Probability and Statistics II Lecture Notes for Computer Science Students

Department of Mathematics, Faculty of Science, South Valley University

First release, September 2024

Contents

Chapter 1

Discrete and Continuous Distributions

This chapter introduces discrete distributions and continuous distributions.

1.1 Discrete Distributions

Next, we introduce the most commonly used discrete distributions.

1.1.1 Binomial Distribution

Definition 1.1 *A variable described as the number of successes in a sequence of independent Bernoulli trials has **Binomial distribution**. Its parameters are n, the number of trials, and p, the probability of success.*

Remark: Binomial probability mass function is

$$P(x) = P\{X = x\} = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n \tag{1.1}$$

which is the probability of exactly x successes in n trials. In this formula, p^x is the probability of x successes, probabilities being multiplied due to independence of trials. Also, q^{n-x} is the probability of the remaining $(n-x)$ trials being failures. Finally, $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the number of elements of the sample space Ω that form the event

$\{X = x\}$. This is the number of possible orderings of x successes and $(n - x)$ failures among n trials, and it is computed as $C(n, x)$

$$\begin{array}{c|lcl|}
\cline{2-4}
 & n & = & \text{number of trials} \\
 & p & = & \text{probability of success} \\
\text{Binomial} & P(x) & = & \binom{n}{x} p^x q^{n-x} \\
\text{Distribution} & \mathbf{E}(X) & = & np \\
 & \mathrm{Var}(X) & = & npq \\
\cline{2-4}
\end{array}$$

Example 1.1 *An exciting computer game is released. Sixty percent of players complete all the levels. Thirty percent of them will then buy an advanced version of the game. Among 15 users, what is the expected number of people who will buy the advanced version? What is the probability that at least two people will buy it?*

Solution. Let X be the number of people (successes), among the mentioned 15 users (trials), who will buy the advanced version of the game. It has Binomial distribution with $n = 15$ trials and the probability of success

$$\begin{aligned}
p &= P\{\text{ buy advanced }\} \\
&= P\{\text{ buy advanced } \mid \text{ complete all levels }\}P\{\text{ complete all levels }\} \\
&= (0.30)(0.60) = 0.18
\end{aligned}$$

Then we have

$$\mathbf{E}(X) = np = (15)(0.18) = \underline{2.7}$$

and

$$P\{X \geq 2\} = 1 - P(0) - P(1) = 1 - (1-p)^n - np(1-p)^{n-1} = \underline{0.7813}.$$

The last probability was computed directly by formula (1.1).

1.1.2 Poisson distribution

Definition 1.2 *The number of rare events occurring within a fixed period of time has Poisson distribution.*

Poisson Distribution	λ	$=$ frequency, average number of events
	$p(x)$	$= e^{-\lambda}\dfrac{\lambda^x}{x!}\ ,\quad x{=}0, 1, 2, \cdots$
	$\mathbf{E}(X)$	$= \lambda$
	$\mathrm{Var}(X)$	$= \lambda$

Example 1.2 *(NEW ACCOUNTS). Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day.*

(a) *What is the probability that more than 8 new accounts will be initiated today?*

(b) *What is the probability that more than 16 accounts will be initiated within 2 days?*

Solution. **(a)** New account initiations qualify as rare events because no two customers open accounts simultaneously. Then the number X of today's new accounts has Poisson distribution with parameter

$\lambda = 10$. From Table A3,

$$P\{X > 8\} = 1 - F_X(8) = 1 - 0.333 = \underline{0.667}.$$

(b) The number of accounts, Y, opened within 2 days does not equal $2X$. Rather, Y is another Poisson random variable whose parameter equals 20. Indeed, the parameter is the average number of rare events, which, over the period of two days, doubles the one-day average. Using Table A3 with $\lambda = 20$,

$$P\{Y > 16\} = 1 - F_Y(16) = 1 - 0.221 = \underline{0.779}.$$

1.2 Continuous Distributions

As in the discrete case, varieties of phenomena can be described by relatively few families of continuous distributions. Here, we shall discuss Exponential Normal distributions.

1.2.1 Exponential Distribution

Exponential distribution has density

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0.$$

$$\begin{aligned}
F(x) &= \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t}dt = 1 - e^{-\lambda x} \quad (x > 0), \\
\mathbf{E}(X) &= \int tf(t)dt = \int_0^\infty t\lambda e^{-\lambda t}dt = \frac{1}{\lambda}, \\
\mathrm{Var}(X) &= \int t^2 f(t)dt - \mathbf{E}^2(X) \\
&= \int_0^\infty t^2\lambda e^{-\lambda t}dt - \left(\frac{1}{\lambda}\right)^2 \\
&= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.
\end{aligned}$$

The quantity λ is a parameter of Exponential distribution, and its meaning is clear from $\mathbf{E}(X) = 1/\lambda$. This λ has the same meaning as the parameter of Poisson distribution.Then we can compute the cdf of T as

$$F_T(t) = 1 - e^{-\lambda t}, \tag{1.2}$$

Example 1.3 *Jobs are sent to a printer at an average rate of 3 jobs per hour.*

(a) *What is the expected time between jobs?*

(b) *What is the probability that the next job is sent within 5 minutes?*

Solution. Job arrivals represent rare events, thus the time T between them is Exponential with the given parameter $\lambda = 3\text{hrs}^{-1}$ (jobs per hour).

(a) $\mathbf{E}(T) = 1/\lambda = 1/3$ hours or 20 minutes between jobs;

(b) Convert to the same measurement unit: 5 min = (1/12)hrs. Then,

$$\boldsymbol{P}\{T < 1/12\text{hrs}\} = F(1/12) = 1 - e^{-\lambda(1/12)} = 1 - e^{-1/4} = \underline{0.2212}.$$

Exponential Distribution	λ	= frequency parameter, the number of events per time unit
	$p(x)$	$= \lambda e^{-\lambda x}\ ,\ \ x > 0$
	$\mathbf{E}(X)$	$= \frac{1}{\lambda}$
	$\mathrm{Var}(X)$	$= \frac{1}{\lambda^2}$

1.2.2 Normal distribution

Normal distribution plays a vital role in Probability and Statistics, mostly because of the Central Limit Theorem, according to which sums and averages often have approximately Normal distribution. Due to this fact, various fluctuations and measurement errors that consist of accumulated number of small terms appear normally distributed.

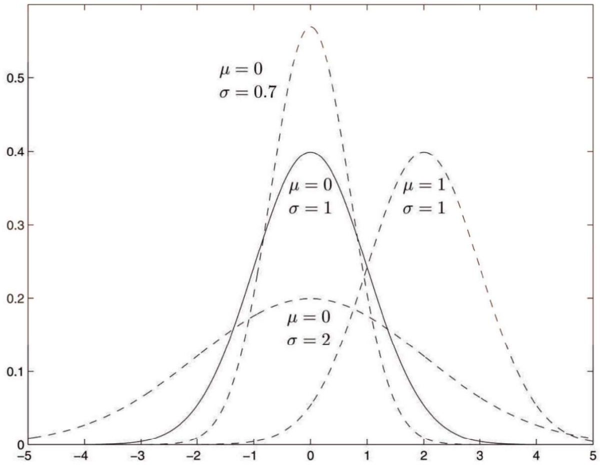


FIGURE 4.6: Normal densities with different location and scale parameters.

Normal distribution has a density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < +\infty$$

where parameters μ and σ have a simple meaning of the expectation $\mathbf{E}(X)$ and the standard deviation $\text{Std}(X)$. This density is known as the bell-shaped curve, symmetric and centered at μ, its spread being controlled by σ. As seen in Figure 4.6, changing μ shifts the curve to the left or to the right without affecting its shape, while changing σ makes it more concentrated or more flat. Often μ and σ are called location and scale parameters.

Normal Distribution	μ	$=$ expectation, location parameter
	σ	$=$ standard deviation, scale parameter
	$f(x)$	$= \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left\{\dfrac{-(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty$
	$\mathbf{E}(X)$	$= \mu$
	$\text{Var}(X)$	$= \sigma^2$

Standard Normal distribution

Definition 1.3 *Normal distribution with ”standard parameters” $\mu = 0$ and $\sigma = 1$ is called Standard Normal distribution.*

NOTATION:

$Z =$ Standard Normal random variable

$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, Standard Normal pdf

$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$, Standard Normal cdf

Example 1.4 (Computing Standard Normal Probabilities.)
For a Standard Normal random variable Z,

$$\begin{aligned}
\boldsymbol{P}\{Z < 1.35\} &= \Phi(1.35) = 0.9115 \\
\boldsymbol{P}\{Z > 1.35\} &= 1 - \Phi(1.35) = 0.0885 \\
\boldsymbol{P}\{-0.77 < Z < 1.35\} &= \Phi(1.35) - \Phi(-0.77) \\
&= 0.9115 - 0.2206 = 0.6909
\end{aligned}$$

According to Table A4. Notice that $\boldsymbol{P}\{Z < -1.35\} = 0.0885 = \boldsymbol{P}\{Z > 1.35\}$, which is explained by the symmetry of the Standard Normal density in Figure 4.6. Due to this symmetry, ”the left tail,” or the area to the left of (-1.35) equals ”the right tail,” or the area to the right of 1.35 .

In fact, the symmetry of the Normal density, mentioned in this example, allows to obtain the first part of Table A4, directly from the second part,

$$\Phi(-z) = 1 - \Phi(z) \quad \text{for } -\infty < z < +\infty$$

To compute probabilities about an arbitrary Normal random variable

X, we have to standardize it first, as in (4.16), then use Table A4.

Example 1.5 (Computing non-standard normal probabilities.) *Suppose that the average household income in some country is 900 coins, and the standard deviation is 200 coins. Assuming the Normal distribution of incomes, compute the proportion of "the middle class," whose income is between 600 and 1200 coins.*

Solution. *Standardize and use Table A4. For a Normal*($\mu = 900, \sigma = 200$) *variable* X *,*

$$\boldsymbol{P}\{600 < X < 1200\} = \boldsymbol{P}\left\{\frac{600-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{1200-\mu}{\sigma}\right\}$$
$$= \boldsymbol{P}\left\{\frac{600-900}{200} < Z < \frac{1200-900}{200}\right\} = \boldsymbol{P}\{-1.5 < Z < 1.5\}$$
$$= \Phi(1.5) - \Phi(-1.5) = 0.9332 - 0.0668 = \underline{0.8664}$$

Chapter 2

Introduction to Sampling Distributions

The sampling distribution of a statistic is the distribution of all possible values taken by the statistic when all possible samples of a fixed size n are taken from the population. It is a theoretical idea—we do not actually build it. The sampling distribution of a statistic is the probability distribution of that statistic.

Suppose you randomly sampled 10 people from the population of women in a city, between the ages of 21 and 35 years and computed the mean height of your sample. You would not expect your sample mean to be equal to the mean of all women in the city. It might be somewhat lower or it might be somewhat higher, but it would not equal the population mean exactly. Similarly, if you took a second sample of 10 people from the same population, you would not expect the mean of this second sample to equal the mean of the first sample.

Recall that inferential statistics concern generalizing from a sample to a population. A critical part of inferential statistics involves determining how far sample statistics are likely to vary from each other and from the population parameter. (In this example, the sample statistics are the sample means and the population parameter is the population mean.) As the later portions of this chapter show, these determinations

Outcome	Ball 1	Ball 2	Mean
1	1	1	1.0
2	1	2	1.5
3	1	3	2.0
4	2	1	1.5
5	2	2	2.0
6	3	2	2.5
7	3	1	2.0
8	2	3	2.5
9	3	3	3.0

Table 1. All possible outcomes when two balls are sampled with replacement.

are based on sampling distributions.

Discrete Distributions

We will illustrate the concept of sampling distributions with a simple example. Figure 1 shows three pool balls, each with a number on it. Suppose two of the balls are selected randomly (with replacement) and the average of their numbers is computed. All possible outcomes are shown below in Table 1.

Figure 1. The pool balls.

Notice that all the means are either $1.0, 1.5, 2.0, 2.5,$ or 3.0 . The frequencies of these means are shown in Table 2. The relative frequencies are equal to the frequencies divided by nine because there are nine possible outcomes.

Mean	Frequency	Relative Frequency
1.0	1	0.111
1.5	2	0.222
2.0	3	0.333
2.5	2	0.222
3.0	1	0.111

Table 2. Frequencies of means for N = 2.

Figure 2 shows a relative frequency distribution of the means based on Table 2. This distribution is also a probability distribution since the Y-axis is the probability of obtaining a given mean from a sample of two balls in addition to being the relative frequency.

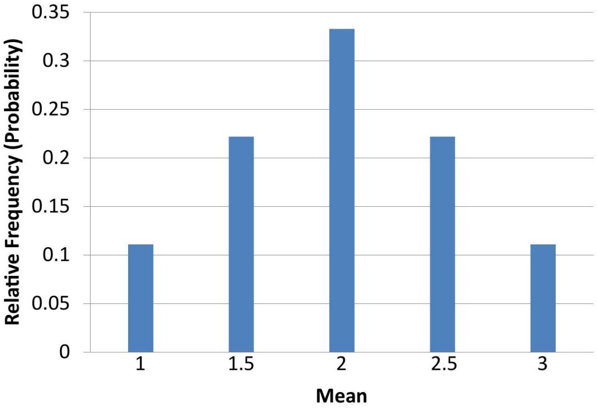


Figure 2. Distribution of means for N = 2.

The distribution shown in Figure 2 is called the sampling distribution of the mean. Specifically, it is the sampling distribution of the mean for a sample size of 2(N = 2). For this simple example, the distribution of pool balls and the sampling distribution are both discrete distributions.

The pool balls have only the values 1,2 , and 3 , and a sample mean can have one of only five values shown in Table 2.

There is an alternative way of conceptualizing a sampling distribution that will be useful for more complex distributions. Imagine that two balls are sampled (with replacement) and the mean of the two balls is computed and recorded. Then this process is repeated for a second sample, a third sample, and eventually thousands of samples. After thousands of samples are taken and the mean computed for each, a relative frequency distribution is drawn. The more samples, the closer the relative frequency distribution will come to the sampling distribution shown in Figure 2. As the number of samples approaches infinity, the relative frequency distribution will approach the sampling distribution. This means that you
can conceive of a sampling distribution as being a relative frequency distribution based on a very large number of samples. To be strictly correct, the relative frequency distribution approaches the sampling distribution as the number of samples approaches infinity.

It is important to keep in mind that every statistic, not just the mean, has a sampling distribution. For example, Table 3 shows all possible outcomes for the range of two numbers (larger number minus the smaller number). Table 4 shows the frequencies for each of the possible ranges and Figure 3 shows the sampling distribution of the range.

Outcome	Ball 1	Ball 2	Range
1	1	1	0
2	1	2	1
3	1	3	2
4	2	1	1
5	2	2	0
6	3	3	1
7	3	2	0
8	3	3	0
9	2	1	

Table 3. All possible outcomes when two balls are sampled with replacement.

Range	Frequency	Relative Frequency
0	3	0.333
1	4	0.444
2	2	0.222

Table 4. Frequencies of ranges for N = 2.

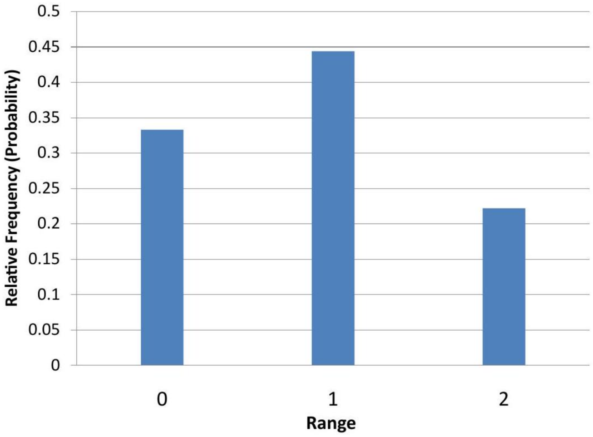


Figure 3. Distribution of ranges for N = 2.

It is also important to keep in mind that there is a sampling distribution for various sample sizes. For simplicity, we have been using N = 2.

The sampling distribution of the range for N = 3 is shown in Figure 4.

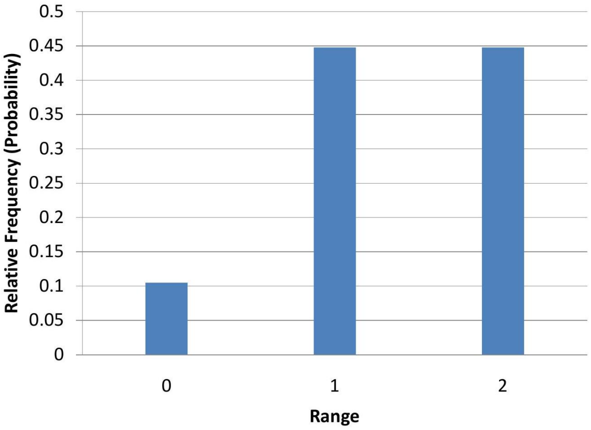


Figure 4. Distribution of ranges for N = 3.

Continuous Distributions

In the previous section, the population consisted of three pool balls. Now we will consider sampling distributions when the population distribution is continuous. What if we had a thousand pool balls with numbers ranging from 0.001 to 1.000 in equal steps? (Although this distribution is not really continuous, it is close enough to be considered continuous for practical purposes.) As before, we are interested in the distribution of means we would get if we sampled two balls and computed the mean of these two balls. In the previous example, we started by computing the mean for each of the nine possible outcomes. This would get a bit tedious for this example since there are $1,000,000$ possible outcomes (1,000 for the first ball x 1,000 for the second). Therefore, it is more convenient to use our second conceptualization of sampling distributions which conceives of sampling distributions in terms of relative frequency distributions. Specifically, the relative frequency distribution that would occur if samples of two balls were repeatedly taken and the mean of each sample computed.

When we have a truly continuous distribution, it is not only impractical but actually impossible to enumerate all possible outcomes. Moreover, in continuous
distributions, the probability of obtaining any single value is zero. Therefore, as discussed in the section ”Distributions” in Chapter 1, these values are called probability densities rather than probabilities.

2.1 Sampling Distributions and Inferential Statistics

As we stated in the beginning of this chapter, sampling distributions are important for inferential statistics. In the examples given so far, a population was specified and the sampling distribution of the mean and the range were determined. In practice, the process proceeds the other way: you collect sample data, and from these data you estimate parameters of the sampling distribution. This knowledge of the sampling distribution can be very useful. For example, knowing the degree to which means from different samples would differ from each other and from the population mean would give you a sense of how close your particular sample mean is likely to be to the population mean. Fortunately, this information is directly available from a sampling distribution. The most common measure of how much sample means differ from each other is the standard deviation of the sampling distribution of the mean. This standard deviation is called the standard error of the mean. If all the sample means were very close to the population mean, then the standard error of the mean would be small. On the other hand, if the sample means varied considerably, then the standard error of the mean would be large.

To be specific, assume your sample mean were 125 and you estimated that the standard error of the mean were 5 (using a method shown in a later section). If you had a normal distribution, then it would be likely that your sample mean would be within 10 units of the population mean since most of a normal distribution is within two standard deviations of the mean.

Keep in mind that all statistics have sampling distributions, not just the mean. In later sections we will be discussing the sampling distribution of the variance, the sampling distribution of the difference between means, and the sampling distribution of Pearson's correlation, among others.

2.2 Sampling Distribution of the Mean

Mean

The mean of the sampling distribution of the mean is the mean of the population from which the scores were sampled. Therefore, if a population has a mean μ, then the mean of the sampling distribution of the mean is also μ. The symbol μ_M is used to refer to the mean of the sampling distribution of the mean. Therefore, the formula for the mean of the sampling distribution of the mean can be written as:

$$\mu_M = \mu$$

Variance

The variance of the sampling distribution of the mean is computed as follows:

$$\sigma_m^2 = \frac{\sigma^2}{N}$$

That is, the variance of the sampling distribution of the mean is the population variance divided by N , the sample size (the number of scores used to compute a mean). Thus, the larger the sample size, the smaller the variance of the sampling distribution of the mean.

(optional paragraph) This expression can be derived very easily from the variance sum law. Let's begin by computing the variance of the sampling distribution of the

sum of three numbers sampled from a population with variance σ^2. The variance of the sum would be $\sigma^2 + \sigma^2 + \sigma^2$. For N numbers, the variance would be $N\sigma^2$. Since the mean is 1/N times the sum, the variance of the sampling distribution of the mean would be 1/N^2 times the variance of the sum, which equals σ^2/N.

The standard error of the mean is the standard deviation of the sampling distribution of the mean. It is therefore the square root of the variance of the sampling distribution of the mean and can be written as:

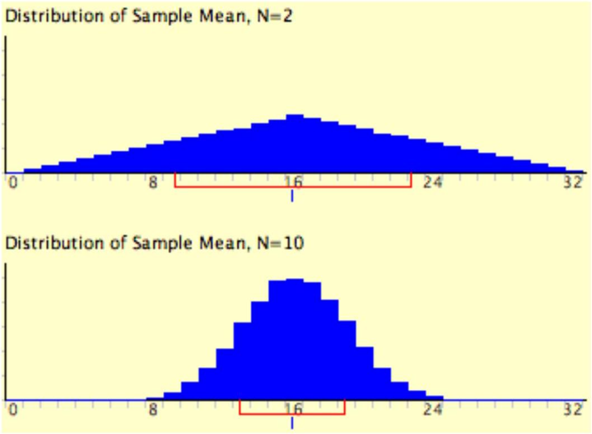
$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

The standard error is represented by a σ because it is a standard deviation. The subscript (M) indicates that the standard error in question is the standard error of the mean.

2.3 Central Limit Theorem

The central limit theorem states that:
Given a population with a finite mean μ and a finite nonzero variance σ^2, the sampling distribution of the mean approaches a normal distribution with a mean of μ and a variance of σ^2/N as N, the sample size, increases.

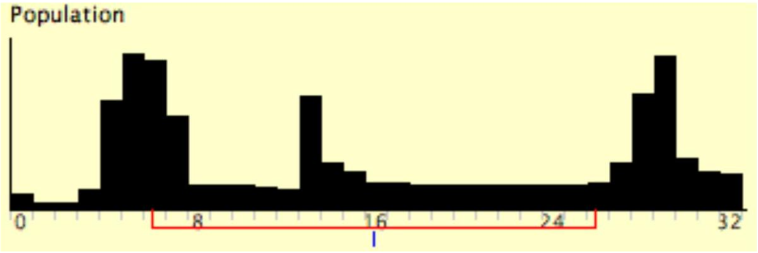
The expressions for the mean and variance of the sampling distribution of the mean are not new or remarkable. What is remarkable is that regardless of the shape of the parent population, the sampling distribution of the mean approaches a normal distribution as N increases. If you have used the "Central Limit Theorem Demo," (external link; requires Java) you have already seen this for yourself. As a reminder, Figure 1 shows the results of the simulation for N = 2 and N = 10. The parent population was a uniform distribution. You can see that the distribution for N = 2 is far from a normal distribution. Nonetheless, it does show that the scores are denser in the middle than in the tails. For N = 10 the distribution is quite close to a normal distribution. Notice that the means of the two distributions are the same, but that the spread of the distribution for N = 10 is smaller.



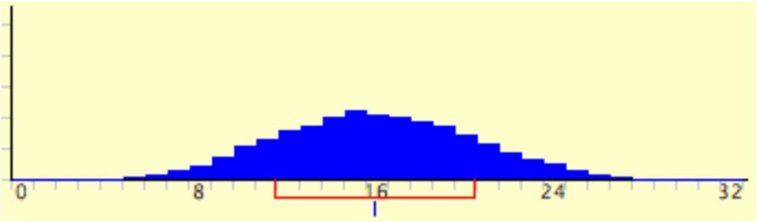
A simulation of a sampling distribution.

The parent population is uniform. The blue line under ” 16 ” indicates that 16 is the mean. The red line extends from the mean plus and minus one standard deviation.

Figure 2 shows how closely the sampling distribution of the mean approximates a normal distribution even when the parent population is very non-normal. If you look closely you can see that the sampling distributions do have a slight positive skew. The larger the sample size, the closer the sampling distribution of the mean would be to a normal distribution.



Distribution of Sample Mean, N = 5



Distribution of Sample Mean, N = 25

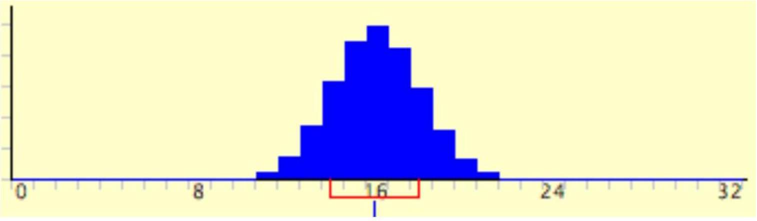


Figure 2. A simulation of a sampling distribution. The parent population is very non-normal.

Theorem 2.1 (The Central Limit Theorem - First form)
If $X_1, X_2, \ldots .X_n$, is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2 and if $\bar{X}$ is the sample mean, the limiting form of the distribution of

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

as $n \to \infty$, is the standard normal distribution.

Note

- The normal approximation for $\bar{X}$ will generally be good if $\boldsymbol{n} \geq 30$, provided the population distribution is not terribly skewed.

- If $\mathbf{n} < 30$, the approximation is good only if the population is not too different from a normal distribution and if the population is known to be normal, the sampling distribution of $\bar{X}$ will follow a normal distribution exactly, no matter how small the size of the samples.

- The sample size $\mathbf{n} = 30$ is a guideline to use for the Central Limit Theorem.

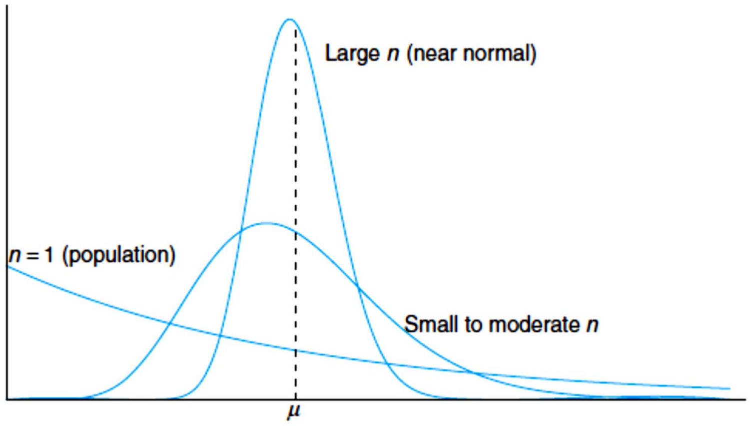
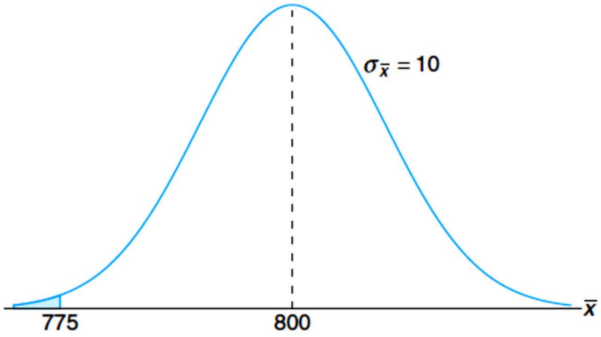


Illustration of the Central Limit Theorem (distribution of $\bar{X}$ for $n = 1$, moderate n, and large n

Example 2.1 *An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of* 40 *hours. Find the* *probability that a random sample of* 16 *bulbs will have an average life of less than* 775 *hours.*

Solution. *The sampling distribution of $\bar{X}$ will be approximately normal, with $\mu_{\bar{X}} = 800$ and*

$$\sigma_{\bar{X}} = 40/\sqrt{16} = 10$$
$$z = \frac{775 - 800}{10} = -2.5$$
$$P(\bar{X} < 775) = P(Z < -2.5) = 0.0062$$



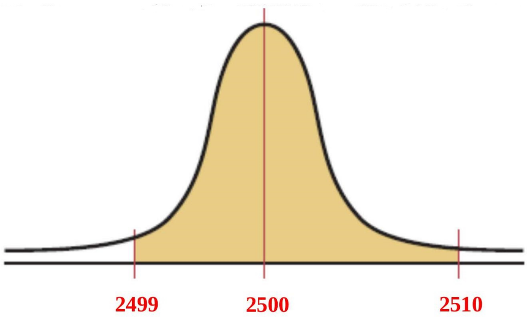
Example 2.2 *The compression strength of concrete is normally distributed with $\mu = 2500$ psi and $\sigma = 50$ psi. Find the probability that a random sample of $n = 5$ specimens will have a sample mean diameter that falls in the interval from 2499 psi to* 2510 *psi.*

Solution.

$$\mu = 2500 \ psi \ \sigma = 50\text{psi} n = 5$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{50}{\sqrt{5}} = 22.3607$$

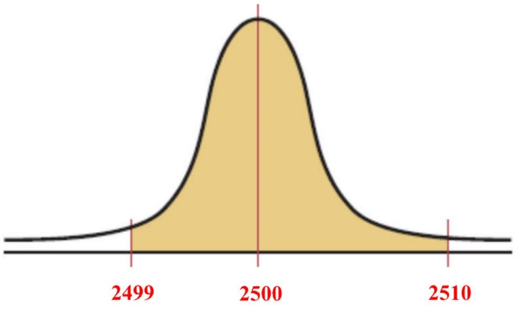


$$P[2499 < \bar{X} < 2510] = P\left[\frac{2499-\mu}{\sigma_{\bar{x}}} < \frac{\bar{X}-\mu}{\sigma_{\bar{x}}} < \frac{2510-\mu}{\sigma_{\bar{x}}}\right]$$

$$P[2499 < \bar{X} < 5510] = P\left[\frac{2499-2500}{22.3607} < Z < \frac{2510-2500}{22.3607}\right]$$
$$= P[-0.0447 < Z < 0.4472]$$

$$Z_1 = -0.0447\,, \qquad Z_2 = 0.4472$$

The area for $\mathbf{Z1} = -0.0447$ *is 0.484047. The area for* $\mathrm{Z2} = 0.4472$ *is 0.673645*



Hence the area between the two values is $= 0.6736 - 0.4840 = 0.1896$ *or* 18.96%.

Therefore,

$$P[2499 < \bar{X} < 2510] = 18.96\%$$

2.3.1 Finite Population Correction Factor

Since sampling with replacement is for the most part unrealistic, a correction factor is necessary for computing the standard error of the mean for samples drawn without replacement from a finite population.

$$\text{correction factor} \ = \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Where : N is the population size and n is the sample size.
This correction factor is necessary if relatively large samples are taken from a small population, because the sample mean will then more accurately estimate the population mean and there will be less error in the estimation.

Finally, the formula for the z value becomes

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}}$$

Example 2.3 *A population of size* 20 *is sampled without replacement. The standard deviation of the population is* 0.35. *We require the standard error of the mean to be no more than* 0.15. *What is the minimum sample size?*

Solution.

$$\mathrm{N} = 20 \quad \sigma = 0.35 \quad \sigma_{\bar{x}} = 0.15$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$0.15 = \frac{0.35}{\sqrt{n}} \sqrt{\frac{20-n}{20-1}} \Rightarrow \sqrt{\frac{20-n}{n}} = \frac{0.15\sqrt{19}}{0.35} = 1.868$$

$$\Rightarrow 20 - n = 3.490n \Rightarrow n = \frac{20}{4.490} = 4.45 \approx 5$$

The central limit theorem can be written as.

Theorem 2.2 (The Central Limit Theorem - Second form) *Let $X_1, X_2, \ldots$ be independent random variables with the same expectation $\mu = \mathbf{E}(X_i)$ and the same standard deviation $\sigma = \mathrm{Std}(X_i)$, and let*

$$S_n = \sum_{i=1}^{n} X_i = X_1 + \ldots + X_n$$

As $n \to \infty$, the standardized sum

$$Z_n = \frac{S_n - \mathbf{E}(S_n)}{\operatorname{Std}(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to a Standard Normal random variable, that is,

$$F_{Z_n}(z) = P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right\} \to \Phi(z)$$

for all z.

This theorem is very powerful because it can be applied to random variables $X_1, X_2, \ldots$ having virtually any thinkable distribution with finite expectation and variance. As long as n is large (the rule of thumb is $n > 30$), one can use Normal distribution to compute probabilities about S_n. Theorem 1 is only one basic version of the Central Limit Theorem. Over the last two centuries, it has been extended to large classes of dependent variables and vectors, stochastic processes, and so on.

Example 2.4 (Allocation OF Disk SPACE) . *A disk has free space of 330 megabytes. Is it likely to be sufficient for 300 independent images, if each image has expected size of 1 megabyte with a standard deviation of 0.5 megabytes?*

Solution. *We have $n = 300, \mu = 1, \sigma = 0.5$. The number of images n is large, so the Central Limit Theorem applies to their total* *size S_n. Then,*

$$\boldsymbol{P}\{\text{ sufficient space }\} = \boldsymbol{P}\left\{S_n \leq 330\right\} = \boldsymbol{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{330-(300)(1)}{0.5\sqrt{300}}\right\}$$
$$\approx \Phi(3.46) = 0.9997$$

This probability is very high, hence, the available disk space is very likely to be sufficient.

In the special case of Normal variables $X_1, X_2, \ldots$, the distribution of S_n is always Normal, and (4.18) becomes exact equality for arbitrary, even small n.

Example 2.5 (Elevator) . *You wait for an elevator, whose capacity is 2000 pounds. The elevator comes with ten adult passengers. Suppose your own weight is 150 lbs , and you heard that human weights are normally distributed with the mean of 165 lbs and the standard deviation of 20 lbs . Would you board this elevator or wait for the next one?*

Solution. *In other words, is overload likely? The probability of an overload equals*

$$P\left\{S_{10} + 150 > 2000\right\} = P\left\{\frac{S_{10}-(10)(165)}{20\sqrt{10}} > \frac{2000-150-(10)(165)}{20\sqrt{10}}\right\}$$
$$= 1 - \Phi(3.16) = 0.0008$$

So, with probability 0.9992 it is safe to take this elevator. It is now for you to decide.

2.4 Sampling Distribution of Difference Between Means

Statistical analyses are very often concerned with the difference between means. A typical example is an experiment designed to compare the mean of a control group with the mean of an experimental group. Inferential statistics used in the analysis of this type of experiment depend on the sampling distribution of the difference between means.

The sampling distribution of the difference between means can be thought of as the distribution that would result if we repeated the following three steps over and over again: (1) sample n_1 scores from Population 1 and n_2 scores from Population 2, (2) compute the means of the two samples (M_1 and M_2), and (3) compute the difference between means, $\mathrm{M}_1 - \mathrm{M}_2$. The distribution of the differences between means is the sampling distribution of the difference between means.

As you might expect, the mean of the sampling distribution of the difference between means is:

$$\mu_{M_1-M_2} = \mu_1 - \mu_2$$

which says that the mean of the distribution of differences between sample means is equal to the difference between population means. For example, say that the mean test score of all 12-year-olds in a population is 34 and the mean of 10-yearolds is 25 . If numerous samples were taken from each age group and the mean

difference computed each time, the mean of these numerous differences between sample means would be $34 - 25 = 9$.

From the variance sum law, we know that:

$$\sigma^2_{M_1-M_2} = \sigma^2_{M_1} + \sigma^2_{M_2}$$

which says that the variance of the sampling distribution of the difference between means is equal to the variance of the sampling distribution of the mean for Population 1 plus the variance of the sampling distribution of the mean for Population 2. Recall the formula for the variance of the sampling distribution of the mean:

$$\sigma^2_M = \frac{\sigma^2}{N}$$

Since we have two populations and two samples sizes, we need to distinguish between the two variances and sample sizes. We do this by using the subscripts 1 and 2. Using this convention, we can write the formula for the variance of the sampling distribution of the difference between means as:

$$\sigma^2_{M_1-M_2} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}$$

Since the standard error of a sampling distribution is the standard deviation of the sampling distribution, the standard error of the difference between means is:

$$\sigma_{M_1-M_2} = \sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}$$

Just to review the notation, the symbol on the left contains a sigma (

σ), which means it is a standard deviation. The subscripts $\mathrm{M}_1 - \mathrm{M}_2$ indicate that it is the standard deviation of the sampling distribution of $\mathrm{M}_1 - \mathrm{M}_2$.

Now let's look at an application of this formula. Assume there are two species of green beings on Mars. The mean height of Species 1 is 32 while the mean height of Species 2 is 22 . The variances of the two species are 60 and 70,
respectively, and the heights of both species are normally distributed. You randomly sample 10 members of Species 1 and 14 members of Species 2. What is the probability that the mean of the 10 members of Species 1 will exceed the mean of the 14 members of Species 2 by 5 or more? Without doing any calculations, you probably know that the probability is pretty high since the difference in population means is 10 . But what exactly is the probability?

First, let's determine the sampling distribution of the difference between means. Using the formulas above, the mean is

$$\mu_{M_1-M_2} = 32 - 22 = 10$$

The standard error is:

$$\sigma_{M_1-M_2} = \sqrt{\frac{60}{10} + \frac{70}{14}} = 3.317$$

The sampling distribution is shown in Figure 1. Notice that it is normally distributed with a mean of 10 and a standard deviation of 3.317. The area above 5 is shaded blue.

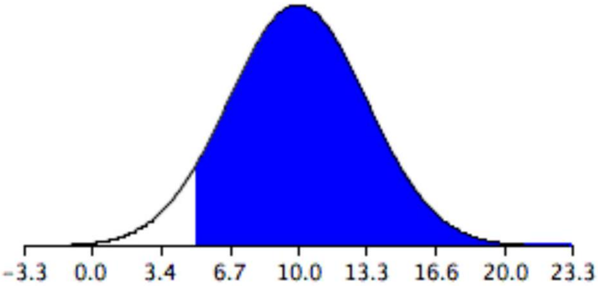


Figure 1. The sampling distribution of the difference between means.

The last step is to determine the area that is shaded blue. Using either a Z table or the normal calculator, the area can be determined to be 0.934 . Thus the probability that the mean of the sample from Species 1 will exceed the mean of the sample from Species 2 by 5 or more is 0.934 .

As shown below, the formula for the standard error of the difference between means is much simpler if the sample sizes and the population variances

are equal. When the variances and samples sizes are the same, there is no need to use the subscripts 1 and 2 to differentiate these terms.

$$\sigma_{M_1-M_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}=\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{n}}=\sqrt{\frac{2\sigma^2}{n}}$$

This simplified version of the formula can be used for the following problem: The mean height of 15 -year-old boys (in cm) is 175 and the variance is 64 . For girls, the mean is 165 and the variance is 64 . If eight boys and eight girls were sampled, what is the probability that the mean height of the sample of girls would be higher than the mean height of the sample of boys? In other words, what is the probability that the mean height of girls minus the mean height of boys is greater than 0 ?

As before, the problem can be solved in terms of the sampling distribution of the difference between means (girls - boys). The mean of the distribution is 165 175 $= -10$. The standard deviation of the distribution is:

$\sigma_{M_1 - M_2} = \sqrt{\frac{2\sigma^2}{n}} = \sqrt{\frac{(2)(64)}{8}} = 4$

A graph of the distribution is shown in Figure 2. It is clear that it is unlikely that the mean height for girls would be higher than the mean height for boys since in the population boys are quite a bit taller. Nonetheless it is not inconceivable that the girls' mean could be higher than the boys' mean.

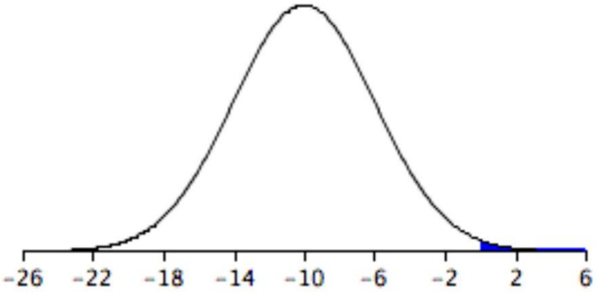


Figure 2. Sampling distribution of the difference between mean heights.

A difference between means of 0 or higher is a difference of $10/4 = 2.5$ standard deviations above the mean of -10 . The probability of a score 2.5 or more standard deviations above the mean is 0.0062 .

Chapter 3

Estimation Theory

After taking a general look at the data, we are ready for more advanced and more informative statistical analysis.

In this chapter, we learn how

- to estimate parameters of the distribution.
- to construct confidence intervals. Any estimator, computed from a collected random sample instead of the whole population, is understood as only an approximation of the corresponding parameter. Instead of one estimator that is subject to a sampling error, it is often more reasonable to produce an interval that will contain the true population parameter with a certain known high probability.

Results of such statistical analysis are used for making decisions under uncertainty, developing optimal strategies, forecasting, evaluating and controlling performance, and so on.

3.1 Parameter estimation

By now, we have learned a few elementary ways to determine the family of distributions. We take into account the nature of our data, basic description, and range; propose a suitable family of distributions;

and support our conjecture by looking at a histogram.
In this section, we learn how to estimate parameters of distributions. As a result, a large family will be reduced to just one distribution that we can use for performance evaluation, forecasting, etc.

Example 1 (Poisson). For example, consider a sample of computer chips with a certain type of rare defects. The number of defects on each chip is recorded. This is the number of rare events, and thus, it should follow a Poisson distribution with some parameter λ.
We know that $\lambda = \mathbf{E}(X)$ is the expectation of a Poisson variable. Then, should we estimate it with a sample mean $\bar{X}$? Or, should we use a sample variance s^2 because λ also equals $\mathrm{Var}(X)$?

Example 2 (Gamma). Suppose now that we deal with a Gamma(α, λ) family of distributions. Its parameters α and λ do not represent the mean, variance, standard deviation, or any other measures discussed in Chapter 8. What would the estimation algorithm be this time?

Questions raised in these examples do not have unique answers. Statisticians developed a number of estimation techniques, each having certain optimal properties.
Two rather popular methods are discussed in this section:

- method of moments, and
- method of maximum likelihood.

3.1.1 Method of moments

Moments: First, let us define the moments.

Definition 3.1 *The k-th population moment is defined as*

$$\mu_k = \mathbf{E}\left(X^k\right)$$

The k-th sample moment

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

estimates μ_k from a sample $(X_1, \ldots, X_n)$.
The first sample moment is the sample mean $\bar{X}$.

Central moments are computed similarly, after centralizing the data, that is, subtracting the mean.

Definition 3.2 *For $k \geq 2$, the k-th population central moment is defined as*

$$\mu'_k = \mathbf{E}\left(X - \mu_1\right)^k$$

The k-th sample central moment

$$m'_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^k$$

estimates μ_k from a sample $(X_1, \ldots, X_n)$.

Remark 3.1 *The second population central moment is variance* $\mathrm{Var}(X)$*. The second sample central moment is sample variance, although* $(n-1)$

in its denominator is now replaced by n. We mentioned that estimation methods are not unique. For unbiased estimation of $\sigma^2 = \mathrm{Var}(X)$, we use

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2 ;$$

however, method of moments and method of maximum likelihood produce a different version,

$$S^2 = m_2' = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2$$

And this is not all! We'll see other estimates of σ^2 as well.

3.1.2 Estimation

Method of moments is based on a simple idea. Since our sample comes from a family of distributions $\{F(\theta)\}$, we choose such a member of this family whose properties are close to properties of our data. Namely, we shall match the moments.

To estimate k parameters, equate the first k population and sample moments,

$$\left\{ \begin{array}{ccc} \mu_1 & = & m_1 \\ \cdots & \cdots & \cdots \\ \mu_k & = & m_k \end{array} \right.$$

The left-hand sides of these equations depend on the distribution parameters. The righthand sides can be computed from data. The method of moments estimator is the solution of this system of equations.

Example 3.1 (Poisson.) *To estimate parameter* λ *of* Poisson(λ) *distribution, we recall that*

$$\mu_1 = \mathbf{E}(X) = \lambda$$

There is only one unknown parameter, hence we write one equation,

$$\mu_1 = \lambda = m_1 = \bar{X}$$

"Solving" it for λ*, we obtain*

$$\widehat{\lambda} = \bar{X}$$

the method of moments estimator of λ*.*

This does not look difficult, does it? Simplicity is the main attractive feature of the method of moments.

If it is easier, one may opt to equate central moments.

Example 3.2 (Gamma distribution of CPU times.)
The histogram in Figure 6 suggested that CPU times have Gamma distribution with some parameters α *and* λ*. To estimate them, we need two equations. From data, we compute*

$$m_1 = \bar{X} = 48.2333 \textit{ and } m_2' = S^2 = 679.7122$$

and write two equations,

$$\begin{cases} \mu_1 = \mathbf{E}(X) = \alpha/\lambda = m_1 \\ \mu_2' = \mathrm{Var}(X) = \alpha/\lambda^2 = m_2' \end{cases}$$

It is convenient to use the second central moment here because we already know the expression for the variance $m_2' = \mathrm{Var}(X)$ of a Gamma variable.
Solving this system in terms of α and λ, we get the method of moment estimates

$$\begin{cases} \widehat{\alpha} = m_1^2/m_2' = 3.4227 \\ \widehat{\lambda} = m_1/m_2' = 0.0710 \end{cases}$$

Of course, we solved these two examples so quickly because we already knew the moments of Poisson and Gamma distributions.
Consider, for example, Pareto distribution that plays an increasingly vital role in modern internet modeling due to very heavy internet traffic nowadays.

Example 3.3 (Pareto.) *A two-parameter Pareto distribution has a cdf*

$$F(x) = 1 - \left(\frac{x}{\sigma}\right)^{-\theta} \qquad \textit{for } x > \sigma$$

How should we compute method of moments estimators of σ and θ ?
We have not seen Pareto distribution in this book so far, so we'll have to compute its first two moments.

We start with the density

$$f(x) = F'(x) = \frac{\theta}{\sigma}\left(\frac{x}{\sigma}\right)^{-\theta-1} = \theta\sigma^{\theta}x^{-\theta-1}$$

and use it to find the expectation

$$\begin{aligned} \mu_1 = \mathbf{E}(X) &= \int_{\sigma}^{\infty} xf(x)dx = \theta\sigma^{\theta}\int_{\sigma}^{\infty} x^{-\theta}dx \\ &= \theta\sigma^{\theta}\frac{x^{-\theta+1}}{-\theta+1}\bigg|_{x=\sigma}^{x=\infty} = \frac{\theta\sigma}{\theta-1},\ \textit{for } \theta > 1 \end{aligned}$$

and the second moment

$$\mu_2 = \mathbf{E}\left(X^2\right) = \int_{\sigma}^{\infty} x^2 f(x)dx = \theta\sigma^{\theta}\int_{\sigma}^{\infty} x^{-\theta+1}dx = \frac{\theta\sigma^2}{\theta-2},\ \textit{for } \theta > 2$$

For $\theta \le 1$, a Pareto variable has an infinite expectation, and for $\theta \le 2$, it has an infinite second moment.
Then we solve the method of moments equations

$$\begin{cases} \mu_1 = \frac{\theta\sigma}{\theta-1} = m_1 \\ \mu_2 = \frac{\theta\sigma^2}{\theta-2} = m_2 \end{cases}$$

and find that

$$\widehat{\theta} = \sqrt{\frac{m_2}{m_2 - m_1^2}} + 1 \ \textit{and}\ \widehat{\sigma} = \frac{m_1(\widehat{\theta}-1)}{\widehat{\theta}} \tag{3.1}$$

When we collect a sample from Pareto distribution, we can compute sample moments m_1 and m_2 and estimate parameters by (9.1).

On rare occasions, when k equations are not enough to estimate k parameters, we'll consider higher moments.

Example 3.4 (Normal.) *Suppose we already know the mean μ of a Normal distribution and would like to estimate the variance σ^2. Only one parameter σ^2 is unknown; however, the first method of moments equation*

$$\mu_1 = m_1$$

does not contain σ^2 and therefore does not produce its estimate. We then consider the second equation, say,

$$\mu_2' = \sigma^2 = m_2' = S^2$$

which gives us the method of moments estimate immediately, $\widehat{\sigma}^2 = S^2$.

Method of moments estimates are typically easy to compute. They can serve as a quick tool for estimating parameters of interest.

3.1.3 Method of maximum likelihood

Another interesting idea is behind the method of maximum likelihood estimation.

Since the sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ has already been observed, we find such parameters that maximize the probability (likelihood) for this to happen. In other words, we make the event that has already happened to be as likely as possible. This is yet another way to make the chosen distribution consistent with the observed data.

Definition 3.3 *Maximum likelihood estimator is the parameter value* *that maximizes the likelihood of the observed sample. For a discrete distribution, we maximize the joint* pmf *of data* $P\left(X_1, \ldots, X_n\right)$*. For a continuous distribution, we maximize the joint density* $f\left(X_1, \ldots, X_n\right)$*.*

Both cases, discrete and continuous, are explained below.

Discrete case

For a discrete distribution, the probability of a given sample is the joint pmf of data,

$$\boldsymbol{P}\left\{\boldsymbol{X}=\left(X_1, \ldots, X_n\right)\right\}=P(\boldsymbol{X})=P\left(X_1, \ldots, X_n\right)=\prod_{i=1}^n P\left(X_i\right)$$

because in a simple random sample, all observed X_i are independent.

To maximize this likelihood, we consider the critical points by taking derivatives with respect to all unknown parameters and equating them to 0 . The maximum can only be attained at such parameter values θ where the derivative $\frac{\partial}{\partial \theta} P(\boldsymbol{X})$ equals 0 , where it does not exist, or at the boundary of the set of possible values of θ.
A nice computational shortcut is to take logarithms first. Differentiating the sum

$$\ln \prod_{i=1}^n P\left(X_i\right)=\sum_{i=1}^n \ln P\left(X_i\right)$$

is easier than differentiating the product $\prod P\left(X_i\right)$. Besides, logarithm is an increasing function, so the likelihood $P(\boldsymbol{X})$ and the log-likelihood $\ln P(\boldsymbol{X})$ are maximized by exactly the same parameters.

Example 3.5 (Poisson.) *The pmf of Poisson distribution is*

$$P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

and its logarithm is

$$\ln P(x) = -\lambda + x\ln\lambda - \ln(x!)$$

Thus, we need to maximize

$$\ln P(\boldsymbol{X}) = \sum_{i=1}^{n}(-\lambda + X_i\ln\lambda) + C = -n\lambda + \ln\lambda\sum_{i=1}^{n}X_i + C$$

where $C = -\sum\ln(x!)$ is a constant that does not contain the unknown parameter λ.
Find the critical point(s) of this log-likelihood. Differentiating it and equating its derivative to 0 , we get

$$\frac{\partial}{\partial\lambda}\ln P(\boldsymbol{X}) = -n + \frac{1}{\lambda}\sum_{i=1}^{n}X_i = 0$$

This equation has only one solution

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}X_i = \bar{X}$$

Since this is the only critical point, and since the likelihood vanishes (converges to 0) as $\lambda \downarrow 0$ or $\lambda \uparrow \infty$, we conclude that $\widehat{\lambda}$ is the maxi-

mizer. Therefore, it is the maximum likelihood estimator of λ.
For the Poisson distribution, the method of moments and the method of maximum likelihood returned the same estimator, $\widehat{\lambda} = \bar{X}$.

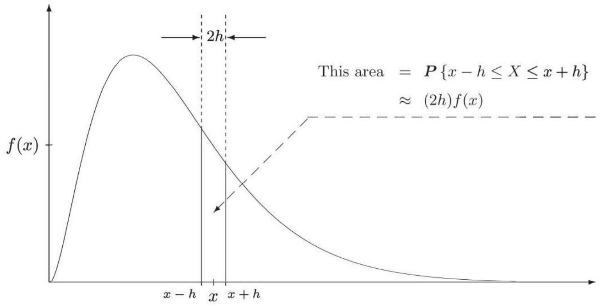


FIGURE 1: Probability of observing "almost" $X = x$.

Continuous case

In the continuous case, the probability to observe exactly the given number $X = x$ is 0. Instead, the method of maximum likelihood will maximize the probability of observing "almost" the same number.
For a very small h,

$$\boldsymbol{P}\{x-h < X < x+h\} = \int_{x-h}^{x+h} f(y)dy \approx (2h)f(x)$$

That is, the probability of observing a value close to x is proportional to the density $f(x)$ (see Figure 1). Then, for a sample $\boldsymbol{X} = (X_1, \ldots, X_n)$, the maximum likelihood method will maximize the joint density $f(X_1, \ldots, X_n)$.

Example 3.6 (Exponential.) *The Exponential density is*

$$f(x) = \lambda e^{-\lambda x}$$

so the log-likelihood of a sample can be written as

$$\ln f(\boldsymbol{X}) = \sum_{i=1}^{n} \ln\left(\lambda e^{-\lambda X_i}\right) = \sum_{i=1}^{n} (\ln\lambda - \lambda X_i) = n\ln\lambda - \lambda\sum_{i=1}^{n} X_i$$

Taking its derivative with respect to the unknown parameter λ, equating it to 0 , and solving for λ, we get

$$\frac{\partial}{\partial\lambda}\ln f(\boldsymbol{X}) = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i = 0$$

resulting in

$$\widehat{\lambda} = \frac{n}{\sum X_i} = \frac{1}{\bar{X}}$$

Again, this is the only critical point, and the likelihood $f(\boldsymbol{X})$ vanishes as $\lambda \downarrow 0$ or $\lambda \uparrow \infty$. Thus, $\widehat{\lambda} = \bar{X}$ is the maximum likelihood estimator of λ. This time, it also coincides with the method of moments estimator.

Sometimes the likelihood has no critical points inside its domain, then it is maximized at the boundary.

Example 3.7 (Uniform.) *Based on a sample from Uniform $(0, b)$ distribution, how can we estimate the parameter b ?*

The Uniform $(0, b)$ *density is*

$$f(x) = \frac{1}{b} \quad \textit{for } 0 \leq x \leq b$$

It is decreasing in b, *and therefore, it is maximized at the the smallest possible value of* b, *which is* x.

For a sample $(X_1, \ldots, X_n)$, *the joint density*

$$f(X_1, \ldots, X_n) = \left(\frac{1}{b}\right)^n \qquad \textit{for } 0 \leq X_1, \ldots, X_n \leq b$$

also attains its maximum at the smallest possible value of b *which is now the largest observation. Indeed,* $b \geq X_i$ *for all* i *only if* $b \geq \max(X_i)$. *If* $b < \max(X_i)$, *then* $f(\boldsymbol{X}) = 0$, *and this cannot be the maximum value.*

Therefore, the maximum likelihood estimator is $\widehat{b} = \max(X_i)$.

When we estimate more than 1 parameter, all the partial derivatives should be equal 0 at the critical point. If no critical points exist, the likelihood is again maximized on the boundary.

Example 3.8 (Pareto.) *For the Pareto distribution in Example 9.5, the log-likelihood is*

$$\ln f(\boldsymbol{X}) = \sum_{i=1}^{n} \ln\left(\theta\sigma^{\theta} X_i^{-\theta-1}\right) = n \ln\theta + n\theta \ln\sigma - (\theta+1) \sum_{i=1}^{n} \ln X_i$$

for $X_1, \ldots, X_n \geq \sigma$. *Maximizing this function over both* σ *and* θ, *we* *notice that it always increases in σ. Thus, we estimate σ by its largest possible value, which is the smallest observation,*

$$\widehat{\sigma} = \min(X_i)\,.$$

We can substitute this value of σ into the log-likelihood and maximize with respect to θ,

$$\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X}) = \frac{n}{\theta} + n\ln\widehat{\sigma} - \sum_{i=1}^{n}\ln X_i = 0$$
$$\widehat{\theta} = \frac{n}{\sum \ln X_i - n\ln\widehat{\sigma}} = \frac{n}{\sum \ln(X_i/\widehat{\sigma})}$$

The maximum likelihood estimates of σ and θ are

$$\widehat{\sigma} = \min(X_i) \ \text{ and } \quad \widehat{\theta} = \frac{n}{\sum \ln(X_i/\widehat{\sigma})}$$

Maximum likelihood estimators are rather popular because of their nice properties. Under mild conditions, these estimators are consistent, and for large samples, they have an approximately Normal distribution. Often in complicated problems, finding a good estimation scheme may be challenging whereas the maximum likelihood method always gives a reasonable solution.

3.1.4 Estimation of standard errors

How good are the estimators that we learned in previous sections? Standard errors can serve as measures of their accuracy. To estimate them, we derive an expression for the standard error and estimate all the unknown parameters in it.

Example 3.9 (Estimation of the Poisson Parameter.) *In Examples 3.3 and 3.7, we found the method of moments and maximum likelihood estimators of the Poisson parameter λ. Both estimators appear to be equal the sample mean $\widehat{\lambda} = \bar{X}$. Let us now estimate the standard error of $\widehat{\lambda}$.*

Solution. *There are at least two ways to do it.*

On one hand, $\sigma = \sqrt{\lambda}$ for the Poisson(λ) *distribution, so $\sigma(\widehat{\lambda}) = \sigma(\bar{X}) = \sigma/\sqrt{n} = \sqrt{\lambda/n}$, as we know from (8.2) on p. 219. Estimating λ by $\bar{X}$, we obtain*

$$s_1(\widehat{\lambda}) = \sqrt{\frac{\bar{X}}{n}} = \frac{\sqrt{\sum X_i}}{n}$$

On the other hand, we can use the sample standard deviation and estimate the standard error of the sample mean as in Example 8.17,

$$s_2(\widehat{\lambda}) = \frac{s}{\sqrt{n}} = \sqrt{\frac{\sum \left(X_i - \bar{X}\right)^2}{n(n-1)}}$$

Apparently, we can estimate the standard error of $\widehat{\lambda}$ by two good estimators, s_1 and s_2. $\diamond$

Example 3.10 (Estimation of the Exponential parameter.) *Derive the standard error of the maximum likelihood estimator in Example 8 and estimate it, assuming a sample size $n \geq 3$.*

Solution. *This requires some integration work. Fortunately, we can* *take a shortcut because we know that the integral of any Gamma density is one, i.e.,*

$$\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = 1 \qquad \textit{for any } \alpha > 0, \lambda > 0$$

Now, notice that $\widehat{\lambda} = 1/\bar{X} = n/\sum X_i$, *where* $\sum X_i$ *has Gamma* (n, λ) *distribution because each* X_i *is* Exponential(λ).

Therefore, the k-th moment of $\hat{\lambda}$ *equals*

$$\begin{aligned}
\mathbf{E}\left(\widehat{\lambda}^k\right) = \mathbf{E}\left(\frac{n}{\sum X_i}\right)^k &= \int_0^\infty \left(\frac{n}{x}\right)^k \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx \\
&= \frac{n^k \lambda^n}{\Gamma(n)} \int_0^\infty x^{n-k-1} e^{-\lambda x} dx \\
&= \frac{n^k \lambda^n}{\Gamma(n)} \frac{\Gamma(n-k)}{\lambda^{n-k}} \int_0^\infty \frac{\lambda^{n-k}}{\Gamma(n-k)} x^{n-k-1} e^{-\lambda x} dx \\
&= \frac{n^k \lambda^n}{\Gamma(n)} \frac{\Gamma(n-k)}{\lambda^{n-k}} \cdot 1 = \frac{n^k \lambda^k (n-k-1)!}{(n-1)!}
\end{aligned}$$

Substituting $k = 1$, *we get the first moment,*

$$\mathbf{E}(\widehat{\lambda}) = \frac{n\lambda}{n-1}$$

Substituting $k = 2$, *we get the second moment,*

$$\mathbf{E}\left(\widehat{\lambda}^2\right) = \frac{n^2 \lambda^2}{(n-1)(n-2)}$$

Then, the standard error of $\hat{\lambda}$ is

$$\begin{aligned}\sigma(\widehat{\lambda}) &= \sqrt{\mathrm{Var}(\widehat{\lambda})} = \sqrt{\mathbf{E}\left(\widehat{\lambda}^2\right) - \mathbf{E}^2(\widehat{\lambda})} \\ &= \sqrt{\frac{n^2\lambda^2}{(n-1)(n-2)} - \frac{n^2\lambda^2}{(n-1)^2}} \\ &= \frac{n\lambda}{(n-1)\sqrt{n-2}}\end{aligned}$$

We have just estimated λ by $\hat{\lambda} = 1/\bar{X}$; therefore, we can estimate the standard error $\sigma(\widehat{\lambda})$ by

$$s(\widehat{\lambda}) = \frac{n}{\bar{X}(n-1)\sqrt{n-2}} \quad \textit{or} \quad \frac{n^2}{\sum X_i(n-1)\sqrt{n-2}}$$

3.2 Confidence intervals

When we report an estimator $\widehat{\theta}$ of a population parameter θ, we know that most likely

$$\widehat{\theta} \neq \theta$$

due to a sampling error. We realize that we have estimated θ up to some error. Likewise, nobody understands the internet connection of 11 megabytes per second as exactly 11 megabytes going through the network every second, and nobody takes a meteorological forecast as the promise of exactly the predicted temperature.

Then how much can we trust the reported estimator? How far can it be from the actual parameter of interest? What is the probability

that it will be reasonably close? And if we observed an estimator $\widehat{\theta}$, then what can the actual parameter θ be?
To answer these questions, statisticians use confidence intervals, which contain parameter values that deserve some confidence, given the observed data.

Definition 3.4 *An interval* $[a, b]$ *is a* $(1-\alpha)100\%$ *confidence interval for the parameter* θ *if it contains the parameter with probability* $(1-\alpha)$*,*

$$\boldsymbol{P}\{a \leq \theta \leq b\} = 1 - \alpha.$$

The coverage probability $(1-\alpha)$ *is also called a confidence level.*

Let us take a moment to think about this definition. The probability of a random event $\{a \leq \theta \leq b\}$ has to be $(1-\alpha)$. What randomness is involved in this event?

The population parameter θ is not random. It is a population feature, independent of any random sampling procedure, and therefore, it remains constant. On the other hand, the interval is computed from random data, and therefore, it is random. The coverage probability refers to the chance that our interval covers a constant parameter θ.
This is illustrated in Figure 2. Suppose that we collect many random samples and produce a confidence interval from each of them. If these are $(1-\alpha)100\%$ confidence intervals, then we expect $(1-\alpha)100\%$ of them to cover θ and $100\alpha\%$ of them to miss it.

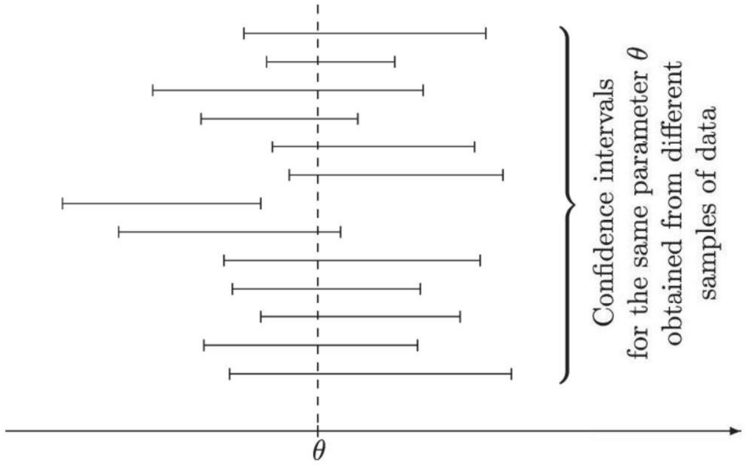


FIGURE 2: Confidence intervals and coverage of parameter θ.

In Figure 2, we see one interval that does not cover θ. No mistake was made in data collection and construction of this interval. It missed the parameter only due to a sampling error.

It is therefore wrong to say, "I computed a 90% confidence interval, it is [3, 6]. Parameter belongs to this interval with probability 90%." The parameter is constant; it either belongs to the interval $[3, 6]$ (with probability 1) or does not. In this case, 90% refers to the proportion of confidence intervals that contain the unknown parameter in a long run.

3.2.1 Construction of confidence intervals

Given a sample of data and a desired confidence level $(1-\alpha)$, how can we construct a confidence interval $[a, b]$ that will satisfy the coverage condition

$$\boldsymbol{P}\{a \leq \theta \leq b\}=1-\alpha$$

in Definition 4?

We start by estimating parameter θ. Assume there is an unbiased estimator $\widehat{\theta}$ that has a Normal distribution. When we standardize it, we get a Standard Normal variable

$$Z = \frac{\widehat{\theta} - \mathbf{E}(\widehat{\theta})}{\sigma(\widehat{\theta})} = \frac{\widehat{\theta} - \theta}{\sigma(\widehat{\theta})} \tag{3.2}$$

where $\mathbf{E}(\widehat{\theta}) = \theta$ because $\widehat{\theta}$ is unbiased, and $\sigma(\widehat{\theta}) = \sigma(\widehat{\theta})$ is its standard error.

This variable falls between the Standard Normal quantiles $q_{\alpha/2}$ and $q_{1-\alpha/2}$, denoted by

$$-z_{\alpha/2} = q_{\alpha/2}, \quad z_{\alpha/2} = q_{1-\alpha/2}$$

with probability $(1 - \alpha)$, as you can see in Figure 3.

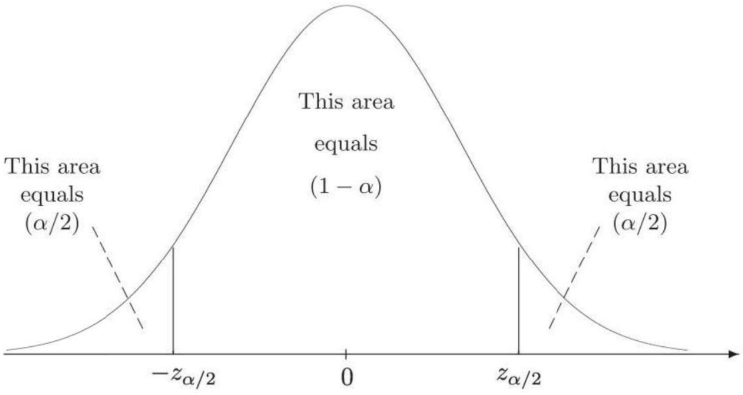


FIGURE 3: Standard Normal quantiles $\pm z_{\alpha/2}$ and partition of the area under the density curve.

Then,

$$\boldsymbol{P}\left\{-z_{\alpha/2} \leq \frac{\widehat{\theta}-\theta}{\sigma(\widehat{\theta})} \leq z_{\alpha/2}\right\} = 1-\alpha$$

Solving the inequality inside $\{\ldots\}$ for θ, we get

$$\boldsymbol{P}\left\{\widehat{\theta} - z_{\alpha/2} \cdot \sigma(\widehat{\theta}) \leq \theta \leq \widehat{\theta} - z_{\alpha/2} \cdot \sigma(\widehat{\theta})\right\} = 1-\alpha$$

The problem is solved! We have obtained two numbers

$$a = \widehat{\theta} - z_{\alpha/2} \cdot \sigma(\widehat{\theta}),\ b = \widehat{\theta} + z_{\alpha/2} \cdot \sigma(\widehat{\theta})$$

such that

$$\boldsymbol{P}\{a \leq \theta \leq b\} = 1-\alpha.$$

3.2.2 Confidence interval, Normal distribution

If parameter θ has an unbiased, Normally distributed estimator $\widehat{\theta}$, then

$$\widehat{\theta} \pm z_{\alpha/2} \cdot \sigma(\widehat{\theta}) = \left[\widehat{\theta} - z_{\alpha/2} \cdot \sigma(\widehat{\theta}), \widehat{\theta} + z_{\alpha/2} \cdot \sigma(\widehat{\theta})\right] \tag{3.3}$$

is a $(1-\alpha)100\%$ confidence interval for θ.
If the distribution of $\widehat{\theta}$ is approximately Normal, we get an approximately $(1-\alpha)100\%$ confidence interval.

In this formula, $\widehat{\theta}$ is the center of the interval, and $z_{\alpha/2} \cdot \sigma(\widehat{\theta})$ is the margin. The margin of error is often reported along with poll and survey results. In newspapers and press releases, it is usually computed for a 95% confidence interval.
We have seen quantiles $\pm z_{\alpha/2}$ in inverse problems. Now, in confidence

estimation, and also, in the next section on hypothesis testing, they will play a crucial role as we'll need to attain the desired confidence level α. The most commonly used values are

$$\begin{gathered} z_{0.10} = 1.282, \quad z_{0.05} = 1.645, \quad z_{0.025} = 1.960 \\ z_{0.01} = 2.326, \quad z_{0.005} = 2.576 \end{gathered} \tag{3.4}$$

NOTATION:

$$z_\alpha = q_{1-\alpha} = \Phi^{-1}(1-\alpha)$$

is the value of a Standard Normal variable Z that is exceeded with probability α.

Several important applications of this general method are discussed below. In each problem, we
(a) find an unbiased estimator of θ,
(b) check if it has a Normal distribution,
(c) find its standard error $\sigma(\widehat{\theta}) = \mathrm{Std}(\widehat{\theta})$,
(d) obtain quantiles $\pm z_{\alpha/2}$ from the table of Normal distribution (Table A4 in the Appendix), and finally,
(e) apply the rule (3).

3.3 Confidence interval for the population mean

Let us construct a confidence interval for the population mean

$$\theta = \mu = \mathbf{E}(X)$$

Start with an estimator,

$$\widehat{\theta} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The rule (3) is applicable in two cases.

1. If a sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ comes from Normal distribution, then $\bar{X}$ is also Normal, and rule (3) can be applied.
2. If a sample comes from any distribution, but the sample size n is large, then $\bar{X}$ has an approximately Normal distribution according to the Central Limit Theorem. Then rule (3) gives an approximately $(1-\alpha)100\%$ confidence interval.

Before, we derived

$$\begin{aligned} &\mathbf{E}(\bar{X}) = \mu \quad \text{(thus, it is an unbiased estimator);} \\ &\sigma(\bar{X}) = \sigma/\sqrt{n}. \end{aligned}$$

Then, (3) reduces to the following $(1-\alpha)100\%$ confidence interval for μ.

Confidence interval for the mean; σ is known

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{3.5}$$

Example 3.11 *Construct a* 95% *confidence interval for the population mean based on a sample of measurements* $2.5, 7.4, 8.0, 4.5, 7.4, 9.2$ *if* *measurement errors have Normal distribution, and the measurement device guarantees a standard deviation of* $\sigma = 2.2$.

Solution. *This sample has size* $n = 6$ *and sample mean* $\bar{X} = 6.50$. *To attain a confidence level of* $1 - \alpha = 0.95$ *we need* $\alpha = 0.05$ *and* $\alpha/2 = 0.025$. *Hence, we are looking for quantiles* $q_{0.025} = -z_{0.025}$ *and* $q_{0.975} = z_{0.025}$.
From (4) or Table A4, we find that $q_{0.975} = 1.960$. *Substituting these values into (5), we obtain a* 95% *confidence interval for* μ,

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 6.50 \pm (1.960)\frac{2.2}{\sqrt{6}} = \underline{6.50 \pm 1.76} \text{ or } \underline{[4.74, 8.26]}$$

The only situation when method (3) cannot be applied is when the sample size is small and the distribution of data is not Normal. Special methods for the given distribution of X are required in this case.

3.4 Confidence interval for the difference between two means

Under the same conditions as in the previous section,

- Normal distribution of data or

- sufficiently large sample size,
  we can construct a confidence interval for the difference between two means.
  This problem arises when we compare two populations. It may be a comparison of two materials, two suppliers, two service providers, two communication channels, two labs, etc. From each population,

a sample is collected (Figure 4),

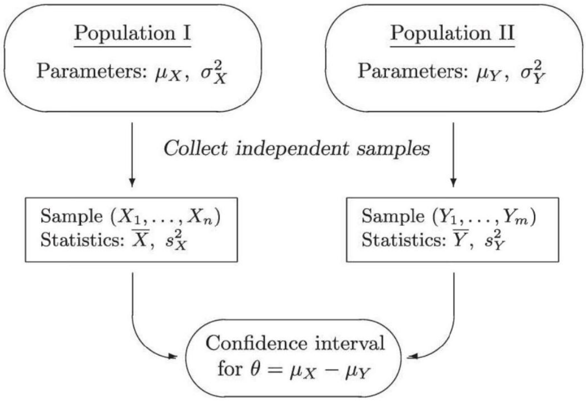

Population I
Parameters: μ_X, σ_X^2
Population II
Parameters: μ_Y, σ_Y^2
Collect independent samples
Sample $(X_1,\dots,X_n)$
Statistics: $\overline{X}$, s_X^2
Sample $(Y_1,\dots,Y_m)$
Statistics: $\overline{Y}$, s_Y^2
Confidence interval
for $\theta = \mu_X - \mu_Y$

FIGURE 4: Comparison of two populations.

$$\begin{aligned} \boldsymbol{X} &= (X_1,\dots,X_n) \quad \text{from one population} \\ \boldsymbol{Y} &= (Y_1,\dots,Y_m) \quad \text{from the other population.} \end{aligned}$$

Suppose that the two samples are collected independently of each other.

To construct a confidence interval for the difference between population means $\theta = \mu_X - \mu_Y$ we complete the usual steps (a)-(e) below.

(a) Propose an estimator of θ,

$$\widehat{\theta} = \bar{X} - \bar{Y}$$

It is natural to come up with this estimator because $\bar{X}$ estimates μ_X and $\bar{Y}$ estimates μ_Y.

(b) Check that $\widehat{\theta}$ is unbiased. Indeed,

$$\mathbf{E}(\widehat{\theta}) = \mathbf{E}(\bar{X} - \bar{Y}) = \mathbf{E}(\bar{X}) - \mathbf{E}(\bar{Y}) = \mu_X - \mu_Y = \theta$$

(c) Check that $\widehat{\theta}$ has a Normal or approximately Normal distribution. This is true if the observations are Normal or both sample sizes m and n are large.
(d) Find the standard error of $\widehat{\theta}$ (using independence of $\boldsymbol{X}$ and $\boldsymbol{Y}$),

$$\sigma(\widehat{\theta}) = \sqrt{\operatorname{Var}(\bar{X} - \bar{Y})} = \sqrt{\operatorname{Var}(\bar{X}) + \operatorname{Var}(\bar{Y})} = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

(e) Find quantiles $\pm z_{\alpha/2}$ and compute the confidence interval according to (3). This results in the following formula.

Confidence interval for the difference of means; known standard deviation

$$\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \tag{3.6}$$

Example 3.12 (Effect of an upgrade.) *A manager evaluates effectiveness of a major hardware upgrade by running a certain process 50 times before the upgrade and 50 times after it. Based on these data, the average running time is 8.5 minutes before the upgrade, 7.2 minutes after it. Historically, the standard deviation has been 1.8 minutes, and presumably it has not changed. Construct a* 90% *confidence interval showing how much the mean running time reduced due to the hardware upgrade.*

Solution. *We have* $n = m = 50, \sigma_X = \sigma_Y = 1.8, \bar{X} = 8.5$, *and* $\bar{Y} = 7.2$. *Also, the confidence level* $(1 - \alpha)$ *equals 0.9 , hence* $\alpha/2 = 0.05$, *and* $z_{\alpha/2} = 1.645$.

The distribution of times may not be Normal; however, due to large *sample sizes, the estimator*

$$\widehat{\theta} = \bar{X} - \bar{Y}$$

is approximately Normal by the Central Limit Theorem. Thus, formula (9.6) is applicable, and a 90% *confidence interval for the difference of means (* $\mu_X - \mu_Y$ *) is*

$$8.5 - 7.2 \pm (1.645)\sqrt{1.8^2\left(\frac{1}{50} + \frac{1}{50}\right)} = \underline{1.3 \pm 0.6 \ \ or \ [0.7, 1.9]}$$

We can say that the hardware upgrade resulted in a 1.3 -minute reduction of the mean running time, with a 90% *confidence margin of 0.6 minutes.*

3.5 Selection of a sample size

Formula (3) describes a confidence interval as " center ± margin , where

$$\text{center } = \widehat{\theta}, \quad \text{margin } = z_{\alpha/2} \cdot \sigma(\widehat{\theta}).$$

We can revert the problem and ask a very practical question: How large a sample should be collected to provide a certain desired precision of our estimator?

In other words, what sample size n guarantees that the margin of a $(1-\alpha)100\%$ confidence interval does not exceed a specified limit Δ ? To answer this question, we only need to solve the inequality

$$\text{margin} \leq \Delta \tag{3.7}$$

in terms of n. Typically, parameters are estimated more accurately based on larger samples, so that the standard error $\sigma(\widehat{\theta})$ and the margin are decreasing functions of sample size n. Then, (7) must be satisfied for sufficiently large n.

3.6 Estimating means with a given precision

When we estimate a population mean, the margin of error is

$$\text{margin} = z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

Solving inequality (7) for n results in the following rule.

Rule: Sample size for a given precision

$$\begin{array}{c}\text{In order to attain a margin of error } \Delta \text{ for estimating} \\ \text{a population mean with a confidence level } (1-\alpha), \\ \text{a sample of size } n \geq \left(\frac{z_{\alpha/2}\cdot\sigma}{\Delta}\right)^2 \text{ is required.}\end{array} \qquad (3.8)$$

When we compute the expression in (8), it will most likely be a fraction. Notice that we can only round it up to the nearest integer sample size. If we round it down, our margin will exceed Δ.

Looking at (8), we see that a large sample will be necessary

- to attain a narrow margin (small Δ);
- to attain a high confidence level (small α); and
- to control the margin under high variability of data (large σ).

In particular, we need to quadruple the sample size in order to half the margin of the interval.

Example 3.13 *In Example 11, we constructed a 95% confidence with the center 6.50 and margin 1.76 based on a sample of size 6 . Now, that was too wide, right? How large a sample do we need to estimate the population mean with a margin of at most 0.4 units with* 95% *confidence?*

Solution. *We have* $\Delta = 0.4, \alpha = 0.05$, *and from Example 9.13,* $\sigma = 2.2$. *By (9.8), we need a sample of*

$$n \geq \left(\frac{z_{0.05/2} \cdot \sigma}{\Delta}\right)^2 = \left(\frac{(1.960)(2.2)}{0.4}\right)^2 = 116.2$$

Keeping in mind that this is the minimum sample size that satisfies Δ, *and we are only allowed to round it up, we need a sample of at least 117 observations.*

Chapter 4

Hypotheses Testing

A vital role of Statistics is in verifying statements, claims, conjectures, and in general - testing hypotheses. Based on a random sample, we can use Statistics to verify whether

- a system has not been infected,
- a hardware upgrade was efficient,
- the average number of concurrent users increased by 2000 this year,
- the average connection speed is 54 Mbps, as claimed by the internet service provider,
- the proportion of defective products is at most 3
- service times have Gamma distribution,
- the number of errors in software is independent of the manager's experience, - etc.

Testing statistical hypotheses has wide applications far beyond Computer Science. These methods are used to prove efficiency of a new medical treatment, safety of a new automobile brand, innocence of a defendant, and authorship of a document; to establish cause-and-effect relationships; to identify factors that can significantly improve the response; to fit stochastic models; to detect information leaks; and so forth.

4.1 Hypothesis and alternative

To begin, we need to state exactly what we are testing. These are hypothesis and alternative.

NOTATION: $H_0 =$ hypothesis (the null hypothesis)
$H_A =$ alternative (the alternative hypothesis)

H_0 and H_A are simply two mutually exclusive statements. Each test results either in acceptance of H_0 or its rejection in favor of H_A.

A null hypothesis is always an equality, absence of an effect or relation, some "normal," usual statement that people have believed in for years. In order to overturn the common belief and to reject the hypothesis, we need significant evidence. Such evidence can only be provided by data. Only when such evidence is found, and when it strongly supports the alternative H_A, can the hypothesis H_0 be rejected in favor of H_A.

Based on a random sample, a statistician cannot tell whether the hypothesis is true or the alternative. We need to see the entire population to tell that. The purpose of each test is to determine whether the data provides sufficient evidence against H_0 in favor of H_A.

This is similar to a criminal trial. Typically, the jury cannot tell whether the defendant committed a crime or not. It is not their task. They are only required to determine if the presented evidence against the defendant is sufficient and convincing. By default, called presumption of innocence, insufficient evidence leads to acquittal.

Example 4.1 *To verify that the the average connection speed is 54 Mbps , we test the hypothesis* $H_0 : \mu = 54$ *against the two-sided alternative* $H_A : \mu \neq 54$*, where* μ *is the average speed of all connections.*

However, if we worry about a low connection speed only, we can conduct a one-sided test of

$$H_0 : \mu = 54 \textit{ vs } H_A : \mu < 54$$

In this case, we only measure the amount of evidence supporting the one-sided alternative $H_A : \mu < 54$*. In the absence of such evidence, we gladly accept the null hypothesis.*

Definition 4.1 *Alternative of the type* $H_A : \mu \neq \mu_0$ *covering regions on both sides of the hypothesis* $(H_0 : \mu = \mu_0)$ *is a two-sided alternative. Alternative* $H_A : \mu < \mu_0$ *covering the region to the left of* H_0 *is one-sided, left-tail.*

Alternative $H_A : \mu > \mu_0$ *covering the region to the right of* H_0 *is one-sided, right-tail.*

Example 4.2 *To verify whether the average number of concurrent* *users increased by 2000, we test*

$$H_0 : \mu_2 - \mu_1 = 2000 \text{ vs } H_A : \mu_2 - \mu_1 \neq 2000$$

where μ_1 *is the average number of concurrent users last year, and* μ_2 *is the average number of concurrent users this year. Depending on the situation, we may replace the two-sided alternative* $H_A : \mu_2 - \mu_1 \neq 2000$ *with a one-sided alternative* $H_A^{(1)} : \mu_2 - \mu_1 < 2000$ *or* $H_A^{(2)} : \mu_2 - \mu_1 > 2000$*. The test of* H_0 *against* $H_A^{(1)}$ *evaluates the amount of evidence that the mean number of concurrent users changed by fewer than 2000. Testing against* $H_A^{(2)}$*, we see if there is sufficient evidence to claim that this number increased by more than 2000 .*

Example 4.3 *To verify if the proportion of defective products is at most* 3%*, we test*

$$H_0 : p = 0.03 \text{ vs } H_A : p > 0.03$$

where p *is the proportion of defects in the whole shipment.*
Why do we choose the right-tail alternative $H_A : p > 0.03$ *? That is because we reject the shipment only if significant evidence supporting this alternative is collected. If the data suggest that* $p < 0.03$*, the shipment will still be accepted.*

4.2 Type I and Type II errors: level of significance

When testing hypotheses, we realize that all we see is a random sample. Therefore, with all the best statistics skills, our decision to accept or to reject H_0 may still be wrong. Four situations are possible,

	Result of the test	
	Reject H_0	Accept H_0
H_0 is true	Type I error	correct
H_0 is false	correct	Type II error

In two of the four cases, the test results in a correct decision. Either we accepted a true hypothesis, or we rejected a false hypothesis. The other two situations are sampling errors.

Definition 4.2 *A type I error occurs when we reject the true null hypothesis.*
A type II error occurs when we accept the false null hypothesis.

Each error occurs with a certain probability that we hope to keep small. A good test results in an erroneous decision only if the observed data are somewhat extreme.

A type I error is often considered more dangerous and undesired than a type II error. Making a type I error can be compared with convicting an innocent defendant or sending a patient to a surgery when (s)he does not need one.

For this reason, we shall design tests that bound the probability of type I error by a preassigned small number α. Under this condition, we may want to minimize the probability of type II error.

Definition 4.3 *Probability of a type I error is the significance level of a test,*

$$\alpha = \boldsymbol{P}\{ \textit{ reject } H_0 \mid H_0 \textit{ is true } \}$$

Probability of rejecting a false hypothesis is the power of the test,

$$p(\theta) = \boldsymbol{P}\{ \textit{ reject } H_0 \mid \theta; H_A \textit{ is true } \}$$

It is usually a function of the parameter θ because the alternative hypothesis includes a set of parameter values. Also, the power is the probability to avoid a Type II error.

Typically, hypotheses are tested at significance levels as small as $0.01, 0.05$, or 0.10 , although there are exceptions. Testing at a low level of significance means that only a large amount of evidence can force rejection of H_0. Rejecting a hypothesis at a very low level of significance is done with a lot of confidence that this decision is right.

4.3 Level α tests: general approach

A standard algorithm for a level α test of a hypothesis H_0 against an alternative H_A consists of 3 steps.

Step 1. Test statistic

Testing hypothesis is based on a test statistic T, a quantity computed from the data that has some known, tabulated distribution F_0 if the hypothesis H_0 is true.

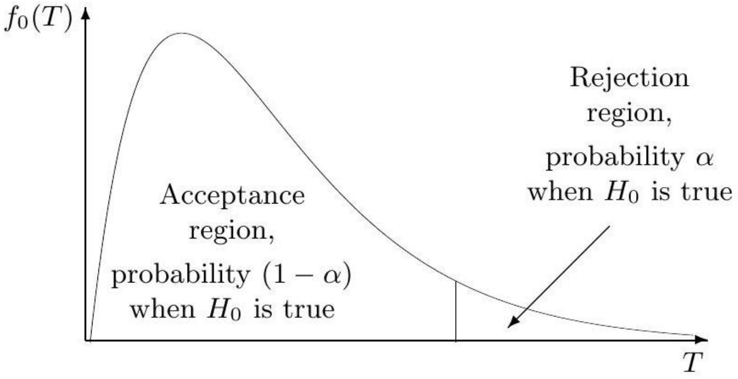


FIGURE 5.1: Acceptance and rejection regions.

Test statistics are used to discriminate between the hypothesis and the alternative. When we verify a hypothesis about some parameter θ, the test statistic is usually obtained by a suitable transformation of its estimator $\widehat{\theta}$.

Step 2. Acceptance region and rejection region
Next, we consider the null distribution F_0. This is the distribution of test statistic T when the hypothesis H_0 is true. If it has a density f_0, then the whole area under the density curve is 1 , and we can always find a portion of it whose area is α, as shown in Figure 1. It is called rejection region ($\mathfrak{R}$).
The remaining part, the complement of the rejection region, is called acceptance region ($\mathfrak{A} = \overline{\mathfrak{R}}$). By the complement rule, its area is $(1-\alpha)$.

These regions are selected in such a way that the values of test statistic T in the rejection region provide a stronger support of H_A

than the values $T \in \mathfrak{A}$. For example, suppose that T is expected to be large if H_A is true. Then the rejection region corresponds to the right tail of the null distribution F_0 (Figure 4.1).

As another example, look at Figure 3 on p. 64. If the null distribution of T is Standard Normal, then the area between $\left(-z_{\alpha/2}\right)$ and $z_{\alpha/2}$ equals exactly $(1-\alpha)$. The interval

$$\mathfrak{A} = \left(-z_{\alpha/2}, z_{\alpha/2}\right)$$

can serve as a level α acceptance region for a two-sided test of $H_0 : \theta = \theta_0$ vs $H_A : \theta \neq \theta_0$. The remaining part consists of two symmetric tails,

$$\mathfrak{R} = \overline{\mathfrak{A}} = \left(-\infty, -z_{\alpha/2}\right] \cup \left[z_{\alpha/2}, +\infty\right);$$

this is the rejection region.

Areas under the density curve are probabilities, and we conclude that

$$\boldsymbol{P}\left\{T \in \text{ acceptance region } \mid H_0\right\} = 1 - \alpha$$

and

$$\boldsymbol{P}\left\{T \in \text{ rejection region } \mid H_0\right\} = \alpha.$$

Step 3: Result and its interpretation

Accept the hypothesis H_0 if the test statistic T belongs to the acceptance region. Reject H_0 in favor of the alternative H_A if T belongs to the rejection region.

Our acceptance and rejection regions guarantee that the significance level of our test is

$$\begin{aligned}\text{Significance level } &= \boldsymbol{P}\{\text{ Type I error }\}\\ &= \boldsymbol{P}\{\text{ Reject } \mid H_0\}\\ &= \boldsymbol{P}\{T \in \mathfrak{R} \mid H_0\}\\ &= \alpha \qquad (4.1)\end{aligned}$$

Therefore, indeed, we have a level α test!

The interesting part is to interpret our result correctly. Notice that conclusions like "My level α test accepted the hypothesis. Therefore, the hypothesis is true with probability $(1-\alpha)$ " are wrong! Statements H_0 and H_A are about a non-random population, and thus, the hypothesis can either be true with probability 1 or false with probability 1.

If the test rejects the hypothesis, all we can state is that the data provides sufficient evidence against H_0 and in favor of H_A. It may either happen because H_0 is not true, or because our sample is too extreme. The latter, however, can only happen with probability α.

If the test accepts the hypothesis, it only means that the evidence obtained from the data is not sufficient to reject it. In the absence of sufficient evidence, by default, we accept the null hypothesis.

NOTATION:

$\alpha =$ level of significance, probability of type I error
$p(\theta) =$ power
$T =$ test statistic
F_0, $f_0 =$ null distribution of T and its density
$\mathfrak{A} =$ acceptance region
$\mathfrak{R} =$ rejection region

4.4 Rejection regions and power

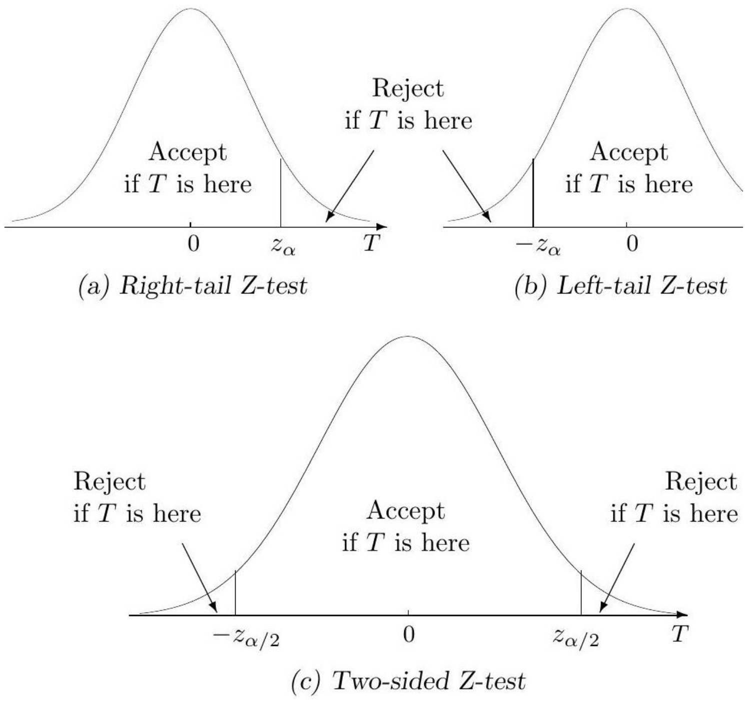
Our construction of the rejection region guaranteed the desired significance level α, as we proved in (4.1). However, one can choose many regions that will also have probability α (see Figure 4.2). Among them, which one is the best choice?

To avoid type II errors, we choose such a rejection region that will likely cover the test statistic T in case if the alternative H_A is true. This maximizes the power of our test because we'll rarely accept H_0 in this case.

Then, we look at our test statistic T under the alternative. Often
(a) a right-tail alternative forces T to be large,
(b) a left-tail alternative forces T to be small,
(c) a two-sided alternative forces T to be either large or small

FIGURE 4.2: Acceptance and rejection regions for a Z-test with (a) a one-sided right-tail alternative; (b) a one-sided left-tail alternative; (c) a two-sided alternative.

(although it certainly depends on how we choose T). If this is the



(a) Right-tail Z-test

(b) Left-tail Z-test

(c) Two-sided Z-test

case, it tells us exactly when we should reject the null hypothesis:
(a) For a right-tail alternative, the rejection region $\mathfrak{R}$ should consist of large values of T. Choose $\mathfrak{R}$ on the right, $\mathfrak{A}$ on the left (Figure 4.2a).
(b) For a left-tail alternative, the rejection region $\mathfrak{R}$ should consist of small values of T. Choose $\mathfrak{R}$ on the left, $\mathfrak{A}$ on the right (Figure 4.2b).
(c) For a two-sided alternative, the rejection region $\mathfrak{R}$ should consist of very small and very large values of T. Let $\mathfrak{R}$ consist of two extreme regions, while $\mathfrak{A}$ covers the middle (Figure 4.2c).

4.5 Standard Normal null distribution (Z-test)

An important case, in terms of a large number of applications, is when the null distribution of the test statistic is Standard Normal.

The test in this case is called a **Z**-test, and the test statistic is usually denoted by Z.
(a) A level α test with a right-tail alternative should

$$\begin{cases} \text{reject } H_0 & \text{if} \quad Z \geq z_\alpha \\ \text{accept } H_0 & \text{if} \quad Z < z_\alpha \end{cases} \tag{4.2}$$

The rejection region in this case consists of large values of Z only,

$$\mathfrak{R} = [z_\alpha, +\infty)\,, \quad \mathfrak{A} = (-\infty, z_\alpha)$$

(see Figure 4.2a).
Under the null hypothesis, Z belongs to $\mathfrak{A}$ and we reject the null hypothesis with probability

$$\boldsymbol{P}\left\{T \geq z_\alpha \mid H_0\right\} = 1 - \Phi\left(z_\alpha\right) = \alpha$$

making the probability of false rejection (type I error) equal α.
For example, we use this acceptance region to test the population mean,

$$H_0 : \mu = \mu_0 \quad \text{vs } H_A : \mu > \mu_0$$

(b) With a left-tail alternative, we should

$$\begin{cases} \text{reject } H_0 & \text{if} \quad Z \leq -z_\alpha \\ \text{accept } H_0 & \text{if} \quad Z > -z_\alpha \end{cases} \tag{4.3}$$

The rejection region consists of small values of Z only,

$$\mathfrak{R} = (-\infty, -z_\alpha]\,, \quad \mathfrak{A} = (-z_\alpha, +\infty)$$

Similarly, $\boldsymbol{P}\{Z \in \mathfrak{R}\} = \alpha$ under H_0; thus, the probability of type I error equals α.
For example, this is how we should test

$$H_0 : \mu = \mu_0 \quad \text{ vs } H_A : \mu < \mu_0$$

(c) With a two-sided alternative, we

$$\begin{cases} \text{reject } H_0 & \text{if} \quad |Z| \geq z_{\alpha/2} \\ \text{accept } H_0 & \text{if} \quad |Z| < z_{\alpha/2} \end{cases} \tag{4.4}$$

The rejection region consists of very small and very large values of Z,

$$\mathfrak{R} = \left(-\infty, z_{\alpha/2}\right] \cup \left[z_{\alpha/2}, +\infty\right)\,, \quad A = \left(-z_{\alpha/2}, z_{\alpha/2}\right)$$

Again, the probability of type I error equals α in this case.
For example, we use this test for

$$H_0 : \mu = \mu_0 \quad \text{ vs } H_A : \mu \neq \mu_0$$

This is easy to remember:

- for a two-sided test, divide α by two and use $z_{\alpha/2}$;
- for a one-sided test, use z_α keeping in mind that the rejection region consists of just one piece.

Now consider testing a hypothesis about a population parameter θ. Suppose that its estimator $\widehat{\theta}$ has Normal distribution, at least approximately, and we know $\mathbf{E}(\widehat{\theta})$ and $\mathrm{Var}(\widehat{\theta})$ if the hypothesis is true.
Then the test statistic

$$Z = \frac{\widehat{\theta} - \mathbf{E}(\widehat{\theta})}{\sqrt{\mathrm{Var}(\widehat{\theta})}} \tag{4.5}$$

has Standard Normal distribution, and we can use (4.2), (4.3), and (4.4) to construct acceptance and rejection regions for a level α test. We call Z a Z-statistic.
Examples of Z-tests are in the next section.

4.6 Z-tests for means and proportions

As we already know,

- sample means have Normal distribution when the distribution of data is Normal;
- sample means have approximately Normal distribution when they are computed from large samples (the distribution of data can be arbitrary);

- sample proportions have approximately Normal distribution when they are computed from large samples;
- this extends to differences between means and between proportions For all these cases, we can use a Z-statistic (4.5) and rejection regions (4.2)-(4.4) to design powerful level α tests.

Example 4.4 (Z-test about a population mean.)
The number of concurrent users for some internet service provider has always averaged 5000 with a standard deviation of 800. After an equipment upgrade, the average number of users at 100 randomly selected moments of time is 5200 . Does it indicate, at a 5% *level of significance, that the mean number of concurrent users has increased? Assume that the standard deviation of the number of concurrent users has not changed.*

Solution. We test the null hypothesis $H_0 : \mu = 5000$ against a one-sided right-tail alternative $H_A : \mu > 5000$, because we are only interested to know if the mean number of users μ has increased.
Step 1: Test statistic. We are given: $\sigma = 800, n = 100, \alpha = 0.05, \mu_0 = 5000$, and from the sample, $\bar{X} = 5200$. The test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{5200 - 5000}{800/\sqrt{100}} = 2.5$$

Step 2: Acceptance and rejection regions. The critical value is

$$z_\alpha = z_{0.05} = 1.645$$

(don't divide α by 2 because it is a one-sided test). With the right-tail alternative, we

$$\begin{cases} \text{reject } H_0 & \text{if} \quad Z \geq 1.645 \\ \text{accept } H_0 & \text{if} \quad Z < 1.645 \end{cases}$$

Step 3: Result. Our test statistic $Z = 2.5$ belongs to the rejection region; therefore, we reject the null hypothesis. The data (5200 users, on the average, at 100 times) provided sufficient evidence in favor of the alternative hypothesis that the mean number of users has increased.

Null hypothesis	Parameter, estimator	If H_0 is true:		Test statistic $Z = \frac{\widehat{\theta}-\theta_0}{\sqrt{\mathrm{Var}(\widehat{\theta})}}$
H_0	$\theta, \widehat{\theta}$	$\mathbf{E}(\widehat{\theta})$	$\mathrm{Var}(\widehat{\theta})$	
One-sample Z-tests for means and proportions, based on a sample of size n				
$\mu = \mu_0$	$\mu, \bar{X}$	μ_0	$\frac{\sigma^2}{n}$	$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
$p = p_0$	$p, \widehat{p}$	p_0	$\frac{p_0(1-p_0)}{n}$	$\frac{\widehat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Two-sample Z-tests comparing means and proportions of two populations, based on independent samples of size n and m				
$\mu_X - \mu_Y = D$	$\mu_X - \mu_Y$ $\bar{X} - \bar{Y}$	D	$\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$	$\frac{\bar{X}-\bar{Y}-D}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}$
$p_1 - p_2 = D$	$p_1 - p_2$ $\widehat{p}_1 - \widehat{p}_2$	D	$\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}$	$\frac{\widehat{p}_1-\widehat{p}_2-D}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n}+\frac{\widehat{p}_2(1-\widehat{p}_2)}{m}}}$
$p_1 = p_2$	$p_1 - p_2$ $\widehat{p}_1 - \widehat{p}_2$	0	$p(1-p)\left(\frac{1}{n}+\frac{1}{m}\right)$ where $p = p_1 = p_2$	$\frac{\widehat{p}_1-\widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}}$ where $\widehat{p} = \frac{n\widehat{p}_1+m\widehat{p}_2}{n+m}$

TABLE 4.1: Summary of Z-tests.

Example 4.5 (Two-SAmple Z-test of Proportions.) *A quality inspector finds 10 defective parts in a sample of 500 parts received from* *manufacturer A. Out of 400 parts from manufacturer B, she finds 12 defective ones. A computer-making company uses these parts in their computers and claims that the quality of parts produced by A and B is the same. At the* 5% *level of significance, do we have enough evidence to disprove this claim?*

Solution. We test $H_0 : p_A = p_B$, or $H_0 : p_A - p_B = 0$, against $H_A : p_A \neq p_B$. This is a two-sided test because no direction of the alternative has been indicated. We only need to verify whether or not the proportions of defective parts are equal for manufacturers A and B.

Step 1: Test statistic. We are given: $\widehat{p}_A = 10/500 = 0.02$ from a sample of size $n = 500$; $\widehat{p}_B = 12/400 = 0.03$ from a sample of size $m = 400$. The tested value is $D = 0$.

As we know, for these Bernoulli data, the variance depends on the unknown parameters p_A and p_B which are estimated by the sample proportions $\widehat{p}_A$ and $\widehat{p}_B$. The test statistic then equals

$$Z = \frac{\widehat{p}_A - \widehat{p}_B - D}{\sqrt{\frac{\widehat{p}_A(1-\widehat{p}_A)}{n} + \frac{\widehat{p}_B(1-\widehat{p}_B)}{m}}} = \frac{0.02 - 0.03}{\sqrt{\frac{(0.02)(0.98)}{500} + \frac{(0.03)(0.97)}{400}}} = -0.945$$

Step 2: Acceptance and rejection regions. This is a two-sided test; thus we divide α by 2 , find $z_{0.05/2} = z_{0.025} = 1.96$, and

$$\begin{cases} \text{reject } H_0 & \text{if} \quad |Z| \geq 1.96 \\ \text{accept } H_0 & \text{if} \quad |Z| < 1.96 \end{cases}$$

Step 3: Result. The evidence against H_0 is insufficient because $|Z| < 1.96$. Although sample proportions of defective parts are unequal, the difference between them appears too small to claim that population proportions are different.

Chapter 5

Variance estimator and Chi-square Distribution

In this section, we'll derive confidence intervals and tests for the population variance $\sigma^2 = \mathrm{Var}(X)$ and for the comparison of two variances $\sigma_X^2 = \mathrm{Var}(X)$ and $\sigma_Y^2 = \mathrm{Var}(Y)$. This will be a new type of inference for us because
(a) variance is a scale and not a location parameter,
(b) the distribution of its estimator, the sample variance, is not symmetric.

Variance often needs to be estimated or tested for quality control, in order to assess stability and accuracy, evaluate various risks, and also, for tests and confidence intervals for the population means when variance is unknown.

We start by estimating the population variance $\sigma^2 = \mathrm{Var}(X)$ from an observed sample $\boldsymbol{X} = (X_1, \ldots, X_n)$. Recall that σ^2 is estimated unbiasedly and consistently by the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

The summands $\left(X_i - \bar{X}\right)^2$ are not quite independent, as the Central Limit Theorem requires, because they all depend on $\bar{X}$. Nevertheless, the distribution of s^2 is approximately Normal, under mild conditions, when the sample is large.

For small to moderate samples, the distribution of s^2 is not Normal at all. It is not even symmetric. Indeed, why should it be symmetric if s^2 is always non-negative!

5.1 Distribution of the sample variance

When observations $X_1, \ldots, X_n$ are independent and Normal with $\mathrm{Var}\,(X_i) = \sigma^2$, the distribution of

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2$$

is Chi-square with $(n-1)$ degrees of freedom

Chi-square distribution, or χ^2, is a continuous distribution with density

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad x > 0$$

where $\nu > 0$ is a parameter that is called degrees of freedom.

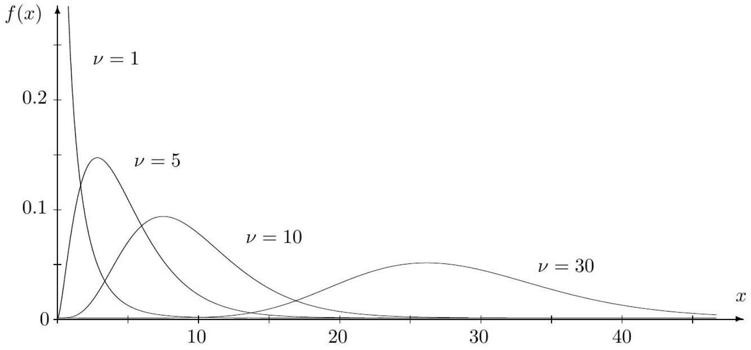


FIGURE 12: Chi-square densities with $\nu = 1, 5, 10$, and 30 degrees of freedom.

Each distribution is right-skewed. For large ν, it is approximately Normal. We see that Chi-square distribution is a special case of Gamma,

$$\text{Chi-square } (\nu) = \text{Gamma}(\nu/2, 1/2)$$

and in particular, the Chi-square distribution with $\nu = 2$ degrees of freedom is Exponential(1/2).

We already know that Gamma(α, λ) distribution has expectation $\mathbf{E}(X) = \alpha/\lambda$ and $\text{Var}(X) = \alpha/\lambda^2$. Substituting $\alpha = \nu/2$ and $\lambda = 1/2$, we get the Chi-square moments,

$$\mathbf{E}(X) = \nu \quad \text{and} \quad \text{Var}(X) = 2\nu$$

5.2 Chi-square distribution (χ^2)

$$\begin{aligned} \nu &= \text{ degrees of freedom} \\ f(x) &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, x > 0 \\ \mathbf{E}(X) &= \nu \\ \mathrm{Var}(X) &= 2\nu \end{aligned} \tag{5.1}$$

Table A6 in the Appendix contains critical values of the Chi-square distribution.

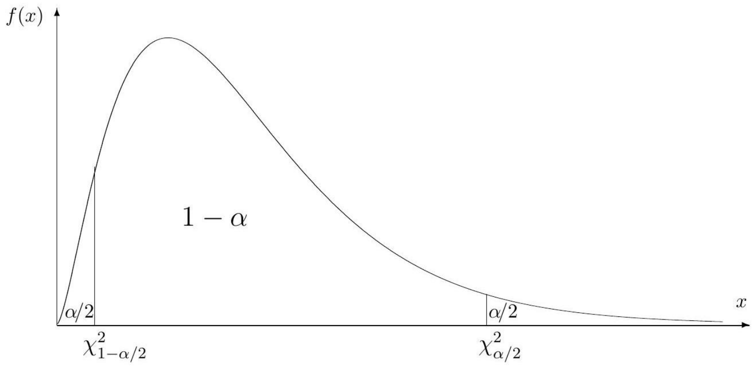


FIGURE 13: Critical values of the Chi-square distribution.

5.2.1 Confidence interval for the population variance

Let us construct a $(1-\alpha)100\%$ confidence interval for the population variance σ^2, based on a sample of size n.

As always, we start with the estimator, the sample variance s^2. However, since the distribution of s^2 is not symmetric, our confidence interval won't have the form "estimator ± margin" as before.

Instead, we use Table A6 to find the critical values $\chi^2_{1-\alpha/2}$ and $\chi^2_{\alpha/2}$ of the Chi-square distribution with $\nu = n-1$ degrees of freedom. These

critical values chop the areas of $(\alpha/2)$ on the right and on the left sides of the region under the Chi-square density curve, as on Figure 13. This is similar to $\pm z_{\alpha/2}$ and $\pm t_{\alpha/2}$ in the previous sections, although these Chisquare quantiles are no longer symmetric. Recall that $\chi^2_{\alpha/2}$ denotes the $(1-\alpha/2)$-quantile, $q_{1-\alpha/2}$.
Then, the area between these two values is $(1-\alpha)$.
A rescaled sample variance $(n-1)s^2/\sigma^2$ has χ^2 density like the one on Figure 13, so

$$\boldsymbol{P}\left\{\chi^2_{1-\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\alpha/2}\right\} = 1-\alpha$$

Solving the inequality for the unknown parameter σ^2, we get

$$\boldsymbol{P}\left\{\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right\} = 1-\alpha$$

A $(1-\alpha)100\%$ confidence interval for the population variance is obtained!

5.2.2 Confidence interval for the variance

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right] \tag{5.2}$$

A confidence interval for the population standard deviation $\sigma = \sqrt{\sigma^2}$ is

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}\right] \tag{5.3}$$

Example 5.1 *A sample of* 6 *measurements* $2.5, 7.4, 8.0, 4.5, 7.4, 9.2$ *is collected from a Normal distribution with mean* μ *and standard deviation* $\sigma = 2.2$. *Let us now rely on the data only and construct a* 90% *confidence interval for the standard deviation. The sample contained* $n = 6$ *measurements,* $2.5, 7.4, 8.0, 4.5, 7.4,$ *and* 9.2.

Solution. *Compute the sample mean and then the sample variance,*

$$\bar{X} = \frac{1}{6}(2.5 + \ldots + 9.2) = 6.5$$
$$s^2 = \frac{1}{6-1}\left\{(2.5 - 6.5)^2 + \ldots + (9.2 - 6.5)^2\right\} = \frac{31.16}{5} = 6.232$$

(actually, we only need $(n-1)s^2 = 31.16$ *).*
From Table A6 of Chi-square distribution with $\nu = n - 1 = 5$ *degrees of freedom, we find the critical values* $\chi^2_{1-\alpha/2} = \chi^2_{0.95} = 1.15$ *and* $\chi^2_{\alpha/2} = \chi^2_{0.05} = 11.1$. *Then,*

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}\right] = \left[\sqrt{\frac{31.16}{11.1}}, \sqrt{\frac{31.16}{1.15}}\right] = [1.68, 5.21]$$

is a 90% *confidence interval for the population standard deviation (and by the way,* $\left[1.68^2, 5.21^2\right] = [2.82, 27.14]$ *is a* 90% *confidence interval for the variance).*

5.2.3 Comparison of two variances. F-Distribution

In this section, we deal with two populations whose variances need to be compared. Such inference is used for the comparison of accuracy, stability, uncertainty, or risks arising in two populations.

Example 5.2 (Efficient upgrade.) *A data channel has the average speed of 180 Megabytes per second. A hardware upgrade is supposed to improve stability of the data transfer while maintaining the same average speed. Stable data transfer rate implies low standard deviation. How can we estimate the relative change in the standard deviation of the transfer rate with* 90% *confidence?*

Example 5.3 (Conservative investment.)
Two mutual funds promise the same expected return; however, one of them recorded a 10% *higher volatility over the last 15 days. Is this a significant evidence for a conservative investor to prefer the other mutual fund? (Volatility is essentially the standard deviation of returns.)*

Example 5.4 (Which method to use?.) *For marketing purposes, a survey of users of two operating systems is conducted. Twenty users of operating system ABC record the average level of satisfaction of 77 on a 100-point scale, with a sample variance of 220 . Thirty users of operating system DEF have the average satisfaction level 70 with a sample variance of 155 . We already know from Section 9.4.8 how to compare the mean satisfaction levels. But what method should we choose? Should we assume equality of population variances,* $\sigma_X^2 = \sigma_Y^2$ *and use the pooled variance? Or we should allow for* $\sigma_X^2 \neq \sigma_Y^2$ *and use Satterthwaite approximation?*

To compare variances or standard deviations, two independent samples $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ are collected, one from each population, as on Figure 4. Unlike population means or proportions, variances are scale factors, and they are compared through their ratio

$$\theta = \frac{\sigma_X^2}{\sigma_Y^2}$$

A natural estimator for the ratio of population variances $\theta = \sigma_X^2/\sigma_Y^2$ is the ratio of sample variances

$$\widehat{\theta} = \frac{s_X^2}{s_Y^2} = \frac{\sum\left(X_i - \bar{X}\right)/(n-1)}{\sum\left(Y_i - \bar{Y}\right)/(m-1)} \tag{5.4}$$

The distribution of this statistic, in standard form, after we divide each sample variance in formula (4) by the corresponding population variance, is called the Fisher-Snedecor distribution or simply F-distribution with $(n-1)$ and $(m-1)$ degrees of freedom.

Distribution of the ratio of sample variances:

For independent samples $X_1, \ldots, X_n$ from Normal (μ_X, σ_X) and $Y_1, \ldots,\ Y_m$ from Normal (μ_Y, σ_Y), the standardized ratio of variances

$$F = \frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} = \frac{\sum\left(X_i - \bar{X}\right)^2/\sigma_X^2/(n-1)}{\sum\left(Y_i - \bar{Y}\right)^2/\sigma_Y^2/(m-1)} \tag{5.5}$$

has F-distribution with $(n-1)$ and $(m-1)$ degrees of freedom.

We know from Section 1 that for the Normal data, both s_X^2/σ_X^2 and s_Y^2/σ_Y^2 follow χ^2 distributions. We can now conclude that the ratio of two independent χ^2 variables, each *divided by its degrees of freedom, has F-distribution.* A ratio of two non-negative continuous random variables, any F-distributed variable is also non-negative and continuous.

F-distribution has two parameters, the numerator degrees of freedom and the denominator degrees of freedom. These are degrees of freedom of the sample variances in the numerator and denominator of the F-ratio (5).

Critical values of F-distribution are in Table A7, and we'll use them to construct confidence intervals and test hypotheses comparing two variances.

One question though... Comparing two variances, σ_X^2 and σ_Y^2, should we divide s_X^2 by s_Y^2 or s_Y^2 by s_X^2 ? Of course, both ratios are ok to use, but we have to keep in mind that in the first case we deal with F$(n-1, m-1)$ distribution, and in the second case with F$(m-1, n-1)$. This leads us to an important general conclusion -

$$\text{If } F \text{ has } F\left(\nu_1, \nu_2\right) \text{ distribution, then the distribution of } \frac{1}{F} \text{ is } F\left(\nu_2, \nu_1\right). \tag{5.6}$$

5.3 Confidence interval for the ratio of population variances

Here we construct a $(1-\alpha)100\%$ confidence interval for the parameter $\theta = \sigma_X^2/\sigma_Y^2$. This is about the sixth time we derive a formula for a confidence interval, so we are well familiar with the method, aren't we?

Start with the estimator, $\widehat{\theta} = s_X^2/s_Y^2$. Standardizing it to

$$F = \frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} = \frac{s_X^2/s_Y^2}{\sigma_X^2/\sigma_Y^2} = \frac{\widehat{\theta}}{\theta}$$

we get an F-variable with $(n-1)$ and $(m-1)$ degrees of freedom. Therefore,

$$\boldsymbol{P}\left\{F_{1-\alpha/2}(n-1,m-1) \le \frac{\widehat{\theta}}{\theta} \le F_{\alpha/2}(n-1,m-1)\right\} = 1-\alpha$$

as on Figure 15. Solving the double inequality for the unknown parameter θ, we get

$$\boldsymbol{P}\left\{\frac{\widehat{\theta}}{F_{\alpha/2}(n-1,m-1)} \le \theta \le \frac{\widehat{\theta}}{F_{1-\alpha/2}(n-1,m-1)}\right\} = 1-\alpha$$

Therefore,

$$\begin{aligned}&\left[\frac{\hat{\theta}}{F_{\alpha/2}(n-1,m-1)}, \frac{\hat{\theta}}{F_{1-\alpha/2}(n-1,m-1)}\right] \\ &\quad = \left[\frac{s_X^2/s_Y^2}{F_{\alpha/2}(n-1,m-1)}, \frac{s_X^2/s_Y^2}{F_{1-\alpha/2}(n-1,m-1)}\right] \qquad (5.7)\end{aligned}$$

is a $(1-\alpha)100\%$ confidence interval for $\theta = \sigma_X^2/\sigma_Y^2$.

The critical values $F_{1-\alpha/2}(n-1,m-1)$ and $F_{\alpha/2}(n-1,m-1)$ come from F-distribution with $(n-1)$ and $(m-1)$ degrees of freedom. However, our Table A7 has only small values of α. What can we do about

$F_{1-\alpha/2}(n-1,m-1)$, a critical value with a large area on the right? We can easily compute $F_{1-\alpha/2}(n-1,m-1)$ by making use of statement (6).

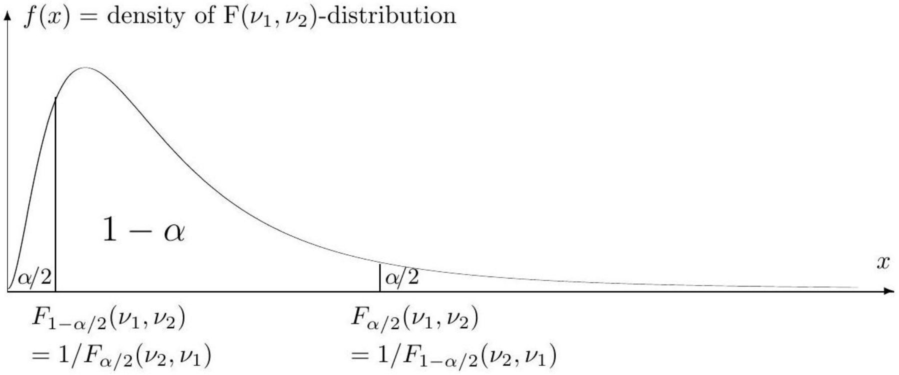


FIGURE 15: Critical values of the F-distribution and their reciprocal property.

Let $F(\nu_1,\nu_2)$ have F-distribution with ν_1 and ν_2 degrees of freedom, then its reciprocal $F(\nu_2,\nu_1) = 1/F(\nu_1,\nu_2)$ has ν_1 and ν_2 degrees of freedom. According to (6),

$$\alpha = \boldsymbol{P}\left\{F(\nu_1,\nu_2) \le F_{1-\alpha}(\nu_1,\nu_2)\right\} = \boldsymbol{P}\left\{F(\nu_2,\nu_1) \ge \frac{1}{F_{1-\alpha}(\nu_1,\nu_2)}\right\}$$

We see from here that $1/F_{1-\alpha}(\nu_1,\nu_2)$ is actually the α-critical value from $F(\nu_2,\nu_1)$ distribution because it cuts area α on the right; see Figure 15. We conclude that

Reciprocal property of F-distribution

The critical values of $F(\nu_1,\nu_2)$ and $F(\nu_2,\nu_1)$ distributions are related as follows,

$$F_{1-\alpha}(\nu_1,\nu_2) = \frac{1}{F_{\alpha}(\nu_2,\nu_1)} \tag{5.8}$$

We can now obtain the critical values from Table A7 and formula (8), plug them into (7), and the confidence interval is ready.

Confidence interval for the ratio of variances

$$\left[\frac{s_X^2}{s_Y^2 F_{\alpha/2}(n-1,m-1)}, \frac{s_X^2 F_{\alpha/2}(m-1,n-1)}{s_Y^2}\right] \tag{5.9}$$

Example 5.5 (Efficient upgrade, continued.) *Refer to Example 9.43. After the upgrade, the instantaneous speed of data transfer, measured at 16 random instants, yields a standard deviation of 14 Mbps . Records show that the standard deviation was 22 Mbps before the upgrade, based on 27 measurements at random times. We are asked to construct a* 90% *confidence interval for the relative change in the standard deviation (assume Normal distribution of the speed).*

Solution. *From the data,* $s_X = 14, s_Y = 22, n = 16$*, and* $m = 27$*. For a* 90% *confidence interval, use* $\alpha = 0.10, \alpha/2 = 0.05$*. Find* $F_{0.05}(15,26) \approx 2.07$ *and* $F_{0.05}(26,15) \approx 2.27$ *from Table A7. Or, alternatively, use functions qf* $(0.95,15,26)$*, qf* $(0.95,26,15)$ *in R or finv* $(0.95,15,26)$*, finv* $(0.95,26,15)$ *in MATLAB to get the exact values, 2.0716 and 2.2722. Then, the* 90% *confidence interval for the ratio of variances* $\theta = \sigma_X^2/\sigma_Y^2$ *is*

$$\left[\frac{14^2}{22^2 \cdot 2.07}, \frac{14^2 \cdot 2.27}{22^2}\right] = [0.20, 0.92]$$

For the ratio of standard deviations $\sigma_X/\sigma_Y = \sqrt{\theta}$*, a* 90% *confidence* *interval is obtained by simply taking square roots,*

$$[\sqrt{0.20}, \sqrt{0.92}] = \underline{[0.44, 0.96]}.$$

Thus, we can assert with a 90% *confidence that the new standard deviation is between* 44% *and* 96% *of the old standard deviation. With this confidence level, the relative reduction in the standard deviation of the data transfer rate (and therefore, the relative increase of stability) is between* 4% *and* 56% *because this relative reduction is* $(\sigma_Y - \sigma_X)/\sigma_Y = 1 - \sqrt{\theta}$.

Example 5.6 (Efficient upgrade, continued again.) *Refer again to Examples 2 and 5. Can we infer that the channel became twice as stable as it was, if increase of stability is measured by the proportional reduction of standard deviation?*

Solution. *The* 90% *confidence interval obtained in Example 9.46 contains 0.5. Therefore, at the* 10% *level of significance, there is no evidence against* $H_0 : \sigma_X/\sigma_Y = 0.5$*, which is a two-fold reduction of standard deviation (recall Section 9.4.9 about the duality between confidence intervals and tests). This is all we can state - there is no evidence against the claim of a two-fold increase of stability. There is no "proof" that it actually happened.*

Testing hypotheses about the ratio of variances or standard deviations is in the next section.

5.4 F-tests comparing two variances

In this section, we test the null hypothesis about a ratio of variances

$$H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = \theta_0 \tag{5.10}$$

against a one-sided or a two-sided alternative. Often we only need to know if two variances are equal, then we choose $\theta_0 = 1$. F-distribution is used to compare variances, so this test is called the **F**-test.

The test statistic for (10) is

$$F = \frac{s_X^2}{s_Y^2} / \theta_0$$

Null Hypothesis $H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = \theta_0$		Test statistic $F_{\text{obs}} = \frac{s_X^2}{s_Y^2}/\theta_0$
Alternative Hypothesis	Rejection region	Use F$(n-1, m-1)$ distribution
$\frac{\sigma_X^2}{\sigma_Y^2} > \theta_0$	$F_{\text{obs}} \geq F_\alpha(n-1, m-1)$	$\boldsymbol{P}\{F \geq F_{\text{obs}}\}$
$\frac{\sigma_X^2}{\sigma_Y^2} < \theta_0$	$F_{\text{obs}} \leq F_\alpha(n-1, m-1)$	$\boldsymbol{P}\{F \leq F_{\text{obs}}\}$
$\frac{\sigma_X^2}{\sigma_Y^2} \neq \theta_0$	$F_{\text{obs}} \geq F_{\alpha/2}(n-1, m-1)$ or $F_{\text{obs}} < 1/F_{\alpha/2}(m-1, n-1)$	$2\min\left(\boldsymbol{P}\{F \geq F_{\text{obs}}\}, \boldsymbol{P}\{F \leq F_{\text{obs}}\}\right)$

TABLE 6: Summary of F-tests for the ratio of population variances.

which under the null hypothesis equals

$$F = \frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2}$$

If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are samples from Normal distributions, this F-statistic has F-distribution with $(n-1)$ and $(m-1)$ degrees of freedom.

Just like χ^2, F-statistic is also non-negative, with a non-symmetric right-skewed distribution. Level α tests and P-values are then developed similarly to χ^2, see Table 9.6.

Example 5.7 (Which Method to USE? Continued.) *In Example 4,* $n = 20, \bar{X} = 77, s_X^2 = 220; m = 30, \bar{Y} = 70,$ *and* $s_Y^2 = 155$*. To compare the population means by a suitable method, we have to test whether the two population variances are equal or not.*

Solution. *Test* $H_0 : \sigma_X^2 = \sigma_Y^2$ *vs* $H_A : \sigma_X^2 \neq \sigma_Y^2$ *with the test statistic*

$$F_{obs} = \frac{s_X^2}{s_Y^2} = 1.42$$

For testing equality of variances, we let the tested ratio $\theta_0 = 1$*. This is a two-sided test, so the* P *-value is*

$$P = 2\min(\boldsymbol{P}\{F \geq 1.42\}, \boldsymbol{P}\{F \leq 1.42\}) = \ldots?$$

How to compute these probabilities for the F-distribution with $n-1 = 19$ *and* $m-1 = 29$ *degrees of freedom? R and MATLAB, as always, can give us the exact answer. Typing 1-pf* $(1.42, 19, 29)$ *in R or* $1-$ $\mathrm{fcdf}(1.42, 19, 29)$ *in MATLAB, we obtain* $\boldsymbol{P}\{F \geq 1.42\} =$ *0.1926 . Then,*

$$P = 2\min(0.1926, 1 - 0.1926) = \underline{0.3852}$$

Table A7 can also be used, for an approximate but a completely satisfactory solution. This table does not have exactly 19 and 29 degrees of freedom and does not have a value $F_\alpha = 1.42$*. However, looking at 15* *and 20 d.f. for the numerator and 25 and 30 d.f. for the denominator, we see that 1.42 is always between $F_{0.25}$ and $F_{0.1}$. This will do it for us.*

It implies that $\boldsymbol{P}\{F \geq .42\} \in (0.1, 0.25), \boldsymbol{P}\{F \leq 1.42\} \in (0.75, 0.9)$, and therefore, the P -value is

$$P = 2\boldsymbol{P}\{F \geq 1.42\} \in (0.2, 0.5)$$

This is a high P-value showing no evidence of different variances. It should be ok to use the exact two-sample T-test with a pooled variance (according to which there is a mild evidence at a 4% level that the first operating system is better, $t = 1.80, P = 0.0388$).

Example 5.8 (Are all the Conditions MET?.) *In Example 3, we are asked to compare volatilities of two mutual funds and decide if one of them is more risky than the other. So, this is a one-sided test of*

$$H_0 : \sigma_X = \sigma_Y \quad vs \quad H_A : \sigma_X > \sigma_Y$$

The data collected over the period of 30 days show a 10% higher volatility of the first mutual fund, i.e., $s_X/s_Y = 1.1$. So, this is a standard F-test, right? A careless statistician would immediately proceed to the test statistic $F_{obs} = s_X^2/s_Y^2 = 1.21$ and the P-value $P = P\{F \geq F_{obs}\} \geq 0.25$ from Table A7 with $n - 1 = 29$ and $m - 1 = 29$ d.f., and jump to a conclusion that there is no evidence that the first mutual fund carries a higher risk.

Indeed, why not? Well, every statistical procedure has its assump-

tions, conditions under which our conclusions are valid. A careful statistician always checks the assumptions before reporting any results.

If we conduct an F-test and refer to the F-distribution, what conditions are required? We find the answer in (5). Apparently, for the F-statistic to have F-distribution under H_0, each of our two samples has to consist of independent and identically distributed Normal random variables, and the two samples have to be independent of each other.

Are these assumptions plausible, at the very least?

1. *Normal distribution - may be. Returns on investments are typically not Normal but log-returns are.*

2. *Independent and identically distributed data within each sample - unlikely. Typically, there are economic trends, ups and downs, and returns on two days in a row should be dependent.*

3. *Independence between the two samples - it depends. If our mutual funds contain stocks from the same industry, their returns are surely dependent.*

Actually, conditions 1-3 can be tested statistically, and for this we need to have the entire samples of data instead of the summary statistics.

The F-test is quite robust. It means that a mild departure from the assumptions 1-3 will not affect our conclusions severely, and we can treat our result as approximate. However, if the assumptions are not met even approximately, for example, the distribution of our data is *asymmetric and far from Normal, then the P -value computed above is simply wrong.*

Discussion in Example 8 leads us to a very important practical conclusion.

Every statistical procedure is valid under certain assumptions. When they are not satisfied, the obtained results may be wrong and misleading. Therefore, unless there are reasons to believe that all the conditions are met, they have to be tested statistically.

Chapter 6

Regression

In this chapter, we study relations among variables. Many variables observed in real life are related. The type of their relation can often be expressed in a mathematical form called regression. Establishing and testing such a relation enables us:

- to understand interactions, causes, and effects among variables;
- to predict unobserved variables based on the observed ones;
- to determine which variables significantly affect the variable of interest.

6.1 Least squares estimation

Regression models relate a response variable to one or several predictors. Having observed predictors, we can forecast the response by computing its conditional expectation, given all the available predictors.

Definition 6.1 ***Response*** *or dependent variable Y is a variable of interest that we predict based on one or several predictors.*

Predictors *or independent variables $X^{(1)}, \dots, X^{(k)}$ are used to predict the values and behavior of the response variable Y.*

***Regression** of Y on $X^{(1)}, \ldots, X^{(k)}$ is the conditional expectation,*

$$G\left(x^{(1)}, \ldots, x^{(k)}\right) = \mathbf{E}\left\{Y \mid X^{(1)} = x^{(1)}, \ldots, X^{(k)} = x^{(k)}\right\}$$

It is a function of $x^{(1)}, \ldots, x^{(k)}$ whose form can be estimated from data.

Examples

Consider several situations when we can predict a dependent variable of interest from independent predictors.

Example 6.1 *Example 11.1 (World population). According to the International Data Base of the U.S. Census Bureau, population of the world grows according to Table 11.1 and data set PopulationWorld. How can we use these data to predict the world population in years 2020 and* 2030?

Figure 1 shows that the population (response) is tightly related to the year (predictor),

$$\textit{population} \ \approx G \ (\textit{year})$$

Population increases every year, and its growth is almost linear. If we estimate the regression function G (the dotted line on Figure 11.1) relating the response and the predictor and extend its graph to the year 2030, the forecast will be ready. We can simply compute $G(2020)$ *and* $G(2030)$.

Year	Population mln. people	Year	Population mln. people	Year	Population mln. people	Year	Population mln. people
1950	2557	1970	3708	1990	5273	2010	6835
1955	2781	1975	4084	1995	5682	2015	7226
1960	3041	1980	4447	2000	6072	2020	?
1965	3347	1985	4844	2005	6449	2030	?

TABLE 1: Population of the world, 1950-2030.

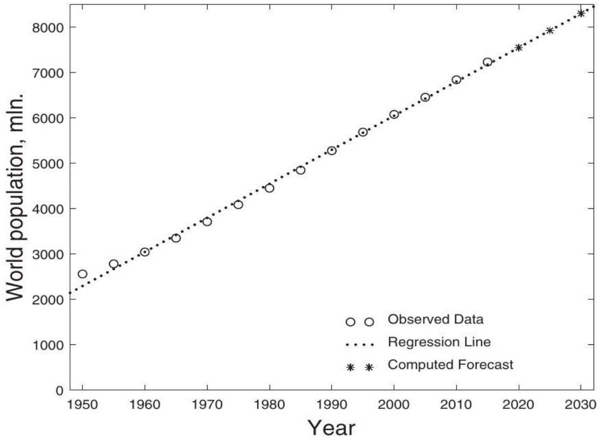


FIGURE 1: World population in 1950-2019 and its regression forecast until 2030.

A straight line that fits the observed data for years 1950-2015 predicts the population of 7.54 billion in 2020, 7.92 billion in 2025 , and 8.29 billion in 2030. It also shows that between 2025 and 2030, around the year 2026, the world population reaches the historical mark of 8 billion.

How accurate is the forecast obtained in this example? The observed population during 1950-2019 appears to grow rather closely to the estimated regression line in Figure 1. It is reasonable to hope that it will continue to do so through 2030.
The situation is different in the next example.

Example 6.2 (House Prices.) *Seventy house sale prices in a certain* *county are depicted in Figure 2 along with the house area.*

***First**, we see a clear relation between these two variables, and in general, bigger houses are more expensive. However, the trend no longer seems linear.*

***Second**, there is a large amount of variability around this trend. Indeed, area is not the only factor determining the house price. Houses with the same area may still be priced differently.*

Then, how can we estimate the price of a 3200 -square-foot house? We can estimate the general trend (the dotted line in Figure 11.2) and plug 3200 into the resulting formula, but due to obviously high variability, our estimation will not be as accurate as in Example 1.

To improve our estimation in the last example, we may take other factors into account: the number of bedrooms and bathrooms, the backyard area, the average income of the

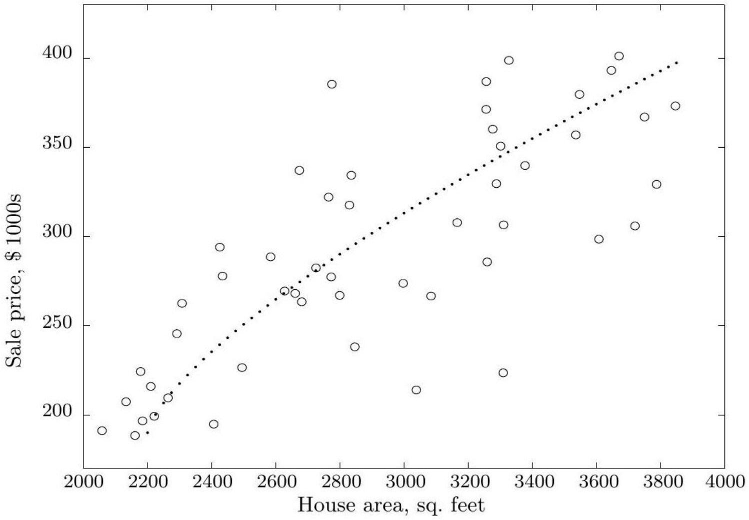


FIGURE 2: House sale prices and their footage.

neighborhood, etc. If all the added variables are relevant for pricing a house, our model will have a closer fit and will provide more accurate predictions.

6.1.1 Method of least squares

Our immediate goal is to estimate the regression function G that connects response variable Y with predictors $X^{(1)}, \ldots, X^{(k)}$. First we focus on univariate regression predicting response Y based on one predictor X. The method will be extended to k predictors later.

In univariate regression, we observe pairs $(x_1, y_1), \ldots, (x_n, y_n)$, shown in Figure 3a.

For accurate forecasting, we are looking for the function $\widehat{G}(x)$ that passes as close as possible to the observed data points. This is achieved by minimizing distances between observed data points $y_1, \ldots, y_n$ and the corresponding points on the fitted regression line,

$$\widehat{y}_1 = \widehat{G}(x_1), \ldots, \widehat{y}_n = \widehat{G}(x_n)$$

(see Figure 3b). Method of least squares minimizes the sum of squared distances.

Definition 6.2

Residuals *$e_i = y_i - \widehat{y}_i$ are differences between observed responses y_i and their fitted values $\widehat{y}_i = \widehat{G}(x_i)$*

Method of least squares *finds a regression function $\widehat{G}(x)$ that min-*

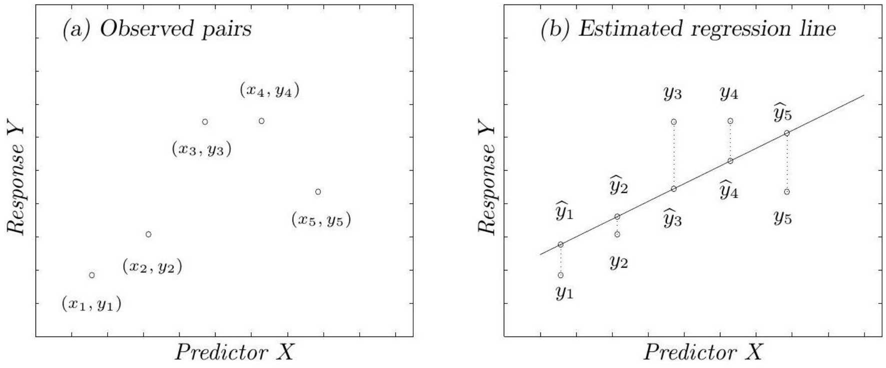


FIGURE 3: Least squares estimation of the regression line.

imizes the sum of squared residuals

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \tag{6.1}$$

Function $\widehat{G}$ is usually sought in a suitable form: linear, quadratic, logarithmic, etc. The simplest form is linear.

6.2 Linear regression

Linear regression model assumes that the conditional expectation

$$G(x) = \mathbf{E}\{Y \mid X = x\} = \beta_0 + \beta_1 x$$

is a linear function of x. As any linear function, it has an intercept β_0 and a slope β_1.

The intercept

$$\beta_0 = G(0)$$

equals the value of the regression function for $x = 0$. Sometimes it has no physical meaning. For example, nobody will try to predict the value of a computer with 0 random access memory (RAM), and nobody will consider the Federal reserve rate in year 0 . In other cases, intercept is quite important. For example, according to the Ohm's Law ($V = RI$) the voltage across an ideal conductor is proportional to the current. A non-zero intercept ($V = V_0 + RI$) would show that the circuit is not ideal, and there is an external loss of voltage.

The slope

$$\beta_1 = G(x+1) - G(x)$$

is the predicted change in the response variable when predictor changes by 1 . This is a very important parameter that shows how fast we can change the expected response by varying the predictor. For example, customer satisfaction will increase by $\beta_1(\Delta x)$ when the quality of produced computers increases by (Δx).

A zero slope means absence of a linear relationship between X and Y. In this case, Y is expected to stay constant when X changes.

6.2.1 Estimation in linear regression

Let us estimate the slope and intercept by method of least squares. Following (1), we minimize the sum of squared residuals

$$Q = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \widehat{G}\left(x_i\right)\right)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

We can do it by taking partial derivatives of Q, equating them to 0 , and solving the resulting equations for β_0 and β_1.
The partial derivatives are

$$\begin{aligned}\frac{\partial Q}{\partial \beta_0} &= -2 \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right) \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right) x_i\end{aligned}$$

Equating them to 0 , we obtain so-called normal equations,

$$\left\{\begin{array}{l}\sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right) = 0 \\ \sum_{i=1}^{n} x_i \left(y_i - \beta_0 - \beta_1 x_i\right) = 0\end{array}\right.$$

From the first normal equation,

$$\beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n} = \bar{y} - \beta_1 \bar{x} \tag{6.2}$$

Substituting this into the second normal equation, we get

$$\begin{aligned}\sum_{i=1}^{n} x_i \left(y_i - \beta_0 - \beta_1 x_i\right) &= \sum_{i=1}^{n} x_i \left(\left(y_i - \bar{y}\right) - \beta_1 \left(x_i - \bar{x}\right)\right) \\ &= S_{xy} - \beta_1 S_{xx} = 0\end{aligned} \tag{6.3}$$

where

$$S_{xx} = \sum_{i=1}^{n} x_i \left(x_i - \bar{x}\right) = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \tag{6.4}$$

and

$$S_{xy} = \sum_{i=1}^{n} x_i \left(y_i - \bar{y}\right) = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) \qquad (6.5)$$

are sums of squares and cross-products. Notice that it is all right to subtract $\bar{x}$ from x_i in the right-hand sides of (4) and (5) because $\sum \left(x_i - \bar{x}\right) = 0$ and $\sum \left(y_i - \bar{y}\right) = 0$. Finally, we obtain the least squares estimates of intercept β_0 and slope β_1 from (2) and (3).

Regression estimates

$$b_0 = \widehat{\beta}_0 = \bar{y} - b_1\bar{x}, \quad b_1 = \widehat{\beta}_1 = S_{xy}/S_{xx}, \qquad (6.6)$$

where

$$S_{xx} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2, \quad S_{xy} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$$

Example 6.3 (World population.) *In Example 1, x_i is the year, and y_i is the world population during that year. To estimate the regression line in Figure 1, we compute*

$$\bar{x} = 1984; \quad \bar{y} = 4843$$
$$S_{xx} = (1950 - \bar{x})^2 + \ldots + (2019 - \bar{x})^2 = 27370$$
$$S_{xy} = (1950 - \bar{x})(2558 - \bar{y}) + \ldots + (2010 - \bar{x})(6864 - \bar{y}) = 2053529$$

Then

$$b_1 = S_{xy}/S_{xx} = 75$$
$$b_0 = \bar{y} - b_1\bar{x} = -144013$$

The estimated regression line is

$$\widehat{G}(x) = b_0 + b_1 x = \underline{-144013 + 75x}.$$

We conclude that the world population grows at the average rate of 75 million every year. We can use the obtained equation to predict the future growth of the world population. Regression predictions for years 2020 and 2030 are

$$\begin{aligned}\widehat{G}(2020) &= b_0 + 2020 b_1 = 7544 \text{ \textit{million people}} \\ \widehat{G}(2030) &= b_0 + 2030 b_1 = \underline{8295 \text{ \textit{million people}}}\end{aligned}$$

6.3 Regression and correlation

Recall, the covariance

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbf{E}\{(X - \mathbf{E}X)(Y - \mathbf{E}Y)\} \\ &= \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)\end{aligned}$$

and correlation coefficient

$$\rho = \frac{\mathrm{Cov}(X,Y)}{(\mathrm{Std}\, X)(\mathrm{Std}\, Y)}$$

measure the direction and strength of a linear relationship between variables X and Y.

Properties

$\mathrm{Var}(aX + bY + c) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y)$

$\mathrm{Cov}(aX + bY, cZ + dW)$

$\quad$=a c $\mathrm{Cov}(X, Z) + ad\,\mathrm{Cov}(X, W) + bc\,\mathrm{Cov}(Y, Z) + bd\,\mathrm{Cov}(Y, W)$

$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$

$\rho(X, Y) \quad = \rho(Y, X),\ -1 \le \rho \le 1$

Example 6.4 *Given*

x	$P_X(x)$	$xP_X(x)$	$x - \mathbf{E}X$	$(x - \mathbf{E}X)^2 P_X(x)$
0	0.5	0	-0.5	0.125
1	0.5	0.5	0.5	0.125
$\mu_X = 0.5$			$\sigma_X^2 = 0.25$	

and

y	$P_Y(y)$	$yP_Y(y)$	y^2	$y^2 P_Y(y)$
0	0.4	0	0	0
1	0.3	0.3	1	0.3
2	0.15	0.3	4	0.6
3	0.15	0.45	9	1.35
$\mu_Y = 1.05$			$\mathbf{E}\left(Y^2\right) = 2.25$	

Result: $\operatorname{Var}(X)=0.25, \operatorname{Var}(Y)=2.25-1.05^2=1.1475, \operatorname{Std}(X)=\sqrt{0.25}=0.5$, and $\operatorname{Std}(Y)=\sqrt{1.1475}=1.0712$. Also,

$$\mathbf{E}(XY)=\sum_x \sum_y xyP(x,y)=(1)(1)(0.1)+(1)(2)(0.1)+(1)(3)(0.1)=0.6$$

(the other five terms in this sum are 0). Therefore,

$$\operatorname{Cov}(X,Y)=\mathbf{E}(XY)-\mathbf{E}(X)\mathbf{E}(Y)=0.6-(0.5)(1.05)=0.075$$

and

$$\rho=\frac{\operatorname{Cov}(X,Y)}{(\operatorname{Std} X)(\operatorname{Std} Y)}=\frac{0.075}{(0.5)(1.0712)}=0.1400$$

Thus, the numbers of errors in two modules are positively and not very strongly correlated.

From observed data, we estimate $\operatorname{Cov}(X,Y)$ and ρ by the sample covariance

$$s_{xy}=\frac{\sum_{i=1}^n\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{n-1}$$

(it is unbiased for the population covariance) and the sample correlation coefficient

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

where

$$s_x=\sqrt{\frac{\sum\left(x_i-\bar{x}\right)^2}{n-1}} \quad \text{and} \quad s_y=\sqrt{\frac{\sum\left(y_i-\bar{y}\right)^2}{n-1}}$$

are sample standard deviations of X and Y. Comparing (3) and (7), we see that the estimated slope b_1 and the sample regression coefficient

r are proportional to each other. Now we have two new formulas for the regression slope.

6.3.1 Estimated regression slope

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{s_{xy}}{s_x^2} = r\left(\frac{s_y}{s_x}\right)$$

Like the correlation coefficient, regression slope is positive for positively correlated X and Y and negative for negatively correlated X and Y. The difference is that r is dimensionless whereas the slope is measured in units of Y per units of X. Thus, its value by itself does not indicate whether the dependence is weak or strong. It depends on the units, the scale of X and Y. We test significance of the regression slope in Section 2.

Table A3. Poisson distribution

$$F(x) = \boldsymbol{P}\left\{X \le x\right\} = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$$

x	λ														
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
0	.905	.819	.741	.670	.607	.549	.497	.449	.407	.368	.333	.301	.273	.247	.223
1	.995	.982	.963	.938	.910	.878	.844	.809	.772	.736	.699	.663	.627	.592	.558
2	1.00	.999	.996	.992	.986	.977	.966	.953	.937	.920	.900	.879	.857	.833	.809
3	1.00	1.00	1.00	.999	.998	.997	.994	.991	.987	.981	.974	.966	.957	.946	.934
4	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.996	.995	.992	.989	.986	.981
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.998	.997	.996
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

x	λ														
	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.202	.183	.165	.150	.135	.122	.111	.100	.091	.082	.074	.067	.061	.055	.050
1	.525	.493	.463	.434	.406	.380	.355	.331	.308	.287	.267	.249	.231	.215	.199
2	.783	.757	.731	.704	.677	.650	.623	.596	.570	.544	.518	.494	.469	.446	.423
3	.921	.907	.891	.875	.857	.839	.819	.799	.779	.758	.736	.714	.692	.670	.647
4	.976	.970	.964	.956	.947	.938	.928	.916	.904	.891	.877	.863	.848	.832	.815
5	.994	.992	.990	.987	.983	.980	.975	.970	.964	.958	.951	.943	.935	.926	.916
6	.999	.998	.997	.997	.995	.994	.993	.991	.988	.986	.983	.979	.976	.971	.966
7	1.00	1.00	.999	.999	.999	.999	.998	.997	.997	.996	.995	.993	.992	.990	.988
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998	.998	.997	.996
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

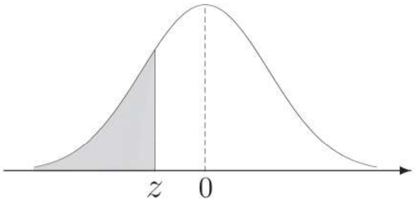
x	λ														
	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5
0	.030	.018	.011	.007	.004	.002	.002	.001	.001	.000	.000	.000	.000	.000	.000
1	.136	.092	.061	.040	.027	.017	.011	.007	.005	.003	.002	.001	.001	.000	.000
2	.321	.238	.174	.125	.088	.062	.043	.030	.020	.014	.009	.006	.004	.003	.002
3	.537	.433	.342	.265	.202	.151	.112	.082	.059	.042	.030	.021	.015	.010	.007
4	.725	.629	.532	.440	.358	.285	.224	.173	.132	.100	.074	.055	.040	.029	.021
5	.858	.785	.703	.616	.529	.446	.369	.301	.241	.191	.150	.116	.089	.067	.050
6	.935	.889	.831	.762	.686	.606	.527	.450	.378	.313	.256	.207	.165	.130	.102
7	.973	.949	.913	.867	.809	.744	.673	.599	.525	.453	.386	.324	.269	.220	.179
8	.990	.979	.960	.932	.894	.847	.792	.729	.662	.593	.523	.456	.392	.333	.279
9	.997	.992	.983	.968	.946	.916	.877	.830	.776	.717	.653	.587	.522	.458	.397
10	.999	.997	.993	.986	.975	.957	.933	.901	.862	.816	.763	.706	.645	.583	.521
11	1.00	.999	.998	.995	.989	.980	.966	.947	.921	.888	.849	.803	.752	.697	.639
12	1.00	1.00	.999	.998	.996	.991	.984	.973	.957	.936	.909	.876	.836	.792	.742
13	1.00	1.00	1.00	.999	.998	.996	.993	.987	.978	.966	.949	.926	.898	.864	.825
14	1.00	1.00	1.00	1.00	.999	.999	.997	.994	.990	.983	.973	.959	.940	.917	.888
15	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.995	.992	.986	.978	.967	.951	.932
16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.996	.993	.989	.982	.973	.960
17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.997	.995	.991	.986	.978
18	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.996	.993	.988
19	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.997	.994
20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.997

Table A3, continued. Poisson distribution

x	λ														
	11	12	13	14	15	16	17	18	19	20	22	24	26	28	30
0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
2	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
3	.005	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
4	.015	.008	.004	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
5	.038	.020	.011	.006	.003	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000
6	.079	.046	.026	.014	.008	.004	.002	.001	.001	.000	.000	.000	.000	.000	.000
7	.143	.090	.054	.032	.018	.010	.005	.003	.002	.001	.000	.000	.000	.000	.000
8	.232	.155	.100	.062	.037	.022	.013	.007	.004	.002	.001	.000	.000	.000	.000
9	.341	.242	.166	.109	.070	.043	.026	.015	.009	.005	.002	.000	.000	.000	.000
10	.460	.347	.252	.176	.118	.077	.049	.030	.018	.011	.004	.001	.000	.000	.000
11	.579	.462	.353	.260	.185	.127	.085	.055	.035	.021	.008	.003	.001	.000	.000
12	.689	.576	.463	.358	.268	.193	.135	.092	.061	.039	.015	.005	.002	.001	.000
13	.781	.682	.573	.464	.363	.275	.201	.143	.098	.066	.028	.011	.004	.001	.000
14	.854	.772	.675	.570	.466	.368	.281	.208	.150	.105	.048	.020	.008	.003	.001
15	.907	.844	.764	.669	.568	.467	.371	.287	.215	.157	.077	.034	.014	.005	.002
16	.944	.899	.835	.756	.664	.566	.468	.375	.292	.221	.117	.056	.025	.010	.004
17	.968	.937	.890	.827	.749	.659	.564	.469	.378	.297	.169	.087	.041	.018	.007
18	.982	.963	.930	.883	.819	.742	.655	.562	.469	.381	.232	.128	.065	.030	.013
19	.991	.979	.957	.923	.875	.812	.736	.651	.561	.470	.306	.180	.097	.048	.022
20	.995	.988	.975	.952	.917	.868	.805	.731	.647	.559	.387	.243	.139	.073	.035
21	.998	.994	.986	.971	.947	.911	.861	.799	.725	.644	.472	.314	.190	.106	.054
22	.999	.997	.992	.983	.967	.942	.905	.855	.793	.721	.556	.392	.252	.148	.081
23	1.00	.999	.996	.991	.981	.963	.937	.899	.849	.787	.637	.473	.321	.200	.115
24	1.00	.999	.998	.995	.989	.978	.959	.932	.893	.843	.712	.554	.396	.260	.157
25	1.00	1.00	.999	.997	.994	.987	.975	.955	.927	.888	.777	.632	.474	.327	.208
26	1.00	1.00	1.00	.999	.997	.993	.985	.972	.951	.922	.832	.704	.552	.400	.267
27	1.00	1.00	1.00	.999	.998	.996	.991	.983	.969	.948	.877	.768	.627	.475	.333
28	1.00	1.00	1.00	1.00	.999	.998	.995	.990	.980	.966	.913	.823	.697	.550	.403
29	1.00	1.00	1.00	1.00	1.00	.999	.997	.994	.988	.978	.940	.868	.759	.623	.476
30	1.00	1.00	1.00	1.00	1.00	.999	.999	.997	.993	.987	.959	.904	.813	.690	.548
31	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.996	.992	.973	.932	.859	.752	.619
32	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.995	.983	.953	.896	.805	.685
33	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.989	.969	.925	.850	.744
34	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.994	.979	.947	.888	.797
35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.996	.987	.964	.918	.843
36	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.992	.976	.941	.880
37	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.995	.984	.959	.911
38	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.990	.972	.935
39	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	.981	.954
40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.996	.988	.968
41	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.992	.978
42	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.995	.985
43	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.990
44	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994
45	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.996
46	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998
47	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
48	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
49	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A4. Standard Normal distribution

$$\Phi(z) = \boldsymbol{P}\left\{Z \leq z\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

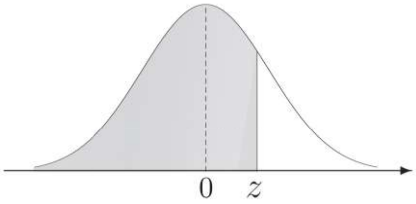


z	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	-0.00
-(3.9+)	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.7	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.6	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
-3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
-3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007
-3.1	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010
-3.0	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
-2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
-2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
-2.7	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
-2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047
-2.5	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062
-2.4	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082
-2.3	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107
-2.2	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139
-2.1	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179
-2.0	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228
-1.9	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287
-1.8	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359
-1.7	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446
-1.6	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548
-1.5	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668
-1.4	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808
-1.3	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968
-1.2	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151
-1.1	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357
-1.0	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587
-0.9	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841
-0.8	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119
-0.7	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420
-0.6	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743
-0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085
-0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446
-0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821
-0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207
-0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602
-0.0	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000

Table A4, continued.

Standard Normal distribution

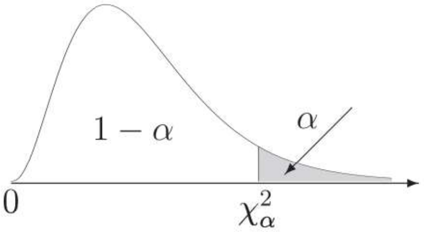
$$\Phi(z) = \boldsymbol{P}\{Z \le z\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9+	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A6. Chi-Square Distribution

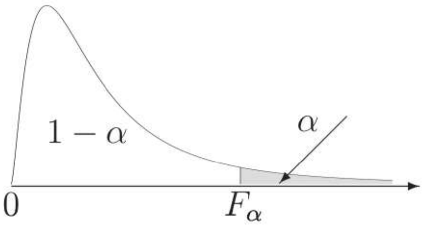
χ^2_α; critical values, such that $\boldsymbol{P}\left\{\chi^2 > \chi^2_\alpha\right\} = \alpha$



ν	α, the right-tail probability													
(d.f.)	.999	.995	.99	.975	.95	.90	.80	.20	.10	.05	.025	.01	.005	.001
1	0.00	0.00	0.00	0.00	0.00	0.02	0.06	1.64	2.71	3.84	5.02	6.63	7.88	10.8
2	0.00	0.01	0.02	0.05	0.10	0.21	0.45	3.22	4.61	5.99	7.38	9.21	10.6	13.8
3	0.02	0.07	0.11	0.22	0.35	0.58	1.01	4.64	6.25	7.81	9.35	11.3	12.8	16.3
4	0.09	0.21	0.30	0.48	0.71	1.06	1.65	5.99	7.78	9.49	11.1	13.3	14.9	18.5
5	0.21	0.41	0.55	0.83	1.15	1.61	2.34	7.29	9.24	11.1	12.8	15.1	16.7	20.5
6	0.38	0.68	0.87	1.24	1.64	2.20	3.07	8.56	10.6	12.6	14.4	16.8	18.5	22.5
7	0.60	0.99	1.24	1.69	2.17	2.83	3.82	9.80	12.0	14.1	16.0	18.5	20.3	24.3
8	0.86	1.34	1.65	2.18	2.73	3.49	4.59	11.0	13.4	15.5	17.5	20.1	22.0	26.1
9	1.15	1.73	2.09	2.70	3.33	4.17	5.38	12.2	14.7	16.9	19.0	21.7	23.6	27.9
10	1.48	2.16	2.56	3.25	3.94	4.87	6.18	13.4	16.0	18.3	20.5	23.2	25.2	29.6
11	1.83	2.60	3.05	3.82	4.57	5.58	6.99	14.6	17.3	19.7	21.9	24.7	26.8	31.3
12	2.21	3.07	3.57	4.40	5.23	6.30	7.81	15.8	18.5	21.0	23.3	26.2	28.3	32.9
13	2.62	3.57	4.11	5.01	5.89	7.04	8.63	17.0	19.8	22.4	24.7	27.7	29.8	34.5
14	3.04	4.07	4.66	5.63	6.57	7.79	9.47	18.2	21.1	23.7	26.1	29.1	31.3	36.1
15	3.48	4.60	5.23	6.26	7.26	8.55	10.3	19.3	22.3	25.0	27.5	30.6	32.8	37.7
16	3.94	5.14	5.81	6.91	7.96	9.31	11.1	20.5	23.5	26.3	28.8	32.0	34.3	39.3
17	4.42	5.70	6.41	7.56	8.67	10.1	12.0	21.6	24.8	27.6	30.2	33.4	35.7	40.8
18	4.90	6.26	7.01	8.23	9.39	10.9	12.9	22.8	26.0	28.9	31.5	34.8	37.2	42.3
19	5.41	6.84	7.63	8.91	10.1	11.7	13.7	23.9	27.2	30.1	32.9	36.2	38.6	43.8
20	5.92	7.43	8.26	9.59	10.9	12.4	14.6	25.0	28.4	31.4	34.2	37.6	40.0	45.3
21	6.45	8.03	8.90	10.3	11.6	13.2	15.4	26.2	29.6	32.7	35.5	38.9	41.4	46.8
22	6.98	8.64	9.54	11.0	12.3	14.0	16.3	27.3	30.8	33.9	36.8	40.3	42.8	48.3
23	7.53	9.26	10.2	11.7	13.1	14.8	17.2	28.4	32.0	35.2	38.1	41.6	44.2	49.7
24	8.08	9.89	10.9	12.4	13.8	15.7	18.1	29.6	33.2	36.4	39.4	43.0	45.6	51.2
25	8.65	10.5	11.5	13.1	14.6	16.5	18.9	30.7	34.4	37.7	40.6	44.3	46.9	52.6
26	9.22	11.2	12.2	13.8	15.4	17.3	19.8	31.8	35.6	38.9	41.9	45.6	48.3	54.1
27	9.80	11.8	12.9	14.6	16.2	18.1	20.7	32.9	36.7	40.1	43.2	47.0	49.6	55.5
28	10.4	12.5	13.6	15.3	16.9	18.9	21.6	34.0	37.9	41.3	44.5	48.3	51.0	56.9
29	11.0	13.1	14.3	16.0	17.7	19.8	22.5	35.1	39.1	42.6	45.7	49.6	52.3	58.3
30	11.6	13.8	15.0	16.8	18.5	20.6	23.4	36.3	40.3	43.8	47.0	50.9	53.7	59.7
31	12.2	14.5	15.7	17.5	19.3	21.4	24.3	37.4	41.4	45.0	48.2	52.2	55.0	61.1
32	12.8	15.1	16.4	18.3	20.1	22.3	25.1	38.5	42.6	46.2	49.5	53.5	56.3	62.5
33	13.4	15.8	17.1	19.0	20.9	23.1	26.0	39.6	43.7	47.4	50.7	54.8	57.6	63.9
34	14.1	16.5	17.8	19.8	21.7	24.0	26.9	40.7	44.9	48.6	52.0	56.1	59.0	65.2
35	14.7	17.2	18.5	20.6	22.5	24.8	27.8	41.8	46.1	49.8	53.2	57.3	60.3	66.6
36	15.3	17.9	19.2	21.3	23.3	25.6	28.7	42.9	47.2	51.0	54.4	58.6	61.6	68
37	16.0	18.6	20,0	22.1	24.1	26.5	29.6	44.0	48.4	52.2	55.7	59.9	62.9	69.3
38	16.6	19.3	20.7	22.9	24.9	27.3	30.5	45.1	49.5	53.4	56.9	61.2	64.2	70.7
39	17.3	20.0	21.4	23.7	25.7	28.2	31.4	46.2	50.7	54.6	58.1	62.4	65.5	72.1
40	17.9	20.7	22.2	24.4	26.5	29.1	32.3	47.3	51.8	55.8	59.3	63.7	66.8	73.4
41	18.6	21.4	22.9	25.2	27.3	29.9	33.3	48.4	52.9	56.9	60.6	65.0	68.1	74.7
42	19.2	22.1	23.7	26.0	28.1	30.8	34.2	49.5	54.1	58.1	61.8	66.2	69.3	76.1
43	19.9	22.9	24.4	26.8	29.0	31.6	35.1	50.5	55.2	59.3	63.0	67.5	70.6	77.4
44	20.6	23.6	25.1	27.6	29.8	32.5	36.0	51.6	56.4	60.5	64.2	68.7	71.9	78.7
45	21.3	24.3	25.9	28.4	30.6	33.4	36.9	52.7	57.5	61.7	65.4	70.0	73.2	80.1
46	21.9	25.0	26.7	29.2	31.4	34.2	37.8	53.8	58.6	62.8	66.6	71.2	74.4	81.4
47	22.6	25.8	27.4	30.0	32.3	35.1	38.7	54.9	59.8	64.0	67.8	72.4	75.7	82.7
48	23.3	26.5	28.2	30.8	33.1	35.9	39.6	56.0	60.9	65.2	69.0	73.7	77.0	84.0
49	24.0	27.2	28.9	31.6	33.9	36.8	40.5	57.1	62.0	66.3	70.2	74.9	78.2	85.4
50	24.7	28.0	29.7	32.4	34.8	37.7	41.4	58.2	63.2	67.5	71.4	76.2	79.5	86.7

Table A7. F-distribution

F_α; critical values such that $\boldsymbol{P}\{F > F_\alpha\} = \alpha$



ν_2, denom. d.f.	α	ν_1, numerator degrees of freedom									
		1	2	3	4	5	6	7	8	9	10
1	0.25	5.83	7.5	8.2	8.58	8.82	8.98	9.1	9.19	9.26	9.32
	0.1	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2
	0.05	161	199	216	225	230	234	237	239	241	242
	0.025	648	799	864	900	922	937	948	957	963	969
	0.01	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056
	0.005	16211	19999	21615	22500	23056	23437	23715	23925	24091	24224
	0.001	405284	499999	540379	562500	576405	585937	592873	598144	602284	605621
2	0.25	2.57	3	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38
	0.1	8.53	9	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
	0.05	18.5	19	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
	0.025	38.5	39	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4
	0.01	98.5	99	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4
	0.005	199	199	199	199	199	199	199	199	199	199
	0.001	999	999	999	999	999	999	999	999	999	999
3	0.25	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44
	0.1	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
	0.05	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
	0.025	17.4	16	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4
	0.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2
	0.005	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7
	0.001	167	149	141	137	135	133	132	131	130	129
4	0.25	1.81	2	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08
	0.1	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
	0.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6	5.96
	0.025	12.2	10.6	9.98	9.6	9.36	9.2	9.07	8.98	8.9	8.84
	0.01	21.2	18	16.7	16	15.5	15.2	15	14.8	14.7	14.5
	0.005	31.3	26.3	24.3	23.2	22.5	22	21.6	21.4	21.1	21
	0.001	74.1	61.2	56.2	53.4	51.7	50.5	49.7	49	48.5	48.1
5	0.25	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89
	0.1	4.06	3.78	3.62	3.52	3.45	3.4	3.37	3.34	3.32	3.3
	0.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
	0.025	10	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
	0.01	16.3	13.3	12.1	11.4	11	10.7	10.5	10.3	10.2	10.1
	0.005	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14	13.8	13.6
	0.001	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2	26.9
6	0.25	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77
	0.1	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94
	0.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.1	4.06
	0.025	8.81	7.26	6.6	6.23	5.99	5.82	5.7	5.6	5.52	5.46
	0.01	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.1	7.98	7.87
	0.005	18.6	14.5	12.9	12	11.5	11.1	10.8	10.6	10.4	10.3
	0.001	35.5	27	23.7	21.9	20.8	20	19.5	19	18.7	18.4
8	0.25	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63
	0.1	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54
	0.05	5.32	4.46	4.07	3.84	3.69	3.58	3.5	3.44	3.39	3.35
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.3
	0.01	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
	0.005	14.7	11	9.6	8.81	8.3	7.95	7.69	7.5	7.34	7.21
	0.001	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12	11.8	11.5
10	0.25	1.49	1.6	1.6	1.59	1.59	1.58	1.57	1.56	1.56	1.55
	0.1	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
	0.05	4.96	4.1	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72
	0.01	10	7.56	6.55	5.99	5.64	5.39	5.2	5.06	4.94	4.85
	0.005	12.8	9.43	8.08	7.34	6.87	6.54	6.3	6.12	5.97	5.85
	0.001	21	14.9	12.6	11.3	10.5	9.93	9.52	9.2	8.96	8.75

Table A7, continued. F-distribution

ν_2, denom. d.f.	α	ν_1, numerator degrees of freedom									
		15	20	25	30	40	50	100	200	500	∞
1	0.25	9.49	9.58	9.63	9.67	9.71	9.74	9.8	9.82	9.84	9.85
	0.1	61.2	61.7	62.1	62.3	62.5	62.7	63	63.2	63.3	63.3
	0.05	246	248	249	250	251	252	253	254	254	254
	0.025	985	993	998	1001	1006	1008	1013	1016	1017	1018
	0.01	6157	6209	6240	6261	6287	6303	6334	6350	6360	6366
	0.005	24630	24836	24960	25044	25148	25211	25337	25401	25439	25464
	0.001	615764	620908	624017	626099	628712	630285	633444	635030	635983	636619
2	0.25	3.41	3.43	3.44	3.44	3.45	3.46	3.47	3.47	3.47	3.48
	0.1	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49	9.49	9.49
	0.05	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
	0.025	39.4	39.4	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5
	0.01	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
	0.005	199	199	199	199	199	199	199	199	199	199
	0.001	999	999	999	999	999	999	999	999	999	999
3	0.25	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47	2.47	2.47
	0.1	5.2	5.18	5.17	5.17	5.16	5.15	5.14	5.14	5.14	5.13
	0.05	8.7	8.66	8.63	8.62	8.59	8.58	8.55	8.54	8.53	8.53
	0.025	14.3	14.2	14.1	14.1	14	14	14	13.9	13.9	13.9
	0.01	26.9	26.7	26.6	26.5	26.4	26.4	26.2	26.2	26.1	26.1
	0.005	43.1	42.8	42.6	42.5	42.3	42.2	42	41.9	41.9	41.8
	0.001	127	126	126	125	125	125	124	124	124	123
4	0.25	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
	0.1	3.87	3.84	3.83	3.82	3.8	3.8	3.78	3.77	3.76	3.76
	0.05	5.86	5.8	5.77	5.75	5.72	5.7	5.66	5.65	5.64	5.63
	0.025	8.66	8.56	8.5	8.46	8.41	8.38	8.32	8.29	8.27	8.26
	0.01	14.2	14	13.9	13.8	13.7	13.7	13.6	13.5	13.5	13.5
	0.005	20.4	20.2	20	19.9	19.8	19.7	19.5	19.4	19.4	19.3
	0.001	46.8	46.1	45.7	45.4	45.1	44.9	44.5	44.3	44.1	44.1
5	0.25	1.89	1.88	1.88	1.88	1.88	1.88	1.87	1.87	1.87	1.87
	0.1	3.24	3.21	3.19	3.17	3.16	3.15	3.13	3.12	3.11	3.1
	0.05	4.62	4.56	4.52	4.5	4.46	4.44	4.41	4.39	4.37	4.36
	0.025	6.43	6.33	6.27	6.23	6.18	6.14	6.08	6.05	6.03	6.02
	0.01	9.72	9.55	9.45	9.38	9.29	9.24	9.13	9.08	9.04	9.02
	0.005	13.1	12.9	12.8	12.7	12.5	12.5	12.3	12.2	12.2	12.1
	0.001	25.9	25.4	25.1	24.9	24.6	24.4	24.1	24	23.9	23.8
6	0.25	1.76	1.76	1.75	1.75	1.75	1.75	1.74	1.74	1.74	1.74
	0.1	2.87	2.84	2.81	2.8	2.78	2.77	2.75	2.73	2.73	2.72
	0.05	3.94	3.87	3.83	3.81	3.77	3.75	3.71	3.69	3.68	3.67
	0.025	5.27	5.17	5.11	5.07	5.01	4.98	4.92	4.88	4.86	4.85
	0.01	7.56	7.4	7.3	7.23	7.14	7.09	6.99	6.93	6.9	6.88
	0.005	9.81	9.59	9.45	9.36	9.24	9.17	9.03	8.95	8.91	8.88
	0.001	17.6	17.1	16.9	16.7	16.4	16.3	16	15.9	15.8	15.7
8	0.25	1.62	1.61	1.6	1.6	1.59	1.59	1.58	1.58	1.58	1.58
	0.1	2.46	2.42	2.4	2.38	2.36	2.35	2.32	2.31	2.3	2.29
	0.05	3.22	3.15	3.11	3.08	3.04	3.02	2.97	2.95	2.94	2.93
	0.025	4.1	4	3.94	3.89	3.84	3.81	3.74	3.7	3.68	3.67
	0.01	5.52	5.36	5.26	5.2	5.12	5.07	4.96	4.91	4.88	4.86
	0.005	6.81	6.61	6.48	6.4	6.29	6.22	6.09	6.02	5.98	5.95
	0.001	10.8	10.5	10.3	10.1	9.92	9.8	9.57	9.45	9.38	9.33
10	0.25	1.53	1.52	1.52	1.51	1.51	1.5	1.49	1.49	1.49	1.48
	0.1	2.24	2.2	2.17	2.16	2.13	2.12	2.09	2.07	2.06	2.06
	0.05	2.85	2.77	2.73	2.7	2.66	2.64	2.59	2.56	2.55	2.54
	0.025	3.52	3.42	3.35	3.31	3.26	3.22	3.15	3.12	3.09	3.08
	0.01	4.56	4.41	4.31	4.25	4.17	4.12	4.01	3.96	3.93	3.91
	0.005	5.47	5.27	5.15	5.07	4.97	4.9	4.77	4.71	4.67	4.64
	0.001	8.13	7.8	7.6	7.47	7.3	7.19	6.98	6.87	6.81	6.76

1. Complete the following statements:

1) If $x \sim B_i(n,p)$ as $n \longrightarrow \infty, p \to 0, np \to \lambda$ then $B_i(n,p) \to \cdots$

2) If $X \sim B_i(n,p)$ as $n \to \infty$

$$\frac{X - np}{\sqrt{npq}} \sim \ldots$$

3) if $X \sim N(50, 10)$

$$\text{at } X = 70 \Longrightarrow z = \cdots$$

4) If $\quad X \sim N(8, 10) \quad$ normal distribution $\Rightarrow$ then: $z = \dfrac{x-8}{10} \sim \cdots.$

2. Suppose that the average household income in some Country is ($\mu = 900$) coins, and the standard deviation is ($\sigma = 20$) coins. Assuming the normal distribution of income, Compute the proportion of the middle class "whose income is between 600 and 1200 coins?

$$\{P(600 < X < 1200)\}$$

3. The regression lines between the random variables X, Y given by equation

$$y = 35.823 + 0.4764x$$
$$x = -3.376 + 1.036y$$

then $r(x,y) = \cdots, \qquad (\overline{X}, \overline{Y}) = \cdots$

4. A firm tested 500 new employees on an aptitude test. the store of each employee was X. Three years Later, they collected supervisor rating of each employee's success on the job. these ratings and denoted by Y. The data yield the following statistics:

$$\overline{X} = 100, S_x = 10, \overline{Y} = 130, S_y = 20 \text{ and } r_{xy} = 0.70$$

Compute the least squares regression line for predicting Y. What is predict for employees who received test scores of $x = 90$ and $x = 125$?

5. If $n = 1000, \sum xy = 30000, \sum x = 3000, \sum x^2 = 14000$ and $\sum y = 5000$.

a) Compute the least square line for that data?

b) If $S_y = 10$ compute the Correlation Coefficient r?

6. Type "F" or "T"

1- Each statistic has some distribution $(\cdots)$

2- Critical region is always on one tail only $(\cdots)$

3- The standard deviation of an estimate and standard error are the same $(\cdots)$

4- Interval estimate is better than point estimate $(\cdots)$

5- t-value, z-value lies between $-\infty$ and ∞ $(\cdots)$

6- x^2-value, F-value lies between o and ∞ $(\cdots)$

7- For any $r.v.$ X the standardized variable $\frac{\overline{X}-E\overline{X}}{S_{\overline{x}}/\sqrt{n}}$ is $N(0,1)$ as $n \to \infty$ $(\cdots)$

8- The variances $s^2 = \frac{1}{n}\sum(x-\overline{x})^2$ is unbiased estimate of σ^2 $(\cdots)$

9- For t-test if $H_0 : \mu_1 = \mu_2$ V.S. $H_1 := \mu_1 > \mu_2$, if $t_{al} > t_{t_{ab}}$ there is no significance between μ_1, μ_2 $(\cdots)$

10- If $r_{xy} = 0.9, \beta_{xy} = 2.04 \quad \& b_{yx} = -3.2$ $(\cdots)$

11- In testing hypothesis if $H_1 : \mu < \mu_0$ the critical region lies on two tailed test $(\cdots)$

12- For two samples from two population the standard error for $(\overline{x}_1 - \overline{x}_2)$ is $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $(\cdots)$

13- If $H_0 : \sigma^2 = \sigma_0^2$ we used χ^2-distribution $(\cdots)$

14- If $H_0 : \mu_1 = \mu_2$ we used F-distribution $(\cdots)$

15- If $H_0 : \sigma_1^2 = \sigma_2^2$ we used t-distribution $(\cdots)$

7. Complete the following statements:

1) The regression line Y on X as $Y = a + b\times$ if $b = 2.79, \overline{X} = 15.4$, $\overline{Y} = 44.667$ then $a = \cdots$

2) For the listing hypothesis $H_0 : \mu_1 = \mu_2$ in small size we used the $\cdots$

3) For F-test the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ if

$$S_1^2 = 148.3, S_2^2 = 24.87 \text{ then } F_{al} = \cdots$$

4) The standard error for the mean $\overline{x}$ is $\cdots$

5) If the regression deficient X on $Y, \beta = 0.131, S_X = 8.147,\ S_Y = 37.57$ then the correlation Coefficient r equal $\cdots$

6) To obtain the confidence interval for proportion p for $n < 30$, we used ...

8. Let $x_1, x_2, \cdots, x_n$ be a random Sample of size n from the distribution $X \sim N\left(\mu, \alpha^2\right)$ find the estimate for the two parameters by

a) The moment estimate for μ and σ^2

b) The maximum likelihood estimate for μ, σ^2 ?

9. A program consists of two modules, the number of errors X in the first module and the number of errors Y in the second module have the joint pmf of X and Y in the following table

$P_{x,y}(X,Y)$		y			
		0	1	2	3
x	0	0.20	0.20	0.05	0.05
	1	0.20	0.10	0.10	0.10

a) Find the marginal distributions of X and Y

b) Find the correlation coefficient between x and Y

c) Is X, Y are independent?

10. The I.Q.'s (intelligence quotients) of 16 students from one area of city showed $\overline{X_1} = 107$ with $S_1 = 10$, while the I.Q.'s of 14 students from another area of the City showed that $\overline{x}_2 = 112$ with $S_2 = 8$.

a) Construct the confidence interval for $\mu_1 - \mu_2$ wish 95% confidence interval

b) Is there a significant difference between the I.Q.'s of the two groups at $\alpha = 0.05, \alpha = 0.01$ level of significance?

11. The mean lifetime of a sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with standard deviation of 120 hours $(\overline{x} = 15,\ S = 120)$. If μ is the mean lifetime of all the bulbs produced by the company

i) Test the hypothesis $\mu = 1600$ hours against the alternate $\mu \neq 1600$ h, using

a level significance of $\alpha = 0.01$.
ii) Test the hypothesis $\mu = 1600$ hours against the alternate $\mu < 1600$ h, using a level significance of $\alpha = 0.05$.
iii) Estimate the confidence internal for the mean μ at 90%.

12. In problem 11 test the hypothesis $\mu = 1600$ h against the alternative $\mu < 1600$ h using a level of significance of $\alpha = 0.05 \,\&\, \alpha = 0.01$

13. The student government of a large college polled a random sample of 325 male students and found that 221 were in favor of a new grading system at the same time 120 out of random sample of 200 female students were in favor of the new system.
a) Construct a 90% confidence interval for the us difference $(P_1 - P_2)$ the proportion of male and female students who favor the new system?
b) Use $\alpha = 0.050, \alpha = 0.01)$ significant difference in the proportions?
14.
a) If $\overline{X} = 14.10, S = 1.67, n = 8$, fund 98% Confidence interns for μ.
b) If $n = 150, \sigma = 6.2, \overline{x} = 69.5$, find 95% Confidence interval for μ ?

15. The overage weight for recruits in the sample $\overline{x} = 160$ pounds, $s = 10$ pounds. Suppose that we want the 95% confidence interval to be equal at most to 5 pounds $(\epsilon = 5)$ what size random sample should you take?
For $n = 90$ establish a 98% confidence interval for the variance σ^2 of all weight the recruits?
16. For $n = 40, s = 0.74$, test the null hypothesis $H_0 : \sigma = 0.80$ against $H_1 : \sigma < 0.80$ at $\alpha = 0.01$ level of Significance?

17.
i) Let $\overline{x} = 16, s = 1.8, n = 25$. Establish a 95% confidence interval for μ ?
(ii) Assume $n = 10, \mu = 0, \sum_1^1 0x^2 = 106.6, \alpha = 0.10$. Find a 90% confidence interval for a^2 ?
(iii) Fund a 95° Confidence interval for σ^2 using the data $n = 9, S^2 = 7.62$?
18. The following random samples are measurements of the head-producing

capacity (in million Calories per tone) of coal from two mines

$$n_1 = n_2 = 5,\ \overline{X_1} = 8178,\ \overline{X_2} = 7788,\ S_1 = 271.1 \text{ and } S_2 = 216.8$$

i) Construct a 95% Confidence interval for the true difference $(\mu_2 - \mu_1)$.
ii) Use the (0.05) and (0.01) level of significance to test where the difference between the means of these two samples is significant? $(t.9._5 = \pm 2.31, t_{.995.5} = \pm 3.36)$.

19.
i) If $n = 250$, $\hat{p} = 0.58$, estimate 98% Confidence interval for the proportion p, and what is the maximum error of the estimate?
ii) If the error of the estimate $\epsilon = 0.04$ and $\hat{p} = 0.25$, find the number of sample at 95% ? $(z_{0.99} = \pm 2.33, Z_{.975} = \pm 1.96)$.

20. A random number table 250 digits, showed the following distribution of digits $0, 1, 2, \ldots, 9$. Does the observed (O) distribution differs significantly from the expected distribution?

Digits	0	1	2	3	4	5	6	7	8	9	$\sum$
O	17	31	29	18	14	20	35	30	20	36	250
W	25	25	25	25	25	25	25	25	25	25	250

21. In the past the standard deviation of weights for Certain 40.0 Newton packages filled by a machine was $\sigma = 0.25$ newtons. A random sample of 20 packages shows a standard deviation of $S = 0.32$ Newtons. Is the apparent increase in variability significant at
a) $\alpha = 0.05$ and **b)** $\alpha = 0.01$ level of Significance. $(H_0 : 0 = 0.25, H_1 = 0 > 0.25)$

22. In 200 tosses of Coin, 115 heads and 85 tails were observed. Test the hypothesis that the Coin is fair using of significance. **a)** $\alpha = 0.05$ & **b)** $\alpha = 0.01$

23. Let $n_1 = 300$, $\hat{P}_1 = 0.27$; $n_2 = 200$, $\hat{P}_2 = 0.2$. F Find a confidence interval for $P_1 - P_2$, if $H_0 : P_1 = P_2$; $H_1 : P_1 > P_2$, $\alpha = 0.01$

24. From appropriate samples, two sets of two scores are obtained:

$$\text{I: } \overline{X}_1 = 104, S_1 = 10, n_1 = 16$$
$$\text{II: } \overline{X}_2 = 112, S_2 = 8, n_2 = 14$$

a) estimate the 98% confidence interval for the difference of sample means? $(\mu_2 - \mu_1)$
b) at the 5% significance level is there a significant difference mean between the two groups? $(t_{.99,28} = \pm 2.467, t_{.975,28} = \pm 2.05)$