24th January

STAT 5023
HW #1

Achyutha Santhoshi
A20314248

(1) A Certain manufactured product is supposed to Contain 23% potassium by weight. A Sample of 10 Specimens of this product had an Average percentage of 23.2 With a Standard deviation of 0.17%. If the mean percentage is found to differ from 23, the manufacturing process will be Recalibrated.

a. State the appropriate null and alternative hypothesis

$H_0 : \mu = 23$
$H_1 : \mu \neq 23$

b. Compute the test Statistic

Test Statistic $Z = \dfrac{\bar{Y} - \mu_0}{\sigma / \sqrt{n}}$

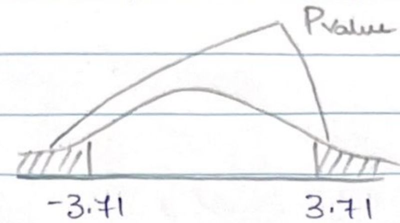
$= \dfrac{23.2 - 23}{\frac{0.17}{\sqrt{10}}} = \dfrac{0.2}{\left(\frac{0.17}{3.16}\right)} = \dfrac{0.2}{0.054} = 3.71$

$Z = 3.7$

$\therefore$ The Computed Value of Test Statistic $= \underline{3.71}$

c. Should the process be Recalibrated. Explain.

Since this is a two tail test



Pvalue = 2 * 0.00243

= 0.00485 (Using TDIST)

$P = 0.00485 < \alpha = 0.05$

∴ We Reject H_0

Hence, we should recalibrate the manufacturing process

d. What Assumptions you made in part (b) regarding the weights percentages of potassium.

We know that if the data is normally distributed & Standard deviation is know/given we Can Use test Statistic

∴ I have assumed that the Data is normally distributed & there is Already Standard deviation is given in the question.

(2) Fifty Specimens of a new Computer Chip were tested for Speed in a Certain application, along With another 50 Specimens of Chips With old design, the average Speed, in MHz, for the new Chips was 495.6 and Standard Deviation was 22.4. The average Speed for the old Chips was 481.2 and the Standard deviation was 20.3

a. Can you Conclude that the mean speed for the new Chips is greater than that of the old Chips? State the appropriate null and alternative hypotheses. and then give your decision Using P-Value.

Test: $H_0: \mu_1 \le \mu_2 \Rightarrow H_0: \mu_1 - \mu_2 \le 0$

$H_1: \mu_1 > \mu_2 \qquad H_1: \mu_1 - \mu_2 > 0$

Test Statistic: $$Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

~~Old Chips~~ ~~New Chips~~

$\cancel{\sigma_2 = 481.2}$ $\cancel{\sigma_1 = 495.6}$

$\cancel{n_2 = 50}$ $\cancel{n_1 = 50}$

$\cancel{\bar{y}_2 = 20.3}$ $\cancel{\bar{y}_1 = 22.4}$

Old Chips μ_2 New Chips μ_1

$y_2 = 481.2$ $y_1 = 495.6$

$n_2 = 50$ $n_1 = 50$

$\sigma_2 = 20.3$ $\sigma_1 = 22.4$

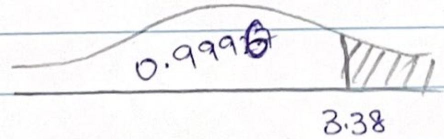
$$Z = \frac{495.6 - 481.2}{\sqrt{\frac{22.4^2}{50} + \frac{20.3^2}{50}}}$$

$$= \frac{14.4}{\sqrt{\frac{501.76 + 412.09}{50}}}$$

$$= \frac{14.4}{\sqrt{18.27}} = 3.3723$$

4.27

Since this is a One tail test



$P_{value} = 1 - 0.9996$

$P = 0.0004 < \alpha = 0.05$

We Reject H_0.

$\therefore \mu_2 \leq \mu_1$

Hence, we Conclude that the mean speed for the new Chips is greater than that of old chips

b. A Sample of 60 even older chips had an average speed of 391.2 MHZ with a Standard deviation of 17.2 MHz. Someone Claims that the new Chips average more than 100 MHZ faster than these very old ones? Do the Data provide Convincing evidence for this claim? State appropriate null & alternative hypotheses and find P Value.

Test: $H_0 : \mu_1 - \mu_2 \leq 100$

$H_1 : \mu_1 - \mu_2 > 100$

Test Static : $Z = \dfrac{\bar{y}_1 - \bar{y}_2 - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

New Chips μ_1 | **60 Sample Old Chips μ_2**

$y_1 = 495.6$ | $y_2 = 391.2$

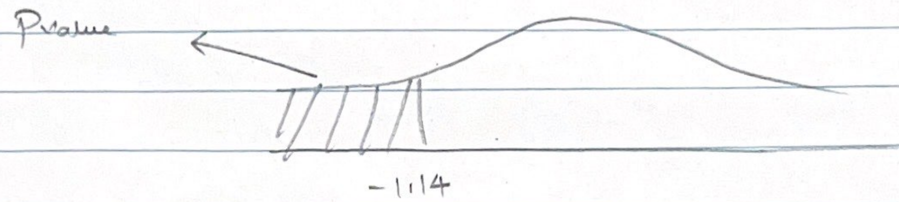
$\sigma_1 = 22.4$ | $\sigma_2 = 17.2$

$n_1 = 50$ | $n_2 = 60$

$$Z = \frac{495.6 - 391.2 - 100}{\sqrt{\frac{22.4^2}{50} + \frac{17.2^2}{60}}} = \frac{4.4}{\sqrt{10.03+4.93}}$$

$Z = 1.14$

Since this is a One tail test, lower tail test



Pvalue $= 0.1271 > \alpha = 0.05$

∴ We don't Reject Ho

We Can Conclude that the Data do not provide Convincing evidence for this Claim

Computer Calculations

3. Table 5.13 (page 237 in the textbook) shows the observed pollution indexes of air samples in two areas of a city. Using a Statistical software, test the hypothesis that the mean pollution indexes are the same for the two areas using $\alpha = 0.05$. Be sure to include the edited computer output and to interpret the results.

Area A	2.92	1.88	5.35	3.81	4.69	4.86	5.81	5.55
Area B	1.84	0.95	4.26	3.18	3.44	3.69	4.95	4.47

Solution: The standard deviations of both the areas A and B are not given, but as of now let us think they are equal. So, we will be using the non-pooled t-test (Equal variances are assumed).

$H_0 \rightarrow \mu_A - \mu_B = 0$

$H_1 \rightarrow \mu_A - \mu_B != 0$

Test Statistics:

Area	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
A		8	4.3588	1.3828	0.4889	1.8800	5.8100
B		8	3.3475	1.3541	0.4787	0.9500	4.9500
Diff (1-2)	Pooled		1.0113	1.3685	0.6843		
Diff (1-2)	Satterthwaite		1.0113		0.6843		

Area	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
A		4.3588	3.2027	5.5148	1.3828	0.9142	2.8143
B		3.3475	2.2154	4.4796	1.3541	0.8953	2.7560
Diff (1-2)	Pooled	1.0113	-0.4563	2.4788	1.3685	1.0019	2.1583
Diff (1-2)	Satterthwaite	1.0113	-0.4564	2.4789			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	14	1.48	0.1616
Satterthwaite	Unequal	13.994	1.48	0.1616

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	7	7	1.04	0.9574

From the output, tValue=1.478 with degree of freedom = 13.994. So, the obtained p-value = 2-tailed = 0.162

Since the obtained p-value = 0.162>0.05= α we fail to reject the null hypothesis

So, there is no clear evidence that the mean index population is the same for the two areas

Code:

```
TITLE 'QUESTION3';
DATA AirSamples;
INPUT Area $ PollutionIndexes @@;
DATALINES;
A 2.92 A 1.88 A 5.35 A 3.81 A 4.69 A 4.86 A 5.81 A 5.55
B 1.84 B 0.95 B 4.26 B 3.18 B 3.44 B 3.69 B 4.95 B 4.47
;

ods graphics off;
ods select Plots SSPlots;
proc univariate data=AirSamples plot;
class Area;
var PollutionIndexes;
run;
```

Area A Output:

QUESTION3

The UNIVARIATE Procedure
Variable: PollutionIndexes
Area = A

```
Stem Leaf                      #             Boxplot
   5 68                        2                |
   5 4                         1             +-----+
   4 79                        2             *-----*
   4                                         |  +  |
   3 8                         1             |     |
   3                                         +-----+
   2 9                         1                |
   2                                            |
   1 9                         1                |
     ----+----+----+----+


                     Normal Probability Plot
5.75+                                         *++  *
    |                                     *+++
    |                              *   *++
    |                              ++++
3.75+                         +*++
    |                      +++
    |                  ++++*
    |              ++++
1.75+          +++    *
     +----+----+----+----+----+----+----+----+----+----+
         -2        -1         0        +1        +2
```

Area B Output:

QUESTION3

The UNIVARIATE Procedure
Variable: PollutionIndexes
Area = B

```
Stem Leaf                      #             Boxplot
   5 0                         1                |
   4 5                         1                |
   4 3                         1             +-----+
   3 7                         1             *-----*
   3 24                        2             |  +  |
   2                                         +-----+
   2                                            |
   1 8                         1                |
   1 0                         1                |
   0
     ----+----+----+----+


                     Normal Probability Plot
5.25+                                            +++
    |                                        ++++ *
    |                                    *+++*
3.75+                                 *++
    |                          *  *+++
    |                          ++++
2.25+                       +++
    |                   ++++*
    |               ++++
0.75+          ++++    *
     +----+----+----+----+----+----+----+----+----+----+
         -2        -1         0        +1        +2
```

Hence From the side by side box plots obtained for both the areas A and B, we can see that the data is relatively symmetric and also normally distributed.

4. In a standard dissolution test for tablets of a particular drug product, the manufacturer must obtain the dissolution rate for a batch of tablets prior to release of the batch. Suppose that the dissolution test consists of assays for 24 randomly selected individual 25 mg tablets. For each test, the tablet is suspended in an acid bath and then assayed after 30 minutes. The results of the 24 assays are given here.

19.5	19.7	19.7	20.4	19.2	19.5	19.6	20.8	19.9	19.2	20.1	19.8
19.8	19.6	19.5	19.3	19.7	19.5	20.6	20.4	19.9	20.0	19.8	20.4

a. Using a graphical display, determine whether the data appear to be a random sample from a normal distribution.

<u>Code:</u>

```
TITLE 'Question4';
DATA Drug;
INPUT dataassays @@;
datalines;
19.5 19.7 19.7 20.4 19.2 19.5 19.6 20.8 19.9 19.2 20.1 19.8
19.8 19.6 19.5 19.3 19.7 19.5 20.6 20.4 19.9 20.0 19.8 20.4
;
proc ttest data=DRUG h0=25;
var dataassays;
run;
```

<u>Output:</u>

Question4

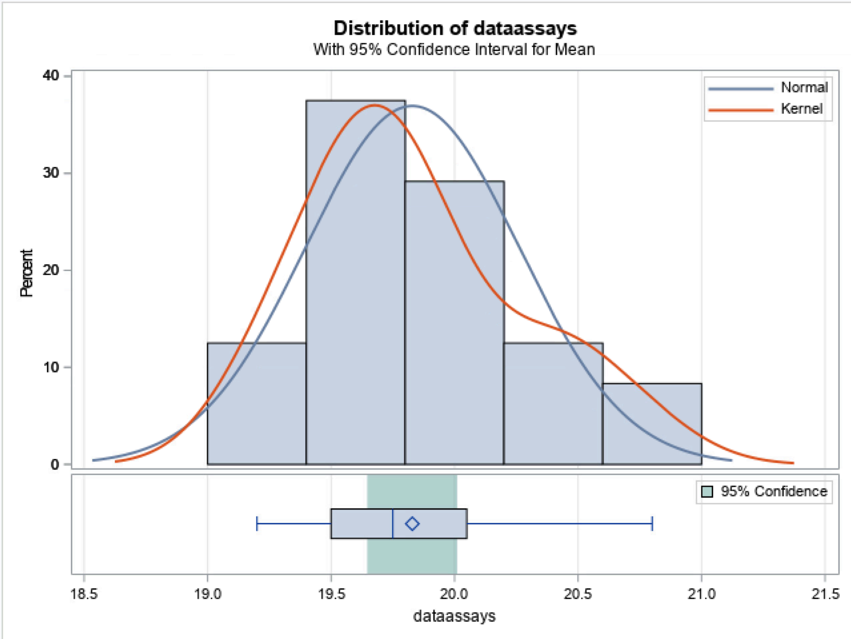
The TTEST Procedure

Variable: dataassays

N	Mean	Std Dev	Std Err	Minimum	Maximum
24	19.8292	0.4319	0.0882	19.2000	20.8000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
19.8292	19.6468	20.0115	0.4319	0.3356	0.6058

DF	t Value	Pr > |t|
23	-58.66	<.0001



The above graph obtained using SAS, it is clear that the data is normally distributed and the data crosses through all the points with a default confidence interval of 95.

b. Estimate the mean and standard deviation of dissolution rate for the batch of tablets and a 99% confidence interval for the mean.

<u>Code:</u>

```
TITLE 'Question4';
DATA Drug;
INPUT dataassays @@;
datalines;
19.5 19.7 19.7 20.4 19.2 19.5 19.6 20.8 19.9 19.2 20.1 19.8
19.8 19.6 19.5 19.3 19.7 19.5 20.6 20.4 19.9 20.0 19.8 20.4
;
proc ttest data=DRUG alpha = 0.01;
var dataassays;
run;
```

Output:

Question4

The TTEST Procedure

Variable: dataassays

N	Mean	Std Dev	Std Err	Minimum	Maximum
24	19.8292	0.4319	0.0882	19.2000	20.8000

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
19.8292	19.5817	20.0766	0.4319	0.3116	0.6806

<table>
<tr><th>DF</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>23</td><td>224.94</td><td><.0001</td></tr>
</table>

Obtained using SAS, the mean dissolution rate for the batch of tablets is 19.8292 and the confidence interval for 99% is (19.5817,20.0766). With a Standard Deviation of 0.4319

c. Is there significant evidence that the batch of pills has a mean dissolution rate less than 20 mg? Explain. Use α =0.01.

From 4.b Solution(alpha=0.01), We can say that the value obtained is less than 20mg.