Z-scores, Probability, & Hypothesis Testing Worksheet

1. A normal distribution of scores has a standard deviation of 10. Find the z-scores corresponding to each of the following values:
	1. A score that is 20 points above the mean.

 20/10 = 2

* 1. A score that is 10 points below the mean.

-10/10 = -1

* 1. A score that is 15 points above the mean

15/10 = 1.5

* 1. A score that is 30 points below the mean.

-30/10 = -3

Because this is a normal distribution, the mean is equal to zero. When a score is below the mean, it is negative.

1. The Welcher Adult Intelligence Test Scale is composed of a number of subtests. On one subtest, the mean is 35 (*SD* = 6) and the distribution is normal.
	1. What number represents the 65th percentile (what number separates the lower 65% of the distribution)? (.39\*6) + 35 = 37.34
	2. What number represents the 90th percentile? (1.28\*6) + 35 = 42.68
	3. What is the probability of getting a raw score between 28 and 38? .3790 + .1915 = .5705 or 57%
	4. What is the probability of getting a raw score between 41 and 44? .4332 - .3413 = .0919 or 9%
2. In 2013, data from collected from in-coming freshman across the state of Georgia. The average age at which children took their first alcoholic drink was 14.6 years old (*SD* = 5). A researcher suspects that a sample of KSU students started to drink at a younger age than that of the population. A sample of 144 KSU students are asked to indicate the age at which they took their first alcoholic drink. Participants reported a mean age of 13.3 years (*SD* = 4.8). Did KSU students initiate drinking at an earlier age in comparison to the population of Georgia students? Using an alpha level of .05, answer the following questions.
	1. What are the null and alternative hypotheses?
* Null Hypothesis, Ho: μdrinking now = 14.6: There is no difference in the age of initial alcohol consumption of KSU students in comparison to the population of Georgia students.
* Alternative Hypothesis, Ha: *μ*drinking now < 14.6. KSU students’ initial age of alcohol consumption is earlier than the population of Georgia students.
	1. Draw a normal curve and indicate where the z-scores corresponding to sample mean fall and indicate where the critical region is located.
	2. Indicate whether you would reject or fail to reject the null hypothesis. Explain your decision.

$$Stand error of the mean=\frac{σ}{√n}=\frac{5}{√144}=\frac{5}{12}=.4166666667$$

$$z= \frac{M-μ}{σ\_{M}}=\frac{13.3-14.6}{.4166666667}=\frac{-1.3}{.4166666667}=-3.12$$

* The alpha level is set at .05. Thus, p-value should be less than .05 before I can reject the null hypothesis. The z-score associated with .05 is -1.645 (between 1.64 and 1.65). Thus, the critical region is beyond -1.645 (i.e., in the tail).
* Our z-score of -3.12 is less than -1.645 and falls in the critical region. In addition, the probability of finding a z-score of 3.12 is .0009, which is less than .05. Based on this information, I would reject null the hypothesis.
	1. Report the results in APA style.
* KSU students began drinking at a younger age (*M* = 13.3, *SD* = 4.8) than the population of Georgia students (*M* = 14.6, *SD* = 5), *z* = -3.12, *p* = .0009.
1. Scores on the SAT form a normal distribution with and. What is the minimum score necessary to be in the top 15% of the SAT distribution?
* First, using the normal unit table we find the z-score associated with the top 15% of scores. Looking at the proportion in the tail (.1492 is the proportion listed on the table that is closest to .15), we see that the z-score associated with .15 is 1.04. Next, we use the z-score formula and solve for X.



The minimum SAT score necessary to be in the top 15% of the distribution is 604.

1. For a normal distribution, find the z-score that separates…
	1. The top 7% from the rest of the distribution.

Looking at the proportion in the tail, we see the normal unit table indicates that .0694 is closest to .07. The z-score associated with .0694 is 1.48.

* 1. The lowest 10% from the rest of the distribution.

Looking at the proportion in the tail, we see the normal unit table indicates that .1003 is closest to .10. The z-score associated with .1003 is 1.28.

1. For the numbers below, find the area between the mean and the z-score:
	1. *z* = 1.17

Looking at the proportion in the column D of the normal unit table, we see that .3790 or 38% of scores fall between the mean and a z-score of 1.17.

* 1. *z* = -1.37

Looking at the proportion in the column D of the normal unit table, we see that .4147 or 41% of scores fall between the mean and a z-score of -1.37.

1. For the z-scores below, find the percentage of individuals scoring below:
	1. -0.47

Because the score is negative, we will look at the value for the tail of the distribution (column C) on the normal unit table. We see that the number in column C that corresponds to a z-score of -.47 is .3192 or 32%.

* 1. 2.24

Because the score is positive, we will look at the value for the body of the distribution (column B) on the normal unit table. We see that the number in column B that corresponds to a z-score of 2.24 is 98.75 or 99%.

1. A normal distribution of scores has a standard deviation of 10. Find the z-scores corresponding to each of the following values:
	1. A score of 60, where the mean score of the sample data values is 40.

60-40/10 = 20/10 = 2

* 1. A score that is 30 points below the mean.

-30/10 = -3

* 1. A score of 80, where the mean score of the sample data values is 30.

80-30/10 = 50/10 = 5

* 1. A score of 20, where the mean score of the sample data values is 50.

20-50/10 = -30/10 = -3

1. A soft-drink bottle vendor claims that its process yields bottles with a mean internal strength of 157 psi (pounds per square inch) (**3 psi), normally distributed. As part of its vendor surveillance, bottles are sampled from the production line to verify the vendor’s claim.
	1. Suppose the bottler randomly selected 40 bottles from the production line. What is the expected mean and the standard error of the mean?

The expected mean of the sample is 157, *n* = 40. That is, we would expect that our same mean looks like population mean. The formula for the standard error of the mean is . So, we plug in our values for the *SD* and *n*, and we get 0.474 for the standard error.

1. In a study of 10,000 male participants who smoke at least two packs of cigarettes daily, the mean life span is 65.3 years (*SD* = 3.4). If a sample of 40 individuals is selected, find the probability that the mean life span of the sample is less than the retirement age of 65 years.

$$Stand error of the mean=\frac{σ}{√n}=\frac{3.4}{√40}=\frac{3.4}{6.32455532}=.5375872023$$

$$z= \frac{M-μ}{σ\_{M}}=\frac{65-65.3}{.5375872023}=\frac{.3}{.5375872023}=.5580489988$$

The probability of the sample mean being less than 65 is .2877 or 28.77%.

1. Your company markets a computerized device for detecting high blood pressure. The device measures an individual’s blood pressure once per hour at a randomly selected time throughout a 12-hour period. Then it calculates the mean systolic (top number) pressure for the sample of measurements. Based on the sample results, the device determines whether there is significant evidence that the individual’s actual mean systolic pressure is greater than 130. If so, it recommends that the person seek medical attention.
	1. State appropriate null and alternative hypotheses in this setting.

Ho: *μ* = 130

Ha: *μ* > 130

* 1. Describe a Type I and a Type II error, and explain the consequences of each.

Type I error: Telling individuals that they have high systolic blood pressure when in fact they do not. This can lead to prescribing medication for an individual that is unneeded.

Type II error: Failing to tell individuals they have high systolic blood pressure when they do. This can lead to serious problems because their condition is going untreated.

* 1. The blood pressure device can be adjusted to decrease one error probability at the cost of an increase in the other error probability. Which error probability do you think is worse in this situation? Explain.

A type II error is worse in this situation because if it goes undiagnosed, it can cause serious health risks. It would be best to minimize this probability.

Questions modified from the following sources:

Questions 1, 2, 5, 6, 7, 8, 9: <http://www.wsfcs.k12.nc.us/cms/lib/NC01001395/Centricity/Domain/9738/Z-score%20worksheet%20solutions.doc>

Question 3: <http://www2.swccd.edu/~bsmith/m119/archives/09summer/lectures/worksheet%20Exercises%20in%20HypothesisTesting.pdf>

Question 4: <http://www.rachaelwelder.com/files/stat-216/Worksheet_6_2_key.pdf>

Questions 10, 11:

<http://www.mceachernhigh.org/member/teachers/class_documents/ma115541/AP-Statistics/Review%20WS%20Chapter%209%20with%20answers.pdf>

Questions 12, 13:

<http://webcache.googleusercontent.com/search?q=cache:7UsqoK3CKWUJ:www.bisd.net/cms/lib02/TX01001322/Centricity/Domain/1353/Type%2520I%2520and%2520II%2520error%2520and%2520power%2520worksheet%2520-%2520ANS.doc+&cd=5&hl=en&ct=clnk&gl=us>