

Hypothesis Testing

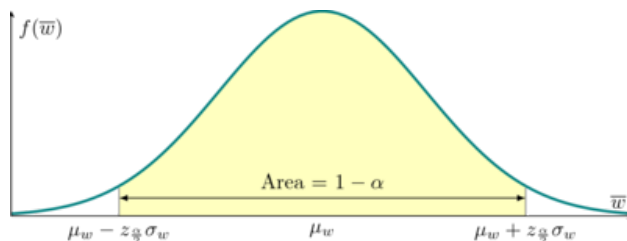
Parthiban Rajendran

October 12, 2018

Note the pre requisite to understand below material is to know confidence intervals for difference between two means as we straight away use the definitions from there. In fact, entire hypothesis testing concept is always understood better after learning confidence intervals and is the typical order in many textbooks.

1 σ known, sample sizes are high

Suppose that we are interested in comparing two approximately normal sampling distributions described by random variables $\bar{X} = N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$ and $\bar{Y} = N(\mu_{\bar{y}}, \sigma_{\bar{y}}^2)$, created from population distributions described by random variables $X(\mu_x, \sigma_x^2)$ and $Y(\mu_y, \sigma_y^2)$. Note that \bar{X} represents collection of sample means from sampled sets sampled from X and similarly for \bar{Y} . Since both \bar{X} and \bar{Y} are normally distributed, and assuming both are independent to each other, the distribution $W = \bar{X} - \bar{Y}$ would be again a normal distribution $W(\mu_w, \sigma_w^2)$, where $\mu_w = \mu_{\bar{x}} - \mu_{\bar{y}}$ and $\sigma_w^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2$.



Then, we already know the confidence intervals could be calculated as below.

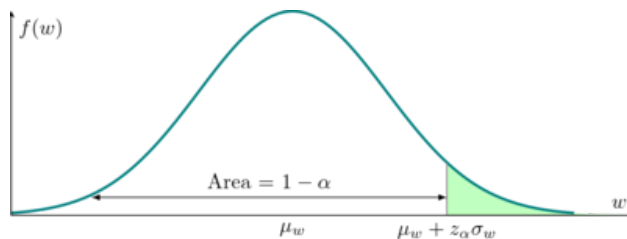
$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{(\bar{X} - \bar{Y}) - (\mu_{\bar{x}} - \mu_{\bar{y}})}{\sqrt{\frac{\sigma_{\bar{x}}^2}{n} + \frac{\sigma_{\bar{y}}^2}{m}}} \leq z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

For Hypothesis testing, let the problem at hand is to wonder, if one mean is greater than the other. For eg, if $\mu_{\bar{x}} > \mu_{\bar{y}}$. This is another way of saying if $\mu_w > 0$. Then we could formulate our hypothesis as follows.

Null hypothesis: $H_0 : \mu_w = 0$

Alternate hypothesis: $H_a : \mu_w > 0$

Then the probability of making Type I error α , would be right hand tail area as follows. Note $z_{\frac{\alpha}{2}}$ becoming z_{α} as now its one side area we are interested in.

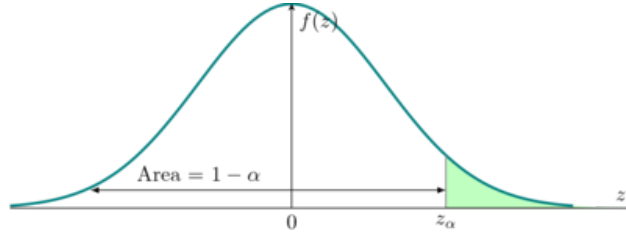


$$\begin{aligned}
P(w \geq \mu_w + z_\alpha \sigma_w) &= \alpha \\
P(w - \mu_w \geq z_\alpha \sigma_w) &= \alpha \\
P\left(\frac{w - \mu_w}{\sigma_w} \geq z_\alpha\right) &= \alpha \\
P\left(\frac{(\bar{X} - \bar{Y}) - (\mu_{\bar{x}} - \mu_{\bar{y}})}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \geq z_\alpha\right) &= \alpha
\end{aligned}$$

Since our null hypothesis is $\mu_w = 0$ or $\mu_{\bar{x}} = \mu_{\bar{y}}$, we could reduce the equation further as,

$$P\left(\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \geq z_\alpha\right) = \alpha \quad (1)$$

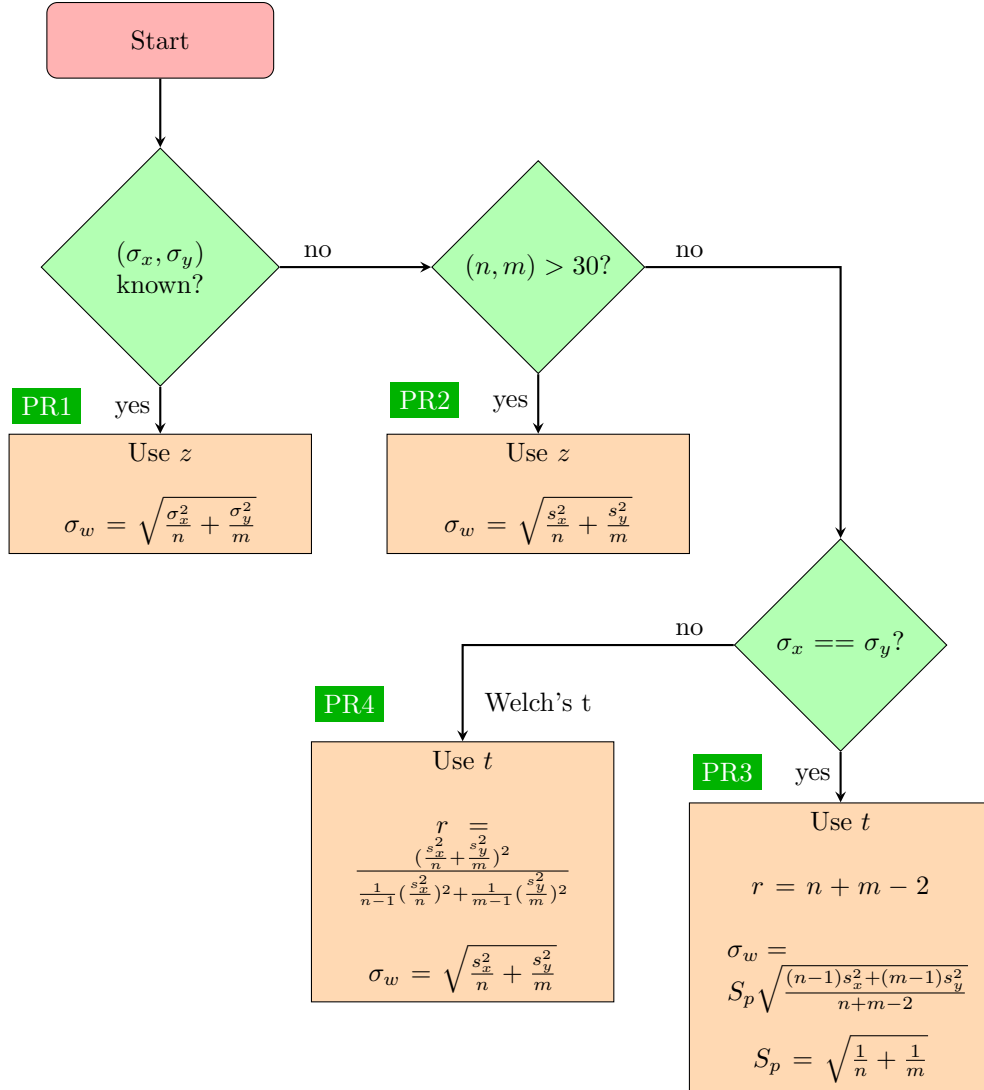
So if Z score of difference between sample means $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$, then the probability of making Type I error α is $P(Z \geq z_\alpha)$. This is depicted below.



So if the calculate Z score from the sample set values (\bar{x}, \bar{y}) exceeds z_α we could straight away **reject null hypothesis** because there is a stronger evidence that the alternate could be true. And we could derive similar Z score for μ decreasing or unequal, but it is much easier to directly tackling the problem than complicating the formula.

2 Visual Summary

Since we use the same components of confidence intervals in hypothesis testing, it helps to recall once the visual summary we have seen there.



3 Examples

3.1 σ unknown, sample sizes are high

As seen in visual summary (**PR2**), in this case, we still could use Z distribution, while we use sample set's unbiased standard deviations (s_x, s_y) in the place of (σ_x, σ_y) as best estimators. Since sample sizes are high, due to CLT, the sampling distribution would still be approximately normal, and our hypothesis testing approximately valid.

Lets assume we have two different ways to lose wieght, and we have to figure out which one is the most effective. We have 10000 people who received treatment A and their average loss is 10 pounds. The standard deviation of their loss is also 10 pounds. Lets consider a second treatment, Treatment B. We also applied it to 10000 people. The average loss in this case is 20 pounds and we also have a standard deviation of 20 pounds. Allowed false positive rate is 5%

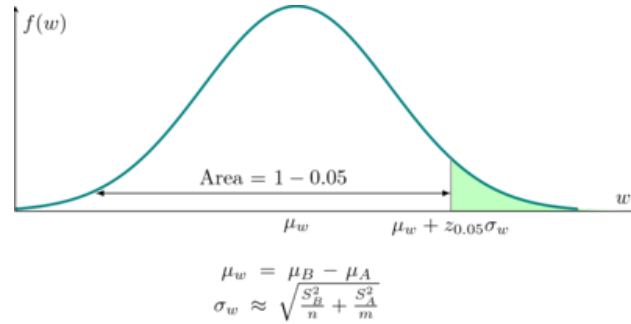
Also given is, null hypothesis $H_0 : \mu_A = \mu_B$

Alternate hypotheis is $H_a : \mu_B > \mu_A$

Solution

Whoa! Sample sizes are so high $\gg \gg 30$. Also $W = B - A$ as we take the hint from alternate hypothesis. So we could rewrite equation 1 as per **PR2** in context as below

$$P\left(\frac{\bar{B} - \bar{A}}{\sqrt{\frac{S_B^2}{n} + \frac{S_A^2}{m}}} \geq z_\alpha\right) = \alpha \quad (2)$$



Given:

B: $n = 10000, \bar{b} = 20, s_B = 20$

A: $m = 10000, \bar{a} = 10, s_A = 10$

5% False positive rate would mean, we could be false 5% of the time while reality is true. This is type I error (rejecting null hypothesis, when null hypothesis is true in reality). Thus, $\alpha = 0.05$. So what would be $z_\alpha = z_{0.05}$?

```
In[7]: import scipy.stats as st
      z_a = st.norm.ppf(1-.05) # as scipy is left tailed by default
      print(z_a)
```

1.6448536269514722

Therefore, $z_\alpha = z_{0.05} = 1.645$.

Let us try to create temporary critical region for W. Our given sample value $\bar{w} = 20 - 10 = 10$. We could say, if our hypothetical next sample means are if or above 10, we would reject the null hypothesis, and then wonder if that is the case, what would be our probability of making Type I error? Will we be in allowed limit of 0.05?

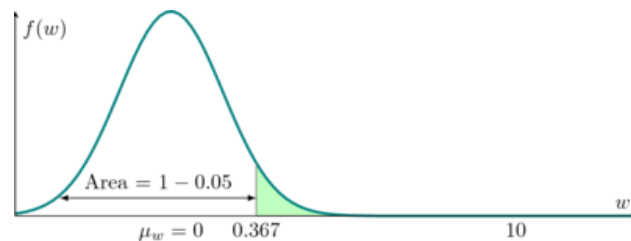
Note that, the critical region for permissible Type I probability of α starts at $(\mu_w + z_{0.05}\sigma_w)$. Since this is when null hypothesis is assumed, it is $z_{0.05}\sigma_w$ which is about 0.367. So any difference beyond 0.367, we could simply reject null hypothesis, that $\mu_A = \mu_B$.

```
In[8]: s_a, s_b, n, m = 10, 20, 10000, 10000
      from math import sqrt
      s_w = sqrt((s_a**2)/m + (s_b**2)/n)
      print(s_w)
      print(s_w*z_a)
```

0.22360679774997896

0.3678004522900572

This situation is depicted below (not drawn at scale on x axis)



Now it would be evident beyond doubt that, we are well within permissible limits of 0.05 for making Type I error, which in fact is almost 0, to choose to reject null hypothesis, and suggest $\mu_B > \mu_A$. If we deployed equation 2 also we would have arrived at same conclusion. We could verify that as well. Rewriting 2,

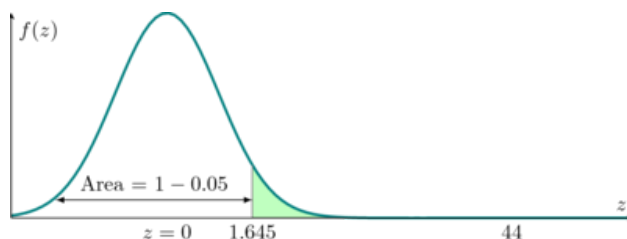
$$P\left(\frac{\bar{b} - \bar{a}}{\sqrt{\frac{s_B^2}{n} + \frac{s_A^2}{m}}} \geq 1.645\right) = 0.05$$

Recall, once the sample set is observed, there is no more probability about it. The calculated Z value is either above Z_α or not.

```
In[10]: b_bar, a_bar = 20, 10
        zs = (b_bar - a_bar)/s_w
        print(zs)
```

44.721359549995796

Our Z score $44 \gg 1.645$, so this again means, while the probability of Z score to be ≥ 1.645 was just 5%, provided null hypothesis was true. Looking at the rarity of this outcome if null hypothesis is true, it would be sane to conclude that this is a strong evidence that alternate hypothesis might be true. This strongly supports alternative hypothesis. This is depicted below (x axis not drawn at scale)



We are thus in a very good position to reject null hypothesis and support alternate hypothesis $H_a : \mu_B > \mu_A$

3.2 σ unknown, unequal and sample sizes are low

As seen in visual summary, we need to use **PR4**, that is Welch's t interval. Note the cumbersome calculation for calculating degrees of freedom. Some textbooks or platforms like Khan,¹ take conservative approach, that is, taking degrees of freedom $r = \min(n, m)$. Nevertheless we will try to use Welch's and see what we get.

Independent random samples of 17 sophomores and 13 juniors attending a large university yield the following data on grade point averages. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean GPAs of sophomores and juniors at the university differ?

Sample Data:

sophomore: $n = 17, \bar{x} = 2.84, s_x = 0.520$

juniors: $m = 13, \bar{y} = 2.9808, s_y = 0.3093$

Solution:

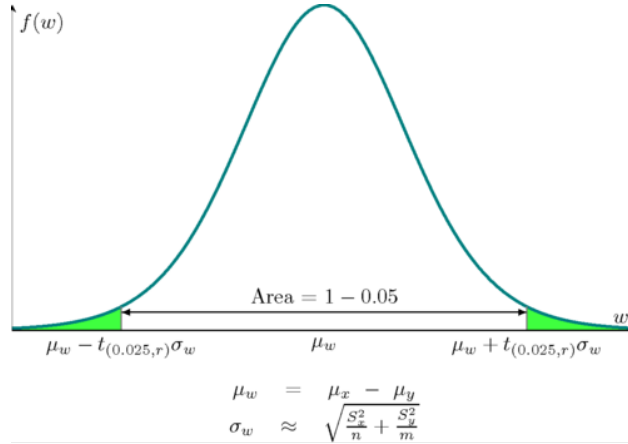
The problem wonders if both the means **differ** so we would need to consider both tails.

Null hypothesis: $\mu_w = 0$ or $\mu_x = \mu_y$

Alternate hypothesis: $\mu_w \neq 0$ or $\mu_x \neq \mu_y$

$\alpha = 0.05$.

¹<https://www.khanacademy.org/math/ap-statistics/two-sample-inference/two-sample-t-test-means/v/two-sample-t-test-for-difference-of-means>



In welch's method, the degrees of freedom, r is the complicated one to calculate. It is given by integer part of below equation.

$$r = \frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2}{\frac{1}{n-1}\left(\frac{s_x^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{s_y^2}{m}\right)^2}$$

```
In[13]: s_x, s_y, n, m = 0.52, 0.3093, 17, 13

num = ( s_x**2/n + s_y**2/m )**2
den1 = (1/(n-1))*( s_x**2/n )**2
den2 = (1/(m-1))*( s_y**2/m )**2
den = den1+den2
r = num/den
print(r)
```

26.629678365237567

The degrees of freedom is the integer part of our result 26.629 which is $r = 26$. Let us then calculate the 't' score for our significance level α ,

```
In[14]: from scipy import stats
ts = stats.t.ppf(0.025, 26) # return value is left tailed by default..
print(ts)
```

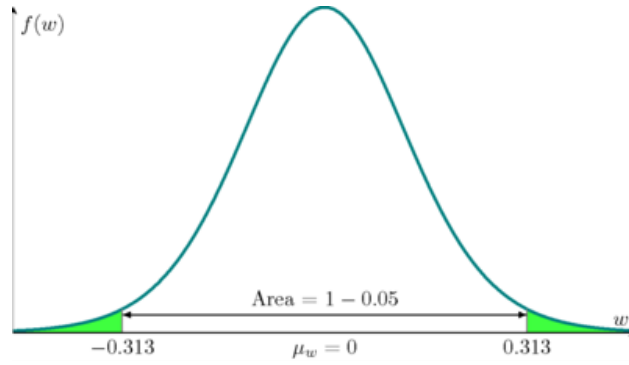
-2.0555294386428713

Therefore, $t_{(\alpha/2,r)} = t_{(0.025,26)} = 2.055$. We could now calculate the limits above or below which Type I error is allowed. Assuming $\mu_w = 0$ due to null hypothesis, $t_{(0.025,26)}\sigma_w$ should give us the limits.

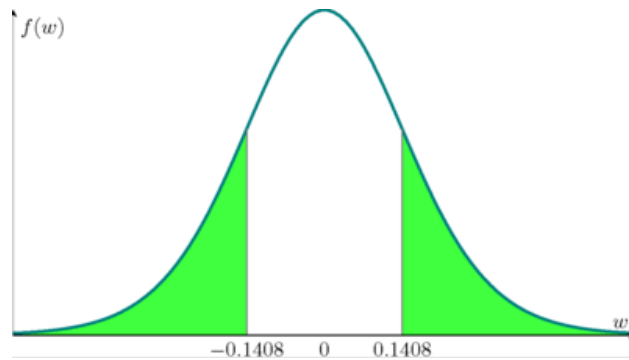
```
In[15]: from math import sqrt
s_w = sqrt( (s_x**2/n + (s_y**2/m) )
print(s_w, s_w*ts)
```

0.15252817156896606 -0.31352614688238034

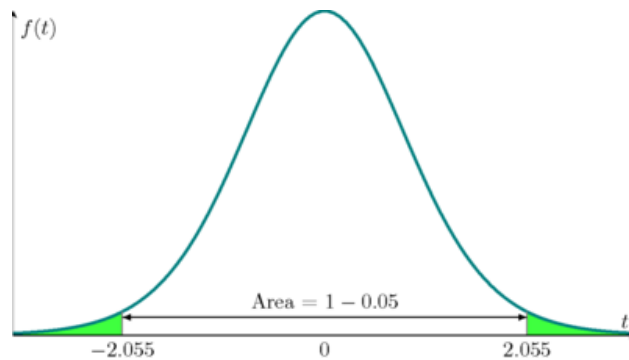
Thus our critical region for given α would be ± 0.313 . Our situation could be depicted as below. We are allowed to reject null hypothesis, if our sample set mean difference is above 0.313 or below -0.313, with $\alpha = 0.05$ probability of making Type I error.



The difference of sample means we got is $\bar{x} - \bar{y} = 2.84 - 2.9808 = -0.1408$. This is far above -0.313, so we **cannot reject null hypothesis**. Taking ± 0.1408 as critical region would increase of probability of Type I error α enormously as shown below.



We could have also taken the difference the other way $\bar{y} - \bar{x} = 2.9808 - 2.84 = 0.1408$, and we still would have arrived at same conclusion because we are interested in only if the sample means of two sampling distributions differ or not (that is why two tails taken in above diagram). Also we could arrive at the same conclusion via 't' values only if we already know $t_{\alpha/2, r}$. We indeed calculated that earlier as 2.055. This means, in units of 't', critical region allowed is ± 2.055 beyond which we are allowed to make Type I error, whose total probability in critical region would be 0.05. This is depicted below.

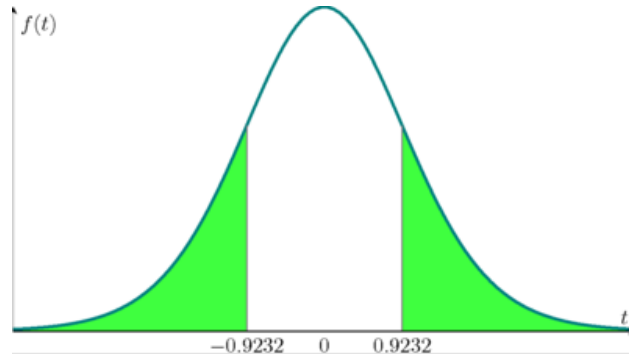


$$\begin{aligned}
P\left(-t_{(\alpha/2,r)} \leq \frac{W - \mu_w}{\sigma_w} \leq t_{(\alpha/2,r)}\right) &= 1 - \alpha \\
2P\left(\left|\frac{W - \mu_w}{\sigma_w}\right| \geq |t_{(\alpha/2,r)}|\right) &= \alpha \\
P\left(\left|\frac{W - \mu_w}{\sigma_w}\right| \geq |t_{(\alpha/2,r)}|\right) &= \frac{\alpha}{2} \\
P\left(\left|\frac{W - \mu_w}{\sigma_w}\right| \geq |t_{(0.025,r)}|\right) &= 0.025 \\
P\left(\left|\frac{W - \mu_w}{\sigma_w}\right| \geq |t_{(0.025,26)}|\right) &= 0.025 \\
P\left(\left|\frac{W - \mu_w}{\sigma_w}\right| \geq 2.055\right) &= 0.025 \\
P\left(\left|\frac{(\bar{X} - \bar{Y}) - (\mu_{\bar{x}} - \mu_{\bar{y}})}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}\right| \geq 2.055\right) &= 0.025 \\
P\left(\left|\frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}\right| \geq 2.055\right) &= 0.025
\end{aligned}$$

When the sample set is observed, we could check if we are in critical region or not by calculating its t score.

$$t = \left| \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \right| = \left| \frac{2.84 - 2.9808}{0.1525} \right| = 0.9232$$

Our t score is 0.9232. This is well outside the critical region towards the null hypothesized zero mean difference, so if we assume this as critical region, we would be making high Type I error, beyond 0.05 as depicted below.



So our conclusion is similar like earlier. We **cannot reject the null hypothesis**.