**PART 1**

**Background of the study**

On January 8, 2003, a 19-seater Beech 1900 turboprop aircraft crashed in Charlotte, North Carolina killing all 21 passengers including the pilot and co-pilot. In response to that, the FAA in January 2013 ordered that passengers be weighed or surveyed about their weight prior to boarding 10 to 19-seat passenger planes. The airlines have been weighing adult passengers and carry-on bags randomly over one-month to try to determine the mean weight per passenger (including luggage). Sample consists weight of 426 people and their luggage.

**Exploratory Data Analysis**

The data has been analyzed to identify patterns with the help of Minitab. Various statistical tools like Mean, Tally, different types of Charts and Graphs and Correlation have been utilized in this analysis. Sample consists of 81% male passengers with an average age of 51.37 and average weight 205.92. On the other hand females constitute 19% of the sample size with average age 45.57 and average weight of 133.41. Because of this uneven gender-wise distribution, sample mean is 192.13 which is much more closer to average weight of male passengers.

**Statistics**

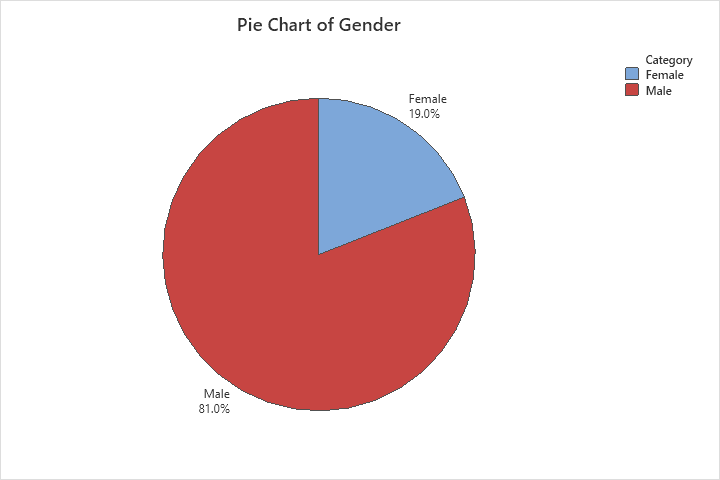
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **N\*** | **Mean** | **SE Mean** | **StDev** | **Minimum** | **Q1** | **Median** | **Q3** | **Maximum** | **Skewness** |
| Weight | 426 | 0 | 192.13 | 1.88 | 38.87 | 90.00 | 167.00 | 197.00 | 221.00 | 282.00 | -0.43 |
| Age | 426 | 0 | 50.268 | 0.417 | 8.605 | 24.000 | 45.000 | 50.000 | 56.000 | 73.000 | -0.27 |

**Statistics**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Gender** | **N** | **N\*** | **Mean** | | **SE Mean** | **StDev** | **Minimum** | **Q1** | **Median** | **Q3** | **Maximum** |
| Weight | Female | 81 | 0 | 133.41 | | 2.21 | 19.93 | 90.00 | 118.50 | 134.00 | 148.00 | 182.00 |
|  | Male | 345 | 0 | 205.92 | | 1.50 | 27.78 | 130.00 | 185.50 | 207.00 | 226.00 | 282.00 |
|  |  |  |  |  | |  |  |  |  |  |  |  |
| Age | Female | 81 | 0 | 45.57 | | 1.08 | 9.69 | 24.00 | 37.00 | 47.00 | 52.00 | 73.00 |
|  | Male | 345 | 0 | 51.371 | | 0.428 | 7.951 | 27.000 | 47.000 | 52.000 | 56.000 | 73.000 |
| **Variable** | **Gender** | **Skewness** | | |
| Weight | Female | 0.07 | | |
|  | Male | -0.08 | | |
|  |  |  | | |
| Age | Female | 0.19 | | |
|  | Male | -0.26 | | |

**Tally**

|  |  |  |
| --- | --- | --- |
| **Gender** | **Count** | **Percent** |
| Female | 81 | 19.01 |
| Male | 345 | 80.99 |
| N= | 426 |  |



As part of Exploratory Data Analysis, we have also tried to find Correlation between age and weight. The analysis using Matrix Plot in Minitab is shown below.

Chart, scatter chart

Description automatically generated

**Pairwise Pearson Correlations**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample 1** | **Sample 2** | **N** | **Correlation** | **99% CI for ρ** | **P-Value** |
| Age | Weight | 426 | 0.160 | (0.036, 0.279) | 0.001 |

We found that Correlation between Age and Weight is 0.160. Since this is close to 0, there is no correlation between the variables. Having said this we might say that variables age and weight have no association with each other.

**Interval estimation of population mean**

Looking at the sample size of 426, we already have sample mean as 192.13 and sample standard deviation as 38.87. Since population standard deviation is not available, we will go for One-Sample t-distribution. Using a 99% confidence level, we have constructed a confidence interval estimate as below.

**Descriptive Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **StDev** | **SE Mean** | **99% CI for μ** |
| 426 | 192.13 | 38.87 | 1.88 | (187.256, 197.002) |

*μ: population mean of Weight*

With 99% confidence level and using t-distribution, we can say with 99% surety that mean weight for all passengers lie between 187.26 and 197.00. This means that possibility of mean weight of all passengers to lie outside the mean interval of 187.26 and 197.00 is only 1%.

**Confidence level being challenged**

It can be argued that 99% confidence level is too high and that population mean could be lesser than the estimated lower bound which was 187.26. One of the reasons for this kind of argument may be that confidence interval is always estimated based on the **assumption** that the population distribution is almost normal. So, population mean estimation could be excessive in case distribution is not normal. Secondly, as has been observed in the airline industry, commonly accepted confidence level is 95%. Any confidence level used more than that would almost always be challenged.

Having said that, we have gone ahead and tried to test this using Hypothesis testing. We have considered the following:

***Null Hypothesis*** : Mean weight of all passengers is more than or equal to 187.26 (lower bound)  
***Alternate Hypothesis*** : Mean weight of all passengers is lower than 187.26 (lower bound)

**Descriptive Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **StDev** | **SE Mean** | **99% Upper Bound for μ** |
| 426 | 192.13 | 38.87 | 1.88 | 196.53 |

*μ: population mean of Weight*

**Test**

|  |  |  |  |
| --- | --- | --- | --- |
| Null hypothesis | | | H₀: μ = 187.26 |
| Alternative hypothesis | | | H₁: μ < 187.26 |
| **T-Value** | **P-Value** |
| 2.59 | 0.995 |

Chart, line chart

Description automatically generated

As per the t-distribution with degree of freedom as 425 (n-1), critical value is coming out to be -2.335 and p-value is 0.995.

Since p-value here is more than the critical value, we failed to reject the Null Hypothesis. Hence, as per the hypothesis testing, FAA official’s claim that mean weight of all passengers is below our lower bound of 187.26 as per 99% confidence level is not correct.

**PART 2**

**Background of the study**

The study is being performed to determine the price that we want to charge for a regular-sized soda for restaurant we have newly opened in the city. We are inclined towards charging the same price which our competitors are currently charging which we believe to be $2.58. For the said purpose, we have collected the soda prices from 35 randomly sampled restaurants.

**Exploratory Data Analysis**

**Statistics**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **N\*** | **Mean** | | **SE Mean** | **StDev** | **Minimum** | **Q1** | **Median** | **Q3** | **Maximum** |
| Soda Price | 35 | 0 | 2.6949 | | 0.0742 | 0.4392 | 1.6800 | 2.3600 | 2.6800 | 3.0700 | 3.4500 |
| **Variable** | **Skewness** | | |
| Soda Price | -0.08 | | |

As per the data sample, the sample mean of the soda price is $2.69 and sample standard deviation as 0.4392.

*Q.1 You are assuming the mean price is $2.58 for all your competitors. When conducting data analysis to test this belief, what is the assumption called?*

This can be tested through Hypothesis testing and the assumption wherein we are considering the mean price as $2.58 is Null Hypothesis since Null Hypothesis uses the sign of equality.

*Q.2 Calculate the t-statistic assuming the mean soda price for all of your competitors is $2.58*

**Descriptive Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **StDev** | **SE Mean** | **95% CI for μ** |
| 35 | 2.6949 | 0.4392 | 0.0742 | (2.5440, 2.8457) |

*μ: population mean of Soda Price*

**Test**

|  |  |  |  |
| --- | --- | --- | --- |
| Null hypothesis | | | H₀: μ = 2.58 |
| Alternative hypothesis | | | H₁: μ ≠ 2.58 |
| **T-Value** | **P-Value** |
| 1.55 | 0.131 |

*Chart, box and whisker chart

Description automatically generated*

*Q.3 Calculate the p-value for your test statistic.*

As per above analysis, p-value is 0.131

*Q.4 Once level of 90%, test whether the mean soda price of all your competitors is $2.58 using the t-statistics.*

**Descriptive Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **StDev** | **SE Mean** | **90% CI for μ** |
| 35 | 2.6949 | 0.4392 | 0.0742 | (2.5693, 2.8204) |

*μ: population mean of Soda Price*

**Test**

|  |  |  |  |
| --- | --- | --- | --- |
| Null hypothesis | | | H₀: μ = 2.58 |
| Alternative hypothesis | | | H₁: μ ≠ 2.58 |
| **T-Value** | **P-Value** |
| 1.55 | 0.131 |

*Chart, line chart

Description automatically generated*

Considering Null Hypothesis that mean soda price for all our competitors is $2.58 and using t-statistic analysis, we plotted Probability Distribution with probability 0.1 (which is 1 – 0.90 confidence level). We found out critical value to be 1.691 which is more than our t-value. Hence, we fail to reject the null hypothesis. This implies mean soda price for all competitors is $2.58 as per 90% confidence level.

*Q.5 Using a confidence level of 90%, test whether the mean soda price of all your competitors is $2.58 using the p-value.*

As per above analysis, p-value is 0.131 which is more than the level of significance i.e. 0.1. Therefore we failed to reject the Null Hypothesis, which implies mean soda price for all competitors is $2.58 with 90% confidence level.

*Q.6 Detail the reasoning behind your decision.*

We have conducted hypothesis testing using t-statistic distribution test and probability distribution plots with confidence levels 95% as well as 90%. As per our findings, mean soda price for all competitors which was assumed to be $2.58 lies within both confidence estimates – Confidence interval $2.54-$2.86 (as per 95% confidence level); and $2.57-$2.82 (as per 90% confidence level).

Our finding has been consistent using both confidence levels.

*Q.7 Is there reason to believe your confidence levels are inaccurate? If so, what assumption(s) may be inaccurate?*

Since standard confidence level that is used across is 95%, and we have used both 95% as well as 90% confidence levels and found the results consistent, we don’t believe that the confidence levels are inaccurate.

**Case study performed by :**

1. Amandeep Singh
2. Abhishek Ponnaganti
3. Maheshwari Banoth
4. Renu Mutha
5. Siva Koti Reddy Kamatham
6. Dhruva Hejamadi Tati