**Simple Regression**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Equation of Simple regression is $y=β\_{0}+β\_{1}+ϵ$

 where $β\_{0}=\overbar{y}-β\_{1}\overbar{x}$ and

 $β\_{1}=\frac{\sum\_{}^{}y\_{i}x\_{i} - \frac{\left(\sum\_{}^{}y\_{i}\right)\left(\sum\_{}^{}x\_{i}\right)}{n}}{\sum\_{}^{}x\_{i}^{2}- \frac{\left(\sum\_{}^{}x\_{i}\right)^{2}}{n}}$

 $\left〈\overbar{y}=\frac{1}{n}\sum\_{}^{}y\_{i} and \overbar{x}=\frac{1}{n}\sum\_{}^{}x\_{i}\right〉$

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Regression Analysis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coef | SE | T | P |
| Dependent | $$\hat{β}\_{0}$$ | See eqn (1) | See eqn (3) | *<p-value>* |
| Independent | $$\hat{β}\_{1}$$ | See eqn (2) | See eqn (4) | *<p-value>* |

 $S\_{xx}=\sum\_{}^{}\left(x\_{i}-\overbar{x}\right)^{2} and S\_{xy}=\sum\_{}^{}y\_{i}\left(x\_{i}-\overbar{x}\right)^{2}$

 i.e. $S\_{xy}=\sum\_{}^{}y\_{i}S\_{xx}$

Least square estimates of the slope and intercept are,

 $\hat{β}\_{1}=\frac{S\_{xy}}{S\_{xx}} and \hat{β}\_{0}=\overbar{y}-\hat{β}\_{1}\overbar{x}$

Estimating $σ^{2}$,

 $SS\_{E}=\sum\_{}^{}\left(y\_{i}-\overbar{y}\right)^{2}$

 therefore,

 $\hat{σ}^{2}=\frac{SS\_{E}}{n-2}$

Estimated error of slope is,

 $SE\left(\hat{β}\_{0}\right)=\sqrt{\hat{σ}^{2}\left[\frac{1}{n}-\frac{\overbar{x}^{2}}{S\_{xx}}\right]}$ (1)

$$SE\left(\hat{β}\_{1}\right)=\sqrt{\frac{\hat{σ}^{2}}{S\_{xx}}} (2)$$

 $T=\frac{\hat{β}\_{0}}{SE\left(\hat{β}\_{0}\right)} (3)$

 $T=\frac{\hat{β}\_{1}}{SE\left(\hat{β}\_{1}\right)} (4)$

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Analysis of Variance (ANOVA)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | SS | MSS | F | P |
| Regression | 1 | $$SS\_{R}=\hat{β}\_{1}S\_{xy}$$ | $$MS\_{R}$$ | $$\frac{MS\_{R}}{MS\_{E}}$$ | <p-value> |
| Residual error | n-2 | $$SS\_{E}=SS\_{T}-\hat{β}\_{1}S\_{xy}$$ | $$MS\_{E}$$ |  |  |
| Total | n-1 | $$SS\_{T}=\sum\_{}^{}\left(y-\overbar{y}\right)$$ |  |  |  |

 where, $MS\_{R}=\frac{SS\_{R}}{n\left(n-1\right)} and MS\_{E}=\frac{SS\_{E}}{n-2} $

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Student’s t test**

$$t\_{0}=\frac{\hat{β}\_{1}}{SE\left(\hat{β}\_{1}\right)}$$

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**CONFIDANCE INTERVAL (CI) for Regression Analysis**

CI on the slope $β\_{1}$ is,

$$\hat{β}\_{1}-t\_{\frac{α}{2},n-2}\sqrt{\frac{\hat{σ}^{2}}{S\_{xx}}\leq }β\_{1}\leq \hat{β}\_{1}+t\_{\frac{α}{2},n-2}$$

CI on the intercept $β\_{0}$ is,

$$\hat{β}\_{1}-t\_{\frac{α}{2},n-2}\sqrt{\hat{σ}\left[\frac{1}{n}-\frac{\overbar{x}^{2}}{S\_{xx}}\right]}\leq β\_{0}\leq \hat{β}\_{1}+t\_{\frac{α}{2},n-2}\sqrt{\hat{σ}\left[\frac{1}{n}-\frac{\overbar{x}^{2}}{S\_{xx}}\right]}$$

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Coefficient of Determination**

$$R^{2}=\frac{SS\_{R}}{SS\_{T}}=1-\frac{SS\_{E}}{SS\_{T}}$$

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_