The mass of steam required to heat a building can be related to the average ambient temperature. Being able to predict the mass of steam required, s, when given the ambient temperature, T, can help in energy planning, and ultimately lead to energy reduction.

The table below lists the mass of steam produced [ton] in the past when the average temperature over a 2 hour period, recorded in K, was observed outside.

Temperature = T [Kelvin]	267	268	272	273	278	281	283	288	289	293	296
Steam produced = s [tons]	220	251	211	210	155	152	122	157	100	64	58

The following calculations have already been performed for you:

- Number of samples, $n = 11$
- Temperature: mean = $\overline{T} = 281$ K and standard deviation is 10.0 K
- Average steam produced, $\overline{s} = 155$ tons, and standard deviation is 64.4 tons.

The modified output from a certain statistical package is:

```
Coefficients:
                           Standard
               Value       Error
-----------------------------------
(Intercept)    1871.6936   183.2183
T                -6.1168     0.6523

Residual standard error: ____ on ____ degrees of freedom
Multiple R-squared: ____
```

A portion of the analysis of variance table is given below:

```
     Analysis of Variance
-----------------------------
                       Sum of
Source                Squares
-----------------------------
Due to the model        37572
Due to error             3845
-----------------------------
Total                   41417
```

1. What is the interpretation of the intercept? Is it a useful piece of knowledge derived from the model?
2. What is the interpretation of the slope coefficient? Is it a useful piece of knowledge derived from the model?
3. What is the multiple R^2 value that would have been calculated for this model?
4. What is the standard error, S_E, value that would have been calculated for this model?
5. How would you interpret your calculated standard error value, and what assumptions are required to match your interpretation?
6. Give a confidence interval for the slope coefficient and interpret what it means.
7. Which other input variable might be added to the linear model to help improve the model's prediction ability?

Temperature	Duration	Speed	Baffles	Yield
82	260	4300	No	51
90	260	3700	Yes	30
88	260	4200	Yes	40
86	260	3300	Yes	28
80	260	4300	No	49
78	260	4300	Yes	49
82	260	3900	Yes	44
83	260	4300	No	59
64	260	4300	No	60
73	260	4400	No	59
60	260	4400	No	57
60	260	4400	No	62
101	260	4400	No	42
92	260	4900	Yes	38

1. The linear model that uses the reactor temperature to predict the yield is: $\hat{y} = 102.5 - 0.69T$. Interpret the slope and intercept term.
2. Build a linear model that uses the impeller speed to predict yield is: $\hat{y} = -20.3 + 0.016S$. Interpret the slope and intercept term.
3. Build a linear model that uses the presence (represent it as 1) or absence (represent it as 0) of baffles to predict yield is: $\hat{y} = 54.9 - 16.7B$. Interpret the slope and intercept term.
4. Which variable(s) would you change to boost the batch yield, at the lowest cost of implementation?

The R code (below) was used to answer all questions.

1. - The model is: $\hat{y} = 102.5 - 0.69T$, where T is tank temperature.
   - Intercept = 102.5 % points is the yield when operating at 0 °C. Obviously not a useful interpretation, because data have not been collected in a range that spans, or is even close to 0 °C. It is likely that this bioreactor system won't yield any product under such cold conditions. Further, a yield greater than 100% is not realizable.
   - Slope = -0.69 $\frac{[\%]}{[°\text{C}]}$, indicating the yield decreases, on average, by about 0.7 units for every degree increase in tank temperature.
2. - The model is: $\hat{y} = -20.3 + 0.016S$, where S is impeller speed.
   - Intercept = -20.3 % points is the yield when operating no agitation. Again, obviously not a useful interpretation, because the data have not been collected under these conditions, and yield can't be a negative quantity.
   - Slope = 0.016 $\frac{[\%]}{[\text{RPM}]}$, indicating the yield increases, on average, by about 1.6 percentage points per 100 RPM increase.
3. - The model is: $\hat{y} = 54.9 - 16.7B$, where B is 1 if baffles are present and $B = 0$ with no baffles.
   - Intercept = 54.9 % points yield is the yield when operating with no baffles (it is in fact the average yield of all the rows that have "No" as their baffle value).
   - Slope = -16.7 %, indicating the presence of baffles decreases the yield, on average, by about 16.7 percentage points.
4. This is an open-ended, and case specific. Some factors you would include are:
   - Remove the baffles, but take into account the cost of doing so. Perhaps it takes a long time (expense) to remove them, especially if the reactor is used to produce other products that do require the baffles.
   - Operate at lower temperatures. The energy costs of cooling the reactor would factor into this.
   - Operate at higher speeds and take that cost into account. Notice however there is one observation at 4900 RPM that seems unusual: was that due to the presence of baffles, or due to temperature in that run? We'll look into this issue with multiple linear regression later on.