Two-Group Hypothesis Tests: Excel 2003 T-TEST Command

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/T-TEST-Command-Excel-2003-6up.pdf

Excel 2003 T-TEST Command

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).

Four Inputs:

1) Array or range of two samples. 2) Alpha cutoff.
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]

Use 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

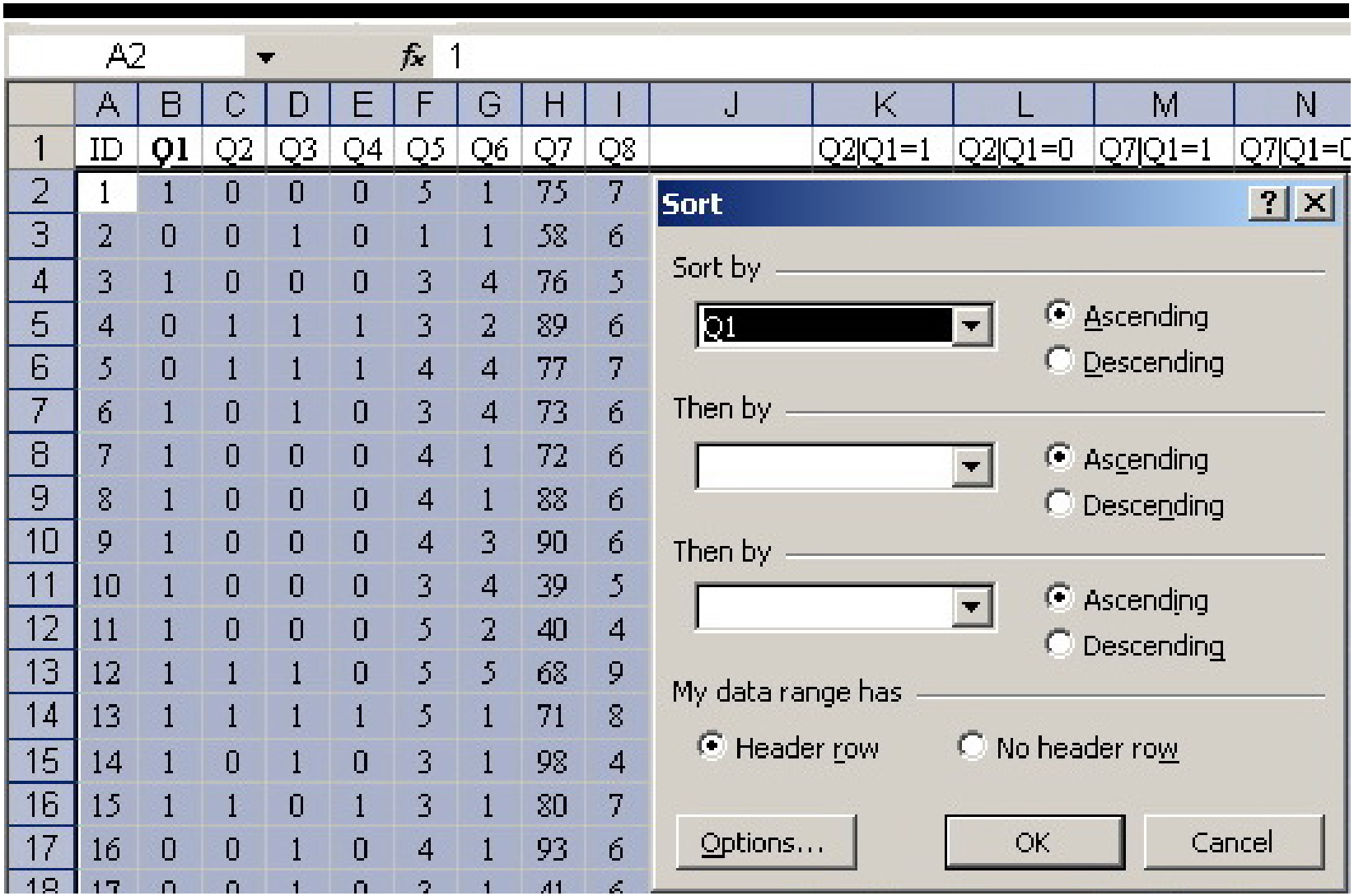
Excel's two-population T-Test command requires that the data be "stacked" (separated into two groups) by the value of the predictor. Predictor must be binary.

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. The T-Test command will not automatically update p-values if data is changed.

A: Select data.
From Data tab, select Sort



B: Sort Data.
Prepare column headings.

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	**Q1**	Q2	Q3	Q4	Q5	Q6	Q7	Q8		Q2|Q1=1	Q2|Q1=0
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			

Create headings that show what column or question is being tested and what column or question is used to split the data into two groups.

Q1 is used as the two-group splitter in this example. Any field with binary data can be used.

The vertical bar "|" means "given" so Q2|Q1=1 indicates the values of Q2 for which Q1 equals 1.

C: Copy stacked data to separate columns

Q1=1

Q1=0.

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8		Q2|Q1=1	Q2|Q1=0	Q7|Q1=1	Q7|Q1=0	Q8|Q1=1	Q8|Q1=0
1	0	0	0	5	1	75	7		0	0	75	58	7	6
1	0	0	0	3	4	76	5		0	1	76	89	5	6
1	0	1	0	3	4	73	6		0	1	73	77	6	7
1	0	0	0	4	1	72	6		0	0	72	93	6	6
1	0	0	0	4	1	88	6		0	0	88	41	6	6
1	0	0	0	4	3	90	6		0	1	90	65	6	7
1	0	0	0	3	4	39	5		0	0	39	70	5	6
1	0	0	0	5	2	40	4		0	0	40	65	4	5
1	1	1	0	5	5	68	9		1	1	68	89	9	7
1	1	1	1	5	1	71	8		1	1	71	64	8	4
1	0	1	0	3	1	98	4		0	0	98	82	4	5
1	1	0	1	3	1	80	7		1	1	80	82	7	4
1	0	1	1	4	2	42	8		0	1	42	75	8	7
1	0	0	0	3	3	39	6		0	1	39	80	6	5
1	0	1	0	5	4	55	6		0	0	55	83	6	6
1	1	0	0	4	2	74	6		1		74		6	
1	0	1	0	5	2	36	4		0		36		4	
1	1	1	1	5	2	49	7		1		49		7	
1	1	0	0	5	1	76	6		1		76		6	
1	0	0	0	3	1	92	4		0		92		4	
1	0	0	0	5	5	62	4		0		62		4	
1	0	0	0	5	4	54	7		0		54		7	
1	0	0	0	5	5	68	5		0		68		5	
1	1	0	1	5	5	60	6		1		60		6	
1	1	0	0	4	2	61	8		1		61		8	
0	0	1	0	1	1	58	6							
0	1	1	1	3	2	89	6							

D: Enable Data Analysis Toolpak

If "Data Analysis" is not shown on Data tab, install it.

- Excel 2010: From File menu, select Options/Add-Ins.
- Excel 2008: From Microsoft button, select "Add Excel Options" in lower right corner.
- Excel 2003 If "Data Analysis Toolpak" is not shown under Tools menu, install it. From TOOLS menu, select ADD-INS.

Check the check box for "Data Analysis Toolpak“. Press the OK command button. The Data Analysis command should be added to the appropriate menu. If not remove and reinstall.

T-TEST Command Procedure Given Separated Data

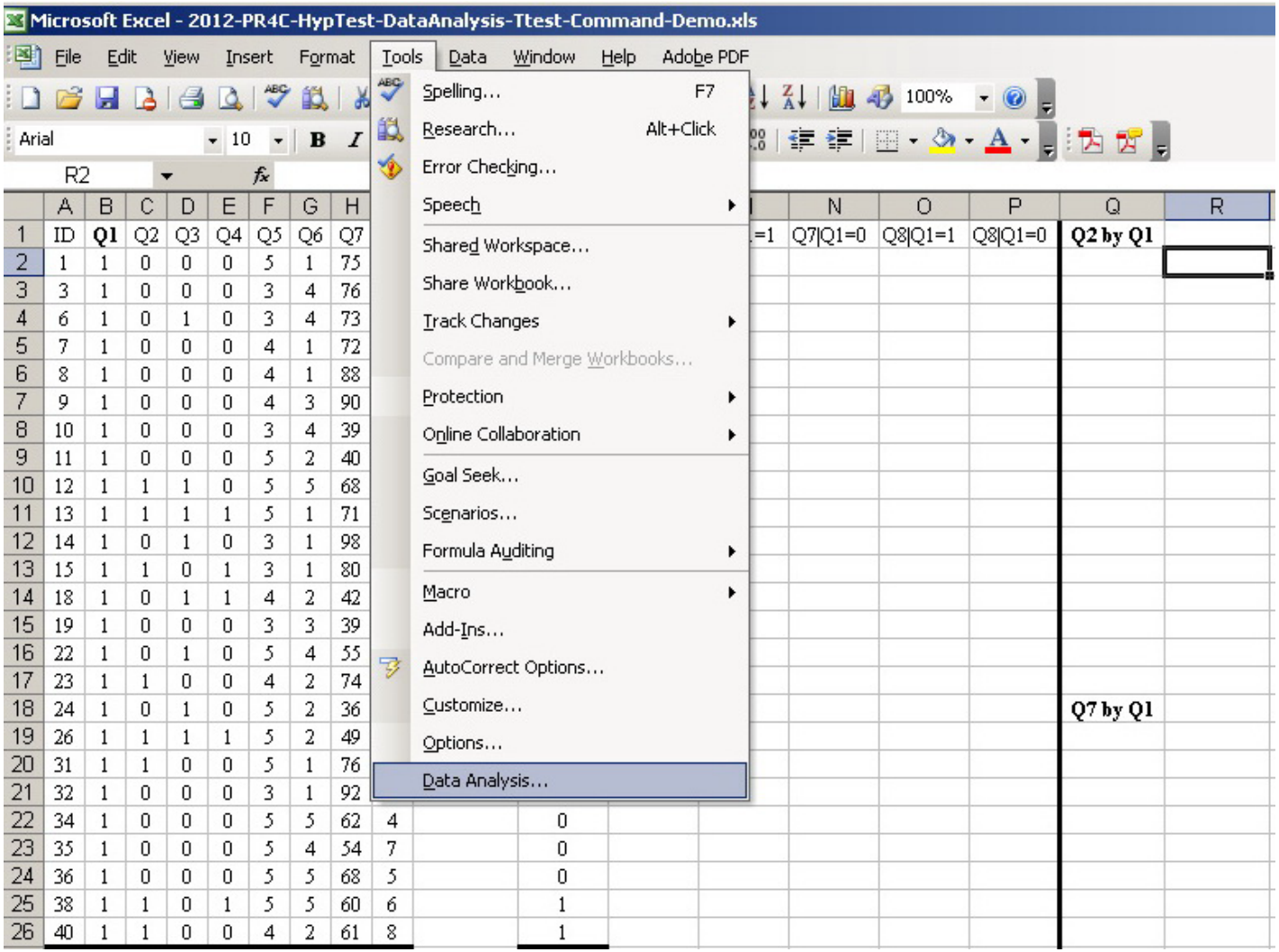
1: From Tool Menu, select “Data Analysis”.

2: From Data Analysis window, select “t-test: Two-sample with unequal variances”.

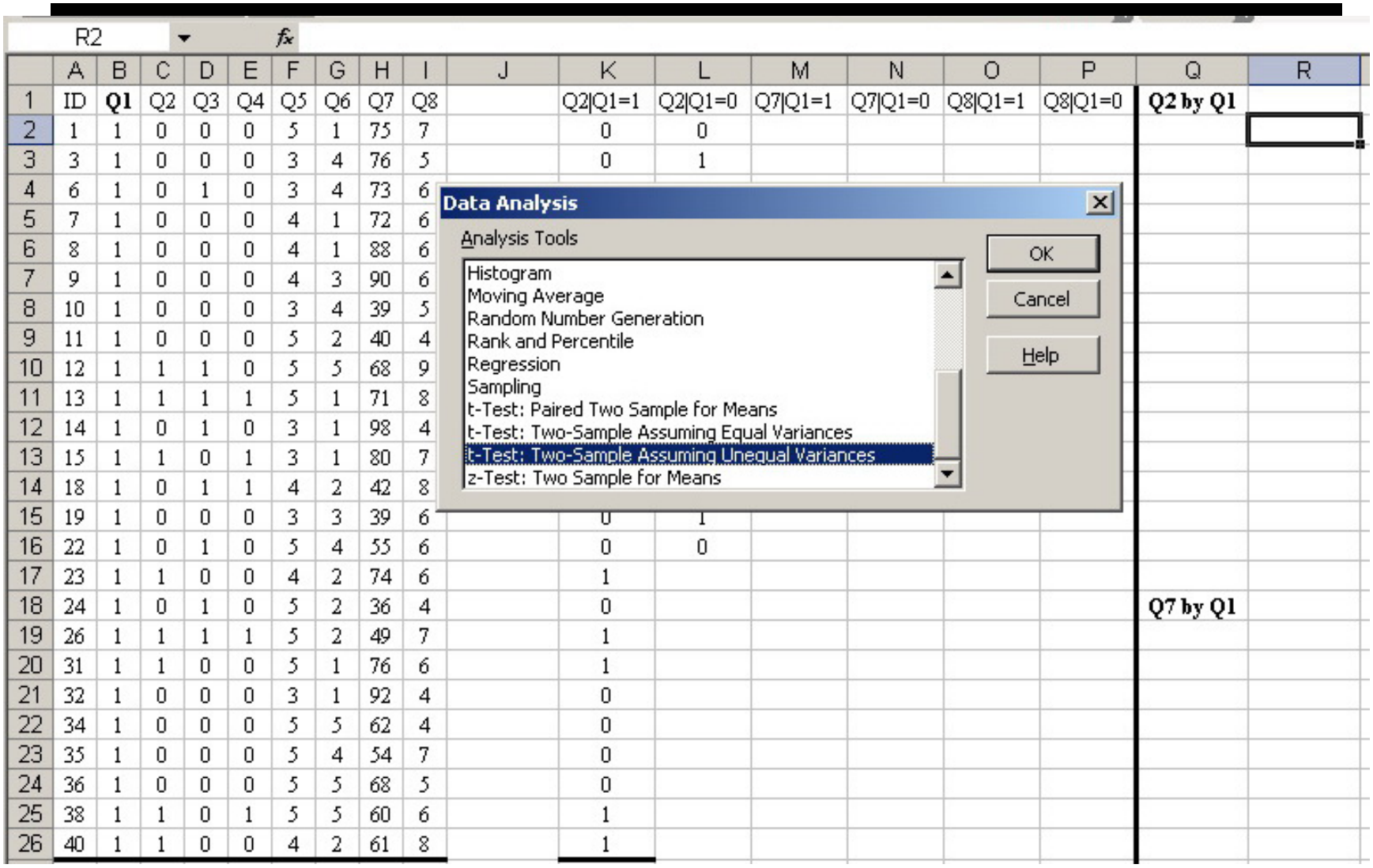
3: In T-Test window, enter input and output options.

4. Obtain results of t-test. Summarize the test results.

1) From the Data menu, select Data Analysis



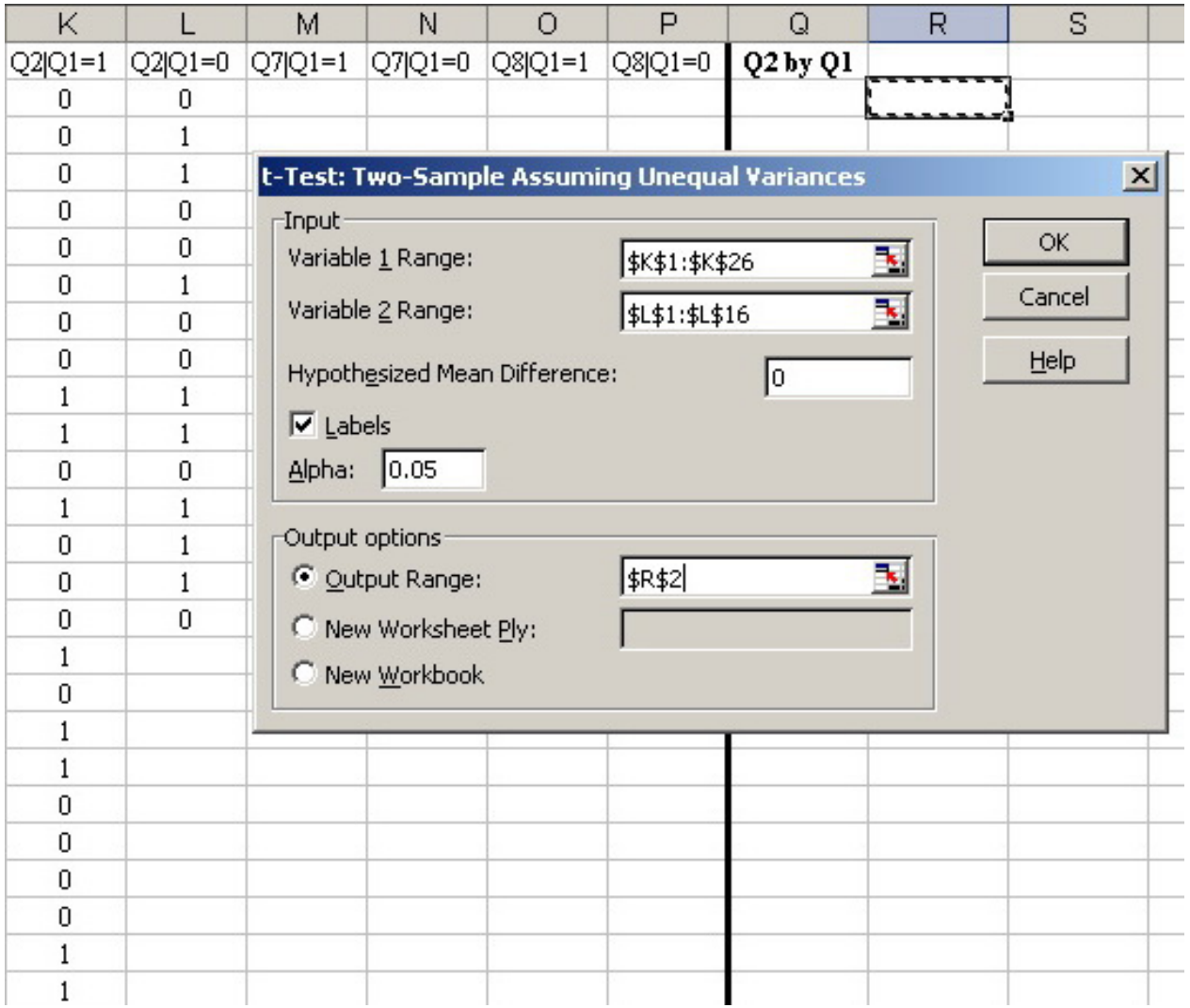
2) Select T-Test Command: Two-Sample; Unequal Variances



R2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	ID	**Q1**	Q2	Q3	Q4	Q5	Q6	Q7	Q8		Q2|Q1=1	Q2|Q1=0	Q7|Q1=1	Q7|Q1=0	Q8|Q1=1	Q8|Q1=0	**Q2 by Q1**	
2	1	1	0	0	0	5	1	75	7		0	0						
3	3	1	0	0	0	3	4	76	5		0	1						
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		0	1						
16	22	1	0	1	0	5	4	55	6		0	0						
17	23	1	1	0	0	4	2	74	6		1							
18	24	1	0	1	0	5	2	36	4		0						**Q7 by Q1**	
19	26	1	1	1	1	5	2	49	7		1							
20	31	1	1	0	0	5	1	76	6		1							
21	32	1	0	0	0	3	1	92	4		0							
22	34	1	0	0	0	5	5	62	4		0							
23	35	1	0	0	0	5	4	54	7		0							
24	36	1	0	0	0	5	5	68	5		0							
25	38	1	1	0	1	5	5	60	6		1							
26	40	1	1	0	0	4	2	61	8		1							

3) Enter Input & Output: Q2 by Q1



K	L	M	N	O	P	Q	R	S
Q2|Q1=1	Q2|Q1=0	Q7|Q1=1	Q7|Q1=0	Q8|Q1=1	Q8|Q1=0	**Q2 by Q1**		
0	0							
0	1							
0	1							
0	0							
0	0							
0	1							
0	0							
0	0							
1	1							
1	1							
0	0							
1	1							
0	1							
0	1							
0	0							
1								
0								
1								
1								
0								
0								
0								
0								
1								
1								

4) Excel-Generated Results: Q2 by Q1

t-Test: Two-Sample Assuming Unequal Variances		
	Q2|Q1=1	Q2|Q1=0
Mean	0.32	0.53
Variance	0.23	0.27
Observations	25	15
Hypothesized Mean Difference	0	
df	28	
t Stat	-1.30	
P(T<=t) one-tail	0.10	
t Critical one-tail	1.70	
P(T<=t) two-tail	0.20	
t Critical two-tail	2.05	

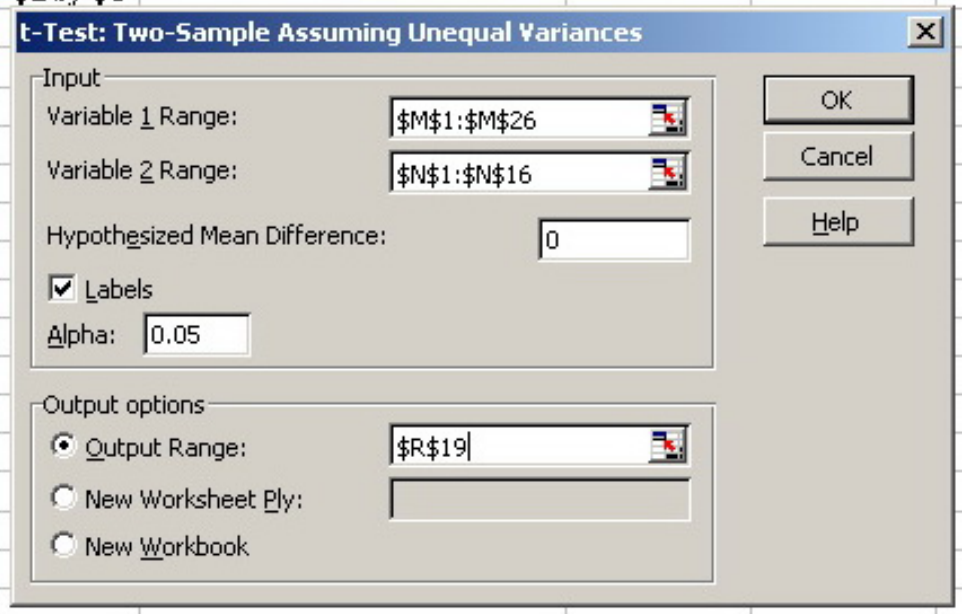
Difference in Q2 proportions by Q1 in not statistically significant

Technical: Fail to reject the null hypothesis for Q2 by Q1.

3) Enter Input and Output: Q7 by Q1

M	N	O	P
Q7|Q1=1	Q7|Q1=0	Q8|Q1=1	Q8|Q1=0
75	58	7	6
76	89	5	6
73	77	6	7
72	93	6	6
88	41	6	6
90	65	6	7
39	70	5	6
40	65	4	5
68	89	9	7
71	64	8	4
98	82	4	5
80	82	7	4
42	75	8	7
39	80	6	5
55	83	6	6
74		6	
36		4	
49		7	
76		6	
92		4	
62		4	
54		7	
68		5	
60		6	
61		8	

Q2 by Q1



Q7 by Q1

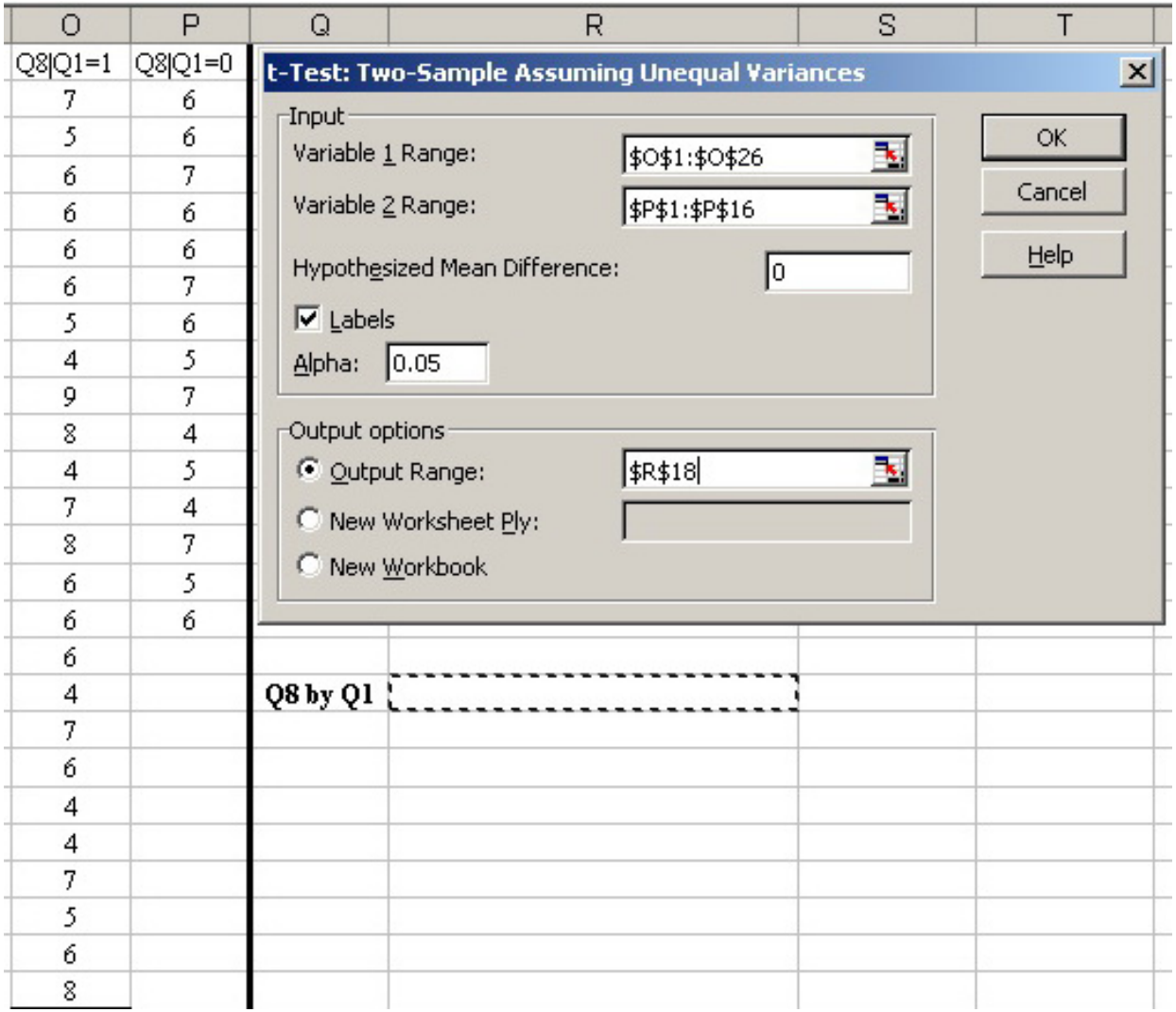
4) Excel-Generated Results: Q7 by Q1

Q7 by Q1			
	t-Test: Two-Sample Assuming Unequal Variances		
		Q7|Q1=1	Q7|Q1=0
	Mean	65.52	74.2
	Variance	318.93	192.03
	Observations	25	15
	Hypothesized Mean Difference	0	
	df	35	
	t Stat	-1.72	
	P(T<=t) one-tail	0.05	
	t Critical one-tail	1.69	
	P(T<=t) two-tail	0.09	
	t Critical two-tail	2.03	

Difference in Q7 means by Q1 IS statistically significant (1 tail)

Technical: Reject the null hypothesis for Q7 by Q1 (1 tail)

3) Enter Input and Output: Q8 by Q1



4) Excel-Generated Results: Q8 by Q1

Q8 by Q1	t-Test: Two-Sample Assuming Unequal Variances		
		Q8|Q1=1	Q8|Q1=0
	Mean	6.00	5.80
	Variance	2.00	1.03
	Observations	25	15
	Hypothesized Mean Difference	0	
	df	37	
	t Stat	0.52	
	P(T<=t) one-tail	0.30	
	t Critical one-tail	1.69	
	P(T<=t) two-tail	0.61	
	t Critical two-tail	2.03	

Difference in Q8 means by Q1 is NOT statistically significant.

Technical: Fail to reject the null hypothesis for Q8 by Q1

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 IS [or IS NOT] statistically significant" is a less-technical conclusion.

Use the less-technical expressions for everyday communication.