Hypothesis Testing Demo

December 14, 2017

The purpose of this demo document is to show the capabilities of Python towards performing hypothesis testing. For a test example, we'll look at a hypothetical scenario where the effect of ascorbic acid is evaluated against the common cold. The data is based on a 1961 experiment where 279 French skiers were given either ascorbic acid or a placebo. After two weeks, they were examined to determine if they still had symptoms of the cold.

Treatment	Cold	No Cold	Total
Placebo	31	109	140
Ascorbic Acid	17	122	139
Total	48	231	279

0.1 Defining the Null and Alternative Hypothesis

To define the null and alternative hypotheses, we must frame them off of what we are looking to investigate. As defined above, we want to see how effective Ascorbic Acid is in preventing cold symptoms (link between ascorbic acid and cold symptoms). Because this is a case-control experiment, we'll define the case as those that have cold symptoms and the control as those without. Let P_1 be the amount of Ascorbic Acid user that are afflicted with the cold and P_2 without.

- $H_0 : P_1 = P_2$
- $H_1 : P_1 \neq P_2$

0.2 Generate Expected Distribution

```
In [60]: # Observed Distribution Array
         obs = {
             "a" : 17, ## Ascorbic Acid Treatment, Cold Symptoms
             "b" : 31, ## Placebo Treatment, Cold Symptoms
             "c" : 122, ## Ascorbic Acid Treatment, No Cold Symptoms
             "d" : 109 ## Placebo Treatment, No Cold Symptoms
         }

         n = sum(obs.values()) ## Sum of participants

         aEXP = (obs['a'] + obs['b']) * (obs['a'] + obs['c']) / n
         bEXP = (obs['a'] + obs['b']) * (obs['b'] + obs['d']) / n
```

```
cEXP = (obs['c'] + obs['d']) * (obs['a'] + obs['c']) / n
dEXP = (obs['c'] + obs['d']) * (obs['b'] + obs['d']) / n

## Expected Distribution Array
exp = {
    "Ascorbic Acid Treatment, Cold Symptoms" : aEXP,
    "Placebo Treatment, Cold Symptoms" : bEXP,
    "Ascorbic Acid Treatment, No Cold Symptoms" : cEXP,
    "Placebo Treatment, No Cold Symptoms" : dEXP
}

import pprint
pprint.PrettyPrinter(indent = 3).pprint(exp)
```

```
{  'Ascorbic Acid Treatment, Cold Symptoms': 23.913978494623656,
   'Ascorbic Acid Treatment, No Cold Symptoms': 115.08602150537635,
   'Placebo Treatment, Cold Symptoms': 24.086021505376344,
   'Placebo Treatment, No Cold Symptoms': 115.91397849462365}
```

0.3 χ^2 Distribution

```
In [61]: arrVals = [
             abs(obs['a'] - aEXP) ** 2 / aEXP,
             abs(obs['b'] - bEXP) ** 2 / bEXP,
             abs(obs['c'] - cEXP) ** 2 / cEXP,
             abs(obs['d'] - dEXP) ** 2 / dEXP,
         ]

         print("Chi-squared value: " + str(sum(arrVals)))
```

```
Chi-squared value: 4.81141264632079
```

P-value (with 1 DF)

```
In [62]: from scipy import stats
         pValueChiSq = 1 - stats.chi2.cdf(sum(arrVals) , 1)

         print("P-value: " + str(pValueChiSq))
```

```
P-value: 0.0282718602468
```

Based on the p-value from the χ^2 test ($p < 0.05$), we can reject the null hypothesis that there is not a difference between Ascorbic Acid and placebo treatments.

0.4 Fisher's Exact Test

```
In [63]: oddsRatio, pValueFET = stats.fisher_exact([
             [obs['a'], obs['c']],
             [obs['b'], obs['d']]
         ], alternative = "less")

         print("P-value: " + str(pValueFET))
```

```
P-value: 0.0205227159928
```

Much like the χ^2 test, we can reject the null hypothesis that there is not a difference between Ascorbic Acid and placebo treatments.

Difference from χ^2 test ($|\chi^2$ p-value - Fisher's p-value$|$)

```
In [64]: print(abs(pValueChiSq - pValueFET))
```

```
0.00774914425407
```

0.5 Relative Risk and Odds Ratio

For the purpose of clarity, we'll define the following variables:

- Exposed: treatment by Ascorbic Acid
- Nonexposed: treatment by Placebo
- Disease: presence of cold symptoms after 2 weeks
- Nondisease: non-presence of cold symptoms

We will also redefine the variables 'a', 'b', 'c', & 'd' as follows:

- a = Exposed group with diseased outcome (17)
- b = Exposed group with non-diseased outcome (122)
- c = Nonexposed group with diseased outcome (31)
- d = Nonexposed group with non-diseased outcome (109)

Relative Risk Ratio

```
In [65]: import math

         a = obs['a']
         b = obs['c']
         c = obs['b']
         d = obs['d']

         ## Relative Risk Ratio
         rr = (a / (a + b)) / (c / (c + d))
```

```
## Standard Error of ln(Relative Risk)
se = math.sqrt(1 / a + 1 / c - 1 / (a + b) - 1 / (c + d))

## 95% Confidence Interval
lowBound = math.exp(math.log(rr) - 1.96 * se)
highBound = math.exp(math.log(rr) + 1.96 * se)

## Standard Normal Deviate
zScore = math.log(rr) / se

## Two-tailed P-value
pValue = stats.norm.sf(abs(zScore)) * 2

print("Relative Risk Ratio: " + str(rr))
print("95% Confidence Interval: ["
      + str(lowBound) + ", " + str(highBound) + "]")
print("P-value: " + str(pValue))
```

```
Relative Risk Ratio: 0.5523323276862382
95% Confidence Interval: [0.32091461822682626, 0.9506298026962117]
P-value: 0.0321321040704
```

Because the Relative Risk ratio is 0.55 (< 1.0) at a 95% CI, we can say that (based on the RR ratio) there is a decreased chance that a risk exists between taking Ascorbic Acid and developing cold symptoms. This is further evidenced by the reported CIs [0.3209, 0.9506] not breaching 1.0.

Odds Ratio

```
In [66]: odds = (a * d) / (b * c)

         ## Standard Error of ln(Odds Ratio)
         se = math.sqrt(1 / a + 1 / b + 1 / c + 1 / d)

         ## 95% Confidence Interval
         lowBound = math.exp(math.log(odds) - 1.96 * se)
         highBound = math.exp(math.log(odds) + 1.96 * se)

         ## Standard Normal Deviate
         zScore = math.log(odds) / se

         ## Two-tailed P-value
         pValue = stats.norm.sf(abs(zScore)) * 2

         print("Odds Ratio: " + str(odds))
         print("95% Confidence Interval: ["
               + str(lowBound) + ", " + str(highBound) + "]")
         print("P-value: " + str(pValue))
```

```
Odds Ratio: 0.48995240613432045
95% Confidence Interval: [0.25693886701875085, 0.9342820066973032]
P-value: 0.030279474415
```

Much like the Relative Risk ratio above, we can say that there is a decreased chance that a risk exists between Ascorbic Acid and cold symptoms (based on the ratio = 0.49 and the confidence interval at 95% [0.2569, 0.9342] not breaching 1.0).