**Assignment 1: Final Exam**

The purpose of this exam is to give you the opportunity to formulate research questions, run the analyses, and interpret the results of the statistics covered in this class. **The final exam is due by midnight on the due date, Friday, October 16, 2015, as a single Microsoft Word document placed in the dropbox. The document must be in APA format.**

**Please read through the entire exam first.**

The exam consists of three sections

I. The research scenario—to provide the context for the data. Please note that the data are “content-neutral,” i.e., they do not refer to a specific discipline or field.

II. The codebook—this identifies the variables (names, labels, and measurement scale) in the database.

III. The exam instructions—for completing the exam. Be sure to read each question carefully and answer each question completely.

**I. Research Scenario:**

An organization wants to know if participants with varying levels of expertise (professionals, paraprofessionals, and nonprofessionals) improve their knowledge after completing a training program.

The organization collected demographic information: gender, age, type of training (professional, paraprofessional, or nonprofessional), location of the worksite (on-site or off-site) and years of experience.

A pre-training test of knowledge, a training program, and post-training test of knowledge was developed. Participants were tested, then participated in the three-week training program, and then were tested again.

The dataset also includes (1) a measure of participant confidence in knowledge and (2) a certification exam score.

The data are discipline-neutral. Therefore, part of your final project is to create a context for the research that is associated with your discipline or area of interest (e.g., training to assess mental health status; training to work with special education children; training to become a technician or consultant).

**II. Codebook**

Variable Information

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Label | Measurement Scale | Catagory Name |
| ID | N/A | N/A | N/A |
| Gender | Gender | Nominal | 1 = Male 2 = Female |
| age | Age in Years | Ratio |  |
| |  | | --- | | qualification | | |  | | --- | | Professional Qualification | | Nominal | 1 = Professional  2 = Paraprofessional  3 = Nonprofessional |
| worksite | Location of Work | Nominal | 1 = On-Site 2 = Off-Site |
| knowledge1 | Level of knowledge before Training | Interval | N/A |
| knowledge2 | Level of knowledge after Training | Interval | N/A |
| years | Years of Experience | Ratio | N/A |
| confidence | Confidence in knowledge | Interval | N/A |
| exam | Certification in knowledge | Interval | N/A |

**III. Exam Instructions**

Overview  
Your task is to review the dataset, formulate a context, and then use your knowledge of statistics to answer the research questions and test hypotheses that will help the organization evaluate the effectiveness of the program.

Part I. Create your context.

Using the research scenario and variables identified in the codebook, create a “story” that describes the purpose and focus of the study. In a few paragraphs describe the intent of your investigation in the form of a **problem background** and **purpose statement**.

Part II. Describe your sample.

Generate frequency tables and bar charts for the nominal variables. Generate and interpret descriptive statistics of central tendency, variability, skewness, and kurtosis for the continuous (scale) variables. Generate frequency tables and histograms with the normal curve superimposed for each scale variable. Label your tables and graphs according to APA format. Conclude with a paragraph summarizing the demographic characteristics of this sample.

1. Gender (nominal)
2. Age (scale)
3. Qualification (nominal)
4. Worksite (nominal)
5. Knowledge1 (scale)
6. Knowledge2 (scale)
7. Years (scale)
8. Confidence (scale)
9. Exam (scale)

Part III. Describe relationships among the variables.

Select the variables that are measured on interval or ratio scales. Create a correlation matrix.  Label the table according to APA format. Identify and discuss the strongest and weakest correlations.

Part IV. Answer any **FOUR**of the nine following research questions. Clearly identify which question(s) you have selected to answer by writing the number associated with the question (1-9) before each answer. Make sure to select four questions that must be answered by 4 **DIFFERENT**tests. For example. One t-test, one ANOVA, one regression and one chi-square; **NOT** a single sample t-test, independent sample t-test and paired sample t-test as these all fall under t-test.

Based on the research scenario and the data, restate the research question, identify the Independent and Dependent Variable(s), formulate the hypothesis, conduct an analysis, and interpret the results for each of the following. Please use the 7 step hypothesis testing model for each analysis, including an APA-style conclusion statement summarizing the findings, interpreting the results and answering the research question.

1. Is there a significant difference in knowledge between on-site and off-site workers ***before*** they take the training?
2. Is there a significant increase in knowledge as a result of the training?
3. Do participants of different classifications (professional, paraprofessional, and nonprofessional) perform differently on the certification exam? Which group performs best?
4. What are the effects of gender and worksite location (on- or off-site) on level of confidence? In other words:
   1. Is there a gender difference in confidence?
   2. Does type of worksite experience affect confidence?
   3. Is there an interaction between gender and worksite in their effect on confidence?
5. After controlling for prior knowledge, is there a difference between professionals, paraprofessionals, and non-professionals in knowledge after the training?
6. Does age predict performance on the certification exam?
7. In addition to age, does confidence improve the ability to predict performance on the certification exam?
8. Does the distribution of the classifications of participants (professional, paraprofessional, and nonprofessional) differ than what one would expect by chance?
9. Does the distribution of the classifications (professional, paraprofessional, and nonprofessional) differ from the State data reported as 15%, 25% and 60%?

Part V. Summarize your findings.

Synthesize the results of your five analyses. Include a brief summary of the sample characteristics and the major findings. Interpret the findings so that the organization’s leaders will have an understanding of the similarities and differences in knowledge, and how effective the training program is in improving knowledge.