Population distribution	σ^2	n	statistic	Sampling Distribution
Any	known	large	$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	standard normal
Normal	known	any	$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	standard normal
Normal	unknown	small ($n<30$)	$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$	T with $(n-1)$ degrees of freedom
Normal	unknown	large $n \geq 30$	$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$	standard normal
Any	known	~~any~~	$\Rightarrow$ cheby	

more empirical data

	μ	σ^2		
Normal	unknown	known	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	χ^2 with $(n-1)$ dof

2 samples

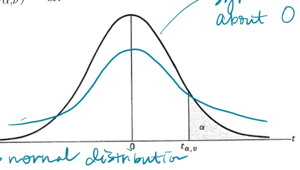
population distribution	σ_1^2, σ_2^2	μ_1, μ_2	statistic	sampling distribution
normal	known	unknown	$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$	F with (n_1-1, n_2-1) df

F & χ^2

- Since the t distribution arises in many important applications, integrals of its probability density function have been extensively tabulated.
- Let $t_{\alpha,\nu}$ be such that the area to its right under the curve of the t distribution with ν degrees of freedom is equal to α. That is, $t_{\alpha,\nu}$ is such that $P(T \geq t_{\alpha,\nu}) = \alpha$.

symmetric about 0

when ν is small.

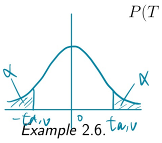


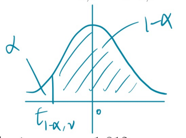
The $t_{\alpha,\nu}$ notation

when $\nu \to \infty$, the distribution $\to$ normal distribution

- Since the probability density function of the t distribution is symmetrical about $t = 0$, thus

$$P(T < -t_{\alpha,\nu}) = \alpha \quad \text{and} \quad -t_{\alpha,\nu} = t_{1-\alpha,\nu}$$





Example 2.6.

$$t_{0.05,10} = 1.812 \quad \text{and} \quad t_{0.95,10} = -1.812.$$

$\alpha \quad n \qquad 1-0.05$

□

Example 2.8. In 16 one-hour test runs, the gasoline consumption of an engine averaged 16.4 liters with a standard deviation of 2.1 liters.

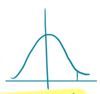
Assume that the distribution of gasoline consumption is approximately normal. Test the claim that the average gasoline consumption of this engine is 12.0 liters per hour.

$\bar{X}\ \mu$ the degrees of freedom $= n-1 = 15$

$\bar{X}\ \sigma$

population is normal $\Rightarrow$ t distrib

$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{16.4-12.0}{2.1/\sqrt{16}} = 8.38$



$t_{\alpha,15}$ unknown

when $\alpha = 0.005$ (smallest no. record in the table),

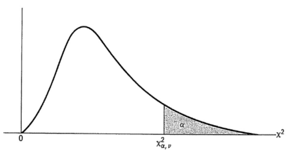
$P(T > t_{0.005} = 2.947) = 0.005$

$\therefore P(T > 8.38) \to 0$

$\therefore$ It won't exceed average 12.0 liters/h

Chi-square Distribution

- Since the chi-square distribution arises in many important applications, integrals of its probability density function have been extensively tabulated.
- Let $\chi^2_{\alpha,\nu}$ be such that the area to its right under the curve of the chi-square distribution with ν degrees of freedom is equal to α. That is, $\chi^2_{\alpha,\nu}$ is such that $P(\boldsymbol{\chi}^2 \geq \chi^2_{\alpha,\nu}) = \alpha$.



The $\chi^2_{\alpha,\nu}$ notation

Applications of the Chi-Square Distribution in Sampling

Theorem 2.12. If S^2 is the variance of a random sample of size n from a normal population with the variance σ^2, then the statistic

$$\boldsymbol{\chi}^2 = \frac{(n-1)S^2}{\sigma^2}$$

has the chi-square distribution with $(n-1)$ degrees of freedom.

2.2.1. Sampling from an Infinite Population

Theorem 2.1. If $\bar{X}$ is the mean of a random sample of size n from an infinite population with the mean μ and the variance σ^2, then

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \quad \text{and} \quad \text{Var}(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

and $\sigma_{\bar{X}}$ (the positive square root of $\sigma^2_{\bar{X}}$) is called the ***standard error of the mean***.

Theorem 2.2 (Chebyshev's Theorem for Sampling Distribution). If $\bar{X}$ is the mean of a random sample of size n from an infinite population with the mean μ and the variance σ^2, then for any positive constant c,

$$P(|\bar{X}-\mu| \geq c) \leq \frac{\sigma^2}{nc^2},$$

or equivalently,

$$P(|\bar{X}-\mu| < c) \geq 1 - \frac{\sigma^2}{nc^2}.$$

μ, σ^2 known

inf population

regardless of the sample size

Theorem 2.3 (The Law of Large Numbers). If $\bar{X}$ is the mean of a random sample of size n from an infinite population with the mean μ and the variance σ^2, then for any positive constant c,

$$P(|\bar{X}-\mu| \geq c) \to 0 \quad \text{as } n \to \infty.$$

Theorem 2.4 (Central Limit Theorem). If $\bar{X}$ is the mean of a random sample of size n from an infinite population with the mean μ and the variance σ^2, then the statistic (called the ***standardized sample mean***)

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

regardless of the distribution of X

is a random variable whose probability density function approaches that of the standard normal distribution as $n \to \infty$.

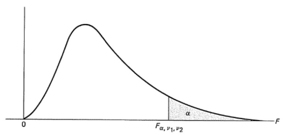
Remark: Although the Central Limit Theorem will work well for small samples in most cases, particularly where the population is continuous, unimodal, and symmetric, larger samples depending on the shape of the population will be required in other situations. In many cases of practical interest, if $n \geq 30$, the normal approximation will be satisfactory regardless of the shape of the population.

$\to$ sample size ≥ 30, can apply Z for $\bar{X}_n$

F-distribution

- The parameters ν_1 and ν_2 are referred as the ***degrees of freedom for numerator and denominator***, respectively.
- In view of its importance, the F distribution has been tabulated extensively.
- Let F_{α,ν_1,ν_2} be such that the area to its right under the curve of the F distribution with ν_1 and ν_2 degrees of freedom is equal to α. That is F_{α,ν_1,ν_2} is such that $P(\mathbf{F} \geq F_{\alpha,\nu_1,\nu_2}) = \alpha$.

The F_{α,ν_1,ν_2} notation



Example 2.12. Find the value of $F_{0.95}$ for $\nu_1 = 10$ and $\nu_2 = 20$ degrees of freedom.

- $F_{1-\alpha,\nu_2,\nu_1} = 1/F_{\alpha,\nu_1,\nu_2}$ order

Solution:

$$F_{0.95,10,20} = \frac{1}{F_{0.05,20,10}} \approx \frac{1}{2.77} \approx 0.361$$

□

Applications of the F Distribution in Sampling

Theorem 2.13. If S_1^2 and S_2^2 are the variances of independent random samples of size n_1 and n_2, respectively, taken from two normal populations with the variances σ_1^2 and σ_2^2, respectively, then the statistic

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

has the F distribution with (n_1-1) and (n_2-1) degrees of freedom.

Exercise 2.11. An optical firm purchases glass to be ground into lenses, and it is known from past experience that the variance of the refractive index of this kind of glass is 1.26×10^{-4}. σ^2

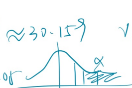
As it is important that the various pieces of glass have nearly the same index of refraction, the firm rejects such a shipment if the sample variance of 20 pieces selected at random exceeds 2.00×10^{-4}. S^2

Assuming that the sample values may be looked upon as a random sample from a normal population, what is the probability that a shipment will be rejected even though $\sigma^2 = 1.26 \times 10^{-4}$? Type I error

$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$

$= \frac{(20-1)(2.00\times10^{-4})}{1.26\times10^{-4}} \approx 30.159 \qquad \nu = 19$

$0.025 < P(\chi^2 > 30.159) < 0.05$



The required probability is between 0.025 and 0.05

2.2.2. Sampling from a Finite Population (w/o replacement)

Theorem 2.5. If $\bar{X}$ is the mean of a random sample of size n from a finite population of size N with the mean μ and the variance σ^2, then

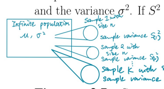
$$E(\bar{X}) = \mu_{\bar{X}} = \mu \quad \text{and} \quad \text{Var}(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}.$$

- It is of interest to note that the formulas we obtained for $\text{Var}(\bar{X})$ of the random sample of size n from an infinite population and that from a finite population of size N differ only by the ***finite population correction factor***
$$\frac{N-n}{N-1}.$$
  - Indeed, when N is large compared to n, then $N >>> n$
$$\frac{N-n}{N-1} \to 1$$
  and thus
$$\text{Var}(\bar{X}) \to \frac{\sigma^2}{n}.$$
  A general rule of thumb is to use this approximation when $n \leq N/20$.
- In practice, we often deal with random samples from populations that are finite, but large enough to be treated as if they were infinite. Thus, most statistical theory and most of the methods we shall discuss apply to samples from infinite populations.

2.3. The Sample Variance

- Suppose that $X_1, X_2, \ldots, X_n$ constitute a random sample of size n. Let $\bar{X}$ be the sample mean. The statistic $s^2 = \frac{n}{n-1}$ (cal theoretical variance)
$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2 \qquad S = \sqrt{S^2}$$
  is called the ***sample variance*** (or the variance of the random sample).
- The statistic S (the positive square root of S^2) is called the ***sample standard deviation*** (or the standard deviation of the random sample).

Theorem 2.6. Suppose that $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from an infinite population which has the mean μ and the variance σ^2. If S^2 is the sample variance, then long run theoretical value



$$E(S^2) = \sigma^2. \qquad E(S^2) = \lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} S_{(i)}^2$$

Theorem 2.7. Suppose that $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from an infinite population which has the mean μ and the variance σ^2. If S^2 is the sample variance, then

$$S^2 = \frac{1}{n-1}\left(\sum_{j=1}^{n} X_j^2 - n\bar{X}^2\right).$$

3.2. Point Estimations of the Population Mean

3.2.1. Estimators

functions of a random sample

- Basically, ***point estimation*** concerns the choosing of a statistic, that is, a single number calculated from sample data for which we have some expectation, or assurance, that it is reasonably close to the parameter it is supposed to estimate.
- The statistic, whose value is used as the point estimate of a parameter, is called an ***estimator***. can be λ, μ, p ... etc
- A statistic $\hat{\theta}$ is said to be an ***unbiased estimator*** of the parameter θ if and only if $E(\hat{\theta}) = \theta$. eg. $E(\bar{X}) = \mu$
- If an estimator is not unbiased, it is said to be ***biased***.
- A statistic $\hat{\theta}_1$ is said to be a more efficient unbiased estimator of the parameter θ than the statistic $\hat{\theta}_2$ if

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta \quad \text{and} \quad \mathrm{Var}(\hat{\theta}_1) \le \mathrm{Var}(\hat{\theta}_2).$$

independent $\Rightarrow \mathrm{cov}(X_1, X_2) = 0$

Example 3.2. In the previous example, the sample mean is a more efficient unbiased estimator of μ than the weighted sample mean.
□ Let $g(k) = \mathrm{Var}\left(\frac{aX_1 + bX_2}{a+b}\right) = \mathrm{Var}[kX_1 + (1-k)X_2]$

$\therefore \mathrm{Var}[kX_1 + (1-k)X_2] \ge \mathrm{Var}[\frac{1}{2}X_1 + (1-\frac{1}{2})X_2]$ $= k^2 \mathrm{Var}(X_1) + (1-k)^2 \mathrm{Var}(X_2)$
$= k^2\sigma^2 + (1-k)\sigma^2 = (2k^2 - 2k + 1)\sigma^2$
$= \mathrm{Var}\left(\frac{X_1 + X_2}{2}\right)$ If k^* is the minimizer of $g(k)$, then
in general $g'(k^*) = (4k^* - 2)\sigma^2 = 0 \Rightarrow k^* = \frac{1}{2}$
$\therefore \frac{X_1 + X_2}{2}$ is more efficient in general. since $g''(k^*) = 4\sigma^2 > 0$
$\therefore k^* = \frac{1}{2}$ is a minimizer of $g(k)$

3.2.2. Maximum Likelihood Estimators

- The definitions of unbiasness and other properties of estimators do not provide any guidance about how good estimators can be obtained.
- In this section, we discuss the method of maximum likelihood for deriving point estimators.
- Recall that the probability distribution/density function $f(x)$ for a random variable X usually has at least one parameter θ associated with it. Assume that we have a random sample $x_1, x_2, \ldots, x_n$ available. The method of maximum likelihood in a sense picks out of all possible values of θ the one most likely to have produced these observations.

P.S. $\mu = np$
$\sigma = \sqrt{np(1-p)}$

eg.

Assume that X has a binomial distribution with $n = 4$ with that p is unknown. Thus observed data

$$f(x; p) = \binom{4}{x} p^x (1-p)^{4-x}, \quad x = 0, 1, 2, 3, 4.$$

Suppose that we observe that $X = 3$ and need to estimate p. Consider

$$f(3; p) = 4p^3(1-p).$$

We now maximize this probability at $X = 3$.

$$\frac{d\, f(3;p)}{dp} = 4(1-p)3p^2 + 4p^3(-1) = 4p^2(3 - 4p)$$

If $\hat{p}$ is the maximizer, then

$$4\hat{p}^2(3 - 4\hat{p}) = 0 \quad \Longrightarrow \quad \hat{p} = 0.75.$$

Since

$$\frac{d^2 f(3;\hat{p})}{dp^2} = 24\hat{p}(1 - 2\hat{p}) = -9 < 0$$

thus $\hat{p} = 0.75$ is the one most likely to have produced the observation $X = 3$.

- We now assume that X is a discrete random variable with probability distribution function $f(x;\theta)$, where θ is a single unknown parameter, and that X has an infinite population. Consider

$$\begin{aligned} &P(X_1 = x_1,\ X_2 = x_2, \ldots, X_n = x_n) \\ &= P(X_1 = x_1) \cdot P(X_2 = x_2) \cdots P(X_n = x_n) \\ &= \prod_{j=1}^{n} f(x_j; \theta). \end{aligned}$$

product; independent; same distribution $\Rightarrow$ same pdf func

This probability is a function of θ because the observed values are fixed numbers. Therefore, we write

$$L(\theta) = \prod_{j=1}^{n} f(x_j; \theta) \longrightarrow \text{maximize it}$$

and call it the ***likelihood function***.

Example 3.4. Let $x_1, x_2, \ldots, x_n$ be the observed values of a random sample of size n from a normal population with unknown mean μ and known variance σ^2, find the maximum likelihood estimator of μ.

known; the only parameter

Solution: Note that the probability density function of X is

$$f(x; \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

Therefore the likelihood function of the random sample is

$$L(\mu) = \prod_{j=1}^{n} f(x_j; \mu) = \prod_{j=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-(x_j-\mu)^2/(2\sigma^2)}\right] = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[\sum_{j=1}^{n} \frac{(x_j-\mu)^2}{-2\sigma^2}\right]$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[\sum_{j=1}^{n} \frac{(x_j-\mu)^2}{-2\sigma^2}\right]$$

and the log likelihood function is minimize

$$\ln\left[L(\mu)\right] = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2. \quad = -\frac{n}{2}\ln(2\pi\sigma^2) + \sum_{j=1}^{n}\frac{(x_j-\mu)^2}{-2\sigma^2}$$

Now

$$\frac{d \ln\left[L(\mu)\right]}{d\mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu) = \frac{1}{\sigma^2}\left(\sum_{j=1}^{n} x_j - n\mu\right).$$

$\frac{d\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right)}{d\mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j-\mu) = \frac{1}{\sigma^2}\left(\sum_{j=1}^{n} x_j - n\mu\right)$

If $\hat{\mu}$ is the maximum likelihood estimator of μ, then

$$\frac{d \ln\left[L(\hat{\mu})\right]}{d\mu} = \frac{1}{\sigma^2}\left(\sum_{j=1}^{n} x_j - n\hat{\mu}\right) = 0.$$

check $\frac{d^2y}{dx^2}$ might be a minimizer

This implies

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} x_j = \bar{x}.$$

remember to check whether concave

Since

$$\frac{d^2 \ln\left[L(\hat{\mu})\right]}{d\mu^2} = \frac{d^2 \ln\left[L(\mu)\right]}{d\mu^2} = -\frac{n}{\sigma^2} < 0,$$

therefore

$$\hat{\mu} = \bar{x}$$

is the maximum likelihood estimator of μ.
□

Confidence interval

- Because different samples will produce different values of l and u, these end-points are values of the random variables L and U, respectively.
- Let $\hat{\Theta}$ be a point estimator for θ. Based on the sampling distribution of $\hat{\Theta}$, we can choose l and u such that the following probability statement is true:

$$P(L < \theta < U) = 1 - \alpha$$

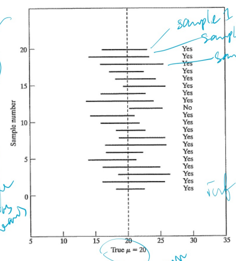
where $0 < \alpha < 1$. $\hookrightarrow \theta$ lies in (L, U)

- Such an interval $l < \theta < u$, computed for a particular sample, is called a $100(1-\alpha)\%$ ***confidence interval***, the fraction $(1-\alpha)$ is called the ***confidence coefficient*** or the ***degree of confidence***, and the end-points l and u are called the lower and upper ***confidence limits***.
- Note that the true value of θ is fixed (although it is unknown) but the end-points l and u are only computed from a particular sample data, the confidence interval $l < \theta < u$ is either correct (true with probability 1) or wrong (false with probability 1).
- The correct interpretation lies in the realization that a confidence interval is a ***random interval*** because in the probability statement defining the end-points of the interval, L and U, are random variables.
- Consequently, the correct interpretation of a $100(1-\alpha)\%$ confidence interval depends on the relative frequency view of probability. Specifically, if an infinite number of random samples are collected and a $100(1-\alpha)\%$ confidence interval for θ is computed from each sample, $100(1-\alpha)\%$ of these intervals will contain the true value of θ.

95% of them contains the true mean

If we repeat many samples, then μ is 95% within the interval (95% of the samples contains the true mean)

Interpretation of the confidence interval for population mean with true mean $\mu = 20$



3.3.1. Confidence Interval for the Mean of a Normal Population with Known Variance

Development of a Confidence Interval

- Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal population with unknown mean μ and known variance σ^2.
- If $\bar{X}$ is the sample mean, then the standardized sample mean

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

has the standard normal distribution, no matter how small the size of the sample.

α, $1-\alpha$, $\frac{\alpha}{2}$, $-Z_\alpha$, $Z_{\frac{\alpha}{2}}$

- Thus, we can write

$$\begin{aligned} 1 - \alpha &= P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\ &= P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\ &= P\left(\bar{X} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

$\le \quad \le$

where $z_{\alpha/2}$ is such that $P(Z > z_{\alpha/2}) = \alpha/2$.

- If $\bar{x}$ is the observed value of the mean of a random sample of size n from a normal population with unknown mean μ and known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is

$$\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}. \Longleftrightarrow \mu = \bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$ [ ] only can use when it is normal

Lengths of Confidence Intervals

length

- If $\bar{x}$ is the value of the mean of a random sample of size n from a normal population with unknown mean μ and known variance σ^2, the length of a $100(1-\alpha)\%$ confidence interval for μ is

$$\left(\bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) = 2\cdot z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}.$$

- Thus, the length of the 99% confidence interval for μ is

$$2\cdot z_{0.005}\cdot\frac{\sigma}{\sqrt{n}} = 2(2.575)\frac{\sigma}{\sqrt{n}} = 5.15\frac{\sigma}{\sqrt{n}} \quad \text{longer}$$

whereas the length of the 95% confidence interval for μ is

$$2\cdot z_{0.025}\cdot\frac{\sigma}{\sqrt{n}} = 2(1.96)\frac{\sigma}{\sqrt{n}} = 3.92\frac{\sigma}{\sqrt{n}}.$$

- Therefore, the 99% confidence interval is longer than the 95% confidence interval. This is why we have a higher degree of confidence in the 99% confidence interval.
- Generally, for a fixed sample size n and standard deviation σ, the higher the degree of confidence, the longer the resulting confidence interval.

Approximate One-Sided Confidence Bounds

- An approximate $100(1-\alpha)\%$ ***lower-confidence bound*** for μ is

$$\bar{x} - z_\alpha\cdot\frac{\sigma}{\sqrt{n}} < \mu$$

and an approximate $100(1-\alpha)\%$ ***upper-confidence bound*** for μ is

$$\mu < \bar{x} + z_\alpha\cdot\frac{\sigma}{\sqrt{n}}.$$

One-Sided Confidence Bounds

- Note that confidence intervals for given parameters are not unique.
- The $100(1-\alpha)\%$ confidence interval for μ: standard normal

$$\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}.$$

is indeed a ***two-sided*** $100(1-\alpha)\%$ confidence interval for μ.

- It is also possible to obtain ***one-sided*** $100(1-\alpha)\%$ confidence intervals/bounds for μ by setting either $l = -\infty$ or $u = +\infty$ and replacing $z_{\alpha/2}$ by z_α.

- The $100(1-\alpha)\%$ ***lower-confidence bound*** for μ is

$$\bar{x} - z_\alpha\cdot\frac{\sigma}{\sqrt{n}} < \mu$$

Z_α

and the $100(1-\alpha)\%$ ***upper-confidence bound*** for μ is

$$\mu < \bar{x} + z_\alpha\cdot\frac{\sigma}{\sqrt{n}}.$$

Z_α

3.3.2. Large-Sample Confidence Interval for the Mean of a Non-Normal Population with Known Variance $\Longrightarrow$ need the sample size ≥ 30

Development of an Approximate Confidence Interval

- Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of size n from a non-normal population with unknown mean μ and known variance σ^2 and that $\bar{X}$ is the sample mean.
- If n is sufficiently large (when $n \ge 30$ in practice), by the central limit theorem, the standardized sample mean

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

has an approximate standard normal distribution.

different from before

- Thus, we can write

$$\begin{aligned} 1 - \alpha &\approx P(-z_{\alpha/2} < Z < z_{\alpha/2}) \\ &= P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \\ &= P\left(\bar{X} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

where $z_{\alpha/2}$ is such that $P(Z > z_{\alpha/2}) = \alpha/2$.

- Let $\bar{x}$ be the value of the mean of a random sample of size n from a non-normal population with unknown mean μ and known variance σ^2. If n is large, then an approximate $100(1-\alpha)\%$ confidence interval for μ is

$$\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}.$$

Choice of Sample Size

- It is desirable to obtain a confidence interval that is short enough for decision-making purposes and that also has adequate degree of confidence. One way to achieve this is by choosing the sample size n to be large enough to give a confidence interval of specified length or precision with prescribed confidence.
- Since we are approximately $100(1-\alpha)\%$ confident that

$$\bar{x} - z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}},$$

this means that in using $\bar{x}$ to estimate μ, the maximum error is approximately

$$\max|\bar{x} - \mu| = z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$$

with confidence $100(1-\alpha)\%$.

- In situation where the sample size can be controlled, we can choose n so that we are approximately $100(1-\alpha)\%$ confident that the maximum error in estimating μ is less than or equal to a specified bound E, that is

$$z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \le E.$$

max error

This implies

$$n \ge \left(z_{\alpha/2}\cdot\frac{\sigma}{E}\right)^2. \longrightarrow \text{normal population}$$

- Thus, for a non-normal population

$$n \ge \left(z_{\alpha/2}\cdot\frac{\sigma}{E}\right)^2 \quad \text{and} \quad n \ge 30.$$

Exercise 3.7. A research worker wants to determine the average time it takes a mechanic to rotate the tires of a car, and she wants to be able to assert with 95% confidence that the mean of her sample is off by at most 0.50 minutes. $\to E$

If she can presume from past experience that $\sigma = 1.6$ minutes, how large a sample will she have to take?

$\alpha = 0.05$ 1.96 Any distribution

$$n \ge \left(z_{0.025}\cdot\frac{1.6}{0.5}\right)^2 = 39.3 \ge 30$$

$\therefore$ the sample of size $n = 40$ is sufficient

3.3.3. Confidence Interval for the Mean of a Normal Population with Unknown Variance

Development of a Confidence Interval (small sample size)

- Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal population with unknown mean μ and unknown variance.
- If $\bar{X}$ and S are the sample mean and the sample standard deviation, then the statistic

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

has the t distribution with $(n-1)$ degrees of freedom.

t distribution $v = n-1$

- Thus, we can write

$$1-\alpha = P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1})$$
$$= P\left(-t_{\alpha/2,n-1} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2,n-1}\right)$$
$$= P\left(\bar{X} - t_{\alpha/2,n-1}\cdot\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\cdot\frac{S}{\sqrt{n}}\right)$$

$-t_{\frac{\alpha}{2},n-1}$ $t_{\frac{\alpha}{2},n-1}$

where $t_{\alpha/2,n-1}$ is such that $P(T > t_{\alpha/2,n-1}) = \alpha/2$.

- If $\bar{x}$ and s are the values of the mean and the standard deviation, respectively, of a random sample of size n from a normal population with unknown mean μ and unknown variance, a $100(1-\alpha)\%$ confidence interval for μ is

$$\bar{x} - t_{\alpha/2,n-1}\cdot\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2,n-1}\cdot\frac{s}{\sqrt{n}}.$$

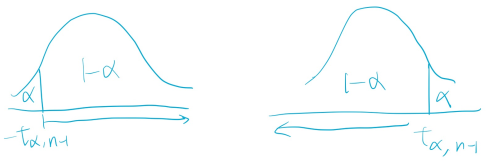
$\mu = \bar{x} \pm t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}$

One-Sided Confidence Bounds

- The $100(1-\alpha)\%$ lower- and upper-confidence bounds for μ are

$$\bar{x} - t_{\alpha,n-1}\cdot\frac{s}{\sqrt{n}} < \mu \quad\text{and}\quad \mu < \bar{x} + t_{\alpha,n-1}\cdot\frac{s}{\sqrt{n}},$$

respectively.



3.4.1. Null Hypothesis versus Alternative Hypothesis

- In statistical decision problems, usually we have to decide between two rival hypotheses.
- The first hypothesis is called the ***null hypothesis*** (denoted by H_0).
- When testing a hypothesis concerning the value of some parameter θ, the statement of equality will ***always*** be included in H_0, that is

$$H_0: \ \theta = \theta_0.$$

any parameter — doesn't matter

Here, θ_0 is called the ***null value***. In this way H_0 pinpoints a specific numerical value that could be the actual value of θ.

- The second hypothesis is called the ***alternative hypothesis*** or the ***motivated hypothesis*** (denoted by H_1). It usually takes one of the three forms: $H_1: \theta \neq \theta_0$, $H_1: \theta < \theta_0$ or $H_1: \theta > \theta_0$.
- It is hoped that the evidence leads us to reject the null hypothesis H_0 and thereby to accept the alternative hypothesis H_1.
- On the other hand, failing to reject H_0 implies that we have not found sufficient evidence to reject H_0. It does not mean that there is a high probability that H_0 is true. It may simply mean that more data are required to reach a strong conclusion. Therefore, rather than saying we "accept H_0," we use the terminology "fail to reject H_0."

failing to reject Ho ~~⇒~~ accept Ho
→ may need more data to reject

- The test procedure, therefore, partitions the possible values of the test statistic into two subsets: a ***critical region*** (also called a ***rejection region***) for H_0 and an ***acceptance region*** for H_0.
- We should reject H_0 and thereby accept H_1 if $\hat{\theta}$, the observed value of the test statistic, falls in the critical region; and we should fail to reject H_0 if $\hat{\theta}$ falls in the acceptance region.

Decision	Actual situation	
	H_0 is true	H_1 is true
Reject H_0	Committed ***type I error***	Made correct decision
Fail to reject H_0	Made correct decision	Committed ***type II error***

Type I Error:

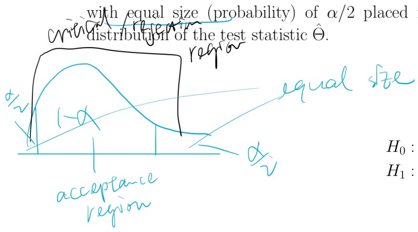
- Rejection of the null hypothesis when it is true is called a type I error.
- The probability of committing a type I error is called the ***significance level*** and is denoted by α.
- The size of a critical region is just the probability α of committing a type I error.

Type II Error:

- Acceptance of the null hypothesis when it is false is called a type II error.
- The probability of committing a type II error is denoted by β.

- Since it appears reasonable to accept H_0 when the point estimate $\hat{\theta}$ of θ is close to θ_0 and to reject it when $\hat{\theta}$ is much larger or much smaller than θ_0, it would be logical to let the critical region consist of both tails of the sampling distribution of the test statistic $\hat{\Theta}$.
- Therefore, the critical region in such a test is split into two parts, with equal size (probability) of $\alpha/2$ placed in each tail of the distribution of the test statistic $\hat{\Theta}$.

critical region



$$H_0: \theta = \theta_0$$
$$H_1: \theta \neq \theta_0$$

α $1-\alpha$ acceptance critical

One-Sided Hypotheses

- A test of hypothesis of one of the two forms:

$\theta \geq \theta_0$ $\qquad$ $\theta \leq \theta_0$

$$H_0: \ \theta = \theta_0 \qquad\qquad H_0: \ \theta = \theta_0$$
$$H_1: \ \theta < \theta_0 \quad\text{or}\quad H_1: \ \theta > \theta_0$$

is called a ***one-sided test*** or a ***one-tailed test***.

- If the alternative hypothesis is $H_1: \theta < \theta_0$, it would seem reasonable to reject H_0 only when $\hat{\theta}$ is much smaller than θ_0. Therefore, in this case it would be logical to let the critical region consist only of the ***lower tail*** of the sampling distribution of $\hat{\Theta}$.
- Whereas if the alternative hypothesis is $H_1: \theta > \theta_0$, we reject H_0 only for large values of $\hat{\theta}$ and the critical region consists only of the ***upper tail*** of the sampling distribution of $\hat{\Theta}$.

acceptance region $1-\alpha$ α critical region

3.5. Tests Concerning Means

3.5.1. Tests on the Mean of a Normal Population with Known Variance

Two-Tailed Tests

- Suppose that we wish to test the hypotheses

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

where μ_0 is a specified constant.

- Suppose further that we have a random sample $X_1, X_2, \ldots, X_n$ from a normal population with known variance σ^2, and that $\bar{X}$ is the sample mean.

standard normal

- If H_0 is true, then $\mu = \mu_0$, and the standardized sample mean

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

has the standard normal distribution. Thus

$\alpha/2$ $1-\alpha$ $\frac{\alpha}{2}$ $-z_{\alpha/2}$ $z_{\alpha/2}$

$$P(Z < -z_{\alpha/2} \ \text{ or } \ Z > z_{\alpha/2}) = \alpha \quad\text{and}$$
$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha,$$

where $z_{\alpha/2}$ is such that $P(Z > z_{\alpha/2}) = \alpha/2$.

- Let α be the type I error probability, and $\bar{x}$ be the value of the mean of a random sample of size n from a normal population with known variance σ^2. We should reject the null hypothesis

$$H_0: \mu = \mu_0$$

and thereby accept the alternative hypothesis

$$H_1: \mu \neq \mu_0$$

if the observed value of the test statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

falls in the critical regions

$$Z < -z_{\alpha/2} \quad\text{or}\quad Z > z_{\alpha/2}$$

and we should fail to reject H_0 if z falls in the acceptance region

$$-z_{\alpha/2} < Z < z_{\alpha/2}.$$

reject region — reject Ho — reject $\alpha/2$ — $\alpha/2$ — $1-\alpha$ — $-z_{\alpha/2}$ — $z_{\alpha/2}$ — do not reject — acceptance region — critical value

Example 3.9. Suppose that it is known from experience that the standard deviation of the weight of 100-gram packages of cookies made by a certain bakery is 2.5 grams.

To check whether its production is under control on a given day, namely, to check whether the true average weight of the packages is 100 grams, they select a random sample of 25 packages and find that their mean weight is $\bar{x} = 98.8$ grams.

Since the bakery stands to lose money when $\mu > 100$ and the customer loses out when $\mu < 100$, test the null hypothesis $\mu = 100$ against the alternative $\mu \neq 100$ using $\alpha = 0.05$.

Assume that the population sampled is normal.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{98.8-100}{2.5/5} = -2.4$$

0.025 0.025 $z_{\frac{\alpha}{2}}$ $z_{\frac{\alpha}{2}}$

Reject Ho if $|z| > z_{\alpha/2} = z_{0.025} = 1.96 < |2.4|$
∴ reject Ho, accept H₁

One-Tailed Tests

- If the alternative hypothesis is

$$H_1: \mu < \mu_0,$$

the critical region of size α should lie in the lower tail of the distribution of the test statistic, that is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha.$$

α $1-\alpha$ $-z_\alpha$ reject — don't reject

- Whereas if the alternative hypothesis is

$$H_1: \mu > \mu_0,$$

the critical region of size α should lie in the upper tail of the distribution of the test statistic, that is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha.$$

$1-\alpha$ α reject Ho — do not reject — z_α

P-Values in Hypothesis Tests

- One way to report the results of a hypothesis test is to state that the null hypothesis was or was not rejected at a specified significance level.
- For example, in the first cookie-weight problem on page 54, we can say that H_0 was rejected at the 0.05 significance level.
- This approach may be unsatisfactory because some decision makers might be uncomfortable with the risks implied by $\alpha = 0.05$ (see the second cookie-weight problem on page 56.)
- To avoid this difficulty the P-value approach has been adopted widely in practice.

upper-tailed
If $z < 0$, ⇒ P is undefined
If $z > 0$, $P = 1 - \Phi(z)$
lower-tailed
$P = \Phi(z)$
If $z > 0$, ⇒ P is undefined

- The P-value is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data. In other words, H_0 would be rejected at any significance level $\alpha > P$-value but it would not be rejected if $\alpha < P$-value.
- Suppose that Z is the test statistic which has the standard normal distribution. If z is the computed value of Z, the P-value is

$$P = \begin{cases} 2\,\Phi(-|z|), & \text{for a two-tailed test with } H_1: \mu \neq \mu_0, \\ \Phi(z), & \text{for a lower-tailed test with } H_1: \mu < \mu_0, \\ 1 - \Phi(z), & \text{for a upper-tailed test with } H_1: \mu > \mu_0, \end{cases}$$

where $\Phi(z) = P(Z < z)$ is the cumulative density function of the standard normal distribution.

Two-tailed
If $z < 0$ $\frac{P}{2} = \Phi(z)$
$P = 2\Phi(z)$
If $z > 0$
$P/2 = \Phi(-z)$
$P = 2\Phi(-z)$

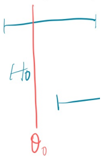
Connection Between Hypothesis Tests and Confidence Intervals

In general, the test at $\alpha\%$ significance level of the hypothesis

$$H_0: \theta = \theta_0$$
$$H_1: \theta \neq \theta_0$$

will lead to rejection of H_0 if and only if θ_0 is not in the $100(1-\alpha)$' confidence interval for θ.

① sample data → confidence interval
if θ_0 within the ↗ ⇒ do not reject H_0



② if out of the confidence interval
⇒ reject

3.5.2. Large-Sample Tests on the Mean of a Non-Normal Population with Known Variance

⇒ sample size should large enough

Two-Tailed Tests

- Suppose that we wish to test the hypotheses

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

where μ_0 is a specified constant.

- If H_0 is true, then $\mu = \mu_0$. If, in addition, n is sufficiently large, that is, when $n \geq 30$, by the central limit theorem, the standardized sample mean

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

has an approximate standard normal distribution. Thus

$$P(Z < -z_{\alpha/2} \text{ or } Z > z_{\alpha/2}) \approx \alpha \qquad \text{and}$$
$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) \approx 1 - \alpha,$$

where $z_{\alpha/2}$ is such that $P(Z > z_{\alpha/2}) = \alpha/2$.

- Let α be the type I error probability, and $\bar{x}$ be the value of the mean of a random sample of size n from a non-normal population with known variance σ^2. If n is large, we should reject the null hypothesis

$$H_0: \mu = \mu_0$$

and thereby accept the alternative hypothesis

$$H_1: \mu \neq \mu_0$$

if the observed value of the test statistic

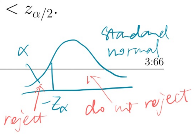
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

falls in the critical regions

$$Z < -z_{\alpha/2} \quad \text{or} \quad Z > z_{\alpha/2}$$

and we should fail to reject H_0 if z falls in the acceptance region

$$-z_{\alpha/2} < Z < z_{\alpha/2}.$$



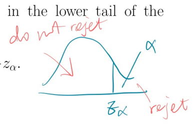
One-Tailed Tests

- If the alternative hypothesis is

$$H_1: \mu < \mu_0,$$

the critical region of size α should lie in the lower tail of the distribution of the test statistic, that is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha.$$



- Whereas if the alternative hypothesis is

$$H_1: \mu > \mu_0,$$

the critical region of size α should lie in the upper tail of the distribution of the test statistic, that is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha.$$

3.5.3. Tests on the Mean of a Normal Population with Unknown Variance

- If H_0 is true, then $\mu = \mu_0$, and the statistic

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

has the t distribution with $(n-1)$ degrees of freedom. Thus

$$P(T < -t_{\alpha/2,n-1} \text{ or } T > t_{\alpha/2,n-1}) = \alpha \qquad \text{and}$$
$$P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) = 1 - \alpha,$$

where $t_{\alpha/2,n-1}$ is such that $P(T > t_{\alpha/2,n-1}) = \alpha/2$.

- Let α be the type I error probability, and $\bar{x}$ and s be the values of the mean and the standard deviation, respectively, of a random sample of size n from a normal population. We should reject the null hypothesis

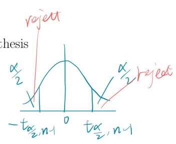
$$H_0: \mu = \mu_0$$

and thereby accept the alternative hypothesis

$$H_1: \mu \neq \mu_0$$

if the observed value of the test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$



falls in the critical regions

$$T < -t_{\alpha/2,n-1} \quad \text{or} \quad T > t_{\alpha/2,n-1}$$

and we should fail to reject H_0 if t falls in the acceptance region

$$-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}.$$

3.5.4. Large-Sample Tests on the Mean of a Normal Population with Unknown Variance

- If H_0 is true, then $\mu = \mu_0$, and the statistic

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

has the t distribution with $(n-1)$ degrees of freedom.

- If n is large, that is, when $n \geq 30$, then the t distribution can be approximated by the standard normal distribution. Thus

$$P(T < -t_{\alpha/2,n-1} \approx -z_{\alpha/2} \text{ or } T > t_{\alpha/2,n-1} \approx z_{\alpha/2}) \approx \alpha \qquad \text{and}$$
$$P(-z_{\alpha/2} \approx -t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1} \approx z_{\alpha/2}) \approx 1 - \alpha,$$

where $z_{\alpha/2}$ is such that $P(Z > z_{\alpha/2}) = \alpha/2$ and $t_{\alpha/2,n-1}$ such that $P(T > t_{\alpha/2,n-1}) = \alpha/2$.

- Let α be the type I error probability, and $\bar{x}$ and s be the values of the mean and the standard deviation, respectively, of a random sample of size n from a normal population. If n is large, we should reject the null hypothesis

$$H_0: \mu = \mu_0$$

and thereby accept the alternative hypothesis

$$H_1: \mu \neq \mu_0$$

if the observed value of the test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

falls in the critical regions

$$T < -t_{\alpha/2,n-1} \approx -z_{\alpha/2} \quad \text{or} \quad T > t_{\alpha/2,n-1} \approx z_{\alpha/2}$$

and we should fail to reject H_0 if z falls in the acceptance region

$$-z_{\alpha/2} \approx -t_{\alpha/2} < T < t_{\alpha/2,n-1} \approx z_{\alpha/2}.$$

One-Tailed Tests

- If the alternative hypothesis is

$$H_1: \mu < \mu_0,$$

the critical region of size α should lie in the lower tail of the distribution of the test statistic, that is

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -t_{\alpha,n-1} \approx -z_\alpha.$$

- Whereas if the alternative hypothesis is

$$H_1: \mu > \mu_0,$$

the critical region of size α should lie in the upper tail of the distribution of the test statistic, that is

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{\alpha,n-1} \approx z_\alpha.$$

$H_0: \mu = 28000$
$H_1: \mu < 28000$

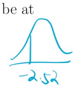
Exercise 3.15. A trucking firm is suspicious of the claim that the average lifetime of certain tires is at least 28000 miles.

To check the claim, the firm puts 40 of these tires on its trucks and gets a mean lifetime of 27463 miles with a standard deviation of 1348 miles.

(a) What can it conclude if the probability of a type I error is to be at most 0.01?

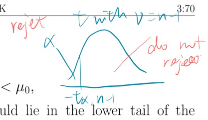
(b) What is the P-value of the test.

b) P-value $= \Phi(-2.52)$
$= 0.0059$



a) $\alpha = 0.01$
$t_{0.01,39} = -2.3267$
Reject H_0 if $t < -z_{0.01} = -2.3267$ where
$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{27463 - 28000}{1348/\sqrt{40}} = -2.52 < -2.3267$
$\therefore H_0$ must be rejected at $\alpha =$ significant level

One-Tailed Tests

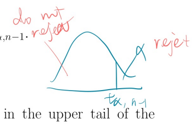


- If the alternative hypothesis is

$$H_1: \mu < \mu_0,$$

the critical region of size α should lie in the lower tail of the distribution of the test statistic, that is

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -t_{\alpha,n-1}.$$



- Whereas if the alternative hypothesis is

$$H_1: \mu > \mu_0,$$

the critical region of size α should lie in the upper tail of the distribution of the test statistic, that is

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{\alpha,n-1}.$$