

Regression properties

what do we need know about

Chuvakin Sergey

«School of Advanced Studies»

January 7, 2021

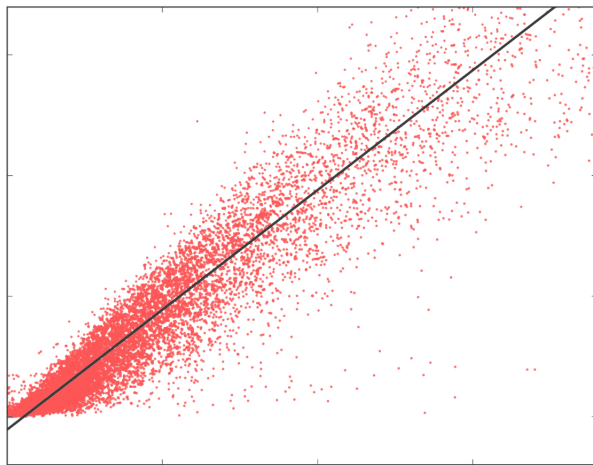
Outline

- ▶ Significance
- ▶ Regression output (table)
- ▶ Regression plot
- ▶ Standart Error
- ▶ T-value
- ▶ P-value
- ▶ Confidence Intervals
- ▶ R-squared

Significance

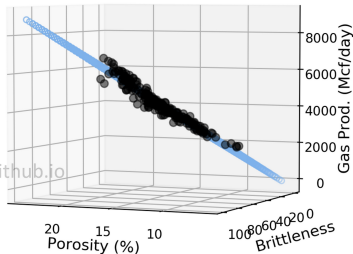
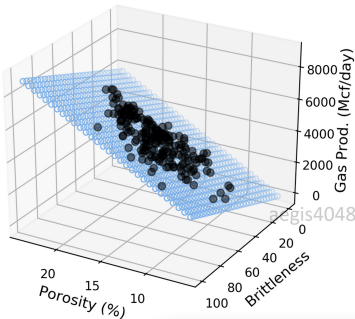
Significance tells if we can make a proper, consistent and robust inference out of our data.

Regression line

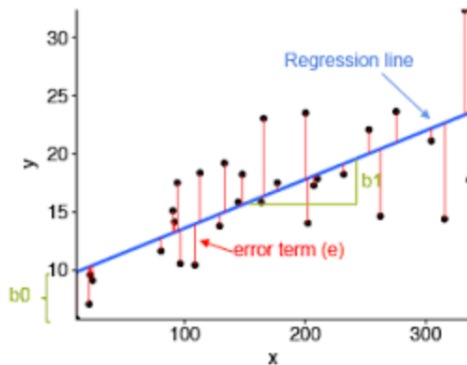


Regression line

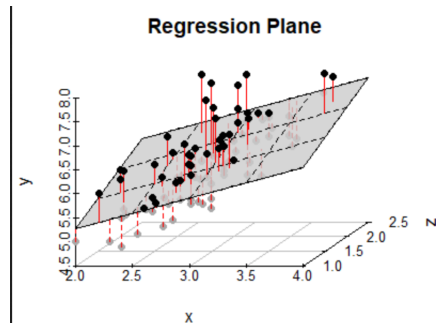
3D multiple linear regression model



Regression line



Regression line



Regression output (table)

Call:

```
lm(formula = Price ~ InMichelin + Food + Decor + Service, data = dfData)
```

Residuals:

Min	1Q	Median	3Q	Max
-20.898	-5.835	-0.755	3.457	105.785

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-57.6004	9.2337	-6.238	3.84e-09 ***
InMichelin	1.9931	2.6357	0.756	0.451
Food	0.2006	0.6683	0.300	0.764
Decor	2.2049	0.3930	5.610	8.76e-08 ***
Service	3.0598	0.5705	5.363	2.84e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.55 on 159 degrees of freedom

Multiple R-squared: 0.6344, Adjusted R-squared: 0.6252

F-statistic: 68.98 on 4 and 159 DF, p-value: < 2.2e-16

Standard Error

SE - can be interpreted as standard deviation of beta coefficient.

$$se(\beta_i) = \frac{\hat{\sigma}}{\sqrt{\sum x_i^2}}$$

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}}$$

$$\hat{u}_i = \bar{y}_i - y_i$$

Standart Error

- ▶ x - predictor
- ▶ β - coeficient
- ▶ u - Residual Sum of Squares
- ▶ y - target
- ▶ \bar{y} - predicted target
- ▶ n - number of degrees of freedom

Student's value

t- value - part of T-distribution that helps to understand if certain coefficient differs from 0.

$$t - value = \frac{\beta}{se(\beta)}$$

What's next?

1. Find number of DF
2. Guess level of significance you need
3. Find threshold in matrix [here](#)
4. If your t-value is higher that means that coef is significant

P-value

P - value - it's just convenient form of t-student value.

- ▶ if P-value is below than 0.05 that means that we can reject null hypothesis (coef is significant)
- ▶ if P-value is above than 0.05 that means that we can not reject null hypothesis (coef is not significant)

Confidence Intervals

Confidence interval can be calculated even for β coefs.

Steps to calculate:

1. Calculate Margin Error

ME = treshhold of t-value * standard error

2. Lower and Upper bounds: $CI = \beta \pm ME$

If CI includes zero, this automatically means, that coef is not significant. The same aproach is applied to get Prediction Interval. Ypu need just change β to \hat{y} .

Prediction Intervals

The same approach is applied to get Prediction Interval. Ypu need just change β to \hat{y} .

$$\hat{y} \pm t \times \sqrt{MSE \times \left(1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)}$$

Prediction Intervals

