Program Evaluation Toolkit

Chapter 2

1 — 2 — 3

Data Analysis Examples

Module 7 Data Analysis

Data Analysis Examples

- Two relatively straightforward examples of data analysis, presented in relation to the fictitious AMMP! evaluation.
- The examples are aligned to specific AMMP! evaluation questions.
  - Example 1 – Quantitative example
  - Example 2 – Qualitative example

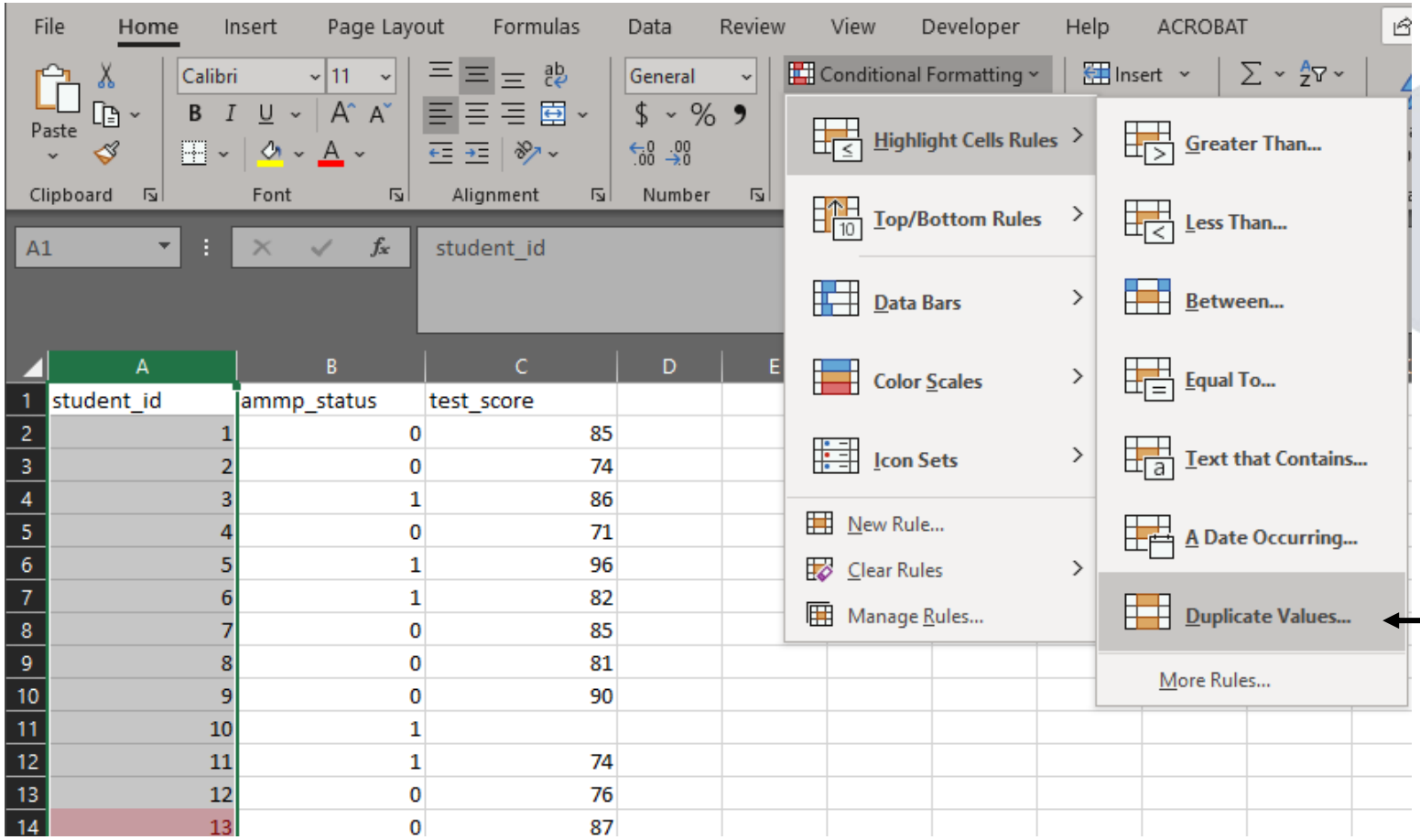
Data Analysis: Quantitative Example

- Consider the following AMMP! evaluation question:
  - How do AMMP! participants' scores on high school math placement tests compare to nonparticipants' scores?
- This question involves examining test scores; therefore, a quantitative method is more suitable to addressing the question.

Data Cleaning

- Do the data contain duplicates?
- Do the data contain missing values?
- Are there entry errors?
- Are there outliers?

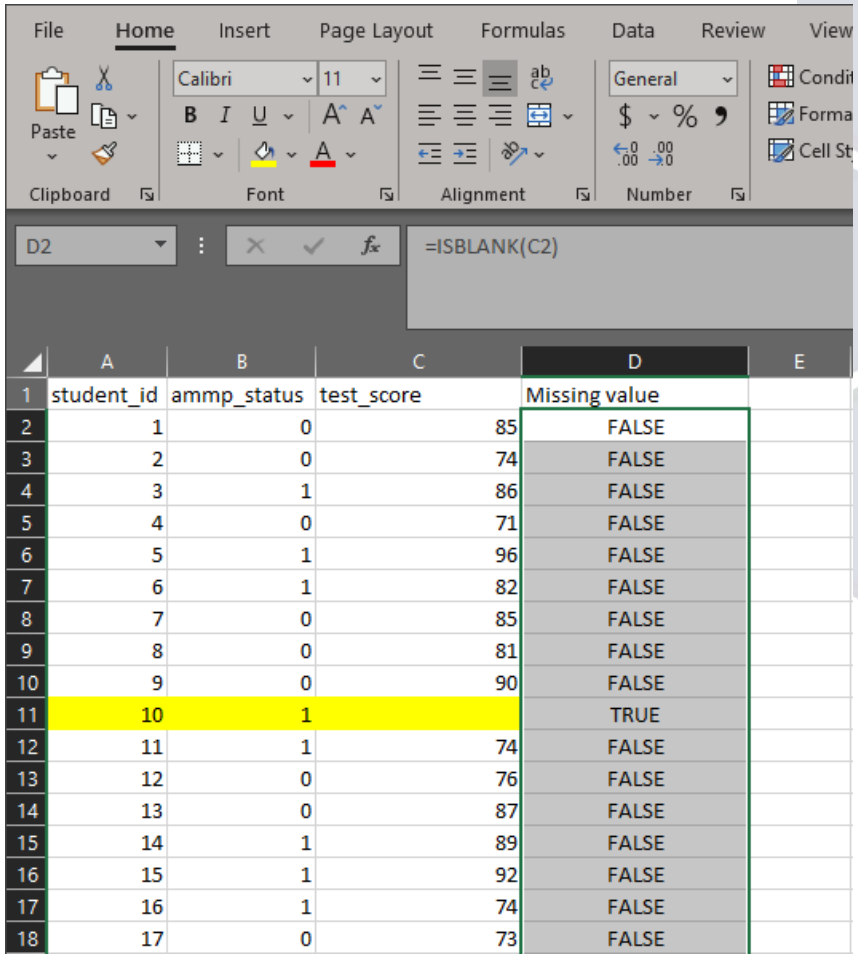
Removing Duplicates



	A	B	C	D	E
1	student_id	ammp_status	test_score		
2	1	0	85		
3	2	0	74		
4	3	1	86		
5	4	0	71		
6	5	1	96		
7	6	1	82		
8	7	0	85		
9	8	0	81		
10	9	0	90		
11	10	1			
12	11	1	74		
13	12	0	76		
14	13	0	87		

Examining Missing Values

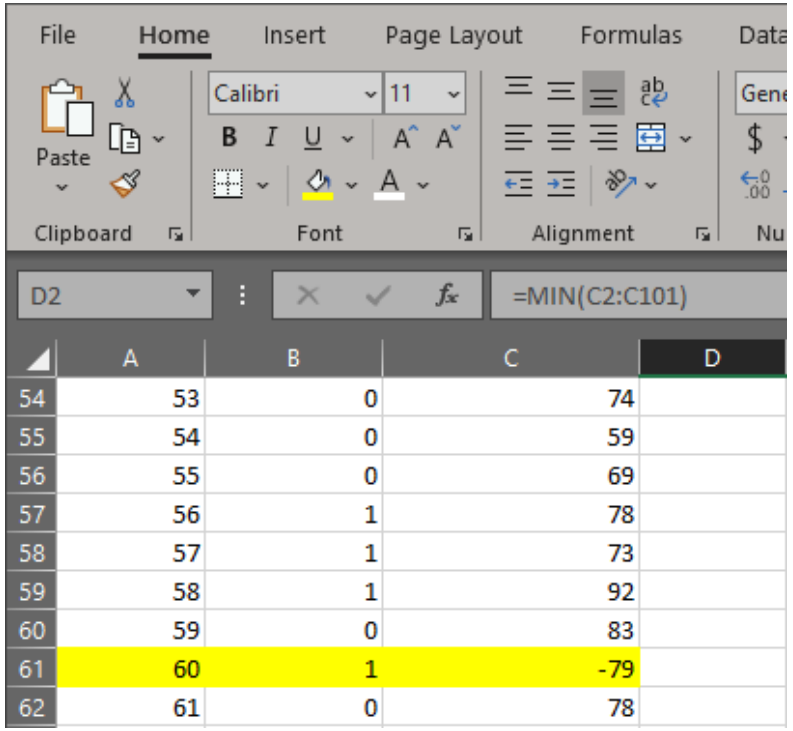
- =ISBLANK(C2)



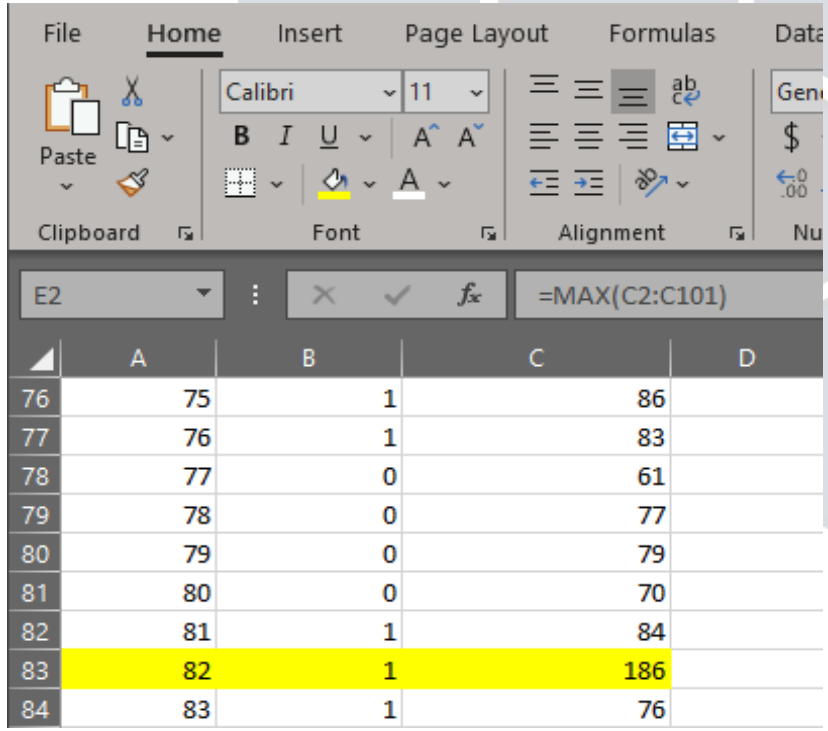
	A	B	C	D	E
1	student_id	ammp_status	test_score	Missing value	
2	1	0	85	FALSE	
3	2	0	74	FALSE	
4	3	1	86	FALSE	
5	4	0	71	FALSE	
6	5	1	96	FALSE	
7	6	1	82	FALSE	
8	7	0	85	FALSE	
9	8	0	81	FALSE	
10	9	0	90	FALSE	
11	10	1		TRUE	
12	11	1	74	FALSE	
13	12	0	76	FALSE	
14	13	0	87	FALSE	
15	14	1	89	FALSE	
16	15	1	92	FALSE	
17	16	1	74	FALSE	
18	17	0	73	FALSE	

Cleaning Data Entry Errors

- =MIN(C2:C101)
- =MAX(C2:C101)



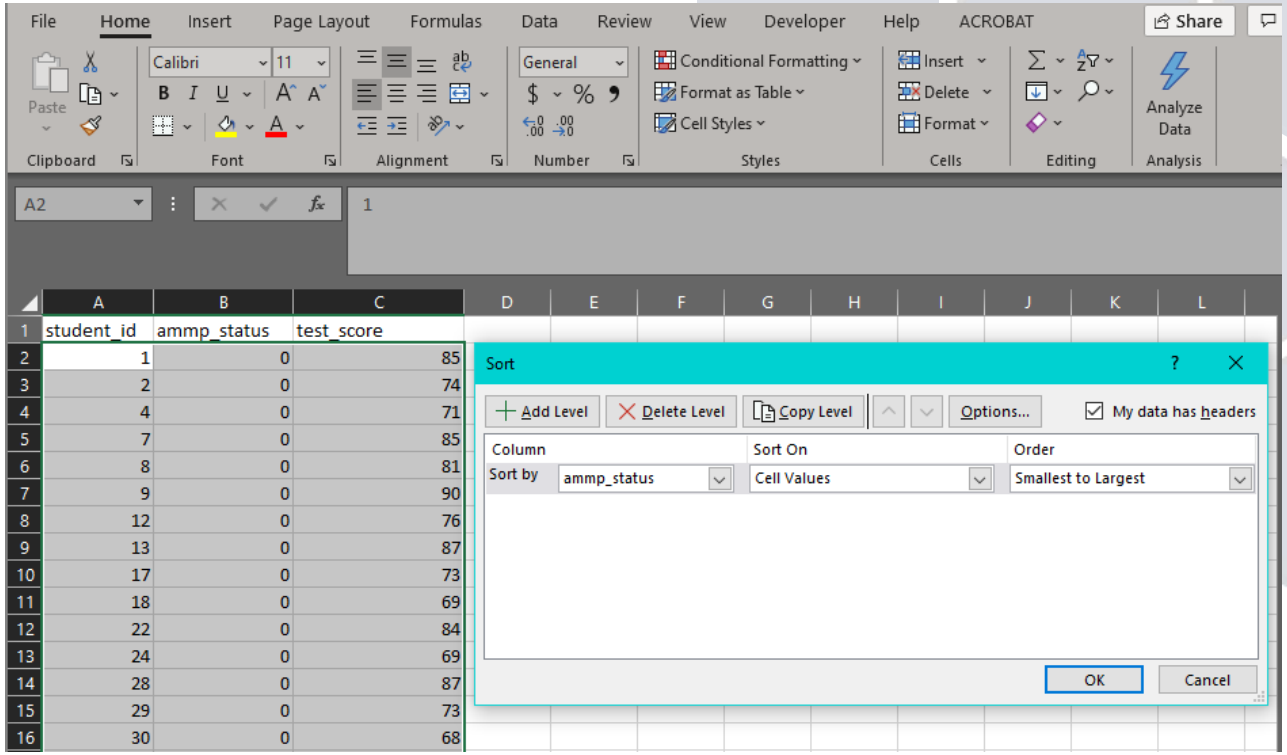
	A	B	C	D
54	53	0	74	
55	54	0	59	
56	55	0	69	
57	56	1	78	
58	57	1	73	
59	58	1	92	
60	59	0	83	
61	60	1	-79	
62	61	0	78	



	A	B	C	D
76	75	1	86	
77	76	1	83	
78	77	0	61	
79	78	0	77	
80	79	0	79	
81	80	0	70	
82	81	1	84	
83	82	1	186	
84	83	1	76	

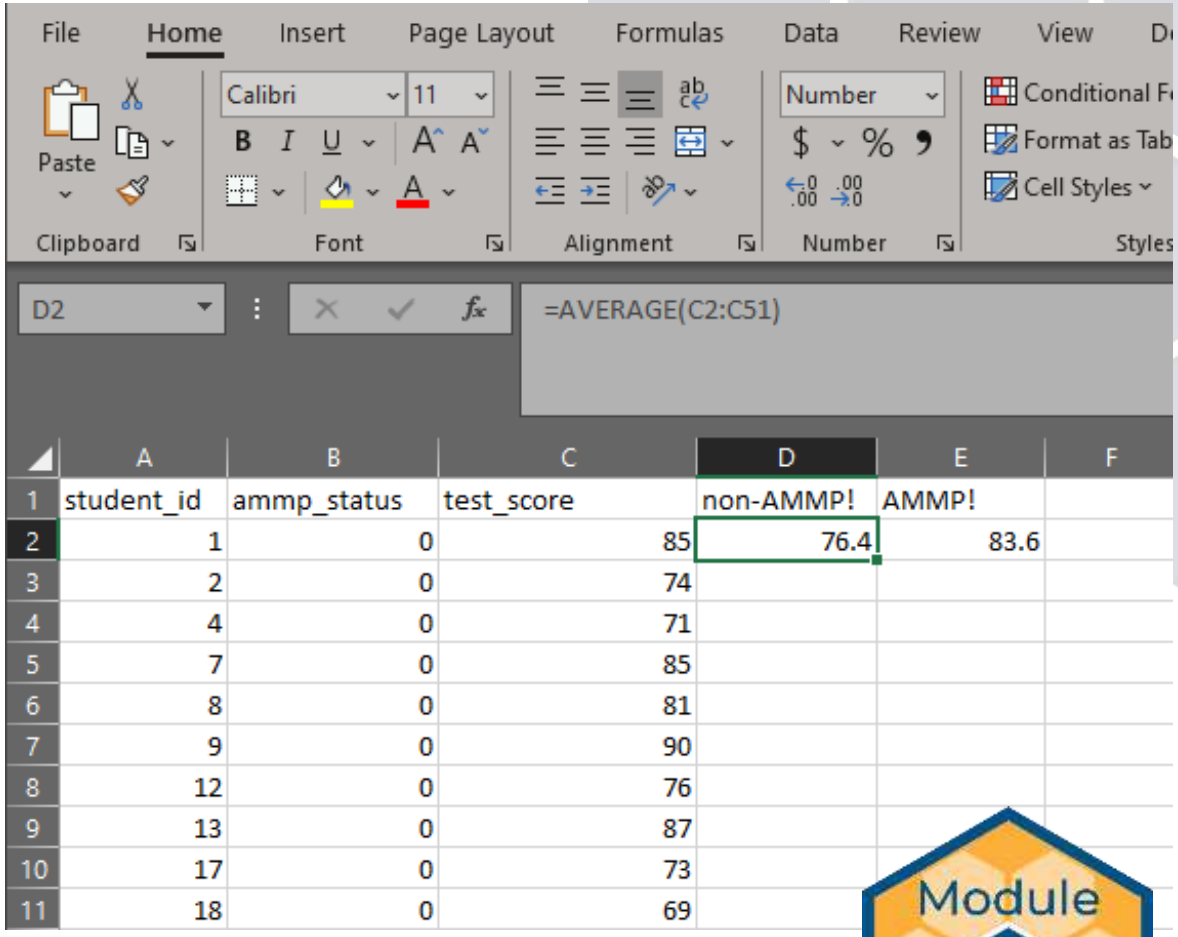
Descriptive Statistics

- Sort data for comparing AMMP! students and non-AMMP! students



Descriptive Statistics: Mean

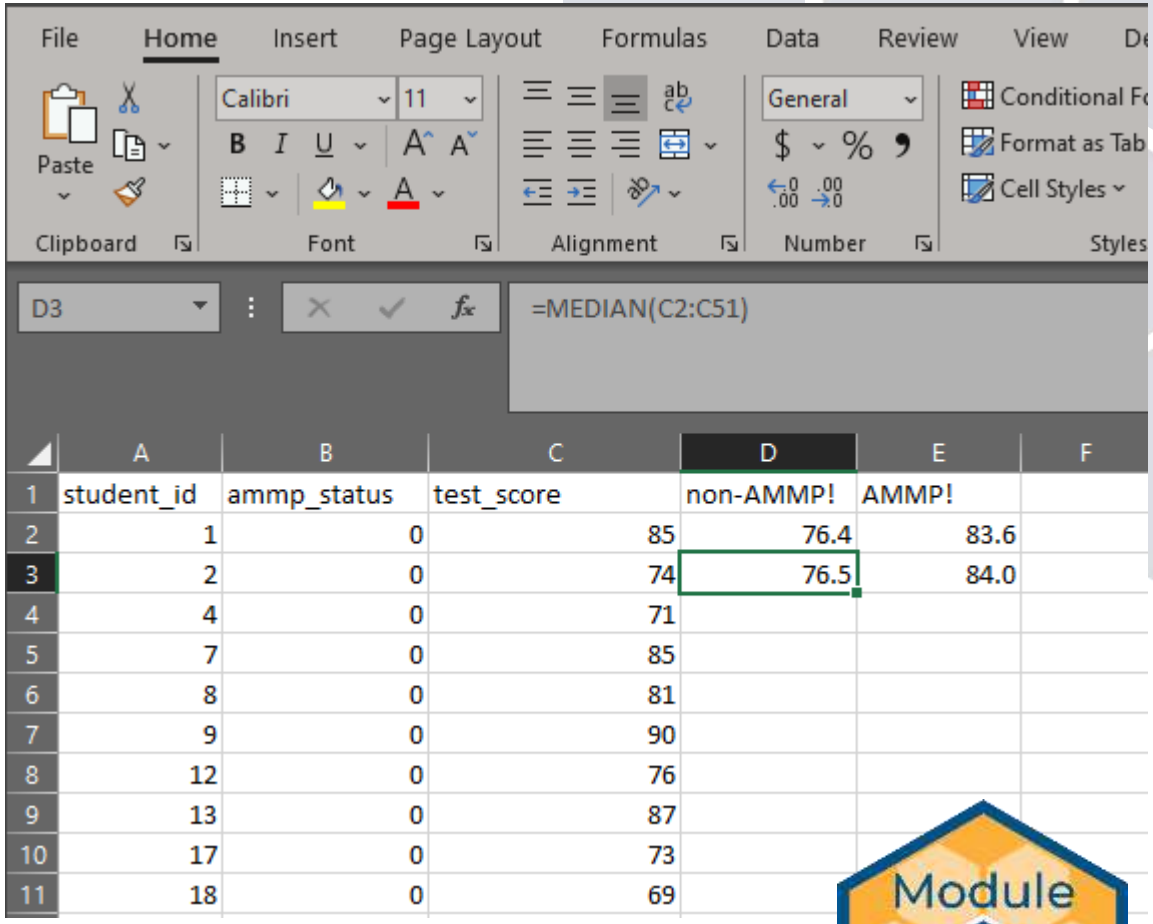
- =AVERAGE(C2:C51)
- =AVERAGE(C52:C101)



	A	B	C	D	E	F
1	student_id	ammp_status	test_score	non-AMMP!	AMMP!	
2	1	0	85	76.4	83.6	
3	2	0	74			
4	4	0	71			
5	7	0	85			
6	8	0	81			
7	9	0	90			
8	12	0	76			
9	13	0	87			
10	17	0	73			
11	18	0	69			

Descriptive Statistics: Median

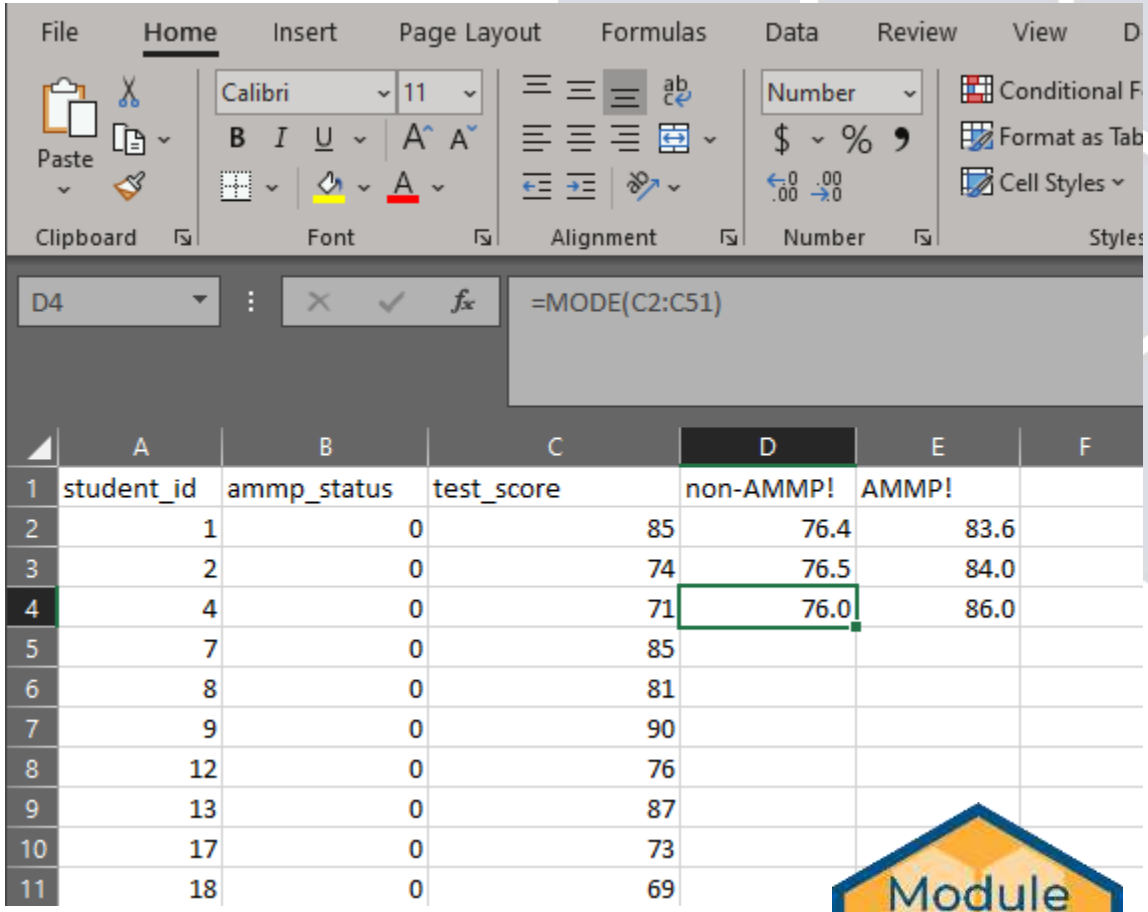
- =MEDIAN(C2:C51)
- =MEDIAN(C52:C101)



	A	B	C	D	E	F
1	student_id	ammp_status	test_score	non-AMMP!	AMMP!	
2	1	0	85	76.4	83.6	
3	2	0	74	76.5	84.0	
4	4	0	71			
5	7	0	85			
6	8	0	81			
7	9	0	90			
8	12	0	76			
9	13	0	87			
10	17	0	73			
11	18	0	69			

Descriptive Statistics: Mode

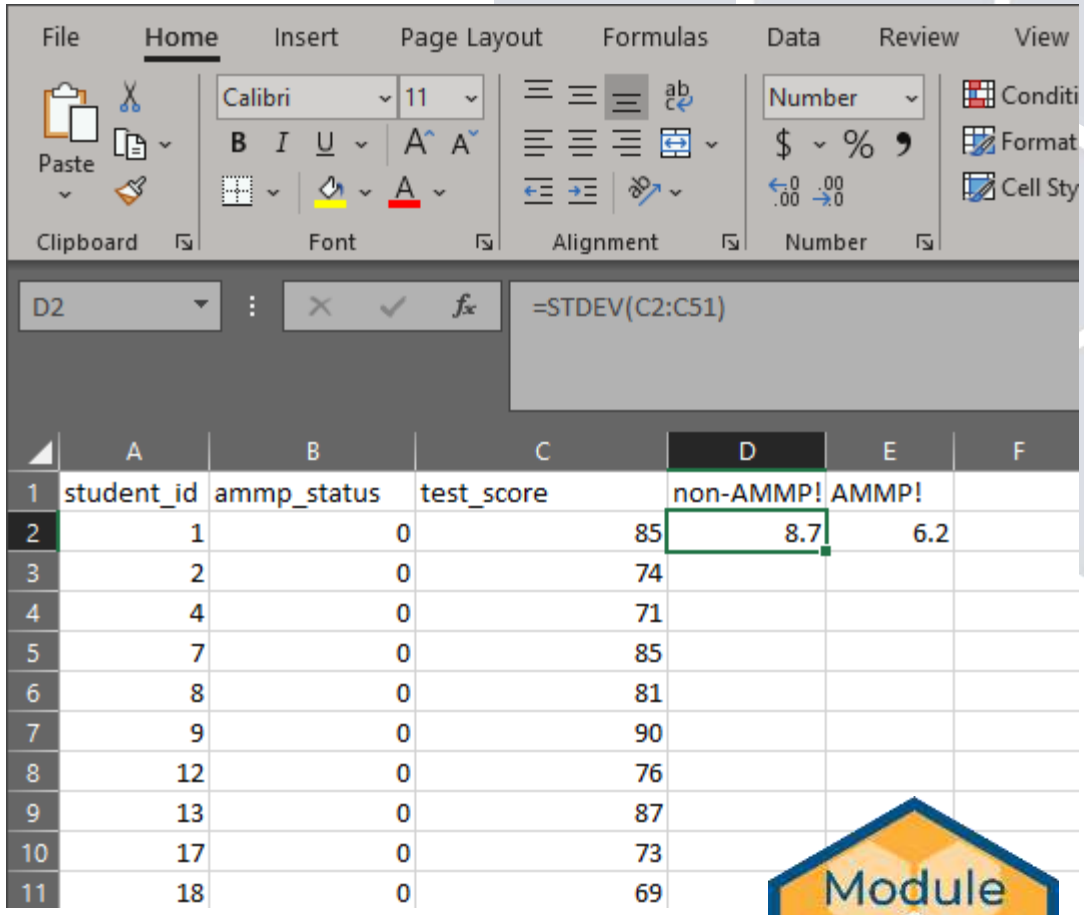
- =MODE(C2:C51)
- =MODE(C52:C101)



	A	B	C	D	E	F
1	student_id	ammp_status	test_score	non-AMMP!	AMMP!	
2	1	0	85	76.4	83.6	
3	2	0	74	76.5	84.0	
4	4	0	71	76.0	86.0	
5	7	0	85			
6	8	0	81			
7	9	0	90			
8	12	0	76			
9	13	0	87			
10	17	0	73			
11	18	0	69			

Descriptive Statistics: Standard Deviation

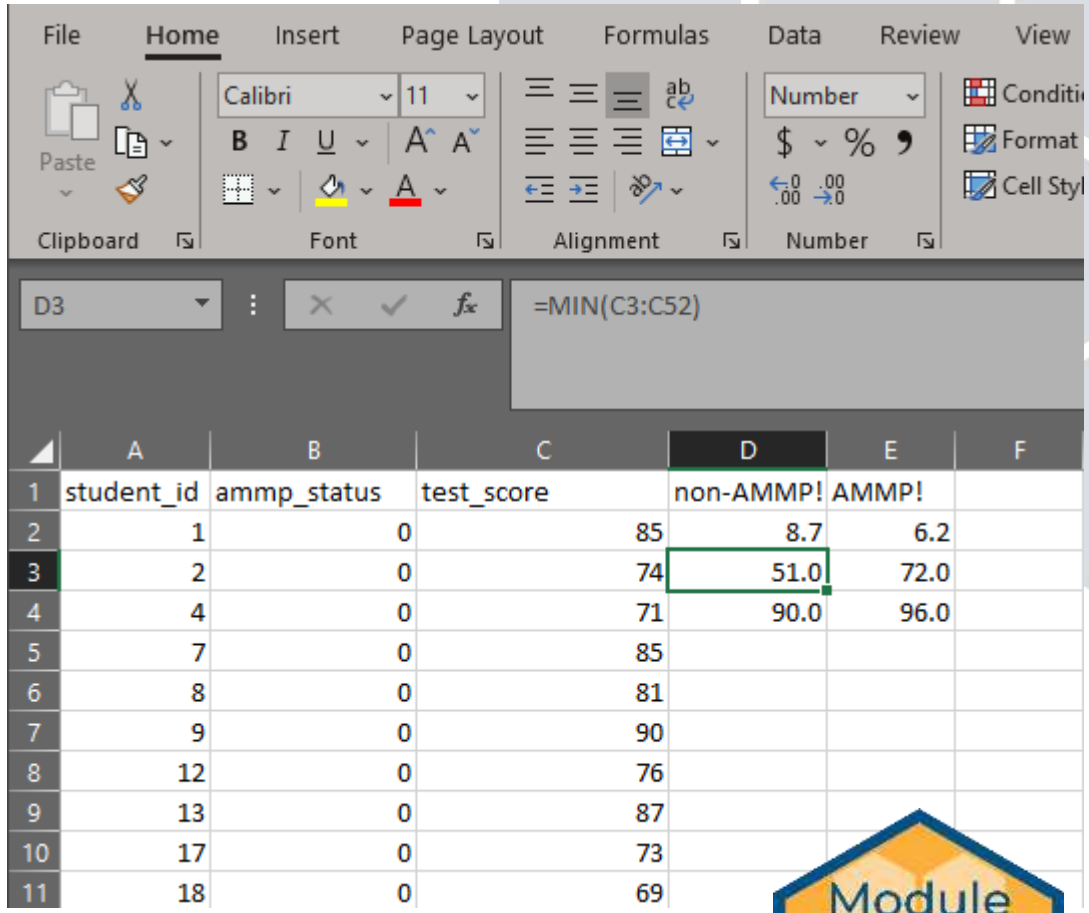
- =STDEV(C2:C51)
- =STDEV(C52:C101)



	A	B	C	D	E	F
1	student_id	ammp_status	test_score	non-AMMP!	AMMP!	
2	1	0	85	8.7	6.2	
3	2	0	74			
4	4	0	71			
5	7	0	85			
6	8	0	81			
7	9	0	90			
8	12	0	76			
9	13	0	87			
10	17	0	73			
11	18	0	69			

Descriptive Statistics: Minimum, Maximum, and Range

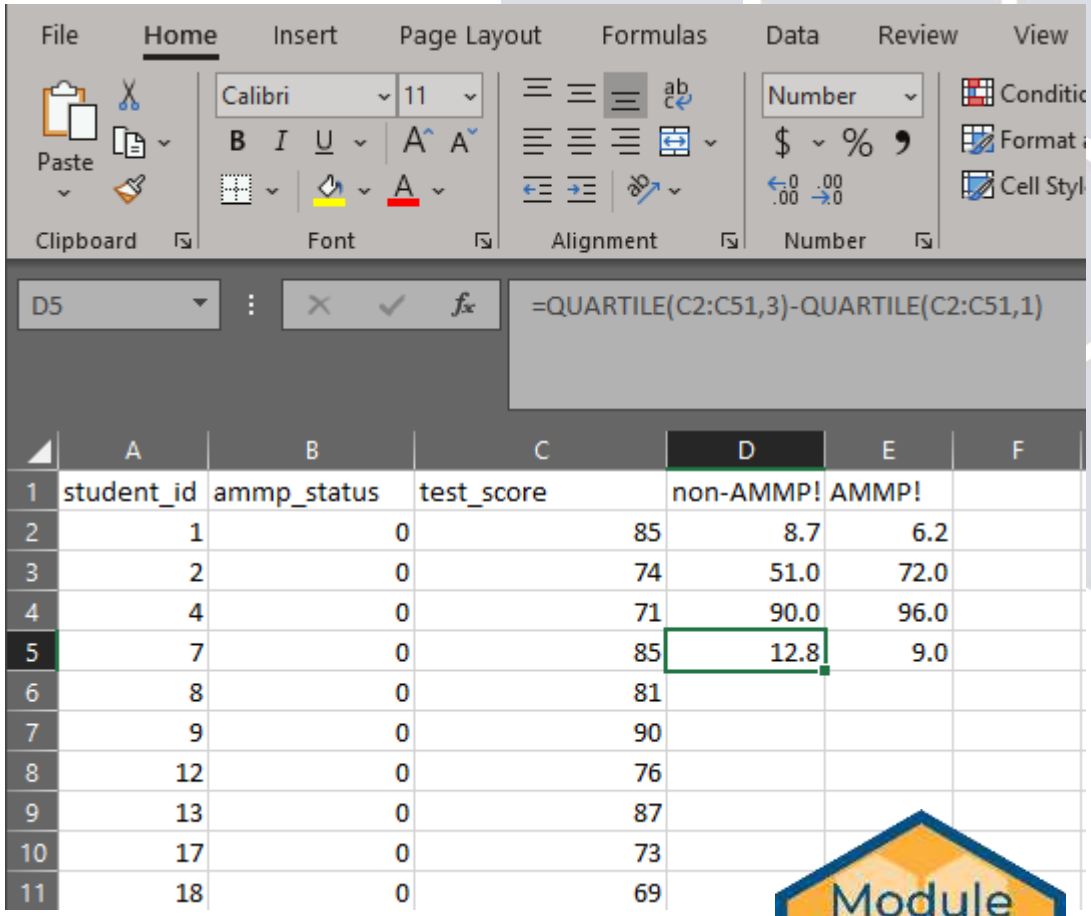
- =MIN(C2:C51)
- =MAX(C2:C51)
- =MIN(C52:C101)
- =MAX(C52:C101)



	A	B	C	D	E	F
1	student_id	ammp_status	test_score	non-AMMP!	AMMP!	
2	1	0	85	8.7	6.2	
3	2	0	74	51.0	72.0	
4	4	0	71	90.0	96.0	
5	7	0	85			
6	8	0	81			
7	9	0	90			
8	12	0	76			
9	13	0	87			
10	17	0	73			
11	18	0	69			

Descriptive Statistics: Interquartile Range

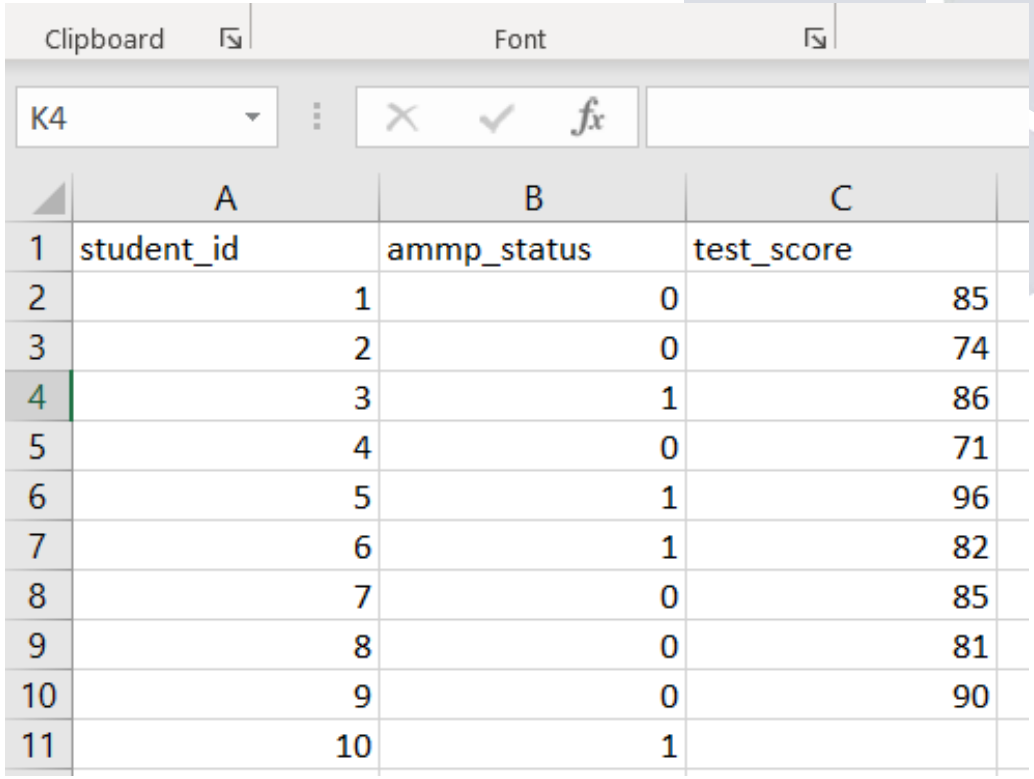
- =QUARTILE(C2:C51,3)-QUARTILE(C2:C51,1)
- =QUARTILE(C52:C101,3)-QUARTILE(C52:C101,1)



	A	B	C	D	E	F
1	student_id	ammp_status	test_score	non-AMMP!	AMMP!	
2	1	0	85	8.7	6.2	
3	2	0	74	51.0	72.0	
4	4	0	71	90.0	96.0	
5	7	0	85	12.8	9.0	
6	8	0	81			
7	9	0	90			
8	12	0	76			
9	13	0	87			
10	17	0	73			
11	18	0	69			

Descriptive Statistics Using the Toolkit Calculator

- Central tendency:
  - Mean
  - Median
  - Mode
- Variation:
  - Standard deviation
  - Minimum
  - Maximum
  - Interquartile range



	A	B	C
1	student_id	ammp_status	test_score
2	1	0	85
3	2	0	74
4	3	1	86
5	4	0	71
6	5	1	96
7	6	1	82
8	7	0	85
9	8	0	81
10	9	0	90
11	10	1	

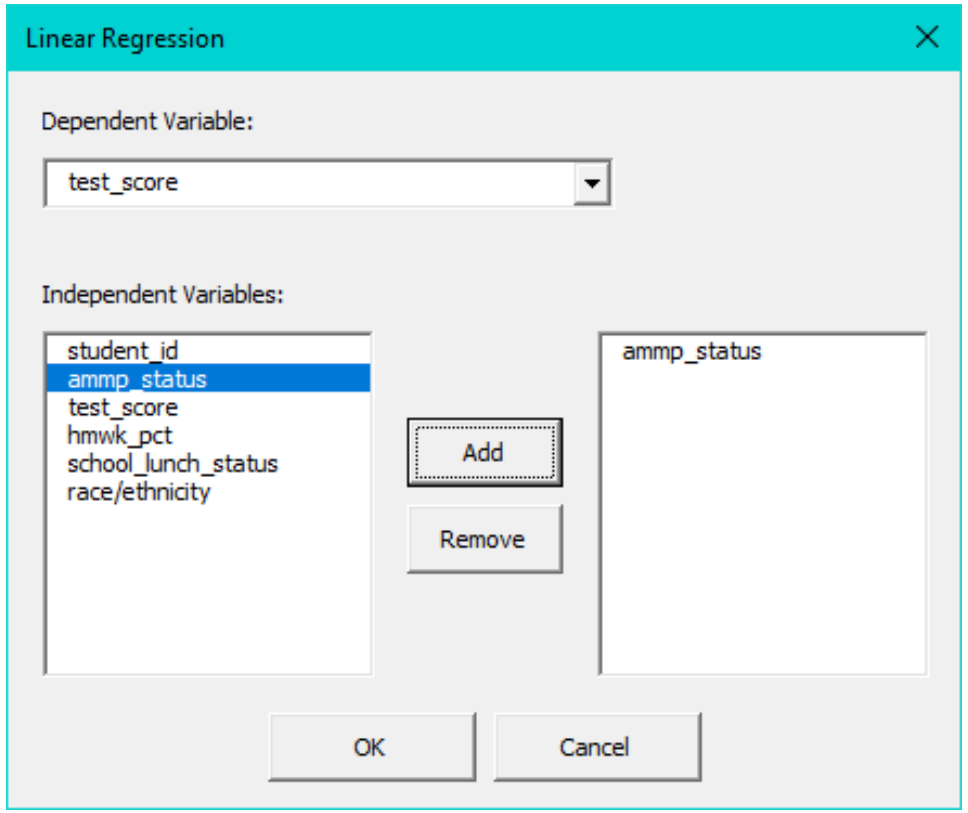
Additional Resources

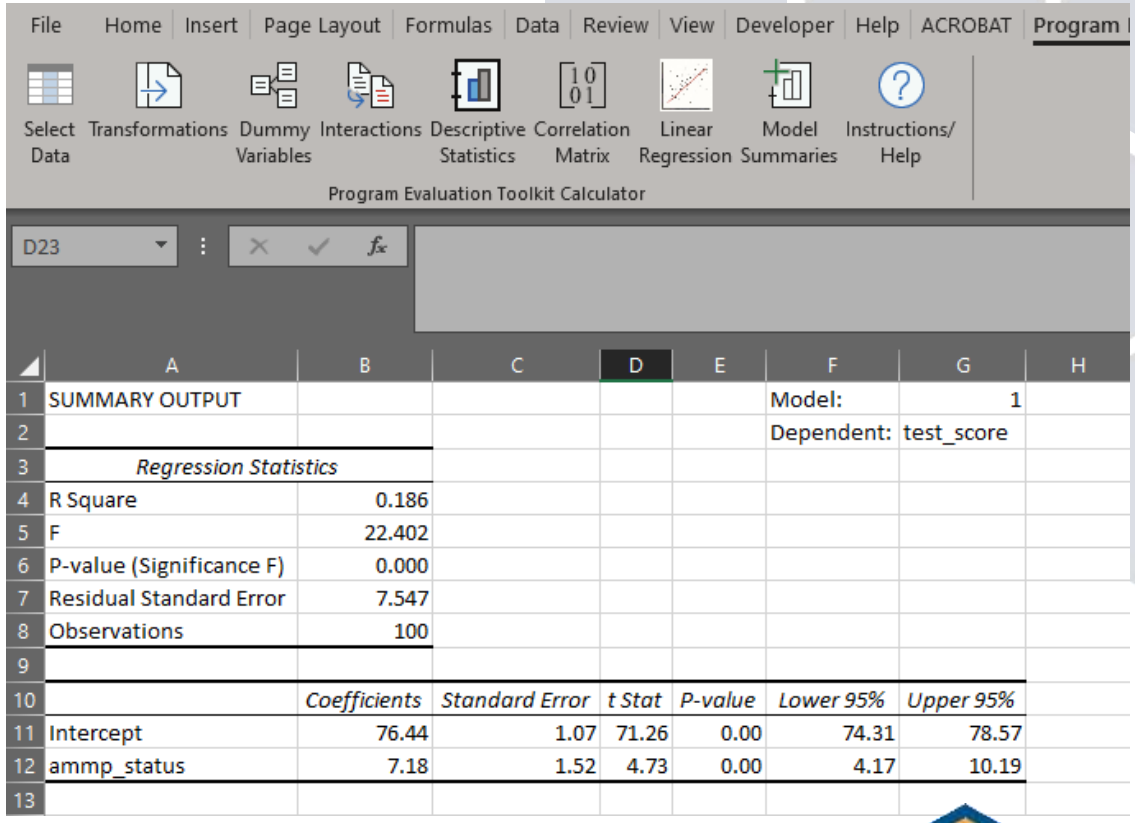
- *Program Evaluation Toolkit Calculator*
- *Program Evaluation Toolkit Calculator: User's Guide*

Inferential Statistics

- Dependent variable: high school math placement test score
- Independent variable: AMMP! participation
- Possible covariates: homework completion percentage, eligibility for the National School Lunch Program, sex, and race/ethnicity
- Possible confounds?
  - Adjust for confounds using multiple regression.
  - If there are unmeasured confounds, results may not be trustworthy.
  - Only adjust for variables that were measured before student exposure to the program (in this case, AMMP!).

Linear Regression Analysis





	A	B	C	D	E	F	G	H
1	SUMMARY OUTPUT					Model:	1	
2						Dependent:	test_score	
3	*Regression Statistics*							
4	R Square	0.186						
5	F	22.402						
6	P-value (Significance F)	0.000						
7	Residual Standard Error	7.547						
8	Observations	100						
9								
10		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	
11	Intercept	76.44	1.07	71.26	0.00	74.31	78.57	
12	ammp_status	7.18	1.52	4.73	0.00	4.17	10.19	
13								

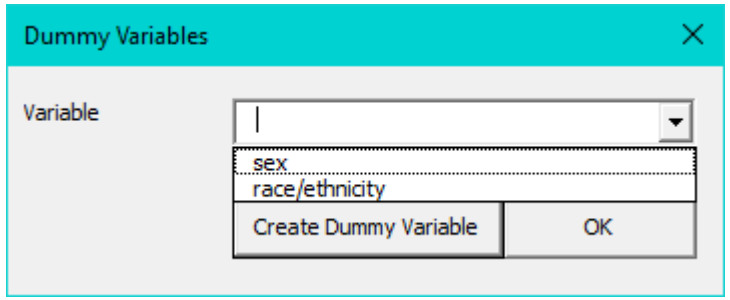
Additional Resources

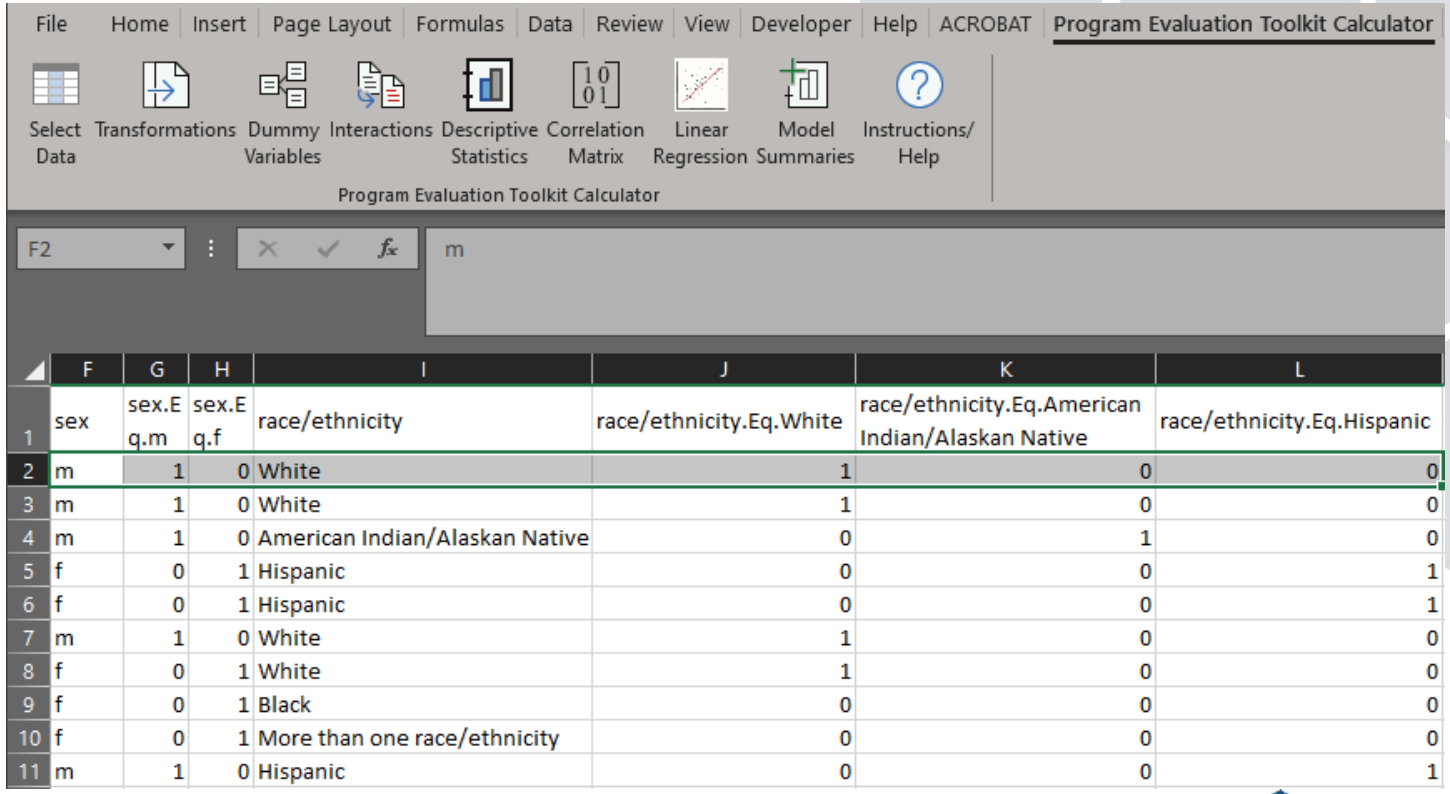
- *Program Evaluation Toolkit Calculator*
- *Program Evaluation Toolkit Calculator: User's Guide*

Interpretation: Regression Coefficients[1]

- The intercept coefficient is the predicted value of the dependent variable when the independent variables are all 0.
  - For example, the average test score for non-AMMP! students.
- The coefficients for the independent variables indicate the predicted change in the dependent variable when an independent variable increases by one unit.
  - For example, the "ammp_status" coefficient is the difference between the average test score of AMMP! students and that of non-AMMP! students.
  - On average, AMMP! students scored 7.18 points higher than non-AMMP! students scored on the math placement exam.

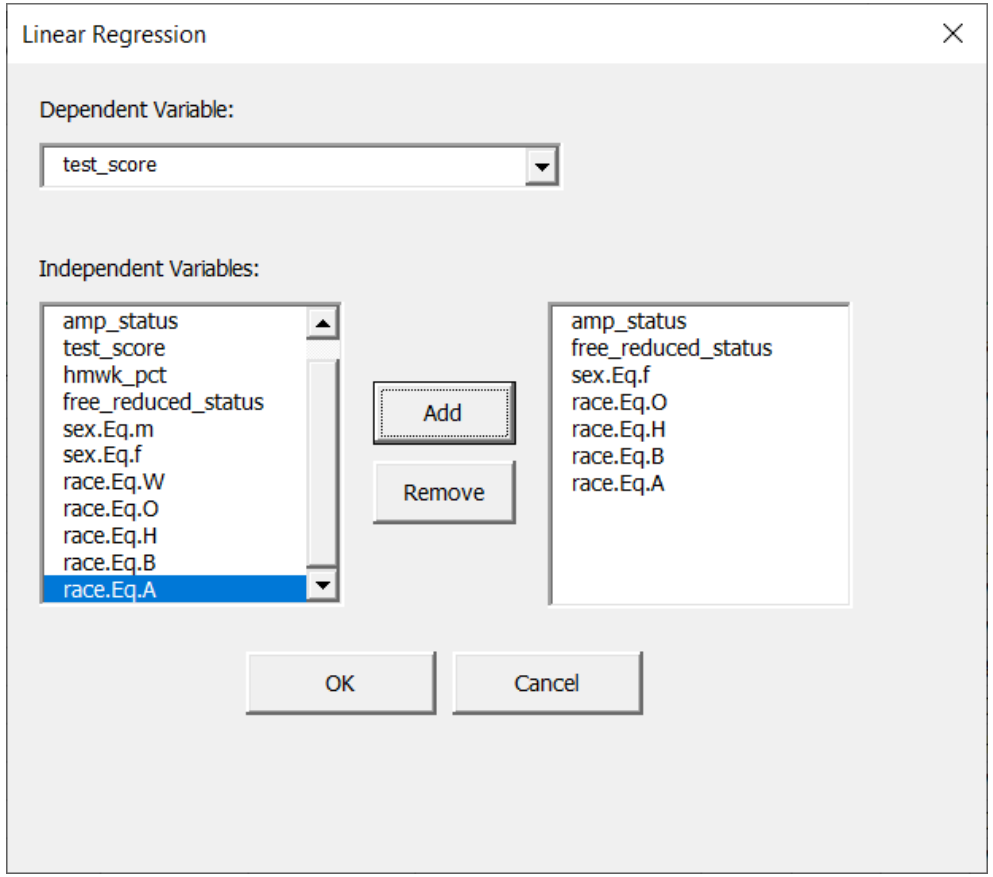
Multiple Regression Analysis[1]

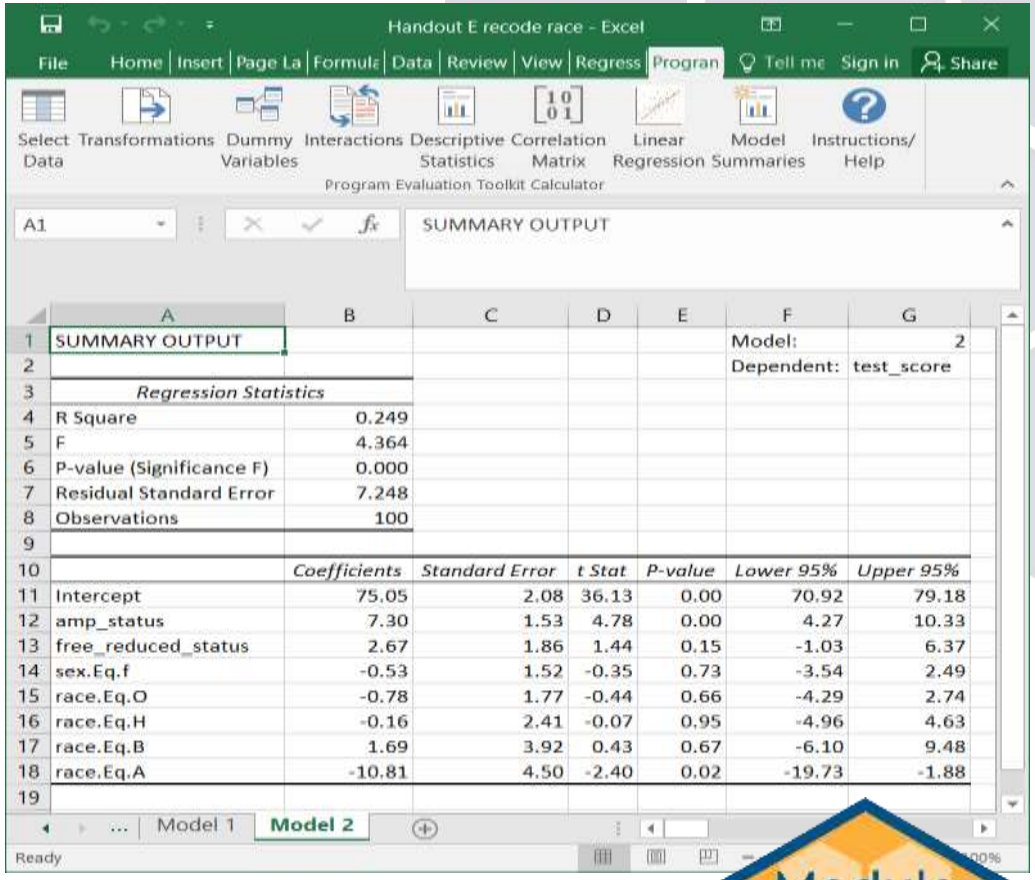




	F	G	H	I	J	K	L
1	sex	sex.Eq.m	sex.Eq.f	race/ethnicity	race/ethnicity.Eq.White	race/ethnicity.Eq.American Indian/Alaskan Native	race/ethnicity.Eq.Hispanic
2	m	1	0	White	1	0	0
3	m	1	0	White	1	0	0
4	m	1	0	American Indian/Alaskan Native	0	1	0
5	f	0	1	Hispanic	0	0	1
6	f	0	1	Hispanic	0	0	1
7	m	1	0	White	1	0	0
8	f	0	1	White	1	0	0
9	f	0	1	Black	0	0	0
10	f	0	1	More than one race/ethnicity	0	0	0
11	m	1	0	Hispanic	0	0	1

Multiple Regression Analysis: Results





	A	B	C	D	E	F	G
1	SUMMARY OUTPUT					Model:	2
2						Dependent:	test_score
3	*Regression Statistics*						
4	R Square	0.249					
5	F	4.364					
6	P-value (Significance F)	0.000					
7	Residual Standard Error	7.248					
8	Observations	100					
9							
10		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
11	Intercept	75.05	2.08	36.13	0.00	70.92	79.18
12	amp_status	7.30	1.53	4.78	0.00	4.27	10.33
13	free_reduced_status	2.67	1.86	1.44	0.15	-1.03	6.37
14	sex.Eq.f	-0.53	1.52	-0.35	0.73	-3.54	2.49
15	race.Eq.O	-0.78	1.77	-0.44	0.66	-4.29	2.74
16	race.Eq.H	-0.16	2.41	-0.07	0.95	-4.96	4.63
17	race.Eq.B	1.69	3.92	0.43	0.67	-6.10	9.48
18	race.Eq.A	-10.81	4.50	-2.40	0.02	-19.73	-1.88
19							

Data Analysis: Qualitative Example[2]

- Consider the following AMMP! evaluation question:
  - What barriers exist that prevent students from completing homework?
- This question likely involves complex subjective judgments for which complete quantitative data do not exist or cannot be easily collected.
  - It is more appropriate to address the question with qualitative data.

Qualitative Analysis Activity

- Interview question for students:
  - What are some things that might keep you from completing homework?
- Interview question for teachers:
  - What do you think are barriers that prevent students from completing homework?
- Identify common themes in the mock responses to the questions.

Common Themes

- Lack of time
- Unsupervised time after school
- Difficulty of homework
- Not seeing the value of homework

Additional Resources[3]

- *Statistical Theory for the* RCT-YES *Software: Design-Based Causal Inference for RCTs* provides support in conducting rigorously designed evaluations.
- Evidence to Insights (e2i) Coach is a free online tool that can help you generate evidence related to programs implemented in your specific context.

Program Evaluation Toolkit

Module 7 Data Analysis

Chapter 2 Complete

1 — 2 — 3

Recommended next: Chapter 3 – From Results to Interpretation to Recommendations

Thank You

Please visit our website and follow us on Twitter
for information about our events, priorities, and research alliances,
and for access to our many free resources.

ies.ed.gov/ncee/edlabs/regions/central/index.asp
@RELCentral
or contact us at
RELCentral@marzanoresearch.com

This presentation was prepared under Contract ED-IES-17-C-0005 by Regional Educational Laboratory Central, administered by Marzano Research. The content does not necessarily reflect the views or policies of IES or the U.S. of Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

1. Glass, G. V., & Hopkins, K. D. (1995). *Statistical methods in education and psychology* (3rd ed.). Allyn & Bacon.
2. Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage.
3. Schochet, P. Z. (2016). *Statistical theory for the* RCT-YES *software: Design-based causal inference for RCTs* (2nd ed.; NCEE 2015–4011). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development. https://ies.ed.gov/ncee/pubs/20154011/
4. Mathematica. (n.d.). Evidence to Insights Coach. https://e2icoach.org/