Hypothesis Testing- Two Population

Comparing two population Means: Large Sample

Hypothesis testing on μ_1, μ_2:
Let a sample of size n_1 and n_2 from a population with means μ_1, μ_2 and variances σ_1^2, σ_2^2 yields sample means $\overline{x}_1, \overline{x}_2$ then the distribution of $\overline{x}_1 - \overline{x}_2$ has approximately normal distribution with

$$\mu_{\overline{x}_1 - \overline{x}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\overline{x}_1 - \overline{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If the two samples are independent.
The test statistic is given by

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

and if $\sigma_1^2 = \sigma_2^2 = \sigma^2$ the test statistic is

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Comparing two population Means: Large Sample

Now, if we are interested in testing the hypothesis

$$H_o : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

we will reject the null hypothesis if $|z| > z_{\alpha/2}$.
Table below summarizes the rejection criteria.
Decision criteria:

Hypothesis			Rejection Criteria
$H_o : \mu_1 = \mu_2$	$Vs.$	$H_1 : \mu_1 > \mu_2$	$z > z_\alpha$
$H_o : \mu_1 = \mu_2$	$Vs.$	$H_1 : \mu_1 < \mu_2$	$z < -z_\alpha$
$H_o : \mu_1 = \mu_2$	$Vs.$	$H_1 : \mu_1 \neq \mu_2$	$|z| > z_{\alpha/2}$

Also note that σ_1^2, σ_2^2 can be approximated by s_1^2 and s_2^2 if both n_1, n_2 are large.

Comparing two population Means: Small Sample

Assumptions for a small sample test of $\mu_1 - \mu_2$
(i) Both populations of interest are normally distributed.
(ii) Both the populations have approximately equal(but unknown) variance
In order to test

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2 \end{aligned}$$

We consider the test statistic

$$t_0 = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ is called pooled variance. Also s_1^2 and s_2^2 are the variances of the first and second samples.

<u>Decision criteria:</u>

Hypothesis	Rejection Criteria
$H_0 : \mu_1 = \mu_2$ Vs. $H_1 : \mu_1 > \mu_2$	$t_0 > t_{\alpha, n_1+n_2-2}$
$H_0 : \mu_1 = \mu_2$ Vs. $H_1 : \mu_1 < \mu_2$	$t_0 < -t_{\alpha, n_1+n_2-2}$
$H_0 : \mu_1 = \mu_2$ Vs. $H_1 : \mu_1 \neq \mu_2$	$\lvert t_0 \rvert > t_{\alpha/2, n_1+n_2-2}$

Comparing two population Means: Small Sample

Assumptions for a small sample test of $\mu_1 - \mu_2$
(i) Both populations of interest are normally distributed.
(ii) The population variances are unequal
In order to test

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2 \end{aligned}$$

We consider the test statistic

$$t_0 = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Also s_1^2 and s_2^2 are the variances of the first and second samples. Note that t_0 has student's t distribution with degrees of freedom given by

$$\nu = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \times \left(\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}\right)^{-1}$$

Decision criteria:

Hypothesis	Rejection Criteria
$H_0 : \mu_1 = \mu_2$ Vs. $H_1 : \mu_1 > \mu_2$	$t_0 > t_{\alpha,\nu}$
$H_0 : \mu_1 = \mu_2$ Vs. $H_1 : \mu_1 < \mu_2$	$t_0 < -t_{\alpha,\nu}$
$H_0 : \mu_1 = \mu_2$ Vs. $H_1 : \mu_1 \neq \mu_2$	$\lvert t_0 \rvert > t_{\alpha/2,\nu}$

Inferences about the differences in Means, Paired t-test

To test

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Let $\mu_d = \mu_1 - \mu_2$ then the above testing problem is equivalent to

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d \neq 0$$

The test statistic for this hypothesis is

$$t_0 = \frac{\overline{d}}{S_d/\sqrt{n}}$$

where $\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ is the mean of the differences and S_d is the sample standard deviation of the differences.

Decision criteria:

Hypothesis	Rejection Criteria
$H_0 : \mu_1 = \mu_2 \quad Vs. \quad H_1 : \mu_1 > \mu_2$	$t_0 > t_{\alpha,n-1}$
$H_0 : \mu_1 = \mu_2 \quad Vs. \quad H_1 : \mu_1 < \mu_2$	$t_0 < -t_{\alpha,n-1}$
$H_0 : \mu_1 = \mu_2 \quad Vs. \quad H_1 : \mu_1 \neq \mu_2$	$\lvert t_0 \rvert > t_{\alpha/2,n-1}$

Example- Two sample Z test

Freshman at public universities work 12.2 hours per week for pay on the average, with a standard deviation of 10.5 hours. At private universities, the average for freshman is 10.2 hours, with a standard deviation of 9.9 hours. The sample size for each is 1,000. Is the difference between the averages real or is it just chance variation. Perform a level 0.05 independent two-sample test to find out.

Solution: Let Let μ_1 and μ_2 denote the average number of hours the public and private university freshman students work per week respectively. We want test

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_a &: \mu_1 \neq \mu_2 \end{aligned}$$

We have $n_1 = 1000, n_2 = 1000, \bar{x}_1 = 12.2, \bar{x}_2 = 10.2, \sigma_1 = 10.5, \sigma_2 = 9.9$
We have the test statistic

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(12.2 - 10.2) - 0}{\sqrt{\frac{10.5^2}{1000} + \frac{9.9^2}{1000}}} = 4.38$$

`p-value=2*pnorm(-4.38)=0.0000118`

Decision: Reject the null hypothesis.

Example- Unknown but Equal Variances

Suppose the recovery time for patients taking a new drug is measured (in days). A placebo group is also used to avoid the placebo effect. The data are as follows:

```
Drug :      15 10 13 7  9 8 21  9 14  8
Placebo:    15 14 12 8 14 7 16 10 15 12
```

Suppose the assumptions of equal variances and normality are valid. Perform the test to show that the drug group has smaller mean.
Let μ_1 and μ_2 respectively denote the mean of drug group and placebo group. We want to test the following

$$H_0 : \mu_1 \geq \mu_2$$
$$H_a : \mu_1 < \mu_2$$

Example

We can use R to test the hypothesis

```
> x = c(15, 10, 13, 7, 9, 8, 21, 9, 14, 8)
> y = c(15, 14, 12, 8, 14, 7, 16, 10, 15, 12)
> t.test(x,y,alt="less",var.equal=TRUE)
        Two Sample t-test
data:  x and y
t = -0.5331, df = 18, p-value = 0.3002
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 2.027436
sample estimates:
mean of x mean of y
     11.4      12.3
```

Decision: Not enough evidence to reject the null hypothesis.

Example: Paired t-test

A study was performed to test whether cars get better mileage on premium gas than on regular gas. Each of 10 cars was first filled with either regular or premium gas, decided by a coin toss, and the mileage for that tank was recorded. The mileage was recorded again for the same cars using the other kind of gasoline.

```
> Premium = c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32)
> Regular = c(16, 20, 21, 22, 23, 22, 27, 25, 27, 28)
> t.test(Premium,Regular, alternative="greater", paired=TRUE)

        Paired t-test
data:  Premium and Regular
t = 4.4721, df = 9, p-value = 0.0007749
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.180207      Inf
sample estimates:
mean of the differences
                      2
```

Decision: Reject the null hypothesis. There is strong evidence of a mean increase in gas mileage between regular and premium gasoline.

Example

Do the volume measurements of tumor change based on physician?
The volume of the tumor was measured by two separate physicians under similar condition. Data below are recorded.
Dr.1 :15.8, 22.3, 14.5, 15.7, 26.8, 24.0, 21.8, 23.0, 29.3, 20.5
Dr.2 :17.2, 20.3, 14.2, 18.5, 28.0, 24.8, 20.3, 25.4, 27.5, 19.7

```
> Dr.1=c(15.8, 22.3, 14.5, 15.7, 26.8, 24.0, 21.8, 23.0, 29.3, 20.5)
> Dr.2=c(17.2, 20.3, 14.2, 18.5, 28.0, 24.8, 20.3, 25.4, 27.5, 19.7)
> t.test(Dr.1,Dr.2, paired=T)

        Paired t-test
data:  Dr.1 and Dr.2
t = -0.3989, df = 9, p-value = 0.6993
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.467632  1.027632
sample estimates:
mean of the differences
                  -0.22
```

Decision: No enough evidence that the volume measurements of tumor change based on physician

Example: Two sample t Vs. Paired t

The composite biodiversity score based on a kick sample of aquatic invertebrates are provided below.

Down: 20 15 10 5 20 15 10 5 20 15 10 5 20 15 10 5

Up: 23 16 10 4 22 15 12 7 21 16 11 5 22 14 10 6

The elements are paired because the two samples were taken on the same river, one upstream and one downstream from the same sewage outfall.

If we ignore the fact that they are paired we may get completely different results

```
> down=c(20,15,10,5,20,15,10,5,20,15,10,5,20,15,10,5)
> up=c(23,16,10,4,22,15,12,7,21,16,11,5,22,14,10,6)
> t.test(down,up)
        Welch Two Sample t-test
data:  down and up
t = -0.4088, df = 29.755, p-value = 0.6856
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.248256  3.498256
sample estimates:
mean of x mean of y
   12.500    13.375
```

Example: Two sample t Vs. Paired t

```
> down=c(20,15,10,5,20,15,10,5,20,15,10,5,20,15,10,5)
> up=c(23,16,10,4,22,15,12,7,21,16,11,5,22,14,10,6)
> t.test(down,up,paired=T)

        Paired t-test
data:  down and up
t = -3.0502, df = 15, p-value = 0.0081
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.4864388 -0.2635612
sample estimates:
mean of the differences
                 -0.875
```

Example

A "new diet and exercise" program has been advertised to be a remarkable way to reduce blood glucose levels in diabetic patients. Ten randomly selected diabetic patients are put on the program and here are the results after one month are given by the following table:

Before:	268	225	252	192	307	228	246	298	231	185
After:	106	186	223	110	203	101	211	176	194	203

Do the data provide sufficient evidence to the claim that the new program reduces blood glucose level in diabetic patients? Use a $= 0.05$.

```
> Before=c( 268, 225, 252, 192, 307, 228, 246, 298, 231, 185)
> After=c( 106, 186, 223, 110, 203,  101, 211, 176, 194, 203)
> t.test(Before, After, alternative="greater", paired=T)
        Paired t-test
data:  Before and After
t = 4.0489, df = 9, p-value = 0.001445
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 39.34775      Inf
sample estimates: mean of the differences
                   71.9
```

Decision: The new diet and exercise program is effective.