STATISTICAL INFERENCES (2cr)
Chapter 8 Sampling Distributions & Data Descriptions

Zhong Guan

Math, IUSB

Outline

Definition

Sampling distribution is the probability distribution of a **statistic.**

For example, the probability distribution of the sampling mean $\bar{X}$ is called the sampling distribution of the mean.

Definition

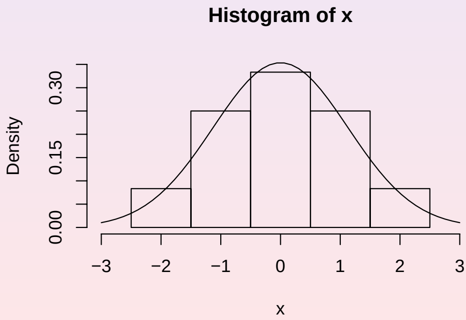
Sampling distribution is the probability distribution of a **statistic.**
For example, the probability distribution of the sampling mean $\bar{X}$ is called the sampling distribution of the mean.

Examples

Example 1. Suppose we have a population containing the following values:

-2, -1, -1, -1, 0, 0, 0, 0, 1, 1, 1, 2

Distribution of the data is close to: $N(0, 7/6)$.

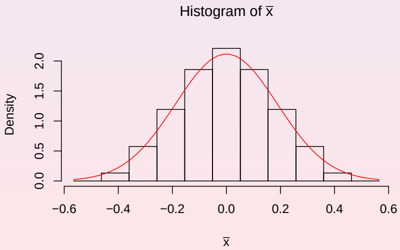


Distribution of All Sample Means

Let $X_1, X_2, \ldots, X_n$ be a sample of size $n = 9$.

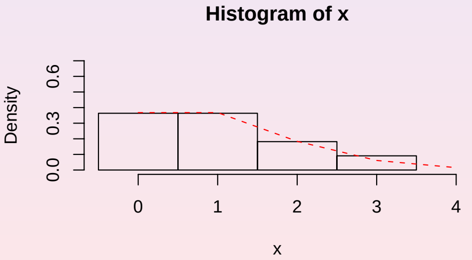
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

There are $\binom{12}{9} = 220$ sample means. The distribution of all the sample means is also normal $N(0, 0.0354)$.



Examples

Example 2. Suppose we have a skewed population containing the following values:
0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 3
Distribution of the data is close to Poisson: $P(1)$.

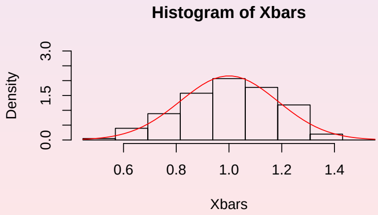


Distribution of All Sample Means

Let $X_1, X_2, \ldots, X_n$ be a sample of size $n = 8$.

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

There are $\binom{11}{8} = 165$ sample means. The distribution of all the sample means is also close to normal $N(1, 0.034)$.



Outline

Exact distribution of sample mean

Let $X_1, X_2, \ldots, X_n$ be a sample of size n. That is, $X_1, X_2, \ldots, X_n$ are independent random variables having the same distribution with mean μ and variance σ^2.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

If the sample is from $N(\mu, \sigma^2)$, then exactly

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{and} \quad \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

What if the sample is not from $N(\mu, \sigma^2)$?

Exact distribution of sample mean

Let $X_1, X_2, \ldots, X_n$ be a sample of size n. That is, $X_1, X_2, \ldots, X_n$ are independent random variables having the same distribution with mean μ and variance σ^2.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

If the sample is from $N(\mu, \sigma^2)$, then exactly

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{and} \quad \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

What if the sample is not from $N(\mu, \sigma^2)$?

Exact distribution of sample mean

Let $X_1, X_2, \ldots, X_n$ be a sample of size n. That is, $X_1, X_2, \ldots, X_n$ are independent random variables having the same distribution with mean μ and variance σ^2.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

If the sample is from $N(\mu, \sigma^2)$, then exactly

$$\overline{X} \sim N(\mu, \tfrac{\sigma^2}{n}) \quad \text{and} \quad \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

What if the sample is not from $N(\mu, \sigma^2)$?

Exact distribution of sample mean

Let $X_1, X_2, \ldots, X_n$ be a sample of size n. That is, $X_1, X_2, \ldots, X_n$ are independent random variables having the same distribution with mean μ and variance σ^2.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

If the sample is from $N(\mu, \sigma^2)$, then exactly

$$\overline{X} \sim N(\mu, \tfrac{\sigma^2}{n}) \quad \text{and} \quad \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

What if the sample is not from $N(\mu, \sigma^2)$?

Exact distribution of sample mean

Let $X_1, X_2, \ldots, X_n$ be a sample of size n. That is, $X_1, X_2, \ldots, X_n$ are independent random variables having the same distribution with mean μ and variance σ^2.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n}$$

If the sample is from $N(\mu, \sigma^2)$, then exactly

$$\overline{X} \sim N(\mu, \tfrac{\sigma^2}{n}) \quad \text{and} \quad \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

What if the sample is not from $N(\mu, \sigma^2)$?

Central Limit Theorem

Central Limit Theorem If $\overline{X}$ is the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution, discrete or continuous, with mean μ and variance σ^2, then the distribution of

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

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

$$\lim_{n\to\infty} P(Z \leqslant z) = \Phi(z).$$

So if $n \geqslant 30$ (or smaller for symmetric distribution), then the distribution of $\overline{X}$ is approximately $N(\mu, \sigma^2/n)$ and

$$P(a < Z \leqslant b) \approx \Phi(b) - \Phi(a), \quad a < b.$$

Central Limit Theorem

Central Limit Theorem If $\overline{X}$ is the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution, discrete or continuous, with mean μ and variance σ^2, then the distribution of

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

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

$$\lim_{n\to\infty} P(Z \leqslant z) = \Phi(z).$$

So if $n \geqslant 30$ (or smaller for symmetric distribution), then the distribution of $\overline{X}$ is approximately $N(\mu, \sigma^2/n)$ and

$$P(a < Z \leqslant b) \approx \Phi(b) - \Phi(a), \quad a < b.$$

Central Limit Theorem

Central Limit Theorem If $\overline{X}$ is the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution, discrete or continuous, with mean μ and variance σ^2, then the distribution of

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

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

$$\lim_{n\to\infty} P(Z \leqslant z) = \Phi(z).$$

So if $n \geqslant 30$ (or smaller for symmetric distribution), then the distribution of $\overline{X}$ is approximately $N(\mu, \sigma^2/n)$ and

$$P(a < Z \leqslant b) \approx \Phi(b) - \Phi(a), \quad a < b.$$

Examples

Example 1. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $U[0,1]$, the uniform distribution on $[0,1]$.

$$\mu = \mathrm{E}(X_1) = \frac{1}{2}, \quad \sigma^2 = \mathrm{Var}(X_1) = \frac{1}{12}.$$

$$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \sqrt{12n}(\overline{X} - \tfrac{1}{2})$$

Examples

Example 1. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $U[0,1]$, the uniform distribution on $[0,1]$.

$$\mu = \mathrm{E}(X_1) = \frac{1}{2}, \quad \sigma^2 = \mathrm{Var}(X_1) = \frac{1}{12}.$$

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{12n}\left(\overline{X} - \frac{1}{2}\right)$$

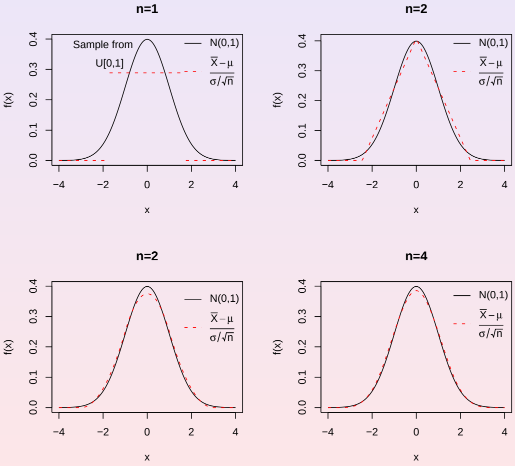
Examples

Example 1. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $U[0, 1]$, the uniform distribution on $[0, 1]$.

$$\mu = \mathrm{E}(X_1) = \frac{1}{2}, \quad \sigma^2 = \mathrm{Var}(X_1) = \frac{1}{12}.$$

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{12n}(\overline{X} - \frac{1}{2})$$

Examples


n=1
Sample from
U[0,1]
N(0,1)
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
f(x)
x
n=2
N(0,1)
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
f(x)
x
n=2
N(0,1)
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
f(x)
x
n=4
N(0,1)
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
f(x)
x

Examples

Example 2. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $EXP(\theta)$, the exponential distribution with

$$\mu = \mathrm{E}(X_1) = \theta, \quad \sigma^2 = \mathrm{Var}(X_1) = \theta^2.$$

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

Examples

Example 2. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $EXP(\theta)$, the exponential distribution with

$$\mu = \mathrm{E}(X_1) = \theta, \quad \sigma^2 = \mathrm{Var}(X_1) = \theta^2.$$

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

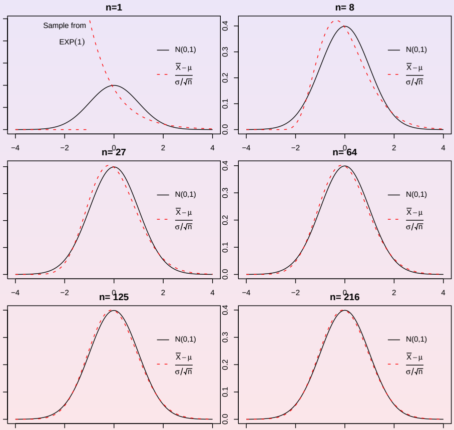
Examples

Example 2. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $EXP(\theta)$, the exponential distribution with

$$\mu = \mathrm{E}(X_1) = \theta, \quad \sigma^2 = \mathrm{Var}(X_1) = \theta^2.$$

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

Examples


n=1
Sample from
EXP(1)
N(0,1)
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
n= 8
n= 27
n= 64
n= 125
n= 216
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Examples

Example 3. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $\chi^2(k)$, the chi-square distribution with

$$\mu = \mathrm{E}(X_1) = k, \quad \sigma^2 = \mathrm{Var}(X_1) = 2k.$$

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - k}{2k/\sqrt{n}}$$

Examples

Example 3. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $\chi^2(k)$, the chi-square distribution with

$$\mu = \mathrm{E}(X_1) = k, \quad \sigma^2 = \mathrm{Var}(X_1) = 2k.$$

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - k}{2k/\sqrt{n}}$$

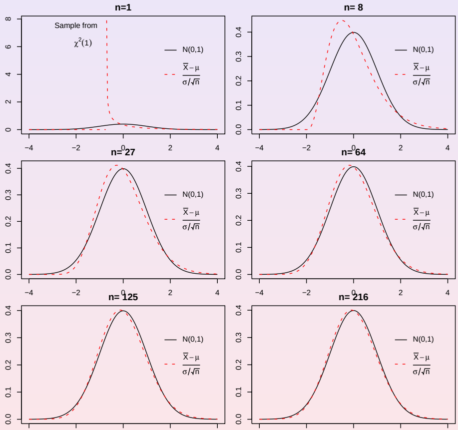
Examples

Example 3. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $\chi^2(k)$, the chi-square distribution with

$$\mu = \mathrm{E}(X_1) = k, \quad \sigma^2 = \mathrm{Var}(X_1) = 2k.$$

$$Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-k}{2k/\sqrt{n}}$$

Examples


n=1
Sample from
$\chi^2(1)$
N(0,1)
$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
n= 8
n= 27
n= 64
n= 125
n= 216

Examples

The central limit theorem is also valid for discrete distribution.

Example 4. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $b(1, p)$ with

$$\mu = p, \quad \sigma^2 = p(1-p).$$

$$Y = \sum_{i=1}^{n} X_i \sim b(n, p).$$

By the CLT,

$$\overline{X} \sim N(\mu, \sigma^2/n), \text{ approximately.}$$

$$Y = n\overline{X} \sim N(n\mu, n^2\sigma^2/n) = N[np, np(1-p)], \text{ approximately.}$$

Examples

The central limit theorem is also valid for discrete distribution.

Example 4. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $b(1, p)$ with

$$\mu = p, \quad \sigma^2 = p(1-p).$$

$$Y = \sum_{i=1}^{n} X_i \sim b(n, p).$$

By the CLT,

$$\overline{X} \sim N(\mu, \sigma^2/n), \text{ approximately.}$$

$$Y = n\overline{X} \sim N(n\mu, n^2\sigma^2/n) = N[np, np(1-p)], \text{ approximately.}$$

Examples

The central limit theorem is also valid for discrete distribution.

Example 4. Let $\overline{X}$ be the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from $b(1, p)$ with

$$\mu = p, \quad \sigma^2 = p(1-p).$$

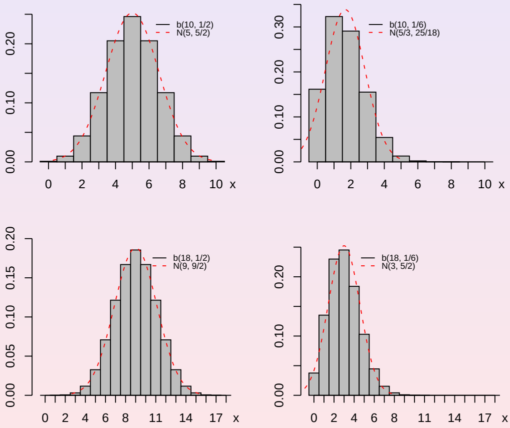
$$Y = \sum_{i=1}^{n} X_i \sim b(n, p).$$

By the CLT,

$$\overline{X} \sim N(\mu, \sigma^2/n), \text{ approximately.}$$

$$Y = n\overline{X} \sim N(n\mu, n^2\sigma^2/n) = N[np, np(1-p)], \text{ approximately.}$$

Examples



Examples

Example 5. If a certain machine makes resistors have a mean resistance of 40 ohms and standard deviation of 2 ohms, what is the probability that a random sample of 36 of these resistors will have an average resistance of more than 40.5 ohms.

Examples

Solution of Example 5. $n = 36$, $\mu = 40$, $\sigma = 2$. By the CLT, $\overline{X}$ has normal distribution with mean $\mu_{\bar{x}} = \mu = 40$ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 2/\sqrt{36} = 2/6 = 1/3$. The z value of $x = 40.5$ is

$$z = \frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{40.5 - 40}{1/3} = 3(0.5) = 1.5$$

Using Table A.3

$$P(\overline{X} > 40.5) = P(Z > 1.5) = 1 - P(Z \leqslant 1.5) = 1 - 0.9332 = 0.0668$$

Examples

Solution of Example 5. $n = 36$, $\mu = 40$, $\sigma = 2$. By the CLT, $\overline{X}$ has normal distribution with mean $\mu_{\bar{x}} = \mu = 40$ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 2/\sqrt{36} = 2/6 = 1/3$. The z value of $x = 40.5$ is

$$z = \frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{40.5 - 40}{1/3} = 3(0.5) = 1.5$$

Using Table A.3

$$P(\overline{X} > 40.5) = P(Z > 1.5) = 1 - P(Z \leqslant 1.5) = 1 - 0.9332 = 0.0668$$

Outline

Examples

Example 6.[Example 5. (Cont.)] An observed sample of 36 of these resistors indicates a sample average of 39.1 ohms. Does this sample information appear to support or refute the conjecture that $\mu = 40$ ohms?

Solution of Example 6. If the conjecture $\mu = 40$ is true, then by the CLT $\overline{X}$ with $n = 36$ is approximately normal with mean $\mu = 40$ and standard deviation $\sigma/\sqrt{n} = 2/\sqrt{36} = 1/3$.

How likely can the value of $\overline{X}$ be as far away from the center $\mu = 40$ as the observed $\bar{x} = 39.1$?

That is, if $\mu = 40$,

$$P(|\overline{X} - 40| \geqslant |39.1 - 40|) = P(|\overline{X} - 40| \geqslant 0.9) = ?$$

Examples

Example 6.[Example 5. (Cont.)] An observed sample of 36 of these resistors indicates a sample average of 39.1 ohms. Does this sample information appear to support or refute the conjecture that $\mu = 40$ ohms?

Solution of Example 6. If the conjecture $\mu = 40$ is true, then by the CLT $\overline{X}$ with $n = 36$ is approximately normal with mean $\mu = 40$ and standard deviation $\sigma/\sqrt{n} = 2/\sqrt{36} = 1/3$.

How likely can the value of $\overline{X}$ be as far away from the center $\mu = 40$ as the observed $\bar{x} = 39.1$?

That is, if $\mu = 40$,

$$P(|\overline{X} - 40| \geqslant |39.1 - 40|) = P(|\overline{X} - 40| \geqslant 0.9) = ?$$

Examples

Example 6.[Example 5. (Cont.)] An observed sample of 36 of these resistors indicates a sample average of 39.1 ohms. Does this sample information appear to support or refute the conjecture that $\mu = 40$ ohms?

Solution of Example 6. If the conjecture $\mu = 40$ is true, then by the CLT $\overline{X}$ with $n = 36$ is approximately normal with mean $\mu = 40$ and standard deviation $\sigma/\sqrt{n} = 2/\sqrt{36} = 1/3$.
How likely can the value of $\overline{X}$ be as far away from the center $\mu = 40$ as the observed $\bar{x} = 39.1$?
That is, if $\mu = 40$,

$$P(|\overline{X} - 40| \geqslant |39.1 - 40|) = P(|\overline{X} - 40| \geqslant 0.9) = ?$$

Examples

Example 6.[Example 5. (Cont.)] An observed sample of 36 of these resistors indicates a sample average of 39.1 ohms. Does this sample information appear to support or refute the conjecture that $\mu = 40$ ohms?

Solution of Example 6. If the conjecture $\mu = 40$ is true, then by the CLT $\overline{X}$ with $n = 36$ is approximately normal with mean $\mu = 40$ and standard deviation $\sigma/\sqrt{n} = 2/\sqrt{36} = 1/3$.

How likely can the value of $\overline{X}$ be as far away from the center $\mu = 40$ as the observed $\bar{x} = 39.1$?

That is, if $\mu = 40$,

$$P(|\overline{X} - 40| \geqslant |39.1 - 40|) = P(|\overline{X} - 40| \geqslant 0.9) = ?$$

Solution of Example 6.

By the CLT, if $\mu = 40$, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-40}{2/\sqrt{36}} = 3(\overline{X} - 40)$.

$$P(|\overline{X} - 40| \geqslant 0.9) = P(|Z| \geqslant 3(0.9)) = P(|Z| \geqslant 2.7)$$

$$= P(Z \geqslant 2.7) + P(Z \leqslant -2.7)$$

$$= P(Z \geqslant 2.7) + P(-Z \geqslant 2.7) = 2P(Z \geqslant 2.7)$$

$$= 2[1 - P(Z < 2.7)] = 2(1 - 0.996533) \approx 0.007$$

One would experience by chance that an $\bar{x}$ is 0.9 ohms from the mean in only 7 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 6.

By the CLT, if $\mu = 40$, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-40}{2/\sqrt{36}} = 3(\overline{X} - 40)$.

$$P(|\overline{X} - 40| \geqslant 0.9) = P(|Z| \geqslant 3(0.9)) = P(|Z| \geqslant 2.7)$$

$$= P(Z \geqslant 2.7) + P(Z \leqslant -2.7)$$

$$= P(Z \geqslant 2.7) + P(-Z \geqslant 2.7) = 2P(Z \geqslant 2.7)$$

$$= 2[1 - P(Z < 2.7)] = 2(1 - 0.996533) \approx 0.007$$

One would experience by chance that an $\bar{x}$ is 0.9 ohms from the mean in only 7 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 6.

By the CLT, if $\mu = 40$, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-40}{2/\sqrt{36}} = 3(\overline{X} - 40)$.

$$P(|\overline{X} - 40| \geqslant 0.9) = P(|Z| \geqslant 3(0.9)) = P(|Z| \geqslant 2.7)$$

$$= P(Z \geqslant 2.7) + P(Z \leqslant -2.7)$$

$$= P(Z \geqslant 2.7) + P(-Z \geqslant 2.7) = 2P(Z \geqslant 2.7)$$

$$= 2[1 - P(Z < 2.7)] = 2(1 - 0.996533) \approx 0.007$$

One would experience by chance that an $\bar{x}$ is 0.9 ohms from the mean in only 7 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 6.

By the CLT, if $\mu = 40$, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-40}{2/\sqrt{36}} = 3(\overline{X} - 40)$.

$$P(|\overline{X} - 40| \geqslant 0.9) = P(|Z| \geqslant 3(0.9)) = P(|Z| \geqslant 2.7)$$

$$= P(Z \geqslant 2.7) + P(Z \leqslant -2.7)$$

$$= P(Z \geqslant 2.7) + P(-Z \geqslant 2.7) = 2P(Z \geqslant 2.7)$$

$$= 2[1 - P(Z < 2.7)] = 2(1 - 0.996533) \approx 0.007$$

One would experience by chance that an $\bar{x}$ is 0.9 ohms from the mean in only 7 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 6.

By the CLT, if $\mu = 40$, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-40}{2/\sqrt{36}} = 3(\overline{X} - 40)$.

$$P(|\overline{X} - 40| \geqslant 0.9) = P(|Z| \geqslant 3(0.9)) = P(|Z| \geqslant 2.7)$$

$$= P(Z \geqslant 2.7) + P(Z \leqslant -2.7)$$

$$= P(Z \geqslant 2.7) + P(-Z \geqslant 2.7) = 2P(Z \geqslant 2.7)$$

$$= 2[1 - P(Z < 2.7)] = 2(1 - 0.996533) \approx 0.007$$

One would experience by chance that an $\bar{x}$ is 0.9 ohms from the mean in only 7 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 6.

By the CLT, if $\mu = 40$, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-40}{2/\sqrt{36}} = 3(\overline{X} - 40)$.

$$P(|\overline{X} - 40| \geqslant 0.9) = P(|Z| \geqslant 3(0.9)) = P(|Z| \geqslant 2.7)$$

$$= P(Z \geqslant 2.7) + P(Z \leqslant -2.7)$$

$$= P(Z \geqslant 2.7) + P(-Z \geqslant 2.7) = 2P(Z \geqslant 2.7)$$

$$= 2[1 - P(Z < 2.7)] = 2(1 - 0.996533) \approx 0.007$$

One would experience by chance that an $\bar{x}$ is 0.9 ohms from the mean in only 7 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 6.

- The probability $P(|\overline{X} - 40| \geqslant 0.9 | \mu = 40)$ is called the *p*-value of the sample mean $\bar{x} = 39.1$.
- Under the condition that the conjecture or hypothesis $\mu = 40$ is true, *p*-value is the probability that we can observed an $\bar{X}$ as extreme as $\bar{x} = 39.1$.
- *p*-value is NOT the probability that the conjecture or hypothesis $\mu = 40$ is true.

Solution of Example 6.

- The probability $P(|\overline{X} - 40| \geqslant 0.9 | \mu = 40)$ is called the p-value of the sample mean $\bar{x} = 39.1$.
- Under the condition that the conjecture or hypothesis $\mu = 40$ is true, p-value is the probability that we can observed an $\bar{X}$ as extreme as $\bar{x} = 39.1$.
- p-value is NOT the probability that the conjecture or hypothesis $\mu = 40$ is true.

Solution of Example 6.

- The probability $P(|\overline{X} - 40| \geqslant 0.9|\mu = 40)$ is called the *p*-value of the sample mean $\bar{x} = 39.1$.
- Under the condition that the conjecture or hypothesis $\mu = 40$ is true, *p*-value is the probability that we can observed an $\bar{X}$ as extreme as $\bar{x} = 39.1$.
- *p*-value is NOT the probability that the conjecture or hypothesis $\mu = 40$ is true.

Outline

Exact Distribution of Difference between Two Means

If $\overline{X}_1$ and $\overline{X}_2$ are the sample means of independent random samples of size n_1 and n_2 from two normal distributions with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the distribution of $\overline{X}_1 - \overline{X}_2$ is exactly normal with mean $\mu_1 - \mu_2$ and variance $(\sigma_1^2/n_1) + (\sigma_2^2/n_2)$. So the distribution of

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

exactly standard normal $N(0, 1)$.

CLT for Difference between Two Means

Central Limit Theorem for Difference between Two Means
If $\overline{X}_1$ and $\overline{X}_2$ are the sample means of independent random samples of size n_1 and n_2 from two nonnormal distributions, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the distribution of

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

approximately standard normal distribution $N(0, 1)$ if $n_1, n_2 \geqslant 30$.

Examples

Example 7. Two independent samples of size 18 are selected from two types of paints, A and B. The average drying time, in hours, is recorded for each sample. Assume that the populations are normal with the same means and the population standard deviations are both known to be 1.0. Find the probability that the difference between the two means is greater than 1.0.

Solution of Example 7. $\mu_A = \mu_B$, $\sigma_A = \sigma_B = 1$, and $n_A = n_B = 18$. Because the populations are normal

$$Z = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{(\sigma_A^2/n_A) + (\sigma_B^2/n_B)}} = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{\sqrt{(1/18) + (1/18)}} = 3(\overline{X}_1 - \overline{X}_2)$$

is exactly standard normal.

Examples

Example 7. Two independent samples of size 18 are selected from two types of paints, A and B. The average drying time, in hours, is recorded for each sample. Assume that the populations are normal with the same means and the population standard deviations are both known to be 1.0. Find the probability that the difference between the two means is greater than 1.0.

Solution of Example 7. $\mu_A = \mu_B$, $\sigma_A = \sigma_B = 1$, and $n_A = n_B = 18$. Because the populations are normal

$$Z = \frac{(\overline{X}_A - \overline{X}_B) - (\mu_A - \mu_B)}{\sqrt{(\sigma_A^2/n_A) + (\sigma_B^2/n_B)}} = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{\sqrt{(1/18) + (1/18)}} = 3(\overline{X}_1 - \overline{X}_2)$$

is exactly standard normal.

Solution of Example 7.

So

$$P(|\overline{X}_A - \overline{X}_B| \geqslant 1) = P(|Z| \geqslant 3) = 2(1 - P(Z < 3)) = 2(1 - 0.9987)$$

$$= 2(0.0013) = 0.0026$$

One would experience by chance that a difference between the two means is bigger than 1 in only 2.6 in 1000 pairs of samples of size 18.

Solution of Example 7.

So

$$P(|\overline{X}_A - \overline{X}_B| \geqslant 1) = P(|Z| \geqslant 3) = 2(1 - P(Z < 3)) = 2(1 - 0.9987)$$

$$= 2(0.0013) = 0.0026$$

One would experience by chance that a difference between the two means is bigger than 1 in only 2.6 in 1000 pairs of samples of size 18.

Examples

Example 7.(Cont.) Suppose a difference of 1.0 in means was observed in real samples.

(a) Does this seem to be a reasonable results if the two population mean drying times truly are equal?

No.

(b) If someone selected 10,000 pairs of samples of size 18 under the condition that $\mu_A = \mu_B$, in how many of these 10,000 experiments would there be a difference $\bar{x}_A - \bar{x}_B$ is as large as 1.0?

26.

Examples

Example 7.(Cont.) Suppose a difference of 1.0 in means was observed in real samples.
(a) Does this seem to be a reasonable results if the two population mean drying times truly are equal?
No.
(b) If someone selected 10,000 pairs of samples of size 18 under the condition that $\mu_A = \mu_B$, in how many of these 10,000 experiments would there be a difference $\bar{x}_A - \bar{x}_B$ is as large as 1.0?
26.

Examples

Example 7.(Cont.) Suppose a difference of 1.0 in means was observed in real samples.

(a) Does this seem to be a reasonable results if the two population mean drying times truly are equal?

No.

(b) If someone selected 10,000 pairs of samples of size 18 under the condition that $\mu_A = \mu_B$, in how many of these 10,000 experiments would there be a difference $\bar{x}_A - \bar{x}_B$ is as large as 1.0?

26.

Examples

Example 7.(Cont.) Suppose a difference of 1.0 in means was observed in real samples.
(a) Does this seem to be a reasonable results if the two population mean drying times truly are equal?
No.
(b) If someone selected 10,000 pairs of samples of size 18 under the condition that $\mu_A = \mu_B$, in how many of these 10,000 experiments would there be a difference $\bar{x}_A - \bar{x}_B$ is as large as 1.0?
26.

Examples

Example 7.(Cont.) Suppose a difference of 1.0 in means was observed in real samples.
(a) Does this seem to be a reasonable results if the two population mean drying times truly are equal?
No.
(b) If someone selected 10,000 pairs of samples of size 18 under the condition that $\mu_A = \mu_B$, in how many of these 10,000 experiments would there be a difference $\bar{x}_A - \bar{x}_B$ is as large as 1.0?
26.