Project Maths
Tionscadal Mata
Development Team

Module 5

Bernoulli Trials

Bernoulli Trials show up in lots of places. There are **4 essential features:**

1. There must be a fixed number of trials, n
2. The trials must be independent of each other
3. Each trial has exactly 2 outcomes called success or failure
4. The probability of success, *p*, is constant in each trial

Where do we see this occurring?

- tossing a coin
- looking for defective products rolling off an assembly line
- shooting free throws in a basketball game

Whenever we are dealing with a Bernoulli trial there is a discrete random variable *X*. This random variable needs to be identified because all probability questions will involve finding the probability of different values of this variable. For example if you toss a coin *n* times, the random variable X could be the number of heads occurring in 3 tosses e.g. *X* can take on the values 0, 1, 2, 3.

We will look at three different types of Problems:

1. calculating the probability of first success after n repeated Bernoulli trials
2. calculating the probability of *k* successes in *n* repeated Bernoulli trials
3. calculating the probability until the k^{th} success in n trials.

First Success After n Repeated Bernoulli Trials

A basketball player has made 80%, of his foul shots during the season. Assuming the shots are independent, find the probability that in tonight's game he:

(a) misses for the first time on his fifth foul shot

(b) makes his first basket on his fourth foul shot

(c) makes his first basket on one of his first 3 foul shots

Solution

Let X = the number of shots until the first missed shot [p = 0.8, q = 0.2]

Let Y = the number of shots until the first made shot [p = 0.2, q = 0.8]

(a) Four shots made followed by a miss:

$$P(X=4)=(0.8)^4(0.2)=0.08192$$

(b) Three misses, then a shot made:

$$P(Y=3)=(0.2)^3(0.8)=0.0064$$

(c) $$P(Y=0)+P(Y=1)+P(Y=2)=\overset{\text{first basket}}{(0.8)}+\overset{\text{1 miss, first basket}}{(0.2)(0.8)^1}+\overset{\text{2 misses, first basket}}{(0.2)^2(0.8)}=0.992$$

k Successes in n Repeated Bernoulli Trials: Binomial Distribution

Problem

A die is tossed 10 times. What is the probability of getting four sixes ?

Solution

$P(6)=\frac{1}{6}$ S

$P(\text{not } 6)=\frac{5}{6}$ F

S	S	S	S	F	F	F	F	F	F
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{5}{6}$	$\frac{5}{6}$	$\frac{5}{6}$	$\frac{5}{6}$	$\frac{5}{6}$	$\frac{5}{6}$

This is only one arragement.

How many arrangements overall?

$$\frac{10!}{4!\times 6!}=\binom{10}{4}={}^{10}C_4 \text{ ways}$$

$$\Rightarrow P(4 \text{ sixes in } 10 \text{ goes})=\binom{10}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^6$$

Problem

A die is tossed n times. What is the probability of getting r sixes ?

Solution

$$\underbrace{S,S,S,S\ldots}_{\text{r sucesses}} \qquad \underbrace{F,F,F,F,F,F\ldots}_{\text{n−r failures}}$$

This is only one arrangement.

But there are $\binom{n}{r}$ ways success can occur.

$$\Rightarrow P(r \text{ successes})=\binom{n}{r}(S)^r(F)^{n-r}$$

Now replace S with p and F with q.

$$P(r \text{ successes})=\binom{n}{r}(p)^r(q)^{n-r}$$

Example 1

A coin is tossed six times, what is the probability of getting four heads.?

We can apply the Binomial Distribution to this question because:

1. There must be a fixed number of trials, n

2. The trials must be independent of each other

3. Each trial has exactly 2 outcomes called success or failure

4. The probability of success, p, is constant in each trial

The Binomial Distrution

$$P(X=r)=\binom{n}{r}(p)^r(q)^{n-r}$$

$p=$ probability of success

$q=1-p=$ probability of a failure

$n=$ total no. of trials

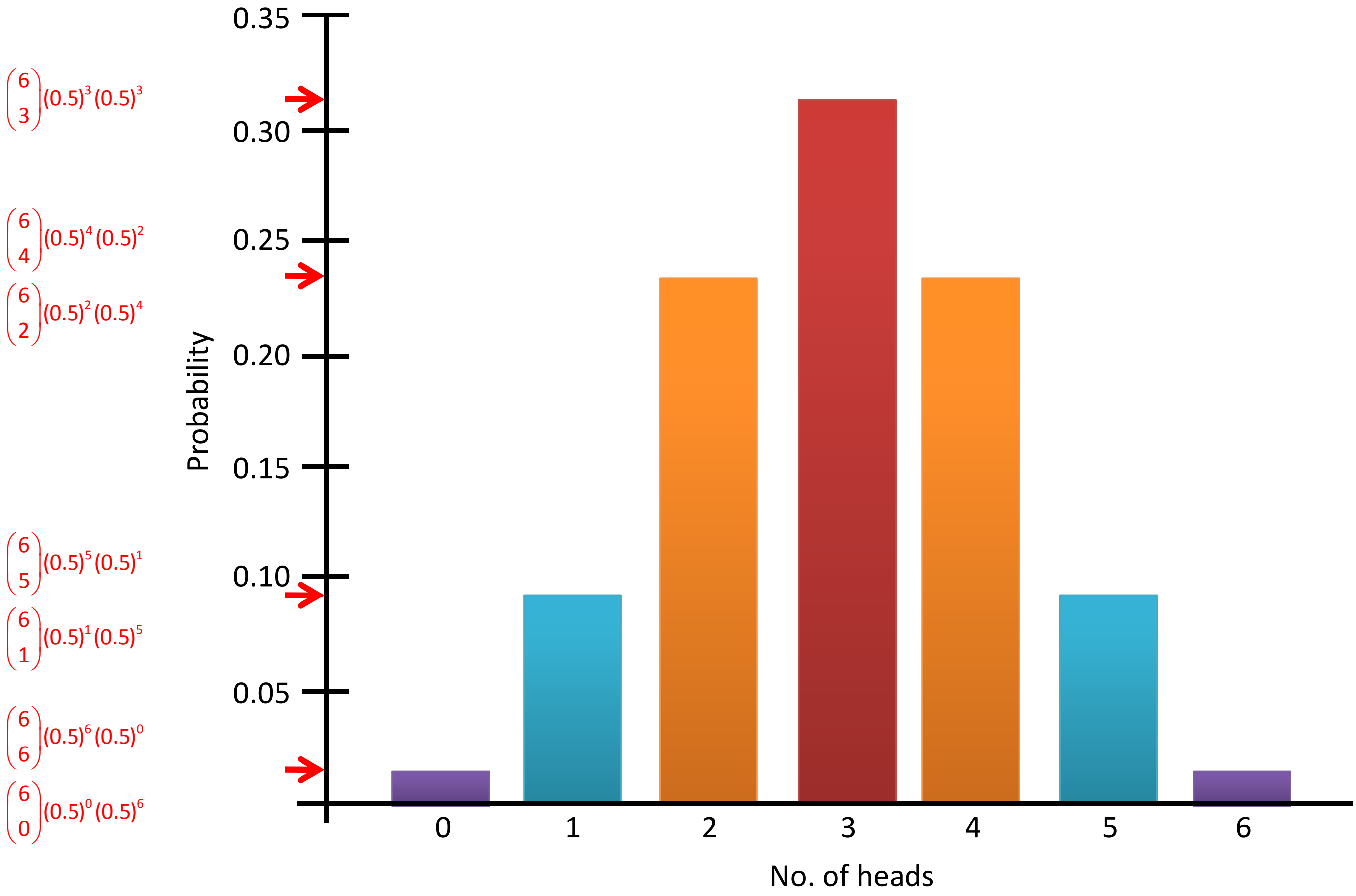
$r=$ number of successes in n trials

Solution

Let $X=$ number of heads $\left[p=\frac{1}{2}, q=\frac{1}{2}\right]$

$$P(X=4)=\binom{6}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^2=\frac{15}{64}=0\cdot 2344$$

Graph of the Distribution



Example 2

In a game of chess against a particular opponent, the probability that Sean wins is $\frac{3}{5}$.

He plays 6 games against his opponent. What is the probability that Sean will:

(i) lose the second game and the 4th game and win the others?

(ii) win exactly four games ?

(iii) lose at least four games?

Solution

(i) The formula does not apply here it is P(w, l, w, l, w, w)

$$P(w, l, w, l, w, w)=\frac{3}{5}\times\frac{2}{5}\times\frac{3}{5}\times\frac{2}{5}\times\frac{3}{5}\times\frac{3}{5}=\frac{324}{15625}$$

In the next two parts a Binomial model is appropriate, where $p=\frac{3}{5}$, $q=\frac{2}{5}$ and $n=6$.

Let X = the number of games won

(ii) $P(X=4)=\binom{6}{4}\left(\frac{3}{5}\right)^4\left(\frac{2}{5}\right)^2=\frac{972}{3125}$

(iii) P(at least 4 losses) = P(no more than 2 wins)

$P(X\le 2)=P(X=0)+P(X=1)+P(X=2)$

$$P(X\le 2)=\binom{6}{0}\left(\frac{3}{5}\right)^0\left(\frac{2}{5}\right)^6+\binom{6}{1}\left(\frac{3}{5}\right)^1\left(\frac{2}{5}\right)^5+\binom{6}{2}\left(\frac{3}{5}\right)^2\left(\frac{2}{5}\right)^4=0.1792$$

Example 3

20% of the items produced by a machine are defective. Four items are chosen at random.

Find the probability that none of the chosen items are defective.

Solution

Let $X =$ number of items that are **not** defective $\quad [p = 0.8 \text{ (not defective)},\ q = 0.2 \text{ (defective)}]$

$$P(X=4) = \binom{4}{4}(0.8)^4(0.2)^0 = \frac{256}{625} = 0.4096$$

Sample Space

0.8	0.8	0.8	0.8	0.4096
0.8	0.8	0.8	0.2	0.1024
0.8	0.8	0.2	0.8	0.1024
0.8	0.2	0.8	0.8	0.1024
0.2	0.8	0.8	0.8	0.1024
0.8	0.8	0.2	0.2	0.0256
0.8	0.2	0.2	0.8	0.0256
0.2	0.2	0.8	0.8	0.0256
0.8	0.2	0.8	0.2	0.0256
0.2	0.8	0.2	0.8	0.0256
0.2	0.8	0.8	0.2	0.0256
0.8	0.2	0.2	0.2	0.0064
0.2	0.2	0.2	0.8	0.0064
0.2	0.8	0.2	0.2	0.0064
0.2	0.2	0.8	0.2	0.0064
0.2	0.2	0.2	0.2	0.0016
Total				1

Example 4

Five unbiased coins are tossed.

(i) Find the probability of getting three heads and two tails.

(ii) The five coins are tossed eight times. Find the probability of getting three heads and two tails exactly four times.

Solution

(i) Let $X =$ number of heads $\quad \left[p = \frac{1}{2},\ q = \frac{1}{2}\right]$

3 heads (and 2 tails) from 5 coins

$$P(X=3) = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5}{16}$$

(ii) The probabilities for this part of the question are got from part **(i)**

Let $X =$ number of times, 3 heads (and 2 tails) occur $\quad \left[p = \frac{5}{16},\ q = \frac{11}{16}\right]$

4 times out of 8 tries

$$P(X=4) = \binom{8}{4}\left(\frac{5}{16}\right)^4\left(\frac{11}{16}\right)^4 = 0\cdot 149$$

Example 5

During a match Owen take a number of penalty shots. The shots are independent of each other and his probability of scoring with each shot is $\frac{4}{5}$.

(i) Find the probability that Owen misses each of his four penalty shots

(ii) Find the probability that Owen scores exactly three of his first four penalty shots

(iii) If Owen takes ten penalty shots during the match, find the probability that he scores at least eight of them

Solution

Let X = number of misses $\left[p=\frac{1}{5},\ q=\frac{4}{5}\right]$

Let Y = number of scores $\left[p=\frac{4}{5},\ q=\frac{1}{5}\right]$

(i) Misses 4 out of 4 shots

$$P(X=4)=\binom{4}{4}\left(\frac{1}{5}\right)^4\left(\frac{4}{5}\right)^0=\frac{1}{625}$$

(ii) Scores 3 out of 4 shots

$$P(Y=3)=\binom{4}{3}\left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)^1=\frac{256}{625}$$

(iii) Scores at least 8 out of 10 shots

$$P(Y\geq 8)=P(Y=8)+P(Y=9)+P(Y=10)$$

$$P(Y\geq 8)=\binom{10}{8}\left(\frac{4}{5}\right)^8\left(\frac{1}{5}\right)^2+\binom{10}{9}\left(\frac{4}{5}\right)^9\left(\frac{1}{5}\right)^1+\binom{10}{10}\left(\frac{4}{5}\right)^{10}\left(\frac{1}{5}\right)^0\approx 0\cdot 678$$

Example 6 (HL)

Ronald is St. Patrick's College best basketball shooter. He is a 70% free throw shooter. Therefore the probability of him scoring on a free throw is 0.7.

What is the probability that Ronald scores his third free throw on his fifth shot?

Solution

His last throw has to be success as we stop when he has 3 free throws after 5 shots.

Let X = number of baskets scored

$$P(X=3)=\overbrace{\binom{4}{2}(0.7)^2(0.3)^2}^{\text{2 baskets out of first 4}}\ \overbrace{(0.7)}^{\text{5th is a basket}}$$

$$P(X=3)=\binom{4}{2}(0.7)^3(0.3)^2=0.18522$$

1st Shot	2nd Shot	3rd Shot	4th Shot	5th Shot	Total
0.7	0.7	0.3	0.3	0.7	0.03087
0.7	0.3	0.7	0.3	0.7	0.03087
0.7	0.3	0.3	0.7	0.7	0.03087
0.3	0.7	0.7	0.3	0.7	0.03087
0.3	0.7	0.3	0.7	0.7	0.03087
0.3	0.3	0.7	0.7	0.7	0.03087
					0.18522

Example 7

What is the probability that Ronald above from St. Patrick's College scores his first free throw on his fifth shot? (This has now become an OL question)

Solution

The only possibility is F F F F S

Let $X =$ number of shots until a score $\quad [p = 0.3,\ q = 0.7]$

$P(X = 4) = (0.3)^4(0.7) = 0.00567$

This is quite low, lower than the last answer because Ronald is quite a sharp shooter and you expect him to have his first score before the 5th shot.

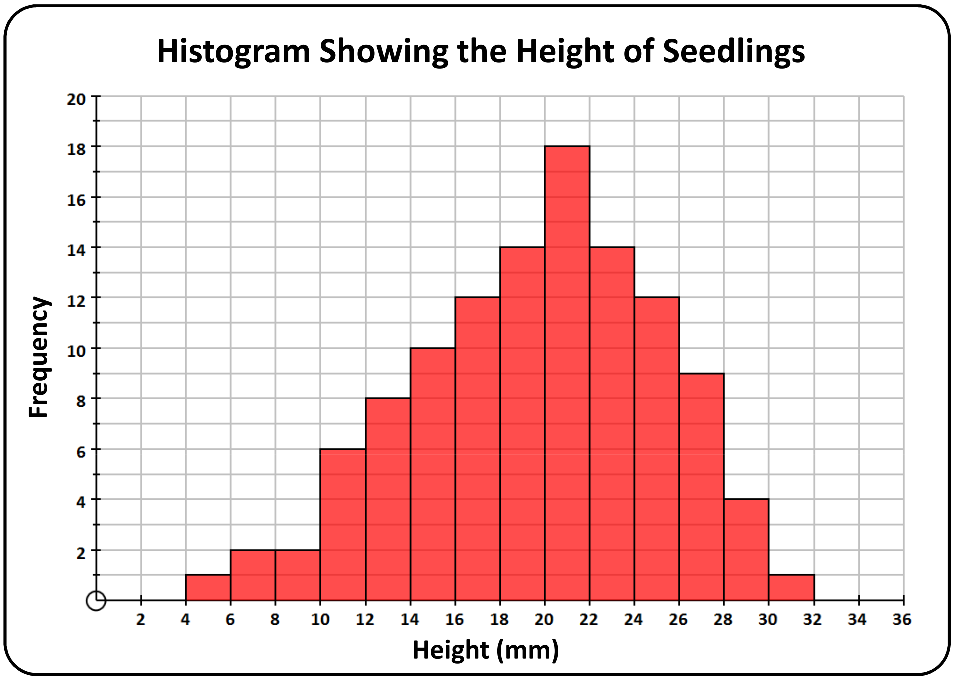
If the probability of him scoring was 20% what would you expect the probability to be that his first free score is on the fifth shot?

Let $X =$ number of shots until a score $\quad [p = 0.8,\ q = 0.2]$

$P(X = 4) = (0.8)^4(0.2) = (0.8)^4(0.2) = 0.08192$

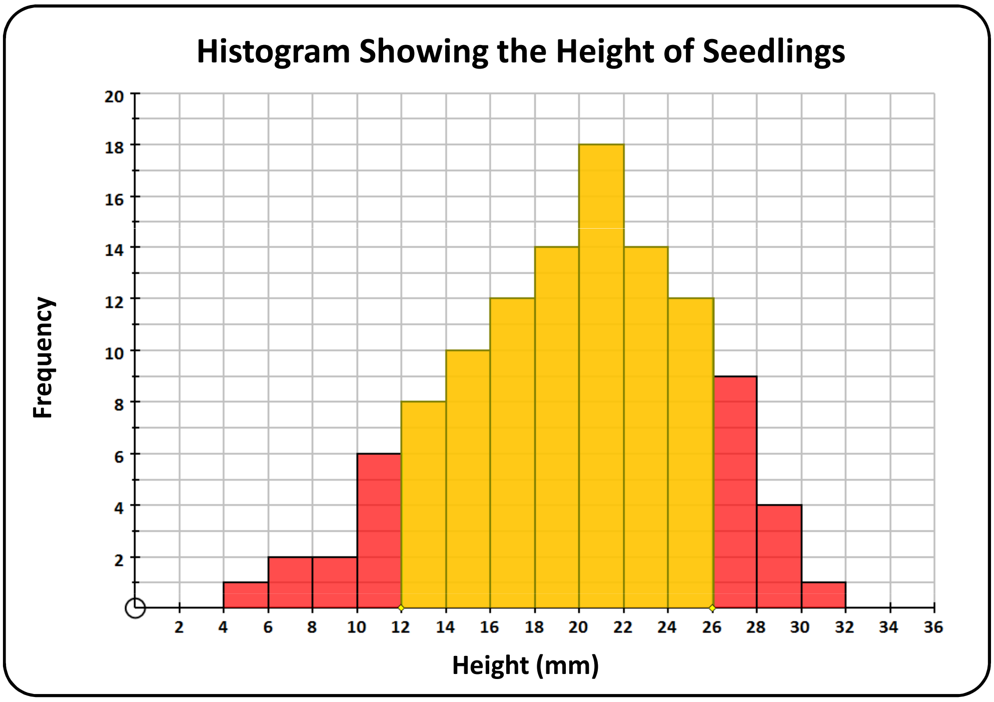
Normal Distribution

Discrete data (golf scores, dice scores) are generally represented by bar charts.
In a bar chart we compare the heights of the bars.
Continuous data (height, weight, physical characteristics) are represented by histograms.
In a histogram we compare the areas of the columns.



The histogram shows that a large quantity of the data is clustered at the centre.

Probability Area



Status Box

Probability by Area: Lower−x = 12, Upper−x = 26, Frequency = 88, Probability = 0.7788

If a seedling is chosen at random it has approximately a 77.88% chance of having a height within the yellow area shown on the histogram i.e. between 12 mm and 26 mm

Sampling Distribution

If another batch of seedlings were taken the picture might look slightly different. It is likely that all batches will follow a common pattern with most of the data clustered around the centre of the histogram. This pattern is common to most measurements in nature. It peaks in the middle and tails at the beginning and end.

To get a perfect model we would need to

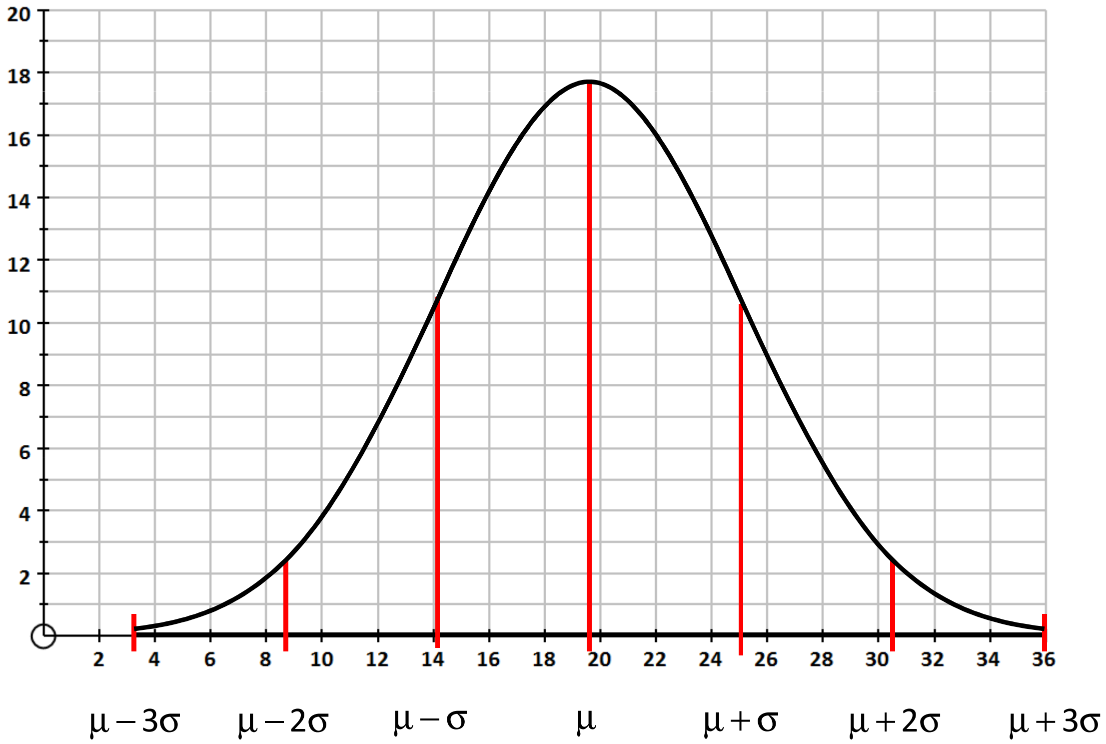
1. Increase the sample size to infinity
2. Take measurements to an infinite number of decimal places
3. Have the widths of the columns approach zero

This is impossible to achieve. We can only create a mathematical model of it. This model is called the **NORMAL DISTRIBUTION**.

What does a Normal Distribution curve look like?

Mean, $\mu = 19.62$ Standard Deviation, $\sigma = 5.46$ of seedlings

If the graph of $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is plotted for the seedling we would get the graph below.



Normal Distribution to Standard Normal Distribution

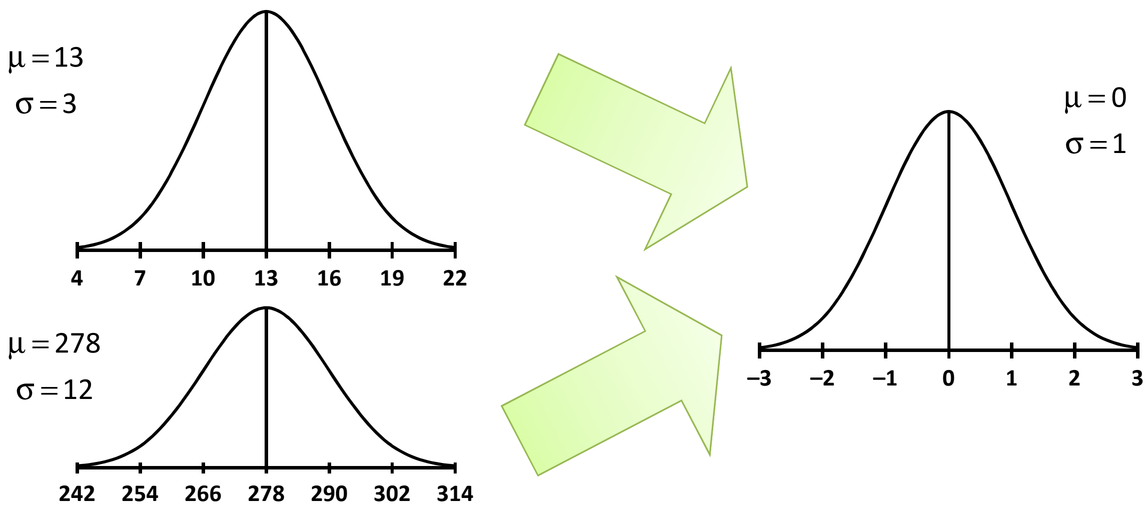
Different sets of data have different means and standard deviations but any that are normally distributed have the same bell-shaped normal distribution type of curves.

Normal Distribution Curve $\rightarrow$ Standard Normal Curve

In order to avoid unnecessary calculations and graphing the scale a Normal Distribution curve is converted to a standard scale called the z score or standard unit scale.

Normal Distributions

Standard Normal Distribution



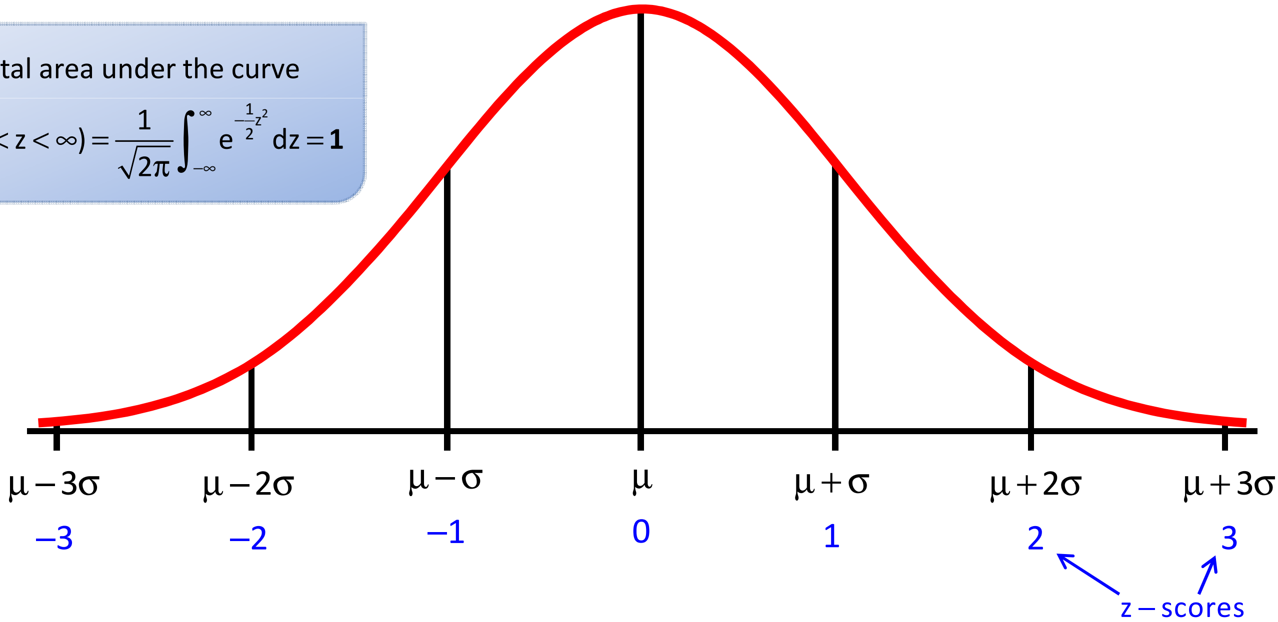
Standard Normal Distribution

If $\mu = 0$ and $\sigma = 1$ we would plot $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$

This graph gives the Standard Normal Graph with a standardised scale.

Total area under the curve

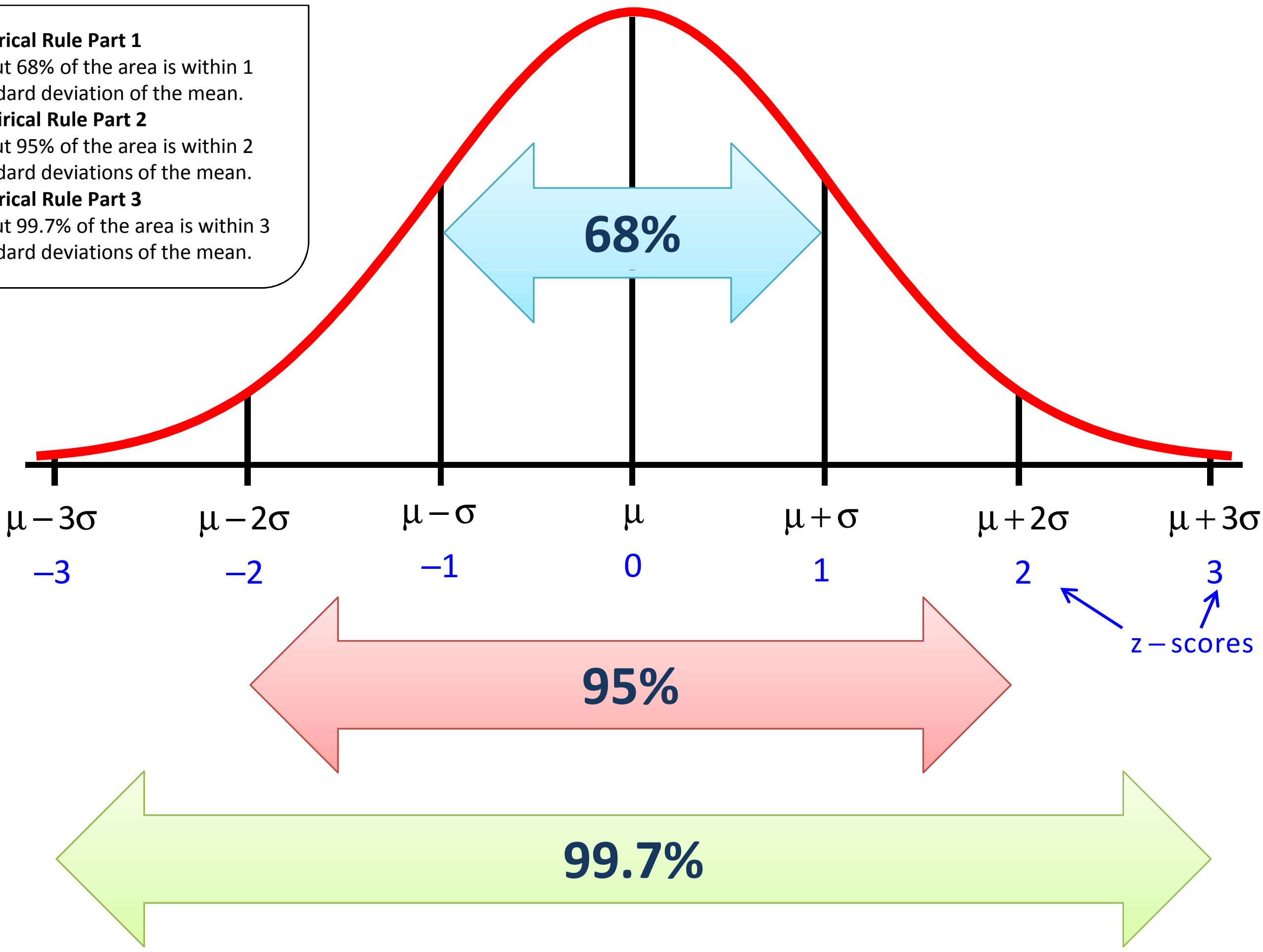
$$P(-\infty < z < \infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\, dz = \mathbf{1}$$



The area between the Standard Normal Curve and the z – axis between $-\infty$ and $+\infty$ is 1.

Empirical Rule

Empirical Rule Part 1
About 68% of the area is within 1 standard deviation of the mean.
Empirical Rule Part 2
About 95% of the area is within 2 standard deviations of the mean.
Empirical Rule Part 3
About 99.7% of the area is within 3 standard deviations of the mean.



Standard Units (z – scores)

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

x is a data point

μ is the population mean

σ is the standard deviation of the population

z – scores define the position of a score in relation to the mean using the standard deviation as a unit of measurement.

z – scores are very useful for comparing data points in different distributions.

The z – score is the number of standard deviations by which the score departs from the mean. This standardises the distribution.

Why do we standardise?

In the 2004 Olympics, Austra Skujte of Lithuania put the shot 16.4 meters, about 3 meters farther than the average of all contestants. Carolina Kluft won the long jump with a 6.78 m jump, about a metre better than the average. Which performance deserves more points for a heptathlon event?

	Long Jump	Shot Put
Mean (all contestants)	6.16 m	13.29 m
SD	0.23 m	1.24 m
n	26	28
Kluft	6.78 m	14.77 m
Skjyte	6.30 m	16.40 m

Both won one event, but Kluft's shot put was second best, while Skujyte's long jump was seventh.

Solution

Standardise the scores, the z – scores can then be added together.

	Long Jump	Shot Put
Kluft	6.78 m	14.77 m
z – score	$\frac{6.78-6.16}{0.23} = 2.70$	$\frac{14.77-13.29}{1.24} = 1.19$
Skjyte	6.30 m	16.40 m
z – score	$\frac{6.30-6.16}{0.23} = 0.61$	$\frac{16.40-13.29}{1.24} = 2.51$

Total z – scores for 2 events:

Kluft : $2.70 + 1.19 = 3.89$

Skjyte : $0.61 + 2.51 = 3.12$

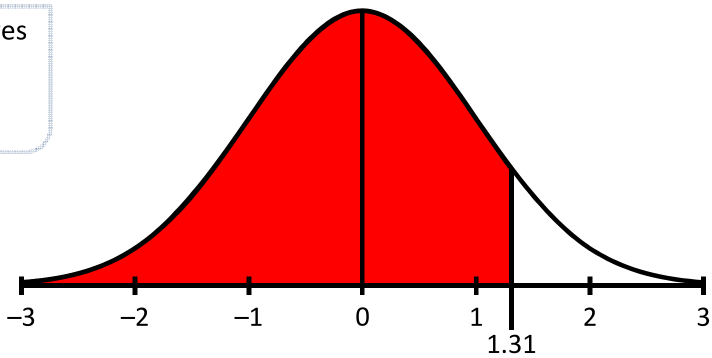
The z – scores measure how far each result is from the event mean in standard deviation units

Example 1

Using the tables find $P(Z \le 1\cdot 31)$.

For a given z, the table gives

$$P(Z \le z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}t}\,dt$$



Pg. 36
Pg. 37

$P(Z \le 1\cdot 31)$ can be read from the tables directly

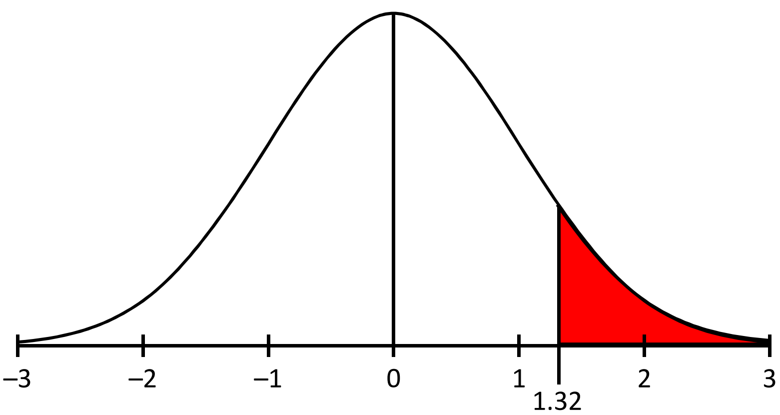
an dáileadh normalach (ar lean) normal distribution (continued)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.1	0.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	0.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	0.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	0.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	0.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	0.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	0.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633

$P(Z \le 1\cdot 31) = 0\cdot 9049 = 90.49\%$

Example 2

Using the tables find $P(Z \ge 1\cdot 32)$



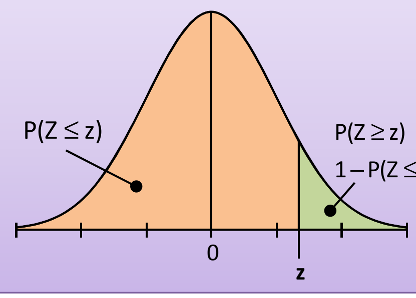
Pg. 36
Pg. 37

$P(Z \ge z)$ is equal to $1 - P(Z \le z)$

$P(Z \ge 1\cdot 32) = 1 - P(Z \le 1\cdot 32)$

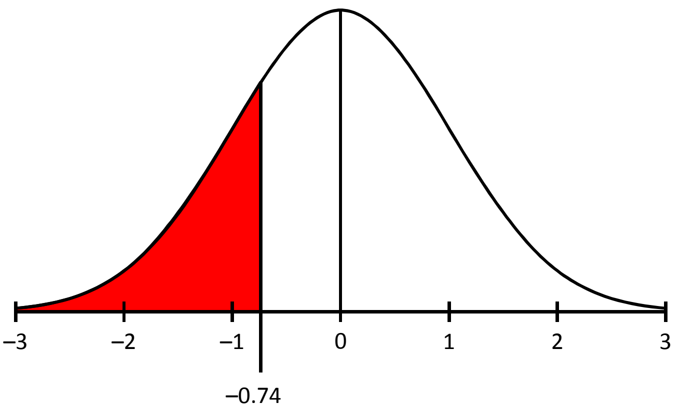
$P(Z \ge 1\cdot 32) = 1 - 0\cdot 9066 = 0\cdot 0934 = 9.34\%$

The table only gives value to the left of z, but the fact that the total area under the curve equals 1, allows us to use, $P(Z \ge z) = 1 - P(Z \le z)$



Example 3

Using the tables find $P(Z \le -0 \cdot 74)$.

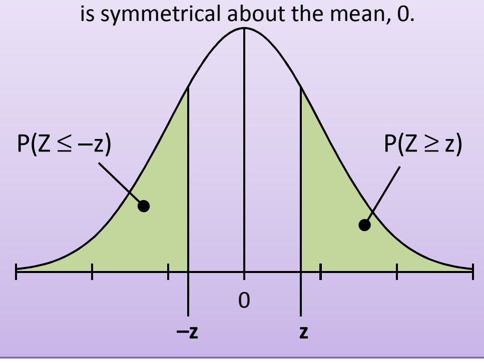


The tables only work for positive values but as the curve is symmetrical about $z = 0$

$P(Z \le -0 \cdot 74) = P(Z \ge 0 \cdot 74)$

$P(Z \le -0 \cdot 74) = 1 - P(Z \le 0 \cdot 74)$

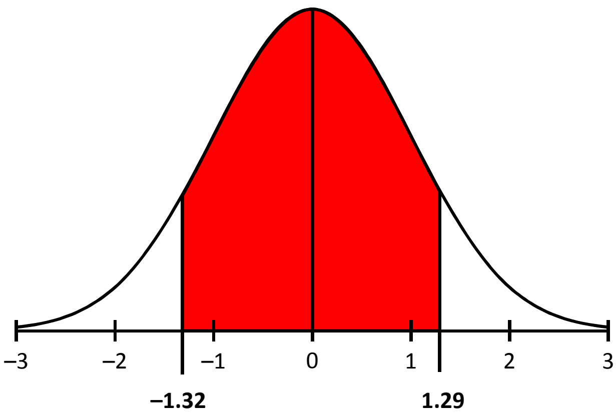
$P(Z \le -0 \cdot 74) = 1 - 0 \cdot 7704 = 0 \cdot 2296 = 22.96\%$

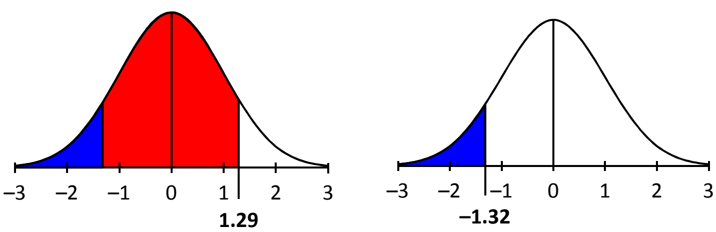


Pg. 36
Pg. 37

Example 4

Using the tables find $P(-1 \cdot 32 \le z \le 1 \cdot 29)$





$P(-1 \cdot 32 \le z \le 1 \cdot 29) =$ Area to the Left of $1 \cdot 29 -$ Area to the left of -1.32

$= P(z \le 1 \cdot 29) - [1 - P(z \le 1 \cdot 32)]$

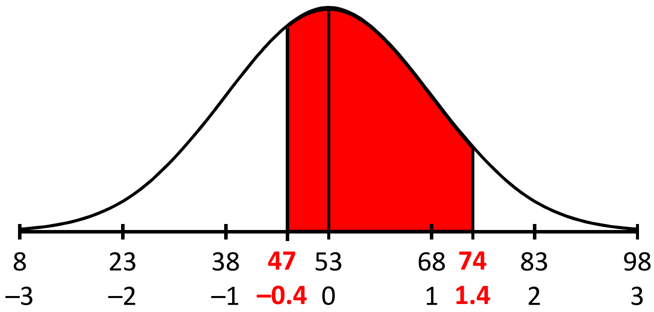
$= 0 \cdot 9015 - [1 - 0 \cdot 9066] = 0 \cdot 8081 = 80.81\%$

Pg. 36
Pg. 37

Example 5

The amounts due on a mobile phone bill in Ireland are normally distributed with a mean of €53 and a standard deviation of €15. If a monthly phone bill is chosen at random, find the probability that the amount due is between €47 and €74.

Solution



Example 6

The mean percentage achieved by a student in a statistic exam is 60%. The standard deviation of the exam marks is 10%.

(i) What is the probability that a randomly selected student scores above 80%?

(ii) What is the probability that a randomly selected student scores below 45%?

(iii) What is the probability that a randomly selected student scores between 50% and 75%?

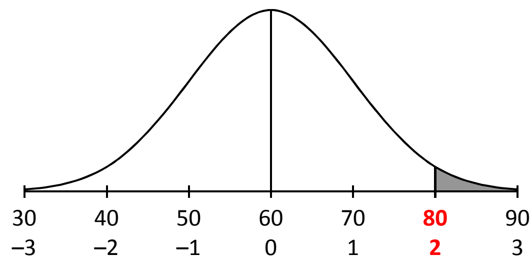
(iv) Suppose you were sitting this exam and you are offered a prize for getting a mark which is greater than 90% of all the other students sitting the exam?
What percentage would you need to get in the exam to win the prize?

Solution

(i) $z = \frac{x-\mu}{\sigma} = \frac{80-60}{10} = 2$

$P(Z>2) = 1 - P(Z<2)$

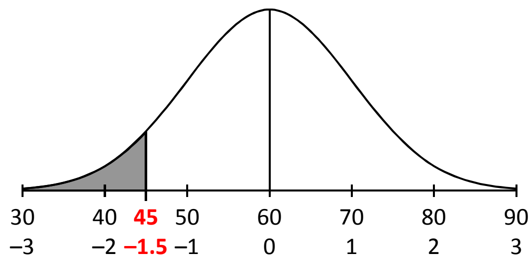
$P(Z>2) = 1 - 0.9772 = 0.0228 = 2.28\%$



(ii) $z = \frac{x-\mu}{\sigma} = \frac{45-60}{10} = -1.5$

$P(Z<-1.5) = P(Z>1.5) \ = 1 - P(Z<1.5)$

$P(Z<-1.5) = 1 - 0.9332 = 0.0668 = 6.68\%$

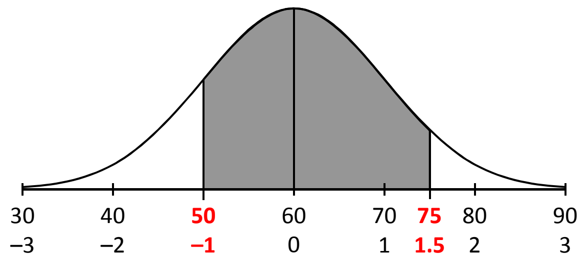


(iii)

$$z_1 = \frac{x-\mu}{\sigma} \qquad z_2 = \frac{x-\mu}{\sigma}$$

$$z_1 = \frac{50-60}{10} \qquad z_2 = \frac{75-60}{10}$$

$$z_1 = -1 \qquad z_2 = 1.5$$



$$P(-1 < Z < 1\cdot 5) = P(Z \le 1\cdot 5) - [1 - P(Z \le 1)]$$

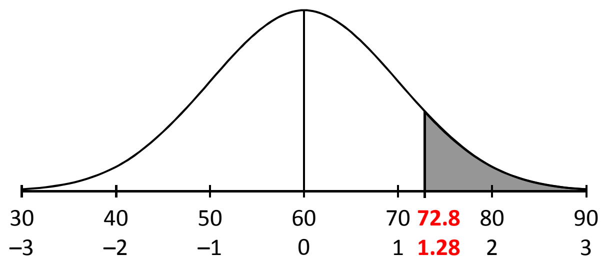
$$P(-1 < Z < 1\cdot 5) = 0.9332 - [1 - 0.8413]$$

$$P(-1 < Z < 1\cdot 5) = 0.7745$$

(iv) From the tables an answer for an area of 90% $(0.9) = 1.28 \Rightarrow Z = 1.28$

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

$$1.28 = \frac{x-60}{10} \Rightarrow x = 72.8 \text{ marks}$$



Hypothesis Testing

Often we need to make a decision about a population based on a sample.

1. **Is a coin which is tossed biased if we get a run of 8 heads in 10 tosses?**
Assuming that the coin is not biased is called a NULL HYPOTHESIS (H_0)
Assuming that the coin is biased is called an ALTERNATIVE HYPOTHESIS (H_1)

2. **During a 5 minute period a new machine produces fewer faulty parts than an old machine.**
Assuming that the new machine is no better than the old one is called a NULL HYPOTHESIS (H_0)
Assuming that the new machine is better than the old one is called an ALTERNATIVE HYPOTHESIS (H_1)

3. **Does a new drug for Hay-Fever work effectively?**
Assuming that the new drug does not work effectively called a NULL HYPOTHESIS (H_0)
Assuming that the new drug does work effectively called an ALTERNATIVE HYPOTHESIS (H_1)

Margin of Error for Population Proportions

A sample of 60 students in a school were asked to work out how much money they spent on mobile phone calls over the last week. If the mean of this sample was found to be €5·80.
Can we say that the mean amount of money spent by the students in the school (population) was €5·80?
The answer is no, (unless the sample size was the same as the population size), we can't say for certain.

However we could say with a certain degree of confidence, if the sample was large enough and representative then the mean of the sample was approximately equal to the mean of the population.

How confident we are is usually expressed as a percentage.
We already saw (from the empirical rule) that approximately 95% of the area of a Normal Curve lies within ± 2 standard deviations of the mean.

This means that we are 95% certain that the population mean is within ± 2 standard deviations of the sample mean. ± 2 standard deviations is our margin of error and the ± 2 standard deviations depends on the sample size.
If n = 1000 the percentage margin of error of ± 3%

Some Notes on Margin of Error

- As the sample size increases the margin of error decreases
- A sample of about 50 has a margin of error of about 14% at 95% level of confidence
  $$\frac{1}{\sqrt{50}} = \pm 14.14\%$$
- A sample of about 1000 has a margin of error of about 3% at 95% level of confidence
  $$\frac{1}{\sqrt{1000}} = \pm 3.16\%$$
- The size of the population does not matter
- If we double the sample size (1000 to 2000) we do not get do not half the margin of error
- Margin of error estimates how accurately the results of a poll reflect the "true" feelings of the population

Example 1

A survey is carried out on 900 randomly selected people and the result is that 40% are in favour of a change of government. The confidence level is cited as 95%.

(i) Calculate the margin of error.

(ii) The following month another survey was carried out on 900 randomly selected people to see if there was a change in support for the government.
The result is that 42% are now in favour of a change of government. State the null hypothesis.
According to this new survey would you accept or reject the null hypothesis?
Give a reason for your conclusion.

Solution

(i) $\text{Margin of Error} = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{900}} = \pm\, 0.03 = \pm\, 3\%$

(ii) Null hypothesis H_0: "There is no change in the support for the government"

Accept H_0 the null hypothesis.

Reason: The result of the first survey was 40% with a margin of error of $+3$ or -3.
The results of the second survey was 42% which is within + or 3% of the first survey so there is no need for the government to be concerned.

Example 2

In a survey I want a margin of error of $+$ or -5% at 95% level of confidence.

What sample size must I pick in order to achieve this?

Solution

$\text{Margin of Error} = \pm\, 0.05$

$$\pm 0.05 = \frac{1}{\sqrt{n}}$$

$$(\pm 0.05)^2 = \frac{1}{n}$$

$$n = \frac{1}{0.0025}$$

$$n = 400$$