Statistics
A2RCC U11D8 Apps

Name Key

1 Normal Distribution and Z-scores

1) **In gym class students have to run a mile. For a sixth grade class the average was 512 seconds with a standard deviation of 68. Assuming the time to run a mile was normally distributed, answer the following questions, rounding to the *nearest thousandth*.**

   a. **What is the probability of a student running in less than 400 seconds?**

   normalcdf(-9999, 400, 512, 68) = .050

   b. **What is the probability of a student running in more than 610 seconds?**

   normalcdf(610, 9999, 512, 68) = .075

   c. **What is the probability of a student running between 475 and 525 seconds?**

   normalcdf(475, 525, 512, 68) = .283

   d. **What percentile is a student if he ran it in 380 seconds?**

   normalcdf(-9999, 380, 512, 68) = .026

   top 3%

2) **On a math test which had a mean of 83 and a standard deviation of 6, what is the 90th percentile score, to the nearest whole number?**

   invnorm(.90, 83, 6) = 91

3) **If Yesenya scored 78 on her AP Euro test which had a mean of 70 and a standard deviation of 3, and she scored an 84 on her Algebra 2 test which had a mean of 80 and a standard deviation of 2, on which test did she score better?**

   AP Euro: $\frac{78-70}{3} = 2.7$ A2: $\frac{84-80}{2} = 2$

   better in AP Euro since the z-score is higher. (Percentile is also higher)

Regressions

Amount of Water Vapor That Will Saturate 1 Cubic Meter of Air at Different Temperatures	
Air Temp (x) °C	Water Vapor (y) (g)
−20	1
−10	2
0	5
10	9
20	17
30	29
40	50

The accompanying table shows the amount of water vapor, y, that will saturate 1 cubic meter of air at different temperatures, x.

A) Write an exponential regression equation for this set of data, rounding all values to the *nearest thousandth.*

$y = 4.194(1.068)^x$

B) Using this equation, predict the amount of water vapor that will saturate 1 cubic meter of air at a temperature of 50°C, and round your answer to the *nearest tenth of a gram*.

$4.194(1.068)^{50} = 112.5$

C) Algebraically determine the air temperature needed, to the *nearest degree*, to obtain 80 grams of water vapor.

$$\frac{80}{4.194} = \frac{4.194(1.068)^x}{4.194}$$

$$1.068^x = 19.07486886$$

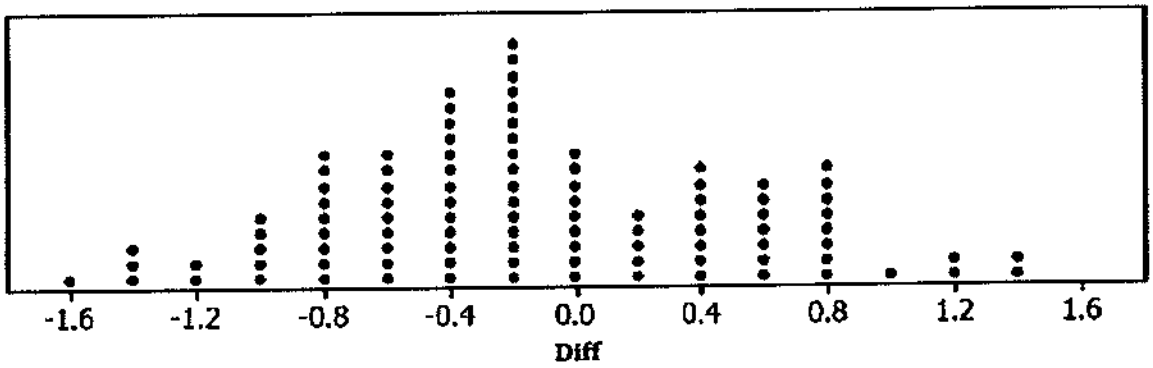
$$x = \log_{1.068}(19.07486886)$$

$$x = 44.816\ldots$$

$45°$

3 Difference of the Means

Medical patients who are in physical pain are often asked to communicate their level of pain on a scale of 0 to 10 where 0 means no pain and 10 means worst pain. (Note: Sometimes a visual device with pain faces is used to accommodate the reporting of the pain score.) Due to the structure of the scale, a patient would desire a lower value on this scale after treatment for pain.



Imagine that 20 subjects participate in a clinical experiment and that a variable of "ChangeinScore" is recorded for each subject as the subject's pain score after treatment minus the subject's pain score before treatment. Since the expectation is that the treatment would lower a patient's pain score, you would desire a *negative* value for "ChangeinScore." For example, a "ChangeinScore" value of -2 would mean that the patient was in less pain (for example, now at a 6, formerly at an 8).

Recall that Diff $= \bar{x}_A - \bar{x}_B$. Although the 20 "ChangeinScore" values for the 20 patients are not shown here, below is a randomization distribution of the value Diff based on 100 random assignments of these 20 observations into two groups of 10 (Group A and Group B).

1. From the distribution above, what is the probability of obtaining a Diff value of -1 or less?

$\frac{11}{100} = .11$

2. With regard to this distribution, would you consider a Diff value of -0.4 to be statistically significant? Explain.

No, Since it is close to the center of the data. $-.4$ is not an unusual or extreme value

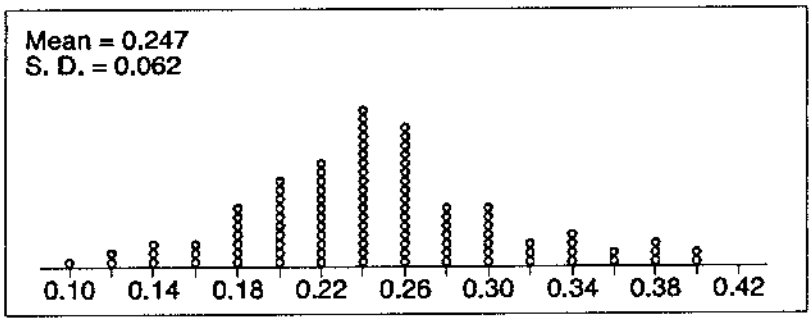
3. a. With regard to how Diff is calculated, if Group A represented a group of patients in your experiment who received a new pain relief treatment and Group B received a pill with no medicine (called a *placebo*), how would you interpret a Diff value of -1.4 pain scale units in context?

The group that received the new pain med had an avg pain reduction of 1.4 pts more than the placebo group.

b. Given the distribution above, if you obtained a Diff value such as -1.4 from your experiment, would you consider that to be significant evidence of the new treatment being effective on average in relieving pain? Explain. Yes, -1.4 is only $\frac{4}{100} = .04 = 4\%$ of the random samples, so it is significant

4 Confidence Intervals

1) Stephen's Beverage Company is considering whether to produce a new brand of cola. The company will launch the product if at least 25% of cola drinkers will buy the product. Fifty cola drinkers are randomly selected to take a blind taste-test of products *A*, *B*, and the new product. Nine out of fifty participants preferred Stephen's new cola to products *A* and *B*. The company then devised a simulation based on the requirement that 25% of cola drinkers will buy the product. Each dot in the graph shown below represents the proportion of people who preferred Stephen's new product, each of sample size 50, simulated 100 times.



Proportion Preferring Stephen's Product

Assume the set of data is approximately normal and the company wants to be 95% confident of its results. Does the sample proportion obtained from the blind taste-test, nine out of fifty, fall within the margin of error developed from the simulation? Justify your answer.

MOE = 2 (.062) = .124

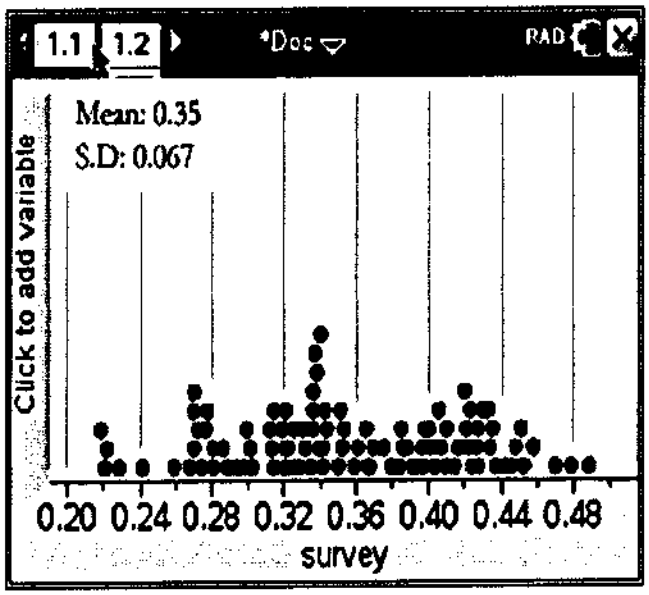
$\frac{9}{50} = .18$ $.247 \pm .124$ $[.123, .371]$

.18 falls in the 95% confidence interval

The company decides to continue developing the product even though only nine out of fifty participants preferred its brand of cola in the taste-test. Describe how the simulation data could be used to support this decision.

The simulation shows only $\frac{9}{100}$ trials were below the 18% likeability, so only 9% of the time the simulation showed less than 18%

2) Fitness Blender, an online exercise channel, is considering whether to create a new 8 week workout challenge. The owners will launch a new program if at least 35% of its users will purchase the program. Fifty users are randomly selected to take a survey. 13 out of fifty participants said they would purchase the new program. The owners of Fitness Blender then devised a simulation based on the requirement that 35% of their users will purchase the program. Each dot in the graph shown below represents the proportion of users who would purchase Fitness Blender's new program, each of sample size 50, simulated 100 times.



Assume the set of data is approximately normal and that the owners of Fitness Blender want to be 95% confident of its results. Does the sample proportion obtained from the survey, 13 out of fifty, fall within the margin of error developed from the simulation? Justify your answer.

MOE = 2(.067) = .134

CI = .35 ± .134 [.216, .484]

$\frac{13}{50}$ = .26 .26 falls within the 95% confidence interval

Fitness Blender decides to move forward with the new program for its users even though only 13 out of 50 participants said they would purchase it. Describe how the simulation data could be used to support this decision.

The owners have evidence that the proportion of users could be .26 or greater. It is unlikely that the proportion will be less than .26.

Extra Practice:

1) A random sample of 30 households was selected as part of a study on electricity usuage, and the number of kilowatt hours (kWh) was recorded for each household in the sample. The average usuage was found to be 375 kWh with a standard deviation if 81 kWh. Assuming the usage is normally distributed, calculate the 95% confidence interval for the mean usuage.

2) An industrial designer wants to determine the average amount of time it takes an adult to assemble an "easy to assemble" toy. A sample of 16 times yielded an average time of 19.92 minutes, with a sample standard deviation of 5.73 minutes. Assuming normally distributed assembly times, state a 95% confidence interval for the mean assembly time.

3) An article in The Artist Magazine, stated that 38% of high school students take advanced art classes. As part of a project for their statistics class, a group of New Paltz High School students decided to verify that claim. They conducted 20 surveys each containing 30 randomly chosen high school students and calculated the proportion, $\hat{p}$, for each sample pertaining to taking advanced art classes.

   1. What is the expected mean, to the nearest hundredth, of the sampling distribution of the sample proportions?

   2. Using the result from part 1, find the standard error, or the standard deviation of the sampling distribution of the sample proportions.

   3. Describe the graph of the sampling distribution of the sample proportions.

   4. Find the 95% confidence interval.