Applying the normal distribution

Applying the Normal Distribution

There are various components which come into play in decision making. Though not the only aspects, there is the research, the hypothesis, and hypothesis testing. All statistical results come are an integral part of the interpretation process in order to the decision making. When one is testing the hypothesis, a cornerstone to decision making, the analysis of the data is the probability. It is the author's opinion, that the best way to describe probability is with the questing beginning with " how likely." Just as Fox, Levin, and Ford (2013) state it, we as a human species uses probability in everyday conversation. Though we may answer the question with vague responses such as " probably," " most likely," or " highly unlikely," without any statistical reasoning to our interpretation, statistical probability offers more precise responses (Fox, Levin, & Forde, 2013). There various way to describe the statistics, including a normal curve, bell curve, central limit theorem, and z-scores which will be discussed below.

Normal Curve

College students, whether they are undertaking the undergraduate studies or that of graduate (masters or doctoral), one thing is certain, there is familiarity with the concept of a curve. At one point in a student's educational career, there has been a component within grading with regards to a curve. The same is true when speaking of statistics. According to Fox, Levin, and Forde (2013), the example of a grade curve is known as the normal distribution or normal curve. There are specific characteristics of a normal curve such as " a smooth, symmetrical distribution that is bell-shapes

and unimodal" (Fox, Levin, & Forde, 2013, p. 88). A normal curve has the mean, median, and mode in the same position, in the middle or center of the curve which is symmetric as " each side is a mirror image of the other" (Weiers, 2011, p. 208). According to Fox, Levin, and Forde (2013), a normal curve is unimodal as it only has one peak or point of maximum likelihood in the middle of the curve. From the maximum point of likelihood or central peak, the curve then begins to fall at both tails " extending indefinitely in either direction and getting closer and closer to the baseline without actually touching it" (Fox, Levin, & Fox, 2013, p. 89). There is the asymptotic tail which signifies the less than or almost zero of the normal distribution and " approaches the x-axis but never reaches it" (Weiers, 2011, p. 208). On the other side is the asymptotic tail indicating a more than average result as it " approaches the x-axis but never reaches it" (Weiers, 2011, p. 208). Thus a normal distribution will result in a normal curve in a bell-shaped curve.

Bell Curve

Why is the bell curve used to represent the normal distribution and not a different shape? The answer according to Fendler and Muzaffar (2008) is simple. It is because of the belief that aspects of the world distribute themselves according to the shape of a bell curve. According to Fendler and Muzaffar (2008), the bell curve is also known mathematically as the Gaussian curve, is indicative of the belief that " most phenomena occur around a middle point, while few occur at either the high or low extreme ends" (p. 63). The bell-shaped curve also comes from the understanding from three key reasons; (1) normal phenomena are approximately normally distributed; (2) can be used in approximating other distributions, including

binomial; and (3) means and proportions tend to be normally distributed (Weiers, 2011). The bell-shaped curve is an indication that there are few instances which occur at the high end of the scale, a few at the low end, and the remaining occurring in the middle. According to Weiers (2011), the shape of a normal curve will depend on its mean and standard deviation, though maintain a variation of a bell-shaped curve. There is also the " Empirical Rule" of a normal distribution is relating to the areas beneath or under the normal curve; (1) 68. 3% of the total area is in the interval of the mean (μ) – standard deviation (σ) to μ + σ; (2) 95. 4% of the total area is in the interval of the μ – 2σ to μ + 2σ; and (3) 99. 7% of the total area is in the interval of μ – 3σ to μ + 3σ (Weiers, 2011).

Importance of the Central Limit Theorem

The central limit theorem is an important part in statistics as it essentially says that if the sample size is large enough, the sampling distribution of the mean for a variable, both independent and random, will be normal. According to Peng (2019), the central limit theorem (CLT), along with the law of large numbers (LLN), are known as two fundamental results in probability theory and statistical analysis. In many instances researchers, or in the author's case, in the homeland security community, we are dealing more commonly with big data. According to Allende-Alonso et al. (2019), the idea of collecting more data is better as " it is supported by the idea that the statistical power is improved by increasing the sample size" (p. 112). Essentially, so long as the sample size is large enough, a normal distribution will occur. Even if it is unknown the shape of the distribution where the data comes from, it is the implication with the central limit theorem that the

sampling distribution is normal. The central limit theorem gives the researcher the confidence that statistical tools will function meaningfully in sample research. According to Kwak and Kim (2017), the central limit theorem is vital to modern statistics. Without the central limit theorem, parametric tests based on sample data of a population " with fixed parameters determining its probability distribution would not exist" (Kwak & Kim, 2019, p. 144). The central limit theorem can assist researchers when analyzing data.

Central Limit Theorem and Sampling Distribution

According to Weiers (2011), with regards to sampling distribution, it is the probability distribution of the sample means. The sampling distribution of the mean is defined as " the probability distribution of the sample means for all possible samples of that particular size" (Weiers, 2011, p. 246). The central limit theorem is the sampling distribution of the sampling means approaches a normal distribution as the sample size gets larger. The sampling distribution is not indicative of the shape of the data distribution. According to Allende-Alonso et al., (2019), there is an experimenter which considers that in experimental research " n> 30 is large enough, saying that sample sizes larger than 30 ensure the researcher that the central limit theorem holds" (p. 113). The central limit theorem and sampling distribution thus go hand and hand.

Published Results

An example is given in which the author recently took an exam for certification in the homeland security field. The certifying agency has

published the results of the exam and 75% of the test takers scored below the average. In a normal distribution, half of the scores would fall above the mean and the other half below. The question raised is how can what the certifying agency published be true? The answer is simple, the distribution is skewed, rather than symmetrical. The results of the exam would be positively skewed or skewed to the right as the curve would have a much longer tail on the right than the left (Fox, Levin, Forde, 2013). According to Fox, Levin, and Forde (2013), a positively skewed distribution is when " more responded receive low rather than high scores, resulting in a longer tail on the right than on the left" (p. 44). Due to the results of the exam indicate the grades were low at 75% below the average, rather than the other 25% who scored average or above average.

Z-Scores

The z-score is the key to standardizing the normal curve and expressing values in terms of their number of standard deviations from the mean (Weiers, 2011). The z-score may also be referenced by the standard score. The z-score, according to Fox, Levin, and Forde (2013), indicate the " direction and degree that any given raw score deviates from the mean of a distribution on a scale of sigma (standard deviation) units" (p. 95). In order to find the percentage of the total areas under the normal curve, the standard deviation distance from the raw score. Obtaining the z-score is done so finding the deviation (X – μ) giving the distance from the raw score from the mean, and then dividing this raw-score deviation by the standard deviation (Fox, Levin, Forde, 2013). The formula used to determine the z-score involves the following; (1) z represents the distance from the mean,

measured in standard deviation units; (2) x represents the value original normal distribution across the x-axis; (3) μ represents the mean of the distribution; and (4) σ represents the standard deviation of the distribution (Weiers, 2011). Researchers may use the z-score to determine probabilities to distinguish between favorable probabilities and potential failures. The z-score can be used to compare the raw scores while taking into account both the mean and standard deviation. The z-scores can also be used in conjunction with the normal curve as it determines the probability of " obtaining any raw score in a distribution" (Fox, Levin, Forde, 2013, p. 97).

An example of the z-score in use is to measure corporate financial distress. According to Agarwal and Taffler (2007), the z-score technique is used " as a proxy for bankruptcy risk in exploring such areas as merger and divestment activity, asset pricing and market efficiency, capital structure determination, the pricing of credit risk, distressed securities, and bond rating and portfolios" (p. 288). With regards to a business being financially health if the z-score is above the calculated score the firm is classified as " financially healthy, if below the cut-off, it is viewed as a potential failure" (Agarwal & Taffler, 2007, p. 290). From business to a more homeland security aspect, Fox, Levin, and Forde (2013) give an excellent example an emergency management agency may look into for preparation of a natural disaster, response times for emergency calls. With the assumption of a normal mean of 5. 6 minutes in response time (the time between dispatch and arrival at the scene) and a standard deviation of 1. 8 minutes. Any raw score can be translated, Fox, Levin, and Forde (2013) use 8. 0 minutes, however, the author will be using 6. 0 minutes for the example. With the formula provided

6. 0 – 5. 6 / 1. 8 equals 0. 22 standard deviation units above the mean response time of 5. 6 minutes. On the other side of the normal distribution would be a response time below the mean. For example, should there be an emergency response time of 4. 5 minutes, within a normal distribution of 5. 6 minutes and a standard deviation of 1. 8 minutes; the equation would read 4. 5 – 5. 6 / 1. 8 = -0. 61 standard deviation units below the mean response time.

Probability

The concept of probability may affect research that is chosen to be undertaken. Why? Simple, because the research will depend on what the dissertation topic is consisted of. The dissertation while built upon research, that research may play a crucial role in decision making down the road. Probability is much more complex than simply saying something is more likely or less likely to occur. In the world of the homeland security community, it is not so easy to say a terrorist act is more likely to occur in one area versus another without the data to back the conclusion. All aspect of statistical probability are an integral part in much more than the dissertation at hand. Understanding probability will aid in the analyzation and interpretation of the data. When undertaking research for a dissertation, the researcher wants to clearly understand what the question is. How and why the researcher plans to answer that question will determine what type of data to gather, and how it will be gathered. The data will then need to be analyzed in terms that is best to answer the question at hand. It not so much about simply answering the question, but how the researcher plans to use that data within the benefit of the dissertation, It is possible that one day

that research will be used by decision makers in creating and implementing

policies, and in the author's case, in the homeland security community.

References

- Agarwal, V., & Taffler, R. J. (2007). Twenty-five years of the Taffler z-score model: Does it really have predictive ability? Accounting and Business Research, 37(4), 285–300. http://dx. doi. org/10. 1080/00014788. 2007. 9663313

- Allende-Alonso, S., Bouza-Herrera, C. N., Rizvi, S. E. H., & Sautto-Vallejo, J. M. (2019). Big Data and the Central Limit Theorem: A Statistical Legend. Investigación Operacional, 40, 112–123. Retrieved fromhttp://proxy1. ncu. edu/login? url= https://search-ebscohost-com. proxy1. ncu. edu/login. aspx? direct= true&db= a9h&AN= 133858001&site= eds-live

- Fendler, L., & Muzaffar, I. (2008). The history of the bell curve: Sorting and the idea of normal. Educational Theory, 58(1), 63–82. Retrieved fromhttps://sites. duke. edu/dukeaccessiblesyllabus/files/2018/04/BellCurve_FendlerMuzaffar. pdf

- Fox, J. A., Levin, J., & Forde, D. R. (2013). Elementary Statistics in Criminal Justice Research. Upper Saddle River, NJ: Prentice Hall.

- Kwak, S. G., & Kim, J. H. (2007). Central limit theorem: the cornerstone of modern statistics. orean journal of anesthesiology, 70(2), 144. http://dx. doi. org/10. 4097/kjae. 2017. 70. 2. 144

- Peng, S. (2019). Law of large numbers and central limit theorem under nonlinear expectations. Probability, Uncertainty and Quantitative Risk, 4(1), 1–8. http://dx. doi. org/10. 1186/s41546-019-0038-2

- Weiers, R. M. (2011). Introduction to business statistics (7th ed.). Boston, MA: Cengage Learning.