Part I. Computer work

Problem. ”Graduate Students Stipend”

Usually, the graduate departments in universities consider a stipend for their students who are willing to act as teaching assistants or TAs. The level of compensation varies from school to school and even over various departments of a particular university. Some departments like Computer Science or Business departments have traditionally more fundings hence they are able to pay their students with relatively higher salaries. In the following, you will be analyzing the data that show the stipends of 1350 students from a certain research-oriented university at east-coast USA. First, you need to download data. The data are in the course web page. The datafile you need is named income.sav.

Notice that the data has two columns. The first column are the incomes in thousands of dollars. The second column includes numbers from 1-5 denoting five different departments of that university. The departments can be identified as follows:

<table>
<thead>
<tr>
<th>Code</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mathematics</td>
</tr>
<tr>
<td>2</td>
<td>Business</td>
</tr>
<tr>
<td>3</td>
<td>Computer</td>
</tr>
<tr>
<td>4</td>
<td>Philosophy</td>
</tr>
<tr>
<td>5</td>
<td>Music</td>
</tr>
</tbody>
</table>
(a) In today’s lab we will practice the concepts behind the confidence interval using the graduate stipend data. The good news is that we know what the population means are so we can test the validity of our confidence intervals.

(b) To be able to analyze these data, we need to obtain the important statistics associated with each department. To do this, go to: Analyze → Descriptive Statistics → Explore. Choose income for the Dependent List and choose code for the Factor List. Click OK. Report the population mean, the population variance, and the population standard deviation for each department.

(c) Now, we need samples from each department!

I have provided the information for each department on separate files. For example, the information for the Math department is available at Math.sav on the webpage. So, the files that you need to look at are: Math.sav, Philosophy.sav, Business.sav, Computer.sav, and Music.sav.

Now, let’s take a simple random sample (SRS) of size 30 from the population of Math stipends. In SPSS this is done by choosing Data → Select Cases from the menu. In the dialog box that opens, choose the bullet for Random Samples of Cases and click the Sample button. Next, choose the bullet for Exactly and type 30 into the first box and 150 into the second box. Click the Continue button. Click the OK button to take the SRS of size 30 from the 150 Mathematics cases. SPSS indicates which cases are included in the SRS by putting a slash through the case number if the case is not included. (NOTE: everyone will likely get a different answer since the simple random sample is random.)

(d) Take a sample of size 30 from each department. For each department calculate the sample mean and the sample standard deviation of the selected cases.

(e) Find the 95% confidence interval for \( \mu \), the average stipend of students in each department. Here, you don’t need to use SPSS. Interpret your results. Note that all numbers are rounded to 1000 Dollars.

(f) Knowing the population means, report whether your 95% confidence intervals covered the actual means of the populations.
(g) Test the null-hypothesis that suggests the mean stipend for each department is $\mu = 10,000$ Dollars. Carry out all tests at $\alpha = 0.05$.

(h) Having the means and the standard deviations of the populations:

1. Find the probability of a philosophy student having an income more than $10,000.
2. Find the probability of a mathematics student’s income being between $7000$ and $9500$.
3. Find the probability of a business student and a computer student each having salaries higher than $13000$.
4. Find the probability of a sample of 30 business students having an average salary higher than $10000$.
5. Find the probability of a sample of 30 computer students having an average salary higher than $10000$.
6. Find the probability of a sample of 30 mathematics students having an average salary higher than $10000$.
7. Find the probability of a sample of 30 philosophy students having an average salary higher than $10000$.
8. Find the probability of a sample of 30 music students having an average salary higher than $10000$.

(i) Since we know the sample mean and the sample standard deviation for each case, we should be able to conduct $t$-confidence intervals as well as $t$-tests.

For each department, construct the 95\% $t$-test for the population mean stipends $\mu$. Also, for each department, test the null hypothesis that suggests the mean population income is 10,000. Carry out those $t$-test at the $\alpha = 0.05$ level.
Part II. Questions relating to the Sample Mean Distribution.

(1) The number of pages of books in the university library has an average of 400 pages with a standard deviation of 50 pages. Suppose we pick 60 books from the library, what is the probability that the average of these 60 books exceed 375 pages?

(2) For a particular brand of cars, the gas consumption on traveling between Bakersfield and Los Angeles is 5 gallons. The gas consumption has a distribution with standard deviation of 0.4 gallons. There are 30 cars of this particular brand that travel together from Bakersfield to Los Angeles. What is the probability that these 30 cars consume less than 145 gallons together?

(3) A manufacturer claims that the average weight of a pack of pastilles is 140 grams with standard deviation of 3 grams. A random sample of 80 packages showed an average of 139 grams. Does this cast a doubt on the manufacturer’s claim that the average weight is 140 grams? Explain.

(4) The effect of anesthetic on the flow of aqueous humour (a fluid in the eye) was investigated in the paper "A Method for Near-Continuous Determination of Aqueous Humour Flow: Effects of Anesthetics, Temperature, and Indomethacin" (1984): 435-53. Summary quantities for aqueous flow rate (1/min) observed under three different anesthetics are given.

<table>
<thead>
<tr>
<th>Anesthetic</th>
<th>Sample size</th>
<th>Sample Mean</th>
<th>Population Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pentobarbital</td>
<td>191</td>
<td>0.99</td>
<td>0.235</td>
</tr>
<tr>
<td>Urethane</td>
<td>52</td>
<td>1.47</td>
<td>0.314</td>
</tr>
<tr>
<td>Ketamine</td>
<td>16</td>
<td>0.99</td>
<td>0.164</td>
</tr>
</tbody>
</table>

For each of the following parts
State the parameter and what is representing
State the null and alternative hypothesis
Calculate the test statistic
Determine the p-value
Make your decision and state your conclusion
(a) Test the hypothesis that the true mean flow rate under the effects of Pentobarbital is greater than 0.96.

(b) Test the hypothesis that the true mean flow rate under the effects of Urethane is less than 1.56.

(c) Test the hypothesis that the true mean flow rate under the effects of Ketamine differs from 1.08.

(5) Water permeability of concrete is an important characteristic in assessing suitability for various applications. Permeability can be measured by letting water flow across the surface and determining the amount lost (inches/hour). Suppose that the permeability index $X$ for a randomly selected concrete specimen of a particular type has a normal distribution with $\mu = 1000$ and $\sigma = 150$.

(a) How likely is it that a single specimen will have a permeability index between 850 and 1300?

(b) How likely is it that the average of four specimens will have a permeability index between 850 and 1300?

(c) How likely is it that the average of forty specimens will have a permeability index between 850 and 1300?