MATHEMATICS 201-510-LW
Business Statistics
Martin Huard
Fall 2008

Two Population Tests with Excel

Hypothesis testing for the mean of two dependent samples

Example 1
A company claims its 12-week exercise program significantly reduces weight. A random sample of 6 persons was selected and their weight (in kg) measured before and after the program.

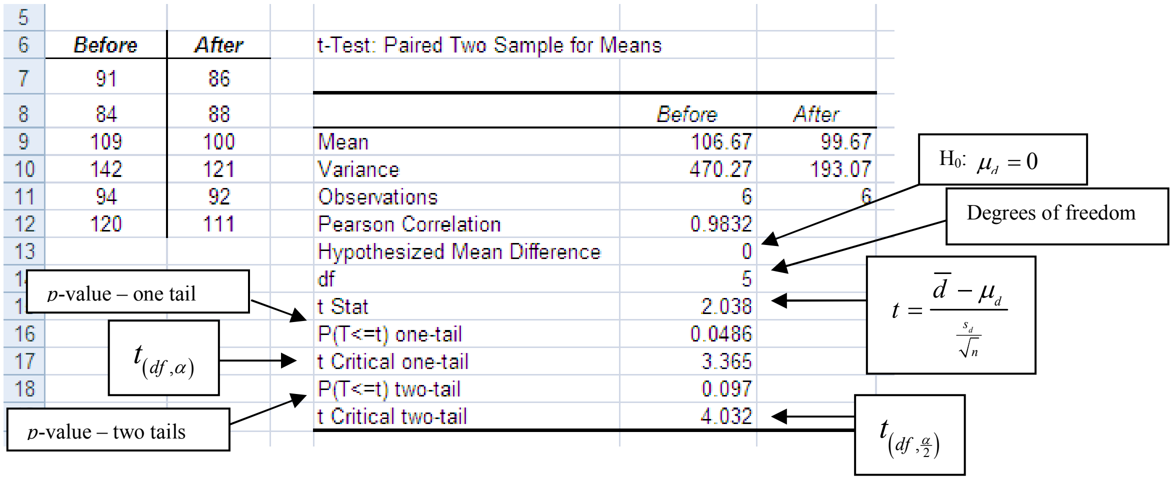
Before	After
91	86
84	88
109	100
142	121
94	92
120	111

At the 1% level of significance, can you conclude that the mean weight loss is greater than zero? Assume the populations are normally distributed. Use the classical approach.

Make the usual heading in cells A1:A4. In cell A6 write “Before” and in cell B6 “After”. Enter the above data in below these two headings, in cells A6:B11.
Go to DATA – DATA ANALYSIS – T-TEST: PAIRED TWO SAMPLE FOR MEANS to access the following dialogue box. If you cannot find DATA ANALYSIS, go to EXCEL OPTIONS, ADD-INS and click on ANALYSIS TOOL PACK. Here is the dialogue box that should appear.

For the VARIABLE 1 RANGE, we use the Before column, so cells A6:A12. For the VARIABLE 2 RANGE, we use the After column. The HYPOTHESIZED MEAN DIFFERENCE is 0 since the null hypothesis is $H_o: \mu_d = 0$. Click on LABELS. Use D6 for the OUTPUT RANGE, and make sure that you put 1% for ALPHA.

Once you have the results, adjust the columns widths accordingly and round some of the numbers. Your Excel worksheet should look like:



	Before	After	t-Test: Paired Two Sample for Means		
5					
6	Before	After	t-Test: Paired Two Sample for Means		
7	91	86			
8	84	88		Before	After
9	109	100	Mean	106.67	99.67
10	142	121	Variance	470.27	193.07
11	94	92	Observations	6	6
12	120	111	Pearson Correlation	0.9832	
13			Hypothesized Mean Difference	0	
14			df	5	
15			t Stat	2.038	
16			P(T<=t) one-tail	0.0486	
17			t Critical one-tail	3.365	
18			P(T<=t) two-tail	0.097	
19			t Critical two-tail	4.032	

You can then complete the 5 steps, making cell references in step 3 and step 4. Here is what your worksheet should look like, where the shaded cells are obtained with an appropriate cell reference.

6	Before	After	t-Test: Paired Two Sample for Means			Step 1	Assumptions: Population normally distributed			
7	91	86				Step 2	H0:	μ_d =	0	
8	84	88		Before	After		HA:	μ_d >	0	
9	109	100	Mean	106.67	99.67	Step 3	Test statistic: *t* with *df* =			5
10	142	121	Variance	470.27	193.07		Right-tailed Test		α =	5%
11	94	92	Observations	6	6		t critical =	3.365		
12	120	111	Pearson Correlation	0.9832		Step 4	t =	2.04		
13			Hypothesized Mean Difference	0		Step 5	*t* is not in the critical region			
14			df	5			Fail to reject Ho			
15			t Stat	2.038			There is not sufficient evidence, at the 5% level			
16			P(T<=t) one-tail	0.0486			of significance to conclude that student heights			
17			t Critical one-tail	3.365			are different than 1.7 m.			
18			P(T<=t) two-tail	0.097						
19			t Critical two-tail	4.032						

Hypothesis testing for the mean of two independent (small) samples

Example 2

A social psychologist was interested in sex differences in the sociability of teenagers. Using the number of good friends as a measure, he compared the sociability of eight female and seven male teenagers. Is there a difference with respect to sociability between teenage females and males? Use a 3% level of significance. Assume that the sociability of female and male teenagers are normally distributed. Use the *p*-value approach.

Females	Males
8	1
3	5
1	8
7	3
7	2
6	1
8	2
5	

As in the previous example, copy this table in cells A6:B14. Proceed as in the previous example, except use T-TEST: TWO-SAMPLE ASSUMING EQUAL VARIANCES in DATA ANALYSIS. Your results should look like this (where the relevant information was highlighted).

Females	*Males*
8	1
3	5
1	8
7	3
7	2
6	1
8	2
5	

t-Test: Two-Sample Assuming Equal Variances

	Females	*Males*
Mean	5.625	3.143
Variance	6.268	6.476
Observations	8	7
Pooled Variance	6.3640	
Hypothesized Mean Difference	0	
df	13	
t Stat	1.901	
P(T<=t) one-tail	0.0398	
t Critical one-tail	2.060	
P(T<=t) two-tail	0.0797	
t Critical two-tail	2.436	

Step 1	Assumptions:	Population normally distributed	
		Samples are independent	
		Variances are equal	
Step 2	H_0:	$\mu_F - \mu_M =$	0
	H_A:	$\mu_F - \mu_M \neq$	0
Step 3	Test statistic: *t* with *df* =		13
	Two-tailed Test with α =		5%
Step 4	t =	1.901	
	t critical =	0.0797	
Step 5	*p*-value > α		
	Fail to reject Ho		
	There is not sufficient evidence, at the 5% level of significance to conclude that there is a difference in sociability between teenage men and women		

$$s_p = \sqrt{\frac{(n_F - 1)s_F^2 + (n_M - 1)s_M^2}{n_F + n_M - 2}}$$

$$t = \frac{\overline{x}_F - \overline{x}_M}{s_p\sqrt{\frac{1}{n_F} + \frac{1}{n_M}}}$$

Hypothesis testing for the mean of two independent (large) samples

Example 3

A social sociologist was interested in sex differences in the number of books a person reads. Two random samples were taken, one of men and the other, and the number of books read during the last month was recorded. Is there a difference with respect to the number of books read by females and males? Use a 3% level of significance.

Females : 8 3 1 7 7 6 8 5 6 5 4 2 9 7 6 8 2 4 5 6 9 1 2 4 6 8 7 5 9 6
Male: 1 5 8 3 2 1 2 6 4 2 9 3 4 8 7 2 6 12 4 5 3 1 3 6 7 4 1 2 6 3 1 2 3

As in the previous example, copy this data in cells A6:B39. Before we use DATA – ANALYSIS, we need to find the variance of both populations. In cell D6 write “Variance (Females)” and in cell D7 “Variance (Males)”. Find the variance for both females and males in cells E6 and E7 using the VAR function. The reason we need this is that this test only works if either the variance is known, or if the sample is large enough that we can use the approximation $\sigma \approx s$, which is our case here.

We can now proceed as earlier by going to TOOLS – DATA ANALYSIS and choosing Z-TEST TWO SAMPLE FOR MEANS. Fill in the information appropriately. Your results should look like this (where the relevant information was highlighted).

Females	*Males*
8	1
3	5
1	8
7	3
7	2
6	1
8	2
5	6
6	4
5	2
4	9
2	3
9	4
7	8
6	7
8	2

Variance (Females): s_F^2 =	5.637
Variance (Males): s_M^2 =	7.235

z-Test: Two Sample for Means

	Females	*Males*
Mean	5.533	4.121
Known Variance	5.637	7.235
Observations	30	33
Hypothesized Mean Difference	0	
z	2.21	
P(Z<=z) one-tail	0.0134	
z Critical one-tail	1.88	
P(Z<=z) two-tail	0.0269	
z Critical two-tail	2.17	

Step 1	Assumptions:	Samples are independent		
		n_F =	30	≥ 30
		n_M =	33	≥ 31
Step 2	H_0:	$\mu_F - \mu_M$ =	0	
	H_A:	$\mu_F - \mu_M \neq$	0	
Step 3	Test statistic: z			
	Two-tailed Test with α =		5%	
	z critical =	2.170		
Step 4	z =	2.21		
Step 5	z is in the critical region			
	Reject Ho			
	There is sufficient evidence, at the 5% level of significance to conclude that there is a difference between the number of books read by men and women.			

$$z = \frac{(\bar{x}_F - \bar{x}_M) - (\mu_F - \mu_M)}{\sqrt{\frac{s_F^2}{n_F} + \frac{s_M^2}{n_M}}}$$