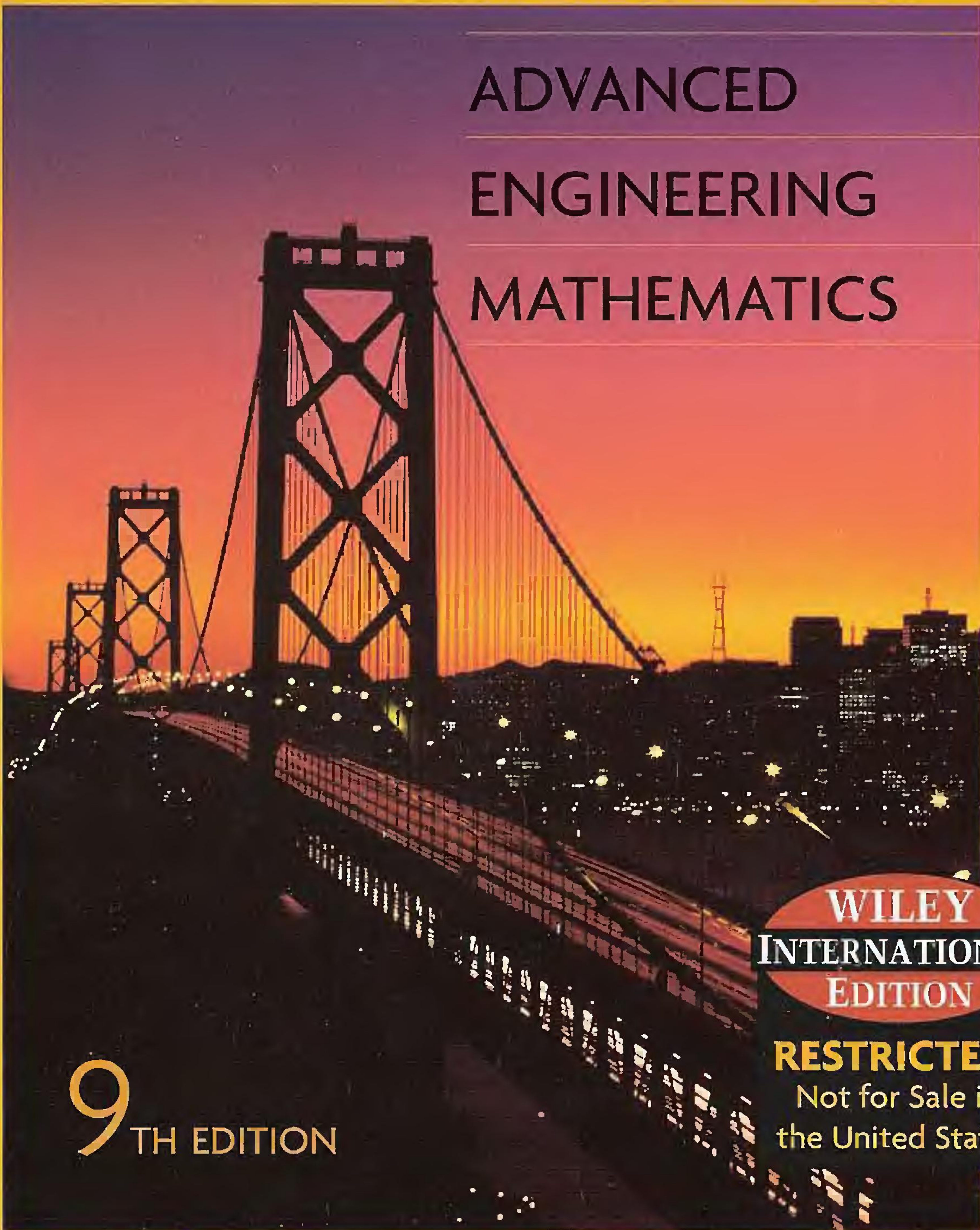
ERWIN KREYSZIG

ADVANCED ENGINEERING MATHEMATICS



9TH EDITION

WILEY INTERNATIONAL EDITION

RESTRICTED! Not for Sale in the United States

CHAPTER 25

Mathematical Statistics

In probability theory we set up mathematical models of processes that are affected by "chance". In mathematical statistics or, briefly, **statistics,** we check these models against the observable reality. This is called **statistical inference.** It is done by **sampling,** that is, by drawing random samples, briefly called **samples.** These are sets of values from a much larger set of values that could be studied, called the **population.** An example is 10 diameters of screws drawn from a large lot of screws. Sampling is done in order to see whether a model of the population is accurate enough for practical purposes. If this is the case, the model can be used for predictions, decisions, and actions, for instance, in planning productions, buying equipment, investing in business projects, and so on.

Most important methods of statistical inference are **estimation of parameters** (Secs. 25.2), determination of **confidence intervals** (Sec. 25.3), and **hypothesis testing** (Secs. 25.4, 25.7, 25.8), with application to *quality control* (Sec. 25.5) and *acceptance sampling* (Sec. 25.6).

In the last section (25.9) we give an introduction to **regression** and **correlation analysis,** which concern experiments involving two variables.

Prerequisite: Chap. 24.
Sections that may be omitted in a shorter course: 25.5, 25.6, 25.8.
References, Answers to Problems, and Statistical Tables: App. 1 Part G, App. 2, App. 5.

25.1 Introduction. Random Sampling

Mathematical statistics consists of methods for designing and evaluating random experiments to obtain information about practical problems, such as exploring the relation between iron content and density of iron ore, the quality of raw material or manufactured products, the efficiency of air-conditioning systems, the performance of certain cars, the effect of advertising, the reactions of consumers to a new product, etc.

Random variables occur more frequently in engineering (and elsewhere) than one would think. For example, properties of mass-produced articles (screws, lightbulbs, etc.) always show **random variation,** due to small (uncontrollable!) differences in raw material or manufacturing processes. Thus the diameter of screws is a random variable X and we have *nondefective screws,* with diameter between given tolerance limits, and *defective screws,* with diameter outside those limits. We can ask for the distribution of X, for the percentage of defective screws to be expected, and for necessary improvements of the production process.

Samples are selected from populations—20 screws from a lot of 1000, 100 of 5000 voters, 8 beavers in a wildlife conservation project—because inspecting the entire population would be too expensive, time-consuming, impossible or even senseless (think of destructive testing of lightbulbs or dynamite). To obtain meaningful conclusions, samples must be **random selections.** Each of the 1000 screws must have the same chance of being sampled (of being drawn when we sample), at least approximately. Only then will the sample mean $\bar{x} = (x_1 + \cdots + x_{20})/20$ (Sec. 24.1) of a sample of size $n = 20$ (or any other n) be a good approximation of the population mean μ (Sec. 24.6); and the accuracy of the approximation will generally improve with increasing n, as we shall see. Similarly for other parameters (standard deviation, variance, etc.).

Independent sample values will be obtained in experiments with an infinite sample space S (Sec. 24.2), certainly for the normal distribution. This is also true in sampling with replacement. It is approximately true in drawing *small* samples from a large finite population (for instance, 5 or 10 of 1000 items). However, if we sample without replacement from a small population, the effect of dependence of sample values may be considerable.

Random uumbers help in obtaining samples that are in fact random selections. This is sometimes not easy to accomplish because there are many subtle factors that can bias sampling (by personal interviews, by poorly working machines, by the choice of nontypical observation conditions, etc.). Random numbers can be obtained from a **random number generator** in Maple, Mathematica, or other systems listed on p. 991. (The numbers are not truly random, as they would be produced in flipping coins or rolling dice, but are calculated by a tricky formula that produces numbers that do have practically all the essential features of true randomness.)

EXAMPLE 1 Random Numbers from a Random Number Generator

To select a sample of size $n = 10$ from 80 given ball bearings, we number the bearings from 1 to 80. We then let the generator randomly produce 10 of the integers from 1 to 80 and include the bearings with the numbers obtained in our sample, for example.

44 55 53 03 52 61 67 78 39 54

or whatever.

Random numbers are also contained in (older) statistical tables. ■

Representing and processing data were considered in Sec. 24.1 in connection with frequency distributions. These are the empirical counterparts of probability distributions and helped motivating axioms and properties in probability theory. The new aspect in this chapter is **randomness:** the data are samples selected **randomly** from a population. Accordingly, we can immediately make the connection to Sec. 24.1, using stem-and-leaf plots, box plots, and histograms for representing samples graphically.

Also, we now call the mean $\bar{x}$ in (5), Sec. 24.1, the **sample mean**

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{1}{n}(x_1 + x_2 + \cdots + x_n). \tag{1}$$

We call n the **sample size,** the variance s^2 in (6), Sec. 24.1, the **sample variance**

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})^2 = \frac{1}{n-1}\left[(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\right], \tag{2}$$

and its positive square root s the **sample staudard deviation.** $\bar{x}$, s^2, and s are called **parameters** *of a sample;* they will be needed throughout this chapter.

25.2 Point Estimation of Parameters

Beginning in this section, we shall discuss the most basic practical tasks in statistics and corresponding statistical methods to accomplish them. The first of them is point estimation of **parameters,** that is, of quantities appearing in distributions, such as p in the binomial distribution and μ and σ in the normal distribution.

A **point estimate** of a parameter is a number (point on the real line), which is computed from a given sample and serves as an approximation of the unknown exact value of the parameter of the population. An **interval estimate** is an interval (*"confidence interval"*) obtained from a sample; such estimates will be considered in the next section. Estimation of parameters is of great practical importance in many applications.

As an approximation of the mean μ of a population we may take the mean $\bar{x}$ of a corresponding sample. This gives the estimate $\hat{\mu} = \bar{x}$ for μ, that is,

$$\hat{\mu} = \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) \tag{1}$$

where n is the sample size. Similarly, an estimate $\hat{\sigma}^2$ for the variance of a population is the variance s^2 of a corresponding sample, that is,

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2. \tag{2}$$

Clearly, (1) and (2) are estimates of parameters for distributions in which μ or σ^2 appear explicity as parameters, such as the normal and Poisson distributions. For the binomial distribution, $p = \mu/n$ [see (3) in Sec. 24.7]. From (1) we thus obtain for p the estimate

$$\hat{p} = \frac{\bar{x}}{n}. \tag{3}$$

We mention that (1) is a special case of the so-called **method of moments.** In this method the parameters to be estimated are expressed in terms of the moments of the distribution (see Sec. 24.6). In the resulting formnlas those moments of the distribution are replaced by the corresponding moments of the sample. This gives the estimates. Here the ***k*th moment of a sample** $x_1, \cdots, x_n$ is

$$m_k = \frac{1}{n}\sum_{j=1}^{n} x_j^{\,k}.$$

Maximum Likelihood Method

Another method for obtaining estimates is the so-called **maximum likelihood method** of R. A. Fisher [*Messenger Math.* **41** (1912), 155–160]. To explain it, we consider a discrete (or continnous) random variable X whose probability function (or density) $f(x)$ depends on a single parameter θ. We take a corresponding sample of n ***independent*** values $x_1, \cdots, x_n$. Then in the discrete case the probability that a sample of size n consists precisely of those n values is

$$l = f(x_1)f(x_2) \cdots f(x_n). \tag{4}$$

In the continuous case the probability that the sample consists of values in the small intervals $x_j \leqq x \leqq x_j + \Delta x$ $(j = 1, 2, \cdots, n)$ is

$$f(x_1)\Delta x\, f(x_2)\Delta x \cdots f(x_n)\Delta x = l(\Delta x)^n. \tag{5}$$

Since $f(x_j)$ depends on θ, the function l in (5) given by (4) depends on $x_1, \cdots, x_n$ and θ. We imagine $x_1, \cdots, x_n$ to be given and fixed. Then l is a function of θ, which is called the **likelihood function.** The basic idea of the maximum likelihood method is quite simple, as follows. We choose *that* approximation for the unknown value of θ for which l is as large as possible. If l is a differentiable function of θ, a necessary condition for l to have a maximum in an interval (not at the boundary) is

$$\frac{\partial l}{\partial \theta} = 0. \tag{6}$$

(We write a *partial* derivative, because l depends also on $x_1, \cdots, x_n$.) A solution of (6) depending on $x_1, \cdots, x_n$ is called a **maximum likelihood estimate** for θ. We may replace (6) by

$$\frac{\partial \ln l}{\partial \theta} = 0, \tag{7}$$

because $f(x_j) > 0$, a maximum of l is in general positive, and $\ln l$ is a monotone increasing function of l. This often simplifies calculations.

Several Parameters. If the distribution of X involves r parameters $\theta_1, \cdots, \theta_r$, then instead of (6) we have the r conditions $\partial l/\partial\theta_1 = 0, \cdots, \partial l/\partial\theta_r = 0$, and instead of (7) we have

$$\frac{\partial \ln l}{\partial \theta_1} = 0, \qquad \cdots, \qquad \frac{\partial \ln l}{\partial \theta_r} = 0. \tag{8}$$

EXAMPLE 1 Normal Distribution

Find maximum likelihood estimates for $\theta_1 = \mu$ and $\theta_2 = \sigma$ in the case of the normal distribution.

Solution. From (1), Sec. 24.8, and (4) we obtain the likelihood function

$$l = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-h} \qquad \text{where} \qquad h = \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2.$$

Taking logarithms, we have

$$\ln l = -n \ln \sqrt{2\pi} - n \ln \sigma - h.$$

The first equation in (8) is $\partial(\ln l)/\partial\mu = 0$, written out

$$\frac{\partial \ln l}{\partial \mu} = -\frac{\partial h}{\partial \mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu) = 0, \qquad \text{hence} \qquad \sum_{j=1}^{n} x_j - n\mu = 0.$$

The solution is the desired estimate $\hat{\mu}$ for μ; we find

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

The second equation in (8) is $\partial(\ln l)/\partial\sigma = 0$, written out

$$\frac{\partial \ln l}{\partial \sigma} = -\frac{n}{\sigma} - \frac{\partial h}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{n}(x_j - \mu)^2 = 0.$$

Replacing μ by $\hat{\mu}$ and solving for σ^2, we obtain the estimate

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

which we shall use in Sec. 25.7. Note that this differs from (2). We cannot discuss criteria for the goodness of estimates but want to mention that for small n, formula (2) is preferable. ■

PROBLEM SET 25.2

1. Find the maximum likelihood estimate for the parameter μ of a normal distribution with known variance $\sigma^2 = \sigma_0^2$.
2. Apply the maximum likelihood method to the normal distribution with $\mu = 0$.
3. **(Binomial distribution)** Derive a maximnm likelihood estimate for p.
4. Extend Prob. 3 as follows. Suppose that m times n trials were made and in the first n trials A happened k_1 times, in the second n trials A happened k_2 times, $\cdots$, in the mth n trials A happened k_m times. Find a maximum likelihood estimate of p based on this information.
5. Suppose that in Prob. 4 we made 4 times 5 trials and A happened 2, 1, 4, 4 times, respectively. Estimate p.
6. Consider X = *Number of independent trials until an event A occurs.* Show that X has the probability function $f(x) = pq^{x-1}$, $x = 1, 2, \cdots$, where p is the probability of A in a single trial and $q = 1 - p$. Find the maximum likelihood estimate of p corresponding to a sample $x_1, \cdots, x_n$ of observed values of X.
7. In Prob. 6 find the maximum likelihood estimate of p corresponding to a single observation x of X.
8. In rolling a die, suppose that we get the first Six in the 7th trial and in doing it again we get it in the 6th trial. Estimate the probability p of getting a Six in rolling that die once.
9. **(Poisson distribution)** Apply the maximum likelihood method to the Poisson distribution.
10. **(Uniform distribution)** Show that in the case of the parameters a and b of the uniform distribution (see Sec. 24.6), the maximum likelihood estimate cannot be obtained by equating the first derivative to zero. How can we obtain maximum likelihood estimates in this case?
11. Find the maximum likelihood estimate of θ in the density $f(x) = \theta e^{-\theta x}$ if $x \geqq 0$ and $f(x) = 0$ if $x < 0$.
12. In Prob. 11, find the mean μ, substitute it in $f(x)$, find the maximum likelihood estimate of μ, and show that it is identical with the estimate for μ which can be obtained from that for θ in Prob. 11.
13. Compute $\hat{\theta}$ in Prob. 11 from the sample 1.8, 0.4, 0.8, 0.6, 1.4. Graph the sample distribution function $\hat{F}(x)$ and the distribution function $F(x)$ of the random variable, with $\theta = \hat{\theta}$, on the same axes. Do they agree reasonably well? (We consider goodness of fit systematically in Sec. 25.7.)

14. Do the same task as in Prob. 13 if the given sample is 0.5, 0.7, 0.1, 1.1, 0.1.

15. CAS EXPERIMENT. Maximum Likelihood Estimates. (MLEs). Find experimentally how much MLEs can differ depending on the sample size. *Hint.* Generate many samples of the same size n, e.g., of the standardized normal distribution, and record $\bar{x}$ and s^2. Then increase n.

25.3 Confidence Intervals

Confidence intervals[1] for an unknown parameter θ of some distribution (e.g., $\theta = \mu$) are intervals $\theta_1 \leqq \theta \leqq \theta_2$ that contain θ, not with certainty but with a high probability γ, which we can choose (95% and 99% are popular). Such an interval is calculated from a sample. $\gamma = 95\%$ means probability $1 - \gamma = 5\% = 1/20$ of being wrong—one of about 20 such intervals will not contain θ. Instead of writing $\theta_1 \leqq \theta \leqq \theta_2$, we denote this more distinctly by writing

$$\text{CONF}_\gamma \{\theta_1 \leqq \theta \leqq \theta_2\}. \tag{1}$$

Such a special symbol, CONF, seems worthwhile in order to avoid the misunderstanding that θ *must* lie between θ_1 and θ_2.

γ is called the **confidence level,** and θ_1 and θ_2 are called the **lower** and **upper confidence limits.** They depend on γ. The larger we choose γ, the smaller is the error probability $1 - \gamma$, but the longer is the confidence interval. If $\gamma \to 1$, then its length goes to infinity. The choice of γ depends on the kind of application. In taking no umbrella, a 5% chance of getting wet is not tragic. In a medical decision of life or death, a 5% chance of being wrong may be too large and a 1% chance of being wrong ($\gamma = 99\%$) may be more desirable.

Confidence intervals are more valuable than point estimates (Sec. 25.2). Indeed, we can take the midpoint of (1) as an approximation of θ and half the length of (1) as an "error bound" (not in the strict sense of numerics, but except for an error whose probability we know).

θ_1 and θ_2 in (1) are calculated from a sample $x_1, \cdots, x_n$. These are n observations of a random variable X. Now comes a **standard trick.** *We regard* $x_1, \cdots, x_n$ *as* ***single*** *observations of* n *random variables* $X_1, \cdots, X_n$ *(with the same distribution, namely, that of* X*).* Then $\theta_1 = \theta_1(x_1, \cdots, x_n)$ and $\theta_2 = \theta_2(x_1, \cdots, x_n)$ in (1) are observed values of two random variables $\Theta_1 = \Theta_1(X_1, \cdots, X_n)$ and $\Theta_2 = \Theta_2(X_1, \cdots, X_n)$. The condition (1) involving γ can now be written

$$P(\Theta_1 \leqq \theta \leqq \Theta_2) = \gamma. \tag{2}$$

Let us see what all this means in concrete practical cases.

In each case in this section we shall first state the steps of obtaining a confidence interval in the form of a table, then consider a typical example, and finally justify those steps theoretically.

[1] JERZY NEYMAN (1894–1981), American statistician, developed the theory of confidence intervals (*Annals of Mathematical Statistics* **6** (1935), 111–116).

Confidence Interval for μ of the Normal Distribution with Known σ^2

Table 25.1 Determination of a Confidence Interval for the Mean μ of a Normal Distribution with Known Variance σ^2

Step 1. Choose a confidence level γ (95%, 99%, or the like).

Step 2. Determine the corresponding c:

γ	0.90	0.95	0.99	0.999
c	1.645	1.960	2.576	3.291

Step 3. Compute the mean $\bar{x}$ of the sample $x_1, \cdots, x_n$.

Step 4. Compute $k = c\sigma/\sqrt{n}$. The confidence interval for μ is

(3) $$\mathrm{CONF}_\gamma \{\bar{x} - k \leqq \mu \leqq \bar{x} + k\}.$$

EXAMPLE 1 Confidence Interval for μ of the Normal Distribution with Known σ^2

Deterimine a 95% confidence interval for the mean of a normal distribution with variance $\sigma^2 = 9$, using a sample of $n = 100$ values with mean $\bar{x} = 5$.

Solution. *Step 1.* $\gamma = 0.95$ is required. *Step 2.* The corresponding c equals 1.960; see Table 25.1. *Step 3.* $\bar{x} = 5$ is given. *Step 4.* We need $k = 1.960 \cdot 3/\sqrt{100} = 0.588$. Hence $\bar{x} - k = 4.412$, $\bar{x} + k = 5.588$ and the confidence interval is $\mathrm{CONF}_{0.95}\{4.412 \leqq \mu \leqq 5.588\}$.

This is sometimes written $\mu = 5 \pm 0.588$, but we shall not use this notation, which can be misleading.

With your CAS you can determine this interval more directly. Similarly for the other examples in this section. ■

Theory for Table 25.1. The method in Table 25.1 follows from the basic

THEOREM 1 Sum of Independent Normal Random Variables

*Let $X_1, \cdots, X_n$ be **independent** normal random variables each of which has mean μ and variance σ^2. Then the following holds.*

(a) *The sum $X_1 + \cdots + X_n$ is normal with mean $n\mu$ and variance $n\sigma^2$.*

(b) The following random variable $\bar{X}$ *is normal with mean μ and variance σ^2/n.*

(4) $$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

(c) *The following random variable Z is normal with mean* 0 *and variance* 1.

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

PROOF The statements about the mean and variance in (a) follow from Theorems 1 and 3 in Sec. 24.9. From this and Theorem 2 in Sec. 24.6 we see that $\bar{X}$ has the mean $(1/n)n\mu = \mu$ and the variance $(1/n)^2 n\sigma^2 = \sigma^2/n$. This implies that Z has the mean 0 and variance 1, by Theorem 2(b) in Sec. 24.6. The normality of $X_1 + \cdots + X_n$ is proved in Ref. [G3] listed in App. 1. This implies the normality of (4) and (5). ■

Derivation of (3) in Table 25.1. Sampling from a normal distribution gives independent sample values (see Sec. 25.1), so that Theorem 1 applies. Hence we can choose γ and then determine c such that

$$P(-c \leqq Z \leqq c) = P\left(-c \leqq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leqq c\right) = \Phi(c) - \Phi(-c) = \gamma. \tag{6}$$

For the value $\gamma = 0.95$ we obtain $z(D) = 1.960$ from Table A8 in App. 5, as used in Example 1. For $\gamma = 0.9, 0.99. 0.999$ we get the other values of c listed in Table 25.1. Finally, all we have to do is to convert the inequality in (6) into one for μ and insert observed values obtained from the sample. We multiply $-c \leqq Z \leqq c$ by -1 and then by $\sigma/\sqrt{n}$, writing $c\sigma/\sqrt{n} = k$ (as in Table 25.1),

$$\begin{aligned} P(-c \leqq Z \leqq c) = P(c \geqq -Z \geqq -c) &= P\left(c \geqq \frac{\mu - \bar{X}}{\sigma/\sqrt{n}} \geqq -c\right) \\ &= P(k \geqq \mu - \bar{X} \geqq -k) = \gamma. \end{aligned}$$

Adding $\bar{X}$ gives $P(\bar{X} + k \geqq \mu \geqq \bar{X} - k) = \gamma$ or

$$P(\bar{X} - k \leqq \mu \leqq \bar{X} + k) = \gamma. \tag{7}$$

Inserting the observed value $\bar{x}$ of $\bar{X}$ gives (3). Here we have regarded $x_1, \cdots, x_n$ as single observations of $X_1, \cdots, X_n$ (the standard trick!), so that $x_1 + \cdots + x_n$ is an observed value of $X_1 + \cdots + X_n$ and $\bar{x}$ is an observed value of $\bar{X}$. Note further that (7) is of the form (2) with $\Theta_1 = \bar{X} - k$ and $\Theta_2 = \bar{X} + k$. ■

EXAMPLE 2 Sample Size Needed for a Confidence Interval of Prescribed Length

How large must n be in Example 1 if we want to obtain a 95% confidence interval of length $L = 0.4$?

Solution. The interval (3) has the length $L = 2k = 2c\sigma/\sqrt{n}$. Solving for n, we obtain

$$n = (2c\sigma/L)^2.$$

In the present case the answer is $n = (2 \cdot 1.960 \cdot 3/0.4)^2 \approx 870$.

Figure 525 shows how L decreases as n increases and that for $\gamma = 99\%$ the confidence interval is substantially longer than for $\gamma = 95\%$ (and the same sample size n). ■

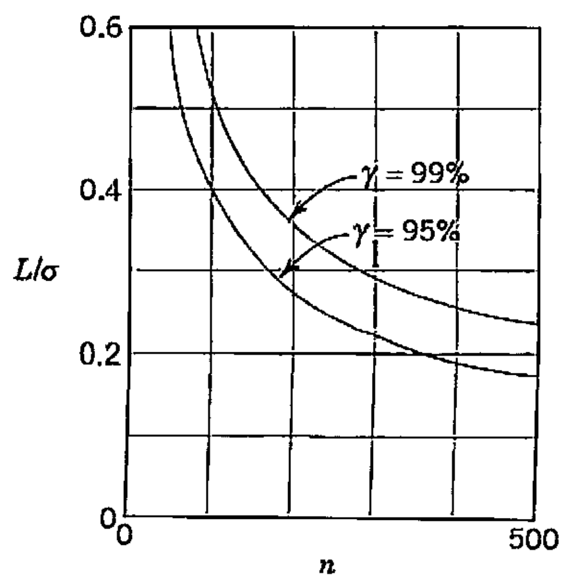


Fig. 525. Length of the confidence interval (3) (measured in multiples of σ) as a function of the sample size n for $\gamma = 95\%$ and $\gamma = 99\%$

Confidence Interval for μ of the Normal Distribution With Unknown σ^2

In practice σ^2 is frequently unknown. Then the method in Table 25.1 does not help and the whole theory changes, although the steps of determining a confidence interval for μ remain quite similar. They are shown in Table 25.2. We see that k differs from that in Table 25.1, namely, the sample standard deviation s has taken the place of the unknown standard deviation σ of the population. And c now depends on the sample size n and must be determined from Table A9 in App. 5 or from your CAS. That table lists values z for given values of the distribution function (Fig. 526)

$$F(z) = K_m \int_{-\infty}^{z} \left(1 + \frac{u^2}{m}\right)^{-(m+1)/2} du \tag{8}$$

of the ***t*-distribution**. Here, m (= 1, 2, $\cdots$) is a parameter, called the **number of degrees of freedom** of the distribution (***abbreviated*** **d.f.**). In the present case, $m = n - 1$; see Table 25.2. The constant K_m is such that $F(\infty) = 1$. By integration it turns out that $K_m = \Gamma(\frac{1}{2}m + \frac{1}{2})/\left[\sqrt{m\pi}\,\Gamma(\frac{1}{2}m)\right]$, where Γ is the gamma function (see (24) in App. A3.1).

Table 25.2 Determination of a Confidence Interval for the Mean μ of a Normal Distribution with Unknown Variance σ^2

Step 1. Choose a confidence level γ (95%, 99%, or the like).

Step 2. Determine the solution c of the equation

$$F(c) = \tfrac{1}{2}(1 + \gamma) \tag{9}$$

from the table of the t-distribution with $n - 1$ degrees of freedom (Table A9 in App. 5; or use a CAS; n = sample size).

Step 3. Compute the mean $\bar{x}$ and the variance s^2 of the sample $x_1, \cdots, x_n$.

Step 4. Compute $k = cs/\sqrt{n}$. The confidence interval is

$$\mathrm{CONF}_\gamma \{\bar{x} - k \leqq \mu \leqq \bar{x} + k\}. \tag{10}$$

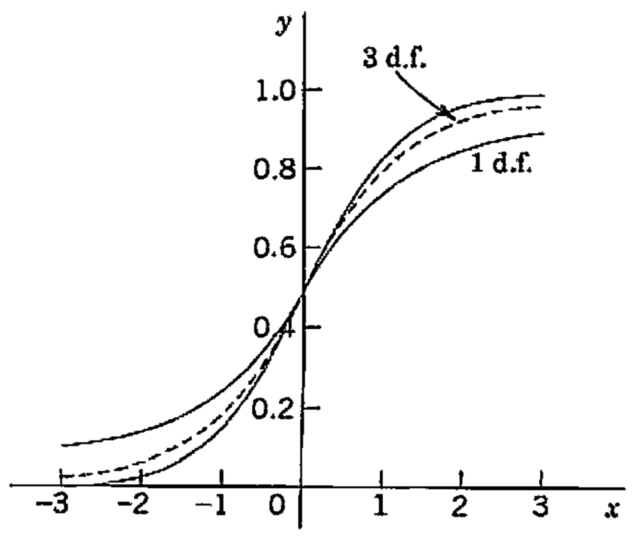


Fig. 526. Distribution functions of the t-distribution with 1 and 3 d.f. and of the standardized normal distribution (steepest curve)

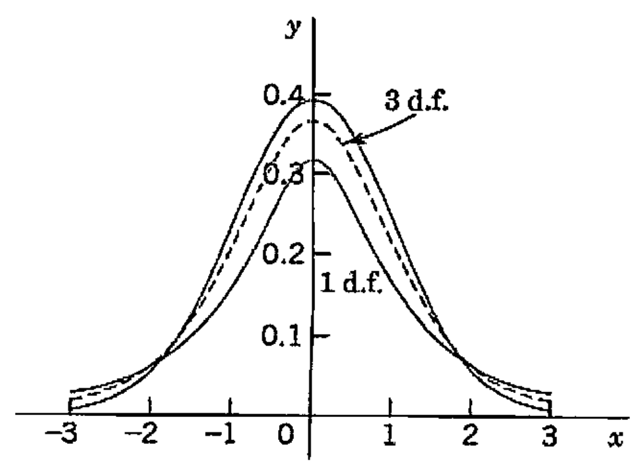


Fig. 527. Densities of the t-distribution with 1 and 3 d.f. and of the standardized normal distribution

Figure 527 compares the curve of the density of the t-distribution with that of the normal distribution. The latter is steeper. This illustrates that Table 25.1 (which uses more information, namely, the known value of σ^2) yields shorter confidence intervals than Table 25.2. This is confirmed in Fig. 528, which also gives an idea of the gain by increasing the sample size.

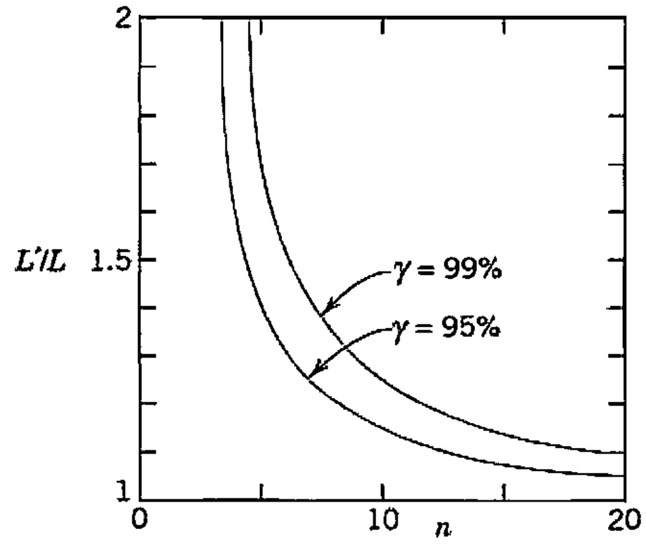


Fig. 528. Ratio of the lengths L' and L of the confidence intervals (10) and (3) with $\gamma = 95\%$ and $\gamma = 99\%$ as a function of the sample size n for equal s and σ

EXAMPLE 3 Confidence Interval for μ of the Normal Distribution with Unknown σ^2

Five independent measurements of the point of inflammation (flash point) of Diesel oil (D-2) gave the values (in °F) 144 147 146 142 144. Assuming normality, determine a 99% confidence interval for the mean.

Solution. *Step 1.* $\gamma = 0.99$ is required.

Step 2. $F(c) = \frac{1}{2}(1 + \gamma) = 0.995$, and Table A9 in App. 5 with $n - 1 = 4$ d.f. gives $c = 4.60$.

Step 3. $\bar{x} = 144.6$. $s^2 = 3.8$.

Step 4. $k = \sqrt{3.8} \cdot 4.60/\sqrt{5} = 4.01$. The confidence interval is $\text{CONF}_{0.99}\{140.5 \leqq \mu \leqq 148.7\}$.

If the variance σ^2 were known and equal to the sample variance s^2, thus $\sigma^2 = 3.8$, then Table 25.1 would give $k = c\sigma/\sqrt{n} = 2.576\sqrt{3.8}/\sqrt{5} = 2.25$ and $\text{CONF}_{0.99}\{142.35 \leqq \mu \leqq 146.85\}$. We see that the present interval is almost twice as long as that obtained from Table 25.1 (with $\sigma^2 = 3.8$). Hence for small samples the difference is considerable! See also Fig. 528. ■

Theory for Table 25.2. For deriving (10) in Table 25.2 we need from Ref. [G3]

THEOREM 2 Student's t-Distribution

Let $X_1, \cdots, X_n$ be independent normal random variables with the same mean μ and the same variance σ^2. Then the random variable

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \tag{11}$$

has a t-distribution [see (8)] *with $n - 1$ degrees of freedom* (d.f.); *here $\bar{X}$ is given by* (4) *and*

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})^2. \tag{12}$$

Derivation of (10). This is similar to the derivation of (3). We choose a number γ between 0 and 1 and determine a number c from Table A9 in App. 5 with $n - 1$ d.f. (or from a CAS) such that

$$P(-c \leqq T \leqq c) = F(c) - F(-c) = \gamma. \tag{13}$$

Since the t-distribution is symmetric, we have

$$F(-c) = 1 - F(c),$$

and (13) assumes the form (9). Substituting (11) into (13) and transforming the result as before, we obtain

$$P(\overline{X} - K \leqq \mu \leqq \overline{X} + K) = \gamma \tag{14}$$

where

$$K = cS/\sqrt{n}.$$

By inserting the observed values $\bar{x}$ of $\overline{X}$ and s^2 of S^2 into (14) we finally obtain (10). ■

Confidence Interval for the Variance σ^2 of the Normal Distribution

Table 25.3 shows the steps, which are similar to those in Tables 25.1 and 25.2.

Table 25.3 Determination of a Confidence Interval for the Variance σ^2 of a Normal Distribution, Whose Mean Need Not Be Known

Step 1. Choose a confidence level γ (95%, 99%, or the like).

Step 2. Determine solutions c_1 and c_2 of the equations

$$F(c_1) = \tfrac{1}{2}(1 - \gamma), \qquad F(c_2) = \tfrac{1}{2}(1 + \gamma) \tag{15}$$

from the table of the chi-square distribution with $n - 1$ degrees of freedom (Table A10 in App. 5; or use a CAS; n = sample size).

Step 3. Compute $(n - 1)s^2$, where s^2 is the variance of the sample $x_1, \cdots, x_n$.

Step 4. Compute $k_1 = (n - 1)s^2/c_1$ and $k_2 = (n - 1)s^2/c_2$. The confidence interval is

$$\text{CONF}_\gamma \{k_2 \leqq \sigma^2 \leqq k_1\}. \tag{16}$$

EXAMPLE 4 Confidence Interval for the Variance of the Normal Distribution

Determine a 95% confidence interval (16) for the variance, using Table 25.3 and a sample (tensile strength of sheet steel in kg/mm^2, rounded to integer values)

$$89 \quad 84 \quad 87 \quad 81 \quad 89 \quad 86 \quad 91 \quad 90 \quad 78 \quad 89 \quad 87 \quad 99 \quad 83 \quad 89.$$

Solution. *Step 1.* $\gamma = 0.95$ is required.

Step 2. For $n - 1 = 13$ we find

$$c_1 = 5.01 \quad \text{and} \quad c_2 = 24.74.$$

Step 3. $13s^2 = 326.9$.

Step 4. $13s^2/c_1 = 65.25$, $13s^2/c_2 = 13.21$.

The confidence interval is

$$\mathrm{CONF}_{0.95}\,\{13.21 \leqq \sigma^2 \leqq 65.25\}.$$

This is rather large, and for obtaining a more precise result, one would need a much larger sample. ■

Theory for Table 25.3. In Table 25.1 we used the normal distribution, in Table 25.2 the t-distribution, and now we shall use the **χ^2-distribution** (*chi-square distribution*), whose distribution function is $F(z) = 0$ if $z < 0$ and

$$F(z) = C_m \int_0^z e^{-u/2} u^{(m-2)/2}\, du \qquad \text{if } z \geqq 0 \qquad \text{(Fig. 529).}$$

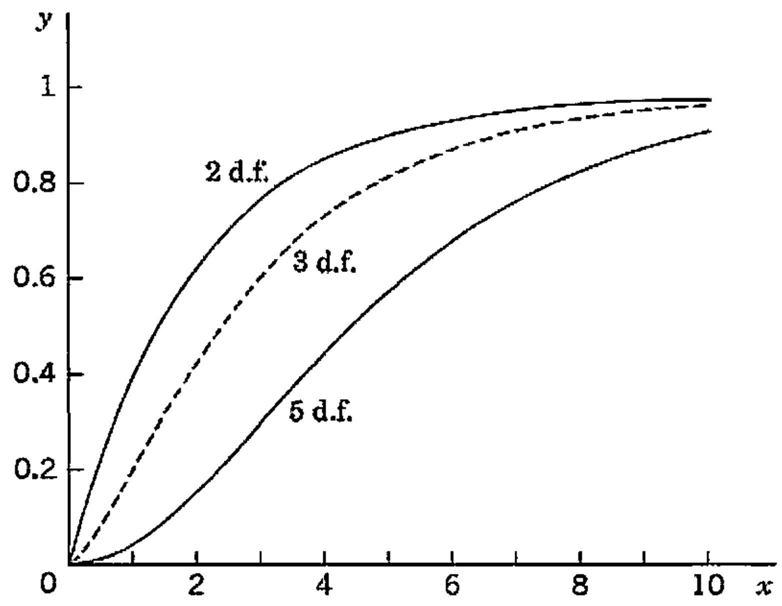


Fig. 529. Distribution function of the chi-square distribution with 2, 3, 5 d.f.

The parameter m ($= 1, 2, \cdots$) is called the **number of degrees of freedom** (d.f.), and

$$C_m = 1/[2^{m/2}\Gamma(\tfrac{1}{2}m)].$$

Note that the distribution is not symmetric (see also Fig. 530).

For deriving (16) in Table 25.3 we need the following theorem.

THEOREM 3

Chi-Square Distribution

Under the assumptions in Theorem 2 the random variable

$$Y = (n-1)\frac{S^2}{\sigma^2} \tag{17}$$

with S^2 given by (12) *has a chi-square distribution with $n - 1$ degrees of freedom.*

Proof in Ref. [G3], listed in App. 1.

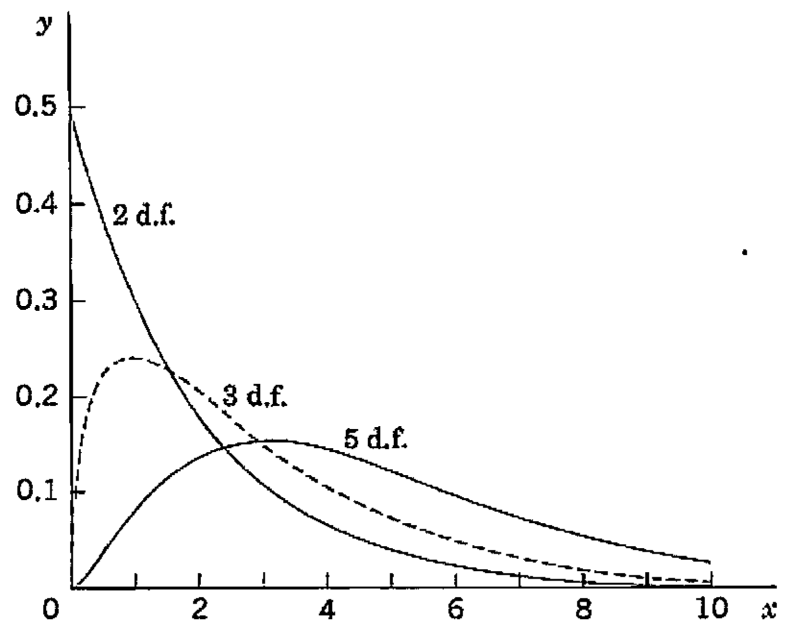


Fig. 530. Density of the chi-square distribution with 2, 3, 5 d.f.

Derivation of (16). This is similar to the derivation of (3) and (10). We choose a number γ between 0 and 1 and determine c_1 and c_2 from Table A10, App. 5, such that [see (15)]

$$P(Y \leqq c_1) = F(c_1) = \tfrac{1}{2}(1 - \gamma), \qquad P(Y \leqq c_2) = F(c_2) = \tfrac{1}{2}(1 + \gamma).$$

Subtraction yields

$$P(c_1 \leqq Y \leqq c_2) = P(Y \leqq c_2) - P(Y \leqq c_1) = F(c_2) - F(c_1) = \gamma.$$

Transforming $c_1 \leqq Y \leqq c_2$ with Y given by (17) into an inequality for σ^2, we obtain

$$\frac{n-1}{c_2} S^2 \leqq \sigma^2 \leqq \frac{n-1}{c_1} S^2.$$

By inserting the observed value s^2 of S^2 we obtain (16). ■

Confidence Intervals for Parameters of Other Distributions

The methods in Tables 25.1–25.3 for confidence intervals for μ and σ^2 are designed for the normal distribution. We now show that they can also be applied to other distributions if we use large samples.

We know that if $X_1, \cdots, X_n$ are independent random variables with the same mean μ and the same variance σ^2, then their sum $Y_n = X_1 + \cdots + X_n$ has the following properties.

(A) Y_n has the mean $n\mu$ and the variance $n\sigma^2$ (by Theorems 1 and 3 in Sec. 24.9).

(B) If those variables are normal, then Y_n is normal (by Theorem 1).

If those random variables are not normal, then **(B)** is not applicable. However, for large n the random variable Y_n is still *approximately* normal. This follows from the central limit theorem, which is one of the most fundamental results in probability theory.

THEOREM 4

Central Limit Theorem

Let $X_1, \cdots, X_n, \cdots$ be independent random variables that have the same distribution function and therefore the same mean μ and the same variance σ^2. Let $Y_n = X_1 + \cdots + X_n$. Then the random variable

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}} \tag{18}$$

is **asymptotically normal** *with mean* 0 *and variance* 1; *that is, the distribution function $F_n(x)$ of Z_n satisfies*

$$\lim_{n\to\infty} F_n(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du.$$

A proof can be found in Ref. [G3] listed in App. 1.

Hence when applying Tables 25.1–25.3 to a nonnormal distribution, we must use ***sufficiently large samples.*** As a rule of thumb, if the sample indicates that the skewness of the distribution (the asymmetry; see Team Project 16(d), Problem Set 24.6) is small, use at least $n = 20$ for the mean and at least $n = 50$ for the variance.

PROBLEM SET 25.3

1–7 MEAN (VARIANCE KNOWN)

1. Find a 95% confidence interval for the mean μ of a normal population with standard deviation 4.00 from the sample 30, 42, 40, 34, 48, 50.
2. Does the interval in Prob. 1 get longer or shorter if we take $\gamma = 0.99$ instead of 0.95? By what factor?
3. By what factor does the length of the interval in Prob. 1 change if we double the sample size?
4. Find a 90% confidence interval for the mean μ of a normal population with variance 0.25, using a sample of 100 values with mean 212.3.
5. What sample size would be needed for obtaining a 95% confidence interval (3) of length 2σ? Of length σ?
6. **(Use of Fig. 525)** Find a 95% confidence interval for a sample of 200 values with mean 120 from a normal distribution with variance 4, using Fig. 525.
7. What sample size is needed to obtain a 99% confidence interval of length 2.0 for the mean of a normal population with variance 25? Use Fig. 525. Check by calculation.

8–12 MEAN (VARIANCE UNKNOWN)

Find a 99% confidence interval for the mean of a normal population from the sample:

8. 425, 420, 425, 435
9. Length of 20 bolts with sample mean 20.2 cm and sample variance 0.04 cm^2
10. Knoop hardness of diamond 9500, 9800, 9750, 9200, 9400, 9550
11. Copper content (%) of brass 66, 66, 65, 64, 66, 67, 64, 65, 63, 64
12. Melting point (°C) of aluminum 660, 667, 654, 663, 662
13. Find a 95% confidence interval for the percentage of cars on a certain highway that have poorly adjusted brakes, using a random sample of 500 cars stopped at a roadblock on that highway, 87 of which had poorly adjusted brakes.
14. Find a 99% confidence interval for p in the binomial distribution from a classical result by K. Pearson, who in 24000 trials of tossing a coin obtained 12012 Heads. Do you think that the coin was fair?

15–20 VARIANCE

Find a 95% confidence interval for the variance of a normal population from the sample:

15. A sample of 30 values with variance 0.0007

16. The sample in Prob. 9

17. The sample in Prob. 11

18. Carbon monoxide emission (grams per mile) of a certain type of passenger car (cruising at 55 mph): 17.3, 17.8, 18.0, 17.7, 18.2, 17.4, 17.6, 18.1

19. Mean energy (keV) of delayed neutron group (Group 3, half-life 6.2 sec.) for uranium U^{235} fission: 435, 451, 430, 444, 438

20. Ultimate tensile strength (k psi) of alloy steel (Maraging H) at room temperature: 251, 255, 258, 253, 253, 252, 250, 252, 255, 256

21. If X is normal with mean 27 and variance 16, what distributions do $-X$, $3X$, and $5X - 2$ have?

22. If X_1 and X_2 are independent normal random variables with mean 23 and 4 and variance 3 and 1, respectively, what distribution does $4X_1 - X_2$ have? *Hint.* Use Team Project 14(g) in Sec. 24.8.

23. A machine fills boxes weighing Y lb with X lb of salt, where X and Y are normal with mean 100 lb and 5 lb and standard deviation 1 lb and 0.5 lb, respectively. What percent of filled boxes weighing between 104 lb and 106 lb are to be expected?

24. If the weight X of bags of cement is normally distributed with a mean of 40 kg and a standard deviation of 2 kg, how many bags can a delivery truck carry so that the probability of the total load exceeding 2000 kg will be 5%?

25. CAS EXPERIMENT. Confidence Intervals. Obtain 100 samples of size 10 of the standardized normal distribution. Calculate from them and graph the corresponding 95% confidence intervals for the mean and count how many of them do not contain 0. Does the result support the theory? Repeat the whole experiment, compare and comment.

25.4 Testing of Hypotheses. Decisions

The ideas of confidence intervals and of tests[2] are the two most important ideas in modern statistics. In a statistical **test** we make inference from sample to population through testing a **hypothesis,** resulting from experience or observations, from a theory or a quality requirement, and so on. In many cases the result of a test is used as a basis for a **decision,** for instance, to buy (or not to buy) a certain model of car, depending on a test of the fuel efficiency (miles/gal) (and other tests, of course), to apply some medication, depending on a test of its effect; to proceed with a marketing strategy, depending on a test of consumer reactions, etc.

Let us explain such a test in terms of a typical example and introduce the corresponding standard notions of statistical testing.

EXAMPLE 1 Test of a Hypothesis. Alternative. Significance Level α

We want to buy 100 coils of a certain kind of wire, provided we can verify the manufacturer's claim that the wire has a breaking limit $\mu = \mu_0 = 200$ lb (or more). This is a test of the **hypothesis** (also called *null hypothesis*) $\mu = \mu_0 = 200$. We shall not buy the wire if the (statistical) test shows that actually $\mu = \mu_1 < \mu_0$, the wire is weaker, the claim does not hold. μ_1 is called the **alternative** (or *alternative hypothesis*) of the test. We shall **accept** the hypothesis if the test suggests that it is true, except for a small error probability α, called the **significance level** of the test. Otherwise we **reject** the hypothesis. Hence α is the probability of rejecting a hypothesis although it is true. The choice of α is up to us. 5% and 1% are popular values.

For the test we need a sample. We randomly select 25 coils of the wire, cut a piece from each coil, and determine the breaking limit experimentally. Suppose that this sample of $n = 25$ values of the breaking limit has the mean $\bar{x} = 197$ lb (somewhat less than the claim!) and the standard deviation $s = 6$ lb.

[2]Beginning around 1930, a systematic theory of tests was developed by NEYMAN (see Sec. 25.3) and EGON SHARPE PEARSON (1895–1980), English statistician, the son of Karl Pearson (see the footnote on p. 1066).

At this point we could only speculate whether this difference $197 - 200 = -3$ is due to randomness, is a chance effect, or whether it is **significant,** due to the actually inferior quality of the wire. To continue beyond speculation requires probability theory, as follows.

We assume that the breaking limit is normally distributed. (This assumption could be tested by the method in Sec. 25.7. Or we could remember the central limit theorem (Sec. 25.3) and take a still larger sample.) Then

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

in (11), Sec. 25.3, with $\mu = \mu_0$ has a **t-distribution** with $n - 1$ degrees of freedom ($n - 1 = 24$ for our sample). Also $\bar{x} = 197$ and $s = 6$ are observed values of $\bar{X}$ and S to be used later. We can now choose a significance level, say, $\alpha = 5\%$. From Table A9 in App. 5 or from a CAS we then obtain a critical value c such that $P(T \leqq c) = \alpha = 5\%$. For $P(T \leqq \tilde{c}) = 1 - \alpha = 95\%$ the table gives $\tilde{c} = 1.71$, so that $c = -\tilde{c} = -1.71$ because of the symmetry of the distribution (Fig. 531).

We now reason as follows—this is the ***crucial idea*** of the test. If the hypothesis is true, we have a chance of only α ($= 5\%$) that we observe a value t of T (calculated from a sample) that will fall between $-\infty$ and -1.71. Hence if we nevertheless do observe such a t, we assert that the hypothesis cannot be true and we reject it. Then we accept the alternative. If, however, $t \geqq c$, we accept the hypothesis.

A simple calculation finally gives $t = (197 - 200)/(6/\sqrt{25}) = -2.5$ as an observed value of T. Since $-2.5 < -1.71$, we reject the hypothesis (the manufacturer's claim) and accept the alternative $\mu = \mu_1 < 200$, the wire seems to be weaker than claimed. ■

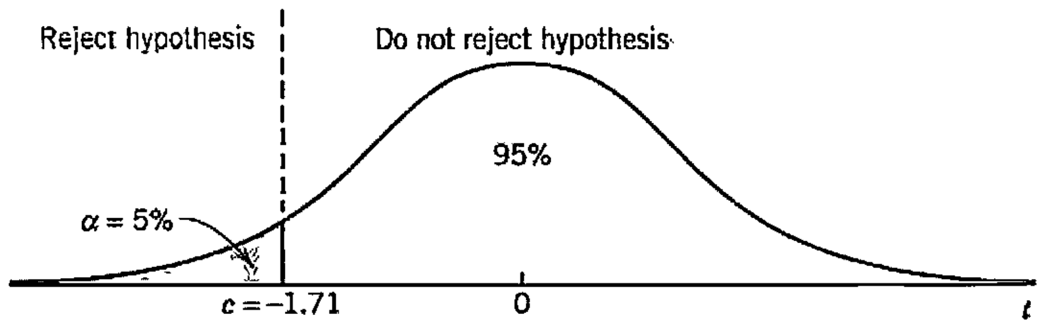


Fig. 531. t-distribution in Example 1

This example illustrates the ***steps of a test:***

1. Formulate the **hypothesis** $\theta = \theta_0$ to be tested. ($\theta_0 = \mu_0$ in the example.)

2. Formulate an **alternative** $\theta = \theta_1$. ($\theta_1 = \mu_1$ in the example.)

3. Choose a **significance level** α (5%, 1%, 0.1%).

4. Use a random variable $\hat{\Theta} = g(X_1, \cdots, X_n)$ whose distribution depends on the hypothesis and on the alternative, and this distribution is known in both cases. Determine a critical value c from the distribution of $\hat{\Theta}$, assuming the hypothesis to be true. (In the example, $\hat{\Theta} = T$, and c is, obtained from $P(T \leqq c) = \alpha$.)

5. Use a sample $x_1, \cdots, x_n$ to determine an observed value $\hat{\theta} = g(x_1, \cdots, x_n)$ of $\hat{\Theta}$. (t in the example.)

6. Accept or reject the hypothesis, depending on the size of $\hat{\theta}$ relative to c. ($t < c$ in the example, rejection of the hypothesis.)

Two important facts require further discussion and careful attention. The first is the choice of an alternative. In the example, $\mu_1 < \mu_0$, but other applications may require $\mu_1 > \mu_0$ or $\mu_1 \neq \mu_0$. The second fact has to do with errors. We know that α (the significance level of the test) is the probability of *rejecting* a *true* hypothesis. And we shall discuss the probability β of *accepting* a *false* hypothesis.

One-Sided and Two-Sided Alternatives (Fig. 532)

Let θ be an unknown parameter in a distribution, and suppose that we want to test the hypothesis $\theta = \theta_0$. Then there are three main kinds of alternatives, namely,

(1) $$\theta > \theta_0$$

(2) $$\theta < \theta_0$$

(3) $$\theta \neq \theta_0.$$

(1) and (2) are **one-sided alternatives,** and (3) is a **two-sided alternative.**

We call **rejection region** (or **critical region**) the region such that we reject the hypothesis if the observed value in the test falls in this region. In ① the critical c lies to the right of θ_0 because so does the alternative. Hence the rejection region extends to the right. This is called a **right-sided test.** In ② the critical c lies to the left of θ_0 (as in Example 1), the rejection region extends to the left, and we have a **left-sided test** (Fig. 532, middle part). These are **one-sided tests.** In ③ we have two rejection regions. This is called a **two-sided test** (Fig. 532, lower part).

All three kinds of alternatives occur in practical problems. For example, (1) may arise if θ_0 is the maximum tolerable inaccuracy of a voltmeter or some other instrument. Alternative (2) may occur in testing strength of material, as in Example 1. Finally, θ_0 in (3) may be the diameter of axle-shafts, and shafts that are too thin or too thick are equally undesirable, so that we have to watch for deviations in both directions.

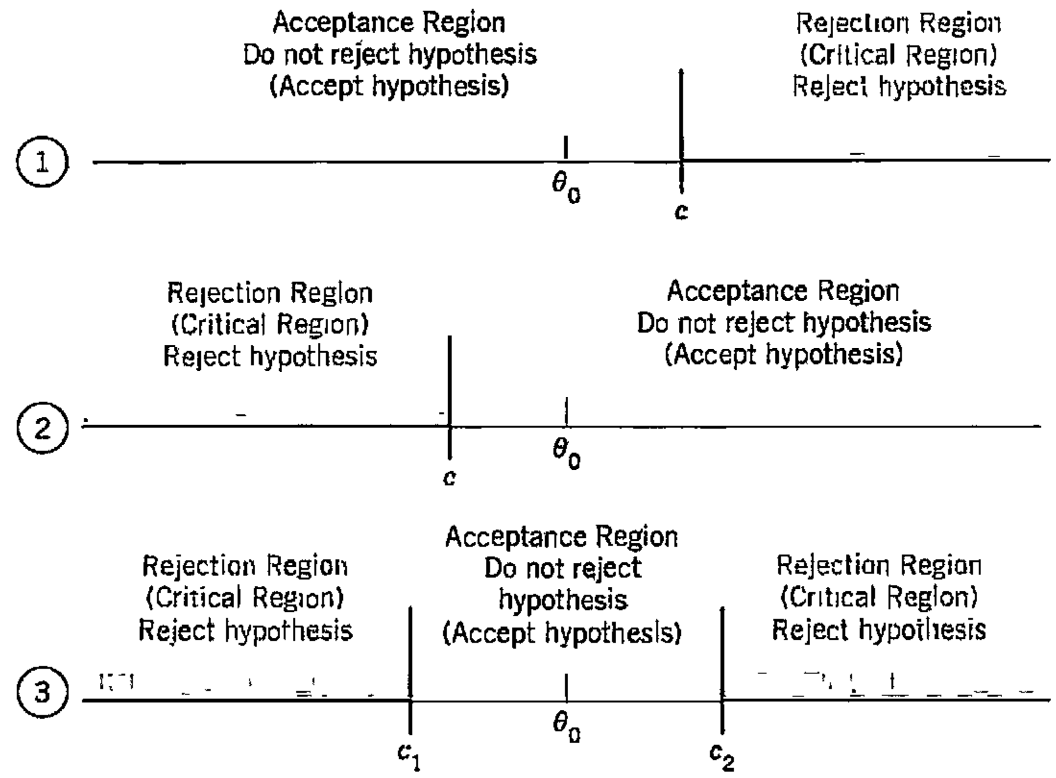


Fig. 532. Test in the case of alternative (1) (upper part of the figure), alternative (2) (middle part), and alternative (3)

Errors in Tests

Tests always involve **risks of making false decisions:**

(I) Rejecting a true hypothesis **(Type I error).**
α = Probability of making a Type I error.

(II) Accepting a false hypothesis **(Type II error).**
β = Probability of making a Type II error.

Clearly, we cannot avoid these errors because no absolutely certain conclusions about populations can be drawn from samples. But we show that there are ways and means of choosing suitable levels of risks, that is, of values α and β. The choice of α depends on the nature of the problem (e.g., a small risk $\alpha = 1\%$ is used if it is a matter of life or death).

Let us discuss this systematically for a test of a hypothesis $\theta = \theta_0$ against an alternative that is a single number θ_1, for simplicity. We let $\theta_1 > \theta_0$, so that we have a right-sided test. For a left-sided or a two-sided test the discussion is quite similar.

We choose a critical $c > \theta_0$ (as in the upper part of Fig. 532, by methods discussed below). From a given sample $x_1, \cdots, x_n$ we then compute a value

$$\hat{\theta} = g(x_1, \cdots, x_n)$$

with a suitable g (whose choice will be a main point of our further discussion; for instance, take $g = (x_1 + \cdots + x_n)/n$ in the case in which θ is the mean). If $\hat{\theta} > c$, we reject the hypothesis. If $\hat{\theta} \leqq c$, we accept it. Here, the value $\hat{\theta}$ can be regarded as an observed value of the random variable

$$\hat{\Theta} = g(X_1, \cdots, X_n) \tag{4}$$

because x_j may be regarded as an observed value of X_j, $j = 1, \cdots, n$. In this test there are two possibilities of making an error, as follows.

Type I Error (see Table 25.4). The hypothesis is true but is rejected (hence the alternative is accepted) because Θ assumes a value $\hat{\theta} > c$. Obviously, the probability of making such an error equals

$$P(\hat{\Theta} > c)_{\theta=\theta_0} = \alpha. \tag{5}$$

α is called the **significance level** of the test, as mentioned before.

Type II Error (see Table 25.4). The hypothesis is false but is accepted because $\hat{\Theta}$ assumes a value $\hat{\theta} \leqq c$. The probability of making such an error is denoted by β; thus

$$P(\hat{\Theta} \leqq c)_{\theta=\theta_1} = \beta. \tag{6}$$

$\eta = 1 - \beta$ is called the **power** of the test. Obviously, the power η is the probability of avoiding a Type II error.

Table 25.4 Type I and Type II Errors in Testing a Hypothesis $\theta = \theta_0$ Against an Alternative $\theta = \theta_1$

		Unknown Truth	
		$\theta = \theta_0$	$\theta = \theta_1$
Accepted	$\theta = \theta_0$	True decision $P = 1 - \alpha$	Type II error $P = \beta$
	$\theta = \theta_1$	Type I error $P = \alpha$	True decision $P = 1 - \beta$

Formulas (5) and (6) show that both α and β depend on c, and we would like to choose c so that these probabilities of making errors are as small as possible. But the important Figure 533 shows that these are conflicting requirements because to let α decrease we must shift c to the right, but then β increases. In practice we first choose α (5%, sometimes 1%), then determine c, and finally compute β. If β is large so that the power $\eta = 1 - \beta$ is small, we should repeat the test, choosing a larger sample, for reasons that will appear shortly.

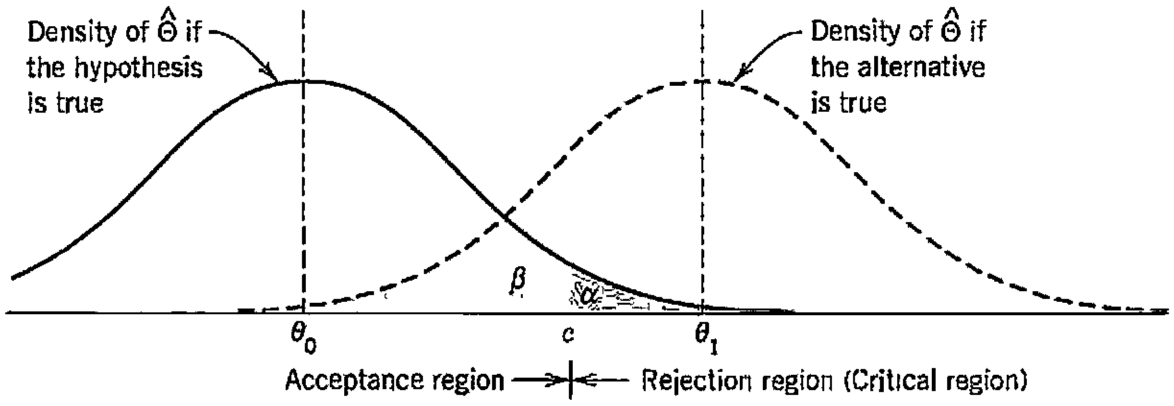


Fig. 533. Illustration of Type I and II errors in testing a hypothesis $\theta = \theta_0$ against an alternative $\theta = \theta_1$ ($> \theta_0$, right-sided test)

If the alternative is not a single number but is of the form (1)–(3), then β becomes a function of θ. This function $\beta(\theta)$ is called the **operating characteristic** (OC) of the test and its curve the **OC curve**. Clearly, in this case $\eta = 1 - \beta$ also depends on θ. This function $\eta(\theta)$ is called the **power function** of the test. (Examples will follow.)

Of course, from a test that leads to the acceptance of a certain hypothesis θ_0, it does ***not*** follow that this is the only possible hypothesis or the best possible hypothesis. Hence the terms **"not reject"** or **"fail to reject"** are perhaps better than the term **"accept."**

Test for μ of the Normal Distribution with Known σ^2

The following example explains the three kinds of hypotheses.

EXAMPLE 2 Test for the Mean of the Normal Distribution with Known Variance

Let X be a normal random variable with variance $\sigma^2 = 9$. Using a sample of size $n = 10$ with mean $\bar{x}$, test the hypothesis $\mu = \mu_0 = 24$ against the three kinds of alternatives, namely,

$$\text{(a)}\quad \mu > \mu_0 \qquad \text{(b)}\quad \mu < \mu_0 \qquad \text{(c)}\quad \mu \neq \mu_0.$$

Solution. We choose the significance level $\alpha = 0.05$. An estimate of the mean will be obtained from

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n).$$

If the hypothesis is true, $\bar{X}$ is normal with mean $\mu = 24$ and variance $\sigma^2/n = 0.9$, see Theorem 1, Sec. 25.3. Hence we may obtain the critical value c from Table A8 in App. 5.

Case (a). *Right-Sided Test.* We determine c from $P(\bar{X} > c)_{\mu=24} = \alpha = 0.05$, that is,

$$P(\bar{X} \leqq c)_{\mu=24} = \Phi\left(\frac{c - 24}{\sqrt{0.9}}\right) = 1 - \alpha = 0.95.$$

Table A8 in App. 5 gives $(c - 24)/\sqrt{0.9} = 1.645$, and $c = 25.56$, which is greater than μ_0, as in the upper part of Fig. 532. If $\bar{x} \leqq 25.56$, the hypothesis is accepted. If $\bar{x} > 25.56$, it is rejected. The power function of the test is (Fig. 534)

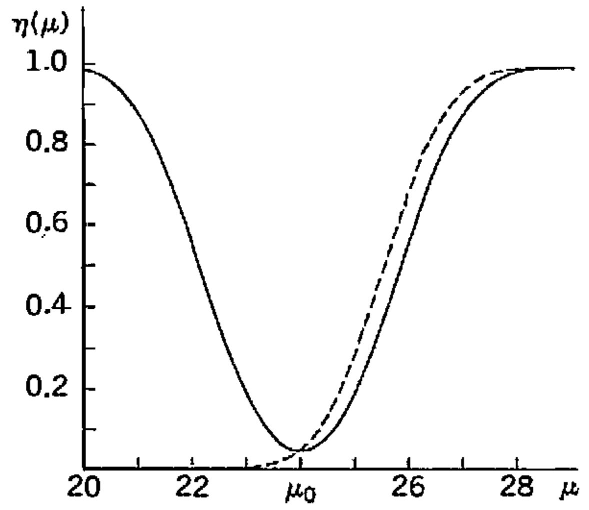


Fig. 534. Power function $\eta(\mu)$ in Example 2, case (a) (dashed) and case (c)

$$
\begin{aligned}
\eta(\mu) &= P(\bar{X} > 25.56)_\mu = 1 - P(\bar{X} \leqq 25.56)_\mu \\
&= 1 - \Phi\left(\frac{25.56 - \mu}{\sqrt{0.9}}\right) = 1 - \Phi(26.94 - 1.05\mu)
\end{aligned} \tag{7}
$$

Case (b). Left-Sided Test. The critical value c is obtained from the equation

$$P(\bar{X} \leqq c)_{\mu=24} = \Phi\left(\frac{c - 24}{\sqrt{0.9}}\right) = \alpha = 0.05.$$

Table A8 in App. 5 yields $c = 24 - 1.56 = 22.44$. If $\bar{x} \geqq 22.44$, we accept the hypothesis. If $\bar{x} < 22.44$, we reject it. The power function of the test is

$$\eta(\mu) = P(\bar{X} \leqq 22.44)_\mu = \Phi\left(\frac{22.44 - \mu}{\sqrt{0.9}}\right) = \Phi(23.65 - 1.05\mu). \tag{8}$$

Case (c). Two-Sided Test. Since the normal distribution is symmetric, we choose c_1 and c_2 equidistant from $\mu = 24$, say, $c_1 = 24 - k$ and $c_2 = 24 + k$, and determine k from

$$P(24 - k \leqq \bar{X} \leqq 24 + k)_{\mu=24} = \Phi\left(\frac{k}{\sqrt{0.9}}\right) - \Phi\left(-\frac{k}{\sqrt{0.9}}\right) = 1 - \alpha = 0.95.$$

Table A8 in App. 5 gives $k/\sqrt{0.9} = 1.960$, hence $k = 1.86$. This gives the values $c_1 = 24 - 1.86 = 22.14$ and $c_2 = 24 + 1.86 = 25.86$. If $\bar{x}$ is not smaller than c_1 and not greater than c_2, we accept the hypothesis. Otherwise we reject it. The power function of the test is (Fig. 534)

$$
\begin{aligned}
\eta(\mu) &= P(\bar{X} < 22.14)_\mu + P(\bar{X} > 25.86)_\mu = P(\bar{X} < 22.14)_\mu + 1 - P(\bar{X} \leqq 25.86)_\mu \\
&= 1 + \Phi\left(\frac{22.14 - \mu}{\sqrt{0.9}}\right) - \Phi\left(\frac{25.86 - \mu}{\sqrt{0.9}}\right) \\
&= 1 + \Phi(23.34 - 1.05\mu) - \Phi(27.26 - 1.05\mu).
\end{aligned} \tag{9}
$$

Consequently, the operating characteristic $\beta(\mu) = 1 - \eta(\mu)$ (see before) is (Fig. 535)

$$\beta(\mu) = \Phi(27.26 - 1.05\mu) - \Phi(23.34 - 1.05\mu).$$

If we take a larger sample, say, of size $n = 100$ (instead of 10), then $\sigma^2/n = 0.09$ (instead of 0.9) and the critical values are $c_1 = 23.41$ and $c_2 = 24.59$, as can be readily verified. Then the operating characteristic of the test is

$$
\begin{aligned}
\beta(\mu) &= \Phi\left(\frac{24.59 - \mu}{\sqrt{0.09}}\right) - \Phi\left(\frac{23.41 - \mu}{\sqrt{0.09}}\right) \\
&= \Phi(81.97 - 3.33\mu) - \Phi(78.03 - 3.33\mu).
\end{aligned}
$$

Figure 535 shows that the corresponding OC curve is steeper than that for $n = 10$. This means that the increase of n has led to an improvement of the test. In any practical case, n is chosen as small as possible but so large that the test brings out deviations between μ and μ_0 that are of practical interest. For instance, if deviations of ± 2 units are of interest, we see from Fig. 535 that $n = 10$ is much too small because when $\mu = 24 - 2 = 22$ or $\mu = 24 + 2 = 26$ β is almost 50%. On the other hand, we see that $n = 100$ is sufficient for that purpose. ■

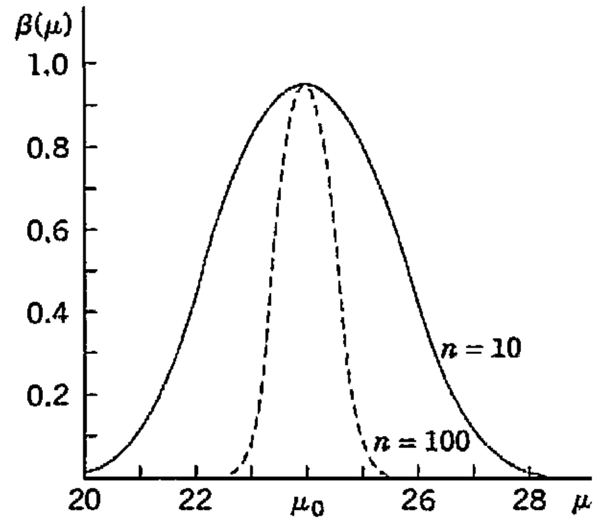


Fig. 535. Curves of the operating characteristic (OC curves) in Example 2, case (c), for two different sample sizes n

Test for μ When σ^2 is Unknown, and for σ^2

EXAMPLE 3 Test for the Mean of the Normal Distribution with Unknown Variance

The tensile strength of a sample of $n = 16$ manila ropes (diameter 3 in.) was measured. The sample mean was $\bar{x} = 4482$ kg, and the sample standard deviation was $s = 115$ kg (N. C. Wiley, 41st Annual Meeting of the American Society for Testing Materials). Assuming that the tensile strength is a normal random variable, test the hypothesis $\mu_0 = 4500$ kg against the alternative $\mu_1 = 4400$ kg. Here μ_0 may be a value given by the manufacturer, while μ_1 may result from previous experience.

Solution. We choose the significance level $\alpha = 5\%$. If the hypothesis is true, it follows from Theorem 2 in Sec. 25.3, that the random variable

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X} - 4500}{S/4}$$

has a t-distribution with $n - 1 = 15$ d.f. The test is left-sided. The critical value c is obtained from $P(T < c)_{\mu_0} = \alpha = 0.05$. Table A9 in App. 5 gives $c = -1.75$. As an observed value of T we obtain from the sample $t = (4482 - 4500)/(115/4) = -0.626$. We see that $t > c$ and accept the hypothesis. For obtaining numeric values of the power of the test, we would need tables called noncentral Student t-tables; we shall not discuss this question here. ■

EXAMPLE 4 Test for the Variance of the Normal Distribution

Using a sample of size $n = 15$ and sample variance $s^2 = 13$ from a normal population, test the hypothesis $\sigma^2 = \sigma_0^{\,2} = 10$ against the alternative $\sigma^2 = \sigma_1^{\,2} = 20$.

Solution. We choose the significance level $\alpha = 5\%$. If the hypothesis is true, then

$$Y = (n - 1)\frac{S^2}{\sigma_0^{\,2}} = 14\,\frac{S^2}{10} = 1.4S^2$$

has a chi-square distribution with $n - 1 = 14$ d.f. by Theorem 3, Sec. 25.3. From

$$P(Y > c) = \alpha = 0.05, \qquad \text{that is,} \qquad P(Y \leqq c) = 0.95,$$

and Table A10 in App. 5 with 14 degrees of freedom we obtain $c = 23.68$. This is the critical value of Y. Hence to $S^2 = \sigma_0{}^2Y/(n-1) = 0.714Y$ there corresponds the critical value $c^* = 0.714 \cdot 23.68 = 16.91$. Since $s^2 < c^*$, we accept the hypothesis.

If the alternative is true, the random variable $Y_1 = 14S^2/\sigma_1{}^2 = 0.7S^2$ has a chi-square distribution with 14 d.f. Hence our test has the power

$$\eta = P(S^2 > c^*)_{\sigma^2=20} = P(Y_1 > 0.7c^*)_{\sigma^2=20} = 1 - P(Y_1 \leqq 11.84)_{\sigma^2=20}.$$

From a more extensive table of the chi-square distribution (e.g. in Ref. [G3] or [G8]) or from your CAS, you see that $\eta \approx 62\%$. Hence the Type II risk is very large, namely, 38%. To make this risk smaller, we would have to increase the sample size. ■

Comparison of Means and Variances

EXAMPLE 5 Comparison of the Means of Two Normal Distributions

Using a sample $x_1, \cdots, x_{n_1}$ from a normal distribution with unknown mean μ_x and a sample $y_1, \cdots, y_{n_2}$ from another normal distribution with unknown mean μ_y, we want to test the hypothesis that the means are equal, $\mu_x = \mu_y$, against an alternative, say, $\mu_x > \mu_y$. The variances need not be known but are assumed to be equal.[3]

Two cases of comparing means are of practical importance:

Case A. *The samples have the same size. Furthermore, each value of the first sample corresponds to precisely one value of the other,* because corresponding values result from the same person or thing (**paired comparison**)—for example, two measurements of the same thing by two different methods or two measurements from the two eyes of the same person. More generally, they may result from pairs of *similar* individuals or things, for example, identical twins, pairs of used front tires from the same car, etc. Then we should form the differences of corresponding values and test the hypothesis that the population corresponding to the differences has mean 0, using the method in Example 3. If we have a choice, this method is better than the following.

Case B. *The two samples are independent and not necessarily of the same size.* Then we may proceed as follows. Suppose that the alternative is $\mu_x > \mu_y$. We choose a significance level α. Then we compute the sample means $\bar{x}$ and $\bar{y}$ as well as $(n_1 - 1)s_x{}^2$ and $(n_2 - 1)s_y{}^2$, where $s_x{}^2$ and $s_y{}^2$ are the sample variances. Using Table A9 in App. 5 with $n_1 + n_2 - 2$ degrees of freedom, we now determine c from

$$P(T \leqq c) = 1 - \alpha. \tag{10}$$

We finally compute

$$t_0 = \sqrt{\frac{n_1 n_2(n_1 + n_2 - 2)}{n_1 + n_2}} \, \frac{\bar{x} - \bar{y}}{\sqrt{(n_1 - 1)s_x{}^2 + (n_2 - 1)s_y{}^2}}. \tag{11}$$

It can be shown that this is an observed value of a random variable that has a t-distribution with $n_1 + n_2 - 2$ degrees of freedom, provided the hypothesis is true. If $t_0 \leqq c$, the hypothesis is accepted. If $t_0 > c$, it is rejected.

If the alternative is $\mu_x \neq \mu_y$, then (10) must be replaced by

$$P(T \leqq c_1) = 0.5\alpha, \qquad P(T \leqq c_2) = 1 - 0.5\alpha. \tag{10*}$$

Note that for samples of equal size $n_1 = n_2 = n$, formula (11) reduces to

$$t_0 = \sqrt{n} \, \frac{\bar{x} - \bar{y}}{\sqrt{s_x{}^2 + s_y{}^2}}. \tag{12}$$

[3] This assumption of equality of variances can be tested, as shown in the next example. If the test shows that they differ significantly, choose two samples of the same size $n_1 = n_2 = n$ (not too small, > 30, say), use the test in Example 2 together with the fact that (12) is an observed value of an approximately standardized normal random variable.

To illustrate the computations, let us consider the two samples $(x_1, \cdots, x_{n_1})$ and $(y_1, \cdots, y_{n_2})$ given by

105 108 86 103 103 107 124 105

and

89 92 84 97 103 107 111 97

showing the relative output of tin plate workers under two different working conditions [J. J. B. Worth, *Journal of Industrial Engineering* **9**, 249–253). Assuming that the corresponding populations are normal and have the same variance, let us test the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x \neq \mu_y$. (Equality of variances will be tested in the next example.)

Solution. We find

$$\bar{x} = 105.125, \qquad \bar{y} = 97.500, \qquad s_x^{\ 2} = 106.125, \qquad s_y^{\ 2} = 84.000.$$

We choose the significance level $\alpha = 5\%$. From (10*) with $0.5\alpha = 2.5\%$, $1 - 0.5\alpha = 97.5\%$ and Table A9 in App. 5 with 14 degrees of freedom we obtain $c_1 = -2.14$ and $c_2 = 2.14$. Formula (12) with $n = 8$ gives the value

$$t_0 = \sqrt{8} \cdot 7.625/\sqrt{190.125} = 1.56.$$

Since $c_1 \leqq t_0 \leqq c_2$, we ***accept the hypothesis*** $\mu_x = \mu_y$ that under both conditions the mean output is the same.

Case A applies to the example because the two first sample values correspond to a certain type of work, the next two were obtained in another kind of work, etc. So we may use the differences

16 16 2 6 0 0 13 8

of corresponding sample values and the method in Example 3 to test the hypothesis $\mu = 0$, where μ is the mean of the population corresponding to the differences. As a logical alternative we take $\mu \neq 0$. The sample mean is $\bar{d} = 7.625$, and the sample variance is $s^2 = 45.696$. Hence

$$t = \sqrt{8}\,(7.625 - 0)/\sqrt{45.696} = 3.19.$$

From $P(T \leqq c_1) = 2.5\%$, $P(T \leqq c_2) = 97.5\%$ and Table A9 in App. 5 with $n - 1 = 7$ degrees of freedom we obtain $c_1 = -2.36$, $c_2 = 2.36$ and ***reject the hypothesis*** because $t = 3.19$ does not lie between c_1 and c_2. Hence our present test, in which we used more information (but the same samples), shows that the difference in output is significant. ■

EXAMPLE 6 Comparison of the Variance of Two Normal Distributions

Using the two samples in the last example, test the hypothesis $\sigma_x^{\ 2} = \sigma_y^{\ 2}$; assume that the corresponding populations are normal and the nature of the experiment suggests the alternative $\sigma_x^{\ 2} > \sigma_y^{\ 2}$.

Solution. We find $s_x^{\ 2} = 106.125$, $s_y^{\ 2} = 84.000$. We choose the significance level $\alpha = 5\%$. Using $P(V \leqq c) = 1 - \alpha = 95\%$ and Table A11 in App. 5, with $(n_1 - 1, n_2 - 1) = (7, 7)$ degrees of freedom, we determine $c = 3.79$. We finally compute $v_0 = s_x^{\ 2}/s_y^{\ 2} = 1.26$. Since $v_0 \leqq c$, we accept the hypothesis. If $v_0 > c$, we would reject it.

This test is justified by the fact that v_0 is an observed value of a random variable that has a so-called ***F*-distribution** with $(n_1 - 1, n_2 - 1)$ degrees of freedom, provided the hypothesis is true. (Proof in Ref. [G3] listed in App. 1.) The F-distribution with (m, n) degrees of freedom was introduced by R. A. Fisher[4] and has the distribution function $F(z) = 0$ if $z < 0$ and

$$F(z) = K_{mn} \int_0^z t^{(m-2)/2}(mt + n)^{-(m+n)/2}\, dt \qquad (z \geqq 0), \tag{13}$$

where $K_{mn} = m^{m/2}n^{n/2}\Gamma(\frac{1}{2}m + \frac{1}{2}n)/\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)$. (For Γ see App. A3.1.) ■

[4]After the pioneering work of the English statistician and biologist, KARL PEARSON (1857–1936), the founder of the English school of statistics, and WILLIAM SEALY GOSSET (1876–1937), who discovered the t-distribution (and published under the name 'Student'), the English statistician Sir RONALD AYLMER FISHER (1890–1962), professor of eugenics in London (1933–1943) and professor of genetics in Cambridge, England (1943–1957) and Adelaide, Australia (1957–1962), had great influence on the further development of modern statistics.

This long section contained the basic ideas and concepts of testing, along with typical applications and you may perhaps want to review it quickly before going on, because the next sections concern an adaption of these ideas to tasks of great practical importance and resulting tests in connection with quality control, acceptance (or rejection) of goods produced, and so on.

PROBLEM SET 25.4

1. Test $\mu = 0$ against $\mu > 0$, assuming normality and using the sample 1, −1, 1, 3, −8, 6, 0 (deviations of the azimuth [multiples of 0.01 radian] in some revolution of a satellite). Choose $\alpha = 5\%$.
2. In one of his classical experiments Buffon obtained 2048 heads in tossing a coin 4040 times. Was the coin fair?
3. Do the same test as in Prob. 2, using a result by K. Pearson, who obtained 6 019 heads in 12 000 trials.
4. Assuming normality and known variance $\sigma^2 = 4$, test the hypothesis $\mu = 30.0$ against the alternative (a) $\mu = 28.5$, (b) $\mu = 30.7$, using a sample of size 10 with mean $\bar{x} = 28.5$ and choosing $\alpha = 5\%$.
5. How does the result in Prob. 4(a) change if we use a smaller sample, say, of size 4, the other data ($\bar{x} = 28.5$, $\alpha = 5\%$, etc.) remaining as before?
6. Detemine the power of the test in Prob. 4(a).
7. What is the rejection region in Prob. 4 in the case of a two-sided test with $\alpha = 5\%$?
8. Using the sample 0.80, 0.81, 0.81, 0.82, 0.81, 0.82, 0.80, 0.82, 0.81, 0.81 (length of nails in inches), test the hypothesis $\mu = 0.80$ in. (the length indicated on the box) against the alternative $\mu \neq 0.80$ in. (Assume normality, choose $\alpha = 5\%$.)
9. A firm sells oil in cans containing 1000 g oil per can and is interested to know whether the mean weight differs significantly from 1000 g at the 5% level, in which case the filling machine has to be adjusted. Set up a hypothesis and an alternative and perform the test, assuming normality and using a sample of 20 fillings with mean 996 g and standard deviation 5 g.
10. If a sample of 50 tires of a certain kind has a mean life of 32 000 mi and a standard deviation of 4000 mi, can the manufacturer claim that the true mean life of such tires is greater than 30 000 mi? Set up and test a corresponding hypothesis at a 5% level, assuming normality.
11. If simultaneous measurements of electric voltage by two different types of voltmeter yield the differences (in volts) 0.8, 0.2, −0.3, 0.1, 0.0, 0.5, 0.7, 0.2, can we assert at the 5% level that there is no significant difference in the calibration of the two types of instruments? (Assume normality.)
12. If a standard medication cures about 70% of patients with a certain disease and a new medication cured 148 of the first 200 patients on whom it was tried, can we conclude that the new medication is better? (Choose $\alpha = 5\%$.)
13. Suppose that in the past the standard deviation of weights of certain 25.0-oz packages filled by a machine was 0.4 oz. Test the hypothesis H_0: $\sigma = 0.4$ against the alternative H_1: $\sigma > 0.4$ (an undesirable increase), using a sample of 10 packages with standard deviation 0.5 oz and assuming normality. (Choose $\alpha = 5\%$.)
14. Suppose that in operating battery-powered electrical equipment, it is less expensive to replace all batteries at fixed intervals than to replace each battery individually when it breaks down, provided the standard deviation of the lifetime is less than a certain limit, say, less than 5 hours. Set up and apply a suitable test, using a sample of 28 values of lifetimes with standard deviation $s = 3.5$ hours and assuming normality; choose $\alpha = 5\%$.
15. Brand A gasoline was used in 9 automobiles of the same model under identical conditions. The corresponding sample of 9 values (miles per gallon) had mean 20.2 and standard deviation 0.5. Under the same conditions, high-power brand B gasoline gave a sample of 10 values with mean 21.8 and standard deviation 0.6. Is the mileage of B significantly better than that of A? (Test at the 5% level; assume normality.)
16. The two samples 70, 80, 30, 70, 60, 80 and 140, 120, 130, 120, 120, 130, 120 are values of the differences of temperatures (°C) of iron at two stages of casting, taken from two different crucibles. Is the variance of the first population larger than that of the second? (Assume normality. Choose $\alpha = 5\%$.)
17. Using samples of sizes 10 and 16 with variances $s_x^2 = 50$ and $s_y^2 = 30$ and assuming normality of the corresponding populations, test the hypothesis H_0: $\sigma_x^2 = \sigma_y^2$ against the alternative $\sigma_x^2 > \sigma_y^2$. Choose $\alpha = 5\%$.
18. Assuming normality and equal variance and using independent samples with $n_1 = 9$, $\bar{x} = 12$, $s_x = 2$, $n_2 = 9$, $\bar{y} = 15$, $s_y = 2$, test H_0: $\mu_x = \mu_y$ against $\mu_x \neq \mu_y$; choose $\alpha = 5\%$.

Figure 540 on p. 1075 shows an example. Since AOQ(0) = 0 and $P(A; 1) = 0$, the AOQ curve has a maximum at some $\theta = \theta^*$, giving the **average outgoing quality limit** (AOQL). This is the worst average quality that may be expected to be accepted under rectification.

PROBLEM SET 25.6

1. Lots of knives are inspected by a sampling plan that uses a sample of size 20 and the acceptance number $c = 1$. What are probabilities of accepting a lot with 1%, 2%, 10% defectives (dull blades)? Use Table A6 in App. 5. Graph the OC curve.
2. What happens in Prob. 1 if the sample size is increased to 50? First guess. Then calculate. Graph the OC curve and compare.
3. How will the probabilities in Prob. 1 with $n = 20$ change (up or down) if we decrease c to zero? First guess.
4. What are the producer's and consumer's risks in Prob. 1 if the AQL is 1.5% and the RQL is 7.5%?
5. Large lots of batteries are inspected according to the following plan. $n = 30$ batteries are randomly drawn from a lot and tested. If this sample contains at most $c = 1$ defective battery, the lot is accepted. Otherwise it is rejected. Graph the OC curve of the plan, using the Poisson approximation.
6. Graph the AOQ curve in Prob. 5. Determine the AOQL, assuming that rectification is applied.
7. Do the work required in Prob. 5 if $n = 50$ and $c = 0$.
8. Find the binomial approximation of the hypergeometric distribution in Example 1 and compare the approximate and the accurate values.
9. In Example 1, what are the producer's and consumer's risks if the AQL is 0.1 and the RQL is 0.6?
10. Calculate $P(A; \theta)$ in Example 1 if the sample size is increased from $n = 2$ to $n = 3$, the other data remaining as before. Compute $P(A; 0.10)$ and $P(A; 0.20)$ and compare with Example 1.
11. Samples of 5 screws are drawn from a lot with fraction defective θ. The lot is accepted if the sample contains (a) no defective screws, (b) at most 1 defective screw. Using the binomial distribution, find, graph, and compare the OC curves.
12. Find the risks in the single sampling plan with $n = 5$ and $c = 0$, assuming that the AQL is $\theta_0 = 1\%$ and the RQL is $\theta_1 = 15\%$.
13. Why is it impossible for an OC curve to have a vertical portion separating good from poor quality?
14. If in a single sampling plan for large lots of spark plugs, the sample size is 100 and we want the AQL to be 5% and the producer's risk 2%, what acceptance number c should we choose? (Use the normal approximation.)
15. What is the consumer's risk in Prob. 14 if we want the RQL to be 12%?
16. Graph and compare sampling plans with $c = 1$ and increasing values of n, say, $n = 2, 3, 4$. (Use the binomial distribution.)
17. Samples of 3 fuses are drawn from lots and a lot is accepted if in the corresponding sample we find no more than 1 defective fuse. Criticize this sampling plan. In particular, find the probability of accepting a lot that is 50% defective. (Use the binomial distribution.)
18. Graph the OC curve and the AOQ curve for the single sampling plan for large lots with $n = 5$ and $c = 0$, and find the AOQL.

25.7 Goodness of Fit. χ^2-Test

To test for **goodness of fit** means that we wish to test that a certain function $F(x)$ is the distribution function of a distribution from which we have a sample $x_1, \cdots, x_n$. Then we test whether the **sample distribution function** $\tilde{F}(x)$ defined by

$$\tilde{F}(x) = \textit{Sum of the relative frequencies of all sample values } x_j \textit{ not exceeding } x$$

fits $F(x)$ "sufficiently well." If this is so, we shall accept the hypothesis that $F(x)$ is the distribution function of the population; if not, we shall reject the hypothesis.

This test is of considerable practical importance, and it differs in character from the tests for parameters (μ, σ^2, etc.) considered so far.

To test in that fashion, we have to know how much $\tilde{F}(x)$ can differ from $F(x)$ if the hypothesis is true. Hence we must first introduce a quantity that measures the deviation of $\tilde{F}(x)$ from $F(x)$, and we must know the probability distribution of this quantity under the assumption that the hypothesis is true. Then we proceed as follows. We determine a number c such that if the hypothesis is true, a deviation greater than c has a small preassigned probability. If, nevertheless, a deviation greater than c occurs, we have reason to doubt that the hypothesis is true and we reject it. On the other hand, if the deviation does not exceed c, so that $\tilde{F}(x)$ approximates $F(x)$ sufficiently well, we accept the hypothesis. Of course, if we accept the hypothesis, this means that we have insufficient evidence to reject it, and this does not exclude the possibility that there are other functions that would not be rejected in the test. In this respect the situation is quite similar to that in Sec. 25.4.

Table 25.7 shows a test of that type, which was introduced by R. A. Fisher. This test is justified by the fact that if the hypothesis is true, then $\chi_0^{\,2}$ is an observed value of a random variable whose distribution function approaches that of the chi-square distribution with $K - 1$ degrees of freedom (or $K - r - 1$ degrees of freedom if r parameters are estimated) as n approaches infinity. The requirement that at least five sample values lie in each interval in Table 25.7 results from the fact that for finite n that random variable has only *approximately* a chi-square distribution. A proof can be found in Ref. [G3] listed in App. 1. If the sample is so small that the requirement cannot be satisfied, one may continue with the test, but then use the result with caution.

Table 25.7 Chi-square Test for the Hypothesis That $F(x)$ is the Distribution Function of a Population from Which a Sample $x_1, \cdots, x_n$ is Taken

Step 1. Subdivide the x-axis into K intervals $I_1, I_2, \cdots, I_K$ such that each interval contains at least 5 values of the given sample $x_1, \cdots, x_n$. Determine the number b_j of sample values in the interval I_j, where $j = 1, \cdots, K$. If a sample value lies at a common boundary point of two intervals, add 0.5 to each of the two corresponding b_j.

Step 2. Using $F(x)$, compute the probability p_j that the random variable X under consideration assumes any value in the interval I_j, where $j = 1, \cdots, K$. Compute

$$e_j = np_j.$$

(This is the number of sample values theoretically expected in I_j if the hypothesis is true.)

Step 3. Compute the deviation

$$\chi_0^{\,2} = \sum_{j=1}^{K} \frac{(b_j - e_j)^2}{e_j}. \tag{1}$$

Step 4. Choose a significance level (5%, 1%, or the like).

Step 5. Determine the solution c of the equation

$$P(\chi^2 \leqq c) = 1 - \alpha$$

from the table of the chi-sqare distribution with $K - 1$ degrees of freedom (Table A10 in App. 5). If r parameters of $F(x)$ are unknown and their maximum likelihood estimates (Sec. 25.2) are used, then use $K - r - 1$ degrees of freedom (instead of $K - 1$). If $\chi_0^{\,2} \leqq c$, accept the hypothesis. If $\chi_0^{\,2} > c$, reject the hypothesis.

Table 25.8 Sample of 100 Values of the Splitting Tensile Strength (lb/in.^{2}) of Concrete Cylinders

320	380	340	410	380	340	360	350	320	370
350	340	350	360	370	350	380	370	300	420
370	390	390	440	330	390	330	360	400	370
320	350	360	340	340	350	350	390	380	340
400	360	350	390	400	350	360	340	370	420
420	400	350	370	330	320	390	380	400	370
390	330	360	380	350	330	360	300	360	360
360	390	350	370	370	350	390	370	370	340
370	400	360	350	380	380	360	340	330	370
340	360	390	400	370	410	360	400	340	360

D. L. IVEY, Splitting tensile tests on structural lightweight aggregate concrete. Texas Transportation Institute, College Station, Texas.

EXAMPLE 1 Test of Normality

Test whether the population from which the sample in Table 25.8 was taken is normal.

Solution. Table 25.8 shows the values (column by column) in the order obtained in the experiment. Table 25.9 gives the frequency distribution and Fig. 541 the histogram. It is hard to guess the outcome of the test—does the histogram resemble a normal density curve sufficiently well or not?

The maximum likelihood estimates for μ and σ^2 are $\hat{\mu} = \bar{x} = 364.7$ and $\tilde{\sigma}^2 = 712.9$. The computation in Table 25.10 yields ${\chi_0}^2 = 2.942$. It is very interesting that the interval $375 \cdots 385$ contributes over 50% of ${\chi_0}^2$. From the histogram we see that the corresponding frequency looks much too small. The second largest contribution comes from $395 \cdots 405$, and the histogram shows that the frequency seems somewhat too large, which is perhaps not obvious from inspection.

Table 25.9 Frequency Table of the Sample in Table 25.8

1 Tensile Strength x [lb/in.^{2}]	2 Absolute Frequency	3 Relative Frequency $\tilde{f}(x)$	4 Cumulative Absolute Frequency	5 Cumulative Relative Frequency $\tilde{F}(x)$
300	2	0.02	2	0.02
310	0	0.00	2	0.02
320	4	0.04	6	0.06
330	6	0.06	12	0.12
340	11	0.11	23	0.23
350	14	0.14	37	0.37
360	16	0.16	53	0.53
370	15	0.15	68	0.68
380	8	0.08	76	0.76
390	10	0.10	86	0.86
400	8	0.08	94	0.94
410	2	0.02	96	0.96
420	3	0.03	99	0.99
430	0	0.00	99	0.99
440	1	0.01	100	1.00

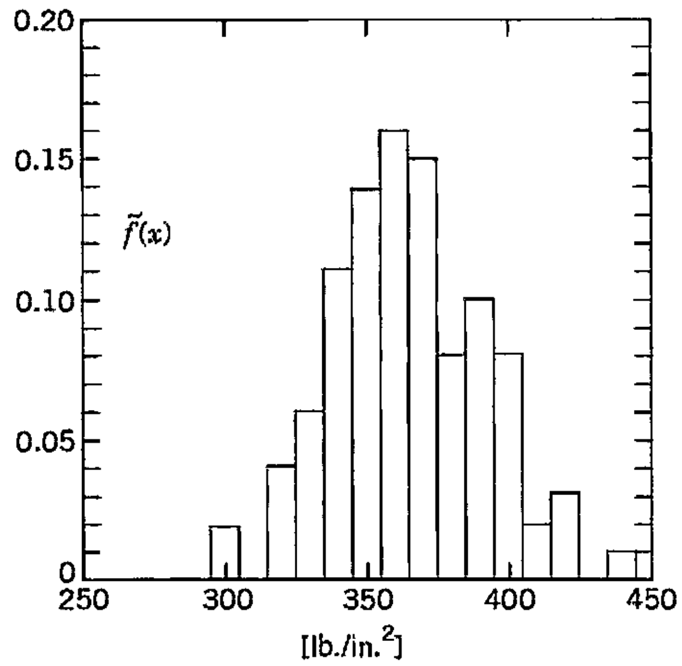


Fig. 541. Frequency histogram of the sample in Table 25.8

We choose $\alpha = 5\%$. Since $K = 10$ and we estimated $r = 2$ parameters we have to use Table A10 in App. 5 with $K - r - 1 = 7$ degrees of freedom. We find $c = 14.07$ as the solution of $P(\chi^2 \leqq c) = 95\%$. Since $\chi_0{}^2 < c$, we accept the hypothesis that the population is normal. ∎

Table 25.10 Computations in Example 1

x_j	$\dfrac{x_j - 364.7}{26.7}$	$\Phi\left(\dfrac{x_j - 364.7}{26.7}\right)$	e_j	b_j	Term in (1)
$-\infty \cdots 325$	$-\infty \cdots -1.49$	$0.0000 \cdots 0.0681$	6.81	6	0.096
$325 \cdots 335$	$-1.49 \cdots -1.11$	$0.0681 \cdots 0.1335$	6.54	6	0.045
$335 \cdots 345$	$-1.11 \cdots -0.74$	$0.1335 \cdots 0.2296$	9.61	11	0.201
$345 \cdots 355$	$-0.74 \cdots -0.36$	$0.2296 \cdots 0.3594$	12.98	14	0.080
$355 \cdots 365$	$-0.36 \cdots 0.01$	$0.3594 \cdots 0.4960$	13.66	16	0.401
$365 \cdots 375$	$0.01 \cdots 0.39$	$0.4960 \cdots 0.6517$	15.57	15	0.021
$375 \cdots 385$	$0.39 \cdots 0.76$	$0.6517 \cdots 0.7764$	12.47	8	1.602
$385 \cdots 395$	$0.76 \cdots 1.13$	$0.7764 \cdots 0.8708$	9.44	10	0.033
$395 \cdots 405$	$1.13 \cdots 1.51$	$0.8708 \cdots 0.9345$	6.37	8	0.417
$405 \cdots \infty$	$1.51 \cdots \infty$	$0.9345 \cdots 1.0000$	6.55	6	0.046

$\chi_0{}^2 = 2.942$

PROBLEM SET 25.7

1. If 100 flips of a coin result in 30 heads and 70 tails, can we assert on the 5% level that the coin is fair?

2. If in 10 flips of a coin we get the same ratio as in Prob. 1 (3 heads and 7 tails), is the conclusion the same as in Prob. 1? First conjecture, then compute.

3. What would be the smallest number of heads in Prob. 1 under which the hypothesis "Fair coin" is still accepted (with $\alpha = 5\%$)?

4. If in rolling a die 180 times we get 39, 22, 41, 26, 20, 32, can we claim on the 5% level that the die is fair?

5. Solve Prob. 4 if the sample is 25, 31, 33, 27, 29, 35.

6. A manufacturer claims that in a process of producing kitchen knives, only 2.5% of the knives are dull. Test the claim against the alternative that more than 2.5% of the knives are dull, using a sample of 400 knives containing 17 dull ones. (Use $\alpha = 5\%$.)

7. Between 1 P.M. and 2 P.M. on five consecutive days (Monday through Friday) a certain service station has 92, 60, 66, 62, and 90 customers, respectively. Test the hypothesis that the expected number of customers during that hour is the same on those days. (Use $\alpha = 5\%$.)

8. Test for normality at the 1% level using a sample of $n = 79$ (rounded) values x (tensile strength [kg/mm^2] of steel sheets of 0.3 mm thickness). $a = a(x)$ = absolute frequency. (Take the first two values together, also the last three, to get $K = 5$.)

x	57	58	59	60	61	62	63	64
a	4	10	17	27	8	9	3	1

9. In a sample of 100 patients having a certain disease 45 are men and 55 women. Does this support the claim that the disease is equally common among men and women? Choose $\alpha = 5\%$.

10. In Prob. 9 find the smallest number (>50) of women that leads to the rejection of the hypothesis on the levels 5%, 1%, 0.5%.

11. Verify the calculations in Example 1 of the text.

12. Does the random variable X = *Number of accidents per week in a certain foundry* have a Poisson distribution if within 50 weeks, 33 were accident-free, 1 accident occurred in 11 of the 50 weeks, 2 in 6 of the weeks and more than 2 accidents in no week? (Choose $\alpha = 5\%$.)

13. Using the given sample, test that the corresponding population has a Poisson distribution. x is the number of alpha particles per 7.5-sec intervals observed by E. Rutherford and H. Geiger in one of their classical experiments in 1910, and $a(x)$ is the absolute frequency (= number of time periods during which exactly x particles were observed). (Use $\alpha = 5\%$.)

x	0	1	2	3	4	5	6
a	57	203	383	525	532	408	273
x	7	8	9	10	11	12	≧13
a	139	45	27	10	4	2	0

14. Can we assert that the traffic on the three lanes of an expressway (in one direction) is about the same on each lane if a count gives 910, 850, 720 cars on the right, middle, and left lanes, respectively, during a particular time interval? (Use $\alpha = 5\%$.)

15. If it is known that 25% of certain steel rods produced by a standard process will break when subjected to a load of 5000 lb, can we claim that a new process yields the same breakage rate if we find that in a sample of 80 rods produced by the new process, 27 rods broke when subjected to that load? (Use $\alpha = 5\%$.)

16. Three samples of 200 rivets each were taken from a large production of each of three machines. The numbers of defective rivets in the samples were 7, 8, and 12. Is this difference significant? (Use $\alpha = 5\%$.)

17. In a table of properly rounded function values, even and odd last decimals should appear about equally often. Test this for the 90 values of $J_1(x)$ in Table A1 in App. 5.

18. Are the 5 tellers in a certain bank equally time-efficient if during the same time interval on a certain day they serve 120, 95, 110, 108, 102 customers? (Use $\alpha = 5\%$.)

19. CAS EXPERIMENT. Random Number Generator. Check your generator experimentally by imitating results of n trials of rolling a fair die, with a convenient n (e.g., 60 or 300 or the like). Do this many times and see whether you can notice any "nonrandomness" features, for example, too few Sixes, too many even numbers, etc., or whether your generator seems to work properly. Design and perform other kinds of checks.

20. TEAM PROJECT. Difficulty with Random Selection. 77 students were asked to choose 3 of the integers 11, 12, 13, ⋯, 30 completely arbitrarily. The amazing result was as follows.

Number	11	12	13	14	15	16	17	18	19	20
Frequ.	11	10	20	8	13	9	21	9	16	8
Number	21	22	23	24	25	26	27	28	29	30
Frequ.	12	8	15	10	10	9	12	8	13	9

If the selection were completely random, the following hypotheses should be true.

(a) The 20 numbers are equally likely.

(b) The 10 even numbers together are as likely as the 10 odd numbers together.

(c) The 6 prime numbers together have probability 0.3 and the 14 other numbers together have probability 0.7. Test these hypotheses, using $\alpha = 5\%$. Design further experiments that illustrate the difficulties of random selection.

25.8 Nonparametric Tests

Nonparametric tests, also called **distribution-free tests,** are valid for any distribution. Hence they are used in cases when the kind of distribution is unknown, or is known but such that no tests specifically designed for it are available. In this section we shall explain the basic idea of these tests, which are based on **"order statistics"** and are rather simple.

If there is a choice, then tests designed for a specific distribution generally give better results than do nonparametric tests. For instance, this applies to the tests in Sec. 25.4 for the normal distribution.

We shall discuss two tests in terms of typical examples. In deriving the distributions used in the test, it is essential that the distributions from which we sample are continuous. (Nonparametric tests can also be derived for discrete distributions, but this is slightly more complicated.)

EXAMPLE 1 Sign Test for the Median

A **median** of the population is a solution $x = \tilde{\mu}$ of the equation $F(x) = 0.5$, where F is the distribution function of the population.

Suppose that eight radio operators were tested, first in rooms without air-conditioning and then in air-conditioned rooms over the same period of time, and the difference of errors (unconditioned minus conditioned) were

$$9 \quad 4 \quad 0 \quad 6 \quad 4 \quad 0 \quad 7 \quad 11.$$

Test the hypothesis $\tilde{\mu} = 0$ (that is, air-conditioning has no effect) against the alternative $\tilde{\mu} > 0$ (that is, inferior performance in unconditioned rooms).

Solution. We choose the significance level $\alpha = 5\%$. If the hypothesis is true, the probability p of a positive difference is the same as that of a negative difference. Hence in this case, $p = 0.5$, and the random variable

$$X = \textit{Number of positive values among n values}$$

has a binomial distribution with $p = 0.5$. Our sample has eight values. We omit the values 0, which do not contribute to the decision. Then six values are left, all of which are positive. Since

$$\begin{aligned} P(X = 6) &= \binom{6}{6}(0.5)^6(0.5)^0 \\ &= 0.0156 \\ &= 1.56\% \end{aligned}$$

we do have observed an event whose probability is very small if the hypothesis is true; in fact $1.56\% < \alpha = 5\%$. Hence we assert that the alternative $\tilde{\mu} > 0$ is true. That is, the number of errors made in unconditioned rooms is significantly higher, so that installation of air conditioning should be considered. ■

EXAMPLE 2 Test for Arbitrary Trend

A certain machine is used for cutting lengths of wire. Five successive pieces had the lengths

$$29 \quad 31 \quad 28 \quad 30 \quad 32.$$

Using this sample, test the hypothesis that there is **no trend,** that is, the machine does not have the tendency to produce longer and longer pieces or shorter and shorter pieces. Assume that the type of machine suggests the alternative that there is *positive trend,* that is, there is the tendency of successive pieces to get longer.

Solution. We count the number of **transpositions** in the sample, that is, the number of times a larger value precedes a smaller value:

29 precedes 28	(1 transposition),
31 precedes 28 and 30	(2 transpositions).

The remaining three sample values follow in ascending order. Hence in the sample there are $1 + 2 = 3$ transpositions. We now consider the random variable

$$T = \textit{Number of transpositions.}$$

If the hypothesis is true (no trend), then each of the $5! = 120$ permutations of five elements 1 2 3 4 5 has the same probability (1/120). We arrange these permutations according to their number of transpositions:

$T = 0$	$T = 1$	$T = 2$	$T = 3$	
1 2 3 4 5	1 2 3 5 4	1 2 4 5 3	1 2 5 4 3	
	1 2 4 3 5	1 2 5 3 4	1 3 4 5 2	
	1 3 2 4 5	1 3 2 5 4	1 3 5 2 4	
	2 1 3 4 5	1 3 4 2 5	1 4 2 5 3	
		1 4 2 3 5	1 4 3 2 5	
		2 1 3 5 4	1 5 2 3 4	
		2 1 4 3 5	2 1 4 5 3	
		2 3 1 4 5	2 1 5 3 4	etc.
		3 1 2 4 5	2 3 1 5 4	
			2 3 4 1 5	
			2 4 1 3 5	
			3 1 2 5 4	
			3 1 4 2 5	
			3 2 1 4 5	
			4 1 2 3 5	

From this we obtain

$$P(T \leqq 3) = \tfrac{1}{120} + \tfrac{4}{120} + \tfrac{9}{120} + \tfrac{15}{120} = \tfrac{29}{120} = 24\%.$$

We accept the hypothesis because we have observed an event that has a relatively large probability (certainly much more than 5%) if the hypothesis is true.

Values of the distribution function of T in the case of no trend are shown in Table A12, App. 5. For instance, if $n = 3$, then $F(0) = 0.167$, $F(1) = 0.500$, $F(2) = 1 - 0.167$. If $n = 4$, then $F(0) = 0.042$, $F(1) = 0.167$, $F(2) = 0.375$, $F(3) = 1 - 0.375$, $F(4) = 1 - 0.167$. and so on.

Our method and those values refer to *continuous* distributions. Theoretically, we may then expect that all the values of a sample are different. Practically, some sample values may still be equal, because of rounding: If m values are equal, add $m(m-1)/4$ (= mean value of the transpositions in the case of the permutations of m elements), that is, $\frac{1}{2}$ for each pair of equal values, $\frac{3}{2}$ for each triple, etc. ■

PROBLEM SET 25.8

1. What would change in Example 1, had we observed only 5 positive values? Only 4?

2. Does a process of producing plastic pipes of length $\mu = 2$ meters need adjustment if in a sample, 4 pipes have the exact length and 15 are shorter and 3 longer than 2 meters? (Use the normal approximation of the binomial distribution.)

3. Do the computations in Prob. 2 without the use of the DeMoivre–Laplace limit theorem (in Sec. 24.8).

4. Test whether a thermostatic switch is properly set to 20°C against the alternative that its setting is too low. Use a sample of 9 values, 8 of which are less than 20°C and 1 is greater than 20°C.

5. Are air filters of type A better than type B filters if in 10 trials, A gave cleaner air than B in 7 cases, B gave cleaner air than A in 1 case, whereas in 2 of the trials the results for A and B were practically the same?

6. In a clinical experiment, each of 10 patients were given two different sedatives A and B. The following table shows the effect (increase of sleeping time, measured in hours). Using the sign test, find out whether the difference is significant.

A	1.9	0.8	1.1	0.1	−0.1	4.4	5.5	1.6	4.6	3.4
B	0.7	−1.6	−0.2	−1.2	−0.1	3.4	3.7	0.8	0.0	2.0
Difference	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

7. Assuming that the populations corresponding to the samples in Prob. 6 are normal, apply a suitable test for the normal distribution.

8. Thirty new employees were grouped into 15 pairs of similar intelligence and experience and were then instructed in data processing by an old method (A) applied to one (randomly selected) person of each pair, and by a new presumably better method (B) applied to

the other person of each pair. Test for equality of methods against the alternative that (B) is better than (A), using the following scores obtained after the end of the training period.

A	60	70	80	85	75	40	70	45	95	80	90	60	80	75	65
B	65	85	85	80	95	65	100	60	90	85	100	75	90	60	80

9. Assuming normality, solve Prob. 8 by a suitable test from Sec. 25.4.

10. Set up a sign test for the lower quartile q_{25} (defined by the condition $F(q_{25}) = 0.25$).

11. How would you proceed in the sign test if the hypothesis is $\tilde{\mu} = \tilde{\mu}_0$ (any number) instead of $\tilde{\mu} = 0$?

12. Check the table in Example 2 of the text.

13. Apply the test in Example 2 to the following data (x = disulfide content of a certain type of wool, measured in percent of the content in unreduced fibers; y = saturation water content of the wool, measured in percent). Test for no trend against negative trend.

x	10	15	30	40	50	55	80	100
y	50	46	43	42	36	39	37	33

14. Test the hypothesis that for a certain type of voltmeter, readings are independent of temperature T [°C] against the alternative that they tend to increase with T. Use a sample of values obtained by applying a constant voltage:

Temperature T [°C]	10	20	30	40	50
Reading V [volts]	99.5	101.1	100.4	100.8	101.6

15. In a swine-feeding experiment, the following gains in weight [kg] of 10 animals (ordered according to increasing amounts of food given per day) were recorded:

20 17 19 18 23 16 25 28 24 22.

Test for no trend against positive trend.

16. Apply the test explained in Example 2 to the following data (x = diastolic blood pressure [mm Hg], y = weight of heart [in grams] of 10 patients who died of cerebral hemorrhage).

x	121	120	95	123	140	112	92	100	102	91
y	521	465	352	455	490	388	301	395	375	418

17. Does an increase in temperature cause an increase of the yield of a chemical reaction from which the following sample was taken?

Temperature [°C]	10	20	30	40	60	80
Yield [kg/min]	0.6	1.1	0.9	1.6	1.2	2.0

18. Does the amount of fertilizer increase the yield of wheat X [kg/plot]? Use a sample of values ordered according to increasing amounts of fertilizer:

41.4 43.3 39.6 43.0 44.1 45.6 44.5 46.7.

25.9 Regression. Fitting Straight Lines. Correlation

So far we were concerned with random experiments in which we observed a single quantity (random variable) and got samples whose values were single numbers. In this section we discuss experiments in which we observe or measure two quantities simultaneously, so that we get samples of *pairs* of values $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. Most applications involve one of two kinds of experiments, as follows.

1. In **regression analysis** one of the two variables, call it x, can be regarded as an ordinary variable because we can measure it without substantial error or we can even give it values we want. x is called the **independent variable,** or sometimes the **controlled variable** because we can control it (set it at values we choose). The other variable, Y, is a random variable, and we are interested in the dependence of Y on x. Typical examples are the dependence of the blood pressure Y on the age x of a person or, as we shall now say, the regression of Y on x, the regression of the gain of weight Y of certain animals on the daily ration of food x, the regression of the heat conductivity Y of cork on the specific weight x of the cork, etc.

2. In **correlation analysis** both quantities are random variables and we are interested in relations between them. Examples are the relation (one says "correlation") between wear X and wear Y of the front tires of cars, between grades X and Y of students in mathematics and in physics, respectively, between the hardness X of steel plates in the center and the hardness Y near the edges of the plates, etc.

Regression Analysis

In regression analysis the dependence of Y on x is a dependence of the mean μ of Y on x, so that $\mu = \mu(x)$ is a function in the ordinary sense. The curve of $\mu(x)$ is called the **regression curve** of Y on x.

In this section we discuss the simplest case, namely, that of a straight **regression line**

$$\mu(x) = \kappa_0 + \kappa_1 x. \tag{1}$$

Then we may want to graph the sample values as n points in the xY-plane, fit a straight line through them, and use it for estimating $\mu(x)$ at values of x that interest us, so that we know what values of Y we can expect for those x. Fitting that line by eye would not be good because it would be subjective; that is, different persons' results would come out differently, particularly if the points are scattered. So we need a mathematical method that gives a unique result depending only on the n points. A widely used procedure is the method of least squares by Gauss and Legendre. For our task we may formulate it as follows.

Least Squares Principle

The straight line should be fitted through the given points so that the sum of the squares of the distances of those points from the straight line is minimum, where the distance is measured in the vertical direction (the y-direction). (Formulas below.)

To get uniqueness of the straight line, we need some extra condition. To see this, take the sample $(0, 1)$, $(0, -1)$. Then all the lines $y = k_1 x$ with any k_1 satisfy the principle. (Can you see it?) The following assumption will imply uniqueness, as we shall find out.

General Assumption (A1)

The x-values $x_1, \cdots, x_n$ in our sample $(x_1, y_1), \cdots, (x_n, y_n)$ are not all equal.

From a given sample $(x_1, y_1), \cdots, (x_n, y_n)$ we shall now determine a straight line by least squares. We write the line as

$$y = k_0 + k_1 x \tag{2}$$

and call it the **sample regression line** because it will be the counterpart of the population regression line (1).

Now a sample point (x_j, y_j) has the vertical distance (distance measured in the y-direction) from (2) given by

$$|y_j - (k_0 + k_1 x_j)| \qquad \text{(see Fig. 542).}$$

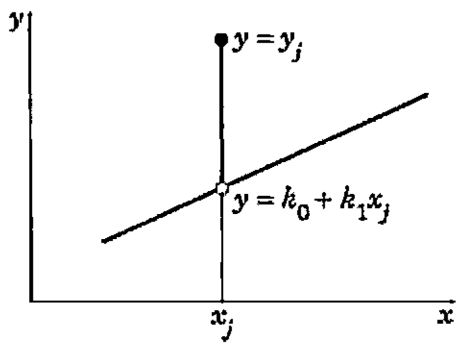


Fig. 542. Vertical distance of a point (x_j, y_j) from a straight line $y = k_0 + k_1 x$

Hence the sum of the squares of these distances is

$$q = \sum_{j=1}^{n} (y_j - k_0 - k_1 x_j)^2. \tag{3}$$

In the method of least squares we now have to determine k_0 and k_1 such that q is minimum. From calculus we know that a necessary condition for this is

$$\frac{\partial q}{\partial k_0} = 0 \quad \text{and} \quad \frac{\partial q}{\partial k_1} = 0. \tag{4}$$

We shall see that from this condition we obtain for the sample regression line the formula

$$y - \bar{y} = k_1(x - \bar{x}). \tag{5}$$

Here $\bar{x}$ and $\bar{y}$ are the means of the x- and the y-values in our sample, that is,

$$\begin{aligned} &\text{(a)} \quad \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) \\ &\text{(b)} \quad \bar{y} = \frac{1}{n}(y_1 + \cdots + y_n). \end{aligned} \tag{6}$$

The slope k_1 in (5) is called the **regression coefficient** of the sample and is given by

$$k_1 = \frac{s_{xy}}{s_x^{\,2}}. \tag{7}$$

Here the **"sample covariance"** s_{xy} is

$$s_{xy} = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y}) = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j y_j - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j\right)\right] \tag{8}$$

and $s_x^{\,2}$ is given by

$$s_x^{\,2} = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2\right]. \tag{9a}$$

From (5) we see that the sample regression line passes through the point $(\bar{x}, \bar{y})$, by which it is determined, together with the regression coefficient (7). We may call s_x^2 the *variance* of the x-values, but we should keep in mind that x is an ordinary variable, not a random variable.

We shall soon also need

$$\text{(9b)} \qquad s_y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n} y_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} y_j\right)^2\right].$$

Derivation of (5) and (7). Differentiating (3) and using (4), we first obtain

$$\frac{\partial q}{\partial k_0} = -2\sum (y_j - k_0 - k_1 x_j) = 0,$$

$$\frac{\partial q}{\partial k_1} = -2\sum x_j(y_j - k_0 - k_1 x_j) = 0$$

where we sum over j from 1 to n. We now divide by 2, write each of the two sums as three sums, and take the sums containing y_j and $x_j y_j$ over to the right. Then we get the **"normal equations"**

$$\text{(10)} \qquad \begin{aligned} k_0 n + k_1 \sum x_j \quad &= \sum y_j \\ k_0 \sum x_j + k_1 \sum x_j^2 &= \sum x_j y_j. \end{aligned}$$

This is a linear system of two equations in the two unknowns k_0 and k_1. Its coefficient determinant is [see (9)]

$$\begin{vmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{vmatrix} = n\sum x_j^2 - \left(\sum x_j\right)^2 = n(n-1)s_x^2 = n\sum (x_j - \bar{x})^2$$

and is not zero because of Assumption (A1). Hence the system has a unique solution. Dividing the first equation of (10) by n and using (6), we get $k_0 = \bar{y} - k_1\bar{x}$. Together with $y = k_0 + k_1 x$ in (2) this gives (5). To get (7), we solve the system (10) by Cramer's rule (Sec. 7.6) or elimination, finding

$$\text{(11)} \qquad k_1 = \frac{n\sum x_j y_j - \sum x_i \sum y_j}{n(n-1)s_x^2}.$$

This gives (7)–(9) and completes the derivation. [The equality of the two expressions in (8) and in (9) may be shown by the student; see Prob. 14]. ■

EXAMPLE 1 Regression Line

The decrease of volume y [%] of leather for certain fixed values of high pressure x [atmospheres] was measured. The results are shown in the first two columns of Table 25.11. Find the regression line of y on x.

Solution. We see that $n = 4$ and obtain the values $\bar{x} = 28\,000/4 = 7000$, $\bar{y} = 19.0/4 = 4.75$, and from (9) and (8)

Table 25.11 Regression of the Decrease of Volume y [%] of Leather on the Pressure x [Atmospheres]

Given Values		Auxiliary Values	
x_j	y_j	x_j^2	$x_j y_j$
4 000	2.3	16 000 000	9 200
6 000	4.1	36 000 000	24 600
8 000	5.7	64 000 000	45 600
10 000	6.9	100 000 000	69 000
28 000	19.0	216 000 000	148 400

$$s_x^2 = \frac{1}{3}\left(216\,000\,000 - \frac{28\,000^2}{4}\right) = \frac{20\,000\,000}{3}$$

$$s_{xy} = \frac{1}{3}\left(148\,400 - \frac{28\,000 \cdot 19}{4}\right) = \frac{15\,400}{3}.$$

Hence $k_1 = 15\,400/20\,000\,000 = 0.000\,77$ from (7), and the regression line is

$$y - 4.75 = 0.000\,77(x - 7000) \qquad \text{or} \qquad y = 0.000\,77x - 0.64.$$

Note that $y(0) = -0.64$, which is physically meaningless, but typically indicates that a linear relation is merely an approximation valid on some restricted interval. ■

Confidence Intervals in Regression Analysis

If we want to get confidence intervals, we have to make assumptions about the distribution of Y (which we have not made so far; least squares is a "geometric principle," nowhere involving probabilities!). We assume normality and independence in sampling:

Assumption (A2)

For each fixed x the random variable Y is normal with mean (1), *that is,*

$$\mu(x) = \kappa_0 + \kappa_1 x \tag{12}$$

and variance σ^2 independent of x.

Assumption (A3)

The n performances of the experiment by which we obtain a sample

$$(x_1, y_1), \qquad (x_2, y_2), \qquad \cdots, \qquad (x_n, y_n)$$

are independent.

κ_1 in (12) is called the **regression coefficient** of the population because it can be shown that under Assumptions (A1)–(A3) the maximum likelihood estimate of κ_1 is the sample regression coefficient k_1 given by (11).

Under Assumptions (A1)–(A3) we may now obtain a confidence interval for κ_1, as shown in Table 25.12.

Table 25.12 Determination of a Confidence Interval for κ_1 in (1) under Assumptions (A1)–(A3)

Step 1. Choose a confidence level γ(95%, 99%, or the like).

Step 2. Determine the solution c of the equation

$$F(c) = \tfrac{1}{2}(1 + \gamma) \tag{13}$$

from the table of the t-distribution with $n - 2$ degrees of freedom (Table A9 in App. 5; n = sample size).

Step 3. Using a sample $(x_1, y_1), \cdots, (x_n, y_n)$, compute $(n - 1)s_x^2$ from (9a), $(n - 1)s_{xy}$ from (8), k_1 from (7),

$$(n - 1)s_y^2 = \sum_{j=1}^{n} y_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} y_j\right)^2 \tag{14}$$

[as in (9b)], and

$$q_0 = (n - 1)(s_y^2 - k_1^2 s_x^2). \tag{15}$$

Step 4. Compute

$$K = c\sqrt{\frac{q_0}{(n - 2)(n - 1)s_x^2}}\,.$$

The confidence interval is

$$\mathrm{CONF}_{\gamma}\,\{k_1 - K \leqq \kappa_1 \leqq k_1 + K\}. \tag{16}$$

EXAMPLE 2 Confidence Interval for the Regression Coefficient

Using the sample in Table 25.11, determine a confidence interval for κ_1 by the method in Table 25.12.

Solution. *Step 1.* We choose $\gamma = 0.95$.

Step 2. Equation (13) takes the form $F(c) = 0.975$, and Table A9 in App. 5 with $n - 2 = 2$ degrees of freedom gives $c = 4.30$.

Step 3. From Example 1 we have $3s_x^2 = 20\,000\,000$ and $k_1 = 0.00077$. From Table 25.11 we compute

$$\begin{aligned} 3s_y^2 &= 102.2 - \frac{19^2}{4} \\ &= 11.95, \\ q_0 &= 11.95 - 20\,000\,000 \cdot 0.00077^2 \\ &= 0.092. \end{aligned}$$

Step 4. We thus obtain

$$\begin{aligned} K &= 4.30\sqrt{0.092/(2 \cdot 20\,000\,000)} \\ &= 0.000\,206 \end{aligned}$$

and

$$\mathrm{CONF}_{0.95}\,\{0.00056 \leqq \kappa_1 \leqq 0.00098\}.$$ ■

Correlation Analysis

We shall now give an introduction to the basic facts in correlation analysis; for proofs see Ref. [G2] or [G8] in App. 1.

Correlation analysis is concerned with the relation between X and Y in a two-dimensional random variable (X, Y) (Sec. 24.9). A sample consists of n ordered pairs of values $(x_1, y_1), \cdots, (x_n, y_n)$, as before. The interrelation between the x and y values in the sample is measured by the sample covariance s_{xy} in (8) or by the sample **correlation coefficient**

$$r = \frac{s_{xy}}{s_x s_y} \tag{17}$$

with s_x and s_y given in (9). Here r has the advantage that it does not change under a multiplication of the x and y values by a factor (in going from feet to inches, etc.).

THEOREM 1

Sample Correlation Coefficient

The sample correlation coefficient r satisfies $-1 \leqq r \leqq 1$. In particular, $r = \pm 1$ if and only if the sample values lie on a straight line. (See Fig. 543.)

The theoretical counterpart of r is the **correlation coefficient** ρ of X and Y,

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{18}$$

where $\mu_X = E(X)$, $\mu_Y = E(Y)$, $\sigma_X{}^2 = E([X - \mu_X]^2)$, $\sigma_Y{}^2 = E([Y - \mu_Y]^2)$ (the means and variances of the marginal distributions of X and Y; see Sec. 24.9), and σ_{XY} is the

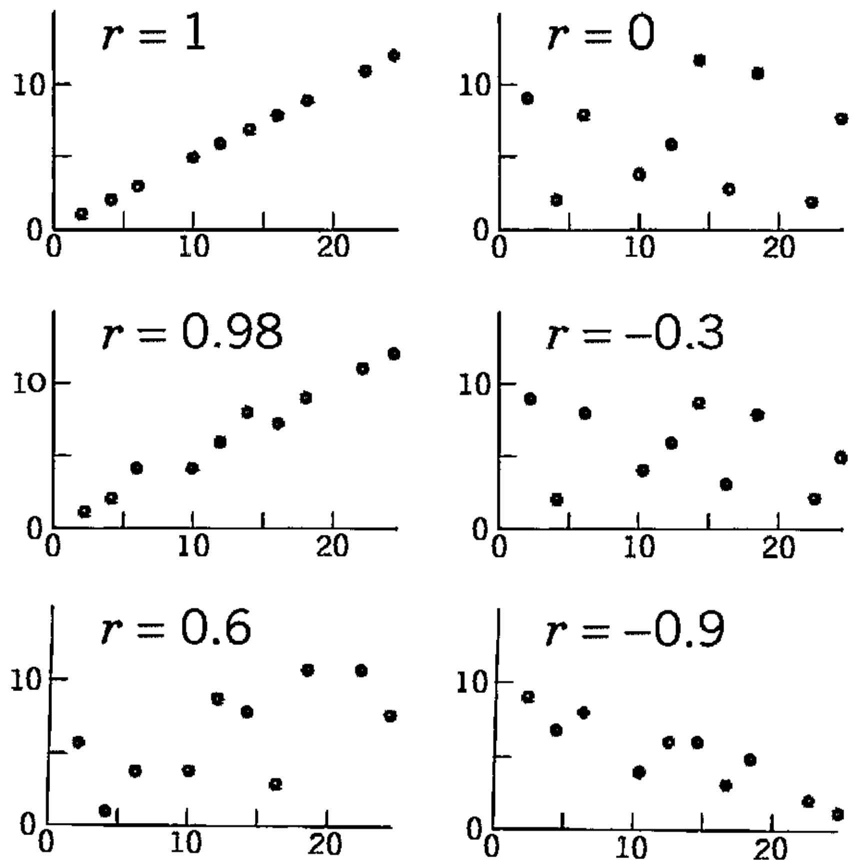


Fig. 543. Samples with various values of the correlation coefficient r

covariance of X and Y given by (see Sec. 24.9)

$$\sigma_{XY} = E([X - \mu_X][Y - \mu_Y]) = E(XY) - E(X)E(Y). \tag{19}$$

The analog of Theorem 1 is

THEOREM 2 **Correlation Coefficient**

The correlation coefficient ρ satisfies $-1 \leqq \rho \leqq 1$. In particular, $\rho = \pm 1$ if and only if X and Y are **linearly related,** *that is, $Y = \gamma X + \delta$, $X = \gamma^* Y + \delta^*$.*

X and Y are called **uncorrelated** if $\rho = 0$.

THEOREM 3 **Independence. Normal Distribution**

(a) *Independent X and Y* (see Sec. 24.9) *are uncorrelated.*

(b) *If (X, Y) is normal* (see below), *then uncorrelated X and Y are independent.*

Here the two-dimensional normal distribution can be introduced by taking two independent standardized normal random variables X^*, Y^*, whose joint distribution thus has the density

$$f^*(x^*, y^*) = \frac{1}{2\pi} e^{-(x^{*2}+y^{*2})/2} \tag{20}$$

(representing a surface of revolution over the x^*y^*-plane with a bell-shaped curve as cross section) and setting

$$X = \mu_X + \sigma_X X^*$$
$$Y = \mu_Y + \rho\sigma_Y X^* + \sqrt{1-\rho^2}\, \sigma_Y Y^*.$$

This gives the general **two-dimensional normal distribution** with the density

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-h(x,y)/2} \tag{21a}$$

where

$$h(x, y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]. \tag{21b}$$

In Theorem 3(b), normality is important, as we can see from the following example.

EXAMPLE 3 **Uncorrelated but Dependent Random Variables**

If X assumes -1, 0, 1 with probability 1/3 and $Y = X^2$, then $E(X) = 0$ and in (3)

$$\sigma_{XY} = E(XY) = E(X^3) = (-1)^3 \cdot \frac{1}{3} + 0^3 \cdot \frac{1}{3} + 1^3 \cdot \frac{1}{3} = 0,$$

so that $\rho = 0$ and X and Y are uncorrelated. But they are certainly not independent since they are even functionally related. ■

Test for the Correlation Coefficient ρ

Table 25.13 shows a test for ρ in the case of the two-dimensional normal distribution. t is an observed value of a random variable that has a t-distribution with $n - 2$ degrees of freedom. This was shown by R. A. Fisher (*Biometrika* **10** (1915), 507–521).

Table 25.13 Test of the Hypothesis $\rho = 0$ Against the Alternative $\rho > 0$ in the Case of the Two-Dimensional Normal Distribution

Step 1. Choose a significance level α (5%, 1%, or the like).

Step 2. Determine the solution c of the equation

$$P(T \leqq c) = 1 - \alpha$$

from the t-distribution (Table A9 in App. 5) with $n - 2$ degrees of freedom.

Step 3. Compute r from (17), using a sample $(x_1, y_1), \cdots, (x_n, y_n)$.

Step 4. Compute

$$t = r\left(\sqrt{\frac{n-2}{1-r^2}}\right).$$

If $t \leqq c$, accept the hypothesis. If $t > c$, reject the hypothesis.

EXAMPLE 4 Test for the Correlation Coefficient ρ

Test the hypothesis $\rho = 0$ (independence of X and Y, because of Theorem 3) against the alternative $\rho > 0$, using the data in the lower left corner of Fig. 543, where $r = 0.6$ (manual soldering errors on 10 two-sided circuit boards done by 10 workers; x = front, y = back of the boards).

Solution. We choose $\alpha = 5\%$; thus $1 - \alpha = 95\%$. Since $n = 10$, $n - 2 = 8$, the table gives $c = 1.86$. Also, $t = 0.6\sqrt{8/0.64} = 2.12 > c$. We reject the hypothesis and assert that there is a **positive correlation.** A worker making few (many) errors on the front side also tends to make few (many) errors on the reverse side of the board. ∎

PROBLEM SET 25.9

1–10 SAMPLE REGRESSION LINE

Find and sketch or graph the sample regression line of y and x and the given data as points on the same axes.

1. $(-1, 1)$, $(0, 1.7)$, $(1, 3)$
2. $(3, 3.5)$, $(5, 2)$, $(7, 4.5)$, $(9, 3)$
3. $(2, 12)$, $(5, 24)$, $(9, 33)$, $(14, 50)$
4. $(11, 22)$, $(15, 18)$, $(17, 16)$, $(20, 9)$, $(22, 10)$
5.

Speed x [mph] of a car	30	40	50	60
Stopping distance y [ft]	150	195	240	295

Also find the stopping distance at 35 mph.

6. x = Deformation of a certain steel [mm], y = Brinell hardness [kg/mm^2]

x	6	9	11	13	22	26	28	33	35
y	68	67	65	53	44	40	37	34	32

7. x = Revolutions per minute, y = Power of a Diesel engine [hp]

x	400	500	600	700	750
y	580	1030	1420	1880	2100

8.

Humidity of air x [%]	10	20	30	40
Expansion of gelatin y [%]	0.8	1.6	2.3	2.8

9.

Voltage x [V]	40	40	80	80	110	110
Current y [A]	5.1	4.8	10.0	10.3	13.0	12.7

Also find the resistance R [Ω] by **Ohms' law** (Sec. 2.9].

10.

Force x [lb]	2	4	6	8
Extension y [in] of a spring	4.1	7.8	12.3	15.8

Also find the spring modulus by **Hooke's law** (Sec. 2.4).

11–13 CONFIDENCE INTERVALS

Find a 95% confidence interval for the regression coefficient κ_1, assuming that (A2) and (A3) hold and using the sample:

11. In Prob. 6

12. In Prob. 7

13. In Prob. 8

14. Derive the second expression for s_x^2 in (9a) from the first one.

15. CAS EXPERIMENT. Moving Data. Take a sample, for instance, that in Prob. 6, and investigate and graph the effect of changing y-values (a) for small x, (b) for large x, (c) in the middle of the sample.

CHAPTER 25 REVIEW QUESTIONS AND PROBLEMS

1. What is a sample? Why do we take samples?

2. What is the role of probability theory in statistics?

3. Will you get better results by taking larger samples? Explain.

4. Do several samples from a certain population have the same mean? The same variance?

5. What is a parameter? How can we estimate it? Give an example.

6. What is a statistical test? What errors occur in testing?

7. How do we test in quality control?

8. What is the χ^2-test? Give a simple example from memory.

9. What are nonparametric tests? When would you apply them?

10. In what tests did we use the t-distribution? The χ^2-distribution?

11. What are one-sided and two-sided tests? Give typical examples.

12. List some areas of application of statistical tests.

13. What do we mean by "goodness of fit"?

14. Acceptance sampling uses principles of testing. Explain.

15. What is the power of a test? What can you do if the power is low?

16. Explain the idea of a maximum likelihood estimate from memory.

17. How does the length of a confidence interval depend on the sample size? On the confidence level?

18. Couldn't we make the error in interval estimation zero simply by choosing the confidence level 1?

19. What is the least squares principle? Give applications.

20. What is the difference between regression and correlation analysis?

21. Find the maximum likelihood estimates of mean and variance of a normal distribution using the sample 5, 4, 6, 5, 3, 5, 7, 4, 6, 5, 8, 6.

22. Determine a 95% confidence interval for the mean μ of a normal population with variance $\sigma^2 = 16$, using a sample of size 400 with mean 53.

23. What will happen to the length of the interval in Prob. 22 if we reduce the sample size to 100?

24. Determine a 99% confidence interval for the mean of a normal population with standard deviation 2.2, using the sample 28, 24. 31, 27, 22.

25. What confidence interval do we obtain in Prob. 24 if we assume the variance to be unknown?

26. Assuming normality, find a 95% confidence interval for the variance from the sample 145.3, 145.1, 145.4, 146.2.

27–29 Find a 95% confidence interval for the mean μ, assuming normality and using the sample:

27. Nitrogen content [%] of steel 0.74, 0.75, 0.73, 0.75, 0.74, 0.72

28. Diameters of 10 gaskets with mean 4.37 cm and standard deviation 0.157 cm

29. Density [g/cm³] of coke 1.40, 1.45, 1.39, 1.44, 1.38

30. What sample size should we use in Prob. 28 if we want to obtain a confidence interval of length 0.1, assuming that the standard deviation of the samples is (about) the same?

31–32 Find a 99% confidence interval for the variance σ^2, assuming normality and using the sample:

31. Rockwell hardness of tool bits 64.9, 64.1, 63.8, 64.0

32. A sample of size $n = 128$ with variance $s^2 = 1.921$

33. Using a sample of 10 values with mean 14.5 from a normal population with variance $\sigma^2 = 0.25$, test the hypothesis $\mu_0 = 15.0$ against the alternative $\mu_1 = 14.4$ on the 5% level.

34. In Prob. 33, change the alternative to $\mu \neq 15.0$ and test as before.

35. Find the power in Prob. 33.

36. Using a sample of 15 values with mean 36.2 and variance 0.9, test the hypothesis $\mu_0 = 35.0$ against the alternative $\mu_1 = 37.0$, assuming normality and taking $\alpha = 1\%$.

37. Using a sample of 20 values with variance 8.25 from a normal population. test the hyothesis $\sigma_0^2 = 5.0$ against the alternative $\sigma_1^2 = 8.1$, choosing $\alpha = 5\%$.

38. A firm sells paint in cans containing 1 kg of paint per can and is interested to know whether the mean weight differs significantly from 1 kg, in which case the filling machine must be adjusted. Set up a hypothesis and an alternative and perform the test, assuming normality and using a sample of 20 fillings having a mean of 991 g and a standard deviation of 8 g. (Choose $\alpha = 5\%$.)

39. Using samples of sizes 10 and 5 with variances $s_x^2 = 50$ and $s_y^2 = 20$ and assuming normality of the corresponding populations, test the hypothesis H_0: $\sigma_x^2 = \sigma_y^2$ against the alternative $\sigma_x^2 > \sigma_y^2$. Choose $\alpha = 5\%$.

40. Assume the thickness X of washers to be normal with mean 2.75 mm and variance 0.00024 mm^2. Set up a control chart for μ, choosing $\alpha = 1\%$, and graph the means of the five samples (2.74, 2.76), (2.74, 2.74), (2.79, 2.81), (2.78, 2.76), (2.71, 2.75) on the chart.

41. What effect on UCL − LCL in a control chart for the mean does it have if we double the sample size? If we switch from $\alpha = 1\%$ to $\alpha = 5\%$?

42. The following samples of screws (length in inches) were taken from an ongoing production. Assuming that the population is normal with mean 3.500 and variance 0.0004, set up a control chart for the mean, choosing $\alpha = 1\%$, and graph the sample means on the chart.

Sample No.	1	2	3	4	5	6	7	8
Length	3.49	3.48	3.52	3.50	3.51	3.49	3.52	3.53
	3.50	3.47	3.49	3.51	3.48	3.50	3.50	3.49

43. A purchaser checks gaskets by a single sampling plan that uses a sample size of 40 and an acceptance number of 1. Use Table A6 in App. 5 to compute the probability of acceptance of lots containing the following percentages of defective gaskets $\frac{1}{4}\%$, $\frac{1}{2}\%$, 1%, 2%, 5%, 10%. Graph the OC curve. (Use the Poisson approximation.)

44. Does an automatic cutter have the tendency of cutting longer and longer pieces of wire if the lengths of subsequent pieces [in.] were 10.1, 9.8, 9.9, 10.2, 10.6, 10.5?

45. Find the least squares regression line to the data (−2, 1), (0, 1), (2, 3), (4, 4), (6, 5).

SUMMARY OF CHAPTER 25
Mathematical Statistics

We recall from Chap. 24 that with an experiment in which we observe some quantity (number of defectives, height of persons, etc.) there is associated a random variable X whose probability distribution is given by a distribution function

$$F(x) = P(X \leqq x) \qquad \text{(Sec. 24.5)} \tag{1}$$

which for each x gives the probability that X assumes any value not exceeding x.

In statistics we take random samples $x_1, \cdots, x_n$ of size n by performing that experiment n times (Sec. 25.1) and draw conclusions from properties of samples about properties of the distribution of the corresponding X. We do this by calculating *point estimates* or *confidence intervals* or by performing a *test* for **parameters** (μ and σ^2 in the normal distribution, p in the binomial distribution, etc.) or by a test for distribution functions.

A **point estimate** (Sec. 25.2) is an approximate value for a parameter in the distribution of X obtained from a sample. Notably, the **sample mean** (Sec. 25.1)

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{1}{n}(x_1 + \cdots + x_n) \tag{2}$$

is an estimate of the mean μ of X, and the **sample variance** (Sec. 25.1)

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2 = \frac{1}{n-1}[(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2] \tag{3}$$

is an estimate of the variance σ^2 of X. Point estimation can be done by the basic **maximum likelihood method** (Sec. 25.2).

Confidence intervals (Sec. 25.3) are intervals $\theta_1 \leqq \theta \leqq \theta_2$ with endpoints calculated from a sample such that with a high probability γ we obtain an interval that contains the unknown true value of the parameter θ in the distribution of X. Here, γ is chosen at the beginning, usually 95% or 99%. We denote such an interval by $\text{CONF}_\gamma\{\theta_1 \leqq \theta \leqq \theta_2\}$.

In a **test** for a parameter we test a *hypothesis* $\theta = \theta_0$ against an *alternative* $\theta = \theta_1$ and then, on the basis of a sample, accept the hypothesis, or we reject it in favor of the alternative (Sec. 25.4). Like any conclusion about X from samples, this may involve errors leading to a false decision. There is a small probability α (which we can choose, 5% or 1%, for instance) that we reject a true hypothesis, and there is a probability β (which we can compute and decrease by taking larger samples) that we accept a false hypothesis. α is called the **significance level** and $1 - \beta$ the **power** of the test. Among many other engineering applications, testing is used in **quality control** (Sec. 25.5) and **acceptance sampling** (Sec. 25.6).

If not merely a parameter but the kind of distribution of X is unknown, we can use the **chi-square test** (Sec. 25.7) for testing the hypothesis that some function $F(x)$ is the unknown distribution function of X. This is done by determining the discrepancy between $F(x)$ and the distribution function $\tilde{F}(x)$ of a given sample.

"Distribution-free" or **nonparametric tests** are tests that apply to any distribution, since they are based on combinatorial ideas. These tests are usually very simple. Two of them are discussed in Sec. 25.8.

The last section deals with samples of **pairs of values,** which arise in an experiment when we simultaneously observe two quantities. In **regression analysis,** one of the quantities, x, is an ordinary variable and the other, Y, is a random variable whose mean μ depends on x, say, $\mu(x) = \kappa_0 + \kappa_1 x$. In **correlation analysis** the relation between X and Y in a two-dimensional random variable (X, Y) is investigated, notably in terms of the **correlation coefficient** ρ.