Table of Contents

Description of problem

Suppose a random sample of 100 US companies taken in 2015 showed that 21 offered high-deductible health insurance plans to their workers. A separate random sample of 120 firms taken in 2016 showed that 30 offered high-deductivle health insurance plans to their workers. Based on the sample results, can you conclude that there is a **higher proportion** of U.S. companies offering high-deductible health insurance plan to their workers **in 2016 than in 2015**? Conduct your hypothesis test at a level of significance alpha = 0.05.

Extraction of information

subject of interest = high-deductible health insurance plans
Random sample 2015 size = 100;
Observed subjects of interest = 21;

Random sample 2016 size = 120;
Observed subjects of interest = 30;
With an alpha of 0.05

Resolution of the problem

Steps for resolution

1. Establish hypotheses
2. Determine testing method
3. Establish rejection criteria
4. Do calculations

Establish hypotheses

The z test statistic is based on the difference between two sample proportions left parenthesis p 1 minus p 2 right parenthesis and the hypothesized value of the difference of the associated population proportions left parenthesis p 1 minus p 2 right parenthesis. Subtracting p 2 from both sides of the null and alternative hypotheses yields the formulations given below.

Note that p 1 and p 2 are the sample proportions of observations with the characteristic of interest in samples taken from populations 1 and 2, respectively; and p 1 and p 2 are the proportions of members with the characteristic of interest in populations 1 and 2, respectively.

Null Hypothesis: p2015 <= p2016
Alternative hypothesis: 2015 > p2016
With an alpha of 0.05

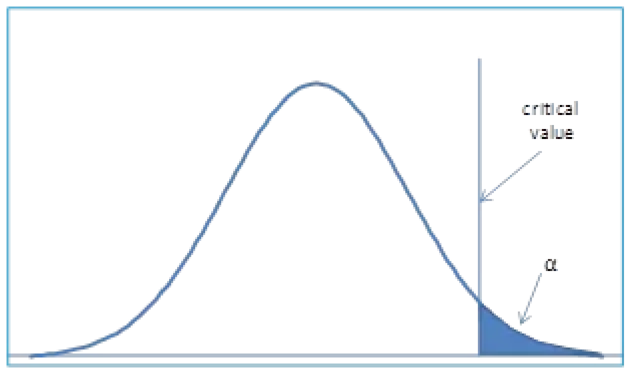
$$H_0: p_1 - p_2 \leq 0$$
$$H_A: p_1 - p_2 > 0$$

Null Hypothesis: p2015 - p2016 <= 0
Alternative hypothesis: p2015 - p2016 > 0

With an alpha of 0.05

Determine testing method

Right tail on alpha of 0.05



Establish rejection criteria

Reject the null hypothesis of the p-value is less than 0.05 alpha

Do calculations

In order to calculate z

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

We need to follow the next steps:

1. Set the variables given by the problem
2. Get sample proportions and review if both samples follow a normal distr
3. Calculate Standard Error of the Difference
   - Calculate p-bar of both populations
4. Calculate z

Set the variables given by the problem and review if both samples follow a normal distr

In [1]:
```
def setVariables(n1,obs1,n2,obs2, alpha):
    n1,obs1,n2,obs2, alpha = n1,obs1,n2,obs2, alpha
    return n1,obs1,n2,obs2, alpha

n1,obs1,n2,obs2,alpha = setVariables(100,21,120,30,.05)
```

Get sample proportions and review if both samples follow a normal distr

Then we need to calculate the sample proportions

In [2]:
```
def getSampleProportions(obs1,obs2,n1,n2):
    ps1 = obs1*1./n1;
    ps2 = obs2*1./n2;
    return ps1,ps2

def checkSamplesNormalDist(n1,ps1,n2,ps2):
    if (n1*ps1 > 5.) & (n1*ps1*(1-ps1)> 5.) & (n2*ps2 > 5.) & (n2*ps2*(1-ps2)> 5.):
        #Checking if both samples are normal
        print "Both samples fulfill the normal distribution"
    return (n1*ps1 > 5.) & (n1*ps1*(1-ps1)> 5.) & (n2*ps2 > 5.) & (n2*ps2*(1-ps2)> 5.)

ps1,ps2 = getSampleProportions(obs1,obs2,n1,n2)
checkSamplesNormalDist(n1,ps1,n2,ps2)
```

```
Both samples fulfill the normal distribution
```

Out[2]:
```
True
```

Get the Standard Error of the Difference

In order to get the Standard Error of the Difference

$$\sigma_{p_1-p_2} = \sqrt{\overline{pq}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

We need to follow these steps:

1. Get the pool proportion
2. Calculate the Standard Error of the Difference

1. Get the pool proportion

Insert variables in the following formula:

$$\overline{p} = \frac{x_1+x_2}{n_1+n_2}$$

Where:

- x1 and n1 are the observations and sample size of 2015 respectively
- x2 and n2 are the the observations and sample size of 2016 respectively

2. Calculate the variability

- Do the calculation

In [3]:
```
def getPoolProportion(obs1,obs2,n1,n2):
    return (obs1+obs2)*1./(n1+n2);

def getSigmaDiffProportions(poolProportion,n1,n2):
    import math
    qBar = 1 - poolProportion;
    numerator = qBar*poolProportion;
    denominator = (1./n1)+(1./n2)
    return math.sqrt(numerator*denominator)

poolProportion = getPoolProportion(obs1,obs2,n1,n2)
sigmaP1P2 = getSigmaDiffProportions(poolProportion,n1,n2)
```

Calculate z

This block provide the steps to:
1) Get the z-score, and
2) Do the hypothesis testing (Here we will need alpha)

$$z = \frac{(\overline{x}_1-\overline{x}_2)-(\mu_1-\mu_2)}{\sigma_{\overline{x}_1-\overline{x}_2}}$$

In [4]:
```
# This block provide the steps to:
# 1) Get the z-score, and
# 2) Do the hypothesis testing (Here we will need alpha)

def getZscore(ps1,ps2,sigmaP1P2):
    diffSampleProps = ps1 - ps2
    diffPopulationProps = 0; # from the definition of the hypotheses where the dif of populations is
 zero.
    zScore = (diffSampleProps*1. - diffPopulationProps)/sigmaP1P2
    return zScore

def get_pValueUpperTail(zScore,alpha):
    import scipy.stats as st
    msg = ""
    if st.norm.sf(abs(zScore)) >= alpha:
        msg = """If the p-value is greater than or equal to alpha, do not reject the null hypothesis.
                 there is not enough statistical evidence that the alternative hypothesis is correct.
                 """
    else:
        msg = """Reject the null hypothesis since the p-value is less than alpha,
                 Therefore ,there is statistical evidence that the alternative hypothesis is correct.
                 """

    return zScore,alpha,st.norm.sf(abs(zScore)),msg, st.norm.sf(abs(zScore)) < alpha

zScore = getZscore(ps1,ps2,sigmaP1P2)
zScore,alpha,pValue,generalConclusion,rejected = get_pValueUpperTail(zScore,alpha)

print "Z-score of %f"%zScore
print "P-value of %f"%pValue
```

```
Z-score of -0.700057
P-value of 0.241946
```

Conclusions

Based on the sample results, can you conclude that there is a **higher proportion** of U.S. companies offering high-deductible health insurance plan to their workers **in 2016 than in 2015**?

Hypotheses:

Null Hypothesis: p2015 <= p2016
Alternative hypothesis: 2015 > p2016
With an alpha of 0.05

$$H_0: p_1 - p_2 \leq 0$$
$$H_A: p_1 - p_2 > 0$$

Null Hypothesis: p2015 - p2016 <= 0
Alternative hypothesis: p2015 - p2016 > 0

With an alpha of 0.05

In [5]:
```
# Conclusion
def conclusion(rejected):
    if rejected:
        print """With a 95% confidence on the sample results, we can conclude that \nthere is statist
ical evidence to support the claim that there is a \nhigher proportion of U.S. companies offering hig
h-deductible health \ninsurance plan to their workers in 2015 than in 2016."""
    else:
        print """With a 95% confidence on the sample results, we can conclude that \nthere is NOT sta
tistical evidence to support the claim that there is a \nhigher proportion of U.S. companies offering
high-deductible health \ninsurance plan to their workers in 2015 than in 2016. In other words, \nther
e is statistical evidence to support the claim that there is a higher \nproportion of U.S. companies
 offering high-deductible health insurance plan \nto their workers in 2016 than in 2015"""
conclusion(rejected)
```

```
With a 95% confidence on the sample results, we can conclude that
there is NOT statistical evidence to support the claim that there is a
higher proportion of U.S. companies offering high-deductible health
insurance plan to their workers in 2015 than in 2016. In other words,
there is statistical evidence to support the claim that there is a higher
proportion of U.S. companies offering high-deductible health insurance plan
to their workers in 2016 than in 2015
```