Undoubtedly, the most widely used model for the distribution of a random variable is a **normal distribution.** Normal probability distribution or commonly called the *normal distribution* is one of the most important continuous theoretical distributions in Statistics. Most of the data relating to economic and business statistics or even in social and physical sciences conform to this distribution.

The normal distribution was first discovered by English Mathematician De-Moivre (1667-1754) in 1733 who obtained the mathematical equation for this distribution while dealing with problems arising in the game of chance. Normal distribution is also known as Gaussian distribution (Gaussian Law of Errors) after Karl Friedrich Gauss (1777-1855) who used this distribution to describe the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies. Today, normal probability model is one of the most important probability models in statistical analysis.

- We say that a random variable X follows the normal distribution if the probability density function of X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

  This is a bell-shaped curve.

- We write $X \sim N(\mu, \sigma)$. We read: X follows the normal distribution (or X is normally distributed) with mean μ, and standard deviation σ.

Definition

A normal random variable with

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

is called a **standard normal random variable** and is denoted as Z.

The cumulative distribution function of a standard normal random variable is denoted as

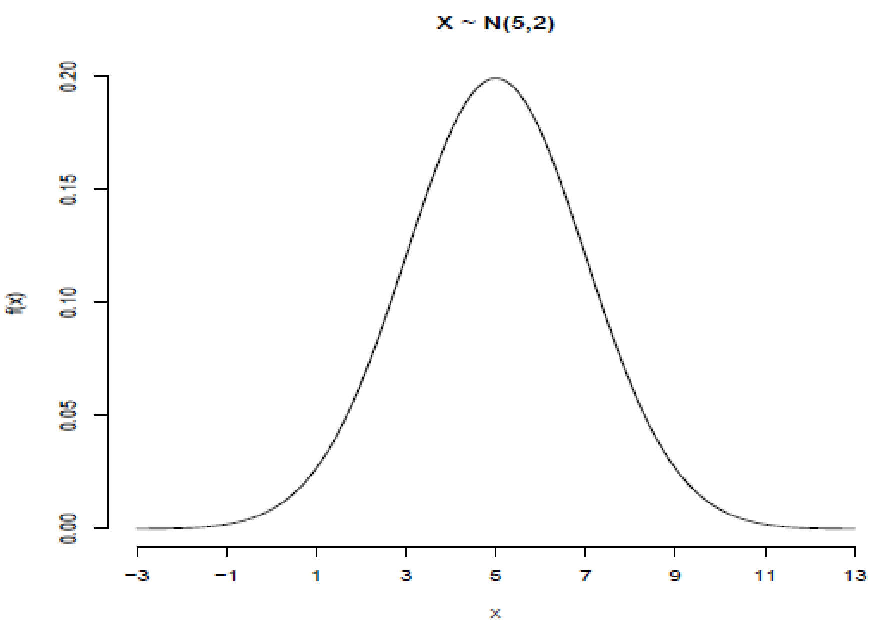
$$\Phi(z) = P(Z \le z)$$

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

$$Z = \frac{X - \mu}{\sigma} \tag{4-10}$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

- The normal distribution can be described completely by the two parameters μ and σ. As always, the mean is the center of the distribution and the standard deviation is the measure of the variation around the mean.
- Shape of the normal distribution. Suppose $X \sim N(5, 2)$.



- The area under the normal curve is 1 (100%).

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

Example 1: Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 $(\text{milliamperes})^2$.

i) What is the probability that a measurement will exceed 13 milliamperes?

ii) What is the probability that a current measurement is between 9 and 11 milliamperes?

Solution:

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$. Let $Z = (X - 10)/2$. The relationship between the several values of X and the transformed values of Z are shown in Fig. 4-15. We note that $X > 13$ corresponds to $Z > 1.5$. Therefore, from Appendix Table II,

$$P(X > 13) = P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - 0.93319 = 0.06681$$

Rather than using Fig. 4-15, the probability can be found from the inequality $X > 13$. That is,

$$P(X > 13) = P\left(\frac{(X - 10)}{2} > \frac{(13 - 10)}{2}\right) = P(Z > 1.5) = 0.06681$$

$$\begin{aligned} P(9 < X < 11) &= P((9 - 10)/2 < (X - 10)/2 < (11 - 10)/2) \\ &= P(-0.5 < Z < 0.5) = P(Z < 0.5) - P(Z < -0.5) \\ &= 0.69146 - 0.30854 = 0.38292 \end{aligned}$$

Example:

Suppose the diameter of a certain car component follows the normal distribution with $X \sim N(10, 3)$. Find the proportion of these components that have diameter larger than 13.4 mm. Or, if we randomly select one of these components, find the probability that its diameter will be larger than 13.4 mm.

Answer:

$$P(X > 13.4) = P(X - 10 > 13.4 - 10) =$$

$$P\left(\frac{X - 10}{3} > \frac{13.4 - 10}{3}\right) = P(Z > 1.13) = 1 - 0.8708 = 0.1292.$$

We read the number **0.8708** from the table. First we find the value of $z = 1.13$ (first column and first row of the table - the first row gives the second decimal of the value of z). Therefore the probability that the diameter is larger than 13.4 mm is 12.92%.

Suppose that the weight of navel oranges is normally distributed with mean $\mu = 8$ ounces, and standard deviation $\sigma = 1.5$ ounces. We can write $X \sim N(8, 1.5)$. Answer the following questions:

a. What proportion of oranges weigh more than 11.5 ounces? (or if you randomly select a navel orange, what is the probability that it weighs more than 11.5 ounces?).

$$P(X > 11.5) = P(Z > \frac{11.5 - 8}{1.5}) = P(Z > 2.33) = 1 - 0.9901 = 0.0099.$$

b. What proportion of oranges weigh less than 8.7 ounces?

$$P(X < 8.7) = P(Z < \frac{8.7 - 8}{1.5}) = P(Z < 0.47) = 0.6808.$$

c. What proportion of oranges weigh less than 5 ounces?

$$P(X < 5) = P(Z < \frac{5 - 8}{1.5}) = P(Z < -2.00) = 0.0228.$$

d. What proportion of oranges weigh more than 4.9 ounces?

$$P(X > 4.9) = P(Z > \frac{4.9 - 8}{1.5}) = P(Z > -2.07) = 1 - 0.0192 = 0.9808.$$

e. What proportion of oranges weigh between 6.2 and 7 ounces?

$$P(6.2 < X < 7) = P(\frac{6.2 - 8}{1.5} < Z < \frac{7 - 8}{1.5}) = P(-1.2 < Z < -0.67) = 0.2514 - 0.1151 = 0.1363.$$

f. What proportion of oranges weigh between 10.3 and 14 ounces?

$$P(10.3 < X < 14) = P(\frac{10.3-8}{1.5} < Z < \frac{14-8}{1.5}) = P(1.53 < Z < 4) \approx 1 - 0.9370 = 0.0630.$$

g. What proportion of oranges weigh between 6.8 and 8.9 ounces?

$$P(6.8 < X < 8.9) = P(\frac{6.8-8}{1.5} < Z < \frac{8.9-8}{1.5}) = P(-0.8 < Z < 0.6) = 0.7257 - 0.2119 = 0.5138.$$

h. Find the 80_{th} percentile of the distribution of X. This question can also be asked as follows: Find the value of X below which you find the lightest 80% of all the oranges.

$$z = \frac{x-\mu}{\sigma} \Rightarrow 0.845 = \frac{x-8}{1.5} \Rightarrow x = 9.27.$$

i. Find the 5_{th} percentile of the distribution of X.

$$z = \frac{x-\mu}{\sigma} \Rightarrow -1.645 = \frac{x-8}{1.5} \Rightarrow x = 5.53.$$

j. Find the interquartile range of the distribution of X.

Normal distribution - Examples

Example 1
The lengths of the sardines received by a certain cannery is normally distributed with mean 4.62 inches and a standard deviation 0.23 inch. What percentage of all these sardines is between 4.35 and 4.85 inches long?

Example 2
A baker knows that the daily demand for apple pies is a random variable which follows the normal distribution with mean 43.3 pies and standard deviation 4.6 pies. Find the demand which has probability 5% of being exceeded.

Example 3
Suppose that the height of $UCLA$ female students has normal distribution with mean 62 inches and standard deviation 8 inches.

a. Find the height below which is the shortest 30% of the female students.

b. Find the height above which is the tallest 5% of the female students.

Example 4
A firm's marketing manager believes that total sales for next year will follow the normal distribution, with mean of \$2.5 million and a standard deviation of $\$300,000$.

a. What is the probability that the firm's sales will fall within \$150000 of the mean?

b. Determine the sales level that has only a 9% chance of being exceeded next year.

Example 5
To avoid accusations of sexism in a college class equally populated by male and female students, the professor flips a fair coin to decide whether to call upon a male or female student to answer a question directed to the class. The professor will call upon a female student if a tails occurs. Suppose the professor does this 1000 times during the semester.

a. What is the probability that he calls upon a female student at least 530 times?

b. What is the probability that he calls upon a female student at most 480 times?

c. What is the probability that he calls upon a female student exactly 510 times?

Example 6
$MENSA$ is an organization whose members possess IQs in the top 2% of the population.

a. If IQs are normally distributed, with mean 100 and a standard deviation of 16, what is the minimum IQ required for admission to $MENSA$?

b. If three individuals are chosen at random from the general population what is the probability that all three satisfy the minimum requirement for $MENSA$?

Example 7
A manufacturing process produces semiconductor chips with a known failure rate 6.3%. Assume that chip failures are independent of one another. You will be producing 2000 chips tomorrow.

a. Find the expected number of defective chips produced.

b. Find the standard deviation of the number of defective chips.

c. Find the probability (approximate) that you will produce less than 135 defects.

EXERCISE 8

Suppose that the height (X) in inches, of a 25-year-old man is a normal random variable with mean $\mu = 70$ inches. If $P(X > 79) = 0.025$ what is the standard deviation of this random normal variable?

EXERCISE 9

Suppose that the weight (X) in pounds, of a 40-year-old man is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population is heavier than 214 pounds what is the mean μ of this distribution?

Problem 10

At *Heinz* ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.1 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of the bottle goes below 35.8 oz. or above 36.2 oz., then the bottle will be declared out of control.

a. If the process is in control, meaning $\mu = 36$ oz. and $\sigma = 0.1$ oz., find the probability that a bottle will be declared out of control.

b. In the situation of (a), find the probability that the number of bottles found out of control in an eight-hour day (16 inspections) will be zero.

c. In the situation of (a), find the probability that the number of bottles found out of control in an eight-hour day (16 inspections) will be exactly one.

d. If the process shifts so that $\mu = 37$ oz and $\sigma = 0.4$ oz, find the probability that a bottle will be declared out of control.

Problem 11

Suppose that a binary message -either 0 or 1- must be trasmitted by wire from location A to location B. However, the data sent over the wire are subject to a channel noise disturbance, so to reduce the possibilty of error, the value 2 is sent over the wire when the message is 1 and the value -2 is sent when the message is 0. If x, $x = \pm 2$, is the value sent from location A, then R, the value received at location B, is given by $R = x + N$, where N is the channel noise disturbance. When the message is received at location B the receiver decodes it according to the following rule:

If $R \geq 0.5$, then 1 is concluded

If $R < 0.5$, then 0 is concluded

If the channel noise follows the standard normal distribution compute the probability that the message will be wrong when decoded.

Normal distribution - Examples
Solutions

Example 1
We are given $X \sim N(4.62, 0.23)$. We want to compute

$$P(4.35 < X < 4.85) = P(\frac{4.35 - 4.62}{0.23} < Z < \frac{4.85 - 4.62}{0.23}) \quad = \\ P(-1.17 < z < 1) = \underline{0.8413 - 0.1210} = 0.7203.$$

Example 2
We are given $X \sim N(43.3, 4.6)$. We want to find the demand d such that $P(X > d) = 0.05$. From the standard normal table this corresponds to $z = 1.645$. Therefore $1.645 = \frac{d-43.3}{4.6} \Rightarrow d = 50.9$ pies.

Example 3
We are given $X \sim N(62, 8)$.

a. We want to find the height h such that $P(X < h) = 0.30$. From the standard normal table this corresponds to $z = -0.525$. Therefore $-0.525 = \frac{h-62}{8} \Rightarrow h = 57.8$ inches.

b. We want to find the height h such that $P(X > h) = 0.05$. From the standard normal table this corresponds to $z = 1.645$. Therefore $1.645 = \frac{h-62}{8} \Rightarrow h = 75.16$ inches.

Example 4
We are given $X \sim N(2500000, 300000)$.

a. $P(2350000 < X < 2650000) = P(\frac{2350000-250000}{300000} < z < \frac{2650000-2500000}{300000}) =$
$P(-0.5 < z < 0.5) = 0.6915 - 0.3085) = 0.3830.$

b. We want to find the sales level s such that $P(X > s) = 0.09$. This corresponds to $z = 1.345$. Therefore $1.345 = \frac{s-2500000}{300000} \Rightarrow s = 2903500$.

Example 5
This is a binomial problem but we are going to use the normal distribution as an approximation. We need μ and σ. These are: $\mu = np = 1000\frac{1}{2} = 500$. And $\sigma^2 = np(1-p) = 1000\frac{1}{2}(1-\frac{1}{2}) = 250 \Rightarrow \sigma = 15.81$.

a. $P(X \geq 530) = P(z > \frac{529.5-500}{15.81}) = P(z > 1.87) = 1 - 0.9693 = 0.0307.$

b. $P(X \leq 480) = P(z < \frac{480.5-500}{15.81}) = P(z < -1.23) = 0.1093.$

c. $P(X = 510) = P(\frac{509.5-500}{15.81} < z < \frac{510.5-500}{15.81}) = P(0.60 < z < 0.66) = 0.7454 - 0.7257 = 0.0197.$

Example 6
We are given $X \sim N(100, 16)$.

a. We want to find the IQ q such that $P(X > q) = 0.02$. This corresponds to $z = 2.055$. Therefore $2.055 = \frac{q-100}{16} \Rightarrow q = 132.88$.

b. This is binomial with $X \sim b(3, 0.02)$. We want $P(X = 3) = \binom{3}{3} 0.2^3 (1-0.02)^0 = 0.000008.$

Example 7
This is binomial with $X \sim b(2000, 0.063)$.

a. $E(X) = np = 2000(0.063) = 126$.

b. $\sigma^2 = np(1-p) = 2000(0.063)(1-0.063) = 118.06 \Rightarrow \sigma = 10.87$.

c. $P(X < 135) = P(z < \frac{134.5-126}{10.87}) = P(z < 0.78) = 0.7823$.

EXERCISE 8
We are given $X \sim N(70, \sigma)$. From $P(X > 79) = 0.025$ we find the corresponding z-value: $z = 1.96$. Therefore $1.96 = \frac{79-70}{\sigma} \Rightarrow \sigma = 4.59$ inches.

EXERCISE 9
We are given $X \sim N(\mu, 20)$. From $P(X > 214) = 0.05$ we find the corresponding z-value: $z = 1.645$. Therefore $1.645 = \frac{214-\mu}{20} \Rightarrow \mu = 181.1$ pounds.

Problem 10
The process is out of control if $P(X < 35.8)$ or $P(X > 36.2)$.

a. We are given $X \sim N(36, 0.1)$. We compute the probability:
$P(X < 35.8) + P(X > 36.2) = P(z < \frac{35.8-36}{0.1}) + P(z > \frac{36.2-36}{0.1}) = P(z < -2) + P(z > 2) = 0.0228 + (1-0.9772) = 0.0456$.

b. This is binomial with $n = 16, p = 0.0456$.
$P(X = 0) = \binom{16}{0}(0.0456)^0(1-0.0456)^{16} = 0.4739$.

c. This is binomial with $n = 16, p = 0.0456$.
$P(X = 1) = \binom{16}{1}(0.0456)^1(1-0.0456)^{15} = 0.3623$.

d. Now $X \sim N(37, 0.4)$. We compute the probability:
$P(X < 35.8) + P(X > 36.2) = P(z < \frac{35.8-37}{0.4}) + P(z > \frac{36.2-37}{0.4}) = P(z < -3) + P(z > -2) = 0.0013 + (1-0.0028) = 0.9785$.

Problem 11
The channel noise N follows the standard normal distribution, $N(0,1)$.
If the message was 1: It will be wrong when decoded if $R < 0.5$. Or $x+N < 0.5 \Rightarrow 2+N < 0.5 \Rightarrow N < -1.5$. This probability is equal to $P(z < -1.5) = 0.0668$. If the message was 0: It will be wrong when decoded if $R \geq 0.5$. Or $x+N \geq 0.5 \Rightarrow -2+N \geq 0.5 \Rightarrow N \geq 2.5$. This probability is equal to $P(z \geq 2.5) = 1-0.9938 = 0.0062$.

Normal distribution - Practice problems

Problem 1

The chickens of the Ornithes farm are processed when they are 20 weeks old. The distribution of their weights is normal with mean 3.8 lb, and standard deviation 0.6 lb. The farm has created three categories for these chickens according to their weight: petite (weight less than 3.5 lb), standard (weight between 3.5 lb and 4.9 lb), and big (weight above 4.9 lb).

a. What proportion of these chickens will be in each category? Show these proportions on the normal distribution graph.

b. Find the 60_{th} percentile of the distribution of the weight. In other words find c such that $P(X < c) = 0.60$.

c. Suppose that 5 chickens are selected at random. What is the probability that 3 of them will be petite?

Problem 2

A television cable company receives numerous phone calls throughout the day from customers reporting service troubles and from would-be subscribers to the cable network. Most of these callers are put "on hold" until a company operator is free to help them. The company has determined that the length of time a caller is on hold is normally distributed with a mean of 3.1 minutes and a standard deviation 0.9 minutes. Company experts have decided that if as many as 5% of the callers are put on hold for 4.8 minutes or longer, more operators should be hired.

a. What proportion of the company's callers are put on hold for more than 4.8 minutes? Should the company hire more operators? Show these probabilities on a sketch of the normal curve.

b. At another cable company (length of time a caller is on hold follows the same distribution as before), 2.5% of the callers are put on hold for longer than x minutes. Find the value of x, and show this on a sketch of the normal curve.

Problem 3

Answer the following questions:

a. Suppose that the height (X) in inches, of a 25-year-old man is a normal random variable with mean $\mu = 70$ inches. If $P(X > 79) = 0.025$ what is the standard deviation of this random normal variable?

b. Suppose that the weight (X) in pounds, of a 40-year-old man is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population weigh less than 160 pounds what is the mean μ of this distribution?

c. Find an interval that covers the middle 95% of $X \sim N(64, 8)$.

Problem 4
A bag of cookies is underweight if it weighs less than 500 grams. The filling process dispenses cookies with weight that follows the normal distribution with mean 510 grams and standard deviation 4 grams.

a. What is the probability that a randomly selected bag is underweight?

b. If you randomly select 5 bags, what is the probability that exactly 2 of them will be underweight?

Problem 5
Answer the following questions:

a. Suppose that X follows the normal distribution with mean $\mu = 5$. If $P(X > 9) = 0.2$ find the variance of X.

b. Let X be a normal random variable with mean $\mu = 12$ and standard deviation $\sigma = 2$. Find the $10th$ percentile of this distribution.

c. The weight X of water melons is normally distributed with mean $\mu = 10$ pounds and standard deviation $\sigma = 2$ pounds. Find c such that $P(X > c) = 0.60$.

d. The montly return of a particular stock follows the normal distribution with mean 0.02 and standard deviation 0.1. Find the 85_{th} percentile of this distribution.

e. Find the probability that the monthly return of the stock in question (b) will be larger that 0.2.

f. Find the probability that in one year (12 months), the return of the stock in question (e) will be larger than 0.2 on exactly 4 months. Assume that the returns are independent from month to month.

g. The annual rainfall X (in inches) at a certain region is normally distributed with mean $\mu = 40$ pounds and standard deviation $\sigma = 4$. What is the probability that starting with this year, it will take more than 10 years before a year occurs having a rainfall of over 50 inches?

h. Let $X \sim N(100, 20)$. Find $P(X > 70|X < 90)$.

Problem 6
The diameters of apples from the Milo Farm follow the normal distribution with mean 3 inches and standard deviation 0.3 inch. Apples can be size-sorted by being made to roll over a mesh screens. First the apples are rolled over a screen with mesh size 2.5 inches. This separates out all the apples with diameters < 2.5 inches. Second, the remaining apples are rolled over a screen with mash size 3.2 inches. Find the proportion of apples with diameters < 2.5 inches, between 2.5 and 3.2 inches, and greater than 3.2 inches. Use only *SOCR* to find the answers and print the appropriate snapshots.

Normal distribution - Practice problems
Solutions

Problem 1

The chickens of the Ortnithes farm are processed when they are 20 weeks old. The distribution of their weights is normal with mean 3.8 lb, and standard deviation 0.6 lb. The farm has created three categories for these chickens according to their weight: petite (weight less than 3.5 lb), standard (weight between 3.5 lb and 4.9 lb), and big (weight above 4.9 lb).

a. What proportion of these chickens will be in each category? Show these proportions on the normal distribution graph.
Answer:
Petite: $P(X < 3.5) = P(Z < \frac{3.5-3.8}{0.6}) = P(Z < -0.50) = 0.3085$.
Standard: $P(3.5 < X < 4.9) = P(\frac{3.5-3.8}{0.6} < Z < \frac{4.9-3.8}{0.6}) = P(-0.5 < Z < 1.83) = 0.9664 - 0.3085 = 0.6579$.
Big: $P(X > 4.9) = P(Z > \frac{4.9-3.8}{0.6}) = P(Z > 1.83) = 1 - 0.9664 = 0.0336$.

b. Find the 60_{th} percentile of the distribution of the weight. In other words find c such that $P(X < c) = 0.60$.
Answer:
From the z table approximately $z = 0.2055$. Therefore, $0.2055 = \frac{x-3.8}{0.6} \Rightarrow x = 3.8 + 0.2055(0.6) = 3.92$.

c. Suppose that 5 chickens are selected at random. What is the probability that 3 of them will be petite?
Answer:
This is binomial with $n = 5, p = 0.3085$. Therefore, $P(Y = 3) = \binom{5}{3}0.3085^3(1 - 0.3085)^2 = 0.1404$.

Problem 2

A television cable company receives numerous phone calls throughout the day from customers reporting service troubles and from would-be subscribers to the cable network. Most of these callers are put "on hold" until a company operator is free to help them. The company has determined that the length of time a caller is on hold is normally distributed with a mean of 3.1 minutes and a standard deviation 0.9 minutes. Company experts have decided that if as many as 5% of the callers are put on hold for 4.8 minutes or longer, more operators should be hired.

a. What proportion of the company's callers are put on hold for more than 4.8 minutes? Should the company hire more operators? Show these probabilities on a sketch of the normal curve.
Answer:
$P(X > 4.8) = P(Z > \frac{4.8-3.1}{0.9}) = P(Z > 1.89) = 1 - 0.9706 = 0.0294$.

b. At another cable company (length of time a caller is on hold follows the same distribution as before), 2.5% of the callers are put on hold for longer than x minutes. Find the value of x, and show this on a sketch of the normal curve.
Answer:
From the z table we find that $z = 1.96$. Therefore, $1.96 = \frac{x-3.1}{0.9} \Rightarrow x = 3.1 + 1.96(0.9) = 4.86$.

Problem 3

Answer the following questions:

a. Suppose that the height (X) in inches, of a 25-year-old man is a normal random variable with mean $\mu = 70$ inches. If $P(X > 79) = 0.025$ what is the standard deviation of this random normal variable?

Answer:

$1.96 = \frac{79-70}{\sigma} \Rightarrow \sigma = \frac{9}{1.96} = 4.59.$

b. Suppose that the weight (X) in pounds, of a 40-year-old man is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population weigh less than 160 pounds what is the mean μ of this distribution?

Answer:

$-1.645 = \frac{160-\mu}{20} \Rightarrow \mu = 160 + 20(1.645) = 192.9.$

c. Find an interval that covers the middle 95% of $X \sim N(64, 8)$.

Answer:

We have 2.5% probability at each one of the two tails. Therefore

$-1.96 = \frac{x-64}{8} \Rightarrow x = 64 - 1.96(8) = 48.32.$

$1.96 = \frac{x-64}{8} \Rightarrow x = 64 + 1.96(8) = 79.68.$

Problem 4

A bag of cookies is underweight if it weighs less than 500 grams. The filling process dispenses cookies with weight that follows the normal distribution with mean 510 grams and standard deviation 4 grams.

a. What is the probability that a randomly selected bag is underweight?

Answer:

$P(X < 500) = P(Z < \frac{500-510}{4}) = P(Z < -2.5) = 0.0062.$

b. If you randomly select 5 bags, what is the probability that exactly 2 of them will be underweight?

Answer:

$P(Y = 2) = \binom{5}{2}0.0062^2(1 - 0.0062)^3 = 0.0004.$