In [1]:

```
import sys
import importlib as imp
if ('Jupytils' in sys.modules):
    reloaded = imp.reload(Jupytils)
else:
    import Jupytils
```

Chi Square test

Chi-square (χ^2) test is used to test null-hypothesis if categorical observed values are same as expected frequencies. The following conditions are met:

- The variable under study is categorical.
- The expected value of the number of sample observations in each level of the variable is at least 5.

This approach consists of four steps: 1) state the hypotheses, 2) formulate an analysis plan, 3) analyze sample data, and 4) interpret results.

Chi-square (χ^2) test is used on two scenarios

- **Goodness of FIT**: How observed frequency fits the expected frequency
- **Test for Independence**: If variables have no influence on each other (i.e. they are independent)

More example to follow

REFERENCES:

- http://math.hws.edu/javamath/ryan/ChiSquare.html (http://math.hws.edu/javamath/ryan/ChiSquare.html)

Goodness of FIT Example

Problem: In order to climb Mount Chesta we need to make some decisions on hiring a guide company.

A guide company claims that 41% succesful attempts last year. It is also known that 33% of 15,000 attempts each year are succefful.

Summit up: 41% succesful attempts of 100 attempt
Generally: 33% succesful attempts of 15,000 attempt

	Successful	unsuccessful	Totals
Expected	33	67	100
Observed	41	59	100
Totals	74	126	200

We want to check if we have a better chance of climbing by hiring the guide company.

We want to check how well the observed proportions fit the population proportions (or expected proportions).!

H0 (Null): expected == observed (i.e. no difference)

H1 (Alternative): Null hypothesis is not true

CHI-square

$\chi^2 = \sum_{k=1}^{n} \frac{(f_o - f_e)^2}{f_e}$

χ^2 Value is smaller when observed value is closer to the expected value.

$\frac{(41-33)^2}{33} + \frac{(59-67)^2}{67} = 2.89$

Degrees of Freedom (df) = 1; (This will be the number of categories minus one)

$p = 0.0889 > 0.05$ - we fail to reject the null-hypothesis and state that *guide company* is not worth hiring. There is about 9% chance we could be wrong and these differences are due to chance

In [2]:

```
A = stats.chisquare([41,59] , [33,67])
Ho = "FAIL TO REJECT: *There is no difference: we fail to reject null hypothesis"
Chi2_c = stats.chi2.ppf(.95, 1);
if (A[1] < 0.05):
    Ho = "We can reject Ho: there is a difference"
print (A, "\n", Ho, "\nChi2 Critical: ", Chi2_c, " Chi2 Value: ", A[0])
print (1-stats.chi2.cdf(2.89,1))
```

```
Power_divergenceResult(statistic=2.8946178199909545, pvalue=0.08887585044058065)
 FAIL TO REJECT: *There is no difference: we fail to reject null hypothesis
Chi2 Critical:  3.841458820694124  Chi2 Value:  2.8946178199909545
0.08913092551708635
```

Test for independece

Chi-square (χ^2) for independece whether or not "two variabled are independent".

Chi-square test for **goodness of FIT** captures data for in two rows. For example: *Which season do you prefer*, in that case, we may have 4 responses *"Summer, Winter, Spring, Fall"

	Response1	**Response2**	**Response3**
Treatment1			
Treatment2			

–

Chi-square (χ^2) also helps for *independece* whether or not "two variables are independent". For example, in above table, we want to know if Response1, Response2, Response3 are independent of Treatment1 or Treatment 2.

In this case, instead of having observed and expected in two rows, we will looking at number of participants (or subjects) fall into variable1 and variable2.

Example

Problem: We want to know if wording of a question influences the answer.

In an experiement, 150 students from University of Washington were shown one minute film clip of a car accident. The students were separated into three equal groups each student chosen randomly into any of the three group and thus each group consisting of 50 students. After the clip was shown:

> **Group 1** was asked : *How fast were the cars going when they* *hit** each other

> **Group 2** was asked : *How fast were the cars going when they* *smashed into** each other

> **Group 3** was asked : *How fast were the cars going* (No question about the speed of the car)

After one week, all students were asked *"Did you see any broken glass"* and the responses were noted as below

> - Group1 *: responded 7 out of 50 answered yes!
> - Group2 *: responded 16 out of 50 answered yes!
> - Group3 *: responded 6 out of 50 answered yes!

This is summarized in the following table:

	HIT	smashed	Control	Total
YES	7	16	6	29
NO	43	34	44	121
TOTAL	50	50	50	150

In this case, 29/150 said *YES*, 121/150 said *NO*. The expected "YES" response is thus $\frac{29}{150}$ for "YES" group; $\frac{121}{150}$ for "NO" group.

Now we can add the expected response to each cell. In HIT group consisting of 50 students, we expect $\frac{29}{150} * 50 = 9.67$ to say YES (and same is true for each group since there are same number of students in each group). Similarly we expect $\frac{121}{150} * 50 = 40.33$ to say "NO" in each group. Lets update the table with the expected values as follows:

–

	HIT	smashed	Control	Total
YES	7 **9.67**	16 **9.67**	6 **9.67**	29
NO	43 **40.33**	34 **40.33**	44 **40.33**	121
TOTAL	**50**	50	50	150

$\chi^2 = \sum_{k=1}^{n} \frac{(f_o - f_e)^2}{f_e}$

χ^2 Value is smaller when observed value is closer to the expected value.

$$\frac{(7-9.67)^2}{9.67} + \frac{(16-9.67)^2}{9.67} + \frac{(6-9.67)^2}{9.67} + \frac{(43-40.33)^2}{40.33} + \frac{(34-40.33)^2}{40.33} + \frac{(44-40.33)^2}{40.33} = 7.7779$$

Degrees of Freedom is 2 ((number of groups -1) * (number of responses -1))

From the Chi-square table: https://people.richland.edu/james/lecture/m170/tbl-chi.html (https://people.richland.edu/james/lecture/m170/tbl-chi.html)
we see the critical value is 5.991. Now Chi2 is 7.777 >> 5.9991 and p-value as calculated from the program below is 0.02 << 0.05.

We reject the null-hypothesis and state that the "**word**" "*SMASHED*" had a difference!

####strength of the relationship:

When we have a contigency table greater than 2x2, we can use * Cramer's V* $(\phi_c) = \sqrt{\frac{\chi^2}{n(k-1)}}$ Where K is the smaller of number of rows or columns: in this case, it is number of rows = 2; n is the total number of subjects = 100

Therefore * Cramer's V* $(\phi_c) = \sqrt{\frac{7.77^2}{100(2-1)}} = .227$

From Cramers V table: for k-1 = 1 indicates small effect!

-

In [3]:

```python
chi2, p, ddof, expected = stats.chi2_contingency( [[7,16,6], [43,34,44]] )
Ho = "KEEP Null Hyp: There is no difference: we fail to reject null hypothesis"
if (p < 0.05):
    Ho = "REJECT H0: there is an effect"

Chi2_c = stats.chi2.ppf(.95, 2);
n = 150;
k = 2;
cramers_v = sqrt(chi2/ ( n * (k-1)))
print ("Chi stat: ", chi2, " Chi Critical: ", Chi2_c, "\np=", p, " ddof: ", ddof, " expected:\n",

# Another way to compute p using CDF
print ("Finding p another way: ", 1-stats.chi2.cdf(chi2,2))

print (" Cramers V: ", cramers_v)
```

```
Chi stat:  7.779994300370478  Chi Critical:  5.991464547107979
p= 0.020445404303469611  ddof:  2  expected:
 [[ 9.67  9.67  9.67]
 [40.33 40.33 40.33]]
 REJECT H0: there is an effect
Finding p another way:  0.02044540430346964
 Cramers V:  0.22774246127838463
```

There are 110 houses in a particular neighborhood.

- Liberals live in 25 of them,
- moderates in 55 of them, and
- conservatives in the remaining 30.

An airplane carrying 65 lb. sacks of flour passes over the neighborhood. For some reason, 20 sacks fall from the plane, each miraculously slamming through the roof of a different house. None hit the yards or the street, or land in trees, or anything like that. Each one slams through a roof. Anyway, 2 slam through a liberal roof, 15 slam through a moderate roof, and 3 slam through a conservative roof.

Null Hypothesis: Sacks of flour hit houses at random?

Should we reject the hypothesis?

–

```
Given the numbers of liberals, moderates and conservative households, we can calculate the
expected number of sacks of flour to crash through each category of house:

20 sacks x 25/110 = 4.55 liberal roofs smashed
20 sacks x 55/110 = 10.00 moderate roofs smashed
20 sacks x 30/110 = 5.45 conservative roofs smashed
Set up the table for the goodness-of-fit test:

Category     Observed     Expected     Obs-Exp     (Obs-Exp)^2 / Exp
Liberal            2         4.55        -2.55          1.43
Moderate          15        10.00         5.00          2.50
Conservative       3         5.45        -2.45          1.10
Total             20        20.00         0             5.03

In a simple test like this, where there are three categories and where the expected values
are not influenced by the observed values, there are two degrees of freedom. Checking the
table of critical values of the chi-square distribution for 2 d.f., we find that 0.05 < p.

That is, there is greater than a 5% probability of getting at least this much departure be
tween observed and expected results by chance.
Therefore, while it appears that moderates have had worse luck than liberals and conservat
ives, we **CANNOT REJECT** the NULL hypothesis - which means sacks struck houses at rando
m.
```

In [4]:

```
observed= [2,15,3]
expected=[25*20./110,55.*20/110,30.*20/110 ]

Chi2_c = stats.chi2.ppf(.95, 2);

A = stats.chisquare(observed, expected)
print ("Chi Critical: ", Chi2_c, " Ch2: ", A[0], " p-value: ", A[1])
print ("Observed/Expected: ", (observed, expected))
```

```
Chi Critical:  5.991464547107979  Ch2:  5.03  p-value:  0.08086291220670366
Observed/Expected:  ([2, 15, 3], [4.545454545454546, 10.0, 5.454545454545454])
```

Another Example

Example from: http://stattrek.com/chi-square-test/independence.aspx?Tutorial=AP

```
                      Voting Preferences              Row total
             Republican      Democrat    Independent
Male              200           150           50          400
Female            250           300           50          600
Column total      450           450          100          1000
```

H0: Gender and voting preferences are independent. Ha: Gender and voting preferences are not independent.

-

Interpret results. Since the P-value (0.0003) is less than the significance level (0.05), we cannot accept the null hypothesis. Thus, we conclude that there is a relationship between gender and voting preference.

In [5]:

```
house = [ [ 200,150,50], [ 250,300,50 ] ]
chi2, p, ddof, expected = stats.chi2_contingency( house )
msg = "Test Statistic: {}\np-value: {}\nDegrees of Freedom: {}\n"
print( msg.format( chi2, p, ddof ) )
print( expected )
```

```
Test Statistic: 16.203703703703702
p-value: 0.0003029775487145488
Degrees of Freedom: 2

[[180. 180.  40.]
 [270. 270.  60.]]
```

In [ ]:

```
```

-