```
  Course: Regression_Models
  Lesson: MultiVar_Examples


- Class: text
  Output: "MultiVar_Examples. (Slides for this and other Data Science courses may be
  found at github https://github.com/DataScienceSpecialization/courses. If you care to
  use them, they must be downloaded as a zip file and viewed locally. This lesson
  corresponds to Regression_Models/02_02_multivariateExamples.)"

- Class: text
  Output: In this lesson, we'll look at some examples of regression models with more
  than one variable. We'll begin with the Swiss data which we've taken the liberty to
  load for you. This data is part of R's datasets package. It was gathered in 1888, a
  time of demographic change in Switzerland, and measured six quantities in 47
  French-speaking provinces of Switzerland. We used the code from the slides (the R
  function pairs) to display here a 6 by 6 array of scatterplots showing pairwise
  relationships between the variables. All of the variables, except for fertility, are
  proportions of population. For example, "Examination" shows the percentage of
  draftees receiving the highest mark on an army exam, and "Education" the percentage
  of draftees with education beyond primary school.

- Class: mult_question
  Output: From the plot, which is NOT one of the factors measured?
  AnswerChoices: Obesity; Catholic; Fertility; Infant Mortality
  CorrectAnswer: Obesity
  AnswerTests: omnitest(correctVal='Obesity')
  Hint: Which of the choices doesn't appear on the plot at all?

- Class: cmd_question
  Output: First, use the R function lm to generate the linear model "all" in which
  Fertility is the variable dependent on all the others. Use the R shorthand "." to
  represent the five independent variables in the formula passed to lm.  Remember the
  data is "swiss".
  CorrectAnswer: all <- lm(Fertility ~ ., swiss)
  AnswerTests: creates_lm_model('all <- lm(Fertility ~ . , swiss)')
  Hint: Type "all <- lm(Fertility ~ ., swiss)" at the R prompt.

- Class: cmd_question
  Output: Now look at the summary of the linear model all.
  CorrectAnswer: summary(all)
  AnswerTests: omnitest(correctExpr='summary(all)')
  Hint: Type "summary(all)" at the R prompt.

- Class: text
  Output: Recall that the Estimates are the coefficients of the independent variables
  of the linear model (all of which are percentages) and they reflect an estimated
  change in the dependent variable (fertility) when the corresponding independent
  variable changes. So, for every 1% increase in percent of males involved in
  agriculture as an occupation we expect a .17 decrease in fertility, holding all the
  other variables constant; for every 1% increase in Catholicism, we expect a .10
  increase in fertility, holding all other variables constant.

- Class: mult_question
  Output: The "*" at the far end of the row indicates that the influence of Agriculture
  on Fertility is significant. At what alpha level is the t-test of Agriculture
  significant?
  AnswerChoices: 0.05; 0.01; 0.1; R doesn't say
  CorrectAnswer: 0.05
  AnswerTests: omnitest(correctVal='0.05')
  Hint: Look at the "Signif. codes" line in the summary output.

- Class: cmd_question
  Output: Now generate the summary of another linear model (don't store it in a new
  variable) in which Fertility depends only on agriculture.
  CorrectAnswer: summary(lm(Fertility ~ Agriculture, swiss))
  AnswerTests: omnitest(correctExpr='summary(lm(Fertility ~ Agriculture, swiss))')
  Hint: Type "summary(lm(Fertility ~ Agriculture, swiss))" at the R prompt.

```

```
- Class: mult_question
  Output: What is the coefficient of agriculture in this new model?
  AnswerChoices: 0.19420; 60.30438; 0.07671; *
  CorrectAnswer: 0.19420
  AnswerTests: omnitest(correctVal='0.19420')
  Hint: Look at the "Estimate" column and "Agriculture" row of the summary data you
  just generated.


- Class: text
  Output: The interesting point is that the sign of the Agriculture coefficient changed
  from negative (when all the variables were included in the model) to positive (when
  the model only considered Agriculture). Obviously the presence of the other factors
  affects the influence Agriculture has on Fertility.

- Class: mult_question
  Output: Let's consider the relationship between some of the factors. How would you
  expect level Education and performance on an Examination to be related?
  AnswerChoices: They would be correlated;  They would be uncorrelated; I would not be
  able to guess without more information
  CorrectAnswer: They would be correlated
  AnswerTests: omnitest(correctVal='They would be correlated')
  Hint: How well would you do on an exam without any class or preparation or swirl
  lesson?

- Class: cmd_question
  Output: Now check your intuition with the R command "cor". This computes the
  correlation between Examination and Education.
  CorrectAnswer: cor(swiss$Examination,swiss$Education)
  AnswerTests:
  ANY_of_exprs('cor(swiss$Examination,swiss$Education)','cor(swiss$Education,swiss$Examin
  ation)')
  Hint: Type "cor(swiss$Examination,swiss$Education)" at the R prompt.

- Class: cmd_question
  Output: The correlation of .6984 shows the two are correlated. Now find the
  correlation between Agriculture  and Education.
  CorrectAnswer: cor(swiss$Agriculture,swiss$Education)
  AnswerTests:
  ANY_of_exprs('cor(swiss$Agriculture,swiss$Education)','cor(swiss$Education,swiss$Agricu
  lture)')
  Hint: Type "cor(swiss$Agriculture,swiss$Education)" at the R prompt.

- Class: text
  Output: The negative correlation (-.6395) between Agriculture and Education might be
  affecting Agriculture's influence on Fertility. I've loaded and sourced the file
  swissLMs.R in your working directory. In it is a function makelms() which generates a
  sequence of five linear models. Each model has one more independent variable than the
  preceding model, so the first has just one independent variable, Agriculture, and the
  last has all 5. I've tried loading the source code in your editor. If I haven't done
  this, open the file manually so you can look at the code.

- Class: cmd_question
  Output: Now run the function makelms() to see how the addition of variables affects
  the coefficient of Agriculture in the models.
  CorrectAnswer: makelms()
  AnswerTests: omnitest(correctExpr='makelms()')
  Hint: Type "makelms()" at the R prompt.

- Class: mult_question
  Output: The addition of which variable changes the sign of Agriculture's coefficient
  from positive to negative?
  AnswerChoices: Education; Catholic; Examination; Infant.Mortality
  CorrectAnswer: Education
  AnswerTests: omnitest(correctVal='Education')
  Hint: The sign changes with the third model. From R code the independent variable
  that appears in the third call but not in the second is ?

- Class: cmd_question
```

```
  Output: Now we'll show what happens when we add a variable that provides no new
  linear information to a model. Create a variable ec that is the sum of
  swiss$Examination and swiss$Catholic.
  CorrectAnswer: ec <- swiss$Examination+swiss$Catholic
  AnswerTests: ANY_of_exprs('ec <- swiss$Examination+swiss$Catholic','ec <-
  swiss$Catholic+swiss$Examination')
  Hint: Type "ec <- swiss$Examination+swiss$Catholic" at the R prompt.

- Class: cmd_question
  Output: Now generate a new model efit with Fertility as the dependent variable and
  the remaining 5 of the original variables AND ec as the independent variables. Use
  the R shorthand ". + ec" for the righthand side of the formula.
  CorrectAnswer: efit <- lm(Fertility ~ . + ec, swiss)
  AnswerTests: creates_lm_model('efit <- lm(Fertility ~ . + ec, swiss)')
  Hint: Type "efit <- lm(Fertility ~ . + ec, swiss)" at the R prompt.

- Class: text
  Output: We'll see that R ignores this new term since it doesn't add any information
  to the model.

- Class: cmd_question
  Output: Subtract the efit coefficients from the coefficients of the first model you
  created, all.
  CorrectAnswer: all$coefficients - efit$coefficients
  AnswerTests: omnitest(correctExpr='all$coefficients - efit$coefficients')
  Hint: Type "all$coefficients-efit$coefficients" at the R prompt.


- Class: mult_question
  Output:  Which is the coefficient of ec?
  AnswerChoices: NA; 0; I haven't a clue.
  CorrectAnswer: NA
  AnswerTests: omnitest(correctVal='NA')
  Hint: Since ec is a linear combination of two othe variables R ignores it so its
  coefficient is Not Available.

- Class: mult_question
  Output:  This tells us that
  AnswerChoices: Adding ec doesn't change the model; Adding ec zeroes out the
  coefficients; R is really cool
  CorrectAnswer: Adding ec doesn't change the model
  AnswerTests: omnitest(correctVal='Adding ec doesn\'t change the model')
  Hint: Since ec is a linear combination of two othe variables it doesn't change the
  model.

- Class: text
  Output: Congrats! You've concluded this first lesson on multivariable linear models.

- Class: mult_question
  Output: "Would you like to receive credit for completing this course on
    Coursera.org?"
  CorrectAnswer: NULL
  AnswerChoices: Yes;No
  AnswerTests: coursera_on_demand()
  Hint: ""

```