Quiz 5 solutions and explanations

Quiz 5 Solutions & Explanations

Everyday Excel, Part 2

Charlie Nuttelman

Hello there! This document is meant to provide clear explanations for the Quiz 5 questions (not the in-video quizzes since they have explanations already). I do NOT provide feedback during the quiz (like I do for the screencasts) because a learner could just guess, obtain the correct answers, then put them back into the quiz and get 100%!

This document is purely for you to learn more and to correct your misconceptions about the material. If you view this document soon after you take the quiz to see why you missed a certain question, it will serve as a great learning tool!

PLEASE DO NOT SHARE THIS DOCUMENT WITH ANYONE! Using this document to complete Quiz 5 is a violation of Coursera's Honor Code (a.k.a. cheating).

NOTE that the order of the answers on Coursera are random and likely different from the order shown here (in general but not always, I like to start with the correct answers followed by the incorrect ones).

Question 1:

Which of the following statements below is true regarding linear regression analysis in Excel? Select all that apply.

A. The **Trendline** tool can be used to create regression equations for polynomial models.

TRUE. Yes, the **Trendline** tool can be used to create polynomial models, although its usefulness is limited by the precision of the coefficients displayed on the chart.

B. The **Trendline** tool can be used to create regression equations for complex, custom-defined models.

FALSE. No, the **Trendline** tool cannot do this.

C. The **Regression** tool can be used to create regression equations for complex, custom-defined models.

TRUE. Yes, this is correct and makes the **Regression** tool a great tool for model building.

D. The **Trendline** tool outputs important statistical information, like confidence intervals and adjusted correlation coefficient.

FALSE. No, the **Trendline** tool does NOT do all of this. Rather, the **Regression** tool does.

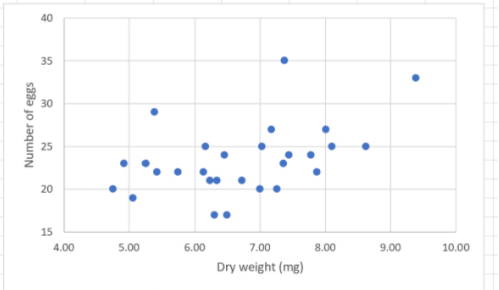
E. The **Regression** tool outputs important statistical information, like confidence intervals and adjusted correlation coefficient.

TRUE. Yes, one really nice thing about the **Regression** tool is that it gives a lot of information related to the goodness of fit, the correlation coefficient, confidence intervals, and whether or not terms in the model are significant (P-values).

Question 2:

The following data presents the number of eggs vs. dry weight in the amphipod *Platorchestia platensis*.
The independent variable is **Dry weight** and the **Number of eggs** is the dependent variable. Which of the following options show a *single* Excel formula that we could use in cell **F4** to estimate the **Number of eggs** as a function of the **Dry weight** in cell **F3**? Cell F3 has been named "**weight**". Select all that apply.

	A	B	C	D	E	F	G	H	I	J
1	dry wt (g)	# of eggs								
2	5.38	29								
3	7.36	23			Dry weight:	6.2 mg				
4	6.13	22			Number of eggs:					
5	4.75	20								
6	8.10	25								
7	8.62	25								
8	6.30	17								
9	7.44	24								
10	7.26	20								
11	7.17	27								
12	7.78	24								
13	6.23	21								
14	5.42	22								
15	7.87	22								
16	5.25	23								
17	7.37	35								
18	8.01	27								
19	4.92	23								
20	7.03	25								
21	6.45	24								
22	5.06	19								
23	6.72	21								
24	7.00	20								
25	9.39	33								
26	6.49	17								
27	6.34	21								
28	6.16	25								
29	5.74	22								



A. **=FORECAST.LINEAR(weight,B2:B29,A2:A29)**

Correct! We can use the **FORECAST.LINEAR** function where the first argument is the x-value (weight) that we wish to predict the y-value for, the second argument is the range of y-values of our data, and the third argument is the range of x-values of our data.

B. **=SLOPE(B2:B29,A2:A29)*weight+INTERCEPT(B2:B29,A2:A29)**

Correct! We can use the slope and intercept, calculated from the **SLOPE** and **INTERCEPT** functions, in a simple equation for a line (y = mx + b where m is the slope and b is the intercept).

C. **=FORECAST.LINEAR(weight,A2:A29,B2:B29)**

Incorrect. We can certainly use the **FORECAST.LINEAR** function, but remember that the correct way to use this function is **=FORECAST.LINEAR(x, known_y's, known_x's)**. Here we have put our x values (# of eggs) as the second argument and our y values (dry wt) as the third argument, so therefore it will not work.

D. **=SLOPE(B2:B29,A2:A29)+INTERCEPT(B2:B29,A2:A29)*weight**

Incorrect. We can use the slope and intercept, calculated from the **SLOPE** and **INTERCEPT** functions, in a simple equation for a line (y = mx + b where m is the slope and b is the intercept). However, here we have mixed around the slope and intercept.

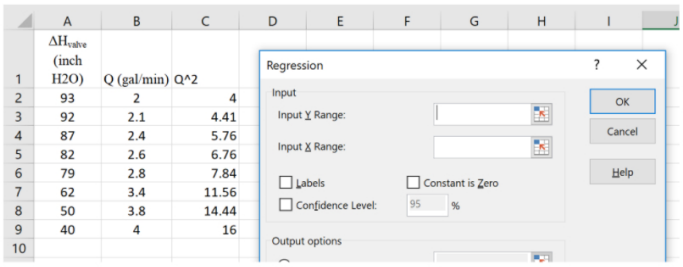
E. **=SLOPE(B2:B29,A2:A29)*weight+FORECAST.LINEAR(F3,B2:B29,A2:A29)**

Incorrect. This will not work unless we use the **INTERCEPT** function for the second term (instead of the **FORECAST.LINEAR** function).

Question 3:

We wish to create the following model relating **ΔH_valve** to flow rate (**Q**). **ΔH_(dead head)** is the intercept.

$$\Delta H_{valve} = \Delta H_{dead\ head} - \alpha Q^2$$



We have set up the spreadsheet above. To perform the regression, what will we place in the **Input X Range** of the **Regression** box?

Correct answer: **$C$2:$C$9**

Explanation: The only regressor function in our model is Q^2 (there is no term with just Q). Therefore, all we need to input into the **Regression** tool is the column for Q^2 (**$C$2:$C$9**).

Question 4:

In which of the following types of regression do we specifically limit the output (dependent variable) to between two values. Note that the independent variable does not have to be limited in range.

Correct answer: **Logistic regression**

Explanation: In logistic regression, we can constrain the output (y-value) to within two values. In the screencast "Introduction to nonlinear regression", I worked through an example of where the output was constrained between 0 and 1. This is typical for the modeling of decision-making processes, like whether a buyer will purchase your product, based on various inputs. The output is a probability between 0 (0%) and 1 (100%) - those who fall somewhere above 50% are those who you might want to advertise to.

Question 5:

The table below presents data on the maximum ice thickness in mm (y), average number of days per year of ice cover (x1), average number of days the bottom temperature is lower than 8°C (x2), and the average snow depth in mm (x3) for 13 lakes in Minnesota. This data set can be found in the attached file entitled "Ice_Thickness.xlsx".

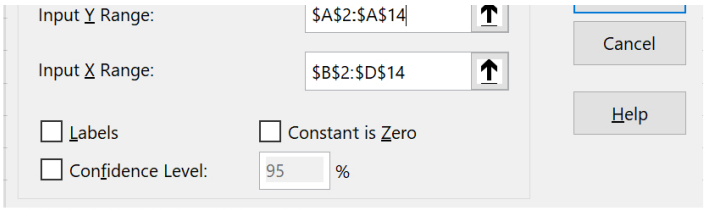
	A	B	C	D
1	y	x_1	x_2	x_3
2	730	152	198	91
3	760	173	201	81
4	850	166	202	69
5	840	161	202	72
6	720	152	198	91
7	730	153	205	91
8	840	166	204	70
9	730	157	204	90
10	650	136	172	47
11	850	142	218	59
12	740	151	207	88
13	720	145	209	60
14	710	147	190	63
15				

Create the following multilinear regression model and use your model to predict the y-value (maximum ice thickness) when x1 = 150, x2 = 200, and x3 = 75. Enter your answer rounded to the nearest ones place (i.e., format = XXX).

Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

Correct answer: **742**

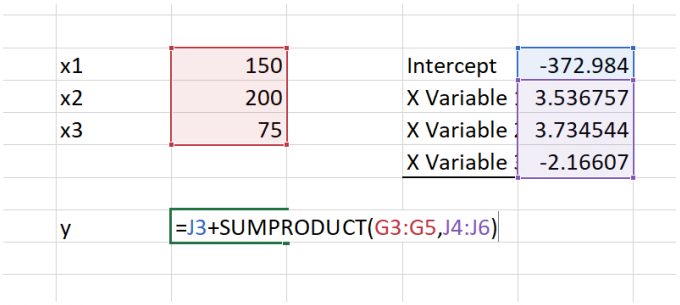
Explanation: We can perform regression using the **Regression** tool in Excel:



And the coefficients from the **Regression** output are:

	Coefficients
Intercept	-372.9838571
X Variable 1	3.536756838
X Variable 2	3.734543655
X Variable 3	-2.166069166

To calculate the response (maximum ice thickness) when x1 = 150, x2 = 200, and x3 = 75, we can just plug those into the model using the coefficients above. This is easily done in Excel using the **SUMPRODUCT** function:



And the result