Two-Group Hypothesis Tests Using Excel T.TEST Function

by
Milo Schield
Member: International Statistical Institute
US Rep: International Statistical Literacy Project
Director, W. M. Keck Statistical Literacy Project

Slides and audio at: www.StatLit.org/pdf/TTEST-Function-Excel-2008-6up.pdf
Audio/TTEST-Function-Excel-2008.mp3

Excel T.TEST Function

Purpose: Calculate likelihood (p-value) of getting the observed difference in two sample means (or more extreme) by chance in random samples – assuming there is no difference in the two population means (the Null Hypothesis).

Note: TTEST function was available in Excel 2003.

Four Inputs:

1) Array or range of 1st sample. 2) Array or range of 2nd sample.
3) Tails: 1 (Excel matches Alternate with sample means) or 2.
4) Type of T.TEST. 1 dependent, matched subjects.
2: population variances unknown but equal. [Often true]
3: population variances unknown & unequal. [Conservative]

Run Hypothesis Tests from this data: B1:I241

Data for Q1-Q4 (B-E) is Binary: 0=No, 1=Yes.
Data for Q5-Q6 (F-G) is Ordinal (discrete): 1-5.
Data for Q7-Q8 (H-I) is Quantitative (ratio).

	A	B	C	D	E	F	G	H	I
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2	1	0	1	0	0	3	5	67	5
3	2	0	1	0	1	4	1	62	4
4	3	0	1	0	1	3	4	60	5
5	4	0	1	1	0	4	5	60	4
6	5	0	0	1	0	3	1	71	3

Excel instructions and data at:
www.StatLit.org/xls/2012Isaacson240Data.xls

Approach

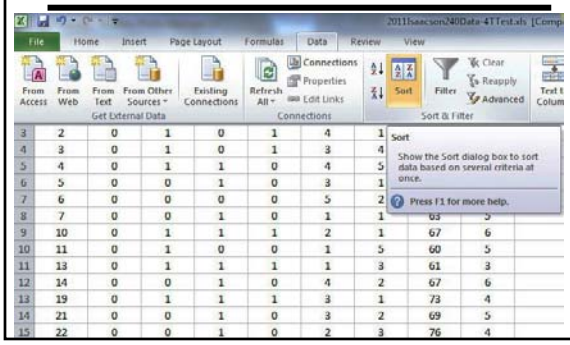
Excel's two-population T-Test function requires that the data be "stacked" (separated into two groups) by the value of the predictor. Predictor must be binary.

If the binary predictor is the answer to Q1, then *the entire data set* must be sorted by Q1.

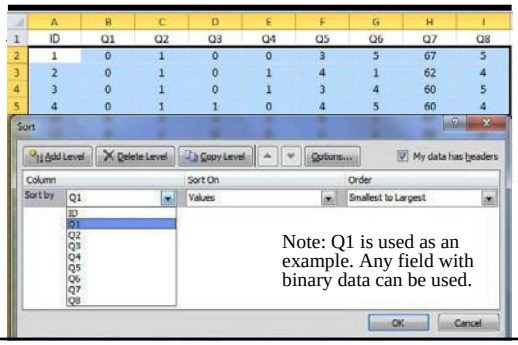
The Excel "Sort" requires that the entire data set be selected **before** invoking the sort command. A common mistake is to sort just a single column rather than the entire dataset.

Unfortunately Excel does not have a "stacked" or conditional T-Test function. T-Test function will automatically update p-values if data is re-sorted.

A: Select data!! From the Home or Data tab, select Sort



B: In Sort dialogue box, select Sort Column by "Q1"



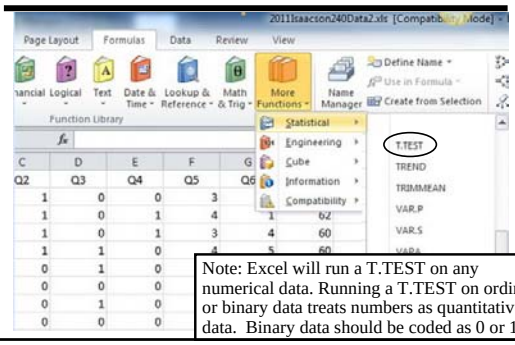
Note: Q1 is used as an example. Any field with binary data can be used.

C: Q2 for Q1=0 from C2 to C131. Q2 for Q1=1 from C132 to C241.

	A	B	C	D	E	F	G	H	I
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2	1	0	1	0	0	3	5	67	5
3	2	0	1	0	1	4	1	62	4
4									
125									
126	228	0	0	1	1	5	1	76	6
127	229	0	0	1	0	1	1	68	6
128	232	0							4
129	233	0							7
130	237	0							5
131	239	0							5
132	8	1							5
133	9	1							7
134	12	1							5
135	15	1	0	0	0	4	2	70	6
136	16	1	0	0	1	5	1	54	5
137	17	1	0	1	1	3	2	48	6

All Excel hypothesis tests require the data to be **stacked: one group on top of the other**. After sort on binary data in column B (Q1), data in columns C through I (Q2-Q8) is **stacked**. **Group 1** in rows 2-131 have Q1=0; **Group 2** in rows 132-241 have Q1=1.

D: Place cursor for results. From *Statistical*, select T.TEST



Note: Excel will run a T.TEST on any numerical data. Running a T.TEST on ordinal or binary data treats numbers as quantitative data. Binary data should be coded as 0 or 1.

T.TEST Procedure Given Stacked Data

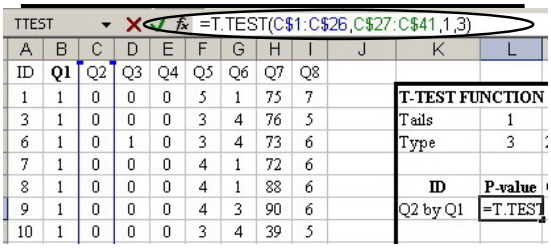
1: Place cursor where T.TEST p-value will be recorded. Locate this cell in a different place for each new test. Label the cell to reflect the T.TEST inputs. E.g., Q2 by Q1.

2: Insert T.TEST in Excel 2008 or newer (TTEST in 2003): Test for a two-group difference in Means (Measures) or in Proportions (Counts)

3: Enter appropriate data or cell references for the T.TEST function arguments. See examples on following slides.

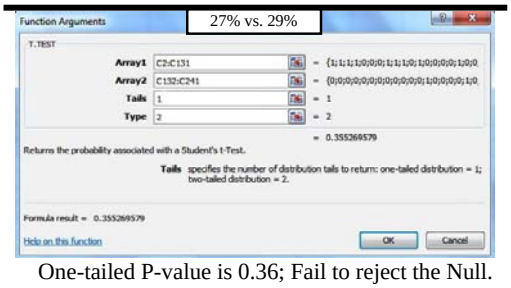
4: T.TEST will change if data is resorted. Three solutions:
A) Put data from each sort in a separate tab.
B) Copy sorted data to a separate place on one worksheet.
C) Copy & Paste/Special/Values with appropriate labeling.

1-3) Insert T.Test Function



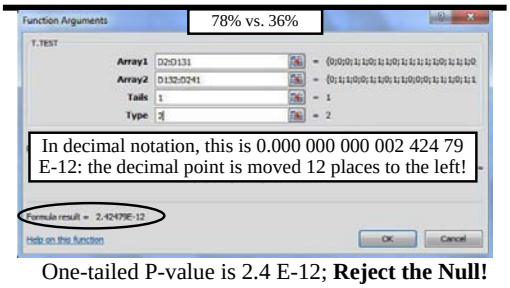
Insert T.TEST function in cell L7.
Enter four arguments: Array1, Array2, Tails and Type.

3a) T.Test for Proportions: Results for Q2 by Q1



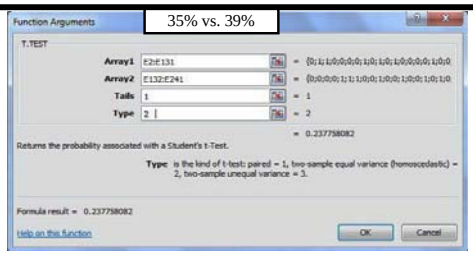
One-tailed P-value is 0.36; Fail to reject the Null.
Difference in Q2 by Q1 is "not statistically significant"

3b) T.Test for Proportions: Results for Q3 by Q1



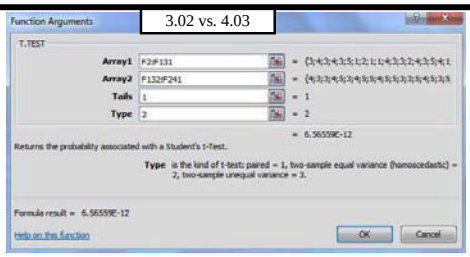
One-tailed P-value is 2.4 E-12; **Reject the Null!**
Difference in Q3 by Q1 **is 'statistically significant'.**

3c) T.Test for Proportions: Results for Q4 by Q1



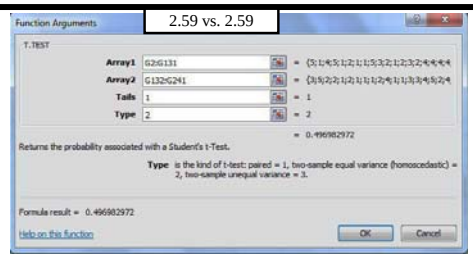
One-tailed P-value is 0.24; Fail to reject the Null.
Difference in Q4 by Q1 is "not statistically significant"

3d) T.Test for Measures: Results for Q5 by Q1



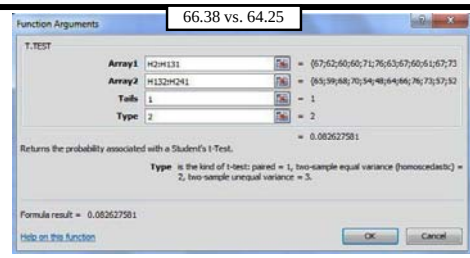
One tailed P-value is 6.5E-12; **Reject the Null**.
Difference in Q5 by Q1 **is "statistically significant".**

3e) T.Test for Measures: Results for Q6 by Q1



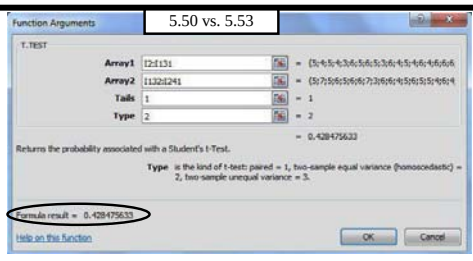
One tailed P-value is 0.50; Fail to reject the Null.
Difference in Q6 by Q1 is "not statistically significant"

3f) T.Test for Measures: Results for Q7 by Q1



One tailed P-value is 0.08; Fail to reject the Null.
Difference in Q7 by Q1 is "not statistically significant"

3g) T.Test for Measures: Results for Q8 by Q1



One tailed P-value is 0.43; Fail to reject the Null.
Difference in Q8 by Q1 is 'not statistically significant'

3h) T.TEST Results: All fields by Q1

Data	Q2 by Q1	Q3 by Q1	Q4 by Q1	Q5 by Q1	Q6 by Q1	Q7 by Q1	Q8 by Q1
P-Value	0.36	2.4E-12	0.24	6.57E-12	0.50	0.08	0.43
Stat. Sig	No	Yes	No	Yes	No	No	No
Average							
Q1=0	0.27	0.78	0.35	3.02	2.59	66.38	5.50
Q1=1	0.29	0.36	0.39	4.03	2.59	64.25	5.53
Differ	0.02	0.42	0.04	1.01	0.00	2.12	0.03
StdDev	0.45	0.49	0.48	1.21	1.42	11.79	1.16
Effect Size	5%	85%	9%	84%	0%	18%	2%

Pooled Std. Dev. = Sqrt(((n1-1)S1^2 + (n2-1)S2^2)/(n1+n2 - 2))
Effect size = Difference in Means / Pooled Std. Deviation

T.TEST Procedure: Step 4

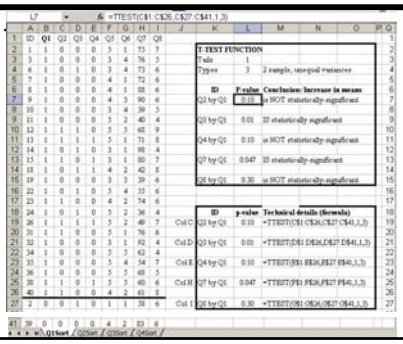
T.TEST function will change if the data is resorted.

There are three solutions:
- 4A) Put data from each sort in a separate tab.
- 4B) Copy sorted data to separate places on one sheet.
- 4C) Copy & Paste/Special/Values with good labels.

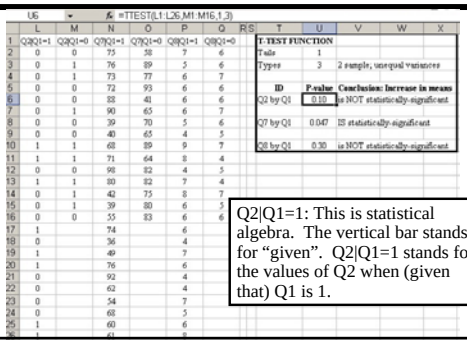
4C is not recommended since there is no clear audit trail.

In a one-tailed test, the T.TEST always tests whether the larger statistic is bigger than the smaller.

4A) Separate tabs for each sort

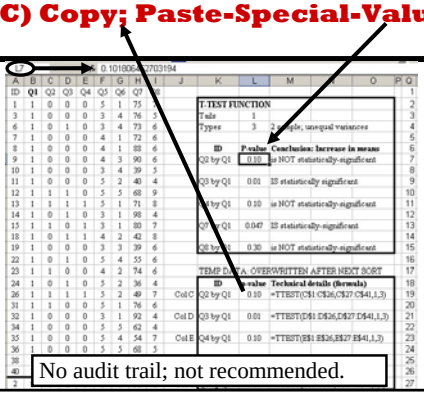


4B) Separate Sorts on 1 sheet



Q2|Q1=1: This is statistical algebra. The vertical bar stands for "given". Q2|Q1=1 stands for the values of Q2 when (given that) Q1 is 1.

4C) Copy; Paste-Special-Values



No audit trail; not recommended.

Summary

In a one-tailed test, T.TEST always tests whether the positive difference between the larger sample statistic and the smaller is statistically-significant.

"Reject the null hypothesis" and "Failure to reject the null hypothesis" are technical conclusions.

"A difference or change IS [or IS NOT] statistically significant" is a non-technical conclusion.

Use the non-technical expressions for everyday communication.

Other Options

In testing sample statistics from two groups for statistical significance, Excel provides two other methods:
- the **t-test command** in the Data Analysis Toolpak, and
- combinations of basic Excel Functions.

The **t-test command** has the clearest documentation (audit trail). All Excel methods require the two-group data be in contiguous blocks.

See statistics textbooks for more on differences between paired or matched subjects. Examples include before-after differences on the same subjects, husband-wife differences, and differences in two appraisals of the same houses.

Two-Group Hypothesis Tests Using Excel T.TEST Function

by
Milo Schield
Member: International Statistical Institute
US Rep: International Statistical Literacy Project
Director, W. M. Keck Statistical Literacy Project

Slides and audio at: www.StatLit.org/pdf/TTEST-Function-Excel-2008-6up.pdf
Audio/TTEST-Function-Excel-2008.mp3

Excel T.TEST Function

Purpose: Calculate likelihood (p-value) of getting the observed difference in two sample means (or more extreme) by chance in random samples – assuming there is no difference in the two population means (the Null Hypothesis).

Note: TTEST function was available in Excel 2003.

Four Inputs:

1) Array or range of 1st sample. 2) Array or range of 2nd sample.
3) Tails: 1 (Excel matches Alternate with sample means) or 2.
4) Type of T.TEST. 1 dependent, matched subjects.
   2: population variances unknown but equal. [Often true]
   3: population variances unknown & unequal. [Conservative]

Run Hypothesis Tests from this data: B1:I241

Data for Q1-Q4 (B-E) is Binary: 0=No, 1=Yes.
Data for Q5-Q6 (F-G) is Ordinal (discrete): 1-5.
Data for Q7-Q8 (H-I) is Quantitative (ratio).

	A	B	C	D	E	F	G	H	I
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2	1	0	1	0	0	3	5	67	5
3	2	0	1	0	1	4	1	62	4
4	3	0	1	0	1	3	4	60	5
5	4	0	1	1	0	4	5	60	4
6	5	0	0	1	0	3	1	71	3

Excel instructions and data at:
www.StatLit.org/xls/2012Isaacson240Data.xls

Approach

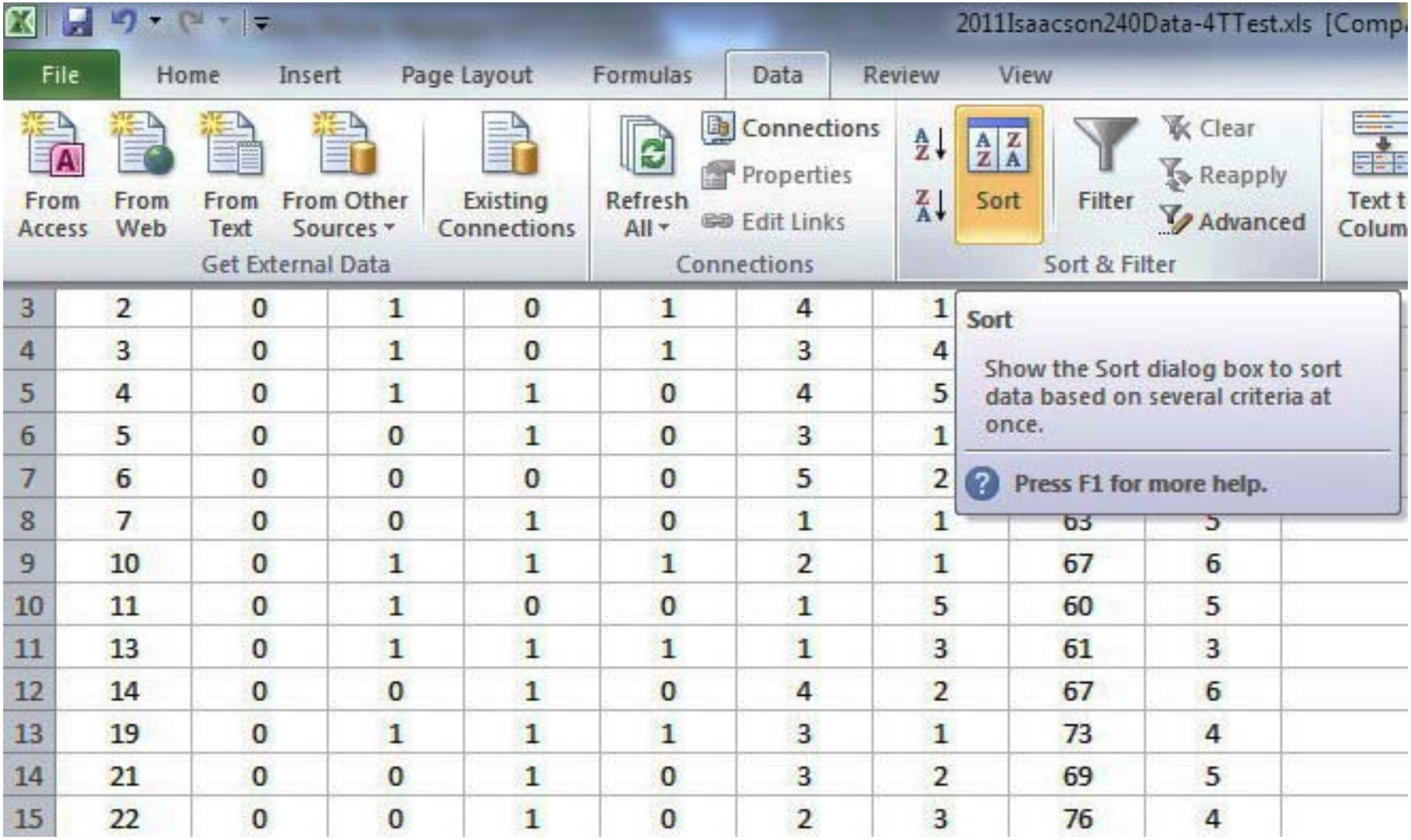
Excel's two-population T-Test function requires that the data be "stacked" (separated into two groups) by the value of the predictor. Predictor must be binary.

If the binary predictor is the answer to Q1, then *the entire data set* must be sorted by Q1.

The Excel "Sort" requires that the entire data set be selected **before** invoking the sort command. A common mistake is to sort just a single column rather than the entire dataset.

Unfortunately Excel does not have a "stacked" or conditional T-Test function. T-Test function will automatically update p-values if data is re-sorted.

A: Select data!! From the Home or Data tab, select Sort



3	2	0	1	0	1	4	1		
4	3	0	1	0	1	3	4		
5	4	0	1	1	0	4	5		
6	5	0	0	1	0	3	1		
7	6	0	0	0	0	5	2		
8	7	0	0	1	0	1	1	63	5
9	10	0	1	1	1	2	1	67	6
10	11	0	1	0	0	1	5	60	5
11	13	0	1	1	1	1	3	61	3
12	14	0	0	1	0	4	2	67	6
13	19	0	1	1	1	3	1	73	4
14	21	0	0	1	0	3	2	69	5
15	22	0	0	1	0	2	3	76	4

B: In Sort dialogue box, select Sort Column by "Q1"

	A	B	C	D	E	F	G	H	I
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2	1	0	1	0	0	3	5	67	5
3	2	0	1	0	1	4	1	62	4
4	3	0	1	0	1	3	4	60	5
5	4	0	1	1	0	4	5	60	4

Sort

Add Level | Delete Level | Copy Level | Options... | My data has headers

Column	Sort On	Order
Sort by Q1	Values	Smallest to Largest

ID
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8

Note: Q1 is used as an example. Any field with binary data can be used.

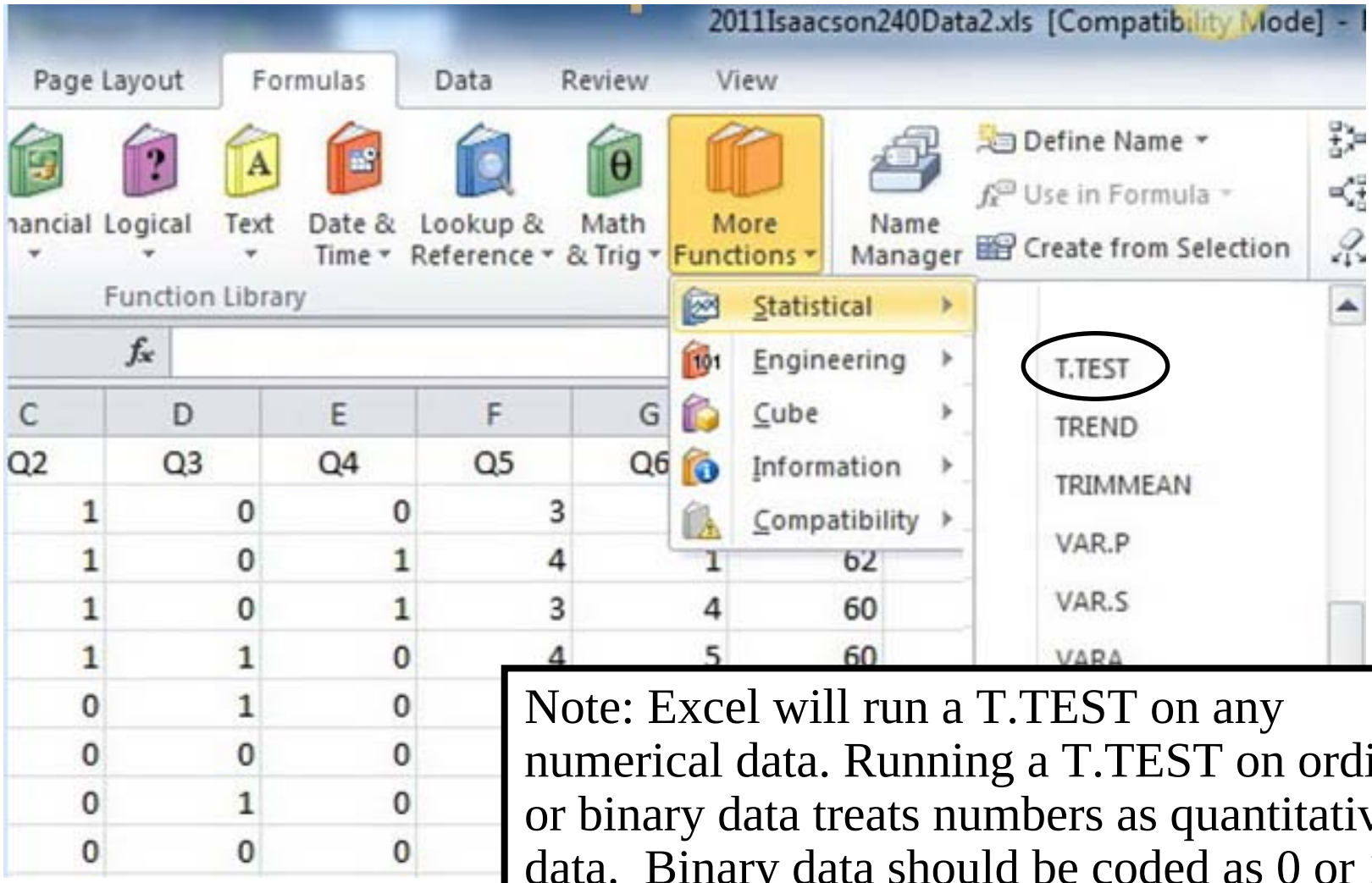
OK | Cancel

C: Q2 for Q1=0 from C2 to C131. Q2 for Q1=1 from C132 to C241.

	A	B	C	D	E	F	G	H	I
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2	1	0	1	0	0	3	5	67	5
3	2	0	1	0	1	4	1	62	4
4									
125									
126	228	0	0	1	1	5	1	76	6
127	229	0	0	1	0	1	1	68	6
128	232	0							4
129	233	0							7
130	237	0							5
131	239	0							5
132	8	1							5
133	9	1							7
134	12	1							5
135	15	1	0	0	0	4	2	70	6
136	16	1	0	0	1	5	1	54	5
137	17	1	0	1	1	3	2	48	6

All Excel hypothesis tests require the data to be **stacked: one group on top of the other**. After sort on binary data in column B (Q1), data in columns C through I (Q2-Q8) is **stacked**. **Group 1** in rows 2-131 have Q1=0; **Group 2** in rows 132-241 have Q1=1.

D: Place cursor for results. From *Statistical*, select T.TEST



Note: Excel will run a T.TEST on any numerical data. Running a T.TEST on ordinal or binary data treats numbers as quantitative data. Binary data should be coded as 0 or 1.

T.TEST Procedure Given Stacked Data

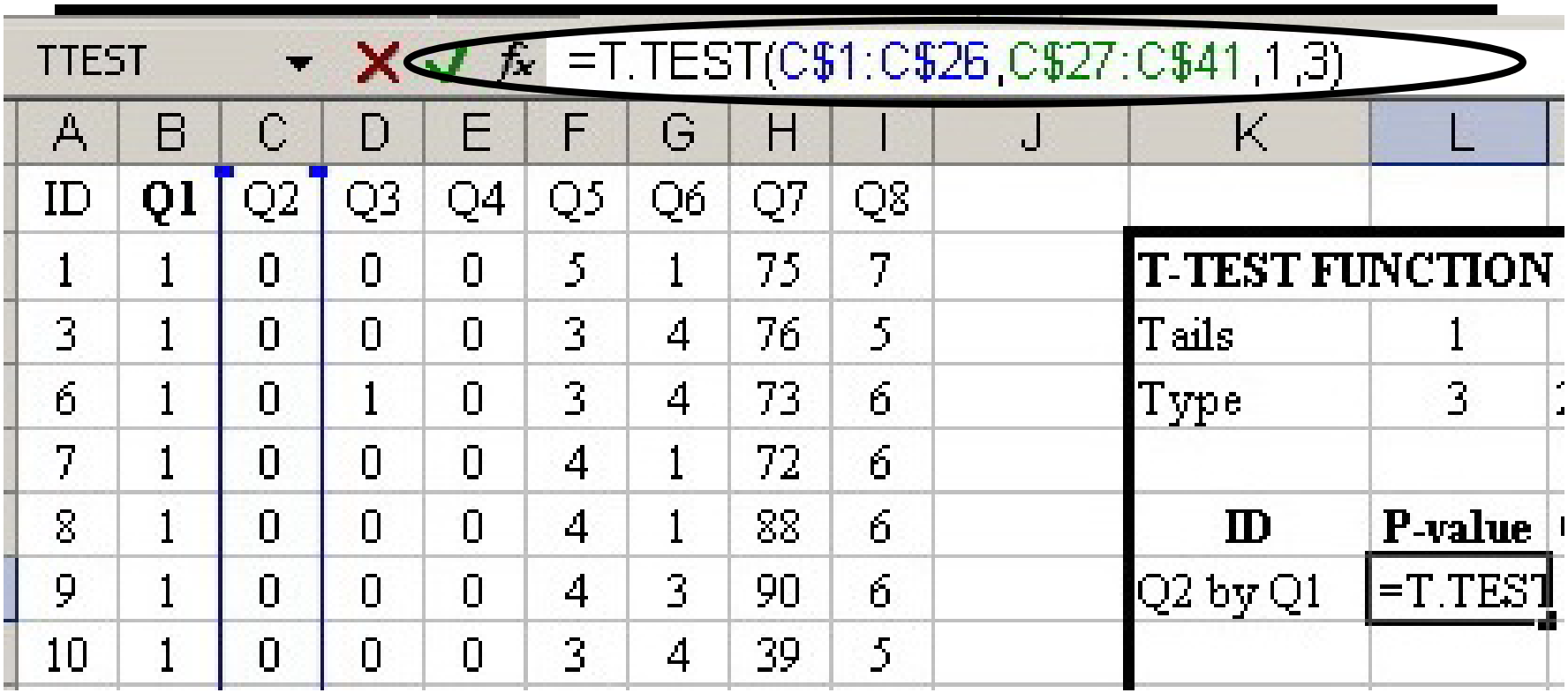
1: Place cursor where T.TEST p-value will be recorded.
Locate this cell in a different place for each new test.
Label the cell to reflect the T.TEST inputs. E.g., Q2 by Q1.

2: Insert T.TEST in Excel 2008 or newer (TTEST in 2003):
Test for a two-group difference in Means (Measures) or
in Proportions (Counts)

3: Enter appropriate data or cell references for the T.TEST
function arguments. See examples on following slides.

4: T.TEST will change if data is resorted. Three solutions:
A) Put data from each sort in a separate tab.
B) Copy sorted data to a separate place on one worksheet.
C) Copy & Paste/Special/Values with appropriate labeling.

1-3) Insert T.Test Function

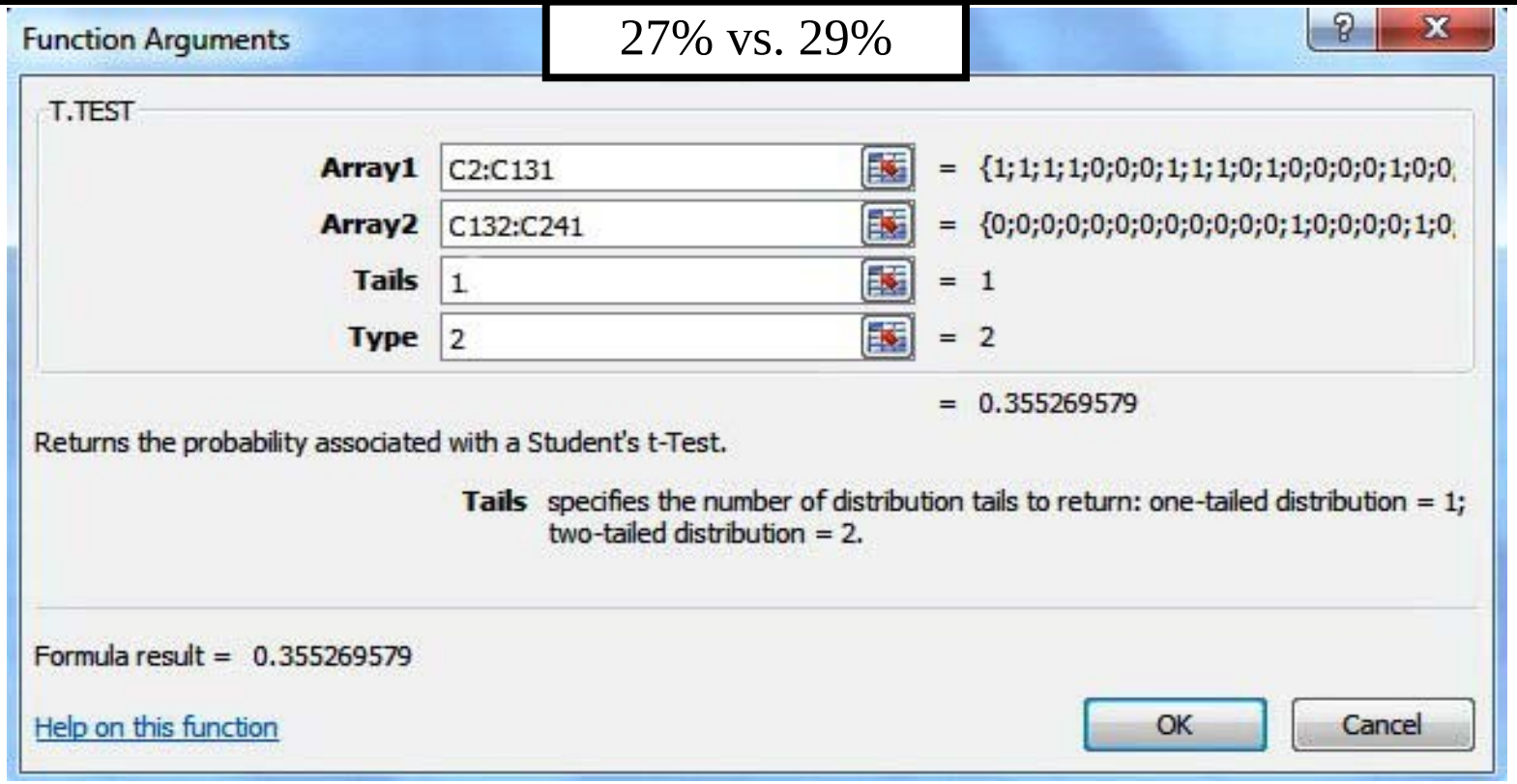


A	B	C	D	E	F	G	H	I	J	K	L
ID	**Q1**	Q2	Q3	Q4	Q5	Q6	Q7	Q8			
1	1	0	0	0	5	1	75	7			
3	1	0	0	0	3	4	76	5			
6	1	0	1	0	3	4	73	6			
7	1	0	0	0	4	1	72	6			
8	1	0	0	0	4	1	88	6			
9	1	0	0	0	4	3	90	6			
10	1	0	0	0	3	4	39	5			

Insert T.TEST function in cell L7.

Enter four arguments: Array1, Array2, Tails and Type.

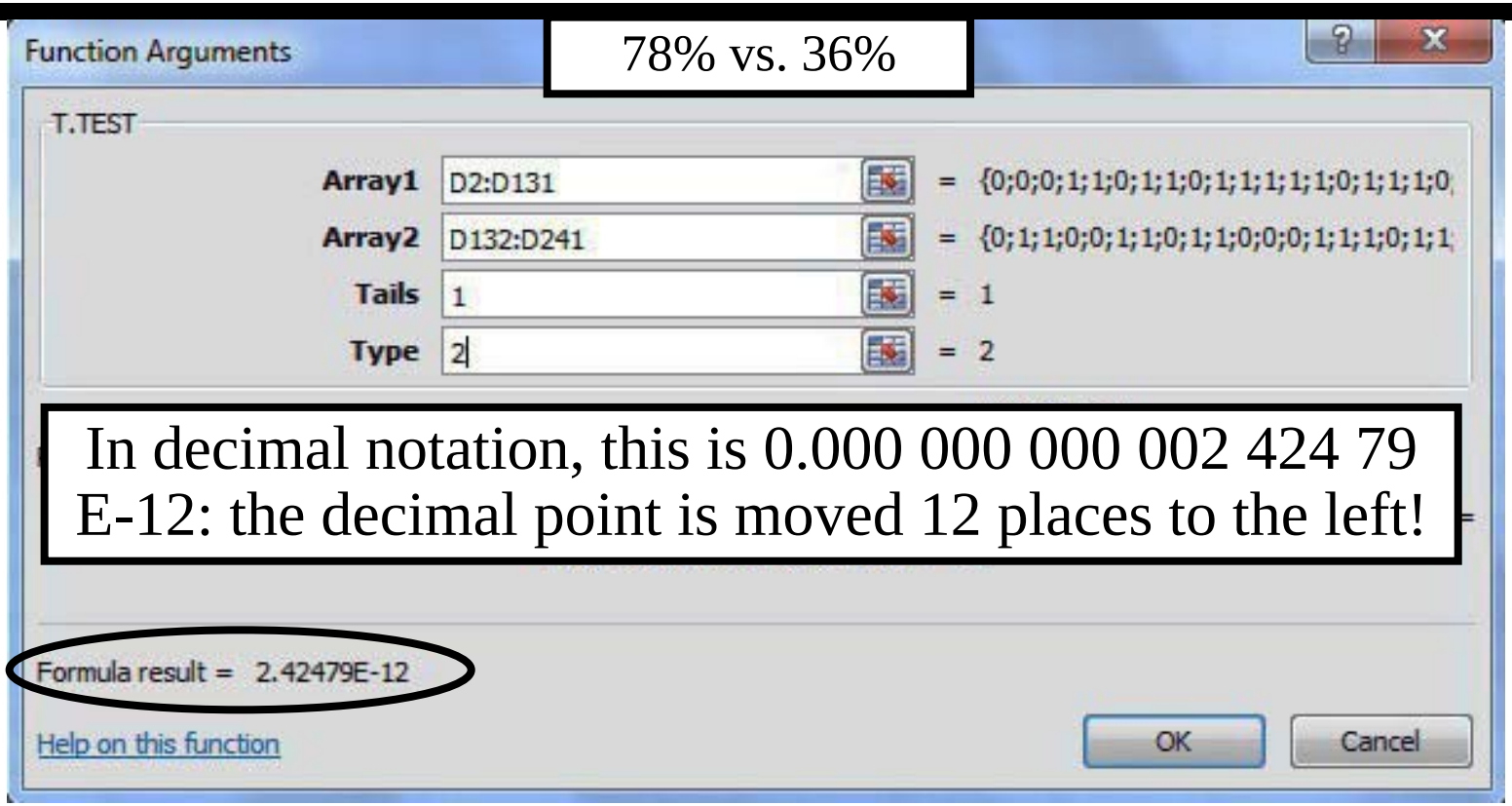
3a) T.Test for Proportions: Results for Q2 by Q1



One-tailed P-value is 0.36; Fail to reject the Null.

Difference in Q2 by Q1 is “not statistically significant”

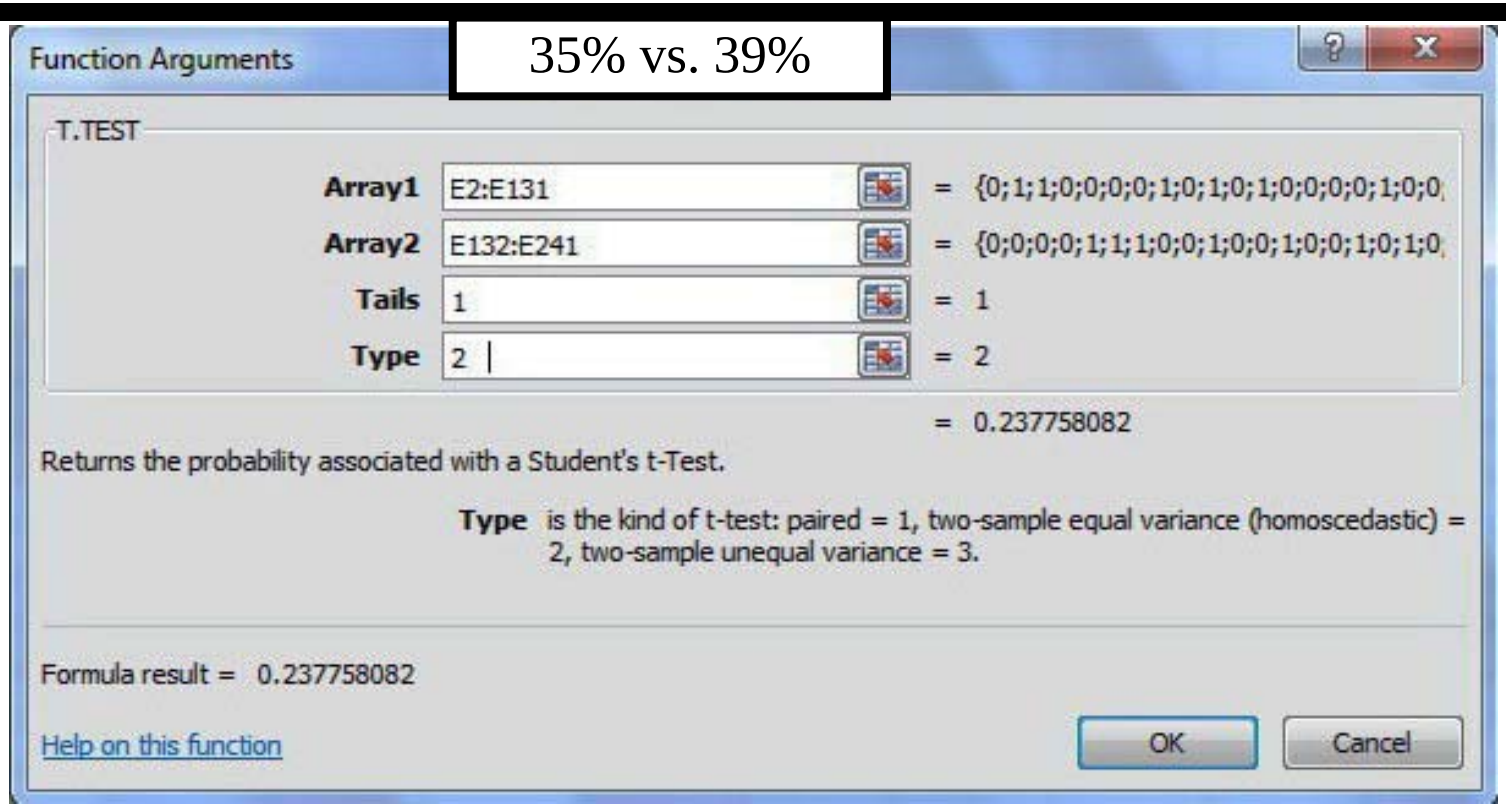
3b) T.Test for Proportions: Results for Q3 by Q1



One-tailed P-value is 2.4 E-12; **Reject the Null!**

Difference in Q3 by Q1 **is 'statistically significant'.**

3c) T.Test for Proportions: Results for Q4 by Q1

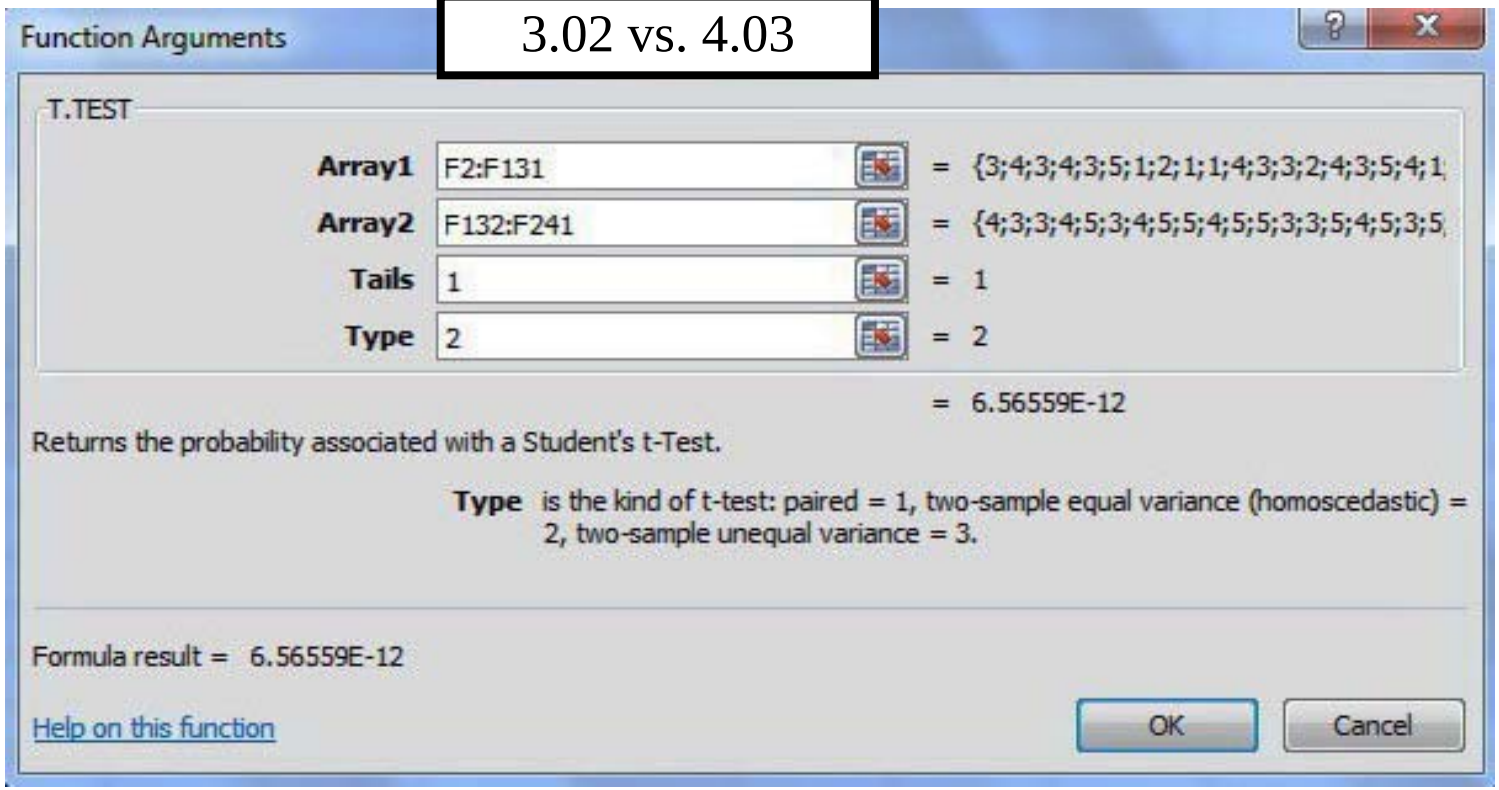


One-tailed P-value is 0.24; Fail to reject the Null.

Difference in Q4 by Q1 is “not statistically significant”

3d) T.Test for Measures: Results for Q5 by Q1

3.02 vs. 4.03

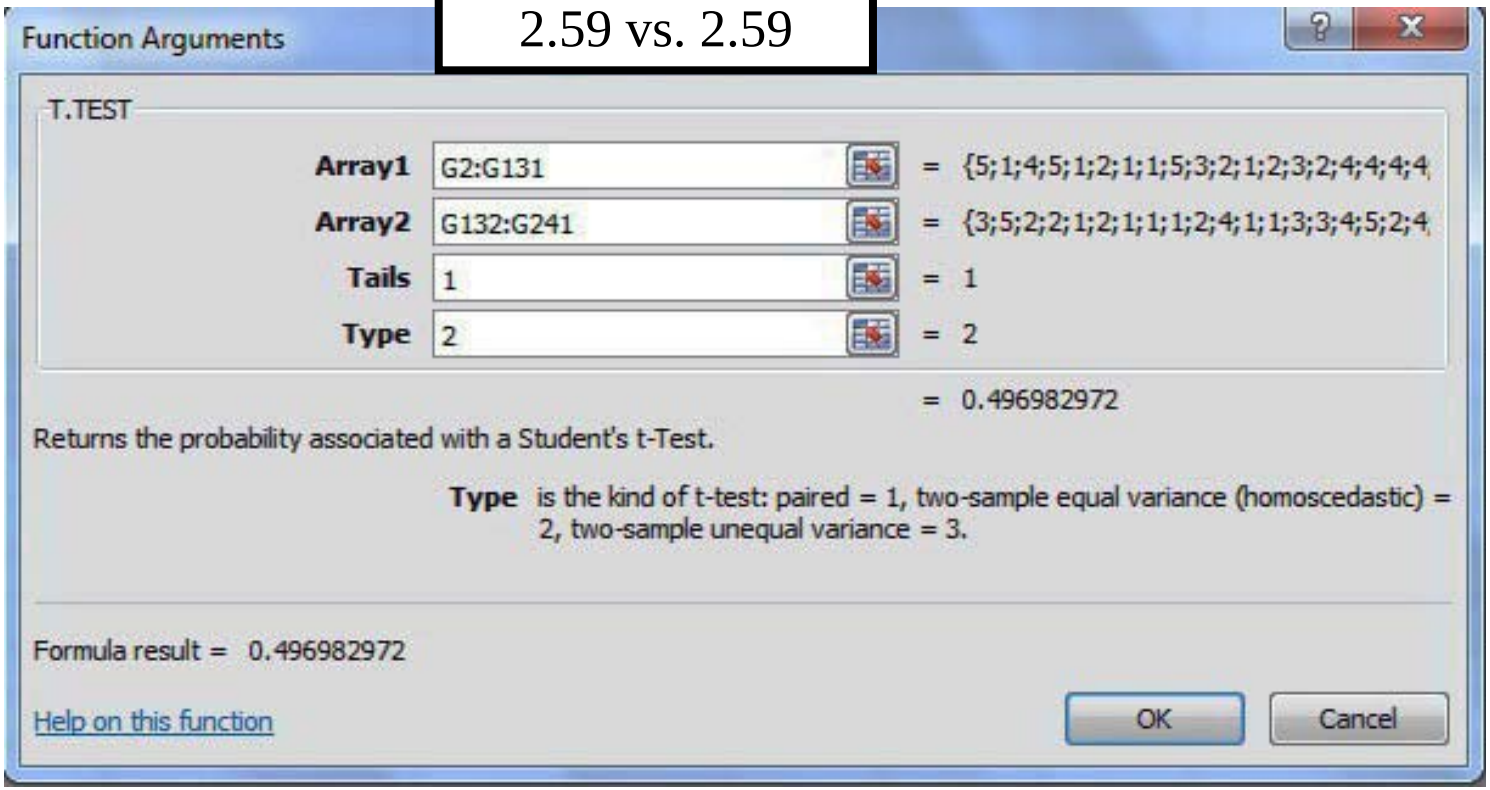


One tailed P-value is 6.5E-12; **Reject the Null**.

Difference in Q5 by Q1 **is "statistically significant".**

3e) T.Test for Measures: Results for Q6 by Q1

2.59 vs. 2.59

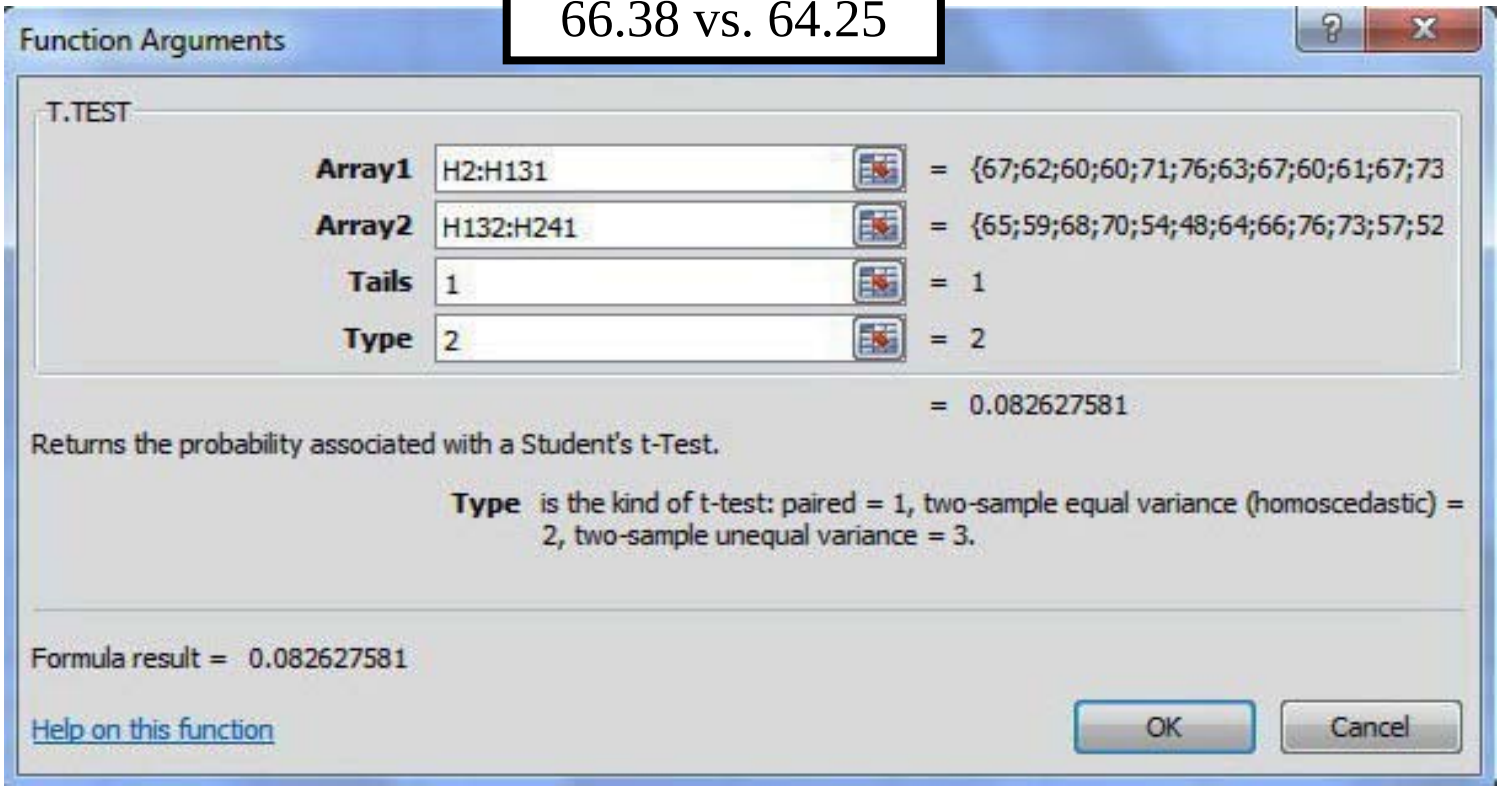


One tailed P-value is 0.50; Fail to reject the Null.

Difference in Q6 by Q1 is “not statistically significant”

3f) T.Test for Measures: Results for Q7 by Q1

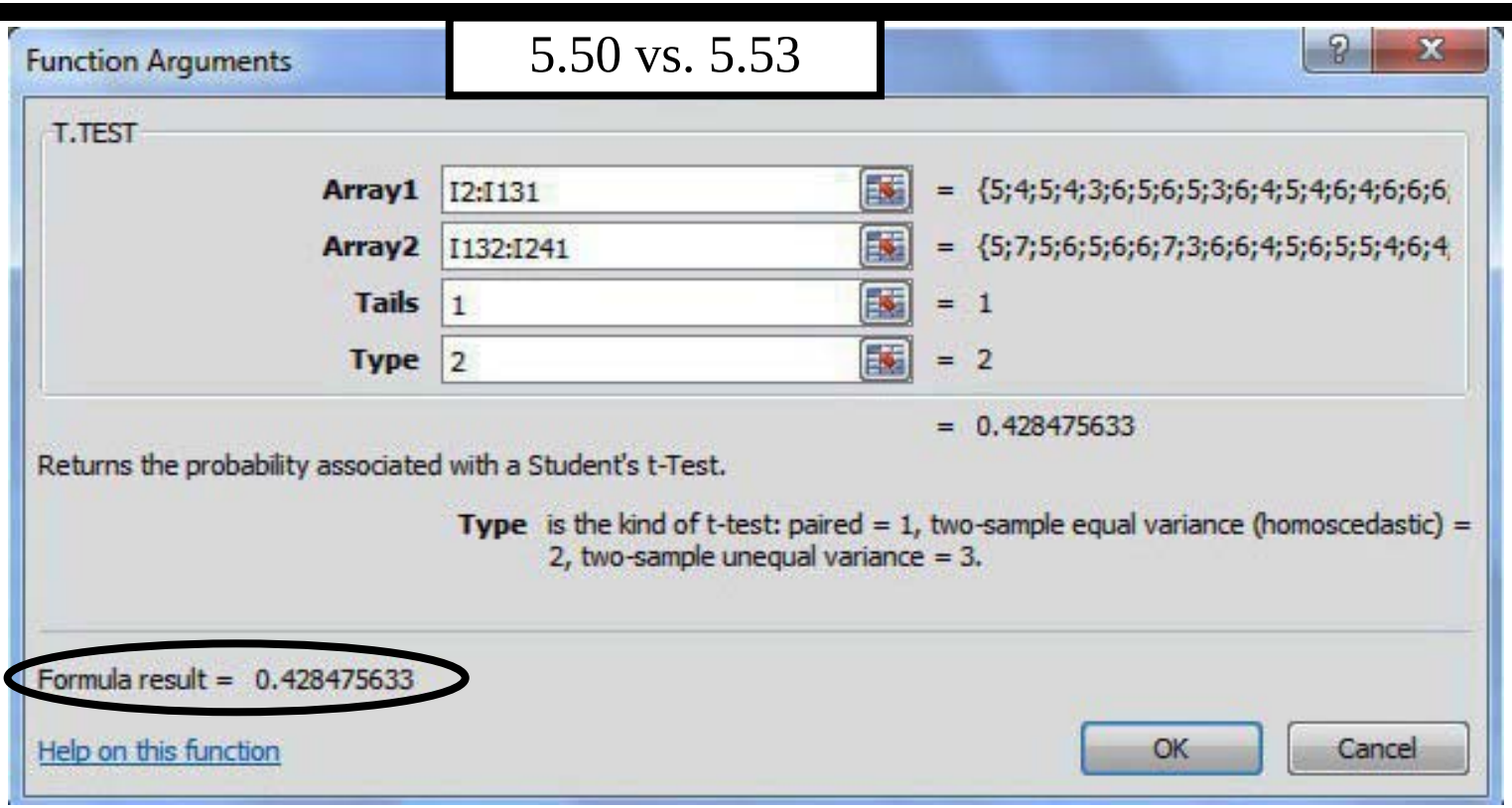
66.38 vs. 64.25



One tailed P-value is 0.08; Fail to reject the Null.

Difference in Q7 by Q1 is “not statistically significant”

3g) T.Test for Measures: Results for Q8 by Q1



One tailed P-value is 0.43; Fail to reject the Null.

Difference in Q8 by Q1 is 'not statistically significant'

3h) T.TEST Results: All fields by Q1

Data	Q2 by Q1	Q3 by Q1	Q4 by Q1	Q5 by Q1	Q6 by Q1	Q7 by Q1	Q8 by Q1
P-Value	0.36	2.4E-12	0.24	6.57E-12	0.50	0.08	0.43
Stat. Sig	No	Yes	No	Yes	No	No	No
Average							
Q1=0	0.27	0.78	0.35	3.02	2.59	66.38	5.50
Q1=1	0.29	0.36	0.39	4.03	2.59	64.25	5.53
Differ	0.02	0.42	0.04	1.01	0.00	2.12	0.03
StdDev	0.45	0.49	0.48	1.21	1.42	11.79	1.16
Effect Size	5%	85%	9%	84%	0%	18%	2%

Pooled Std. Dev. = Sqrt(((n1-1)S1^2 + (n2-1)S2^2)/(n1+n2 - 2))

Effect size = Difference in Means / Pooled Std. Deviation

T.TEST Procedure: Step 4

T.TEST function will change if the data is resorted.

There are three solutions:

4A) Put data from each sort in a separate tab.

4B) Copy sorted data to separate places on one sheet.

4C) Copy & Paste/Special/Values with good labels.

4C is not recommended since there is no clear audit trail.

In a one-tailed test, the T.TEST always tests whether the larger statistic is bigger than the smaller.

4A) Separate tabs for each sort

L7 =TTEST(C$1:C$26,C$27:C$41,1,3)

	A	B	C	D	E	F	G	H	I
1	ID	**Q1**	Q2	Q3	Q4	Q5	Q6	Q7	Q8
2	1	1	0	0	0	5	1	75	7
3	3	1	0	0	0	3	4	76	5
4	6	1	0	1	0	3	4	73	6
5	7	1	0	0	0	4	1	72	6
6	8	1	0	0	0	4	1	88	6
7	9	1	0	0	0	4	3	90	6
8	10	1	0	0	0	3	4	39	5
9	11	1	0	0	0	5	2	40	4
10	12	1	1	1	0	5	5	68	9
11	13	1	1	1	1	5	1	71	8
12	14	1	0	1	0	3	1	98	4
13	15	1	1	0	1	3	1	80	7
14	18	1	0	1	1	4	2	42	8
15	19	1	0	0	0	3	3	39	6
16	22	1	0	1	0	5	4	55	6
17	23	1	1	0	0	4	2	74	6
18	24	1	0	1	0	5	2	36	4
19	26	1	1	1	1	5	2	49	7
20	31	1	1	0	0	5	1	76	6
21	32	1	0	0	0	3	1	92	4
22	34	1	0	0	0	5	5	62	4
23	35	1	0	0	0	5	4	54	7
24	36	1	0	0	0	5	5	68	5
25	38	1	1	0	1	5	5	60	6
26	40	1	1	0	0	4	2	61	8
27	2	0	0	1	0	1	1	58	6
41	39	0	0	0	0	4	2	83	6

Q1Sort | Q2Sort | Q3Sort | Q4Sort

T-TEST FUNCTION				
Tails	1			
Types	3	2 sample; unequal variances		
ID	**P-value**	**Conclusion: Increase in means**		
Q2 by Q1	0.10	is NOT statistically-significant		
Q3 by Q1	0.01	IS statistically significant		
Q4 by Q1	0.10	is NOT statistically-significant		
Q7 by Q1	0.047	IS statistically-significant		
Q8 by Q1	0.30	is NOT statistically-significant		

	ID	**p-value**	**Technical details (formula)**
Col C	Q2 by Q1	0.10	=TTEST(C$1:C$26,C$27:C$41,1,3)
Col D	Q3 by Q1	0.01	=TTEST(D$1:D$26,D$27:D$41,1,3)
Col E	Q4 by Q1	0.10	=TTEST(E$1:E$26,E$27:E$41,1,3)
Col H	Q7 by Q1	0.047	=TTEST(F$1:F$26,F$27:F$41,1,3)
Col I	Q8 by Q1	0.30	=TTEST(G$1:G$26,G$27:G$41,1,3)

4B) Separate Sorts on 1 sheet

- U6 fx =TTEST(L1:L26,M1:M16,1,3)

	L	M	N	O	P	Q
1	Q2|Q1=1	Q2|Q1=0	Q7|Q1=1	Q7|Q1=0	Q8|Q1=1	Q8|Q1=0
2	0	0	75	58	7	6
3	0	1	76	89	5	6
4	0	1	73	77	6	7
5	0	0	72	93	6	6
6	0	0	88	41	6	6
7	0	1	90	65	6	7
8	0	0	39	70	5	6
9	0	0	40	65	4	5
10	1	1	68	89	9	7
11	1	1	71	64	8	4
12	0	0	98	82	4	5
13	1	1	80	82	7	4
14	0	1	42	75	8	7
15	0	1	39	80	6	5
16	0	0	55	83	6	6
17	1		74		6	
18	0		36		4	
19	1		49		7	
20	1		76		6	
21	0		92		4	
22	0		62		4	
23	0		54		7	
24	0		68		5	
25	1		60		6	
26	1		61		8	

T	U	V
T-TEST FUNCTION		
Tails	1	
Types	3	2 sample; unequal variances
ID	**P-value**	**Conclusion: Increase in means**
Q2 by Q1	0.10	is NOT statistically-significant
Q7 by Q1	0.047	IS statistically-significant
Q8 by Q1	0.30	is NOT statistically-significant

Q2|Q1=1: This is statistical algebra. The vertical bar stands for “given”. Q2|Q1=1 stands for the values of Q2 when (given that) Q1 is 1.

4C) Copy; Paste-Special-Values

- L7 → fx 0.1018064[illegible]2703194

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
ID	**Q1**	Q2	Q3	Q4	Q5	Q6	Q7	[illegible]8								1
1	1	0	0	0	5	1	75	7		**T-TEST FUNCTION**						2
3	1	0	0	0	3	4	76	5		Tails	1					3
6	1	0	1	0	3	4	73	6		Types	3	2 sample; unequal variances				4
7	1	0	0	0	4	1	72	6								5
8	1	0	0	0	4	1	88	6		**ID**	**P-value**	**Conclusion: Increase in means**				6
9	1	0	0	0	4	3	90	6		Q2 by Q1	0.10	is NOT statistically-significant				7
10	1	0	0	0	3	4	39	5								8
11	1	0	0	0	5	2	40	4		Q3 by Q1	0.01	IS statistically significant				9
12	1	1	1	0	5	5	68	9								10
13	1	1	1	1	5	1	71	8		Q4 by Q1	0.10	is NOT statistically-significant				11
14	1	0	1	0	3	1	98	4								12
15	1	1	0	1	3	1	80	7		Q7 by Q1	0.047	IS statistically-significant				13
18	1	0	1	1	4	2	42	8								14
19	1	0	0	0	3	3	39	6		Q8 by Q1	0.30	is NOT statistically-significant				15
22	1	0	1	0	5	4	55	6								16
23	1	1	0	0	4	2	74	6		TEMP DATA: OVERWRITTEN AFTER NEXT SORT						17
24	1	0	1	0	5	2	36	4		**ID**	**p-value**	**Technical details (formula)**				18
26	1	1	1	1	5	2	49	7	Col C	Q2 by Q1	0.10	=TTEST(C$1:C$26,C$27:C$41,1,3)				19
31	1	1	0	0	5	1	76	6								20
32	1	0	0	0	3	1	92	4	Col D	Q3 by Q1	0.01	=TTEST(D$1:D$26,D$27:D$41,1,3)				21
34	1	0	0	0	5	5	62	4								22
35	1	0	0	0	5	4	54	7	Col E	Q4 by Q1	0.10	=TTEST(E$1:E$26,E$27:E$41,1,3)				23
36	1	0	0	0	5	5	68	5								24
38																25
40																26
2																27

No audit trail; not recommended.

Summary

In a one-tailed test, T.TEST always tests whether the positive difference between the larger sample statistic and the smaller is statistically-significant.

"Reject the null hypothesis" and "Failure to reject the null hypothesis" are technical conclusions.

"A difference or change IS [or IS NOT] statistically significant" is a non-technical conclusion.

Use the non-technical expressions for everyday communication.

Other Options

In testing sample statistics from two groups for statistical significance, Excel provides two other methods:

- the **t-test command** in the Data Analysis Toolpak, and
- combinations of basic Excel Functions.

The **t-test command** has the clearest documentation (audit trail). All Excel methods require the two-group data be in contiguous blocks.

See statistics textbooks for more on differences between paired or matched subjects. Examples include before-after differences on the same subjects, husband-wife differences, and differences in two appraisals of the same houses.