Algebra 2/Pre-Calculus

Name__________________

Statistics Review, Part 2

In this problem set, we will explore ways that we can use the concept of standard deviations to answer questions about a variety of situations.

1. Yang and Alec both run track. Yang's event is the 600 meter. His average (mean) time is 1:34 (or 94 seconds). Alec is a sprinter. His mean time for the 100 meter dash is 12.8 seconds.

For their upcoming mean, Yang's goal is to run the 600 in 1:30 (or less) and Alec's goal is to run the 100 meter in 12.5 seconds (or less).

a. Whose goal is more ambitious? Do we have enough information to answer this question? Explain.

b. Suppose we are told the following information: Yang's times for the 600 have a standard deviation of 3 seconds and Alec's time for the 100 have a standard deviation of 0.14 seconds.

Now can we determine whose goal is more ambitious?

c. To answer the last question, we can determine how far each of their goals are from the mean. ***Note:*** This is called a "z-score."

$$\text{Yang's z-score: } \frac{90-94}{3} = -1.3333 \qquad \text{Alec's z-score: } \frac{12.5-12.8}{0.14} = -2.1429$$

Yang's goal is 1.3333 standard deviations below his mean, whereas Alec's goal is 2.1429 standard deviations below his mean. So Alec's goal is a lot more ambitious.

Suppose that the race times for Yang and Alec are normally distributed. What is the probability that Yang reaches his goal? What is the probability that Alec reaches his goal? Do you think that it is likely that either of them will reach their goals?

Answer c. Yang: $\text{normalcdf}(0,90,94,3) = 0.0912 = 9.12\%$

Alec: $\text{normalcdf}(0,12.5,12.8,0.14) = 0.0161 = 1.61\%$

Interpretation: Both goals are quite ambitious, especially Alec's.

2. Kaitlin, Heather, Azalea, and Lizzie were swimming the 200 meter medley relay. The following table provides information about their times.

Swimmer	**Stroke**	**Average time (mean)**	**Standard deviation**
Kaitlin	Backstroke	84.1 sec	2.8 sec
Heather	Breaststroke	77.2 sec	2.3 sec
Azalea	Butterfly	75.4 sec	2.4 sec
Lizzie	Freestyle	66.9 sec	1.9 sec

a. What is their mean time for the entire relay? Note: This question is asking for the average time it takes the team to swim the entire race, not the average time for each leg of the race.

b. What is the standard deviation time for the entire relay? ***Note:*** The answer is **not** $2.8+2.3+2.4+1.9=9.4$. (You can look ahead to part **c** if you need a hint.)

c. Find the variance for each leg of the race. (What are the units for the variance?)

d. Find the variance for the entire relay.

e. Find the standard deviation time for the entire relay (if you didn't already find it in part **b**.)

f. The team's goal for the next race is to achieve a time of 5 minutes or better. How many standard deviations better than the mean it this time? (In other words, what is the z-score?)

g. Suppose that we assume that the times are normally distributed and that the swimmers are not influenced by each other's performance. What is the probability that the team reaches their goal?

h. To answer the last question, we made the assumption that the times for each swimmer were independent. This means, we were assuming that the swimmers were not influenced by each other's performance.

Is this a realistic assumption? Explain.

Answers a. 303.6 sec (not 75.9 sec) b. 4.7434 sec c. Backstroke: 7.84 sec^2, Breaststroke: 5.29 sec^2, Butterfly: 5.76 sec^2, Free: 3.61 sec^2. (Note: Since the variance is the square of the standard deviation, the units for the variance are seconds squared.)

d. $7.84+5.29+5.76+3.61=22.5\ \text{sec}^2$ e. 4.7434 seconds f. $\frac{300-303.6}{4.7434}=-0.7589$ (Their goal is to achieve a time that is 0.7589 standard deviations below the mean.)

g. $\text{normalcdf}(0,300,303.6,4.7434)=0.2239$, so they have a 22.39% chance of achieving their goal. h. This assumption is not realistic. For example, if the earlier swimmers are swimming well, the later swimmers may be more highly motivated.

3. The following calculation may be helpful in answering some of the questions for this problem.

$(x + x^2 + x^3 + x^4 + x^5 + x^6)^3 =$

$x^{18} + 3x^{17} + 6x^{16} + 10x^{15} + 15x^{14} + 21x^{13} + 25x^{12} + 27x^{11} + 27x^{10} + 25x^9 + 21x^8 + 15x^7 + 10x^6 + 6x^5 + 3x^4 + x^3$

a. Suppose we roll a standard six-sided die three times. What is the probability that the sum will be a 14?

b. What is the probability that the sum (of the three rolls) will be less than 6?

c. Now suppose instead that the die is rolled 100 times. Find the mean, variance, and standard deviation for the sum.

d. Suppose the die is rolled 100 times. Use the normal curve to estimate the probability that the sum will be at least 375.

e. Suppose the die is rolled 100 times. Use the normal curve to estimate the probability that the sum exactly 345.

Answers a. $\frac{15}{216}$ b. $\frac{10}{216}$ c. mean = 350, variance = 291.6667,

standard deviation = 17.0783 (Hint: Start by finding the mean and variance for 1 roll. Then multiply by 100.) d. $\text{normalcdf}(375, 600, 350, 17.0783) = 7.16\%$

e. $\text{normalpdf}(345, 350, 17.0783) = 2.24\%$

4. Four percent of all males are colorblind. Suppose 25 males are selected at random.

a. Find the probability that exactly two of the males are colorblind.

b. Find the probability that none of the males are colorblind.

c. Find the probability that more than two are colorblind.

Answers a. $({}_{25}C_2)(.04)^2(.96)^{23} = 18.77\%$ b. $({}_{25}C_0)(.04)^0(.96)^{25} = 36.04\%$

c. $1-\left(({}_{25}C_0)(.04)^0(.96)^{25}+({}_{25}C_1)(.04)^1(.96)^{24}+({}_{25}C_2)(.04)^2(.96)^{23}\right)=7.65\%$

5. A manufacturer produces a large number of toasters. From past experience, the manufacturer knows that approximately 2% are defective. In a quality control procedure, we randomly select 20 toasters for testing.

a. Determine the probability that exactly one of the toasters is defective.

b. Find the probability that at most two of the toasters are defective. Include enough details so that it can be understood how you arrived at your answer.

6. The values on a spinner are 3, 4, 6, 8, and 9, but each outcome is not equally likely, as shown on the table below.

a. Find the mean, variance, and standard deviation for a single spin.

Number	***Probability***
3	20%
4	35%
6	5%
8	15%
9	25%

b. Suppose the spinner is spun 200 times. Find the mean, variance, and standard deviation for the sum.

c. Suppose the spinner is spun 200 times. Use the normal curve to estimate the probability that the sum is exactly 1155.

d. Suppose the spinner is spun 200 times. Use the normal curve to estimate the probability that the sum is between 1160 and 1180.

Answers a. mean = $.2(3)+.35(4)+.05(6)+.15(8)+.25(9)=5.75$,variance= $.2(3-5.75)^2+.35(4-5.75)^2+.05(6-5.75)^2+.15(8-5.75)^2+.25(9-5.75)^2=5.9875$, standard deviation = 2.4469 b. mean = $5.75(200)=1150$, variance = $5.9875(200)=1197.5$, standard deviation = 34.6049 c. $\text{normalpdf}(1155,1150,34.6049)=1.14\%$

d. $\text{normalcdf}(1160,1180,1150,34.6049)=19.34\%$

7. A college wishes to find out how successful its graduates are 20 years after their graduation. They wish to determine whether graduates are content with their jobs and how much money they earn.

Identify any bias in the following possible methods for carrying out a survey.

a. A questionnaire is sent to all graduates and the responses tabulated.

b. A questionnaire is sent to all those who have donated money to the university.

c. A questionnaire is sent to all graduates stating that those who respond will be entered in a lottery for 20 years of free football tickets to all home games.

d. A small set of graduates is chosen by drawing their names randomly from the school's database and they are personally called to take part in the survey over the phone.

8. Which if any of the above is likely to produce the most accurate representation of the actual income and level of contentment of the graduates?

10. A teacher is trying to determine if peer tutoring helps students to do better on tests. The teacher finds out which students in his class are getting peer tutoring and compares their test average to the test average for the class as a whole. He finds that the students getting peer tutoring have a lower test average than the class as a whole and concludes that peer tutoring is not helpful.

a. Explain the flaw in this argument.

b. Consider the following experiment. "At the beginning of the school year, each student in the class is randomly assigned to two groups, tutoring and no tutoring. Students in the tutoring group will go to one session of peer tutoring each week, whereas students in the no tutoring group do not have this option. At the end of the year, the teacher compares the scores for the two groups."

Explain why the conclusions drawn from this experiment would be more statistically valid than the conclusions drawn in part **a**.

c. What are the practical and ethical reasons why the experiment described in part **b** would problematic in real life?