

Topic 4. Normal Distribution

Cheng Peng

Contents

| | | |
|----------|--|----------|
| 1 | Introduction | 1 |
| 2 | Standard Normal Distribution | 1 |
| 2.1 | Finding Probabilities | 2 |
| 2.2 | Finding Percentiles | 2 |
| 3 | General Normal Distribution | 4 |
| 3.1 | Finding Probabilities | 4 |
| 3.2 | Finding Percentiles | 4 |
| 4 | Use of Technology | 5 |
| 4.1 | Example 1 - Standard normal distribution | 6 |
| 4.2 | Example 2 - General normal distribution | 6 |

1 Introduction

The standard normal distribution is the most frequently used distribution in this course. We will introduce the general normal distributions and focus on two basic questions: Finding probabilities and percentiles using the standard normal table.

2 Standard Normal Distribution

The standard normal distribution, Z , has a mean of $\mu = 0$ and a standard deviation of $\sigma = 1$. Its probability density curve is

Two basic types of questions need to be answered for any distribution including the standard normal distribution:

1. Finding probability of events such as $P(Z < a)$, $P(Z > c)$, $P(a < Z < b)$, etc.
2. Finding percentiles. For example, finding z_0 for given $P(Z < z_0) = 0.90$.

We have discussed how to find probabilities from the uniform distribution in topic #3 whose density curve is a rectangle. The probabilities of events defined based on uniform distributions are the areas of rectangles using the area formula of a rectangle. We still need to find the probability of events defined based on the standard normal distribution like $P(-0.86 < Z < 0)$ which is still the area of the following shaded region (as outlined in the previous note for any general distribution).

2.1 Finding Probabilities

Unlike uniform distributions whose density curves are rectangles, we can use the formula to calculate the areas of rectangles. In a standard normal distribution, there is no formula to calculate the area of the shaded irregular region in the above figure.

We will use a standard normal distribution table to find the area of the left-hand side tail regions shown below (part of the table).

Before doing examples, we point out the basic facts of the standard normal distribution.

1. The density curve is symmetric concerning the vertical axis.
2. The area between the curve and the horizontal axis is equal to 1. This means that the areas of the left and right regions are equal to 0.5.

Next, we use several examples to illustrate how to use the standard normal table to find the areas of different regions defined based on the standard normal distribution.

Example 1. Find the probabilities indicated, where as always Z denotes a standard normal random variable.

- 1). $P(Z < -1.48)$.
- 2). $P(Z < 0.25)$.

Solution. First of all, we only keep two decimal places for the z value (also called z -score). When using the table, we first locate the integral part and the first decimal place of the z score in the **first column** and the second decimal place in the **top row**. The **left tail area** is on the intersection of the aforementioned row and column. This is explained in the following figure.

Example 2. Find the probabilities indicated, where as always Z denotes a standard normal random variable.

- 1). $P(Z > -1.96)$.
- 2). $P(Z > 0.75)$.

Solution: The probabilities to be found represent the areas of right tail regions. We can use the table to find the area of the left tail region and then subtract it from 1 to get the desired probability. The following figure illustrates the idea to get the “right-tail” probabilities.

Example 3. Find the probabilities indicated, where as always Z denotes a standard normal random variable.

- 1). $P(-1.96 < Z < 0.75)$.

Solution: The general idea is to find the two **left tail areas** and then take the difference to get the area of the region defined by the two z -scores as shown in the following figure.

For 1), the probability is found in the following figure.

Summary of Finding Probabilities

2.2 Finding Percentiles

We have introduced how to find a percentile from a given data set. We basically do the same thing for the standard normal distribution.

Recall that the q -th percentile is the cut-off value such that $100q\%$ data values are less than or equal to the cut-off value (q -th percentile). This means, we find a percentile, and the left tail area is always given. The general formulation of the problem is to find the cut-off k that satisfies

$$P(Z < k) = 0.90,$$

for a given q (such as 90%, etc). This is depicted in the following figure.

The process of finding a percentile is the opposite of the process of finding probability. If the given left tail probability itself is in the main body of the table, we then locate the row and the column to find the z -score (i.e., the percentile).

In general, the given left tail probability is not in the table but is closest to two values in the main body of the table. Each of the two closest table values corresponds to a z -score. The average of the two z -scores is defined to be the desired percentile.

Example 4. Find 90th percentile of the standard normal distribution.

Solution: We go to the normal table and find two values in the main body of the table that is closest to 0.9 (see the figure below).

Example 5. The Precision Scientific Instrument Company manufactures thermometers that are supposed to give readings of $0^\circ C$ at the freezing point of water. Tests on a large sample of these instruments reveal that the freezing point of water is around zero (some thermometers give positive degrees, some thermometers give negative degrees), Assume that the mean reading is $0^\circ C$ and the standard deviation of the readings is $1.00^\circ C$. Assume further that the readings are normally distributed.

1. Find the probability that, at the freezing point of water, the reading is between $0^\circ C$ and $1.58^\circ C$.
2. Find the probability that the reading is between $-2.43^\circ C$ and $0^\circ C$.
3. Find the probability that the reading is between $0.5^\circ C$ and $2.5^\circ C$.
4. Find the probability that the reading is between $-1^\circ C$ and $-2.5^\circ C$.
5. Find the probability that the reading is between $-1.5^\circ C$ and $1^\circ C$.
6. Find the probability that the reading is exactly $0^\circ C$.
7. Find the temperature z corresponding to P_{95} , the 95th percentile (95% of the readings less than z and 5% of the readings are greater than z).
8. Find the 10th percentile.

Solution: Based on the given information, the thermometer readings follow the standard normal distribution. The standard normal distribution table will be used to answer the above questions. We only do questions 5 (finding probability) and 7 (finding percentile) to work and leave the rest of the questions to you to practice.

5). $P(-1.5 < Z < 0) = P(Z < 0) - P(Z < -1.5) = 0.5 - P(Z < -1.5) = 0.5 - 0.0668 = 0.4332$

7). We want to find P_{95} , or equivalently, to find k from $P(Z < k) = 0.95$. We can see from the normal table that 0.9495 and 0.9505 are the two values that are closest to 0.95. The two corresponding z -scores are 1.64 and 1.65. By the convention, the 95th percentile is the average of the two z -scores (see the figure below).

3 General Normal Distribution

In practice, we rarely have a standard normal distribution. Many real-world problems are associated with general normal distribution. We still need to answer the two basic types of questions: finding probabilities and percentile. The following figure illustrates the two types of questions based on the normal distribution with a mean of 500 ($\mu = 500$) and a standard deviation of 100 ($\sigma = 100$)

The question is whether we cannot use the standard normal distribution table to answer the above two types of questions associated with general normal distribution.

We can use z-score transformation to transform general normal distributions to the standard normal distribution to use the table and then transform back the original general normal distribution. The following figure outlines the above idea.

3.1 Finding Probabilities

We use the following example to show the steps for finding the left-tail probabilities.

Example 6. Consider the general normal distribution $N(500, 100)$. Find $P(X < 600) = ?$

Solution. The following figure shows the z-score transformation to obtain the answer.

3.2 Finding Percentiles

We continue to use the previous normal distribution as an example to show how to find a percentile of the general normal distribution.

Example 7. Consider the general normal distribution $N(500, 100)$. Find the 15th percentile.

Solution: We are given that the left tail area is 0.15. After z-score transformation, the left tail area of the standard normal density curve is also 0.15 (see the following figure). We can find Z_0 from $P(Z < Z_0) = 0.15$ using the standard normal table which is $Z_0 \approx -1.04$.

Using the relationship between Z_0 and K in the z-score transformation (see the above figure). We have

$$-1.04 = \frac{K - 500}{100}$$

Solve for K , we have $K = 500 - 1.04 \times 100 = 396$.

Example 8. Tomkins Associates reports that the mean clear height for a Class A warehouse in the United States is 22 feet. Suppose clear heights are normally distributed and that the standard deviation is 4 feet. A Class A warehouse in the United States is randomly selected

- What is the probability that the clear height is greater than 17 feet?
- What is the probability that the clear height is less than 13 feet?
- What is the probability that the clear height is between 25 and 31 feet?
- Find the clear height such that 10% of all clear heights are less than it.

Solution The following figures outline the process of finding the answers to each of the questions.

a). $P(X > 17) = P(Z > -5/4) = 1 - P(Z < -5/4) = 0.8944$.

b). $P(X < 13) = P(Z < 9/4) = 0.012$.

c). $P(25 < X < 31) = P(3/4 < Z < 9/4) = P(Z < 9/4) - P(Z < 3/4) = 0.9878 - 0.9734 = 0.2144$.

d). Since $P(Z < Z_0) = 0.10$, we have $Z_0 = -1.28$ (from the normal table). The desired clear height (10th percentile) is $X = 22 - 1.28 \times 4 = 16.88$ feet.

The next video demonstrates how to use ISLA to answer the above questions.

Example 9. In redesigning jet ejection seats to better accommodate women as pilots, it is found that women's weights are normally distributed with a mean of 143 lb and a standard deviation of 29 lb.

- If a woman is randomly selected, what is the probability that she weighs between 163 lb and 201 lb?
- If the current ejection seat for men weighs between 130 lb and 211 lb, what percentage of women have weights that are within those limits?
- If a woman is randomly selected, what is the probability that she weighs less than 125 lb?
- If a woman is randomly selected, what is the probability that she weighs exactly 143 lb?
- If a woman is randomly selected, what is the probability that she weighs between 90 lb and 130 lb?
- Find the 10th percentile P_{10} , that is, the weight separating the bottom 10% from the top 90%.

Solution The following are brief solutions with graphical explanations.

a). $P(163 < X < 201) = P(0.69 < Z < 2) = P(Z < 2) - P(Z < 0.69) = 0.9772 - 0.7549 = 0.2223$.

b). $P(130 < X < 211) = P(-0.45 < Z < 2.35) = P(Z < 2.35) - P(Z < -0.45) = 0.9906 - 0.3264 = 0.6642$.

c). $P(X < 125) = P(Z < -0.62) = 0.2676$.

d). $P(X=143) = 0$.

e). $P(90 < X < 130) = P(-1.83 < Z < -0.45) = P(Z < -0.45) - P(Z < -1.83) = 0.3264 - 0.0336 = 0.2928$.

f). $P(Z < z_0) = 0.1$, so we get $z_0 = -1.285$. The P_{10} is calculated by $x = 143 - 1.285 \times 29 = 105.73$.

4 Use of Technology

ISLA has two apps to solve standard normal and general normal distribution problems.

- The interactive **standard normal table** can be found at: <https://wcupeng.shinyapps.io/ZTable/>
- Apps for solving general normal distribution problems can be found at <https://pengdsci.github.io/ISLA/ISLA-03.html>.

4.1 Example 1 - Standard normal distribution

We still use the **Thermometer** example with the following two questions.

- (1) Randomly select a thermometer, what is the probability the reading of this thermometer in the ice water is bigger than 0.5?
- (2) what is the cut-off reading that 75% of the readings of this type of thermometers in the ice water are higher than this cut-off?

4.2 Example 2 - General normal distribution

Blood Pressure The distribution of diastolic blood pressure for men is normally distributed with a mean of about 80 and a standard deviation of 20.

- (1) Randomly select a man from the population, what is the probability that his diastolic blood pressure is higher than 95.
- (2) What is the cut-off diastolic blood pressure that 90% diastolic blood pressures are higher than it?