# Statistics Notes

Devesh

December 21, 2024

## 1 Measurements of Central Tendency

### 1.1 Mean

The mean, or average, is the sum of all values divided by the number of values. It is a measure of the central tendency of a set of numbers.

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n}$$

### 1.2 Median

The median is the middle value in a list of numbers. To find the median, the numbers must be arranged in numerical order. If there is an even number of observations, the median is the average of the two middle numbers.

$$\text{Median} = \begin{cases} x_{(n+1)/2} & \text{if } n \text{ is odd} \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2} & \text{if } n \text{ is even} \end{cases}$$

### 1.3 Mode

The mode is the value that appears most frequently in a data set. A set of numbers may have one mode, more than one mode, or no mode at all.

$$\text{Mode} = \text{most frequent value in the data set}$$

### 1.4 Variance

Variance measures how far a set of numbers are spread out from their average value. It is the average of the squared differences from the mean.

$$\text{Variance} = \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

### 1.5 Standard Deviation

The standard deviation is the square root of the variance. It provides a measure of the average distance from the mean.

$$\text{Standard Deviation} = \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

### 1.6 Skewness

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. If the skewness is negative, the data are skewed to the left, meaning the left tail is longer or

fatter than the right tail. If the skewness is positive, the data are skewed to the right, meaning the right tail is longer or fatter than the left tail. A skewness of zero indicates that the data are perfectly symmetrical.

$$\text{Skewness} = \frac{\sum_{i=1}^{n}(x_i - \mu)^3}{n\sigma^3}$$

### 1.7 Standard Error

The standard error is the standard deviation of the sampling distribution of a statistic, most commonly of the mean. It provides an estimate of the variability of the sample mean.

$$\text{Standard Error} = \text{SE} = \frac{\sigma}{\sqrt{n}}$$

### 1.8 Gaussian Distribution

The Gaussian distribution, also known as the normal distribution, is a continuous probability distribution characterized by a bell-shaped curve. It is defined by two parameters: the mean ($\mu$) and the standard deviation ($\sigma$). The probability density function (PDF) of a Gaussian distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where: - $x$ is the variable - $\mu$ is the mean - $\sigma$ is the standard deviation
The Gaussian distribution is symmetric about the mean, and its shape is determined by the standard deviation. A larger standard deviation results in a wider and flatter curve, while a smaller standard deviation results in a narrower and taller curve.

## 2 Central Limit Theorem

The Central Limit Theorem (CLT) states that the distribution of the sample mean of a large number of independent, identically distributed random variables approaches a normal distribution, regardless of the original distribution of the variables. This theorem is fundamental in statistics because it allows for the use of normal distribution approximations in various statistical methods.
Formally, if $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with mean $\mu$ and variance $\sigma^2$, then the sample mean $\bar{X}$ is approximately normally distributed with mean $\mu$ and variance $\frac{\sigma^2}{n}$ for sufficiently large $n$. Mathematically, this can be expressed as:

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

The CLT is important because it justifies the use of the normal distribution in many practical applications, even when the underlying data do not follow a normal distribution.

## 3 Z-Score

The Z-score measures how many standard deviations an element is from the mean, allowing for comparison between different data sets.
The Z-score formula is:

$$Z = \frac{X - \mu}{\sigma}$$

where: - $X$ is the value - $\mu$ is the mean - $\sigma$ is the standard deviation
A Z-score of 0 indicates the element is at the mean. Positive or negative Z-scores indicate the element is above or below the mean, respectively.

# 4 P-Values

A p-value is the probability of obtaining test results at least as extreme as the observed results, assuming that the null hypothesis is true. It is used in hypothesis testing to determine the significance of the results. The p-value is calculated using the cumulative distribution function (CDF) of the test statistic under the null hypothesis. A smaller p-value indicates stronger evidence against the null hypothesis.

## 4.1 Example: Probability of Sampling at Least 2.5 Standard Deviations from the Mean

To find the probability of sampling at least 2.5 standard deviations from the mean in a standard normal distribution, we can use the Z-score and the properties of the normal distribution.
The Z-score formula is:

$$Z = \frac{X - \mu}{\sigma}$$

For a standard normal distribution, $\mu = 0$ and $\sigma = 1$. Therefore, the Z-score for 2.5 standard deviations from the mean is:

$$Z = 2.5$$

The probability of sampling at least 2.5 standard deviations from the mean is the sum of the probabilities in the two tails of the distribution:

$$P(|Z| \geq 2.5) = P(Z \leq -2.5) + P(Z \geq 2.5)$$

Using the standard normal distribution table or a calculator, we find:

$$P(Z \geq 2.5) \approx 0.0062$$

Since the normal distribution is symmetric:

$$P(Z \leq -2.5) = P(Z \geq 2.5) \approx 0.0062$$

Therefore, the total probability is:

$$P(|Z| \geq 2.5) = 2 \times 0.0062 = 0.0124$$

So, the probability of sampling at least 2.5 standard deviations from the mean is approximately 0.0124, or 1.24

# 5 Null Hypothesis

The null hypothesis, denoted as $H_0$, is a statement that there is no effect or no difference, and it serves as the default or starting assumption in hypothesis testing. It is the hypothesis that researchers aim to test against.
In hypothesis testing, the null hypothesis is typically tested against an alternative hypothesis, denoted as $H_a$ or $H_1$, which represents a new effect or difference that the researcher wants to prove.
The steps to test a null hypothesis are as follows:

- Formulate the null hypothesis ($H_0$) and the alternative hypothesis ($H_a$).
- Choose a significance level ($\alpha$), commonly set at 0.05.
- Collect data and calculate a test statistic.
- Determine the p-value, which is the probability of observing the test statistic or something more extreme under the null hypothesis.
- Compare the p-value to the significance level:

- If $p \leq \alpha$, reject the null hypothesis ($H_0$).
- If $p > \alpha$, fail to reject the null hypothesis ($H_0$).

Rejecting the null hypothesis suggests that there is sufficient evidence to support the alternative hypothesis. Failing to reject the null hypothesis suggests that there is not enough evidence to support the alternative hypothesis.

### 5.1 Example 1

Suppose we want to test whether a new drug is effective in lowering blood pressure. The null and alternative hypotheses might be:

$$H_0 : \text{The new drug has no effect on blood pressure.}$$

$$H_a : \text{The new drug lowers blood pressure.}$$

Let's say we conduct an experiment with 30 participants and measure their blood pressure before and after taking the drug. The mean decrease in blood pressure is found to be 5 mmHg with a standard deviation of 8 mmHg.
We perform a one-sample t-test to determine if the mean decrease is significantly different from zero. The test statistic is calculated as:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{5 - 0}{8/\sqrt{30}} \approx 3.42$$

Using a t-distribution table with 29 degrees of freedom, we find the p-value corresponding to $t = 3.42$. The p-value is approximately 0.002.
Since the p-value (0.002) is less than the chosen significance level (0.05), we reject the null hypothesis and conclude that the new drug is effective in lowering blood pressure.

### 5.2 Example 2

With a fair coin, the probability of throwing six heads or six tails in a six-coin-flip experiment is 0.03125 ($\rho = -0.015625$ for either of six heads or six tails). If a friend of yours hands you a coin, the null hypothesis (the baseline assumed by the fair-toss distribution) would be that the coin is fair. If you test this coin by flipping it six times and it comes up heads on all six or tails on all six, this observation would suggest that you should reject the null hypothesis (theres a good chance that the coin is not fair) because chance alone would facilitate such an observation less than 5% of the time, i.e., .

## 6 Binomial Distribution

The binomial distribution is a discrete probability distribution that describes the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success. It is defined by two parameters: $n$ (the number of trials) and $p$ (the probability of success in each trial).
The probability mass function (PMF) of the binomial distribution is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where: - $X$ is the random variable representing the number of successes - $k$ is the number of successes - $\binom{n}{k}$ is the binomial coefficient, calculated as $\frac{n!}{k!(n-k)!}$ - $p$ is the probability of success - $1 - p$ is the probability of failure

### 6.1 Example: Coin Toss

Suppose we have a fair coin (i.e., $p = 0.5$) and we flip it 10 times. We want to find the probability of getting exactly 6 heads.
Here, $n = 10$ and $p = 0.5$. The number of successes $k = 6$.
Using the binomial formula:

$$P(X = 6) = \binom{10}{6}(0.5)^6(0.5)^4$$

First, calculate the binomial coefficient:

$$\binom{10}{6} = \frac{10!}{6!4!} = 210$$

Then, calculate the probability:

$$P(X = 6) = 210 \times (0.5)^{10} = 210 \times \frac{1}{1024} \approx 0.205$$

## 7 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a method used to estimate the parameters of a statistical model. The goal of MLE is to find the parameter values that maximize the likelihood function, which measures how well the model explains the observed data.
Given a set of independent and identically distributed data points $X = \{x_1, x_2, \ldots, x_n\}$ and a probability density function $f(x; \theta)$ parameterized by $\theta$, the likelihood function $L(\theta)$ is defined as:

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

The log-likelihood function, which is often easier to work with, is given by:

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$$

The MLE for $\theta$ is the value that maximizes the log-likelihood function:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \ell(\theta)$$

### 7.1 Example: Estimating the Mean of a Normal Distribution

Suppose we have a sample $X = \{x_1, x_2, \ldots, x_n\}$ from a normal distribution with unknown mean $\mu$ and known variance $\sigma^2$. The probability density function is:

$$f(x; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

The log-likelihood function is:

$$\ell(\mu) = \sum_{i=1}^{n} \log f(x_i; \mu) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

To find the MLE for $\mu$, we take the derivative of $\ell(\mu)$ with respect to $\mu$ and set it to zero:

$$\frac{\partial \ell(\mu)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

Solving for $\mu$, we get:

$$\hat{\mu}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Thus, the MLE for the mean $\mu$ of a normal distribution is the sample mean.
So, the probability of getting exactly 6 heads in 10 flips of a fair coin is approximately 0.205, or 20.5