PITCH DROP EXPERIMENT

In Question 1, we have been given a dataset from one of the longest running experiments from the University of Queensland, beginning in 1927. The experiment studies the viscosity of pitch, recording the observations of the flow of pitch through a funnel. The data in the dataset in the number of years for each drop to occur.

We have been tasked with investigating whether the newly added air conditioning has had a significant impact on the experiment. In order to investigate the impact, we will have to employ a technique called hypothesis testing.

a) In order to do a hypothesis test, we first have to create a null and alterative hypothesis. We are told in the question that after the A/C unit was installed the average number of years for the pitch to drop was 12.85.

We can then say that the null hypothesis is that the air conditioning unit has had no effect on the amount of time taken for the pitch to drop. Subsequently, the alternative hypothesis is that the air conditioning has had an effect on the amount of time taken for the pitch to drop. In order to represent this in a more precise way we can employ mathematical notion:

$$H_0 : \mu = 12.85$$

$$H_1 : \mu \neq 12.85$$

b) Looking at the distribution of the sample, we can see that it follows somewhat of a normal distribution, having just one mode in the centre and tails either side.

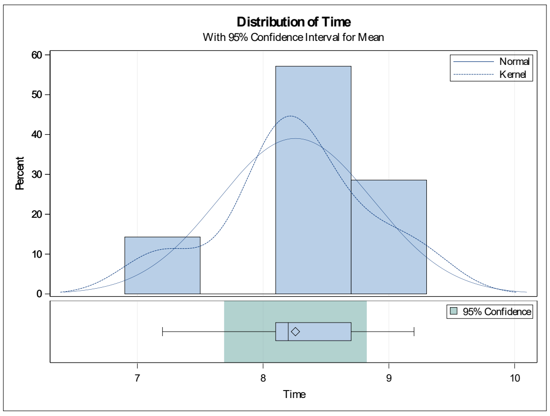


Figure 1 - Distribution of Time Histogram and Box Plot

Looking at the box plot distribution, we can see the distribution is skewed to the right. To gain a better understanding of how the sample compares to a normal distribution we can use the QQ-plot:

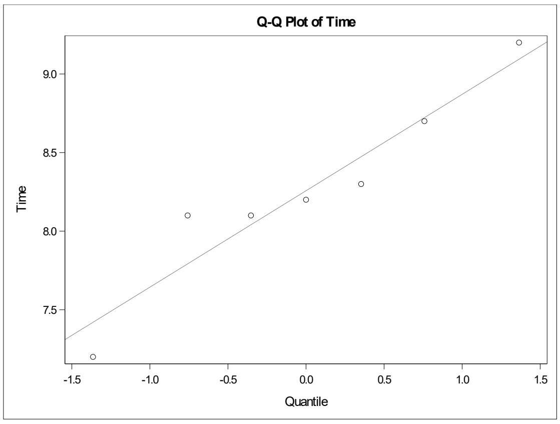


Figure 2 - QQ-plot of the Dataset

We can see here, that the points do follow the general shape of the normal line, however, no points intersect the line. Looking at the first and last plots on the graph, we can see that they are far away from the normal line suggesting fat tails.

c) In order to report the p-value, we first have to compute the test statistic, which is computed using the following formula:

$$t = \frac{\bar{x} - \mu_0}{S_x / \sqrt{n}}$$

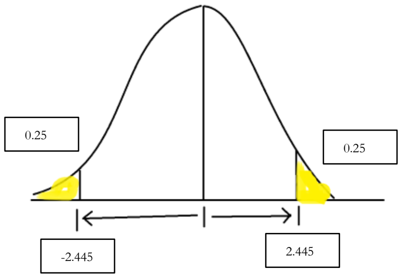
Where,

- $\bar{x}$ is the sample mean,
- μ_0 is the population mean for the null hypothesis,
- S_x is the standard deviation of the sample,
- $\sqrt{n}$ is the square root of the number of entries in the sample.

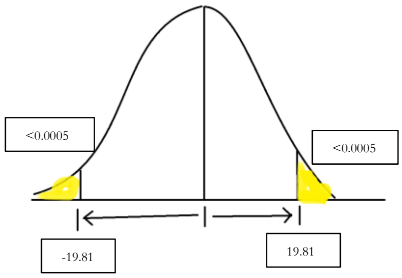
If we put the problem specific numbers into the formula we get:

$$t = \frac{8.257 - 12.85}{0.613 / \sqrt{7}} = -19.812$$

Now that we have the t value, we can use the student t's distribution to find the critical values. The first step of doing this is to compute the degrees of freedom, which is $n - 1$, making DoF = 6. As we are looking for a 95% Confidence Level, the critical value of the t distribution for a DoF of 6 is 2.447. Graphically this would look like:



However, with the computed test statistic of -19.812 ≠ -2.445, the graph would look something like:



We can say that the p-value is less than 0.001 as for a DoF of 6, the critical value for $\alpha = 0.001$ is 5.9588, and 5.9588 < 19.812.

Using SAS, we can calculate the test statistic (t-value) and the p – value using the t-test procedure. The results are as follows:

Table 1 - Stat Summary and 95% CL's

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
8.2571	7.6899	8.8244	0.6133	0.3952	1.3506

Table 2 – Hypothesis Testing Results

DF	t Value	Pr > |t|
6	-19.81	<.0001

Looking at the tables, we can confirm that the Degree of Freedom is 6, the test statistic or t-value is -19.81 and the p-value is <.0001. SAS was able to compute a more accurate p-value as the t-distribution table in the mathematical formulae booklet only went up to significance level of 0.001 for a two tailed test.

d) Since the p-value < 0.0001, which is less than 0.05, we have to reject H_0 at the 5% significance level. We have sufficient evidence to reject the assertion that the mean time spent for the pitch to drop is 12.85. Rejecting H_0, means that we have to accept H_1 which is that the mean time spent for the pitch to drop is not equal to 12.85.

e) Looking at the hypothesis testing, we have to conclude that the air conditioning has had quite an impact on the time spent for the pitch to drop. We were able to reject H_0, the null hypothesis of the assertion that the mean time spent was equal to 12.85, therefore accepting H_1, which was that the mean time spent wasn't equal to 12.85.

If we delve into the 95% CL for the sample data we can see it ranges from 7.6899 – 8.8244, which is no where near 12.85, strengthening our claim that the air conditioning has had an impact on the results.

CODE

```
/*Create Data Set for Analysis*/
DATA PitchDrop;
      /*Create 2 Variables - Stress (Y/N) & Percentage*/
      INPUT Time;
      /*Put Values in Variables*/
      DATALINES;
      8.1
      8.2
      7.2
      8.1
      8.3
      8.7
      9.2
      ;
RUN;
/*Run a T-TEST procedure using
      an alpha of 0.05,
      a null hyp of 12.85
      and making it a 2 tailed test
*/
PROC TTEST DATA=pitchdrop alpha=0.05 H0=12.85 SIDES=2;
RUN;
```

LOADED DIE AT CASINO

In question 2, we have been given the results of 190 rolls of a die in a Casino. They suspect that the die is loaded, as the records say that "six" was rolled 56 out of the 190. Using the appropriate hypothesis test, we have to determine whether the die is loaded or not. Due to the nature of the data, with records of being a six or not, we will use the binomial hypothesis test.

a) In order to do a hypothesis test, we first have to create a null and alterative hypothesis. From the data we have been given, Let X be the R.V. 'Number of occurrences of Six' where X~B(190,p). Using this we can state that our Null Hypothesis would be p = 0.1667, which means that six occurs 1/6th of the time. Our alternative hypothesis would be p > 0.1667, meaning that a six occurs more than 1/6th of the time. Mathematically, this would look like:

$$H_0 : p = 0.1667$$

$$H_1 : p > 0.1667$$

Finally, we would like to do the hypothesis test with a significance level of 5%, α = 0.05.

b) Using SAS, we can compute the results of the hypothesis test by using the FREQ procedure. Within the procedure, we state that the data is Binomial using the key word, plus stating the null hypothesis, using the p key word. The results of the test are as follow:

Binomial Proportion	
Six = YES	
Proportion	0.2947
ASE	0.0331
95% Lower Conf Limit	0.2299
95% Upper Conf Limit	0.3596
Exact Conf Limits	
95% Lower Conf Limit	0.2309
95% Upper Conf Limit	0.3651

Test of H0: Proportion = 0.1667	
ASE under H0	0.0270
Z	4.7353
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	<.0001

Sample Size = 190

c) Since we have a one tailed alternative hypothesis, the p-value is <0.0001. Our significance level for the test was α = 0.05, and 0.0001 < 0.05 therefore giving us enough evidence to reject the null hypothesis at a significance level of 5%.

d) From our testing, we can conclude that the die at the casino is, in fact, loaded. Since we rejected the null hypothesis, we have to accept the alternative hypothesis that states that the occurrence of six on the die is greater than 1/6. Furthermore, using the sample size and probability of a six occurring on a fair die, we can state that 190 * 1/6 = 31.673 is the expected value of occurrences of six, which is a lot less than the observed 56.

CODE

```
/*Create Data Set for Analysis*/
DATA CasinoDie;
     /*Create 2 Variables - SIX(Y/N) & # of OCC*/
     INPUT Six$ Occ;
     /*Put Values in Variables*/
     DATALINES;
     YES 56
     NO 134
     ;
RUN;
/*Call the FREQ Procedure */
/*Using ORDER = DATA to get the YES value*/
PROC FREQ DATA=CasinoDie ORDER=DATA;
     /*Specifying the data is Binomial*/
     WEIGHT Occ;
     TABLES Six / BINOMIAL(p=0.1667);
RUN;
```