Regression analysis is a form of inferential statistics. The p-values help determine whether the relationships (that you observe in your sample) also exist in the larger population. The p-value for each independent variable tests the null hypothesis (that the variable has no correlation with the dependent variable. If there is no correlation, there is no association between the changes in the independent variable and the shifts in the dependent variable.) In other words, there is insufficient evidence to conclude that there is effect at the population level.

모집단

연관

X: null hypothesis : zero correlation

(If the p-value for a variable is less than your significance level,) your sample data provide enough evidence to reject the "null hypothesis (for the entire population.) Your data favor the hypothesis (that there *is* a non-zero correlation.) Changes in the independent variable *are* associated with changes in the response at the population level. This variable is statistically significant and probably a worthwhile addition to your regression model.

On the other hand, a p-value (that is greater than the significance level) indicates that there is insufficient evidence in your sample (to conclude that a "non-zero correlation" exists.)

상관 관계가 0.이 아님

"상관관계가 0이 아님'이라고 단정짓기에는 해당 증거가 불충분함."

The sign of a regression coefficient tells you "whether there is a positive or negative correlation between each independent variable the dependent variable." A positive coefficient indicates that as the value of the independent variable increases, the mean of the dependent variable also tends to increase. A negative coefficient suggests that as the independent variable increases, the dependent variable tends to decrease.

· 계수의 부호가 `해당 독립변수와 종속변수가 양의 관계인지 또는 음의 관계인지를 나타낸다.

· 해당 계수는 `되어는 상관계수'가 아빈, `가중치'이다.

The regression output example below shows that the South and North predictor variables are statistically significant because their p-values equal 0.000. On the other hand, East is not statistically significant because its p-value (0.092) is greater than the usual significance level of 0.05.

[handwritten: P-value]

[handwritten (pink): 'coefficient' 표준오차 로 T-value를 산출하고, 해당 T-value를 통해서 각 coefficient들의 p-value를 결정한다.]

```
Coefficients

Term      Coef     SE Coef        T      P
Constant  389.166  66.0937    5.8881   0.000
East        2.125   1.2145    1.7495   0.092
South       5.318   0.9629    5.5232   0.000
North     -24.132   1.8685  -12.9153   0.000
```
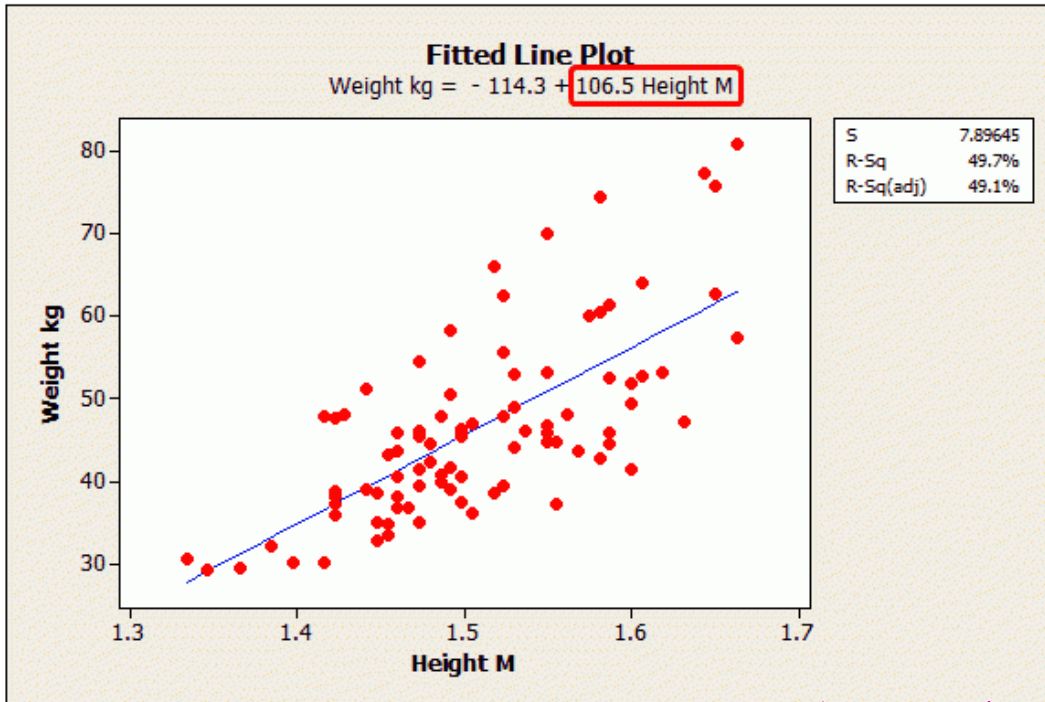
It is standard practice to use the coefficient p-values to decide whether to include variables in the final model. For the results above, we would consider removing East. Keeping variables that are not statistically significant can reduce the model's precision.

```
Coefficients

Term           Coef   SE Coef        T      P
Constant   -114.326   17.4425  -6.55444  0.000
Height M    106.505   11.5500   9.22117  0.000
```

Fitted Line Plot

Weight kg = - 114.3 + 106.5 Height M

S	7.89645
R-Sq	49.7%
R-Sq(adj)	49.1%

coefficient : 계수 (해당 독립변수가 종속변수에 미치는 영향력)

The height "coefficient" in the regression equation is 106.5. This coefficient represents the mean increase of weight in kilograms for every additional one meter in height. If your height increases by 1 meter, the average weight increases by 106.5 kilograms.

The regression line on the graph visually displays the same information. If you move to the right along the x-axis by one meter, the line increases by 106.5 kilograms. Keep in mind that it is only safe to interpret regression results within the observation space of your data. In this case, the height and weight data were collected from middle-school girls and range from 1.3 m to 1.7 m. Consequently, we can't shift along the line by a full meter for these data.