Chapter 1

SAMPLING DISTRIBUTIONS

1.1 Introduction

Statistics concerns itself mainly with conclusions and predictions resulting from chance outcomes that occur carefully experiments or investigations. In the finite case, these chance outcomes constitute a subset or ***sample*** of measurements or observations from a larger set of values called the ***population***. In the continuous case they are usually values of i.i.d (independent identically distributed) random variables, whose distribution we refer to as the ***population distribution***, or the ***infinite population sampled***. The word "infinite" implies that there is, logically speaking, no limit to the number of values we could observe.

Definition 1.1 Population

The totality of elements which are under discussion or investigation and about which information is desired will be called the ***target population.***

Definition 1.2 Random sample

If $X_1, X_2, ..., X_n$ are i.i.d. r.v.'s, we say that they constitute a random sample (abbreviated by R.S.) from the infinite population given by their common distribution

An important part f the definition of a R.S. is the meaning of the r.v.'s $X_1, X_2, ..., X_n$. The r.v. X_i is a representation for the numerical value that the i^{th} item (or element) sampled will be assumed. After the sample is observed, the actual values of $X_1, X_2, ..., X_n$ are known, we denote these observed values by $x_1, x_2, ..., x_n$.

One of central problems in statistics is the following:

If it is desired to study a population which has a known density function but it contains some unknown parameters. For example, suppose that we have a population, which has the normal distribution, but the parameters μ and σ^2 are unknown.

The procedure is to take a random sample (R.S.) $X_1, X_2, ..., X_n$ of size n from this population and let the value of some function represent or estimate the unknown parameter. This function is called a ***statistic***. Since many random samples are possible from the same population, we would expect every statistic to vary somewhat from sample to sample. Hence a statistic is a random variable, and as such it must have a probability distribution.

Definition 1.3

If $\mathbf{X_1, X_2, ..., X_n}$ are i.i.d. r.v.'s, we say that they constitute a random sample from the infinite population given by their common distribution.

If $f(x_1, x_2, ..., x_n)$ is the joint p.m.f. (or p.d.f.) of such a set of r.v.'s , we can write

$$\mathbf{f(x_1, x_2, ..., x_n) = f(x_1)\, f(x_2) \ldots f(x_n) = \prod_{i=1}^{n} f(x_i)}$$

where $\mathbf{f(x_i)}$ is the common p.m.f. (or p.d.f.) of each X_i (or of the population sampled).

Definition 1.4

A statistic is a function of observable r.v.'s, which is itself an observable r.v. and does not contain any unknown parameter.

For example, if $\mathbf{X_1, X_2, ..., X_n}$ is a r.s., then the sample mean

$$\mathbf{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

and the sample variance

$$\mathbf{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

are statistics.

Since statistics are r.v.'s, their values will vary from sample to sample, and it is customary to refer to their distributions as ***sampling distributions***. Note that;

$$\mathbf{E\left(\overline{X}\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n} n\mu = \mu}, \text{ and}$$

$$\mathbf{Var\left(\overline{X}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}}$$

1.2 Stochastic Convergence

Sometimes the distribution of a r.v. (perhaps a statistic) depends upon a +ve integer n. Clearly, the distribution function (CDF) **F** of that r.v. will also depends upon n and denote it by $\mathbf{F_n}$. We now define a limiting distribution of a r.v. whose distribution depends upon n.

Definition 1.5

Let the CDF $\mathbf{F_n}(x)$ of the r.v. $\mathbf{X_n}$ depends upon **n**. If

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every point y at which F(y) is continuous where F(x) is a CDF, then the r.v. $\mathbf{X_n}$ is said to have **a limiting distribution with CDF F(x).**

Theorem 1.1 *(Weak law of large numbers)*

Let $\mathbf{(X_n)}_{\mathbf{n=1,2,\ldots}}$ be a sequence of independent r.v's such that all have the same expectation, $\mathbf{E(X_n) = \mu}$ and the same variance $\mathbf{Var(X_n) = \sigma^2}$ then we say that the r.v. $\mathbf{X_n}$ converges stochastically (or in probability) to the constant μ iff, for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbf{P}(|\mathbf{X_n} - \mu| < \varepsilon) = \mathbf{1}$$

and we may write

$$\mathbf{X_n} \xrightarrow{\mathbf{p}} \mu$$

We should like to point out a simple but useful fact. Clearly

$$\mathbf{P}(|\mathbf{X_n} - \mu| < \varepsilon) + \mathbf{P}(|\mathbf{X_n} - \mu| \geq \varepsilon) = \mathbf{1}$$

Thus the limit of $P(|X_n - \mu| < \varepsilon)$ is equal to 1 when and only when

$$\lim_{n\to\infty} \mathbf{P}(|\mathbf{X_n} - \mu| \geq \varepsilon) = \mathbf{0}$$

That is, the last limit is also a necessary and sufficient condition for the stochastic convergence of the r.v. $\mathbf{X_n}$ to μ.

In addition, a very important result on law of large numbers, is:

Let $\bar{\mathbf{X}}_\mathbf{n}$ be the mean of a random sample of size n, then;

$$\bar{\mathbf{X}}_\mathbf{n} \xrightarrow{\mathbf{p}} \mu \quad \Leftrightarrow \quad \lim_{n\to\infty} \mathbf{P}(|\bar{\mathbf{X}}_\mathbf{n} - \mu| < \varepsilon) = \mathbf{1}$$

1.3 The Distribution of $\bar{X}$

Theorem 1.2

If $\mathbf{X_1, X_2, \ldots, X_n}$ constitute a R.S. from a normal population with mean μ and variance σ^2, then $\bar{\mathbf{X}} \sim \mathbf{N(\mu, \sigma^2/n)}$ i.e. $\mathbf{Z} = \dfrac{\bar{\mathbf{X}} - \mu}{\sigma/\sqrt{\mathbf{n}}} \sim \mathbf{N(0,1)}$.

Proof

Since $\overline{X}$ is the mean of a r.s. , then $X_1, X_2, \ldots, X_n$ are i.i.d. with common M.G.F.,

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Now, the M.G.F. of $\overline{X}$ is,

$$M_{\overline{X}}(t) = E\left[e^{\overline{X} t}\right] = E[e^{(\frac{1}{n}\sum_{i=1}^{n} X_i)t}]$$

$$= E[\ e^{X_1\left(\frac{t}{n}\right)}\ ]\, E[\ e^{X_2\left(\frac{t}{n}\right)}\ ] \ldots E[\ e^{X_n\left(\frac{t}{n}\right)}\ ]$$

$$= M_{X_1}\left(\frac{t}{n}\right) M_{X_2}\left(\frac{t}{n}\right) \ldots M_{X_n}\left(\frac{t}{n}\right)$$

$$= \left[M_X\left(\frac{t}{n}\right)\right]^n$$

$$= \left(e^{\mu(\frac{t}{n}) + \frac{1}{2}\sigma^2(\frac{t}{n})^2}\right)^n$$

$$= e^{\mu t + \frac{1}{2}(\frac{\sigma^2}{n})t^2}$$

which is the M.G.F. of $N(\mu, \sigma^2/n)$. Therefore $\overline{X} \sim N(\mu, \sigma^2/n)$ i.e. $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

From this theorem we also conclude that the mean and variance of $\overline{X}$ are given by

$$\mu_{\overline{X}} = E(\overline{X}) = \mu, \qquad \text{and} \qquad \sigma^2_{\overline{X}} = var(\overline{X}) = \frac{\sigma^2}{n} \qquad \textbf{(1.1)}$$

Theorem 1. 3 (Central Limit Theorem)

If random samples of size n are drawn from any infinite population with mean μ and variance σ^2, the limiting distribution

$$Z = \frac{\overline{X} - \mu}{\sigma\sqrt{n}}$$

as $n \to \infty$ is the standard normal distribution N(0,1).

The Proof is omitted.

The Normal approximation in the central limit theorem will be good if n ≥ 30 regardless of the shape of the population. If the population variance σ2 is unknown, the central limit theorem still valid when we replace σ2 by the sample variance $\mathbf{S^2}$, i.e. for large **n** enough, we have

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$$

It is interested to note that when the population we are sampling is normal, the distribution of $\bar{X}$ is a normal distribution (see theorem 1.1) regardless of the size of n.

Example 1.1

Certain tubes manufactured by a company have a mean lifetime of 900 hr and standard deviation of 50 hr. Find the probability that a random sample of 64 tubes taken from the group will have a mean lifetime between 895 and 910 hrs.

Solution

Here we have μ = 900, σ = 50. Let $\bar{X}$ denotes the sample mean lifetime of the tubes and since n = 64 is large enough, then by the central limit theorem

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Thus

$$P(895 < \bar{X} < 910) = P\left(\frac{895 - 900}{50/8} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{910 - 900}{50/8}\right) = P(-8.0 < Z < 1.6)$$

$$= \Phi(1.6) - \Phi(-0.8) = \Phi(1.6) - 1 + \Phi(0.8) = 0.733$$

Case of two populations

If two **independent** random samples of sizes $\mathbf{n_1}$ and $\mathbf{n_2}$ are drawn from any two populations with means $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$, and variances σ_1^2 and σ_2^2, respectively, the sampling distribution of $\bar{X}_1 - \bar{X}_2$ will be approximately distributed with mean and variance given by

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2 , \quad \text{and} \quad \sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Hence,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

If both n_1 and n_2 are greater than or equal to 30, the normal approximation for the distribution of $\bar{X}_1 - \bar{X}_2$ will be good regardless of the shapes of the two populations. Similarly, if the variances σ_1^2 and σ_2^2 are unknown, the central limit theorem still valid with using the sample variances S_1^2 and S_2^2 instead of σ_1^2 and σ_2^2. Therefore

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$$

Example 1.2

The electric light bulbs of manufacturer A have a mean lifetime of 1400 hrs with a standard deviation of 200 hrs, while those of manufacturer B have a mean lifetime of 1200 hours with a standard deviation of 100 hours. If random samples of 125 bulbs of each brand are tested, what is the probability that the brand A bulbs will have a mean lifetime which is at least 160 hours more than the brand B bulbs?

Solution

Let $\bar{X}_A$ and $\bar{X}_B$ denote the mean lifetimes of samples A and B respectively. Then the variable

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - \mu_{\bar{X}_A - \bar{X}_B}}{\sigma_{\bar{X}_A - \bar{X}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 200}{20} \sim N(0,1)$$

The required probability is then, given by

$$P(\bar{X}_A - \bar{X}_B \geq 160) = P\left(Z \geq \frac{160 - 200}{20}\right) = P(Z \geq -2.0) = 1 - \Phi(-2.0) = \Phi(2.0) = 0.977$$

1.4 The Chi-Squared Distribution

If $Z_1, Z_2, \ldots, Z_v$ are independent r.v.'s having standard normal distribution N(0,1), then the r.v.

$$U = Z_1^2 + Z_2^2 + \ldots + Z_\nu^2 \quad \textbf{(1.2)}$$

has the so called Chi-Squared Distribution (often denoted by χ^2 distribution) with **ν** degrees of freedom (d.f.) and it has the following properties;

1- The mean and variance of the χ^2_ν distribution are

$$\mu = \nu \quad \text{and} \quad \sigma^2 = 2\,\nu$$

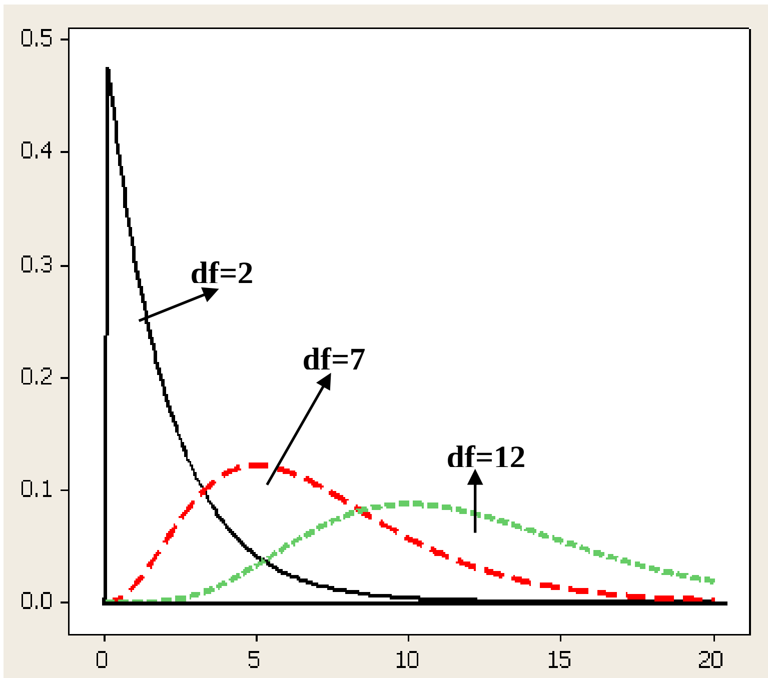


Fig. 1.1 χ^2_ν distribution curves for various values of **ν**

3- If $\mathbf{U_1}, \mathbf{U_2}, \ldots, \mathbf{U_k}$ are independent r.v.'s having chi-squared distributions with $v_1, v_2, \ldots, v_k$ d.f., then

$$\mathbf{Y} = \sum_{i=1}^{k} \mathbf{U}_i$$

has the chi-squared distribution with $\boldsymbol{\nu} = \boldsymbol{\nu_1} + \boldsymbol{\nu_2} + \ldots + \boldsymbol{\nu_k}$ d.f.

4- The percentage points of the χ^2_ν distribution have been extensively tabulated. Define $\chi^2_{\alpha,\nu}$ as the percentage point or value of the chi-square r.v. **U** with **ν** d.f. such that

$$P(U \ge \chi^2_{\alpha,\nu}) = \int_{\chi^2_{\alpha,\nu}}^{\infty} f_{\chi^2}(u)\,du = \alpha$$

This probability is shown as the shaded area in Fig.1.2

Note that if $\mathbf{X_1, X_2, \dots, X_n}$ constitute a R.S. from a normal population with mean μ and variance σ^2, then $\mathbf{Z_i = \dfrac{X_i - \mu}{\sigma} \sim N(0,1)}$ and therefore the variable

$$\mathbf{U = \sum_{i1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2} .$$

Has the χ^2 distribution with $\mathbf{v}$ = n d.f.

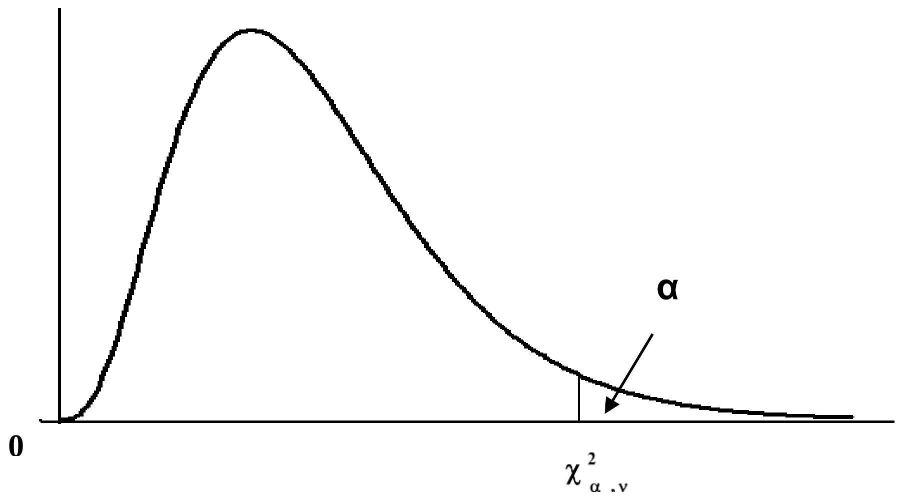
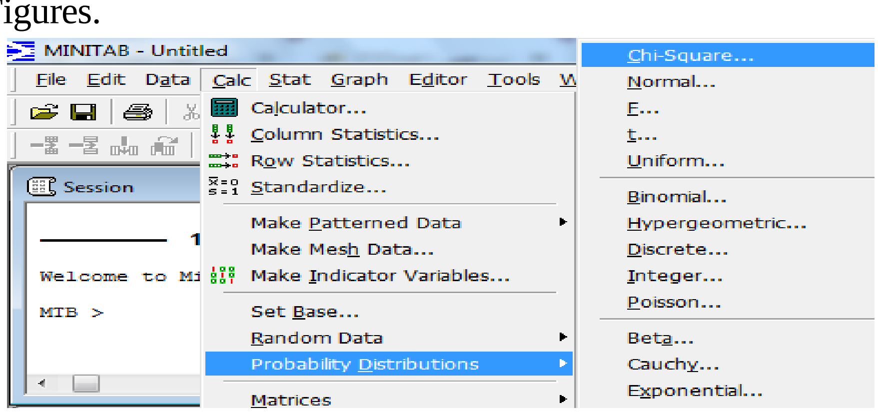
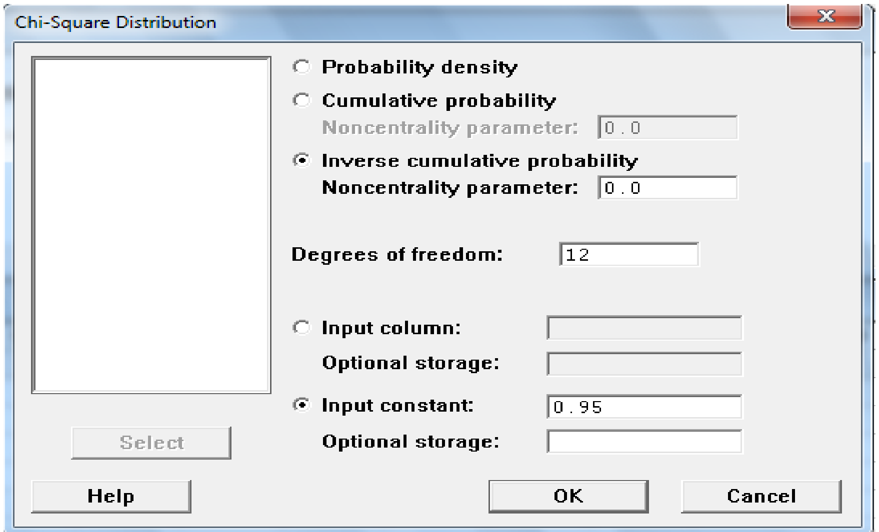


Fig. 1.2 Percentages point of the chi-squared distribution

Using MINITAB

Suppose we want to find $\chi^2_{.05,12}$, then its CDF is F($\chi^2_{.05,12}$) = 1- $\boldsymbol{\alpha}$ = 0.95. Now press ***Calc*** $\longrightarrow$ ***Probabilty Distributions*** $\longrightarrow$ ***Chi-Square*** then click on "inverse cumulative distribution" and write 12 for the degrees of freedom and 0.95 for "input constant" as in the following Figures.





Click on " "ok" we obtain:

```
Chi-Square with 12 DF

P( X <= x )          x
     0.95          21.0261
```

i.e. $\chi^2_{.05,12} = 21.02$.

The Distribution of S^2

Theorem 1.4

If $X_1, X_2, \ldots, X_n$ are a random sample from $N(\mu,\sigma^2)$, then

1- $\bar{X}$ and the terms $X_i - \bar{X}$; i=1,..., n are independent,

2- $\bar{X}$ and S^2 are independent.

The proof is omitted.

Theorem 1.5

If $\bar{X}$ and S^2 are the mean and variance of a r.s. of size n from a population having $N(\mu,\sigma^2)$, then

1- $\bar{X}$ and S^2 are independent;

2- the r.v. $U = \dfrac{(n-1)S^2}{\sigma^2}$ has the chi-squared distribution with ν = n-1 d.f.

Proof

First note that by adding and subtracting $\bar{X}$ and then expanding, we obtain the relationship

$$U = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} + \frac{n(\overline{X} - \mu)^2}{\sigma^2} \qquad (6.3)$$

$$= \quad U_1 \quad + \quad U_2$$

Since

$$U = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

and

$$\overline{X} \sim N(\mu, \sigma^2/n) \Rightarrow Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \Rightarrow Z^2 = \frac{(\overline{X} - \mu)^2}{\frac{\sigma^2}{n}} = \frac{n(\overline{X} - \mu)^2}{\sigma^2} = U_2 \sim \chi_1^2$$

Thus,

$$U \sim \chi_n^2 \quad \text{and} \quad U_2 \sim \chi_1^2$$

Therefore by property (3) of the chi-square distribution we have

$$U_1 = U - U_2 \sim \chi_{n-1}^2$$

Corollary

Since the mean and variance of the χ_{n-1}^2 are respectively, (n-1) and 2(n-1), it follows that

$$E\left(\frac{(n-1)S^2}{\sigma^2} \right) = n - 1 \quad \Rightarrow \quad E(S^2) = \sigma^2$$

and

$$Var\left(\frac{(n-1)S^2}{\sigma^2} \right) = 2(n-1) \quad \Rightarrow \quad var(S^2) = \frac{2\sigma^4}{n-1}$$

1.5 The t- Distribution

We know that if $X_1, X_2, ..., X_n$ are a R.S. from a normal population with mean μ and variance σ^2, then $\overline{X} \sim N(\mu, \sigma^2/n)$ i.e.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Most of the time we are not fortunate enough to know the variance of the population from which we select our random samples. For samples of size n < 30, a good estimate of σ^2 is provided by calculating S^2. What then happens to the distribution of the Z values in the central limit theorem if we replace σ^2 by S^2? As long as S^2 is a good estimate of σ^2 and does not vary much from sample to sample, which is usually the case for $n \geq 30$, the values

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

are still approximately distributed as a standard normal variable, and central limit theorem is valid. If the sample size is small (n<30), the values of S^2 fluctuate considerably from sample to sample and the distribution of the values $(\bar{X} - \mu)/(S/\sqrt{n})$ is no longer a standard normal distribution.

Thus the theory which follows leads to the exact distribution of

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

for r.s.'s from normal populations.

Theorem 1.6

Let Z and U be two r.v.'s with

1- $Z \sim N(0,1)$,

2- $U \sim \chi^2_r$

3- Z and **U** are independent.

Then the distribution of

$$T = \frac{Z}{\sqrt{U/r}}$$

is called the t-distribution with **r** degrees of freedom and its p.d.f. is given by

$$f(x) = \frac{\Gamma\left(\frac{r}{2} + 1\right)}{\sqrt{\pi r}\;\Gamma\left(\frac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}} \qquad -\infty < x < \infty$$

The t-distribution is also known as the student-t distribution. The t-distribution is similar to the N(0,1) distribution in that they both are symmetric about a mean of zero. Both distributions are bell shaped but the t-distribution is more variable. The areas under the curve have been tabulated in sufficient detail to meet the requirements of most problems. The distribution of t is similar to the distribution of Z, in that they both are symmetric about a mean of zero. Both distributions are bell shaped but the t distribution is more variable. The distribution of t differs from that of Z in that the t-distribution depends on the degrees of freedom r and is always greater than 1. Only when $r \to \infty$ (or r large > 30) will the two distributions become the same. In Figure 1, we show the relationship between a standard normal distribution ($r = \infty$), and t distribution with 4 and 8 degrees of freedom

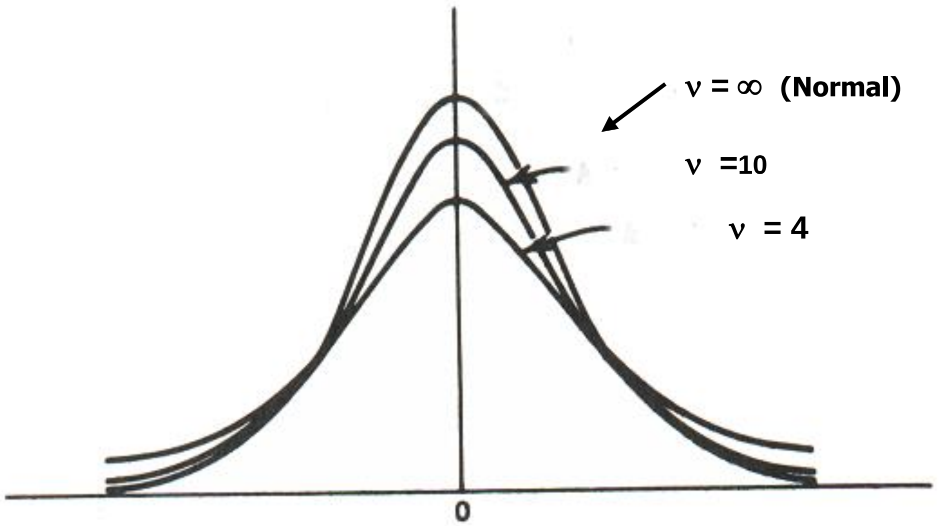


Fig. 1.3 t distribution curves for ν = 4, 10 and ∞.

Theorem 1.7

If $\overline{\mathbf{X}}$ and $\mathbf{S^2}$ are the mean and variance, respectively, of a random sample of size n taken from a population that is normally distributed with mean μ and unknown variance σ^2. Then the variable

$$\mathbf{T} = \frac{\overline{\mathbf{X}} - \mu}{\mathbf{S}/\sqrt{\mathbf{n}}}$$

has a t-distribution with $\boldsymbol{\nu = n-1}$ degrees of freedom.

Proof

If $\mathbf{X_1, X_2, ..., X_n}$ are a random sample from $\mathbf{N(\mu,\sigma^2)}$, then

1- $\overline{\mathbf{X}} \sim \mathbf{N(\mu, \sigma^2/n)}$ i.e. $\mathbf{Z} = \dfrac{\overline{\mathbf{X}} - \mu}{\sigma/\sqrt{\mathbf{n}}} \sim \mathbf{N(0, 1)}$

2- $\mathbf{U} = \dfrac{(\mathbf{n}-1)\mathbf{S}^2}{\sigma^2} \sim \chi^2_{n-1}$

3- $\overline{\mathbf{X}}$ and S^2 are independent, thus also Z and U are independent.

Therefore by theorem 1.7, we have

$$T = \frac{Z}{\sqrt{U/r}} = \frac{\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2} / (n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

For a t-distribution with (n-1) degrees of freedom the symbol t_α denotes the t-value leaving area of α to the right. t_α is the upper α- point of the t-distribution with (n-1) degrees of freedom (see Fig. 1.4). The t-table is arranged to give the values t_α for several frequently used values of α and different values of **ν** = (n-1).

Since the t-distribution is symmetrical about the value t = 0, the lower points can be obtained form the, upper points. The relationship between the lower and upper points is

$$t_{1-\alpha} = -t_\alpha$$

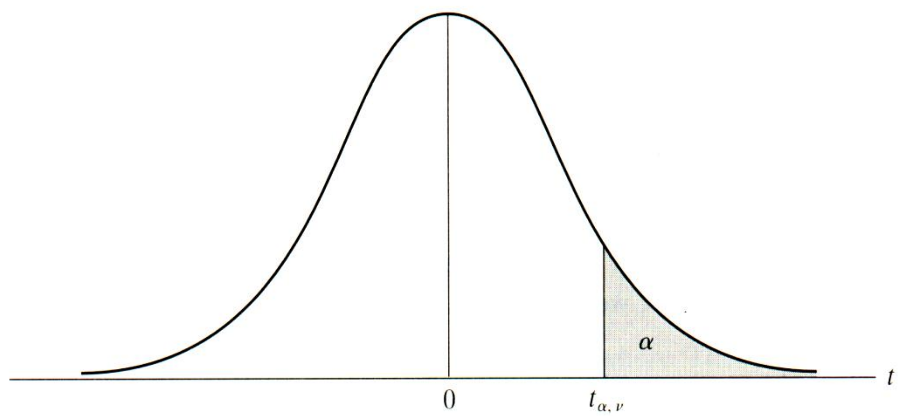


Fig. 1.4 α- point of the t-distribution with (n-1) d.f.

For example; if **ν** = n-1 = 5, then from the t-table we have

$t_{0.025} = 2.57$ therefore $t_{0.975} = -2.57$

Using MINITAB

Suppose we want to find $t_{0.025,5}$, then its CDF is $F(t_{0.025,5}) = 1 - \alpha/2 = 0.975$. Now press ***Calc* ⟶ *Probabilty Distributions* ⟶ *t*** then click on "inverse cumulative distribution" and write 5 for the degrees of freedom and 0.975 for "input constant". Click on "ok" we obtain:

```
Student's t distribution with 5 DF

P( X <= x )        x
   0.975       2.57058
```

i.e. $t_{0.025,5} = 2.57058$.

1.6 The F- Distribution

Another derived distribution of great importance in statistics is called the F distribution.

Theorem 1.8

Let U_1 and U_2 be two r.v.'s with

1- $U_1 \sim \chi^2_{r_1}$,

2- $U_2 \sim \chi^2_{r_2}$

3- U_1 and U_2 are independent.

Then the distribution of the r.v.

$$F = \frac{U_1 / r_1}{U_2 / r_2}$$

is called the F distribution with (r_1 , r_2) degrees of freedom and its p.d.f. is given by

$$f(x) = \frac{\Gamma\left(\frac{r_1 + r_2}{2}\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}\left(\frac{r_1}{r_2}\right)^{r_1/2} x^{\frac{r_1}{2}-1}\left(1 + \frac{r_1}{r_2}x\right)^{-\frac{r_1+r_2}{2}} \qquad -\infty < x < \infty$$

Corollary

Let $X_1, X_2,\ldots,X_n$ and $Y_1, Y_2,\ldots,Y_n$ be independent random samples from populations with respective distributions $X_i \sim N(\mu_1, \sigma_1^2)$ and $Y_j \sim N(\mu_2, \sigma_2^2)$. If $r_1 = n_1$-1 and $r_2 = n_2$-1, then

$$U_1 = \frac{(n_1 - 1) S_1^2}{\sigma_1^2} \sim \chi^2_{n_1 - 1} \qquad \text{and} \qquad U_2 = \frac{(n_2 - 1) S_2^2}{\sigma_2^2} \sim \chi^2_{n_2 - 1}$$

so that

$$F = \frac{U_1 / (n_1 - 1)}{U_2 / (n_2 - 1)} = \frac{S_1^2 \ \sigma_2^2}{S_2^2 \sigma_1^2} \sim F_{n_1 - 1, n_2 - 1}$$

<+><+><+><+><+><+><+><+><+>

EXERCISES

[1] Certain tubes manufactured by a company have a mean lifetime of 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 36 tubes taken from the group will have a mean lifetime.

a- Between 790 and 810 hours, **b-** More than 815 hours.

[2] A and B manufacture two types of cables, having mean breaking strengths of 4000 and 4500 pounds and standard deviations of 300 and 200 pounds respectively. If 100 cables of brand A and 50 cables of brand B are tested, what is the probability that the mean breaking strength of B will be

a- At least 600 pounds more than A, **b-** At least 450 pounds more than A.

[3] Find **a-** $P(-t_{0.005} < t < t_{0.01})$ **b-** Find $P(t > -t_{0.025})$.

[4] Given a random sample of size 24 from a normal distribution, find, K such that

a- $P(-2.069 < t < K) = 0.965$ **b-** $P(K < t < 2.807) = 0.095$.

c- $P(-K < t < K) = 0.90$.

[5] Consider the four independent random variables X, Y, U and V such that $X \sim N(0,16)$, $Y \sim N(5,4)$, $U \sim \chi^2(4)$ and $V \sim \chi^2(16)$.

State the distribution of each of the following variables

a- $\dfrac{X^2}{16} + \dfrac{(Y-5)^2}{4}$ **b-** $\dfrac{X}{\sqrt{V}}$ **c-** $\dfrac{4U}{V}$ **d-** $X+2Y$ **e-** $2X-Y$

[6] If $X_1, X_2, \ldots, X_n$ are i.i.d. $N(0,\sigma^2)$, state the distribution of each of the following variables:

a- $U = 3X_1 - 5X_2 + 8$ **b-** $V = \sum_{i=1}^{n} X_i$

c- $W = \left(\sum_{i=1}^{n} X_i\right)^2 \Big/ n\sigma^2$ **d-** $Y = \dfrac{2X_1^2}{X_2^2 + X_3^2}$ **e-** $Y = \dfrac{\Sigma X_i}{\sqrt{\Sigma X_i^2}}$

[7] If $X_1, X_2, \ldots, X_n$ are i.i.d. $N(0,\sigma^2)$, state the distribution of each of the following variables:

a- $Y = 5X_1 - 7X_2 + 2$ **b-** $Y = \dfrac{2X_1^2}{X_2^2 + X_3^2}$ **c-** $Y = \dfrac{\Sigma X_i}{\sqrt{\Sigma X_i^2}}$

[8] Suppose that X_1, X_2, X_3 and X_4 are i.i.d. $N(0,\sigma^2)$, then the distribution of the random variable $\mathbf{Y} = \dfrac{\mathbf{X_1 + X_2}}{\sqrt{\mathbf{X_3^2 + X_4^2}}}$ **is**

a. χ^2 **(2)** **b.** t(2) **c.** F(2,2) **d.** None of the above.

[9] Consider the three independent random variables X, U and V such that **X ~ N(0,1)**, **U** ~ χ^2**(4)** and **V** ~ χ^2**(16)**. Find the distribution of $\mathbf{W = X^2 + U + V}$.

[10] Let $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ be sample means of two independent random samples of sizes 10 and 20 from N(4,9) and N(5,16) respectively. Find mean, variance and distribution of $\mathbf{Z = \overline{X} - 2\overline{Y} + 3}$.

[11] Show that if **X** has a t distribution with v d.f., then $\mathbf{Y=X^2}$ has an **F** distribution with $\boldsymbol{\nu}_1=1$ and $\boldsymbol{\nu}_2 = \boldsymbol{\nu}$ d.f.

[12] Circle the best answer from each of the following multiple-choice questions:

Let X ~ N(1,16), Y ~ N(0,4) and U ~ χ^2 (15) be three independent r.v's.

a- The distribution of 2X-3Y+5 is

i. N(7,28) ii. N(7,100) iii. N(2,105) iv. None of the above.

b- One of the following r.v.'s has F(16,1)

i. $\dfrac{\mathbf{U + Y^2/4}}{\mathbf{(X-1)^2/16}}$ ii. $\dfrac{\mathbf{U + Y^2/4}}{\mathbf{(X-1)^2}}$ iii. $\dfrac{\mathbf{(U + Y^2/4)/16}}{\mathbf{(X-1)^2}}$ iv. None of the above.

c- The distribution of $\dfrac{\mathbf{X-1}}{\sqrt{\mathbf{Z^2+U}}}$ is

i. t(3) **ii.** t(15) **iii.** t(16) iv. None of the above.

<+><+><+><+><+><+><+><+><+>