STA 218: Statistics for Management

Al Nosedal.
University of Toronto.

Fall 2018

My momma always said: "Life was like a box of chocolates. You never know what you're gonna get."

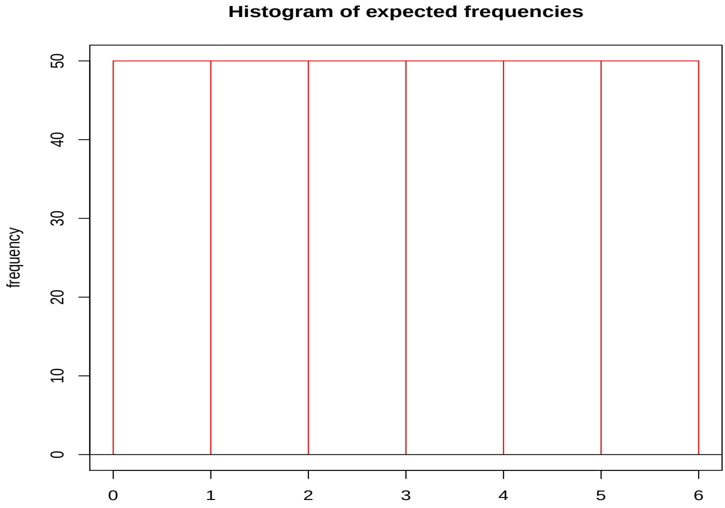
Forrest Gump.

Simple Example

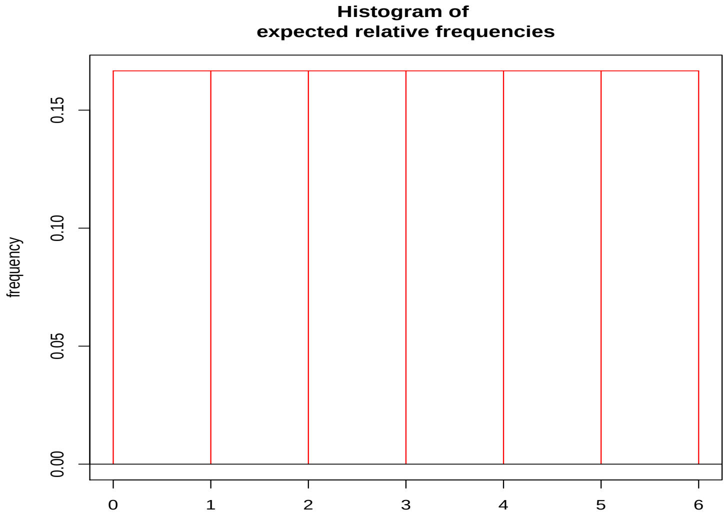
Random Experiment: Rolling a fair die 300 times.

Class	Expected Frequency	Expected Relative Freq
$0 < x \leq 1$	50	$1/6$
$1 < x \leq 2$	50	$1/6$
$2 < x \leq 3$	50	$1/6$
$3 < x \leq 4$	50	$1/6$
$4 < x \leq 5$	50	$1/6$
$5 < x \leq 6$	50	$1/6$

Histogram of Expected Frequencies



Histogram of Expected Relative Frequencies



Density Curve

A density curve is a curve that is always on or above the horizontal axis, and has area exactly 1 underneath it.
A density curve describes the overall pattern of a distribution. The area under the curve and above any range of values is the proportion of all observations that fall in that range.
Note. No set of real data is exactly described by a density curve. The curve is an idealized description that is easy to use and accurate enough for practical use.

Example

The amount of gasoline sold daily at a service station is uniformly distributed with a minimum of 2,000 gallons and a maximum of 5,000 gallons.

a. Find the probability that daily sales will fall between 2,500 and 3,000 gallons.

b. What is the probability that the service station will sell at least 4,000 gallons?

c. What is the probability that the station will sell exactly 2,500 gallons?

Solution

a. $P(2500 \leq X \leq 3000) = (3000 - 2500)\left(\frac{1}{3000}\right) = 0.1667.$
b. $P(X \geq 4000) = (5000 - 4000)\left(\frac{1}{3000}\right) = 0.3333.$
c. $P(X = 2500) = 0.$

Normal Density Function

The probability density function of a **Normal random variable** is

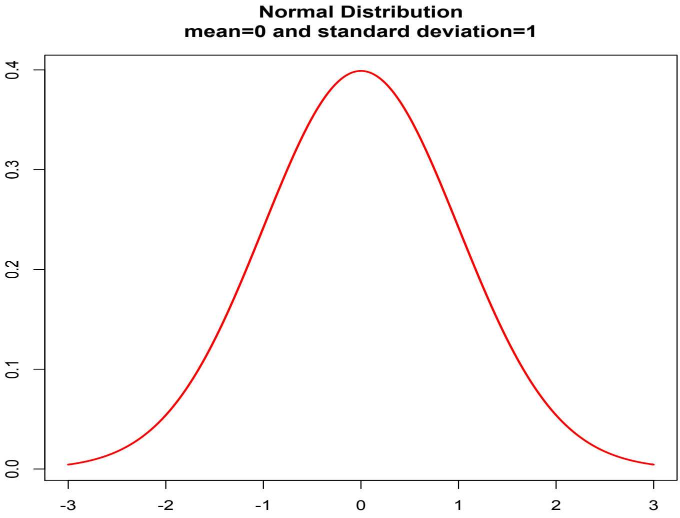
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where $-\infty < x < \infty$, $e = 2.71828...$ and $\pi = 3.14159...$

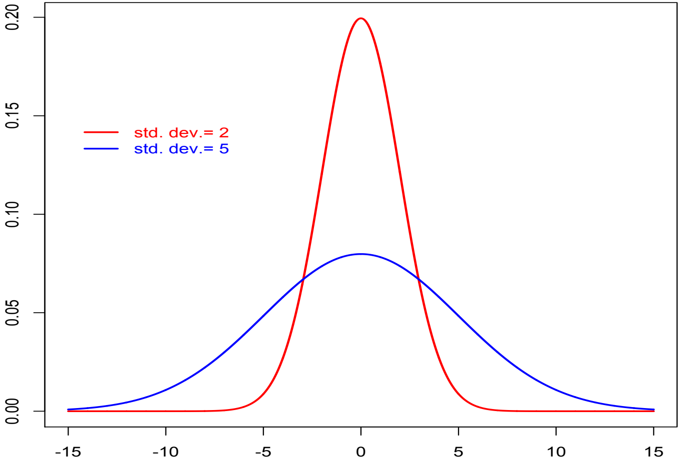
Normal Distributions

A Normal Distribution is described by a Normal density curve. Any particular Normal distribution is completely specified by two numbers, its mean μ and standard deviation σ.
The mean of a Normal distribution is at the center of the symmetric Normal curve. The standard deviation is the distance from the center to the change-of-curvature points on either side.

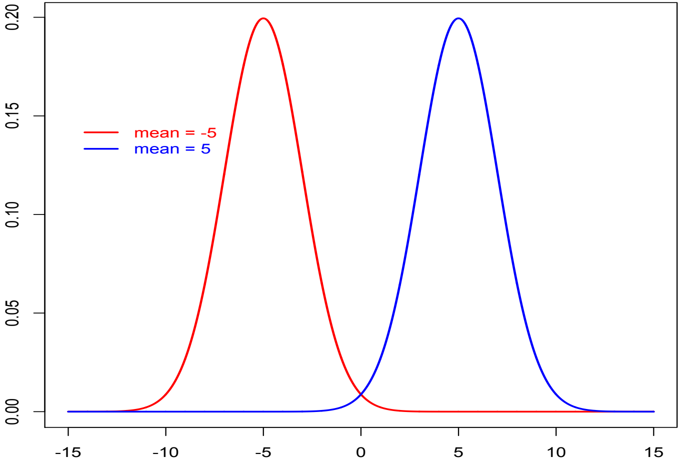
Standard Normal Distribution



Two Different Standard Deviations



Two Different Means



The 68-95-99.7 rule

In a Normal distribution with mean μ and standard deviation σ:
Approximately 68% of the observations fall within σ of the mean μ.
Approximately 95% of the observations fall within 2σ of μ.
Approximately 99.7% of the observations fall within 3σ of μ.

Standard Normal Distribution

The standard Normal distribution is the Normal distribution $N(0, 1)$ with mean 0 and standard deviation 1.
If a variable x has any Normal distribution N(μ,σ) with mean μ and standard deviation σ, then the standardized variable

$$z = \frac{x - \mu}{\sigma}$$

has the standard Normal distribution.

SAT vs ACT

In 2010, when she was a high school senior, Alysha scored 670 on the Mathematics part of the SAT. The distribution of SAT Math scores in 2010 was Normal with mean 516 and standard deviation 116. John took the ACT and scored 26 on the Mathematics portion. ACT Math scores for 2010 were Normally distributed with mean 21.0 and standard deviation 5.3. Find the standardized scores for both students. Assuming that both tests measure the same kind of ability, who had the higher score?

Solution

Alysha's standardized score is

$$z_A = \frac{670 - 516}{116} = 1.33.$$

John's standardized score is

$$z_J = \frac{26 - 21}{5.3} = 0.94.$$

Alysha's score is relatively higher than John's.

Men's and women's heights

The heights of women aged 20 to 29 are approximately Normal with mean 64.3 inches and standard deviation 2.7 inches. Men the same age have mean height 69.9 inches with standard deviation 3.1 inches. What are the z-scores for a woman 6 feet tall and a man 6 feet tall? Say in simple language what information the z-scores give that the original nonstandardized heights do not.

Solution

We need to use the same scale, so recall that 6 feet = 72 inches.
A woman 6 feet tall has standardized score

$$z_W = \frac{72 - 64.3}{2.7} = 2.85$$

(quite tall, relatively).
A man 6 feet tall has standardized score

$$z_M = \frac{72 - 69.9}{3.1} = 0.68.$$

Hence, a woman 6 feet tall is 2.85 standard deviations taller than average for women. A man 6 feet tall is only 0.68 standard deviations above average for men.

Using the Normal table

Use table 3 to find the proportion of observations from a standard Normal distribution that satisfies each of the following statements. In each case, sketch a standard Normal curve and shade the area under the curve that is the answer to the question.

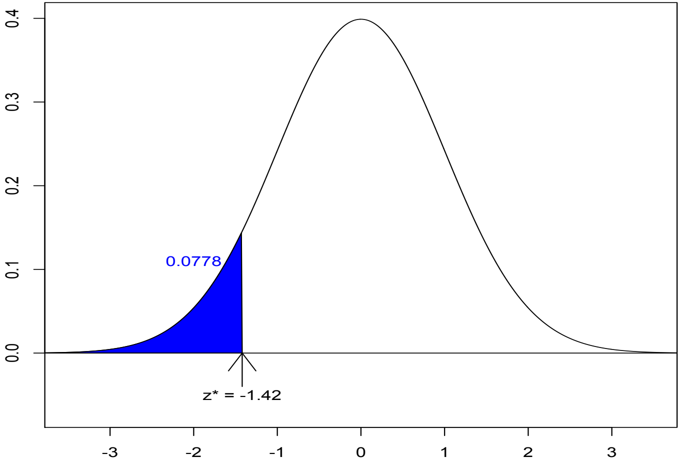
a) $z < -1.42$

b) $z > -1.42$

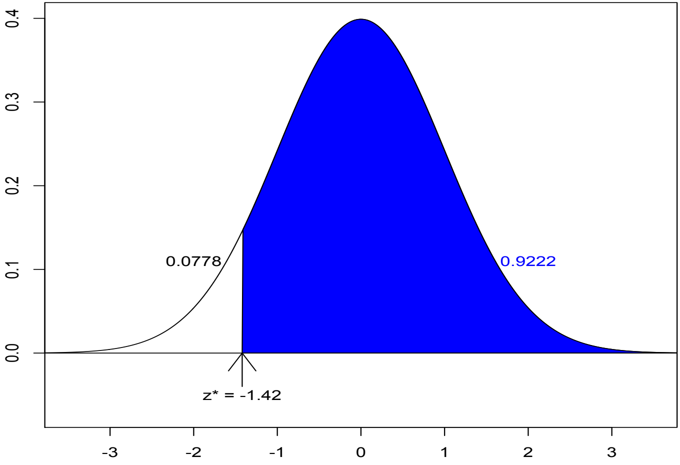
c) $z < 2.35$

d) $-1.42 < z < 2.35$

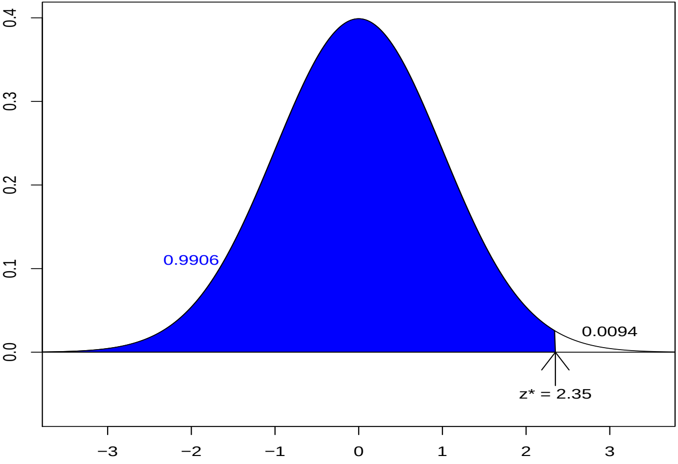
Solution a) 0.0778



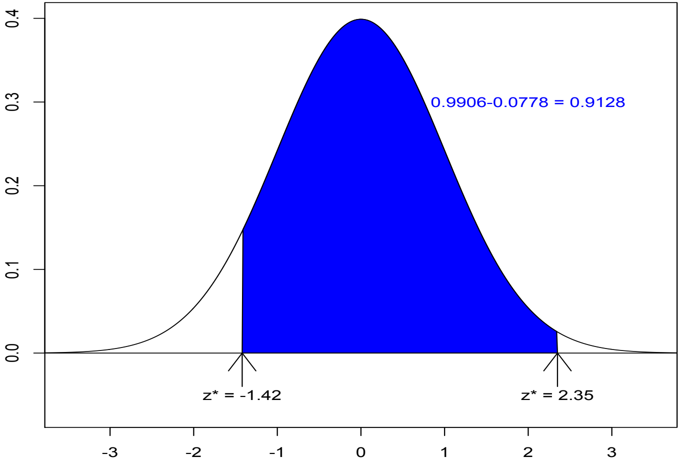
Solution b) 0.9222



Solution c) 0.9906



Solution d) 0.9966 - 0.0778 = 0.9128



Example

The annual rate of return on stock indexes (which combine many individual stocks) is approximately Normal. Since 1945, the Standard & Poor's 500 stock index has had a mean yearly return of about 12%, with a standard deviation of 16.5%. Take this Normal distribution to be the distribution of yearly returns over a long period.

a) The market is down for the year if the return on the index is less than zero. In what proportion of years is the market down?

b) What percent of years have annual rates of return between 12% and 50%?

Solution a)

1. State the problem. Call the annual rate of return for Standard & Poor's 500 stock index x. The variable x has the $N(12, 16.5)$ distribution. We want the proportion of years with $x < 0$.
2. Standardize. Subtract the mean, then divide by the standard deviation, to turn x into a standard Normal z.
Hence $x < 0$ corresponds to $z < \frac{0-12}{16.5} = -0.73$.
3. Use the table. From Table A, we see that the proportion of observations less than -0.73 is 0.2327. The market is down on an annual basis about 23.27% of the time.

Solution b)

1. State the problem. Call the annual rate of return for Standard & Poor's 500 stock index x. The variable x has the $N(12, 16.5)$ distribution. We want the proportion of students with
$12 \leq x \leq 50$.
2. Standardize. Subtract the mean, then divide by the standard deviation, to turn x into a standard Normal z.
$12 \leq x \leq 50$ corresponds to $\frac{12-12}{16.5} \leq z \leq \frac{50-12}{16.5}$, or $0 \leq z \leq 2.30$.
3. Use the table. Using Table A, the area is
$0.9893 - 0.5000 = 0.4893$, or 48.93%.
About 49% of years have annual rates of return between 12% and 50%.

The Medical College Admission Test

Almost all medical schools in the United States require students to take the Medical College Admission Test (MCAT). The exam is composed of three multiple-choice sections (Physical Sciences, Verbal Reasoning, and Biological Sciences). The score on each section is converted to a 15-point scale so that the total score has a maximum value of 45. The total scores follow a Normal distribution, and in 2010 the mean was 25.0 with a standard deviation of 6.4. There is little change in the distribution of scores from year to year.

a) What proportion of students taking the MCAT had a score over 30?

b) What proportion had scores between 20 and 25?

Solution a)

1. State the problem. Let x be the MCAT score of a randomly selected student. The variable x has the $N(25, 6.4)$ distribution. We want the proportion of students with $x > 30$.
2. Standardize. Subtract the mean, then divide by the standard deviation, to turn x into a standard Normal z.
Hence $x > 30$ corresponds to $z > \frac{30-25}{6.4} = 0.78$.
3. Use the table. From Table A, we see that the proportion of observations less than 0.78 is 0.7823. Hence, the answer is $1 - 0.7823 = 0.2177$, or 21.77%.

Solution b)

1. State the problem. Let x be the MCAT score of a randomly selected student. The variable x has the $N(25, 6.4)$ distribution. We want the proportion of students with $20 \leq x \leq 25$.
2. Standardize. Subtract the mean, then divide by the standard deviation, to turn x into a standard Normal z.
$20 \leq x \leq 25$ corresponds to $\frac{20-25}{6.4} \leq z \leq \frac{25-25}{6.4}$, or $-0.78 \leq z \leq 0$.
3. Use the table. Using Table A, the area is
$0.5 - 0.2177 = 0.2833$, or 28.33%.

Using a table to find Normal proportions

Step 1. State the problem in terms of the observed variable x. Draw a picture that shows the proportion you want in terms of cumulative proportions.

Step 2. Standardize x to restate the problem in terms of a standard Normal variable z.

Step 3. Use Table A and the fact that the total are under the curve is 1 to find the required area under the standard Normal curve.

Example

Consider an investment whose return is Normally distributed with a mean of 10% and a standard deviation of 5%.
a. Determine the probability of losing money.
b. Find the probability of losing money when the standard deviation is equal to 10%.

Solution a)

$X =$ return. We know that X is Normally distributed with mean (μ) 0.10 and standard deviation (σ) 0.05. Another way of writing this is: X has a $N(0.10, 0.05)$ distribution.

a. $P(X < 0) = P\left(\frac{X-\mu}{\sigma} < \frac{0-0.1}{0.05}\right) = P(Z < -2) = 0.0228.$

Therefore, the probability of losing money is 2.28%.

(Using the Empirical Rule gives 2.5%)

Solution b)

(If we increase the standard deviation to 10%).
$Y =$ return. We know that Y is Normally distributed with mean (μ) 0.10 and standard deviation (σ) 0.10. Another way of writing this is: Y has a $N(0.10, 0.10)$ distribution.
a. $P(Y < 0) = P\left(\frac{Y-\mu}{\sigma} < \frac{0-0.1}{0.1}\right) = P(Z < -1) = 0.1587$.
Therefore, the probability of losing money is 15.87%.

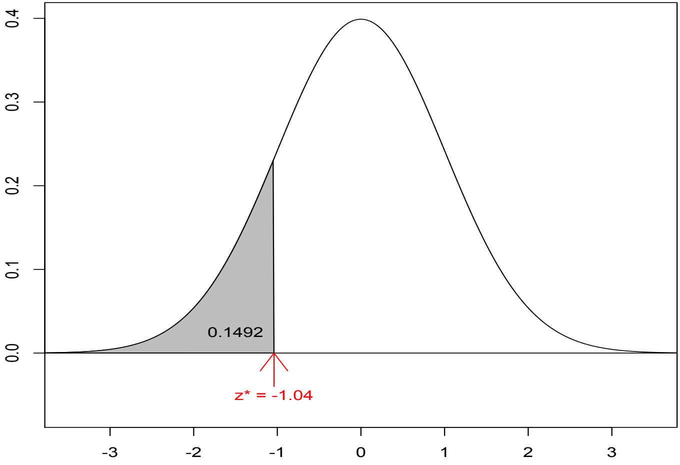
Table 3 (also known as Table A)

Use Table 3 to find the value $z*$ of a standard Normal variable that satisfies each of the following conditions. (Use the value of $z*$ from Table 3 that comes closest to satisfying the condition.) In each case, sketch a standard Normal curve with your value of $z*$ marked on the axis.

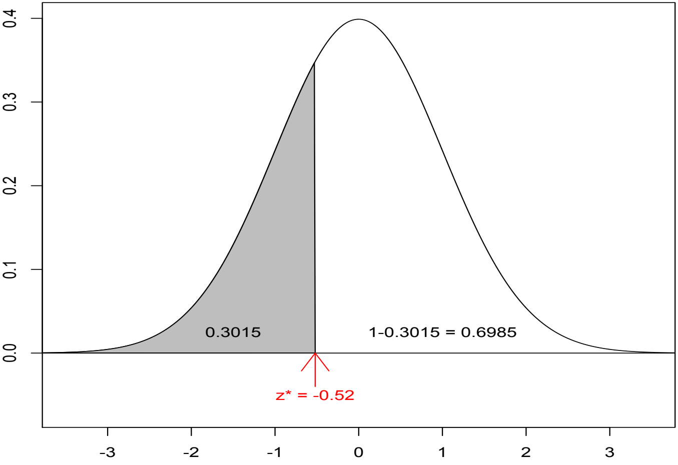
a) The point $z*$ with 15% of the observations falling below it.

b) The point $z*$ with with 70% of the observations falling above it.

Solution a) z* = -1.04



Solution b) z* = -0.52



The Medical College Admission Test

The total scores on the Medical College Admission Test (MCAT) follow a Normal distribution with mean 25.0 and standard deviation 6.4. What are the median and the first and third quartiles of the MCAT scores?

Solution: Finding the median

Because the Normal distribution is symmetric, its median and mean are the same. Hence, the median MCAT score is 25.

Solution: Finding Q_1

1. State the problem. We want to find the MCAT score x with area 0.25 to its left under the Normal curve with mean $\mu = 25$ and standard deviation $\sigma = 6.4$.
2. Use the table. Look in the body of Table A for the entry closest to 0.25. It is 0.2514. This is the entry corresponding to $z* = -0.67$. So $z* = -0.67$ is the standardized value with area 0.25 to its left.
3. Unstandardize to transform the solution from the $z*$ back to the original x scale. We know that the standardized value of the unknown x is $z* = -0.67$.
So x itself satisfies

$$\frac{x - 25}{6.4} = -0.67$$

Solving this equation for x gives
$x = 25 + (-0.67)(6.4) = 20.71$

Solution: Finding Q_3

1. State the problem. We want to find the MCAT score x with area 0.75 to its left under the Normal curve with mean $\mu = 25$ and standard deviation $\sigma = 6.4$.
2. Use the table. Look in the body of Table A for the entry closest to 0.75. It is 0.7486. This is the entry corresponding to $z* = 0.67$. So $z* = 0.67$ is the standardized value with area 0.75 to its left.
3. Unstandardize to transform the solution from the $z*$ back to the original x scale. We know that the standardized value of the unknown x is $z* = 0.67$.
So x itself satisfies

$$\frac{x - 25}{6.4} = 0.67$$

Solving this equation for x gives
$x = 25 + (0.67)(6.4) = 29.29$

Exercise 8.70

The demand for a daily newspaper at a newsstand at a busy intersection is known to be Normally distributed with a mean of 150 and a standard deviation of 25. How many newspapers should the newsstand operator order to ensure that he runs short on no more than 20% of days?

Solution

D = demand. We know that D has a Normal distribution with mean 150 and standard deviation 25.

1. State the problem. We want to find a number of newspapers, n^*, with area 0.80 to its left under the Normal curve with mean $\mu = 150$ and standard deviation $\sigma = 25$.

2. Use the table. Look in the body of Table A for the entry closest to 0.80. It is .7995. This is the entry corresponding to $z* = 0.84$. So $z* = 0.84$ is the standardized value with area 0.80 to its left.

Solution

3. Unstandardize to transform the solution from the $z*$ back to the original D scale. We know that the standardized value of the unknown n^* is $z* = z* = 0.84$.
So n^* itself satisfies

$$\frac{n^* - 150}{25} = 0.84$$

Solving this equation for x gives
$n^* = 150 + (0.84)(25) = 171$
The newsstand operator should order 171 newspapers.

Example

During the spring, the demand for electric fans at a large home-improvement store is quite strong. The company tracks inventory using a computer system so that it knows how many fans are in the inventory at any time. The policy is to order a new shipment of 250 fans when the inventory level falls to the reorder point, which is 150.

Example

However, this policy has resulted in frequent shortages and thus lost sales because both lead time and demand are highly variable. The manager would like to reduce the incidence of shortages so that only 5% of orders will arrive after inventory drops to 0 (resulting in a shortage). This policy is expressed as a 95% service level. From previous periods, the company has determined that demand during lead time is Normally distributed with a mean of 200 and a standard deviation of 50. Find the reorder point.

Solution

D = demand during lead time. We know that D has a Normal distribution with mean 200 and standard deviation 50.
1. State the problem. We want to find the reorder point (ROP), ROP with area 0.95 to its left under the Normal curve with mean $\mu = 200$ and standard deviation $\sigma = 50$.
2. Use the table. Look in the body of Table A for the entry closest to 0.95. It is 0.9505. This is the entry corresponding to $z* = 1.65$. So $z* = 1.65$ is the standardized value with area 0.95 to its left.

Solution

3. Unstandardize to transform the solution from the $z*$ back to the original ROP scale. We know that the standardized value of the unknown ROP is $z* = 1.65$.
So ROP itself satisfies

$$\frac{ROP - 200}{50} = 1.65$$

Solving this equation for x gives
$ROP = 200 + (1.65)(50) = 282.5$
which we round up to 283. The policy is to order a new batch of fans when there are 283 fans left in inventory.

Finding a value when given a proportion

1. State the problem.
2. Use the table.
3. Unstandardize to transform the solution from the $z*$ back to the original x scale.

Exercise 8.66

Mensa is an organization whose members possess IQs that are in the top 2% of the population. It is known that IQs are Normally distributed with a mean of 100 and a standard deviation of 16. Find the minimum IQ needed to be a Mensa member.

Solution

$IQ^* = \mu + z^*\sigma = 100 + 2.05(16) = 132.8$
We could round it to 133.

R Code

```
qnorm(0.98, mean = 100, sd = 16);

## [1] 132.86
```

Example

The summer monsoon rains in India follow approximately a **Normal** distribution with mean 852 mm of rainfall and standard deviation 82 mm. a) In the drought year 1987, 697 mm of rain fell. In what percent of all years will India have 697 mm or less of monsoon rain? b) "Normal rainfall" means within 20% of the long-term average, or between 683 and 1022 mm. In what percent of all years is the rainfall normal?

R code

Just type the following:

```
# a)
pnorm(697, mean = 852, sd = 82);

## [1] 0.02936267

# b)
pnorm(1022, 852, 82) - pnorm(683, 852, 82);

## [1] 0.9612691
```