**Homework #3:**
**Properties and Assumptions for Inferential Statistics & Introduction to Hypothesis Testing**

**<u>Descriptive Statistics</u>**
Problems 1 – 2

**<u>Properties for Inferential Statistics:</u>**
Problems 1 – 5
*Optional*: 6

**<u>Hypothesis Testing</u>**:
7, 8 (Hint: Think about how sampling might affect decision errors),
9 (do two-tailed tests only), 10, 11

*Optional*: Explain your answer to Question 10 to someone who has never had a course in statistics (but who is familiar with mean, standard deviation, and Z scores).

**<u>Statistical Tests</u>**: None this week.

**<u>SPSS Assignment</u>**: No assignment this week.

**Hints for Hypothesis Testing:**

- When answering, always remember to refer back to the research question when interpreting findings. Also, remember to state your conclusion in terms of the variables of the study (e.g., drug dose, gender of participant, etc.).

- Remember that no matter how appealing to believe, we never test the alternative hypothesis (e.g., Cognitive Behavioral Therapy will reduce anxiety). Instead, we test the *null hypothesis*. Thus, we can "reject" the null hypothesis in favor of our best thought-out alternative, the alternative hypothesis. Such an outcome, however, *does not prove* that the alternative hypothesis is correct. Terminology is very important on this assignment.

- If the question does not ask you to explain your answer, please make sure you read and understand the explanation given in the answers.

- When doing hypothesis testing, label the 4 major steps of hypothesis testing clearly (e.g., Step 1, Step 2, etc.) to differentiate them from the problems. Labeling will help you conceptualize the process of hypothesis testing and will help you identify a way to check that you have completed all parts.

**Homework #3:**
Questions

**Descriptive Statistics**

1. How do *outliers* (extreme scores) affect measures of central tendency?

2. What information does the variance and the mean provide about a distribution of scores?

**Properties for Inferential Statistics:**

1. Six months after a divorce, the former wife and husband each take a test that measures divorce adjustment. The wife's score is 63, and the husband's score is 59; over all the mean score for divorced women on this test is 60 with a standard deviation of 6; the mean score for divorced men on this test is 55 with a standard deviation of 4. If greater scores represent better adjustment, has the wife or the husband adjusted better to the divorce in relation to other divorced women and men of the same gender? Please explain your answer to yourself verbally as a means of testing whether you are able to describe the concepts appropriately.

2. A psychologist has studied eye fatigue using a particular measure, which she administers to students after they have worked for one hour typing on a computer. On this measure, she has found that the distribution follows a normal curve. Using a normal curve table, what percentage of students have Z scores that are:
    a. Above -1.5?
    b. Below 2.10?
    c. Above .45?
    d. Below -1.78?

3. Following from the eye fatigue problem above, the test of eye fatigue has a mean of 15 and a standard deviation of five. Using a normal curve table, what percentage of students have scores that are:
    a. Above 16?
    b. Below 18?
    c. Below 14?

4. Following from above, use a normal curve table to determine what the lowest score on the eye fatigue measure a person has to have to be in:
    a. The top 40%?
    b. The top 30%?
    c. The top 20%?

5. Suppose that you are designing an instrument panel for a large industrial machine. The machine requires the person using it to reach 2 feet from a particular position. The reach for this position for adult women is known to have a mean of 2.8 feet with a standard deviation of .5 feet. The reach for adult men is known to have a mean of 3.1 feet with a standard deviation of .6 feet. Both women's and men's reach from this position is distributed normally. If this machine design is implemented:
   a. What percentage of women will be unable to work on this instrument panel?
   b. What percentage of men will be unable to work on this instrument panel?

6. A large study of how people make future plans and the relation of this to their life satisfaction (Prenda & Lachman, 2001) recruited participants "through random digit dialing procedures". These are procedures by which phone numbers to call potential participants are randomly generated by a computer. Explain to a person who has never had a course in statistics :
   a. Why this method of sampling might be used?
   b. Why this method of sampling might be problematic if not everyone who the experimenters called agreed to be interviewed?

**<u>Introduction to Hypothesis Testing</u>:**

7. When a statistical result is not *extreme* enough to *reject* the *null hypothesis*, why is stating that your results support the null hypothesis incorrect?

8. Is the level of income for residents of a particular city different from the level of income for people in the region? For this research question:
   a. State which two populations are being compared.
   b. State the research hypothesis.
   c. State the null hypothesis.
   d. State whether you should use a one-tailed or two-tailed test and explain why.

9. Based on the information given for each of the following studies, decide whether to reject the null hypothesis. For each study, provide:
   a. the Z-score cutoff (or cutoffs if two tailed) on the comparison distribution at which the null hypothesis should be rejected
   b. the Z score on the comparison distribution for the sample score
   c. your conclusion about the study (e.g., reject or fail to reject the null hypothesis)

| Study | Population $\mu$ | Population $\sigma$ | Sample Score (not the sample mean) | $\alpha$ | Tails of Test |
|---|---|---|---|---|---|
| A | 10 | 2 | 14 | .05 | 1 |
| B | 10 | 2 | 14 | .05 | 2 |
| C | 10 | 2 | 14 | .01 | 1 |
| D | 10 | 2 | 14 | .01 | 2 |
| E | 10 | 4 | 14 | .05 | 1 |

10. A psychologist studying the senses of taste and smell has carried out many studies in which students are given each of 20 different foods. She administers each food by dropping a liquid on the tongue. Based on her past research, she knows that for students overall at the university, the mean number of the 20 foods that students can identify correctly is 14, with a standard deviation of 4, and she knows that the distribution of scores follows a normal curve. The psychologist wants to know whether people's accuracy on the task has more to do with smell than with taste. In other words, she wants to test whether people perform better or worse on the task when they are only able to taste the liquid compared to when they can both taste it and smell it. She sets up a special procedure that keeps a person from using the sense of smell during the task. The psychologist then tries the procedure on one student selected at random from her class. This student identifies only 5 correctly.
    a. Using the .05 significance level, what should the psychologist conclude?

11. Explain why you cannot say the result "proves" the research hypothesis when a result is statistically significant.

BONUS PRACTICE

1. The data from the table below are saved in a datafile named "*tuition_campus.sav*". Use SPSS to determine if you would consider Campus #13 to be an outlier. If so, on what basis would you decide this?

| AVERAGE % TUITION INCREASE PER CAMPUS | NUMBER OF FTE STUDENTS PER CAMPUS |
|---|---|
| Campus 1: 2.2% | 8,037 |
| Campus 2: 2.3% | 7,241 |
| Campus 3: 2.8% | 11,760 |
| Campus 4: 2.8% | 13,789 |
| Campus 5: 3.4% | 6,452 |
| Campus 6: 3.7% | 4,239 |
| Campus 7: 3.7% | 12,365 |
| Campus 8: 3.7% | 9,452 |
| Campus 9: 4.6% | 3,563 |
| Campus 10: 4.8% | 4,223 |
| Campus 11: 5.1% | 3,335 |
| Campus 12: 47.3% | 423 |
| Campus 13: 89.8% | 217 |
| Average % increase = 13.55% | |

**Homework #3:**
Answers

**Descriptive Statistics:**

1. **Effects of Outliers**
   Measures of central tendency are the mean, median, and mode.
   a. The mean represents that arithmetic average of all scores in the distribution; calculated by taking the sum of all scores in a distribution, divided by number of scores in a distribution, $\frac{\sum_{i=1}^{n} X_i}{N}$. Thus, outliers can influence the mean greatly.

   b. The median represents the point at or below which 50% of the scores fall in a distribution when scores are arranged in numerical order. When there is an odd number of scores the median is the middle score that exists in the data set. When there is an even number of scores the median is the average of the two middle scores in the data set. Outliers do not influence the median.

   c. The mode represents the most frequently occurring score(s) in a distribution. Outliers do not influence the mode.

2. **Variance and Mean**

   Mean and Variance tell you different things about a distribution. Mean is a measure of central tendency, a representative value that tells you about the center of a group of scores. Variance tells you about the dispersion of scores in a distribution around the mean. A small variance informs you that scores are close to the mean; scores that are far away from the mean are not common. By contrast, a large variance informs you that scores are distributed farther from the mean; scores that are far away from the mean are common. Mean and variance are descriptive statistics and are used in inferential statistics.

**Properties and Assumptions for Inferential Statistics:**

1. We first need to determine how many standardized units both the ex-wife and the ex-husband deviate from the average of divorced women and divorced men, respectively. Using respective group means and standard deviations:

   Wife: $Z = \frac{X - \overline{X}}{SD} = \frac{63 - 60}{6} = .5$ Husband: $Z = \frac{X - \overline{X}}{SD} = \frac{59 - 55}{4} = 1$

   Therefore, the ex-husband has adjusted better in relation to other divorced men than his ex-wife has adjusted in relation to other divorced women.

   For wives, a score of 63 is three points higher than the average of 60 for divorced women. There is, of course, some variation in scores among divorced women. The approximate average amount that women's scores differ from the average score is six points; this is the standard deviation referred to in the problem. Thus, the wife's score is only half is far above the mean of women's scores than the husband's score is above the mean of men's scores. Her Z score of .5 allows us to compare her adjustment to other divorced women's adjustment. We can see that she has adjusted slightly above average for women. Using the same logic, the husband's divorce-adjustment Z score of 1 is as much above the mean as the average amount that men differ from the mean (e.g., 1 SD). Although the husband's raw score of 55 is lower than his wife's score of 63, he has adjusted better than his wife, relative to other men and women. However, because both of them have positive Z scores, they have adjusted better than their corresponding gender average.

**2.**

- **a.** A Z score of -1.5 corresponds to 43.32% in the negative tail of the Z distributions, but you also must add the 50% from the positive tail; thus 43.32 + 50 = 93.32%
- **b.** A Z score of 2.10 corresponds to 1.79 in the positive tail; thus below 2.10 is 100-1.79 = 98.21%
- **c.** A Z score above .45 = 32.64%
- **d.** A Z score below -1.78 = 3.75%

*Tips to address common errors*: If you calculate the percent in the upper tail for a positive Z score, the percentage cannot exceed 50%. If you are calculating in the lower tail for a negative Z score the percentage cannot exceed 50%.

**3.** Use $Z = \dfrac{X - \bar{X}}{SD}$

- **a.** $Z = \dfrac{16-15}{5} = .2$

  From the normal curve table, a percentage in the tail for a Z score of .2 = 42.07%

- **b.** $Z = \dfrac{18-15}{5} = .6 \rightarrow 32.57 + 50 = 72.57\%$

- **c.** $Z = \dfrac{14-15}{5} = -.20 \rightarrow 42.07\%$

**4.** Use $Z = \dfrac{X - \bar{X}}{SD}$; solving for X=$(Z)(SD) + \bar{X}$

- **a.** The top 40% means that 40% of the distribution remains in the tail; the nearest Z score from the normal curve table for 40% in the tail is Z = .25 (at 40.13%) . A Z score of .25 equals a raw score of $(Z)(SD) + \bar{X} = X; (.25)(5) + 15 = 16.25$
- **b.** 17.6
- **c.** 19.2

**5.** The reach must be 2 feet, so we need to determine what percentage of women have a reach of less than 2 feet. To do so, calculate the Z score and then the corresponding percentage in the tail beyond the 2 feet point. Make sure to describe the calculations and outcome in terms of the variables under study.

a. For women who will be unable to work on the machine, $Z = \frac{2-2.8}{.5} = \frac{-.8}{.5} = -1.6 \rightarrow 5.48\%$. Women who have a *reach* LESS than *2 feet*, which corresponds to a Z score of -1.6, will be unable to use the machine. No job for you!

b. The reach is the same, but the distribution changes. $Z = \frac{2-3.1}{.6} = \frac{-1.1}{.6} = -1.8333 \rightarrow 3.36\%$. Therefore, 3.36% of men cannot work the machine because their reach is shorter than 2 feet. No job for you!

**6.**

a. Researchers might choose to use random number dialing to sample because phone surveys are potentially less expensive than other methods of collecting samples for surveys, and because many people own phones.

b. Random number dialing can create biased sampling. Due to changes in telephone ownership habits in recent years, many people do not own a home phone or are more likely to use cell phones if they do own a home phone. Because each person in the population does not have an equal likelihood of being selected in random number dialing, this method is not truly random sampling. If a researcher chose to limit the population of their study to individuals with home phones, self-selection bias would be an additional problem. If a sample only includes people who agree to be in a study, it is likely that these people systematically differ from people who choose not to participate. These systematic differences may include age, personality, or other variables that could be confounds for the study.

**<u>Hypothesis Testing</u>:**

**7.** The research hypothesis may be correct, but the result in a particular sample was not extreme enough for the rejection of the null hypothesis. In other words, the research design or the statistical analysis might not have been *powerful* enough to detect differences that truly do exist in nature. The concern is one of statistical power.

**8.**

a. Population 1: people who live in a particular city; Population 2: all people who live in the region.

b. Populations 1 and 2 have different mean incomes; $\mu_1 = \mu_2$

c. Populations 1 and 2 have the same mean incomes; $\mu_1 \neq \mu_2$

d. Two-tailed because the question being asked is whether the income of the people in the city is different from those in the region as a whole; a difference in either direction is of interest. The test would be one-detailed if the researchers were interested in a directional outcome between incomes.

**9.** Consider, the similarity $Z = \frac{X - \overline{X}}{SD} \approx z = \frac{\overline{X} - \mu}{\sigma}$

| Study | (a) Cutoff(s) | (b) Z score on Comparison Distribution | (c) Decision |
|---|---|---|---|
| A | Closest to 5% (one-tailed) in the tail of Z dist is 5.05%, so +1.64 | $z = \frac{14-10}{2} = \frac{4}{2} = 2$ | Reject the null hypothesis |
| B | Closest to 5% (two-tailed) in the tail of Z dist is 2.5%, so $\pm 1.96$ | 2 | Reject the null hypothesis |
| C | +2.33 | 2 | Fail to reject null hypothesis; Inconclusive |
| D | ±2.57 | 2 | Fail to reject null hypothesis; Inconclusive |
| E | +1.64 | $z = \frac{14-10}{4} = \frac{4}{4} = 1$ | Fail to reject null hypothesis; Inconclusive |

**10.** **<u>Step 1: Restate the question as a research hypothesis and formulate a null hypothesis about the population.</u>**

Pop. 1 = Students who are prevented from using their sense of smell (her student).

Pop. 2 = Students who use sense of smell and taste (students in general).

*Research hypothesis*: Students prevented from using their sense of smell will perform worse on the task than students in general.

*Null hypothesis*: Students prevented from using their sense of smell will not perform worse on the task than students in general.

**<u>Step 2: Select an appropriate statistic for testing your hypothesis and determine the characteristics of the comparison distribution.</u>**

We are comparing an individual score (X) to the Population of ALL students in general (Pop #2). Pop. 2 takes the shape of a normal distribution; we adopt the properties of the normal distribution to Pop #2. The comparison distribution is Population #2, with $\mu$= 14 and $\sigma$= 4.

**<u>Step 3: Determine the cutoff sample score(s) on the comparison distribution at which the null hypothesis should be rejected.</u>**

For a two-tailed test at $\alpha = .05$ (2.5% in each tail), the Z cutoff is ±1.96. The cutoff is either positive or negative because the outcome could lead to better or worse performance.

**<u>Step 4: Obtain data and compare your data to the comparison distribution so that you can make a decision about rejecting the null hypothesis.</u>**

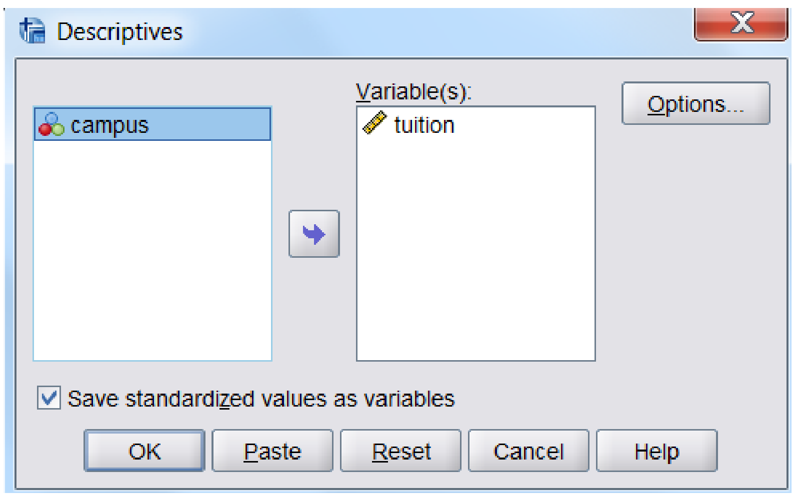
The individual score was 5. Converting this score to a Z score, Z = (5-14)/4= -2.25.

A Z score of -2.25 is more extreme than the alpha cutoff Z score ±1.96. You should reject the null hypothesis.  If this student represents the population, he or she would likely not score this extremely; the negative score indicates that removing smell reduces performance.

**11.** We cannot state that a statistically-significant result *proves* a research hypothesis, or proves anything because with statistical analyses, we deal only with *statements of probability*, or likelihood, of events occurring by chance factors alone. We can, however, state that a statistically-significant result supports, or indicates, the difference between a sample mean and a population $\left(\overline{X}-\mu\right)$ is "likely" great enough that it must not have occurred by chance alone. In fact, the likelihood may be 5% or 1% (or .05 or .01 alpha-level probabilities). In addition, perhaps a different research hypothesis is the best or accurate explanation for the cause of a statistical difference between groups or populations.

BONUS PRACTICE

1. Go to Analyze →Descriptive Statistics and move the `tuition` variable to the variables box. Click on "Save standardized values as variables". Choose your options and make sure you select the mean and standard deviation.




Campus #13 is almost 3 SD above the sample mean of 13.55%. Scores that are this extreme are not like the other scores. They are unlikely to occur and they influence the mean greatly. In fact, the mean is probably not the best measure of central tendency because the distribution is skewed positively. Without Campus #13, the mean is only 7.2.

Additional Questions for HW4

1. Suppose you want to study the number of errors people make on a simple counting task. You ask participant judges to count and then report the number of people entering a major department store in one morning between the hours of 8 and 11 am. We know that the distribution of the number of reported shoppers (by the judges) is distributed normally with mean of 975 and a standard deviation of 15. One judge counted only 950 shoppers entering the store. You want to determine whether this judge's response reflects the typical error expected by a conscientious counter. Is this judge's response a surprising answer if he counted conscientiously?
(Statistical Methods for Psychology, 8th edition, Howell, p. 79)

2. What is a p-value?

**1. Surprising Judge Problem**
Answer this problem by calculating z-score in order to determine the probability of a certain score deviating from the sample mean.

Mean= 975
Standard Deviation= 15
Judge's Individual Score = 950

$$Z = \frac{X - \overline{X}}{SD}$$

$$Z = \frac{950 - 975}{15} = -1.67$$

Next you find this value on a Z table to find the probability of getting counts this low or lower. With a two-tailed α = .05, we see that the smaller portion of the curve 0.04746. We would expect counts as low as 950 people only 4.75% of the time.

Using a p-value of 0.05, this is a significant result. Our judge had a surprising answer if he is a conscientious counter.

**2. Define a p-value.**
A p-value is the probability of an event occurring, given that the null hypothesis is true and all assumptions of the statistical test are met. Different statistical tests have different assumptions that must be met in order to have valid results. Z-tests assume random sampling from a normally distributed population.

When your data is randomly sampled from a normally distributed population, the p-value is an accurate probability of the event occurring under the null hypothesis. When a population is not normally distributed (e.g. an extremely skewed population), the p-value is misleading. It is important to check that your data fit the assumptions of a statistical test before running analyses and drawing conclusions.