# **Habermanv2 Dataset**

**Hypothesis 1:** Patient who can survive 5-years after operations depends on patient’s age at time of operation

Regression Summary:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.067950324 |  |  |  |  |  |  |  |
| R Square | 0.004617247 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.001342961 |  |  |  |  |  |  |  |
| Standard Error | 0.441602292 |  |  |  |  |  |  |  |
| Observations | 306 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 0.274997775 | 0.274997775 | 1.410153991 | 0.235957851 |  |  |  |
| Residual | 304 | 59.28382575 | 0.195012585 |  |  |  |  |  |
| Total | 305 | 59.55882353 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 1.118905107 | 0.125348117 | 8.926381449 | 4.26208E-17 | 0.872245315 | 1.365564899 | 0.872245315 | 1.365564899 |
| Age | 0.002779407 | 0.002340555 | 1.187499049 | 0.235957851 | -0.001826333 | 0.007385146 | -0.001826333 | 0.007385146 |

* R- Square -> 0.004617247

That means 0.46 % of the population in the dependent variable (survival) can be explained by the independent variable(Age) which is a very very small amount and thus we cannot accept this mode.

* P-value -> 0.235957851

Ha: Patient who can survive 5-years after operations depends on patient’s age at time of operation

H0: Patient who can survive 5-years after operations **does not** depends on patient’s age at time of operation

In order to prove our hypothesis H0,

p-value(H0) > a

Here, a is equal to 0.05 and p-value is greater than a.

Therefore, our H0 is accepted and Ha is rejected

That means,

**This model cannot be accepted**

* Significat-F -> 0.235957851

It represents the collective p-value of all the independent variables used in the model.

As there is only one I.V used in our model, the analysis of significant-F would be same as p-value.i.e.,

**This model cannot be accepted**

* Equation:

Survival = 1.1189 + 0.0028 \* Age

**To Conclude, the variable Survive is independent of the Age after 5 years.**

**Hypothesis 2:** Patient who can survive 5-years after operations depends on  the year the operation was performed

Regression Summary:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.00476818 |  |  |  |  |  |  |  |
| R Square | 2.27355E-05 |  |  |  |  |  |  |  |
| Adjusted R Square | -003266663 |  |  |  |  |  |  |  |
| Standard Error | 0.442620298 |  |  |  |  |  |  |  |
| Observations | 306 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 0.001354102 | 0.001354102 | 0.006911762 | 0.933797316 |  |  |  |
| Residual | 304 | 59.55746943 | 0.195912728 |  |  |  |  |  |
| Total | 305 | 59.55882353 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 1.30546245 | 0.490886314 | 2.659398748 | 0.008242927 | 0.339497273 | 2.271427626 | 0.339497273 | 2.271427626 |
| Ops Yr | -0.000648443 | 0.007799695 | -0.083137007 | 0.933797316 | -0.015996668 | 0.014699782 | -0.015996668 | 0.014699782 |

* **.** R- Square -> 2.27355E-05

That means 1.56 % of the population in the dependent variable (survival) can be explained by the independent variable(Ops Yr) which is a very very small amount and thus we cannot accept this mode.

* P-value -> 0.933797316
* Ha: Patient who can survive 5-years after operations depends on the year the operation was performed
* H0: Patient who can survive 5-years after operations **does not** depends on the year the operation was performed

In order to prove our hypothesis H0,

p-value(H0) > a

Here, a is equal to 0.05 and p-value is greater than a.

Therefore, our H0 is accepted and Ha is rejected

That means,

**This model cannot be accepted**

* Significat-F -> 0.933797316

It represents the collective p-value of all the independent variables used in the model.

As there is only one I.V used in our model, the analysis of significant-F would be same as p-value.i.e.,

**This model cannot be accepted**

* Equation:

Survival = 1.3054 - 0.0006 \* Ops Yr

**To Conclude, the variable Survive is independent of the Operation Years**

**Hypothesis 3:** Patient who can survive 5-years after operations depends on  the number of nodes infected

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Regression Summary: |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.286767567 |  |  |  |  |  |  |  |
| R Square | 0.082235638 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.079216676 |  |  |  |  |  |  |  |
| Standard Error | 0.42403515 |  |  |  |  |  |  |  |
| Observations | 306 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 4.897857825 | 4.897857825 | 27.23970862 | 3.33539E-07 |  |  |  |
| Residual | 304 | 54.6609657 | 0.179805808 |  |  |  |  |  |
| Total | 305 | 59.55882353 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 1.19374247 | 0.027793357 | 42.95063943 | 3.9567E-131 | 1.139050754 | 1.248434187 | 1.139050754 | 1.248434187 |
| Nodes | 0.017625653 | 0.0033771 | 5.219167426 | 3.33539E-07 | 0.010980201 | 0.024271105 | 0.010980201 | 0.024271105 |

* R- Square -> 0.082235638

That means 8.22 % of the population in the dependent variable (survival) can be explained by the independent variable(Nodes) which is a very very small amount and thus we cannot accept this mode.

* P-value -> 3.33539E-07
* Ha: Patient who can survive 5-years after operations depends on the nodes of the operation
* H0: Patient who can survive 5-years after operations **does not** depends on the nodes of the operation

In order to prove our hypothesis H0,

p-value(H0) > a

Here, a is equal to 0.05 and p-value is greater than a.

Therefore, our H0 is accepted and Ha is rejected

That means,

**This model cannot be accepted**

* Significat-F -> 3.33539E-07

It represents the collective p-value of all the independent variables used in the model.

As there is only one I.V used in our model, the analysis of significant-F would be same as p-value.i.e.,

**This model cannot be accepted**

* Equation:

Survival = 1.1937 + 0.0176 \* Nodes

**To Conclude, the variable Survive is independent of the Nodes of operation**

# **tempData Dataset**

**Hypothesis 1:** January Temperature of the 50 States is associated with latitude

Regression Summary:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  | | | | | |
| Multiple R | 0.848035206 |  |  |  |  |
| R Square | 0.71916371 |  |  |  |  |
| Adjusted R Square | 0.713963038 |  |  |  |  |
| Standard Error | 7.155817489 |  |  |  |  |
| Observations | 56 |  |  |  |  |
|  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 1 | 7080.87305 | 7080.87305 | 138.2828423 | 1.62396E-16 |
| Residual | 54 | 2765.109093 | 51.20572394 |  |  |  |  |  |  |  |  |
| Total | 55 | 9845.982143 |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |  |  |  |
| Intercept | 108.7277422 | 7.056103859 | 15.4090337 | 2.03402E-21 | 94.58110567 | 122.8743786 | 94.58110567 | 122.8743786 |  |  |  |
| Lat | -2.109587848 | 0.179396294 | -11.75937253 | 1.62396E-16 | -2.469255762 | -1.749919935 | -2.469255762 | -1.749919935 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

* R- Square -> 0.71916371

That means 71.9 % of the population in the dependent variable (survival) can be explained by the independent variable(Nodes) which is good and pretty accurate and thus we can accept this model.

* P-value -> 1.62396E-16

Ha: January Temperature of the 50 States is associated with latitude

H0: January Temperature of the 50 States **is not** associated with latitude

In order to prove our hypothesis H0,

p-value(H0) > a

Here, a is equal to 0.05 and p-value is less than a.

Therefore, our H0 is rejected and Ha is accepted

That means,

**This model can be accepted**

* Significat-F -> 1.62396E-16

It represents the collective p-value of all the independent variables used in the model.

As there is only one I.V used in our model, the analysis of significant-F would be same as p-value.i.e.,

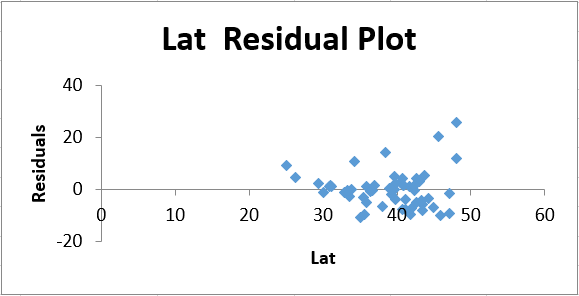
**This model can be accepted**

* Equation:

JanTemp = 108.7277 - 2.1095 \* Lat

* Residual Plot:

We can see that the data is randomly distributed and therefore, we can accept this model



**To Conclude, the variable JanTemp is dependent on the Latitude**