Summary formula sheet for simple linear regression

Slope $b = \sum (Y_i - \bar{Y})(X_i - \bar{X}) / \sum (X_i - \bar{X})^2$

Variance $\sigma^2 / \sum (X_i - \bar{X})^2$

Intercept $a = \bar{Y} - b\,\bar{X}$

Variance of a $\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2}\right] \sigma^2$

Estimated mean at X_0 $\quad a + b\,X_0$

Variance $\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}\right] \sigma^2$

Estimated individual at X_0 $\quad a + b\,X_0$

Variance $\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}\right] \sigma^2$

Total SS $= \sum (Y_i - \bar{Y})^2$

Regression SS =

$[\sum (Y_i - \bar{Y})(X_i - \bar{X})]^2 / \sum (X_i - \bar{X})^2$

Error SS = Total SS - Regression SS

R^2 = Regression SS/ Total SS = "proportion explained"

MSE = error *mean* square = estimate of σ^2 = Error SS/ df
df= degrees of freedom = n-2 for simple linear.

Example

Data points $(x_1,y_1), (x_2,y_2), ..., (x_n,y_n)$
(1,5), (2,7), (3,9), (4,6), (5,8)
$\bar{x} = 15/5=3, \quad \bar{y} = 7$
Corrected sum of squares for x:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = S_{xx} = (1-3)^2+...+(5-3)^2 = 10$$

Corrected sum of squares for y:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = S_{yy} = (5-7)^2+...+(8-7)^2 = 10$$

Corrected sum of cross products = S_{xy}=

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i-\bar{y}) =$$

(-2)(-2)+(-1)(0)+...+(2)(1) = 5 =

$$\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 110-5(3)(7)$$

Slope: $b = S_{xy}/S_{xx} = 5/10 = 0.5$

Intercept:

$\bar{y} - b\,\bar{x} = 7-0.5(3) = 5.5$

$\hat{y}=5.5 + 0.5x$

y	5	7	9	6	8
$\hat{y}$	6	6.5	7	7.5	8
$r=y-\hat{y}$	-1	0.5	2	-1.5	0

$\sum_{i=1}^{n} r_i^2$ = "Error sum of squares" =

SSE = 1+0.25+4+2.25=7.5

SSE is also $S_{yy} - S_{xy}^2/S_{xx} = S_{yy} - b^2 S_{xx} = 10-5^2/10$

Variance of b:

$MSE/S_{xx} = 2.5/10 = 0.25.$

$\sqrt{MSE/S_{xx}}$ is called "standard error" of b.

Task: test H_0: true slope is 0

$t = b/\sqrt{0.25} = 1$ which is not an unusual t.

```
data a; input x y  @@; cards;
1 5     2 7      3 9     4 6     5 8
 ;
proc reg; model Y =X / p;
run;
```

```
                    Dependent Variable: y

                    Analysis of Variance

                     Sum of        Mean
Source        DF    Squares      Square F Value    Pr > F
Model          1    2.50000     2.50000    1.00    0.3910
Error          3    7.50000     2.50000
Corr Total     4   10.00000


Root MSE              1.58114    R-Square     0.2500
Dependent Mean        7.00000    Adj R-Sq     0.0000
Coeff Var            22.58770

                    Parameter Estimates

                   Parameter  Standard
Variable      DF   Estimate     Error   t Value   Pr > |t|
Intercept      1    5.50000   1.65831      3.32     0.0452
x              1    0.50000   0.50000      1.00     0.3910

                   Output Statistics
```

Obs	Dep Var y	Predicted Value	Residual
1	5.0000	6.0000	-1.0000
2	7.0000	6.5000	0.5000
3	9.0000	7.0000	2.0000
4	6.0000	7.5000	-1.5000
5	8.0000	8.0000	0