Regression Addendum: Regression Tables

When you use a computer to do regression analysis it will combine all of this information into a series of tables... It will usually look something like the following.

Predictor	Coeff	Std. Error	t	P
Const	1	0.25	4	.0004
x	2	0.447	4.472	.0001

Source	df	SS	MS	F	P
Regression	1	40	40	20	.0001
Residual	30	60	2		
Total	31	100	100/31		

Goodness of Fit	
R^2	$2/5$
S	$\sqrt{2}$

The "Predictor" table gives information about the regression line and its coefficients

$$\text{"Const Coeff"} = \hat{\beta}_0$$

$$\text{"x Coeff"} = \hat{\beta}_1$$

(If you did non-linear regression then other coefficients would also be listed.)

In this case the regression line is

$$\hat{y}(x) = 1 + 2x$$

The "Std. Error" column gives $\sigma_{\hat{\beta}_0}$ & $\sigma_{\hat{\beta}_1}$ which you would use for confidence interval

$$y(x) = \hat{y}(x) \pm qt(\alpha/2, 30)\sqrt{\sigma_{\hat{\beta}_0}^2 + \sigma_{\hat{\beta}_1}^2 (x-\bar{x})^2}$$

In this case

$$\sigma_{\hat{\beta}_0} = .25 \qquad \sigma_{\hat{\beta}_1} = .447$$

The "t" column gives t-score of coefficients

$$\left.\begin{array}{l} t_{\hat{\beta}_0} = \hat{\beta}_0 / \sigma_{\hat{\beta}_0} \\ t_{\hat{\beta}_1} = \hat{\beta}_1 / \sigma_{\hat{\beta}_1} \end{array}\right\} \begin{array}{l} \text{Testing } H_0 : \beta_0 = 0 \\ H_0 : \beta_1 = 0 \end{array}$$

The "p" column gives two-tailed p-values.

Note: # df of β_0 & β_1 is same as MSE

The "Source" table tracks variance.

("Source of Variance")

It is a usual ANOVA table.

"# df (Regression)" = # coefficients − 1 ← = 1 for Linear Regression

"# df (Total)" = # samples − 1

Recall: "Residual" = "Error"

As usual df adds to give Total

SS adds to give Total

$MS = SS/df$ & $F = MSR/MSE$

MSE is "pooled sample variance"

If $Y(x) = (\beta_0 + \beta_1 x) + \varepsilon$ then

$$Var[\varepsilon] = MSE$$

$$\sigma_\varepsilon = \sqrt{MSE}$$

MST is "total sample variance"

The p-value is $1 - pf(\underbrace{MSR/MSE}_{F}, \underbrace{dfR}_{1}, \underbrace{dfE}_{n-2})$

→ This is the p-value for $H_0 : \beta_1 = 0$

(Note: It is same as p-value in "Predictor" table.)

The Goodness of Fit table records values which tell how "close" samples are to regression line $\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$.

$$R^2 = 1 - SSE/SST$$

$$= SSR/SST$$ ← square of "sample correlation"

$$S = \sqrt{MSE}$$ ← "pooled sample standard deviation"

Sometimes other values will also be given in the "Goodness of Fit" table. For example, "Adjusted R^2" which is a modification of R^2 used for non-linear regression (it is modified to account for # coefficients because adding more coefficients will artificially increase R^2...)