

## Linear Regression &amp; Computer Output: Interpreting Important Variables

I. Minitab/ Computer Printouts

Below is a computer output. You will be expected to use and interpret computer output on the AP Exam. This output is from Minitab, however most computer output looks very similar. We will discuss which numbers you need to know, what they mean, and how to interpret them.

MINITAB OUTPUT FOR BOX 2.1 Analysis of the trees dataset: regression				
Regression Analysis: VOLUME versus HEIGHT				
The regression equation is				
VOLUME = - 87.1 + 1.54 HEIGHT				
Predictor	Coef	SE Coef	T	P
Constant	-87.12	29.27	-2.98	0.006
HEIGHT	1.5433	0.3839	4.02	0.000
S = 13.40		R-Sq = 35.8%	R-Sq (adj) = 33.6%	
Analysis of Variance				
-----				
SE Coef, = 0.3839 represents the standard deviation of the slope				
S = 13.40 represents standard deviation of residuals				

**Constant -----→ -87.12**

This is the y-intercept. This is the value of the response variable when the explanatory variable is 0. Check the context of the situation. Often, there can be no such value.

In this case, it is not possible to have a volume that is negative nor is it possible to have a height of zero.

**Height/Slope ---→ 1.5433**

This is the coefficient of the explanatory variable, thus it is the slope.

This entire line of numbers deals with regression for slope.

For each increase in height of one unit, the volume is expected to increase by *approximately* 1.5433 units. (Actual units were not provided)

**Prediction equation-----→  $y = -87.1 + 1.54x$  (ie. Least Squares Regression Line)**

This is an equation used to make predictions and is based on only one sample.

**SE Coef -----→ 0.3839**

This is the standard deviation of the slope. Remember, this data came from only one sample. We would expect the slope to vary a little from sample to sample. Thus,

If we gathered repeated samples, we would expect the slope of the volumes of the trees to vary by approximately 0.3839 units.

**S -----→ 13.40**

This is the standard deviation of the residuals. The average amount that the observed values differ from the predicted values is 13.40. The average amount that the observed volumes of trees differ from the predicted volumes is approximately 13.40 units.

**Linear Regression & Computer Output: Interpreting Important Variables**

$r^2$  -----→ 35.8%

This is the correlation of determination, which is the fraction or proportion of variation in the y values that is explained by the least squares regression of y on x.

About 35.8% in the variation in volume can be explained by the least squares regression of y( volume) on x ( height).

$r = \sqrt{r^2} = \sqrt{.358} = .598$

This is the correlation coefficient. It tells you strength and direction of the relationship.

With an r value of .598, there is a weak, positive relationship between height and volume of trees.

T -----→ 4.02

This is the test statistic which =  $test\ statistic = \frac{statistic - parameter}{std.dev\ of\ statistic} = \frac{b}{SE}$

P-----→ 0

This is the p-value of a Linear Regression t test.

With a p-value of approx.. 0 less than any alpha level (.05, .01), reject the null. There is evidence that there is a relationship between the volume of a tree and its height.

Example 2: Minitab / Computer Printouts

**Regression Analysis: Height versus Mother Height**

The regression equation is

Height = 24.7 + 0.640 Mother Height

(dependent variable, y) (intercept,  $b_0$ ) (slope,  $b_1$ ) (independent variable, x)

← the estimated regression equation:  $\hat{y} = b_0 + b_1x$

Predictor	(estimates) Coef	(sd of ests.) SE Coef	(test statistics) T	(p-values) P	
Constant	24.690	8.978	2.75	0.009	IGNORE these values tests $H_0: \beta_1 = 0$ , vs. two-tailed altern. (the latter is equivalent to testing for linear correlation between x and y)
Mother H intercept, $b_0$	0.6405	0.1394	4.59	0.000	
slope, $b_1$					

S = 2.973  
(standard error of estimate,  $s_e$ )

R-Sq = 35.7%  
(coefficient of linear determination,  $r^2$ )

R-Sq (adj) = 34.0%  
(adjusted  $r^2$ , used for multiple regression)

\* (the coefficient of linear correlation is the square root of  $r^2$ , with the same sign as the slope,  $b_1$ )

for a simple linear regression minitab models only conduct two-tailed alternatives

S - 2.973 standard error of the estimate, or the standard deviation of the residuals - actual values versus the predicted values

SE Coef - 0.1394 - standard deviation of the slope

**Linear Regression & Computer Output: Interpreting Important Variables**[Practice Minitab / Computer Printouts](#)

1. A sample of men agreed to participate in a study to determine the relationship between several variables including height, weight, waste size, and percent body fat. A scatterplot with percent body fat on the y-axis and waist size (in inches) on the horizontal axis revealed a positive linear association between these variables. Computer output for the regression analysis is given below:

```
Dependent variable is: %BF
R-squared = 67.8%
S = 4.713 with 250-2 = 248 degrees of freedom
```

<b>Variable</b>	<b>Coefficient</b>	<b>se of coeff</b>	<b>t-ratio</b>	<b>prob</b>
Constant	-42.734	2.717	-15.7	<.0001
Waist	1.70	0.0743	22.9	<.0001

(a) Write the equation of the regression line (be sure to use correct notation and define your variables):

(b) Explain/interpret the information provided by R-squared in the context of this problem. Be specific.

(c) Calculate and interpret the correlation coefficient ( $r$ ).

(d) One of the men who participated in the study had waist size 35 inches and 10% body fat. Calculate the residual associated with the point for this individual.

Linear Regression & Computer Output: Interpreting Important Variables

Determine the LSRL, Standard deviation for slope, correlation coefficient and the standard error of the residuals for each:

#2.

MINITAB OUTPUT FOR BOX 2.1 Analysis of the trees dataset: regression

---

Regression Analysis: VOLUME versus HEIGHT  
 The regression equation is

Predictor	Coef	SE Coef	T	P
Constant	-87.12	29.27	-2.98	0.006
HEIGHT	1.5433	0.3839	4.02	0.000

S = 13.40      R-Sq = 35.8%      R-Sq (adj) = 33.6%

#3.

MINITAB OUTPUT FOR BOX 2.3 Analysis of the scores dataset: regression

---

Regression Analysis: MATHS versus ESSAYS  
 The regression equation is

Predictor	Coef	SE Coef	T	P
Constant	27.57	22.26	1.24	0.251
ESSAYS	0.6548	0.3154	2.08	0.072

S = 9.140      R-Sq = 35.0%      R-Sq (adj) = 26.9%