Student's-t– distribution (W.S. Gosset-1908)

The entire topic is divided into 7 sub-divisions

1. **Introduction and definition**
2. **Derivation of the pdf of Student's t-distribution**
3. **Probability curve of Student's t-distribution**
4. **Derivation of the pdf of Fisher's t-distribution**
5. **Properties of t- distribution**
6. **Applications of t-distribution**
7. **Summary**

1. Introduction and definition of Students t-distribution:

When the sample size n is small (say below 30) and standard deviation is unknown, the distribution of the statistic $t = \frac{\bar{x} - \mu}{S / \sqrt{n}}$ is far from normality, and as such, normal test can not be applied. In such cases exact sample tests like students t-test, F-test and etc., could be used.

i.a. Definition:

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n, from $N(\mu, \sigma^2)$ population. Then the student's t-statistic is defined as

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}} \sim \text{Student's t distribution with (n-1) d.f.} \qquad (1)$$

with p.d.f. defined by

$$f(t) = \frac{1}{\sqrt{n-1}\, B(\frac{1}{2}, \frac{n-1}{2})} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}}; \ for \ -\infty < t < \infty$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$, which is an unbiased estimator of the population variance and $B(\frac{1}{2}, \frac{n-1}{2}) = \frac{\Gamma\frac{1}{2} \times \Gamma\frac{n-1}{2}}{\Gamma(\frac{n}{2})}$.

2. Derivations of the pdf of Student's t-distribution:

Equation (1) defined above, can be written as

$$t^2 = \frac{n(\bar{x} - \mu)^2}{S^2} = \frac{n(\bar{x} - \mu)^2}{ns^2/(n-1)}$$

$$\Rightarrow \frac{t^2}{n-1} = \frac{(\bar{x} - \mu)^2}{\sigma^2/n} \times \frac{1}{ns^2/\sigma^2} = \frac{(\bar{x} - \mu)^2/(\sigma^2/n)}{ns^2/\sigma^2}.$$

Since $x_1, x_2, \ldots, x_n$ be a random sample of size n, from $N(\mu, \sigma^2)$ population, then $\bar{x} \sim N(\mu, \sigma^2/n)$. $\Rightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$. Hence $\frac{(\bar{x} - \mu)^2}{\sigma^2/n}$ being the square of a standard normal variate, is a chi-square variate with 1 d.f. Also, (ns^2/σ^2) is a chi-square variate with (n-1) d.f. Further, since $\bar{x}$ and s^2 are independently distributed , $\frac{t^2}{n-1}$ being the ratio of two independent χ^2 variates with 1 and (n - 1) d.f. respectively, is a β_2[1/2, (n-1)/2] variate and its distribution is given by

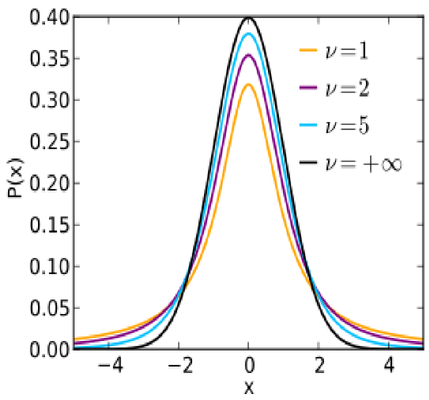
$$dF(t) = \frac{1}{B(\frac{1}{2}, \frac{n-1}{2})} \cdot \frac{(\frac{t^2}{n-1})^{1/2 - 1}}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}} dt; \quad \text{for } 0 < t^2 < \infty$$

$$= \frac{1}{\sqrt{n-1}\, B(\frac{1}{2}, \frac{n-1}{2})} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}} dt; \quad for \quad -\infty < t < \infty.$$

Which is the p.d.f. of Student's t-distribution with (n-1) d.f., where the factor 2 disappears, because, the integral form $-\infty < t < \infty$ must be unity.

3. Probability curve of Student's t-distribution:

Like the normal distribution, the curve of *t*-distribution is symmetric and bell-shaped, that is, f(-t)=f(t), the probability curve is symmetrical about the line t=0. But it has heavier tails, that is, it is more flat to produce values that fall far from its mean. As t increases, f(t) decreases rapidly. Here we observe that, for the degrees of freedom n = 1 and 2, the probability curve show more flat than when n-- >∞.



1. **Derivations of the pdf of Fisher's t-distribution:**

Prof. R.A. Fisher gave the, rigorous proof for the statistic t's sampling distribution. According to Fisher, the statistic 't' is the ratio of a standard normal variate to the square root of an independent chi-square variate , divided by its degree of freedom. Thus, the statistic 't' is defined as

$t = \frac{Z}{\sqrt{X/n}}$ ~student's t - distribution with n d.f. where

Z ~N(0, 1) and X~ $\chi^2_{(n)}$ d.f.

Since X and Z are independent, their joint p.d.f. is given by

$$f(x,z)=\frac{1}{\sqrt{2\pi}\,2^{n/2}\Gamma(n/2)}e^{-z^2/2}e^{-x/2}x^{\frac{n}{2}-1}, \quad for\ 0<x<\infty \text{ and } -\infty<z<\infty$$

Consider the following transformation in variables:

Let U = X, and we have $t = \frac{Z}{\sqrt{X/n}}$ => Z = t √(u/n)

Now, the Jacobian transformation J is given by

$$J = \frac{\partial(Z,X)}{\partial(t,U)} = \begin{vmatrix} \sqrt{u/n} & t/(2\sqrt{un}) \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{n}}$$

The joint p.d.f. of t and u becomes

$$g(t,u)=\frac{1}{\sqrt{2\pi}2^{n/2}\Gamma(n/2)\sqrt{n}}e^{-\frac{u}{2}(1+t^2/n)}u^{\frac{n-1}{2}}du$$

(1)

Since X ≥ 0, and $-\infty < z < \infty$, *we have* $u \geq 0$ *and* $-\infty < t < \infty$. Integrating (1) w. r. t. u over the range 0 to ∞, the marginal p.d.f. g(t) of t becomes

$$g(t)=\frac{1}{\sqrt{2\pi}\,2^{n/2}\Gamma(n/2)\sqrt{n}}\int_0^\infty e^{-\frac{u}{2}(1+t^2/n)]}u^{\frac{n-1}{2}}du$$

$$=\frac{1}{\sqrt{2\pi}\,2^{n/2}\Gamma(n/2)\sqrt{n}}\frac{\Gamma[(n+1)/2]}{\left[\frac{1}{2}\left(1+\frac{t^2}{n}\right)\right]^{(n+1)/2}}$$

$$=\frac{\Gamma[(n+1)/2]}{\sqrt{n}\,\Gamma(n/2)\Gamma(1/2)}\frac{1}{\left[1+\frac{t^2}{n}\right]^{(n+1)/2}},$$

$$\therefore g(t)=\frac{1}{\sqrt{n}\,B(\frac{1}{2},\frac{n}{2})}\cdot\frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}},\quad \textit{for}\quad -\infty < t < \infty.$$

which is the pdf of Student's t- distribution with n d.f., where $B(\frac{1}{2},\frac{n}{2})=\frac{\Gamma\frac{1}{2}\times\Gamma\frac{n}{2}}{\Gamma(\frac{n+1}{2})}$.

Note : Students t-distribution is a particular case of Fisher's t- distribution.

5. Properties of t- distribution with n d.f. :

1. When n = 1, *f(t)* reduces to

   $$f(t)=\frac{1}{\pi}\frac{1}{\left(1+t^2\right)};\quad \text{for } -\infty < \text{t} < \infty$$

   which is the pdf of Cauchy distribution,
2. All odd ordered moments are zero and all the even ordered moments exist and are constants.

3. Mean = 0 and Variance = n /(n-2), for n > 2.

4. Mean deviation about mean is
$$\frac{\sqrt{n}\,\Gamma[(n-1)/2]}{\sqrt{\pi}\,\Gamma(n/2)}$$

5.a. Moments of t-distribution :

Since f(t) is symmetric about the line t=0, all the moments of odd order about origin vanish, that is

$\mu'_{2r+1}(about\ origin) = 0,\ for\ r = 0,1,2,...$

In particular, $\mu'_1(about\ origin) = 0 = mean.$ Therefore, the central moments coincide with moments about origin => $\mu_{2r+1} = 0,\ for\ r = 0,1,2,...$

The moments of even order are given by

$\mu_{2r} = \mu'_{2r}(about\, origin) = \int_{-\infty}^{\infty} t^{2r} f(t)dt = 2\int_0^{\infty} t^{2r} f(t)dt$

$$\therefore \mu_{2r} = 2.\frac{1}{\sqrt{n}\,B(\frac{1}{2},\frac{n}{2})}\int_0^{\infty}\frac{t^{2r}}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}dt$$

The above integral is convergent, if 2r < n.

Now let $\left(1+\frac{t^2}{n}\right) = \frac{1}{x} \Rightarrow t^2 = \frac{n(1-x)}{x} \Rightarrow 2tdt = -\frac{n}{x^2}dx$

When t = 0, x = 1 and when t = ∞, x = 0. Therefore, we have

$$\mu_{2r} = 2.\frac{1}{\sqrt{n}\,B(\frac{1}{2},\frac{n}{2})}\int_1^{0}\frac{t^{2r}}{\left(\frac{1}{x}\right)^{(n+1)/2}}.\frac{-n}{2tx^2}dx$$

$$= \frac{\sqrt{n}}{B(\frac{1}{2}, \frac{n}{2})} \int_0^1 (t^2)^{(2r-1)/2} . x^{\frac{(n+1)}{2} - 2} dx$$

$$= \frac{\sqrt{n}}{B(\frac{1}{2}, \frac{n}{2})} \int_0^1 \left(\frac{n(1-x)}{x} \right)^{r-1/2} . x^{\frac{(n+1)}{2} - 2} dx$$

$$= \frac{n^r}{B(\frac{1}{2}, \frac{n}{2})} \int_0^1 (1-x)^{r-1/2} . x^{\frac{n}{2}-r-1} \quad dx$$

$$= \frac{n^r}{B(\frac{1}{2}, \frac{n}{2})} B(\tfrac{n}{2} - r, \ r + \tfrac{1}{2})$$

$$\therefore \mu_{2r} = \frac{n^r \Gamma[(n/2) - r].\Gamma(r + 1/2)}{\Gamma(\frac{1}{2}).\Gamma(\frac{n}{2})}$$

$$= \frac{n^r (r - \frac{1}{2})(r - \frac{3}{2}) .... \frac{3}{2} . \frac{1}{2} \Gamma(\frac{1}{2}) \Gamma(\frac{n}{2} - r)}{\Gamma(\frac{1}{2}).(\frac{n}{2} - 1).(\frac{n}{2} - 2)...(\frac{n}{2} - r)\Gamma(\frac{n}{2} - r)}$$

$$= \frac{n^r (2r-1)(2r-3)....3.1}{(n-2).(n-4)...(n-2r)}, \ n > 2r.$$

In particular,

$$\mu_2 = \frac{n}{n-2} = \text{variance}, \ \ for \ \ n > 2.$$

Note : MGF of t-distribution does not exist, because all the moments of order n > 2r exist, but the moments of order n ≤ 2r do not exist.

5.b. Mean deviation about mean of t-distribution:

Since t ~$t_{(n)}$ d.f. then E(t) = 0 and hence,

Mean deviation (about mean)= $\int_{-\infty}^{\infty} |t| \ f(t) dt$

$$= \frac{1}{\sqrt{n}\, B(\frac{1}{2}, \frac{n}{2})} \int_{-\infty}^{\infty} \frac{|t|}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}} dt$$

$$= 2.\frac{1}{\sqrt{n}\, B(\frac{1}{2}, \frac{n}{2})} \int_{0}^{\infty} \frac{t}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}} dt$$

$$= \frac{\sqrt{n}}{B(\frac{1}{2}, \frac{n}{2})} \int_{0}^{\infty} \frac{1}{(1+x)^{(n+1)/2}} dx, \; where \;\; x = \frac{t^2}{n}$$

$$= \frac{\sqrt{n}}{B(\frac{1}{2}, \frac{n}{2})} \int_{0}^{\infty} \frac{x^{1-1}}{(1+x)^{(n-1)/2\,+\,1}} dx,$$

$$= \frac{\sqrt{n}}{B(\frac{1}{2}, \frac{n}{2})} .B(\tfrac{n\,-\,1}{2}, 1) = \frac{\sqrt{n}\;\; \Gamma[(n-1)/2]}{\sqrt{\pi}\;\; \Gamma(n/2)}.$$

5.c. Limiting form of t- distribution :

Here we discuss about the form t-distribution when t-- > ∞.

Proof: Given t ~ $t_{(n)}$ d.f., then we have

$$f(t) = \frac{1}{\sqrt{n}\, B(\frac{1}{2}, \frac{n}{2})} \frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, \;\; for \quad -\infty < t < \infty$$

Now, taking limit as n-- > ∞, on both sides, we get,

$$\lim_{n\to\infty} f(t) = \lim_{n\to\infty} \left(\frac{1}{\sqrt{n}\, B(\frac{1}{2}, \frac{n}{2})} . \frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}} \right)$$

$$= \lim_{n \mapsto \infty} \frac{\Gamma[(n+1)/2}{\sqrt{n}\,\Gamma\frac{1}{2}.\Gamma\frac{n}{2}} . \lim_{n \to \infty}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \qquad (1)$$

Since, $\Gamma\frac{1}{2} = \sqrt{\pi}, \quad and \quad \lim_{n\to\infty}\frac{\Gamma(n+k)}{\Gamma n} = n^k$, therefore equation (1) reduces to

$$= \frac{1}{\sqrt{2\pi}} . \lim_{n\to\infty}\left[\left(1+\frac{t^2}{n}\right)^n\right]^{-1/2} \times \lim_{n\to\infty}\left(1+\frac{t^2}{n}\right)^{-1/2}$$

$$\therefore \lim_{n\to\infty} f(t) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{t^2}{2}}, \; for \quad -\infty < t < \infty.$$

which is the pdf of a standard normal variate. Hence, for large n, that is, when n--> ∞, t distribution ~ N(0, 1) distribution asymptotically.

6. Applications of t-distribution

The t- distribution has a wide range of applications, some of which are

i. To test the significance of sample mean from the population mean or simply, t-test for single mean.
ii. To test the significance of the difference between two independent sample means.
iii. Paired t-test for difference between two sample means.
iv. To test the significance of an observed sample correlation coefficient
v. To test the significance of an observed sample regression coefficient.
vi. To test the significance of an observed partial correlation coefficient.

Critical values of t-distribution:

The critical or significant values of t at level of significance α and d.f.(υ = n-1) for two tailed test are given by the equation

$P[\,|t| > t_\upsilon(\alpha)\,] = \alpha \quad \Rightarrow P[\,|t| \le t_\upsilon(\alpha)\,] = 1-\alpha.$

The values of $t_\upsilon(\alpha)$ can be obtained from student's t-table. Since t- distribution is symmetric about t = 0, we have

$$P[\,t > t_\upsilon(\alpha)\,] + P[\,t < -t_\upsilon(\alpha)] = \alpha$$

$\Rightarrow 2.\,P[\,t > t_\upsilon(\alpha)\,] = \alpha \;\Rightarrow P[\,t > t_\upsilon(\alpha)\,] = \alpha/2.$

Therefore, $P[\,t > t_\upsilon(2\alpha)\,] = \alpha$, where, t_{υ} (2α) gives the significant value of t for a singe tail test(right or left) at level of significance α and d.f.(υ).

7. Summary :

So far we have discussed in deep, the definition, the nature of probability curve, derivation of the p.d.f., and proved, some of the properties of t- distribution. Also, we have listed out some of the applications of t- distribution.