Chapter 7. Sampling Distributions

FPT Education
FPT UNIVERSITY

February 18, 2022

Statistical inference is concerned with making decisions about a population based on the information contained in a random sample from that population.

Before the data are collected, the observations are considered to be random variables, say $X_1, X_2, \ldots, X_n$.

Definition

(1) Statistic is a function of the observation. Statistic is also a random variable. For example, the sample mean and the sample variance are statistic. $(\bar{X}, S)$

(2) Sampling distribution is the probability distribution of a statistic.

Before the data are collected, the observations are considered to be random variables, say $X_1, X_2, \ldots, X_n$.

Definition

(1) Statistic is a function of the observation. Statistic is also a random variable. For example, the sample mean and the sample variance are statistic. $(\bar{X}, S)$

(2) Sampling distribution is the probability distribution of a statistic.

(3) A parameter is any numerical quantity that characterizes a given population such as: μ, σ^2, σ. The symbol θ can represent any parameter.

Before the data are collected, the observations are considered to be random variables, say $X_1, X_2, \ldots, X_n$.

Definition

(1) Statistic is a function of the observation. Statistic is also a random variable. For example, the sample mean and the sample variance are statistic. $(\bar{X}, S)$

(2) Sampling distribution is the probability distribution of a statistic.

(3) A parameter is any numerical quantity that characterizes a given population such as: μ, σ^2, σ. The symbol θ can represent any parameter.

(4) If X is a random variable with probability distribution , characterized by the unknown parameter θ, and if $X_1, X_2, \ldots, X_n$ is a random sample of size n from X, the statistic $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ is called a point estimator of θ.

Before the data are collected, the observations are considered to be random variables, say $X_1, X_2, \ldots, X_n$.

Definition

(1) Statistic is a function of the observation. Statistic is also a random variable. For example, the sample mean and the sample variance are statistic. $(\bar{X}, S)$

(2) Sampling distribution is the probability distribution of a statistic.

(3) A parameter is any numerical quantity that characterizes a given population such as: μ, σ^2, σ. The symbol θ can represent any parameter.

(4) If X is a random variable with probability distribution , characterized by the unknown parameter θ, and if $X_1, X_2, \ldots, X_n$ is a random sample of size n from X, the statistic $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ is called a point estimator of θ.

Populations (parameters need estimate)	Samples of size n (point estimators)
Mean μ	sample mean $\bar{x}$
Variance σ^2	Sample variance s^2
Proposition p of interested items	Sample proposition $\hat{p} = x/n$, x the number of interested items
Difference of means $\mu_1 - \mu_2$	Difference of sample means $\bar{x_1} - \bar{x_2}$
Difference of proposition $p_1 - p_2$	Difference of sample propositions $\hat{p_1} - \hat{p_2}$

FPT Education FPT UNIVERSITY

Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a normal population (either finite or infinite) with mean μ and finite variance σ^2. The the sample mean (random variable)

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

has a normal distribution with

- *mean*

$$\mu_{\bar{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

- *variance*

$$\sigma^2_{\bar{X}} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Note. If X and Y are two independent random variables then

$$\mu_{aX+bY} = a\mu_X + b\mu_Y$$
$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

Note. If X and Y are two independent random variables then

$$\mu_{aX+bY} = a\mu_X + b\mu_Y$$
$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

Example

The amount of time required for an oil and filter change on an automobile is normally distributed with a mean of 28 minutes and a standard deviation of 5 minutes. A random sample of 16 cars is selected. What is the probability that the sample mean is greater than 29?

Note. If X and Y are two independent random variables then

$$\mu_{aX+bY} = a\mu_X + b\mu_Y$$
$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

Example

The amount of time required for an oil and filter change on an automobile is normally distributed with a mean of 28 minutes and a standard deviation of 5 minutes. A random sample of 16 cars is selected. What is the probability that the sample mean is greater than 29?

Note. If X and Y are two independent random variables then

$$\mu_{aX+bY} = a\mu_X + b\mu_Y$$
$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

Example

The amount of time required for an oil and filter change on an automobile is normally distributed with a mean of 28 minutes and a standard deviation of 5 minutes. A random sample of 16 cars is selected. What is the probability that the sample mean is greater than 29?

Answer: The sample mean has a normal distribution with:

$$\mu_{\bar{X}} = 28, \sigma_{\bar{X}} = \frac{5}{\sqrt{16}} = 1.25$$

We want to compute $P(\bar{X} > 29)$. (Set $Z = \frac{\bar{X}-28}{1.25}$ is standard normal)

$$P(\bar{X} > 29) = P\left(Z > (29-28)/1.25\right) = P(Z > 0.8) = 1 - \Phi(0.8) = 0.212$$

Note. If X and Y are two independent random variables then

$$\mu_{aX+bY} = a\mu_X + b\mu_Y$$
$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

Example

The amount of time required for an oil and filter change on an automobile is normally distributed with a mean of 28 minutes and a standard deviation of 5 minutes. A random sample of 16 cars is selected. What is the probability that the sample mean is greater than 29?

Answer: The sample mean has a normal distribution with:

$$\mu_{\bar{X}} = 28, \sigma_{\bar{X}} = \frac{5}{\sqrt{16}} = 1.25$$

We want to compute $P(\bar{X} > 29)$. (Set $Z = \frac{\bar{X}-28}{1.25}$ is standard normal)

$$P(\bar{X} > 29) = P\left(Z > (29-28)/1.25\right) = P(Z > 0.8) = 1 - \Phi(0.8) = 0.212$$

Example

To determine the difference , if any, between two brands of radial tires, 16 tires of first brand (X) and 25 tires of second brand (Y) are tested. Assume that the lifetimes of both brands of tires come from the same normal distribution $N(90000, 25000)$. What is the distribution of the difference of the sample means $\bar{X} - \bar{Y}$?

Example

To determine the difference , if any, between two brands of radial tires, 16 tires of first brand (X) and 25 tires of second brand (Y) are tested. Assume that the lifetimes of both brands of tires come from the same normal distribution $N(90000, 25000)$. What is the distribution of the difference of the sample means $\bar{X} - \bar{Y}$?

Answer. We have $\bar{X}, \bar{Y}$ has normal distributions with

$$\mu_{\bar{X}} = 90000, \sigma^2_{\bar{X}} = \frac{25000}{16} = 1562.5$$
$$\mu_{\bar{Y}} = 90000, \sigma^2_{\bar{Y}} = \frac{25000}{25} = 1000$$

Hence $\bar{X} - \bar{Y}$ has a normal distribution with

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = 0$$
$$\sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}} = 1562.5 + 1000 = 2562.5$$

Example

To determine the difference , if any, between two brands of radial tires, 16 tires of first brand (X) and 25 tires of second brand (Y) are tested. Assume that the lifetimes of both brands of tires come from the same normal distribution $N(90000, 25000)$. What is the distribution of the difference of the sample means $\bar{X} - \bar{Y}$?

Answer. We have $\bar{X}$, $\bar{Y}$ has normal distributions with

$$\mu_{\bar{X}} = 90000, \sigma^2_{\bar{X}} = \frac{25000}{16} = 1562.5$$
$$\mu_{\bar{Y}} = 90000, \sigma^2_{\bar{Y}} = \frac{25000}{25} = 1000$$

Hence $\bar{X} - \bar{Y}$ has a normal distribution with

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = 0$$
$$\sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}} = 1562.5 + 1000 = 2562.5$$

Theorem (Central Limit Theorem)

Suppose $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2. Let

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

be the sample mean. Then

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

approximate standard normal when n large ($n \geq 30$).
It means as $n \geq 30$, we can consider $\bar{X}$ has a normal distribution with mean μ and variance σ^2/n.

Example

One year, professional sports players salaries averaged 1.5 million with a standard deviation of 0.9 million. Suppose a sample of 100 major league players was taken. Find the approximate probability that the average salary of the 100 players not exceeded 1.4 million.

Example

One year, professional sports players salaries averaged 1.5 million with a standard deviation of 0.9 million. Suppose a sample of 100 major league players was taken. Find the approximate probability that the average salary of the 100 players not exceeded 1.4 million.

Answer: The average salary $\bar{X}$ of the 100 players has

$$\mu_{\bar{X}} = 1.5, \sigma_{\bar{X}} = \frac{0.6}{\sqrt{100}} = 0.09$$

Hence

$$Z = \frac{\bar{X} - 1.5}{0.09}$$

is approximated the standard normal. Thus,

$$P(\bar{X} \leq 1.4) = P\left(Z \leq (1.4 - 1.5)/0.09\right)$$
$$= P(Z \leq -1.111) = \Phi(-1.111) = 0.133$$

Example

One year, professional sports players salaries averaged 1.5 million with a standard deviation of 0.9 million. Suppose a sample of 100 major league players was taken. Find the approximate probability that the average salary of the 100 players not exceeded 1.4 million.

Answer: The average salary $\bar{X}$ of the 100 players has

$$\mu_{\bar{X}} = 1.5, \sigma_{\bar{X}} = \frac{0.6}{\sqrt{100}} = 0.09$$

Hence

$$Z = \frac{\bar{X} - 1.5}{0.09}$$

is approximated the standard normal. Thus,

$$P(\bar{X} \leq 1.4) = P\left(Z \leq (1.4 - 1.5)/0.09\right)$$
$$= P(Z \leq -1.111) = \Phi(-1.111) = 0.133$$