Scope

This document shows the solution for the data set in the table below. Your data set will almost certainly be different than this one, but the methods illustrated here are still valid for your data set. To refresh your memory, we begin with the general formulas involved. If you want to skip the formulas, click here.

Formulas

There are n bivariate observations, $(X_i, Y_i),\ i = 1, \ldots, n$.

Means

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Standard Deviations

$$S_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}, \quad S_Y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

Pearson Correlation

$$r = \frac{1}{n-1}\sum_{i=1}^{n} X_i' Y_i',$$

where X_i' and Y_i' are the standardized variates

$$X_i' = \frac{X_i - \overline{X}}{S_X} \text{ and } Y_i' = \frac{Y_i - \overline{Y}}{S_Y}$$

Least Squares Estimators

$$\hat{\beta}_1 = r\frac{S_Y}{S_X}, \quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$$

Residuals

$$e_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i),\ i = 1, \ldots, n$$

Sums of Squares and Mean Squares

SSTO= $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$

SSE=$\sum_{i=1}^{n} e_i^2$, MSE=SSE/$(n-2)$

SSR=SSTO$-$SSE, MSR=SSR

Standard Errors of Least Squares Estimators

Define $\hat{\sigma} = \sqrt{\text{MSE}}$. Then the standard errors are

$$\begin{aligned} \hat{\sigma}(\hat{\beta}_0) &= \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}} \\ \hat{\sigma}(\hat{\beta}_1) &= \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}} \end{aligned}$$

Hypothesis Test

You can use a t test to test $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$. The p-value of the test equals $p\pm = P(|T| \geq |t^*|)$, where T has a t distribution with 1 degree of freedom, and

$$t^* = \frac{\hat{\beta}_1}{\hat{\sigma}(\hat{\beta}_1)}$$

Data

There are $n = 6$ bivariate observations given in the table below, which also includes their standardized values, the means and standard deviations of both variables and their Pearson correlation.

Age (X)	Salary (Y)	X'	Y'	$X'Y'$
50	449	0.66664	1.34039	0.89356
41	224	−0.57469	−0.51554	0.29627
35	157	−1.40224	−1.06819	1.49786
44	201	−0.16091	−0.70525	0.11348
45	265	−0.02299	−0.17734	0.00408
56	423	1.49419	1.12593	1.68236
$\overline{X} = 45.\overline{16}$	$\overline{Y} = 286.50$			$r = 0.8975$
$S_X = 7.25$	$S_Y = 121.23$			

SAS Code

The following SAS code will produce all the output needed to answer questions a.)-f.) (and more). The data step reads the data into the SAS data set *salary*. The SAS procedure *proc reg* performs the regression and outputs results. The corr option produces a correlation matrix.

```
data salary;
 input age salary;
datalines;
50 449
41 224
35 157
44 201
45 265
56 423
;
run;
proc reg data=salary corr;
 model salary=age;
run;
```

Solutions

Below, we show how the SAS output can be used to answer questions a.)-f.).

a.) The Pearson correlation, $r = 0.8975$ can be read from the correlation matrix in the SAS output:

Correlation

Variable	age	salary
age	1.0000	0.8975
salary	0.8975	1.0000

b.) The least squares estimates, $\hat{\beta}_0 = -391.34369$ and $\hat{\beta}_1 = 15.00761$ are found in the Parameter Estimates table in the SAS output:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-391.34369	168.29730	-2.33	0.0807
age	1	15.00761	3.68677	4.07	0.0152

c.) and d.) The error sum of squares, SSE= 14290 can be found in the Analysis of Variance table in the SAS output, as can the mean squared error, MSE= 3572.49620:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	59198	59198	16.57	0.0152
Error	4	14290	3572.49620		
Corrected Total	5	73488			

e.) The estimated standard errors, $\hat{\sigma}(\hat{\beta}_0) = 168.29730$ and $\hat{\sigma}(\hat{\beta}_1) = 3.68677$ can be found in the Parameter Estimates table shown above.

f.) The value of the t test statistic used to test $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$ is 4.07 and is found in the Parameter Estimates table above. To its right in the table, you can see the p-value $p\pm = 0.0152$. Since $p\pm > 0.01$, we do not reject H_0 at the 0.01 significance level, and conclude that there is not a statistically significant relationship between age and salary.