SPS 2349 Test of Hypothesis

Sampling From Two Independent Normal Populations

Dr. Mutua Kilai

Kirinyaga University

Introduction

Let X and Y be two independently distributed random variables with distribution $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ respectively.

Let $X_1, X_2, ..., X_m$ be a random sample of size m from X and let $y_1, y_2, ..., y_n$ be another random variable of size n from Y.

We wish to test the equality of means and equality of variance.

Testing Equality of Means

Assuming $\sigma_1^2 = \sigma_2^2 = \sigma^2$ in the two populations.

We wish to test the hypothesis

$$H_0 : \mu_1 = \mu_2$$

against

$$H_1 : \mu_1 \neq \mu_2$$

Now

$$f(x_1, \mu_1 \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_i - \mu_1)^2}$$

$$f(y, \mu_2, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y_i - \mu_2)^2}$$

The Likelihood function is:

$$L(X, \mu_1, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(X_i - \mu_1)^2}$$

$$L(Y, \mu_2, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu_2)^2}$$

Since X and Y are independent,

$$f(x, y, \mu_1, \mu_2, \sigma^2) = f(x, \mu_1, \sigma^2) f(y, \mu_2, \sigma^2)$$

and

$$L(x,y,\mu_1,\mu_2,\sigma^2)=\Big(\frac{1}{2\pi}\Big)^{\frac{m+n}{2}}\Big(\frac{1}{\sigma^2}\Big)^{\frac{m+n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu_1)^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_2)^2}$$

Now

$$\Omega=\{(\mu_1,\mu_2,\sigma^2):-\infty<\mu_1,\mu_2<\infty,\sigma^2>0\}$$

$$\Omega_0=\{(\mu_1,\mu_2,\sigma^2):\mu_1=\mu_2=\mu,\sigma^2>0\}$$

Therefore the maximum likelihood estimates for μ_1,μ_2,σ^2 are

$$\hat{\mu_1}=\bar{x},\hat{\mu_2}=\bar{y},\hat{\sigma^2}=\frac{1}{m+n}\Big[\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2\Big]$$

This implies that under Ω:

$$\begin{aligned}L(\hat{\omega})&=\Big(\frac{1}{2\pi}\Big)^{\frac{m+n}{2}}\Big(\frac{1}{\hat{\sigma^2}}\Big)^{\frac{m+n}{2}}e^{-\frac{1}{2\sigma^2}\Big[\sum_{i=1}^{m}(x_i-\hat{\mu_1})^2+\sum_{i=1}^{n}(y_i-\hat{\mu_2})^2\Big]}\\&=\Big(\frac{1}{2\pi}\Big)^{\frac{m+n}{2}}\Big[\frac{m+n}{\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2}\Big]^{\frac{m+n}{2}}e^{-\frac{1}{2}\Big[\frac{\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2}\Big]\times(m+n)}\\&=\Big(\frac{1}{2\pi}\Big)^{\frac{m+n}{2}}\Big[\frac{m+n}{\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2}\Big]^{\frac{m+n}{2}}e^{-\frac{m+n}{2}}\end{aligned}\tag{1}$$

Now in

Ω_0 $\mu_1=\mu_2=\mu(say)$ when H_0 is true

Then

$$L=\Big(\frac{1}{2\pi}\Big)^{\frac{m+n}{2}}\Big(\frac{1}{\hat{\sigma^2}}\Big)^{\frac{m+n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

Then the ML estimates of μ,σ^2 are

$$\hat{\mu}=\frac{\sum x_i+\sum y_i}{m+n}=\frac{m\bar{x}+n\bar{y}}{m+n}$$

and

$$\hat{\sigma^2}=\frac{\sum_{i=1}^{m}(x_i-\hat{\mu})^2+\sum_{i=1}^{n}(y_i-\hat{\mu})^2}{m+n}=\frac{\sum_{i=1}^{m}(x_i-\frac{m\bar{x}+n\bar{y}}{m+n})^2+\sum_{i=1}^{n}(y_i-\frac{m\bar{x}+n\bar{y}}{m+n})^2}{m+n}$$

Note:

$$\begin{aligned}\sum_{i=1}^{m}(x_i - (\frac{m\bar{x}+n\bar{y}}{m+n}))^2 &= \sum_{i=1}^{m}\left[(x_i-\bar{x}) + (\bar{x} - \frac{m\bar{x}+n\bar{y}}{m+n})\right]^2 \\ &= \sum_{i=1}^{m}(x_i-\bar{x})^2 + (\bar{x} - \frac{m\bar{x}+n\bar{y}}{m+n})^2 \\ &= \sum_{i=1}^{m}(x_i-\bar{x})^2 + \frac{mn^2}{(m+n)^2}(\bar{x}-\bar{y})^2\end{aligned} \tag{2}$$

Similarly

$$\sum_{i}^{n}(y_i - (\frac{m\bar{x}+n\bar{y}}{m+n}))^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{m^2n}{(m+n)^2}(\bar{x}-\bar{y})^2$$

Hence

$$\frac{\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{mn}{m+n}(\bar{x}-\bar{y})^2}{m+n}$$

$$L(\Omega_0) = \left(\frac{1}{2\pi}\right)^{\frac{m+n}{2}}\left[\frac{m+n}{\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{mn}{m+n}(\bar{x}-\bar{y})^2}\right]^{\frac{m+n}{2}} \times e^{-\frac{m+n}{2}}$$

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \left[\frac{\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2}{\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{mn}{m+n}(\bar{x}-\bar{y})^2}\right]^{\frac{m+n}{2}}$$

Now $\bar{x}-\bar{y} \sim N(\mu_1-\mu_2, \sigma^2(\frac{1}{m}+\frac{1}{n}))$

$$Z = \frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2(\frac{1}{m}+\frac{1}{n})}} = \frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2(m+n)}{mn}}}$$

$$\sqrt{\frac{mn}{m+n}}\left[\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sigma}\right] \sim N(0,1)$$

also $\frac{\sum_{i=1}^{m}(x_i-\bar{x})^2}{\sigma^2} \sim \chi^2_{m-1}$ and $\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$

$$U = \frac{\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2}{\sigma^2} \sim \chi^2_{m+n-1}$$

Therefore when H_0 is true the statistic

$$T = \frac{\sqrt{\frac{mn}{m+n}}(\frac{\bar{x}-\bar{y}}{\sigma})}{\sqrt{\frac{\sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{i=1}^{n}(y_i-\bar{y})^2}{\sigma^2(m+n-2)}}} \sim t_{m+n-2}$$

Therefore

$$\lambda = \left[\frac{1}{1 + T^2_{m+n-2}} \right]^{\frac{m+n}{2}}$$

We reject H_0 whenever $T < T_1$ or $T > T_2$ where T_1 and T_2 are constants satisfying

$$P[T < T_1 \ \ or \ \ T > T_2 | H_0] = \alpha$$

Therefore the LRT for $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$ is to compute T and reject H_0 at α level of significance whenever

$$T < -t_{\frac{\alpha}{2}} \ \ or \ \ T > t_{\frac{\alpha}{2}} \ \ or \ \ |T| \geq t_{\frac{\alpha}{2}}$$

Example

A botanist is interested in comparing the growth responses for dwarf pea stems to show different levels of hormone indocletic acid. Let X denote the growth attributed to level 1 on the stem and Y to level II. 11 observations of X were taken and 13 observations of Y as follows

$$X : 0.8, 1.0, 0.1, 0.9, 1.7, 1.0, 1.4, 0.9, 1.2, 0.5, 1.8$$

$$Y : 1.0, 0.8, 1.6, 2.6, 1.3, 1.1, 2.4, 1.8, 2.5, 1.4, 1.9, 2.0, 1.2$$

Assuming there are random samples from normal population where $X \sim N(\mu_1, \sigma^2)$ and $Y \sim N(\mu_2, \sigma^2)$ i.e $\sigma_1^2 = \sigma_2^2$

Test the hypothesis:

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

take $\alpha = 0.05$

Solution

The test statistic is

$$T = \frac{\sqrt{\frac{mn}{m+n}} (\bar{x} - \bar{y})}{\sqrt{\frac{\sum_{i=1}^{m} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2}{m+n-2}}}$$

From the data we obtain the following:

$$\bar{x} = 1.027, \sum_{i=1}^{n} (x_i - \bar{x})^2 = 2.442, m = 11$$

$$\bar{y} = 1.662, \sum_{i=1}^{n} (y_i - \bar{y})^2 = 4.231, n = 13$$

Hence

$$T = \frac{\sqrt{\frac{11\times 13}{11+13}}(1.027-1.662)}{\sqrt{\frac{2.442+4.231}{11+13-2}}} = -2.81$$

From the table $t_{22,0.025} = 2.074$

$$|T| = 2.81 > 2.074$$

Thus we reject H_0 at 5% level of significance.

Tests on Variance of two Normal Population

Let X and Y be two independent random variables where $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$

Suppose we want to test the hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2 \;\; against \;\; H_1 : \sigma_1^2 \neq \sigma_2^2$$

Let $x_1, x_2, ..., x_m$ and $y_1, y_2, ..., y_n$ be two independent random samples from X and Y respectively, the likelihood function is

$$L(\Omega) = \left(\frac{1}{2\pi\sigma_1^2}\right)^{\frac{m}{2}} \left(\frac{1}{2\pi\sigma_2^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^m (x_i-\mu_1)^2 - \frac{1}{2\sigma_2^2}\sum_{i=1}^n (y_i-\mu_2)^2}$$

The parameter spaces are:

$$\Omega = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : -\infty < \mu_1, \mu_2 < \infty, \sigma_1^2, \sigma_2^2 > 0\}$$

and

$$\Omega_0 = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : -\infty < \mu_1, \mu_2 < \infty, \sigma_1^2 = \sigma_2^2 = \sigma^2 > 0\}$$

The ML estimates of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ under Ω are

$$\hat{\mu_1} = \bar{x}, \hat{\mu_2} = \bar{y}, \hat{\sigma_1^2} = \frac{1}{m}\sum_{i=1}^m (x_i - \bar{x})^2, \hat{\sigma_2^2} = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2$$

Hence

$$L(\hat{\Omega}) = \left(\frac{1}{2\pi}\right)^{\frac{m+n}{2}} \left(\frac{1}{\hat{\sigma_1^2}}\right)^{\frac{m}{2}} \left(\frac{1}{\hat{\sigma_2^2}}\right)^{\frac{n}{2}} e^{-\frac{m+n}{2}}$$

Under the $Omega_0$ the ML estimates of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ are

$$\hat{\mu_1} = \bar{x}, \hat{\mu_2} = \bar{y}, \hat{\sigma^2} = \frac{1}{m+n}\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2$$

Hence

$$L(\hat{\Omega_0}) = \left(\frac{1}{2\pi}\right)^{\frac{m+n}{2}} \left(\frac{1}{\hat{\sigma^2}}\right) e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\bar{x})^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$L(\hat{\Omega_0}) = \left(\frac{1}{2\pi}\right)^{\frac{m+n}{2}} \left(\frac{1}{\hat{\sigma^2}}\right)^{\frac{m+n}{2}} e^{-\frac{m+n}{2}}$$

The likelihood ratio is

$$\lambda = \frac{L(\hat{\Omega_0})}{L(\hat{\Omega})} = \frac{\left(\sigma_1^2\right)^{\frac{m}{2}} \left(\hat{\sigma_2^2}\right)^{\frac{n}{2}}}{(\hat{\sigma^2})^{\frac{m+n}{2}}}$$

$$\begin{aligned}
\lambda &= \frac{\left[\sum_{i=1}^{m}(x_i-\bar{x})^2\right]^{\frac{m}{2}} \left[\sum_{i=1}^{n}(y_i-\bar{y})^{\frac{n}{2}}\right](m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}} n^{\frac{n}{2}} \left[\sum_{i=1}^{m}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2\right]^{\frac{m+n}{2}}} \\
&= \frac{(m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}} n^{\frac{n}{2}}} \left[\frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2}\right]^{\frac{m}{2}} \left[\frac{\sum(y_i-\bar{y})^2}{\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2}\right]^{\frac{n}{2}} \qquad (3) \\
&= \frac{(m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}} n^{\frac{n}{2}}} \left[\frac{\sum(x_i-\bar{x})^2/\sum(y_i-\bar{y})^2}{1+\sum(x_i-\bar{x})^2/\sum(y_i-\bar{y})^2}\right]^{\frac{m}{2}} \left[\frac{1}{1+\sum(x_i-\bar{x})^2/\sum(y_i-\bar{y})^2}\right]^{\frac{n}{2}}
\end{aligned}$$

Now let

$$F = \frac{\sum(x_i-\bar{x})^2/m-1}{\sum(y_i-\bar{y})^2/n-1}$$

The above variable gives:

$$\begin{aligned}
\lambda &= \frac{(m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}} n^{\frac{n}{2}}} \left[\frac{\frac{m-1}{n-1}F}{1+\frac{m-1}{n-1}F}\right]^{\frac{m}{2}} \left[\frac{1}{1+\frac{m-1}{n-1}F}\right]^{\frac{n}{2}} \qquad (4) \\
&= \frac{(m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}} n^{\frac{n}{2}}} \frac{\left(\frac{m-1}{n-1}F\right)^{\frac{m}{2}}}{(1+\frac{m-1}{n-1}F)^{\frac{m+n}{2}}}
\end{aligned}$$

The variable

$$V = \frac{\sum(x_i-\bar{x})^2}{\sigma_1^2(m-1)} / \frac{\sum(y_i-\bar{y})^2}{\sigma_2^2(n-1)}$$

has a Fishers F distribution with $(m-1)$ and $(n-1)$ df

Now under H_0

$$F = \frac{\sum(x_i-\bar{x})^2/(m-1)}{\sum(y_i-\bar{y})^2/(n-1)} = \frac{S_1^2}{S_2^2}$$

has $F_{m-1,n-1}$

The test is to compute F and reject H_0 whenever $F < F_1$ or $F > F_2$ where $F_2 = F_{\frac{\alpha}{2}}$ and $F_1 = \frac{1}{F_{\frac{\alpha}{2}}}$

Remark

- If $H_1 : \sigma_1^2 > \sigma_2^2$ Compute F and reject H_0 if $F > F_{1-\alpha}$
- If $H_1 : \sigma_1^2 < \sigma_1^2$ Compute F and reject H_0 if $F < \frac{1}{F_\alpha}$
- $F_{1-\frac{\alpha}{2}(m,n)} = \frac{1}{F_{\frac{\alpha}{2}}(m,n)}$

F-test for equal population variance

One-tailed test

$$H_0 : \sigma_1^2 = \sigma_2^2 \ \ vs \ \ H_1 : \sigma_1^2 < \sigma_2^2 \ \ (H_1 : \sigma_1^2 > \sigma_2^2)$$

Test statistic

$$F = \frac{S_2^2}{S_1^2}$$

or

$$F = \frac{S_1^2}{S_2^2}$$

when $\sigma_1^2 > \sigma_2^2$

Rejection region $F > F_{\alpha(m,n)}$

Two-tailed Test

$$H_0 : \sigma_1^2 = \sigma_2^2 \ \ vs \ \ H_1 : \sigma_1^2 \neq \sigma_2^2$$

Test-statistic

$$F = \frac{larger \ \ sample \ \ variance}{smaller \ \ sample \ \ variance}$$

$$F = \frac{S_1^2}{S_2^2}$$

where $s_1^2 > s_2^2$

or $\frac{s_2^2}{s_1^2}$ where $s_2^2 > s_1^2$

Rejection region

$$F > F_{\frac{\alpha}{2},(m,n)} \ \ or \ \ F < \frac{1}{F_{\frac{\alpha}{2}}}$$

Assumptions

- Both sampled populations are normally distributed
- The samples are random and independent

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Example

Independent random samples were selected from each of two normally distributed population $n_1 = 16$ from population 1 and $n_2 = 25$ from population 2. The mean and variance for the two population for the two sample are shown below:

sample 1

$$n_1 = 16, \bar{x_1} = 22.5, s_1^2 = 2.87$$

$$n_2 = 25, \bar{x_2} = 28.2, s_2^2 = 9.85$$

a. Test the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative $H_1 : \sigma_1^2 \neq \sigma_2^2$ use $\alpha = 0.05$
b. Test the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative $H_1 : \sigma_1^2 < \sigma_2^2$ use $\alpha = 0.05$

Solution: a

Test statistic

$$F = \frac{S_2^2}{S_1^2} = \frac{9.85}{2.87} = 3.43$$

$$F_1 = \frac{1}{F_{0.025(15,24)}} = 0.41$$

$$F_2 = F_{\frac{\alpha}{2}(24,15)} = F_{0.025(24,15)} = 2.70$$

The computed value exceeds the rejection value of 2.70 thus we reject H_0

Solution: b

$$F_{0.95(24,15)} = \frac{1}{F_{0.05(15,24)}} = \frac{1}{2.11} = 0.4739$$

Computed $F > Ftable$ thus we do not reject H_0

Testing for Independence in a Bivariate Normal Population

Suppose that the random variables $X\&Y$ have the joint distribution given by bivariate normal distribution:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right) - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{(y-\mu_2)}{\sigma_2}\right)^2\right]}$$

Where

$$E(x) = \mu, Var(X) = \sigma_1^2, E[Y] = \mu_2, Var(Y) = \sigma_2^2, Cov(X,Y) = \rho\sigma_1\sigma_2$$

For bivariate normal densities $\rho = 0$ implies that $X\&Y$ are independent a result not true for other distributions.

To test the independence of $X\&Y$ it is enough to test $\rho = 0$ thus we have:

$$H_0 : \rho = 0 \;\; vs \;\; H_1 : \rho \neq 0$$

Let $y_1, y_2, ..., y_n$ be independent observations on Y such that

$$y_i = \alpha + \beta x_i + e_i$$

The joint distribution $y_1, y_2, ..., y_n$ is

$$L(Y|X) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha+\beta x_i)^2}$$

We are interested in finding out whether the assumed linear relationship between $Y\&X$ exists i.e after finding the best estimates of $\alpha\&\beta$ we must test the hypothesis

$$H_0 : \beta = 0 \;\; vs \;\; H_1 : \beta \neq 0$$

The parameter spaces in this case are:

$$\Omega = \{(\alpha, \beta, \sigma^2) : -\infty < \alpha, \beta < \infty, \sigma^2 > 0\}$$

$$\Omega_0 : \{(\alpha, \beta, \sigma^2) : -\infty < \alpha < \infty, \beta = 0, \sigma^2 > 0\}$$

The MLE under Ω are

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}X_i)^2}{n}$$

Under Ω_0 the MLEs are

$$\hat{\alpha} = \bar{Y}, \hat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$$

The likelihood ratio is

$$\lambda = \frac{L(\hat{\Omega_0})}{L(\hat{\Omega})} = \frac{(\frac{1}{2\pi\hat{\sigma^2}})^{\frac{n}{2}} e^{-\frac{1}{2\hat{\sigma^2}} \sum_{i=1}^{n}(y_i - \hat{\alpha})^2}}{(\frac{1}{2\pi\hat{\sigma^2}})^{\frac{n}{2}} e^{-\frac{1}{2\hat{\sigma^2}}} \sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta x_i})^2}$$

$$\lambda = \left[\frac{\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}X_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right]^{\frac{n}{2}}$$

$$\lambda = \left[\frac{\sum_{i=1}^{n}(y_i - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right]^{\frac{n}{2}}$$

$$\lambda = \left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\hat{\beta}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) + \hat{\beta}^2\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{2}(Y_i - \bar{Y})^{2]}}\right]$$

Substituting for $\hat{\beta}$ we obtain

$$\lambda = \left[\sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{2[\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{[\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]^{\frac{n}{2}}$$

$$\lambda = \left[1 - \frac{[\sum(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\right]^{\frac{n}{2}} = (1 - V^2)^{\frac{n}{2}}$$

Where

$$V = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Since $\hat{\beta}$ is a function of a normal random variable then it is normally distributed with

$$E(\hat{\beta}) = E\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(\alpha + \beta x_i - \beta\bar{X} - \alpha)}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\beta\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \beta$$

$$Var(\hat{\beta}) = Var\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] = \frac{\sum(X_i - \bar{X})^2 Var(Y_i - \bar{Y})}{(\sum(x_i - \bar{x})^2)^2} = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

This implies that

$$\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum(X_i - \bar{X})^2})$$

Hence

$$Z = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum(x_i - \bar{x})^2}}} \sim N(0, 1)$$

Also

$$Q = \frac{n\hat{\sigma^2}}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \sim \chi^2_{n-2}$$

Hence

$$T = \frac{Z}{\sqrt{Q/(n-2)}} \sim t_{n-2}$$

If H_0 is true then $\beta = 0$

$$T = \frac{\hat{\beta}\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}{\sqrt{n\sigma^2/(\hat{n-2})}} \sim t_{n-2}$$

Substituting $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma^2}$ in T we have:

$$T = \frac{V\sqrt{n-2}}{\sqrt{1-V^2}}$$

where

$$V = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Hence

$$\lambda = (1 - V^2)^{\frac{n}{2}}$$

gives

$$\lambda = \left[\frac{1}{1 + T^2/n-2}\right]^{\frac{n}{2}}$$

The test is therefore to compute

$$T = \frac{V\sqrt{n-2}}{\sqrt{1-V^2}}$$

and reject H_0 if $|T| > c$

Example

Suppose the following data were observed

X	2	7	3	5	10	9
Y	1	6	2	5	8	8

Test the hypothesis $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ take $\alpha = 0.05$

Solution

The test statistic is

$$T = \frac{V\sqrt{n-2}}{\sqrt{1-V^2}}$$

Where

$$V = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
2	1	-4	-4	16	16	16
7	6	1	1	1	1	1
3	2	-3	-3	9	9	9
5	5	-1	0	1	0	0
10	8	4	3	16	9	12
9	8	3	3	9	9	9
36	30	0	0	52	44	47

$$V = \frac{47}{\sqrt{52 \times 44}} = 0.9826$$

$$T = \frac{V\sqrt{n-2}}{\sqrt{1 - V^2}} = \frac{0.9826\sqrt{4}}{\sqrt{1 - 0.9826}} = 10.58$$

From Tables we have

$$t_{0.025,4} = 2.776$$

Therefore the computed value is greater than the table value thus we reject H_0 at $\alpha = 0.05$ significance level.