that by increasing the rate of growth, they'll be able to fulfill an unusually high demand for tomatoes. They plant seedlings under both conditions and measure the height of each plant after 8 weeks:

| | **Sunlight** | **Artificial light** |
|---|---|---|
| **Height (cm)** | 67 | 91 |
| | 55 | 87 |
| | 62 | 101 |
| | 71 | 98 |
| | 58 | 106 |
| | 69 | 88 |

   a. Calculate $\bar{x}$, SEM and 95% CI for each of the two groups (round to nearest whole number).
   b. Construct two bar graphs comparing the means of each group. In one graph plot error bars using the SEM. In the other, plot error bars using the 95% CI.
   c. Looking at the error bars, would you conclude there's a statistically significant difference in the growth rate of tomatoes grown in sunlight versus artificial light? WHY?
   d. When comparing the two groups, what would the null hypothesis ($H_0$) be?
   e. Perform a Student's *t*-test : What is $t_{obs}$ and what is $t_{crit}$ ($\alpha$=0.05)?
   f. Based on the Student's *t*-test, would you accept or reject the null hypothesis? Is the difference between the two groups statistically significant?

7. Molecular biologists are attempting to clone a DNA polymerase, R1, from a highly replicative strain of bacteria. They are hoping that by isolating this strain, scientists will be able to synthesize DNA constructs in half the time it currently takes using the standard *Taq* polymerase. In separate tubes, they incubate equal amounts of Taq polymerase and R1 polymerase with a DNA template and radioactive nucleotides under conditions optimal for DNA polymerization using both enzymes. After allowing each reaction to incubate for 30 minutes, the free radioactive nucleotides are washed away and the radioactive signal from DNA polymers are measured (in units of radioactive counts per minute – or CPM). The experiments were repeated 6 times, however, in two instances the purity of R1 was questionable and therefore only 4 values were collected for R1:

| | **Taq** | **R1** |
|---|---|---|
| **CPM (counts per minute)** | 398 | 500 |
| | 421 | 467 |
| | 490 | 482 |
| | 438 | 431 |
| | 390 | N/A |
| | 560 | N/A |

   a. Calculate $\bar{x}$, SEM and 95% CI for each of the two groups (round to nearest whole number).
   b. Construct two bar graphs comparing the means of each group. In one graph plot error bars using the SEM. In the other, plot error bars using the 95% CI.
   c. Looking at the error bars, would you conclude there's a statistically significant difference in the polymerization rates between the Taq and R1 groups? WHY?
   d. When comparing the two groups, what would the null hypothesis ($H_0$) be?
   e. Perform a Student's *t*-test : What is $t_{obs}$ and what is $t_{crit}$ ($\alpha$=0.05)?
   f. Based on the Student's *t*-test, would you accept or reject the null hypothesis? Is there a statistically significant increase in the rate of DNA polymerization with the R1 polymerase compared to the regular Taq polymerase?