MAT 243 Final Project Part (1): Summary Report

Milestone *(Number I)*

Scenario (B): Solar Power

Jo De Leon

jo.deleon@snhu.edu

Southern New Hampshire University

**1. Statement of the Problem**

*The power generated in solar farms that are installed in the cities must be analyzed. By doing so, cities with maximum power generation can be identified which will refine the marketing play in order to increase sales of solar panels.*

*The available data for analysis includes five cities where the company has active solar forms. These data include the average monthly solar power generation (in kWh) per panel for 204 months beginning from the year 2000.*

*In order to analyze these data, calculating the descriptive statistics will be needed: mean, median, variance, standard deviation, minimum, maximum, 25th percentile, and the 75th percentile.*

*These statistics will be used to perform hypothesis tests. The tests performed were for population proportion, which a 99% confidence interval for the proportion of days with solar power generation above 43 kWh for City A was constructed, and for population mean, which a 95% confidence interval for average solar power generation for City B was constructed.*

*Creating and executing the Python functions in Codio was almost seamless. The check it button was a great tool as a check and balance of the results.*

**2. Descriptive Statistics**

Table 1. Sample Descriptive Statistics for City A

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Mean | 43.1177  Considering the median, this mean of approximately 43.1177 depicts the accurate solar generation for City A. |
| Median | 43.105  Considering the mean, this median also gives a significant insight of range of the solar generation in City A. Since it is approximately “close” to the mean, the average and the median presents to be a reliable statistic for this particular study and location. |
| Mode | 44.052  This mode also sheds light regarding the relationship between the other descriptive statistics and shows stability and reliability in the data |
| Minimum | 39.418  This minimum gives us the lowest solar power generation in terms of kWh. |
| Maximum | 47.644  This maximum gives is the highest solar power generation in terms of kWh for City A. |
| Variance | 2.2025  The variance shows us that the data gathered have a tendency to vary away from the mean. |
| Standard Deviation | 1.4841  This standard deviation shows how spread out the data are in reference from the mean. This helps indicate if a data point is considered normal or is an outlier. |
| 25th Percentile | 42.0000  25% of the solar power generation occur below 42.0000. |
| 75th Percentile | 44.1743  75% of the solar power generation occur below 44.1743. |

Table 2. Sample Descriptive Statistics for City B

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Mean | 55.4802  Considering the median, this mean of approximately 55.4802 depicts the accurate solar generation for City B. |
| Median | 55.5005  Considering the mean, this median also gives a significant insight of range of the solar generation in City B. Since it is approximately “close” to the mean, the average and the median presents to be a reliable statistic for this particular study and location. |
| Mode | 60.470  This mode also sheds light regarding the relationship between the other descriptive statistics. This particular mode is insightful because it shows that City B has a higher solar power generation compared to City A. |
| Minimum | 46.364  This minimum gives us the lowest solar power generation in terms of kWh and is significantly much higher than City A. |
| Maximum | 66.072  This maximum gives is the highest solar power generation in terms of kWh for City A and is significantly much higher than City A. |
| Variance | 12.4963  The variance shows us that the data gathered have a tendency to vary away from the mean in large intervals. |
| Standard Deviation | 3.5350  This standard deviation shows how spread out the data are in reference from the mean. This helps indicate if a data point is considered normal or is an outlier. |
| 25th Percentile | 53.2415  25% of the solar power generation occur below 53.0000 – which is a lot higher than City A. |
| 75th Percentile | 57.6575  75% of the solar power generation occur below 44.1743 – which is a lot higher than City A. |

*The mean is calculated by using python function mean(x) where x is in the input which results into the average of the data.*

*The median is calculated by using python function median(x) where x is in the input which results into the median of the data.*

*The mode is calculated by using python function mode(x) where x is in the input which results into the mode of the data.*

*The minimum is calculated by using python function min(x) where x is in the input which results into the minimum of the data.*

*The maximum is calculated by using python function max(x) where x is in the input which results into the maximum of the data.*

*The median is calculated by using python function median(x) where x is in the input which results into the median of the data.*

*The variance is calculated by using python function var(x) where x is in the input which results into the variance of the data.*

*The standard deviation is calculated by using python function std(x) where x is in the input which results into the standard deviation of the data.*

*The 25th and 75th percentile is calculated by using python function describe(x) where x is in the input which results into the 25th and 75th percentile of the data.*

**3. Confidence Intervals**

*Identify the variable(s) and Python functions used in the scripts to estimate Confidence Intervals. Summarize all Confidence Intervals for each variable in a formatted table as shown below.*

Table 3. Sample Confidence Interval for City A

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| (39.418, 47.644) | (0.4394, 0.6194)  This confidence interval shows that there is a 99% confidence that the solar power generation population proportion falls within this interval. |

The confidence interval python function for the population proportion (st.norm.interval()) is executed with the parameters of the confidence, probability, and stderror.

Table 4. Sample Confidence Interval for City B

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| (46.3640,66.0720) | (54.9922, 55.9682)  This confidence interval shows that there is a 95% confidence that the solar power generation population mean falls within this interval. |

The confidence interval python function for the population mean (st.t.interval()) is executed with the parameters of the confidence, degrees of freedom, mean, and stderror.

**4. Hypothesis Tests (1-sample and 2-sample)**

*Identify the variable(s) and Python functions used in the scripts to perform Hypothesis Testing. Summarize statistics for each variable in a formatted table as shown below.*

Table 5. Sample Hypothesis Test for City A

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistics | -2.0199 |
| P-Value | 0.0217 |

* *Variable Name: City A*
* *Null hypothesis: The proportion is greater than 60%? City A > 43kWh*
* *Alternative hypothesis: The proportion is smaller than 60% City A < 43 kWh*
* *Level of significance: 0.05*
* *Test Statistic: -2.0199*
* *With a p-value of 0.02170, since it is less than the level of significance, it suggests that there is insufficient evidence that the proportion is less than 60%.*

*The hypothesis test for population proportion python function (prop\_1samp\_ztest()) is executed with parameters of the x which is the proportion greater than 43kWh, n which is the sample, the null value which is 60%, and the alternative which is ‘smaller’.*

Table 6. Sample Hypothesis Test for City B

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistics | -2.1000 |
| P-Value | 0.0370 |

* *Variable Name: City B*
* *Null hypothesis: The average solar power generation in City B is 56 kWh. City B = 56 kWh*
* *Alternative hypothesis: The average solar power generation in City B is not 56 kWh.*
* *Level of significance: 0.01*
* *Test Statistic: -2.1000*
* *With a p-value of 0.0370, since it is greater than the level of significance, it suggests that there is sufficient evidence that the average solar power generation in City B is 56 kWh.*

*The hypothesis test for population mean python function (means\_1samp\_ttest()) is executed with parameters of the mean, std\_dev, n, null\_value, and alternative.*

**5. Results**

*Based on the results, there is not sufficient evidence that the proportion for City A is less than 60% that are above 43 kWh. However, for City B, there is sufficient evidence that City B averages approximately 56 kWh for solar power generation. With that being said, the hypothesis test and the mean and median for City B, so far, indicates that it generates more of the solar power generation than compared to City A. So far, it seems as if the solar power generation for City A is comparatively lower than that of City B – this can be analyzed by comparing the minimum and maximum, with the mean, from City A and City B.*

MAT 243 Final Project Part (2): Summary Report

Milestone *(Number II)*

Scenario (B): Solar Power

Jo De Leon

jo.deleon@snhu.edu

Southern New Hampshire University

**1. Statement of the Problem**

*The power generated in solar farms that are installed in the cities must be analyzed. By doing so, cities with maximum power generation can be identified which will refine the marketing play in order to increase sales of solar panels.*

*The available data for analysis includes five cities where the company has active solar forms. These data include the average monthly solar power generation (in kWh) per panel for 204 months beginning from the year 2000.*

*In order to analyze these data, calculating the descriptive statistics will be needed: mean, median, variance, standard deviation, minimum, maximum, 25th percentile, and the 75th percentile.*

*These statistics will be used to perform hypothesis tests. The tests performed were for population proportion, which a 99% confidence interval for the proportion of days with solar power generation above 43 kWh for City A was constructed, and for population mean, which a 95% confidence interval for average solar power generation for City B was constructed.*

*Creating and executing the Python functions in Codio was almost seamless. The check it button was a great tool as a check and balance of the results.*

**2. Hypothesis Tests (difference of two population proportions)**

*Identify the variable(s) and Python functions used in the scripts to perform Hypothesis Testing. Summarize statistics for each variable in a formatted table as shown below.*

Table 7. Sample Hypothesis Test for City A – Population Proportions

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistics | -3.2344 |
| P-Value | 0.0012 |

* *Variable Name: City A*
* *Null hypothesis: The proportion with power generation over 44.4 kWh is the same for the month of June and July. June = July*
* *Alternative hypothesis: The proportion with power generation over 44.4 kWh is not the same for the month of June and July. June =/= July*
* *Test Statistic:* -3.2344
* *With a p-value of* 0.0012, there is insufficient evidence to reject that the power generation is the same for the month of June and July.

Table 8. Sample Hypothesis Test for City B

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistics | -3.7639 |
| P-Value | 0.0002 |

* *Variable Name: City B*
* *Null hypothesis: The proportion with power generation over 55.2 kWh is the same for the month of June and July. June = July*
* *Alternative hypothesis: The proportion with power generation over 55.2 kWh is not the same for the month of June and July. June =/= July*
* *Level of significance: 0.01*
* *Test Statistic: -3.7639*
* *With a p-value of* 0.0002, there is insufficient evidence to reject that the power generation is the same for the month of June and July.

*The hypothesis test for the difference of two population proportions python function (proportions\_ztest()) is executed with the parameters counts and n (between the two months in June and July in this case) – this is used for Table 7 and Table 8.*

Table 9. Sample Hypothesis Test for City A

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistics | -165.7428 |
| P-Value | 4.2679e-37 |

* *Variable Name: City A*
* *Null hypothesis: The average power generation is not the same for July and August. July =/= August*
* *Alternative hypothesis: The average power generation is the same for July and August. July = August*
* *Test Statistic:* 165.7428
* *With a p-value of* 4.2679e-37*,* there is insufficient evidence to reject that the average power generation is not the same for the month of June and July.

Table 10. Sample Hypothesis Test for City B

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistics | -224.6878 |
| P-Value | 1.4082e-52 |

* *Variable Name: City B*
* *Null hypothesis: The average power generation is less in February than August. February < August*
* *Alternative hypothesis: The average power generation in February is greater than or equal to August. February >= August*
* *Level of significance: 0.01*
* *Test Statistic: -*224.6878
* *With a p-value of* 1.4082e-52, there is insufficient evidence to reject that the average power generation in February is less than August.

*The hypothesis test for the difference of two population means python function (st.ttest\_ind()) is executed with the parameters of data from particular months – this is used for Table 9 and Table 10. The months compared for City A is between June and July and for City B is between February and August.*

**3. Results**

*Based on the results, there is insufficient evidence to claim that the power generation is over 44.4 kWh for both months of June and July. There is also insufficient evidence to claim that the average power generation is February is less than the average power generation in August.*