|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DATA ANALYTICS FOR INFERENTIAL STATISTICS** ***Parameter Estimation and Hypothesis Testing******--Show All Work--***     The ***purposes*** of this assignment are to demonstrate some of the *processes* and *procedures* used in statistical inference.  You will perform *three* major tasks:  (1) construct an empirical sampling distribution, (2) select a simple random sample and construct confidence intervals, and (3) test several hypotheses.***A.  Constructing a Sampling Distribution*** 1.  The concept of *sampling distribution* is one of the most important concepts in inferential statistics.  The goal of this exercise is to demonstrate the construction and meaningfulness of an *empirical sampling distribution*. a. Suppose you have a population (N=5) comprised of the following five sums:      2, 3, 4, 5, and 6.* Compute the *population mean* and *population standard deviation*.
* Construct the *sampling distribution* for these five amounts (hint: See handout entitled "Sampling Distribution: CT, V, and Form" and arrange the information in a data matrix) as if you were selecting samples two (i.e., n=2) at a time with replacement.
* Compute the *mean of the sampling distribution* and the *standard error of the mean* for this empirical sampling distribution.

**NOTE**:  Pay special attention to the correspondence between the population mean and sampling distribution mean and the population standard deviation and the standard deviation of the sampling distribution (called the *standard error of the mean*).  This relationship is the key to understanding how sampling from a population (rather than studying it exhaustively) is so important and useful in statistics and social research.* Construct a *bar graph* depicting the distribution of these sample means. This illustrates (part of) the *central limit theorem.*

***B.  Drawing a Simple Random Sample*** 2.  Select a *simple random sample* ("*SRS*") of 15 teachers (n=15) from the universe of teachers.  This necessitates consulting a table of random digits (provided in the Appendix of **Data Analytics**) and selecting 15 teachers.  Begin with Row #00000 and use the third five-digit column.  Sample *without* replacement using the handout for selecting a simple random sample as a guide.**NOTE**:  When selecting a probability sample the difference between the *population parameter* and the *sample statistic* is due to what is called *random sampling error*.* Then compute the *sample mean* (http://lilt.ilstu.edu/wleonard/images/image007.gif) and *sample standard deviation* (*s)* for the 15 randomly selected teachers for the salary variable.
* Construct the 99%, 95%, and 90% *confidence intervals* for your sample and express it using the proper notation, e.g., C (u) = .99 or .95 or .90.  Notice what happens as the confidence interval is lowered from the 99% to 90%.
* Interpret the meaning of these confidence intervals.

**NOTE**:  The purpose of this assignment is to estimate the population mean and standard deviation (the *parameter* values for all teachers based on a randomly selected sample) for the variable "salary".  This inferential technique is known as *parameter estimation*.  Since you know the mean for *all* teachers, you can assess how similar the *sample* mean is to the *population* mean.  If the sample mean is similar to the population mean, you have discovered for yourself why sampling -- taking a "part" rather than the "whole" -- is such an important and viable research procedure.***C.  Computing and Interpreting Measures of Statistical Significance (Hypothesis Testing)*** 3.  Determine whether there is a statistically significant difference between the teachers' salaries (#'s 1 through 11) and (#'s 30 through 41) (Hint: consider these two "groups" to be randomly selected samples from two populations: hence, you will employ a two-sample test–the *t-ratio (or z-test)*--appropriate for the level of measurement of the variable).* State the *null* (Ho) and *alternative* (H1) hypotheses.
* Test these data at the .001, .01 and .05 *levels of significance (p-values)*.
* c. What is your decision?  Do you accept or reject the null hypothesis at the various *p* levels?

 4.  Using the data in the 3 x 3 *contingency table* for question #5 of “Data Analytics for Bivariate Distributions”, determine whether the variables are independent or related.   Use the *chi square test of independence* to answer this question.* State the *null* (Ho) and *alternative* (H1) hypotheses.
* Test these data at the .001, .01 and .05 *levels of significance (p-values)*.
* What is your decision?  Do you accept or reject the null hypothesis at the various levels?

|  |  |
| --- | --- |
|  |  |
|  |
|  |  |  |  |  |  |  |
|  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

 |
|  |
|  |

5. For this statistical problem you are playing the role of a gambler (in Las Vegas) on horseracing. You want to know if the postpositions of horses makes a difference in their winning. You task is to determine if there is a statistically significant difference in postpositions.

Here are the post positions and the number of wins in the Kentucky Derby:

Post Wins

1 – 7
2 – 7
3 – 8
4 – 11
5 – 11
6 – 16
7 – 6
8 – 11
9 – 4
10 – 10
11 – 13
12 – 3
13 – 4
14 – 2
15 – 3
16 – 3
17 – 0
18 – 0
19 – 3
20 - 2

* State the *null* (Ho) and *alternative* (H1) hypotheses.
* Test these data at the .001, .01 and .05 *levels of significance (p-values)*.
* What is your decision?  Do you accept or reject the null hypothesis at the various levels?

 6. Suppose that the federal government wants to test the effect of three different training programs on the socialization of G.I.'s. Ten recruits are randomly assigned to three different training programs. "Effective military socialization scores" are deter­mined at the end of the training period. The results are below. For these data, follow the steps listed in completing Fisher's *analysis of variance test* for *k* independent samples.

|  |  |  |
| --- | --- | --- |
|  | *Training Programs* |  |
| *Type I**x* | *Type II**x* | *Type III**x* |
| 5 | 3 | 2 |
| 7 | 8 | 2 |
| 4 | 4 | 8 |
| 2 | 6 | 6 |
| 5 | 10 | 9 |
| 9 | 6 | 10 |
| 10 | 8 | 15 |
| 4 | 7 | 4 |
| 6 | 9 | 7 |
| 3 | 4 | 5 |
|

|  |  |  |
| --- | --- | --- |
| 6 | 7 | 2 |
| 7 | 8 | 2 |
| 3 | 4 | 7 |
| 7 | 8 | 5 |
| 5 | 11 | 4 |
| 9 | 6 | 10 |
| 8 | 8 | 5 |
| 4 | 8 | 4 |
| 6 | 9 | 5 |
| 7 | 7 | 6 |

 |  |  |

* 1. Calculate the total sum of squares.
	2. Compute the within sum of squares.
	3. Calculate the between sum of squares.
	4. Determine *df for* *TSS, BSS and* *WSS.*
	5. Determine the variance estimates for *BSS* and *WSS.*
	6. Compute the *F* ratio.

g. Is there a significant difference at the .01 level?

h. Is there a significant difference at the .05 level?

* State the *null* (Ho) and *alternative* (H1) hypotheses.
* Test these data at the .01 and .05 *levels of significance (p-values)*.
* What is your decision?  Do you accept or reject the null hypothesis at the various levels?

7. Using the Factorial ANOVA handout complete the steps listed and provide a PASW printout and interpretation of the findings