

## EXPECTED VALUES AND REPEATED MEASURES: DEMO NOTES

### Chi-square test of independence

File: housing.xlsx

Is there a relationship between homes that have air conditioning and homes that have a full basement?

1. Insert a PivotTable from the source data. Place airco in the Columns section, fullbase in the Rows section, and count of id in the Values section.

The screenshot shows an Excel spreadsheet with a PivotTable and the PivotTable Fields task pane. The PivotTable is located in the range O1:P7 and shows the count of 'id' for different combinations of 'airco' (Columns) and 'fullbase' (Rows). The data is as follows:

Count of id	airco no	airco yes	Grand Total
fullbase no	248	107	355
fullbase yes	125	66	191
Grand Total	373	173	546

The PivotTable Fields task pane is open on the right side of the screen. It shows the following configuration:

- Choose fields to add to report:** A list of fields including 'id', 'price', 'lotsize', 'bedrooms', 'bathrms', 'stories', 'driveway', 'recroom', 'fullbase', and 'gashw'. The 'fullbase' field is checked.
- Drag fields between areas below:**
  - Filters:** (Empty)
  - Columns:** 'airco' is selected.
  - Rows:** 'fullbase' is selected.
  - Values:** 'Count of id' is selected.
- Defer Layout Update:** (Unchecked)
- Update:** (Button)

2. These are our *actual* values. We will now calculate the *expected* values based on overall proportions. Copy and paste the data from the PivotTable and add formulas to sum the row and column totals.
  - a. It can also be helpful to relabel the row and column headers to know exactly which category is which.

R11     $\times$      $\checkmark$      $f_x$      $=SUM(R9,R10)$

	N	O	P	Q	R
1		Count of id	Column Labels		
2		Row Labels	ac-no	ac-yes	Grand Total
3		fullbase-no	248	107	355
4		fullbase-yes	125	66	191
5		Grand Total	373	173	546
6					
7		Count of id	Column Labels		
8		Row Labels	ac-no	ac-yes	Grand Total
9		fullbase-no	248	107	355
10		fullbase-yes	125	66	191
11		Grand Total	373	173	546

3. We will now calculate what would be expected for each category based on a straight proportion.

P9     $\times$      $\checkmark$      $f_x$      $=($R$3*$P$5)/$R$5$

	N	O	P	Q	R	S
1		Count of id	Column Labels			
2		Row Labels	ac-no	ac-yes	Grand Total	
3		fullbase-no	248	107	355	
4		fullbase-yes	125	66	191	
5		Grand Total	373	173	546	
6						
7		Count of id	Column Labels			
8		Row Labels	ac-no	ac-yes	Grand Total	
9		fullbase-no	242.518315	112.481685	355	
10		fullbase-yes	130.481685	60.51831502	191	
11		Grand Total	373	173	546	
12						



4. We will now find the p-value of the chi-square test using the Excel function CHISQ.TEST(). The function takes two arguments: the first is the range of *actual* values (which come from the PivotTable), and the second from the *expected* (which we calculated).
  - a. In this case, since the p-value exceeds .05, we fail to reject the null. There is no significant relationship between these categories.

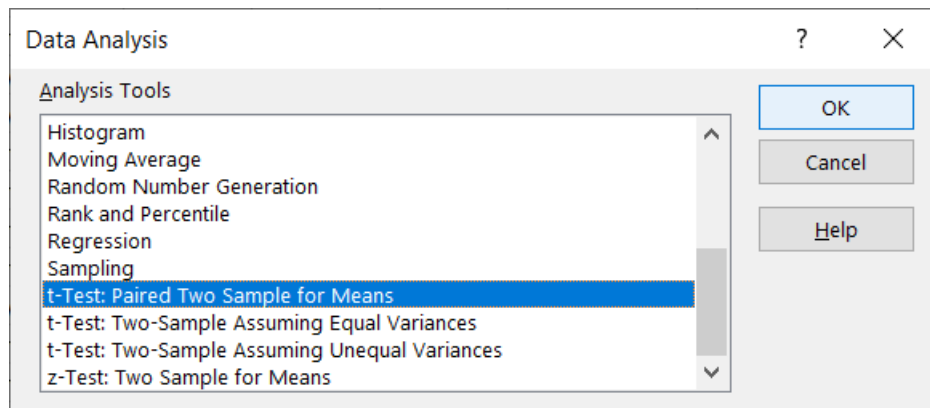
	N	O	P	Q	R
1		Count of id	Column Labels		
2		Row Labels	ac-no	ac-yes	Grand Total
3		fullbase-no	248	107	355
4		fullbase-yes	125	66	191
5		Grand Total	373	173	546
6					
7		Count of id	Column Labels		
8		Row Labels	ac-no	ac-yes	Grand Total
9		fullbase-no	242.518315	112.481685	355
10		fullbase-yes	130.481685	60.51831502	191
11		Grand Total	373	173	546
12					
13		0.290378605	=CHISQ.TEST(P3:Q4,P9:Q10)		
14					

## Paired sample t-test

Demo: bp.xlsx

Is there a difference in blood pressures after the intervention?

1. For this test, we will use the Data Analysis ToolPak. Find the t-test: Paired Two Sample for Means option.



- Select the measurements for Time 1 and Time 2. The Hypothesized Mean Difference will set to zero by default, which is what we want.

t-Test: Paired Two Sample for Means

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

☒ Labels

Alpha:

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

- Based on the p-value for the two-tailed t-test, we reject the null. Blood pressures are significantly lower on average after the intervention.

t-Test: Paired Two Sample for Means		
	<i>bp_before</i>	<i>bp_after</i>
Mean	156.45	151.3583333
Variance	129.7285714	201.004972
Observations	120	120
Pearson Correlation	0.159118103	
Hypothesized Mean Difference	0	
df	119	
t Stat	3.337187051	
P(T<=t) one-tail	0.000564896	
t Critical one-tail	1.657759285	
P(T<=t) two-tail	0.001129791	
t Critical two-tail	1.980099876	



## Wilcoxon signed-rank test

Demo: cortisol.xlsx

For this test, we will multiply the sign of each observation (+1 or -1) by the relative ranking of its absolute value, ranked ascending.

1. Enter the following formulas for columns D-F:

D	Sign	=SIGN(B3-C3)
E	Absolute value	=ABS(B3-C3)
F	Signed-rank	=RANK.AVG(E3,\$E\$3:\$E\$22,1)*D3

2. Now we take the sum of all positive and negative ranks, and find their absolute values:

	H	I	J	K	L	M	N
1							
2		Sum ranks	Absolute value				
3	Positive sum (Sum of all positive signed-ranks)	197	197	=SUMIF(\$F\$3:\$F\$22,">0")	=ABS(I3)		
4	Negative sum (Sum of all negative signed-ranks)	-13	13	=SUMIF(\$F\$3:\$F\$22,"<0")	=ABS(I4)		

3. Our test statistic is whatever of these two values is smaller, and our critical value is found from a lookup of the critical-values worksheet based on our sample size.

	H	I	J	K	L	M
1						
2		Sum ranks	Absolute value			
3	Positive sum (Sum of all positive signed-ranks)	197	197	=SUMIF(\$F\$3:\$F\$22,">0")	=ABS(I3)	
4	Negative sum (Sum of all negative signed-ranks)	-13	13	=SUMIF(\$F\$3:\$F\$22,"<0")	=ABS(I4)	
5						
6	Test statistic	13	=MIN(J3:J4)			
7	Critical value	52	=VLOOKUP(COUNT(\$A\$3:\$A\$22),'critical-values'!\$A\$1:\$B\$22,2,FALSE)			
8						



4. If the test statistic is *less than* the critical value, we reject the null.

	H	I	J	K	L	M
1						
2		Sum ranks	Absolute value			
3	Positive sum (Sum of all positive signed-ranks)	197	197	=SUMIF(\$F\$3:\$F\$22,">0")	=ABS(I3)	
4	Negative sum (Sum of all negative signed-ranks)	-13	13	=SUMIF(\$F\$3:\$F\$22,"<0")	=ABS(I4)	
5						
6	Test statistic	13	=MIN(J3:J4)			
7	Critical value	52	=VLOOKUP(COUNT(\$A\$3:\$A\$22),'critical-values'!\$A\$1:\$B\$22,2,FALSE)			
8	Reject the null?	Yes	=IF(I6<I7,"Yes","No")			
9						
10						

