III-2-6. Student' s t distribution

Student's t distribution is used student's t test. Student's t test is test for significance of difference between two sample populations. Requirement of the test is normal distribution of date in each population. The shape is resembling to that of normal distribution. When the degree of freedom increases enough the shape is approximately same as normal distribution. T distribution varies with degree of freedom. This is the largest difference between t distribution and normal distribution. T distribution is produced by combination of normal distribution and chai square distribution.

T distribution is formulized by William Gosset. He was engineer working in Guinness company and was prohibited publication of scientific paper by the company. Student was his pen name. Probably, his question is similar as our question. "It is understandable that we can evaluate likeliness of the data to mean of parent population by the position of the observed value in standard normal distribution. For this purpose, we standardize the date by deviation of parent population σ. However, we do not know σ. We can only estimate the σ from the observed data and the estimated value fluctuates stochastically. We have to consider the stochastic variation "

We know normal distribution.

$$W(u) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}u^2}$$

$$u = \frac{x - \mu}{\sigma}$$

μ: average of parent population

σ: standard deviation of parent population

However, we cannot know σ and μ.

We estimate the average and standard deviation parent population from average and standard deviation of sample population.

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2$$

$$\mu \approx m$$

$$\sigma^2 \approx s^2$$

μ: mean of parent population

m: mean of sample population

σ^2: variance of parent population

s^2: estimated variance of parent population from sample data

Here, we discuss assuming the number of data is same between two sample populations. Actually, we have to compare two sample populations different in sample size. In such case we need to consider how to combine sample sizes. This skill will introduce in IV-2. Student's t test.

Here,

$$v = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2$$

$$u = \frac{x - \mu}{\sigma}$$

The function v is chai square.

$$v = {\chi^2}_n$$

$$\sigma^2 v = \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\sigma^2 = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n}$$

When we presume $\sigma = s$

$$s^2 = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n}$$

$$\sigma^2 v = \sum_{i=1}^{n} (x_i - \mu)^2 = ns^2$$

$$s^2 = \frac{1}{n}\sigma^2 v$$

$$s = \sigma\sqrt{\frac{v}{n}}$$

$$\frac{\sigma}{s} \approx \sqrt{\frac{n}{v}}$$

We presume x distribute in normal distribution and distribution of u is standard normal distribution.

$$u = \frac{x - \mu}{\sigma}$$

Most important discover by Gosset is independence of v and u (When v changes, u does not change. When u changes, v does not change.) So, we can calculate the possibility by products of possibility of u and v.

When we consider $x-\mu$ is difference between two expectation value such as average of two sub population. This is a presumption that one mean value is correct μ and the other is observed mean of the other sub population obtained by random sampling from the same parent population. However, we do not know correct σ. We have to use s as estimated value of σ instead of σ.

$$t=\frac{\bar{x}-\mu}{s}$$

However, we can transform this formula as follow

$$t=\frac{\bar{x}-\mu}{\sigma}\cdot\frac{\sigma}{s}=u\sqrt{\frac{n}{v}}$$

$$\left(\because u=\frac{x-\mu}{\sigma}, \qquad \frac{\sigma}{s}=\sqrt{\frac{n}{v}}\right)$$

Actually, Gosset used $z=\frac{t}{\sqrt{n-1}}$ in his publication, and t is used in the book of Fisher lately, however basic idea to combine u and v as t was proposed by Gosset.

Possibility distribution of u is normal distribution, and possibility distribution of v is chai square distribution.

$$\mathrm{W}(u)=\frac{1}{\sqrt{2\pi}}e^{\frac{-1}{2}u^2}$$

$$\mathrm{P}(v)=\frac{z^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{\frac{-v}{2}}$$

So, possibility of t is as follow, because u and v are independent each other.

$$\mathrm{S}(t)=\mathrm{W}(u)\mathrm{P}(v)$$

This formula expresses basic structure of t distribution, though S(t) is not expressed by t. We have to converse coordinate using the relation $t=u\sqrt{\frac{n}{v}}$

Here, we have to convert integration is u-v coordinate to integration in s-t coordinate. For this we need knowledge of Jacobian. Jacobian is a kind of expansion ratio with the conversion. The readers who need knowledge of Jacobian and coordinate conversion, please read III-3-3. Jacobian, III-3-4. Polar coordinate, III-3-5. Multiple integral.

Right side members are independent each other, and the sum of the probability is 1.

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \mathrm{W}(u)\mathrm{P}(v)dudv \ = 1$$

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \mathrm{W}(u)\mathrm{P}(v)dudv$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \mathrm{J}(u, v/t, s)\mathrm{W}(u)\mathrm{P}(v)dsdt$$

$\mathrm{J(u, v/t, s)}$ is Jacobian

$$v = w$$

$$t = u\sqrt{\frac{n}{v}}$$

$$\frac{du}{dt} = \frac{\sqrt{v}}{\sqrt{n}}, \quad \frac{du}{dw} = o, \ \frac{dv}{dt} =, \ \frac{dv}{dw} = 1$$

$$\mathrm{J} = \begin{bmatrix} \dfrac{du}{dt} & \dfrac{du}{dw} \\ \dfrac{dv}{dt} & \dfrac{dv}{dw} \end{bmatrix} = \begin{bmatrix} \dfrac{\sqrt{v}}{\sqrt{n}} & 0 \\ \dfrac{-2nu^2}{t^{-3}} & 1 \end{bmatrix} = \frac{\sqrt{v}}{\sqrt{n}}$$

$$\mathrm{J}(u, v/t, w) = \frac{\sqrt{v}}{\sqrt{n}}$$

$$\mathrm{W(u)} = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}u^2}$$

$$\mathrm{P}(v) = \frac{z^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{\frac{-v}{2}}$$

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}u^2} \cdot \frac{v^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{\frac{-v}{2}} dudv$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{\sqrt{v}}{\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}u^2} \cdot \frac{v^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{\frac{-v}{2}} dwdt$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot v^{\frac{1}{2}} v^{\frac{n}{2}-1} e^{\frac{-1}{2}u^2} e^{\frac{-v}{2}} dwdt$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot v^{\frac{n-1}{2}} e^{\frac{-(u^2+v)}{2}} dwdt$$

$$\mathrm{t} = \frac{\sqrt{n}u}{\sqrt{v}}, \ w = v,$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot w^{\frac{n-1}{2}} e^{\frac{-(\frac{t^2}{n^2}v+v)}{2}} dwdt$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot w^{\frac{n-1}{2}} e^{\frac{-(\frac{t^2}{n^2}+1)w}{2}} dwdt$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot w^{\frac{n-1}{2}} e^{-(\frac{t^2}{n^2}+1)\frac{w}{2}} dsdt$$

$$(\frac{t^2}{n^2}+1)\frac{w}{2} = q$$

$$\frac{dq}{dw} = (\frac{t^2}{n^2}+1)\frac{1}{2}$$

$$w = \frac{2q}{(\frac{t^2}{n^2}+1)}$$

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot w^{\frac{n-1}{2}} e^{-(\frac{t^2}{n^2}+1)\frac{w}{2}} dwdt$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot \left(\frac{2q}{(\frac{t^2}{n^2}+1)}\right)^{\frac{n-1}{2}} e^{-w} \frac{2}{(\frac{t^2}{n^2}+1)} dqdt$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \cdot \left(\frac{2}{(\frac{t^2}{n^2}+1)}\right)^{\left(\frac{n+1}{2}\right)} q^{\frac{n-1}{2}} e^{-q} dqdt$$

$$= \int_{-\infty}^{\infty} \frac{2^{\frac{n}{2}+1}}{\sqrt{2n\pi}2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)(\frac{t^2}{n^2}+1)^{\frac{n}{2}+1}} \left(\int_{0}^{\infty} q^{\left(\frac{n+1}{2}-1\right)} e^{-q} dq\right) dt$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)(\frac{t^2}{n^2}+1)^{\frac{n}{2}+1}} \left(\int_{0}^{\infty} q^{\left(\frac{n+1}{2}-1\right)} e^{-q} dq\right) dt$$

$$= \int_{-\infty}^{\infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)(\frac{t^2}{n^2}+1)^{\frac{n}{2}+1}} dt$$

$\because \int_0^\infty w^{\left(\frac{n+1}{2}-1\right)} e^{-w} dw$ is Γ function.

$$\int_0^\infty q^{\left(\frac{n+1}{2}-1\right)} e^{-q} dq = \Gamma\left(\frac{n+1}{2}\right)$$

$$\mathrm{S(t)} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)(\frac{t^2}{n^2}+1)^{\frac{n}{2}+1}}$$

We can simplify the formula using β function. β function is combination of Γ function.

$$\beta(p,q)) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

Formula 22

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$S(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)(\frac{t^2}{n^2}+1)^{\frac{n}{2}+1}}$$

$$S(t) = \frac{1}{\sqrt{n}\beta\left(\frac{n}{2},\frac{1}{2}\right)(\frac{t^2}{n^2}+1)^{\frac{n}{2}+1}}$$

Formula 23

Estimation of t value of data from data is as follow.

$$\mathrm{t} = \frac{\bar{x}-\mu}{\sigma}\cdot\frac{\sigma}{s} = \frac{\bar{x}-\mu}{s}$$

When we estimate quadratic moment around μ form actual data, quadratic moment around μ is $\frac{\sigma^2}{n}$, so we have to replace σ by $\frac{\sigma}{\sqrt{n}}$.

$$\mathrm{t} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\cdot\frac{\sigma}{s} = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

Formula 24