

Hypothesis Testing

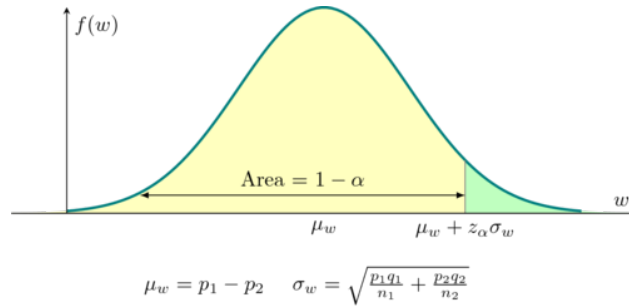
Parthiban Rajendran

October 12, 2018

1 When Successes and Failures are high enough

1.1 (p_1, p_2) known

Suppose that we are interested in comparing two approximately normal sampling distributions described by random variables $\frac{Y_1}{n_1} = N\left(p_1, \frac{p_1 q_1}{n_1}\right)$ and $\frac{Y_2}{n_2} = N\left(p_2, \frac{p_2 q_2}{n_2}\right)$, created from population distributions which are Bernoulli distributions. Note that Y_1 represents the sum of *successes* in a sample set, and thus $\frac{Y_1}{n_1}$ represents sample proportions. For example, for any k th sample set of $\frac{Y_1}{n_1}$, we calculate sample proportion statistic, $\frac{Y_{1k}}{n_1} = \frac{1}{n} \sum_{i=1}^n Y_{1ki}$, where Y_{1ki} is i th sample in k th sample set of sampling distribution described by $\frac{Y_1}{n_1}$. Similarly for $\frac{Y_2}{n_2}$. Then, if no of success and failures are high enough ¹, that is at least > 10 , as a general rule, we could assume that the random variable $W = \frac{Y_1}{n_1} - \frac{Y_2}{n_2}$ has approximately normal distribution $W = N(p_w, \sigma_w^2)$ where $p_w = p_1 - p_2$ and $\sigma_w = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ and has shown below, before standardization to Z. We *destandardize* from Z, because, each α could be linked to corresponding z score, which further could be linked to actual w or x axis in question.



The significance level α , corresponds to the rest of $1 - \alpha$ area, that is green area as shown above.

¹<https://www.khanacademy.org/math/ap-statistics/two-sample-inference/two-sample-z-test-proportions/v/hypothesis-test-for-difference-in-proportions>

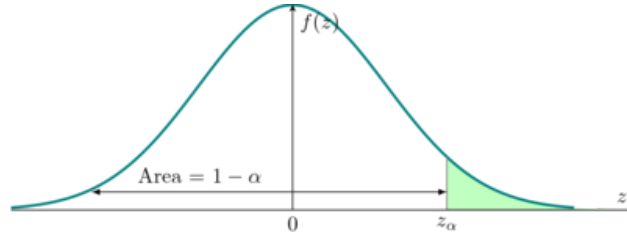
$$\begin{aligned}
P(W \geq \mu_w + z_\alpha \sigma_w) &= \alpha \\
\therefore P\left(\frac{W - \mu_w}{\sigma_w} \geq z_\alpha\right) &= \alpha \\
P\left(\frac{\left(\frac{Y_1}{n_1} - \frac{Y_2}{n_2}\right) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \geq z_\alpha\right) &= \alpha
\end{aligned}$$

Typically, null hypothesis is $p_1 = p_2$, so, assigning it to a common p , i.e $p_1 = p_2 = p$,..

$$\begin{aligned}
P\left(\frac{\frac{Y_1}{n_1} - \frac{Y_2}{n_2}}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \geq z_\alpha\right) &= \alpha \\
P\left(\frac{\frac{Y_1}{n_1} - \frac{Y_2}{n_2}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \geq z_\alpha\right) &= \alpha
\end{aligned}$$

Thus the z score for given sample data would be $z = \frac{\frac{Y_1}{n_1} - \frac{Y_2}{n_2}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

So if our alternate hypothesis is that $H_a : p_1 > p_2$, then we could calculate Z score as above and if that is beyond z_α we could reject null hypothesis.



We could similarly derive for $H_a : p_1 < p_2$, and $H_a : p_1 \neq p_2$.

1.2 (p_1, p_2) unknown

Of course, the above section was for pedagogical purposes, to illustrate the concept. In reality, the individual p_1 and p_2 are not hypothesized typically, and usually compared only to see if there is significant evidence that if one is greater/smaller/different from the other. In which case we simply could use our best estimator \hat{p} for calculating standard deviation in place of p . There are usually two ways, here.

Way 1: Calculate weighted p

This is usually given as $\hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2}$. And then, 1 beocmes

$$P\left(\frac{\frac{Y_1}{n_1} - \frac{Y_2}{n_2}}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \geq z_\alpha\right) = \alpha \tag{1}$$

At the time of this writing, I could not find a derivation for the same, so over to next one.

Way 2: Use sample \hat{p}_1, \hat{p}_2

This is straight forward approach directly from 1, with $p_1 = p_2$

$$P\left(\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \geq z_\alpha\right) = \alpha \quad (2)$$

Tips

- Equation 2 would be the one mostly used for almost any of difference of proportions problems (of course adapted to right or left or both tails as needed)

Example

A machine shop that manufactures toggle levers has both a day and a night shift. A toggle lever is defective if a standard nut cannot be screwed onto the threads. Let p_1 and p_2 be the proportion of defective levers among those manufactured by the day and night shifts, respectively. We shall test the null hypothesis, $H_0 : p_1 = p_2$, against a two-sided alternative hypothesis based on two random samples, each of 1000 levers taken from the production of the respective shifts.

(a) Define the test statistic and a critical region that has an $\alpha = 0.05$ significance level. Sketch a standard normal pdf illustrating this critical region.

(b) If $y_1 = 37$ and $y_2 = 53$ defectives were observed for the day and night shifts, respectively, calculate the value of the test statistic. Locate the calculated test statistic on your figure in part (a) and state your conclusion.

This example was taken from exercise 8.3-11 in ?]

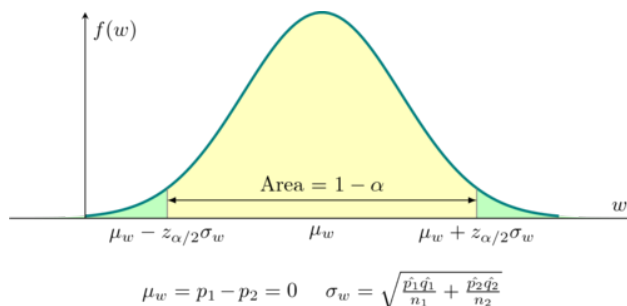
Solution:

Day: $y_1 = 37, n_1 = 1000, \hat{p}_1 = \frac{Y_1}{n_1} = \frac{37}{1000} = 0.037$

Night: $y_2 = 53, n_2 = 1000, \hat{p}_2 = \frac{Y_2}{n_2} = \frac{53}{1000} = 0.053$

(a)

It is said, "two sided alternative hypothesis", so below is our required test statistic. note we have used our best estimators (\hat{p}_1, \hat{p}_2) so result is only approximate.



Calculating the values σ_w , we could arrive at $w = \mu_w \pm z_{\alpha/2} \sigma_w = \pm z_{\alpha/2} \sigma_w$ value beyond which we could define critical region α . Since it is double tailed, we already know $z_{0.025} = 1.96$.

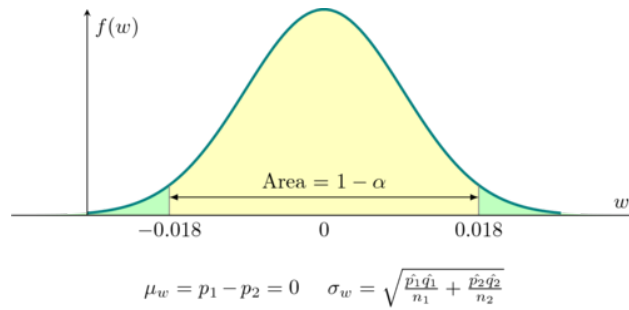
$$\sigma_w = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} = \sqrt{\frac{(0.037)(1-0.037)}{1000} + \frac{(0.053)(1-0.053)}{1000}}$$

```
In[14]: p_1_hat, q_1_hat, p_2_hat, q_2_hat, n_1, n_2 = 0.037, 1-0.037, 0.053, 1-0.053, 1000,
1000
z_0025 = 1.96

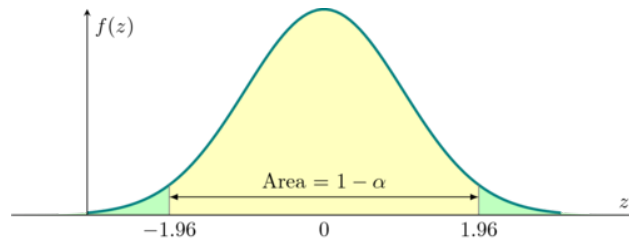
from math import sqrt
s_w = sqrt( (p_1_hat*q_1_hat/n_1) + (p_2_hat*q_2_hat/n_2) )
```

```
print(s_w*z_0025)
```

0.018157472158866164



We could already take a call on our null hypothesis, Our $\hat{p}_1 - \hat{p}_2 = 0.037 - 0.053 = 0.016 < 0.0018$, so we cannot reject H_0 . Our standardized test statistic would be simply z distribution as below.



(b)

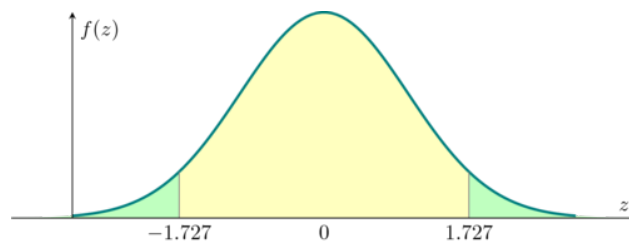
We have already kinda finished the solution, but for question's sake we could complete it fully by calculating the Z score.

Using 2, $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$

```
In[19]: num = p_1_hat - p_2_hat
        den = s_w
        num/den
```

Out[19]: -1.7271126578424703

Being double tailed operation, our z score is thus ± 1.727 . And since $\pm 1.727 < \pm 1.96$, we again **cannot reject null hypothesis** because then our probability of making Type I error would be more than allowed limit of $\alpha = 0.05$. Our standardized test statistic, with $\pm z_{\alpha/2} = \pm 1.727$ is shown below.



Though visibly not clear, one could use z table to find that $z_{1.727}$ takes more area than 0.05 which corresponds to $z_{1.96}$. Thus we conclude our answer.

2 Conditions Summary

