Sampling distributions

Tiago M. Magalhães

Department of Statistics - ICE-UFJF

Juiz de Fora, November 04, 2026

Outline

1. Sample mean
2. Sample variance
3. Joint distribution
4. Two populations
5. Bibliography

Outline

1. Sample mean
2. Sample variance
3. Joint distribution
4. Two populations
5. Bibliography

Sample mean

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (IID) random variables (RVs), with distribution function $\mathcal{D}$, expected value μ, and variance σ^2, that is,

$$X_i \sim \mathcal{D}(\mu, \sigma^2),$$

$i = 1, 2, \ldots, n$. Such a sequence of RVs is said to constitute a random sample from the distribution $\mathcal{D}$.

Sample mean

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (IID) random variables (RVs), with distribution function $\mathcal{D}$, expected value μ, and variance σ^2, that is,

$$X_i \sim \mathcal{D}(\mu, \sigma^2),$$

$i = 1, 2, \ldots, n$. Such a sequence of RVs is said to constitute a random sample from the distribution $\mathcal{D}$.

Sample mean

The quantity,

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

is called the sample mean.

Sample mean

Note that,

$$\mathbb{E}\left(\bar{X}\right) = \mathbb{E}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu,$$

that is, the expected value of a sample mean is μ, the mean of the distribution.

Sample mean

Note that,

$$\mathbb{E}\left(\bar{X}\right) = \mathbb{E}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu,$$

that is, the expected value of a sample mean is μ, the mean of the distribution. When the mean of the distribution is unknown, the sample mean is frequently used to estimate it.

Sample mean

Note that,

$$\mathbb{E}\left(\bar{X}\right) = \mathbb{E}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu,$$

that is, the expected value of a sample mean is μ, the mean of the distribution. When the mean of the distribution is unknown, the sample mean is frequently used to estimate it.

Sample mean

Additionally,

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

$$\overset{\text{Ind.}}{=} \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}.$$

Deviations

The quantities,

$$X_i - \bar{X},$$

$i = 1, 2, \ldots, n,$

Deviations

The quantities,

$$X_i - \bar{X},$$

$i = 1, 2, \ldots, n$, are called deviations and they are equal to the differences between the individual values of the data and the sample mean.

Deviations

The quantities,

$$X_i - \bar{X},$$

$i = 1, 2, \ldots, n$, are called deviations and they are equal to the differences between the individual values of the data and the sample mean.

Deviations

And we also have the following relationship,

$$\begin{aligned}
\text{Cov}\left(X_i - \bar{X}, \bar{X}\right) &= \text{Cov}\left(X_i, \bar{X}\right) - \text{Cov}\left(\bar{X}, \bar{X}\right) \\
&= \text{Cov}\left(X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right) - \text{Var}\left(\bar{X}\right) \\
&= \frac{1}{n}\sum_{j=1}^{n} \text{Cov}\left(X_i, X_j\right) - \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0. \qquad (2)
\end{aligned}$$

Deviations

Although $\bar{X}$ and each individual deviation $X_i - \bar{X}$ are uncorrelated, they are not, in general, independent.

However, in the special case where the RVs X_i are normal,

Deviations

Although $\bar{X}$ and each individual deviation $X_i - \bar{X}$ are uncorrelated, they are not, in general, independent.

However, in the special case where the RVs X_i are normal, $\bar{X}$ is independent of the entire collection of deviations $X_i - \bar{X}$, $i = 1, \ldots, n$.

Deviations

Although $\bar{X}$ and each individual deviation $X_i - \bar{X}$ are uncorrelated, they are not, in general, independent.

However, in the special case where the RVs X_i are normal, $\bar{X}$ is independent of the entire collection of deviations $X_i - \bar{X}$, $i = 1, \ldots, n$.

Outline

1. Sample mean
2. Sample variance
3. Joint distribution
4. Two populations
5. Bibliography

Sample variance

Let $X_i \sim \mathcal{D}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ and $\bar{X}$ be the sample mean defined in (1). The RV,

$$S^2 = \sum_{i=1}^{n} \frac{\left(X_i - \bar{X}\right)^2}{n-1} \tag{3}$$

is called the sample variance.

Sample variance

Let $X_i \sim \mathcal{D}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ and $\bar{X}$ be the sample mean defined in (1). The RV,

$$S^2 = \sum_{i=1}^{n} \frac{\left(X_i - \bar{X}\right)^2}{n-1} \tag{3}$$

is called the sample variance.

Sample variance

We can write (3) in the following way:

$$
\begin{aligned}
(n-1)\,S^2 &= \sum_{i=1}^{n}\left(X_i - \mu + \mu - \bar{X}\right)^2 = \sum_{i=1}^{n}\left[(X_i-\mu) - \left(\bar{X}-\mu\right)\right]^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2 - 2\left(\bar{X}-\mu\right)\sum_{i=1}^{n}(X_i-\mu) + \sum_{i=1}^{n}\left(\bar{X}-\mu\right)^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2 - 2\left(\bar{X}-\mu\right)n\left(\bar{X}-\mu\right) + n\left(\bar{X}-\mu\right)^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2 - n\left(\bar{X}-\mu\right)^2. \qquad (4)
\end{aligned}
$$



Sample variance

Calculating the expected value of the previous equation, we obtain:

$$
\begin{aligned}
(n-1)\,\mathbb{E}\left(S^2\right) &= \sum_{i=1}^{n} \mathbb{E}\left[(X_i-\mu)^2\right] - n\mathbb{E}\left[\left(\bar{X}-\mu\right)^2\right] \\
&= n\sigma^2 - n\,\mathrm{Var}\left(\bar{X}\right) = (n-1)\,\sigma^2 \\
\Rightarrow \mathbb{E}\left(S^2\right) &= \sigma^2.
\end{aligned}
$$

Sample variance

Calculating the expected value of the previous equation, we obtain:

$$
\begin{aligned}
(n-1)\,\mathbb{E}\left(S^2\right) &= \sum_{i=1}^{n} \mathbb{E}\left[(X_i-\mu)^2\right] - n\mathbb{E}\left[\left(\bar{X}-\mu\right)^2\right] \\
&= n\sigma^2 - n\,\mathrm{Var}\left(\bar{X}\right) = (n-1)\,\sigma^2 \\
\Rightarrow \mathbb{E}\left(S^2\right) &= \sigma^2.
\end{aligned}
$$

Thus, the expected value of the sample variance is the variance of the distribution $\mathcal{D}$.

Sample variance

Calculating the expected value of the previous equation, we obtain:

$$
\begin{aligned}
(n-1)\,\mathbb{E}\left(S^2\right) &= \sum_{i=1}^{n} \mathbb{E}\left[(X_i-\mu)^2\right] - n\mathbb{E}\left[\left(\bar{X}-\mu\right)^2\right] \\
&= n\sigma^2 - n\,\mathrm{Var}\left(\bar{X}\right) = (n-1)\,\sigma^2 \\
\Rightarrow \mathbb{E}\left(S^2\right) &= \sigma^2.
\end{aligned}
$$

Thus, the expected value of the sample variance is the variance of the distribution $\mathcal{D}$.

Outline

1. Sample mean
2. Sample variance
3. Joint distribution
4. Two populations
5. Bibliography

Joint distribution of the sample mean and sample variance

Let $X_1, X_2, \ldots, X_n$ be normal IID RVs with expected value μ and variance σ^2. Let also $\bar{X}$ be the sample mean defined in (1).

Joint distribution of the sample mean and sample variance

Let $X_1, X_2, \ldots, X_n$ be normal IID RVs with expected value μ and variance σ^2. Let also $\bar{X}$ be the sample mean defined in (1). We have that,

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right). \tag{5}$$

Joint distribution of the sample mean and sample variance

Let $X_1, X_2, \ldots, X_n$ be normal IID RVs with expected value μ and variance σ^2. Let also $\bar{X}$ be the sample mean defined in (1). We have that,

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right). \tag{5}$$

Joint distribution of the sample mean and sample variance

To prove (5), we can use the moment-generating function (MGF). Recalling that the MGF of $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$, is given by:

$$M_{X_i}(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}.$$

Joint distribution of the sample mean and sample variance

To prove (5), we can use the moment-generating function (MGF). Recalling that the MGF of $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$, is given by:

$$M_{X_i}(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}.$$

Joint distribution of the sample mean and sample variance

The MGF of $\bar{X}$, $M_{\bar{X}}(t)$, is given by:

$$
\begin{aligned}
M_{\bar{X}}(t) &= \mathbb{E}\left[e^{t\bar{X}}\right] = \mathbb{E}\left[e^{t\sum_{i=1}^{n}\frac{X_i}{n}}\right] = \mathbb{E}\left[e^{\frac{1}{n}tX_1} \times e^{\frac{1}{n}tX_2} \times \cdots \times e^{\frac{1}{n}tX_n}\right] \\
&\overset{\text{Ind.}}{=} \mathbb{E}\left[e^{\frac{1}{n}tX_1}\right] \times \mathbb{E}\left[e^{\frac{1}{n}tX_2}\right] \times \cdots \times \mathbb{E}\left[e^{\frac{1}{n}tX_n}\right] \\
&= M_{X_1}\left(\frac{t}{n}\right) \times M_{X_2}\left(\frac{t}{n}\right) \times \cdots \times M_{X_n}\left(\frac{t}{n}\right) \\
&\overset{\text{I.D.}}{=} \left[M_{X_1}\left(\frac{t}{n}\right)\right]^n = \left\{\exp\left[\mu\left(\frac{t}{n}\right) + \frac{\sigma^2}{2}\left(\frac{t}{n}\right)^2\right]\right\}^n \\
&= \exp\left\{n\left[\mu\left(\frac{t}{n}\right) + \frac{\sigma^2}{2}\left(\frac{t}{n}\right)^2\right]\right\} = \exp\left\{\mu t + \frac{1}{2}\frac{\sigma^2}{n}t^2\right\}.
\end{aligned}
$$

Thus, $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$.

Joint distribution of the sample mean and sample variance

The MGF of $\bar{X}$, $M_{\bar{X}}(t)$, is given by:

$$
\begin{aligned}
M_{\bar{X}}(t) &= \mathbb{E}\left[e^{t\bar{X}}\right] = \mathbb{E}\left[e^{t\sum_{i=1}^{n}\frac{X_i}{n}}\right] = \mathbb{E}\left[e^{\frac{1}{n}tX_1} \times e^{\frac{1}{n}tX_2} \times \cdots \times e^{\frac{1}{n}tX_n}\right] \\
&\overset{\text{Ind.}}{=} \mathbb{E}\left[e^{\frac{1}{n}tX_1}\right] \times \mathbb{E}\left[e^{\frac{1}{n}tX_2}\right] \times \cdots \times \mathbb{E}\left[e^{\frac{1}{n}tX_n}\right] \\
&= M_{X_1}\left(\frac{t}{n}\right) \times M_{X_2}\left(\frac{t}{n}\right) \times \cdots \times M_{X_n}\left(\frac{t}{n}\right) \\
&\overset{\text{I.D.}}{=} \left[M_{X_1}\left(\frac{t}{n}\right)\right]^n = \left\{\exp\left[\mu\left(\frac{t}{n}\right) + \frac{\sigma^2}{2}\left(\frac{t}{n}\right)^2\right]\right\}^n \\
&= \exp\left\{n\left[\mu\left(\frac{t}{n}\right) + \frac{\sigma^2}{2}\left(\frac{t}{n}\right)^2\right]\right\} = \exp\left\{\mu t + \frac{1}{2}\frac{\sigma^2}{n}t^2\right\}.
\end{aligned}
$$

Thus, $\bar{X} \sim \mathrm{N}\left(\mu, \sigma^2/n\right)$.

Joint distribution of the sample mean and sample variance

Similarly, let $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ and S^2 be the sample variance as defined in (3). To find the distribution of S^2, first, we need to rewrite it as follows, see (4):

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

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

$$\Rightarrow \frac{(n-1)\,S^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 - \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 .$$

Joint distribution of the sample mean and sample variance

Similarly, let $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ and S^2 be the sample variance as defined in (3). To find the distribution of S^2, first, we need to rewrite it as follows, see (4):

$$
\begin{aligned}
(n-1)\,S^2 &= \sum_{i=1}^{n} (X_i - \mu)^2 - n\left(\bar{X} - \mu\right)^2 \\
\Rightarrow \frac{1}{\sigma^2}(n-1)\,S^2 &= \frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - \mu)^2 - \frac{1}{\sigma^2} n\left(\bar{X} - \mu\right)^2 \\
\Rightarrow \frac{(n-1)\,S^2}{\sigma^2} &= \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 - \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 .
\end{aligned}
$$

Joint distribution of the sample mean and sample variance

Since $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$,

$$\frac{X_i - \mu}{\sigma} \sim \mathrm{N}(0,1), \quad \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_1, \text{ and } \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n.$$

And, since $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, it follows that

$$\frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \sim \mathrm{N}(0,1), \text{ and } \left[\frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma}\right]^2 \sim \chi^2_1.$$

Joint distribution of the sample mean and sample variance

Since $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$,

$$\frac{X_i - \mu}{\sigma} \sim \mathrm{N}(0,1), \quad \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_1, \text{ and } \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n.$$

And, since $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, it follows that

$$\frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \sim \mathrm{N}(0,1), \text{ and } \left[\frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma}\right]^2 \sim \chi^2_1.$$

Joint distribution of the sample mean and sample variance

Therefore,

$$\frac{(n-1)\,S^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2_{n-1}. \tag{6}$$

Remark: Since $\bar{X}$ and the deviations $X_i - \bar{X}$ have zero covariance, see (2), the assumption of normal populations guarantees their independence and, consequently, that $\bar{X}$ and S^2 are independent RVs.

Joint distribution of the sample mean and sample variance

Therefore,

$$\frac{(n-1)\,S^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2_{n-1}. \tag{6}$$

Remark: Since $\bar{X}$ and the deviations $X_i - \bar{X}$ have zero covariance, see (2), the assumption of normal populations guarantees their independence and, consequently, that $\bar{X}$ and S^2 are independent RVs.

Joint distribution of the sample mean and sample variance

Result

Let $X_1, X_2, \ldots, X_n$ be normal IID RVs with expected value μ and variance σ^2, then the sample mean $\bar{X}$ and the sample variance S^2 are independent.

Joint distribution of the sample mean and sample variance

Result

Let $X_1, X_2, \ldots, X_n$ be normal IID RVs with expected value μ and variance σ^2, then the sample mean $\bar{X}$ and the sample variance S^2 are independent.

$\bar{X}$ is a normal RV with mean μ and variance σ^2/n and S^2 is a chi-squared RV with $n-1$ degrees of freedom.

Joint distribution of the sample mean and sample variance

Result

Let $X_1, X_2, \ldots, X_n$ be normal IID RVs with expected value μ and variance σ^2, then the sample mean $\bar{X}$ and the sample variance S^2 are independent.

$\bar{X}$ is a normal RV with mean μ and variance σ^2/n and S^2 is a chi-squared RV with $n-1$ degrees of freedom.

Outline

1. Sample mean
2. Sample variance
3. Joint distribution
4. Two populations
5. Bibliography

Two populations

Let there be two independent samples such that $X_{11}, X_{12}, \ldots, X_{1n_1}$ are normal IID RVs with expected value μ_1 and variance σ_1^2 and $X_{21}, X_{22}, \ldots, X_{2n_2}$ are normal IID RVs with expected value μ_2 and variance σ_2^2. By (6), we have that,

$$Q_1 = \frac{(n_1 - 1)\, S_1^2}{\sigma_1^2} \sim \chi^2_{n_1 - 1} \text{ and } Q_2 = \frac{(n_2 - 1)\, S_2^2}{\sigma_2^2} \sim \chi^2_{n_2 - 1},$$

Two populations

Let there be two independent samples such that $X_{11}, X_{12}, \ldots, X_{1n_1}$ are normal IID RVs with expected value μ_1 and variance σ_1^2 and $X_{21}, X_{22}, \ldots, X_{2n_2}$ are normal IID RVs with expected value μ_2 and variance σ_2^2. By (6), we have that,

$$Q_1 = \frac{(n_1 - 1)\, S_1^2}{\sigma_1^2} \sim \chi^2_{n_1 - 1} \text{ and } Q_2 = \frac{(n_2 - 1)\, S_2^2}{\sigma_2^2} \sim \chi^2_{n_2 - 1},$$

where S_1^2 and S_2^2 are the sample variances of populations 1 and 2, respectively.

Two populations

Let there be two independent samples such that $X_{11}, X_{12}, \ldots, X_{1n_1}$ are normal IID RVs with expected value μ_1 and variance σ_1^2 and $X_{21}, X_{22}, \ldots, X_{2n_2}$ are normal IID RVs with expected value μ_2 and variance σ_2^2. By (6), we have that,

$$Q_1 = \frac{(n_1 - 1)\, S_1^2}{\sigma_1^2} \sim \chi^2_{n_1 - 1} \text{ and } Q_2 = \frac{(n_2 - 1)\, S_2^2}{\sigma_2^2} \sim \chi^2_{n_2 - 1},$$

where S_1^2 and S_2^2 are the sample variances of populations 1 and 2, respectively.

Two populations

A very important result states that the ratio between the ratios of the RVs Q_1 and Q_2 to their respective degrees of freedom follows Snedecor's F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, that is,

$$\frac{Q_1/(n_1-1)}{Q_2/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathrm{F}(n_1-1, n_2-1). \tag{7}$$

Two populations

A very important result states that the ratio between the ratios of the RVs Q_1 and Q_2 to their respective degrees of freedom follows Snedecor's F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, that is,

$$\frac{Q_1/(n_1-1)}{Q_2/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathrm{F}(n_1 - 1, n_2 - 1). \tag{7}$$

The ratio (7) and, consequently, the F-distribution are frequently used in **analysis of variance**.

Two populations

A very important result states that the ratio between the ratios of the RVs Q_1 and Q_2 to their respective degrees of freedom follows Snedecor's F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, that is,

$$\frac{Q_1/(n_1-1)}{Q_2/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathrm{F}(n_1 - 1, n_2 - 1). \tag{7}$$

The ratio (7) and, consequently, the F-distribution are frequently used in **analysis of variance**.

Outline

1. Sample mean
2. Sample variance
3. Joint distribution
4. Two populations
5. Bibliography

Bibliography

Meyer, P. L. (2017). *Probabilidade: aplicações à Estatística*. Livros Técnicos e Científicos Editora Ltda, Rio de Janeiro, 2nd edition.

Ross, S. (2019). *A first course in probability*. Pearson Education, Malaysia, 10th edn. global edition.

Thank you!

✉ tiago.magalhaes@ufjf.br

ufjf.br/tiago_magalhaes

Department of Statistics, Room 319