Homework 1. Solutions

Q1) The following are the 3 exam scores for 6 students.

<table>
<thead>
<tr>
<th>Name</th>
<th>Exam1</th>
<th>Exam2</th>
<th>Exam3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joe</td>
<td>75</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>Mary</td>
<td>88</td>
<td>88</td>
<td>97</td>
</tr>
<tr>
<td>Jim</td>
<td>65</td>
<td>05</td>
<td>100</td>
</tr>
<tr>
<td>Jane</td>
<td>100</td>
<td>99</td>
<td>78</td>
</tr>
<tr>
<td>Mike</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Sue</td>
<td>10</td>
<td>60</td>
<td>80</td>
</tr>
</tbody>
</table>

Use SAS to establish a file and print out of the six students, their 3 exam scores and their averages, in descending order of their averages.

Solution:

```
DATA Students;
INPUT Name $ Exam1 Exam2 Exam3;
Average = (Exam1 + Exam2 + Exam3)/3;
DATALINES;
    Joe    75    86    90
    Mary   88    88    97
    Jim    65    05    100
    Jane   100   99    78
    Mike   90    90    90
    Sue    10    60    80
;
PROC SORT DATA = Students;
   BY DESCENDING Average;
RUN;
PROC PRINT DATA = Students;
   TITLE 'Ranking from Three Exams';
RUN;
```
Q2) To determine whether glaucoma affects the corneal thickness, measurements were made in 8 people affected by glaucoma in one eye but not in the other. The corneal thickness (in microns) were as follows:

<table>
<thead>
<tr>
<th>Person</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye affected</td>
<td>488</td>
<td>478</td>
<td>480</td>
<td>426</td>
<td>440</td>
<td>410</td>
<td>458</td>
<td>460</td>
</tr>
<tr>
<td>Eye not affected</td>
<td>484</td>
<td>478</td>
<td>492</td>
<td>444</td>
<td>436</td>
<td>398</td>
<td>464</td>
<td>476</td>
</tr>
<tr>
<td>Difference</td>
<td>4</td>
<td>0</td>
<td>-12</td>
<td>-18</td>
<td>4</td>
<td>12</td>
<td>-6</td>
<td>-16</td>
</tr>
</tbody>
</table>

(a) According to the data, can you conclude, at the significance level of 0.10, that the corneal thickness is not equal for affected versus unaffected eyes?
(b) Calculate a 90% confidence interval for the mean difference in thickness.
(c) Please write the entire SAS code to check the assumptions necessary in (a) and to perform the test asked for in (a). (*Part C was not given as part of the quiz today.)*

**Solution:**

(a) Using $\bar{d} = -4$ and $s_d = 10.744$, the test statistic is

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{-4 - 0}{10.744/\sqrt{8}} = -1.053$$

Since $|t| < t_{8-0.05} = 1.895$, we can NOT reject $H_0$ at $\alpha = 0.10$. That is, we do NOT have enough evidence to support the claim that the average corneal thicknesses are affected by glaucoma.
(b) A 90% CI for $\mu_1 - \mu_2$ is given by

$$\bar{d} \pm t_{n-1,\alpha/2} \cdot \frac{s_d}{\sqrt{n}} = -4 \pm 1.895 \times \frac{10.744}{\sqrt{8}}$$

That is, $[-11.198, 3.198]$

(c) The SAS Code:

```sas
DATA Eyes;
INPUT Bad Good;
Diff = Bad - Good;
DATALINES;
488 484
478 478
480 492
426 444
440 436
410 398
458 464
460 476
RUN;
PROC UNIVARIATE DATA = Eyes NORMAL ALPHA = 0.1;
VAR Diff;
RUN;
PROC TTEST DATA = Eyes ALPHA = 0.1;
VAR Diff;
RUN;
PROC TTEST DATA = Eyes ALPHA = 0.1;
PAIRED Bad*Good;
RUN;

*** Alternatively, we can enter the data as follows:

DATA Eyes;
INPUT Bad Good @@;
Diff = Bad - Good;
DATALINES;
488 484 478 478 480 492 426 444 440 436 410 398 458 464 460 476
RUN;
```
Selected SAS Output:

The UNIVARIATE Procedure
Tests for Normality

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>W</td>
<td>0.94404 Pr &lt; W 0.6512</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.146749 Pr &gt; D &gt;0.1500</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.037488 Pr &gt; W-Sq &gt;0.2500</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.244817 Pr &gt; A-Sq &gt;0.2500</td>
</tr>
</tbody>
</table>

The TTEST Procedure

Mean 90% CL Mean Std Dev 90% CL Std Dev

DF t Value Pr > |t|
7 -1.05 0.3273

Interpretation:

T-test assumes normality. Null Hypothesis and Alternative Hypothesis for normality test:

\[ H_0: \text{The Population is Normal}, \]
\[ H_1: \text{The Population is Not Normal}. \]

The p-value for Shapiro-Wilk test is 0.6512. Since 0.6512 > \( \alpha = 0.1 \), we fail to reject the null hypothesis and conclude that the normality assumption is met and it is proper to use t-test.

Null Hypothesis and Alternative Hypothesis for t-test:

\[ H_0: \mu_d = 0 \]
\[ H_1: \mu_d \neq 0 \]

Since \( t_{0.05,7} = 1.895 \), and \( t_0 = -1.05 \), which is less than \( t_{0.05,7} \), we fail to reject \( H_0 \) at the significance level \( \alpha = 0.10 \). That is, we do NOT have enough evidence to support the claim that the average corneal thicknesses are affected by glaucoma.
Q3) Over the past 5 years, the mean time for a warehouse to fill a buyer’s order has been 25 minutes. Officials of the company believe that the length of time has increased recently, either due to a change in the workforce or due to a change in customer purchasing policies. The processing time (in minutes) was recorded for a random sample of 15 orders processed over the past month.

28 25 27 31 10 26 30 15 55 12 24 32 28 42 38

Use SAS to answer the questions:

(a) Please check the normality of the data.
(b) Please test the research hypothesis at the significance level $\alpha = 0.05$.

Solution:

```sas
DATA Buyers;
INPUT Time @@;
DATALINES;
28 25 27 31 10 26 30 15 55 12 24 32 28 42 38 ;
PROC UNIVARIATE DATA = Buyers NORMAL MU0 = 25;
   VAR Time;
RUN;
PROC TTEST DATA = Buyers H0 = 25 SIDES = U;
   VAR Time;
RUN;
```

Selected SAS Output:

```
The UNIVARIATE Procedure
   Variable: Time

Tests for Normality

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk</td>
<td>W</td>
<td>0.941665</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>D</td>
<td>0.169887</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>W-Sq</td>
<td>0.07693</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>A-Sq</td>
<td>0.428232</td>
</tr>
</tbody>
</table>
```
The TTEST Procedure
Variable: Time

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Std Err</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>15</td>
<td>28.2000</td>
<td>11.4405</td>
<td>2.9539</td>
<td>10.0000</td>
</tr>
</tbody>
</table>

Mean 95% CL Mean Std Dev 95% CL Std Dev

DF  t Value  Pr > |t|
14  1.08  0.1485

Interpretation:
(a) Null Hypothesis and Alternative Hypothesis:

\[ H_0 : \text{The Population is Normal,} \]
\[ H_1 : \text{The Population is Not Normal.} \]

The P-value of the Shapiro-Wilk test for normality is 0.4038. Since 0.4038 > \( \alpha = 0.1 \), we fail to reject \( H_0 \) the null hypothesis and conclude that the normality assumption is met and it is proper to use the t-test.

(b) Null Hypothesis and Alternative Hypothesis:

\[ H_0 : \mu = 25 \]
\[ H_1 : \mu > 25 \]

The significance level of this test is \( \alpha = 0.05 \). Since the p-value (one-sided test) is 0.1485, which is greater than \( \alpha = 0.05 \), we fail to reject \( H_0 \) at the significance level \( \alpha = 0.05 \). That is, we do NOT have enough evidence to conclude that the mean time for a warehouse to fill a buyer’s order is more than 25 minutes.

Note: Dear students, please install SAS, R and G*Power (**Go to Google and type: GPower – you will be given the link to the free download site) on your computers.