ANSWERS TO TEST NUMBER 8

The data are as follows:

IMF INTERNATIONAL FINANCIAL STATISTICS FROM CHASS DATA CENTER

156/64 CANADA / CONSUMER PRICES (Index number)
156/67R CANADA / UNEMPLOYMENT RATE (Percent per annum)

	DATE	CPI	UEM	INF	DUEM	DINF	DUEMSQ	DINFSQ	DUEM x DINF
	1984	69.2							
	1985	71.9	10.50	3.95	1.04	0.97	1.08	0.94	1.01
	1986	74.9	9.60	4.17	0.14	1.19	0.02	1.43	0.17
	1987	78.2	8.90	4.36	-0.56	1.38	0.31	1.92	-0.78
	1988	81.3	7.80	4.02	-1.66	1.04	2.76	1.09	-1.73
	1989	85.4	7.50	4.99	-1.96	2.01	3.84	4.06	-3.95
	1990	89.5	8.10	4.76	-1.36	1.78	1.85	3.19	-2.43
	1991	94.5	10.40	5.62	0.94	2.64	0.88	6.94	2.48
	1992	95.9	11.30	1.51	1.84	-1.47	3.39	2.17	-2.71
	1993	97.7	11.20	1.84	1.74	-1.14	3.03	1.30	-1.98
	1994	97.9	10.40	0.19	0.94	-2.79	0.88	7.81	-2.63
	1995	100.0	9.55	2.17	0.09	-0.81	0.01	0.66	-0.07
	1996	101.6	9.70	1.57	0.24	-1.41	0.06	1.97	-0.34
	1997	103.2	9.22	1.62	-0.24	-1.36	0.06	1.84	0.33
	1998	104.2	8.34	0.99	-1.12	-1.99	1.25	3.97	2.23
SUM			132.51	41.77	0.00	0.00	19.42	39.29	-10.40
MEAN			9.46	2.98					

CPI = CONSUMER PRICE INDEX
INF = (CPI(t) - CPI(t-1))/CPI(t-1)
UEM = UNEMPLOYMENT RATE
DINF = INF - MEAN(INF)
DUEM = UEM - MEAN(UEM)
DUEMSQ = DUEM SQUARED
DINFSQ = DINF SQUARED
DUEM x DINF = DUEM TIMES DINF

Since the unemployment rate is thought to respond to the inflation rate, it would seem appropriate to let `UEM` be the dependent variable Y and `INF` be the independent variable X. The question requires, however, that we find out whether there is a significant negative *relationship* between the two variables—we can do this by letting either variable be the dependent variable. The regression equation can be written

$$Y = \beta_0 + \beta_1 X + \epsilon.$$

Letting `UEM` be Y and `INF` be X we can calculate b_1, the point estimate of β_1 , using the numbers from the `SUM` row of the above table, as follows:

$$b_1 = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{-10.40}{39.29} = -.265$$

If we run the regression the other way, letting `INF` be the dependent variable Y, we obtain a value of b_1 equal to

$$b_1 = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{-10.40}{19.42} = -.535$$

The correlation coefficient between X and Y is

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}} = \frac{-10.40}{\sqrt{19.42}\sqrt{39.29}} = \frac{-10.40}{(4.41)(6.29)} = -.376$$

We can square r obtain $R^2 = .141$ and then obtain the sum of squared residuals (SSE) for the regression with `UEM` as the dependent variable from the fact that

$$SSR = R^2\, SSTO = R^2\, 19.42 = (.141)(19.42) = 2.74$$

whence

$$SSE = SSTO - SSR = 19.42 - R^2\, 19.42 = (1 - .141)(19.42) = 16.68$$

When `INF` is the dependent variable, the calulation is

$$SSE = SSTO - SSR = (1 - R^2)\, 39.29 = (.859)(39.29) = 33.75$$

The mean square errors (MSE) in the two cases are

$$MSE = \frac{SSE}{n-2} = \frac{16.68}{14-2} = \frac{16.68}{12} = 1.39$$

and

$$MSE = \frac{SSE}{n-2} = \frac{33.75}{12} = 2.81$$

respectively.

[Give yourself 10 points for understanding how to calculate each of b_1, R^2, SSR, SSE and MSE.]

The variance of b_1 in the case where `UEM` is the dependent variable is therefore

$$Var\{b_1\} = \frac{MSE}{\sum (X - \bar{X})^2} = \frac{1.39}{39.29} = .0354$$

yielding a standard deviation of b_1 of $s_{b_1} = \sqrt{.0354} = .188$ and a t-ratio for the test of the null hypothesis that $\beta_1 \geq 0$ of

$$t = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-.265 - 0}{.188} = \frac{-.265}{.188} = -1.409$$

In reverse case where `INF` is the dependent variable the corresponding variance of b_1 is

$$Var\{b_1\} = \frac{MSE}{\sum (X - \bar{X})^2} = \frac{2.81}{19.42} = .145$$

yielding $s_{b_1} = \sqrt{.145} = .381$ and the t-ratio

$$t = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-.535 - 0}{.381} = \frac{-.535}{.3.81} = -1.404$$

Since the t-ratio tests the significance of the relationship between the two variables, which should be the same regardless of which happens to be the dependent variable, it is not suprising that the two alternative calculations of that t-ratio differ by rounding error—in the absense of rounding error, they would be the same.

Using Xlispstat we can calculate the critical value of t for $\alpha = .05$ as

```
> (t-quant .95 12)
1.782287554577734
```

Since this critical value exceeds the absolute value of our t-statistic, we cannot reject the null hypothesis that $\beta_1 \geq 0$. Since the P-value turns out to be approximately

```
> (t-cdf -1.4065 12)
0.09247055367274187
```

we could reject the null hypothesis at the 10% level of significance.

[Give yourself 10 points for understanding how to calculate s_{b_1} and a further 10 points for understanding how to calculate the t-statistic. Give yourself 10 points for understanding what the correct null hypothesis is, 10 more points for calculating the correct critical values and 10 points for reaching the correct statistical decision.]

It is useful to plot the data to visually examine the possible relationship between the variables. This is done using Xlispstat with the following code

```
> (load "cpiuedat");
loading cpiuedat.lsp
T
> (plot-points uem inf)
#<Object: 816e2a0, prototype = SCATTERPLOT-PROTO>>
```

The first command loads the data from the file `cpiuedat.lsp` while the second creates the plot, which is shown in Figure 1.

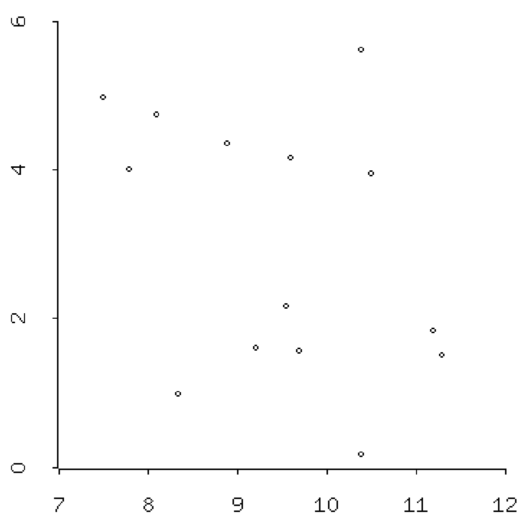

Figure 1: Scatterplot of the Canadian unemployment rate (horizontal axis) vs. the Canadian CPI inflation rate (vertical axis): Annual data from 1985 through 1998.

The scatter indicates a very weak relationship between the two variables.

The regression-results calculated using the (`regression-model`) function in Xlispstat are

```
>(def reguem (regression-model inf uem))

Least Squares Estimates:
Constant                    10.2537      (0.643865)

Variable 0                -0.264407      (0.188245)
R Squared:                 0.141193
Sigma hat:                  1.17889
Number of cases:                 14
Degrees of freedom:              12

REGUEM

>(def reginf (regression-model uem inf))

Least Squares Estimates:

Constant                    8.03715      (3.62617)
Variable 0                -0.533998      (0.380181)

R Squared:                 0.141193
Sigma hat:                  1.67536
Number of cases:                 14
Degrees of freedom:              12

REGINF
```

The first regression uses `UEM` as the dependent variable and the second uses `INF` as the dependent variable. `Sigma hat` is the square root of MSE. The results differ from our calculations above only due to rounding error.

Finally, let us plot the residuals.

```
> (def residuals (send reguem :residuals))
RESIDUALS
> (def dates (iseq 85 98))
DATES
> (def cpiuem2 (plot-lines dates residuals))
```

CPIUEM2

They are shown in Figure 2. The residuals appear to be serially correlated.

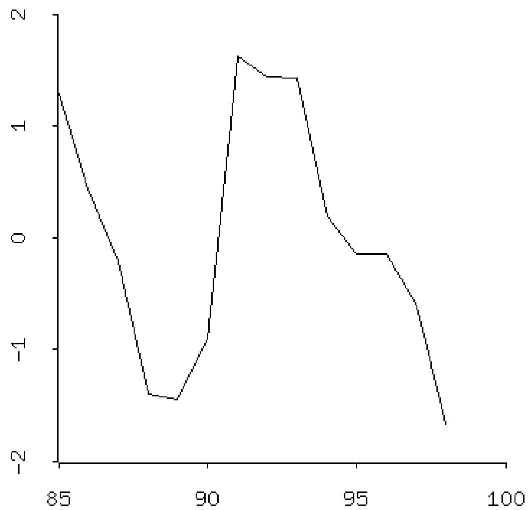

Figure 2: Residuals from regression of Canadian unemployment rate on the Canadian CPI inflation rate.