Lecture 4
Sampling Distributions

Review of the normal curve

The behaviour of sample means

The behaviour of sample proportions

Review of the normal curve

In the Last Lecture

- The **standard normal distribution** has a mean of 0 and standard deviation of 1. **Any normal distribution** can be standardised:

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

- For normal distributions:
  - A z-score can be used to find the probability of obtaining a value which is either greater or smaller than a particular value y.

    given y
  - A percentile can be found using the z-score corresponding to the cut-off for the specified area/probability.

    given area/probability

Review Quiz 1

According to an article published on the web site www.PCMag.com, Facebook users spend an average of 190 minutes per month on their Facebook pages. Suppose that the current distribution of time spent per month by on a member's Facebook page is normally distributed with a mean of 190 minutes and a standard deviation of 53.4 minutes.

a. Determine the probability that a randomly selected Facebook user spends more than 3 hours per month on his/her Facebook page.

b. Determine the time spent per month on his/her Facebook page for a user who is on the 25^{th} percentile of this distribution.

Solution to Review Quiz 1a

Solution to Review Quiz 1b

Review Quiz 2

The two primary standardised tests used by college admissions in America are the SAT (Scholastic Assessment Test) and the ACT (American College Test). **SAT scores have a mean of 500 and a standard deviation of 100** while **ACT scores have a mean of 21 and a standard deviation of 4.7**. Both scores follow normal distributions. Bernard applies to college after scoring 650 on his SAT exam and Maria also applies to the same college after scoring 30 on her ACT exam. Which student performed better overall? The college only accepts students who score in the highest 5% for either the SAT or the ACT exam. Will both students be accepted into the college?

Source: Agresti, Franklin (2009). Statistics

Solution to Review Quiz 2

The behaviour of sample means

Why Are We Interested in Averages?

We use averages in everyday life to summarise populations in which we are interested.

Examples:

- Life expectancy:

  This measures the **average** expected length of life of a country's population.

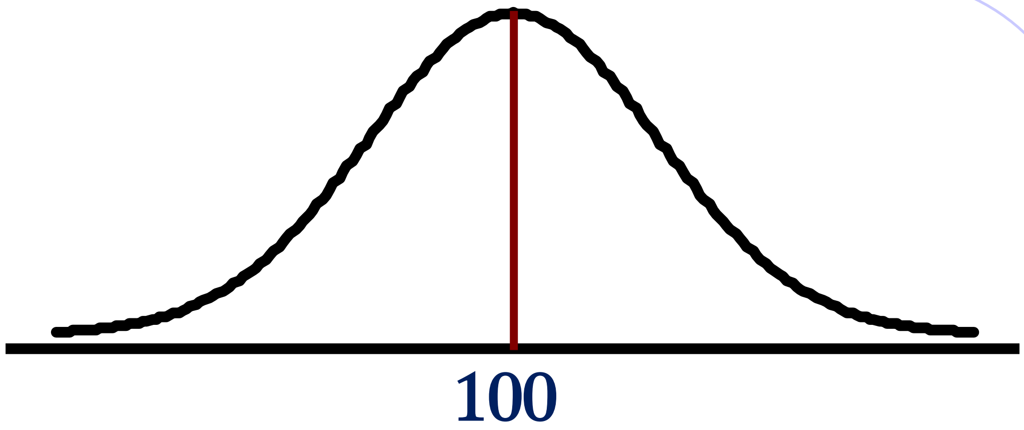
- Olympic results in diving & gymnastics:

  Eight judges may each give a score and the **average** is used, after discarding the two most extreme scores.

How do Sample Means Behave in Repeated Sampling?

Recall:

IQ scores are normally distributed with μ = 100, σ = 15.

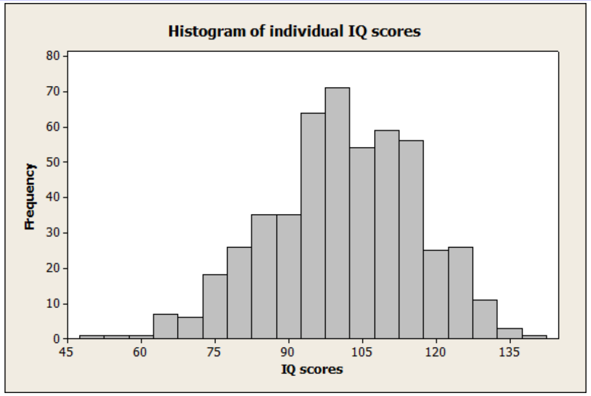


Consider taking many random samples from this population, each of size n = 4. We then calculate the mean of each of these samples.

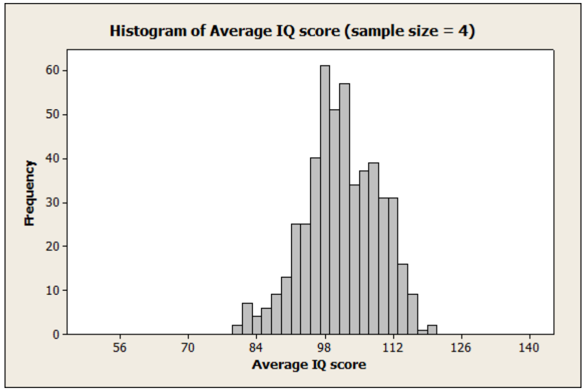
If we constructed a histogram of the means of these samples, what shape should the distribution follow?

What would happen if the sample size was increased to size n = 10 for each sample?

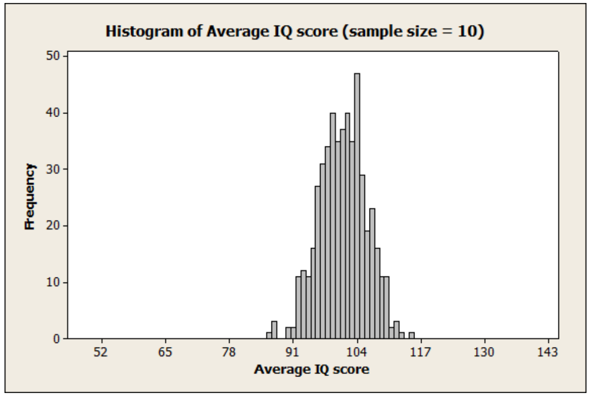
Three Simulation Studies: IQ Scores: μ=100, σ=15



Individual scores



Averages from many samples of size 4



Averages from many samples of size 10

Simulation ... tatistics

This is just the number of times each experiment was repeated.

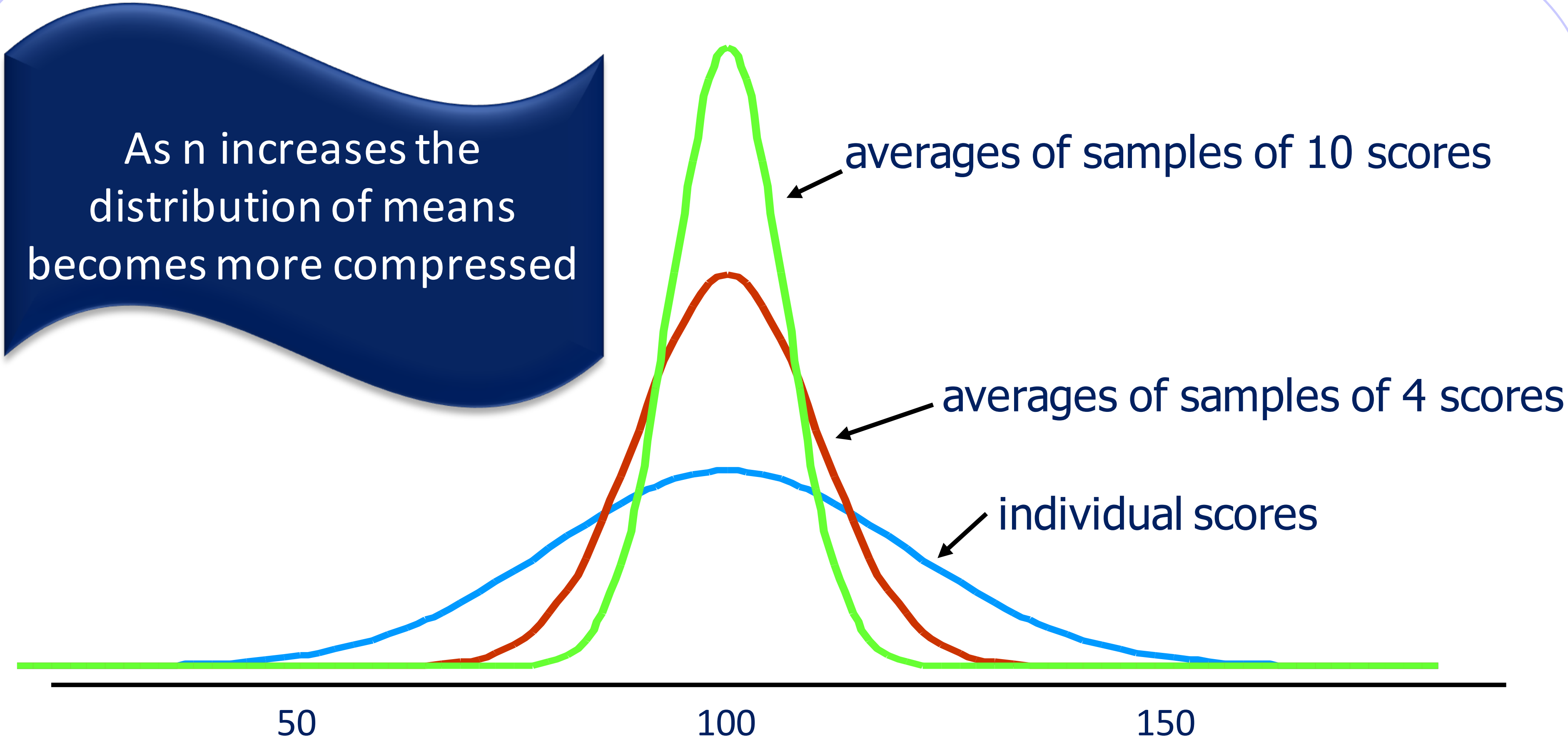
IQ scores
$\mu = 100$
$\sigma = 15$

Numerical Summaries

Variable	# Samples	Mean	StDev	Min	Max
Individuals (n=1)	500	101.13	15.08	50	141
Means (n = 4)	500	101.36	7.61	80	121
Means (n = 10)	500	101.07	4.75	87	115

- The descriptive statistics show that the overall means of the individual scores and of the sample means are all close to 100.
- What is happening to the standard deviations? They are decreasing - the distributions are becoming more compact. This is clear from both the descriptive statistics and the histograms.
- The shape of each histogram (slide 5.9) appears normal.

What Did We Expect? Plots of the Distributions



The Accuracy of an Average

How much more accurate is an average (compared to a single measurement)?

We can use the standard deviation to find out.

The standard deviation of the *averages* obtained from random samples of size n, is given by:

$$\frac{\sigma}{\sqrt{n}} = standard\ deviation\ of\ \bar{y}$$
$$= standard\ error\ of\ \bar{y} = \sigma_{\bar{y}}$$

where σ is the standard deviation of an *individual* measurement.

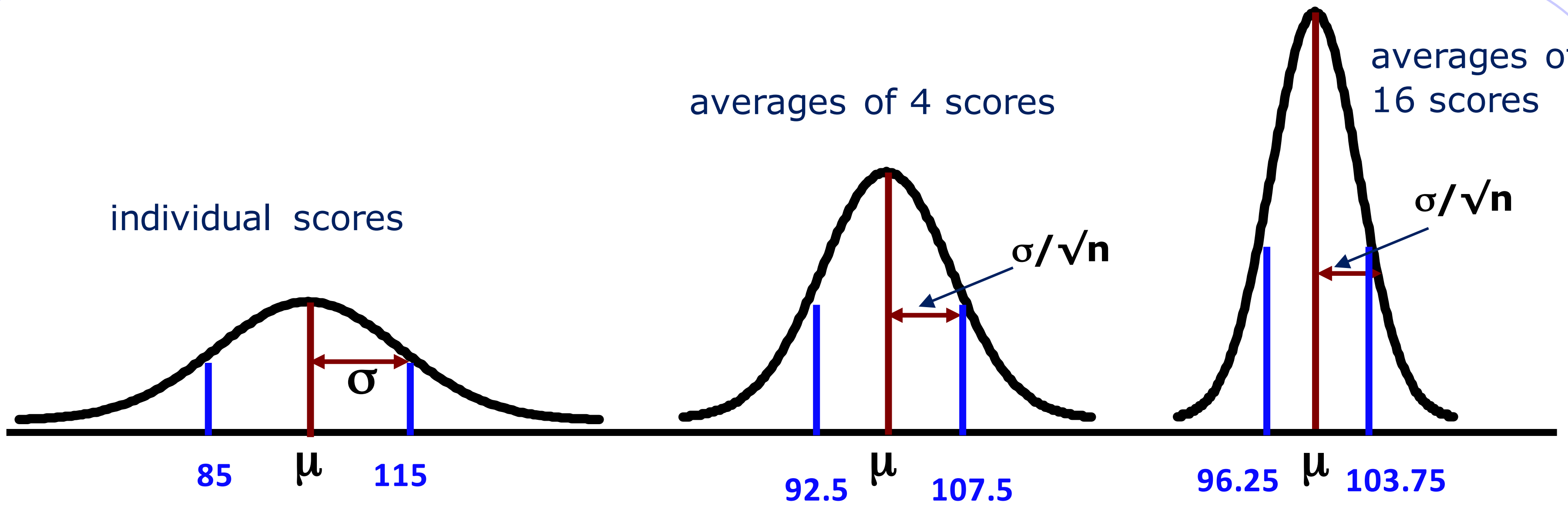
So to *increase* the accuracy of an estimate you would need to reduce the standard error. To do this, you would need to *increase* the sample size.

Sampling Distributions: IQ Scores

number of IQ scores in each sample	standard deviation of average
1	15.00 $\left(\sigma/\sqrt{1} = \sigma\right)$
2	10.61 $\left(\sigma/\sqrt{2}\right)$
4	7.50 $\left(\sigma/\sqrt{4}\right)$
8	5.30 $\left(\sigma/\sqrt{8}\right)$
16	3.75 $\left(\sigma/\sqrt{16}\right)$

Quadrupling the sample size doubles the accuracy. eg. the variation among sample means from samples of size 16 is half the variation among sample means from samples of size 4.

Normal Population Distributions: IQ Scores



- 68% of people have IQs between **85 and 115**
- 68% of samples of size 4 have mean IQs between **92.5 and 107.5**
- 68% of samples of size 16 have mean IQs between **96.25 and 103.75**

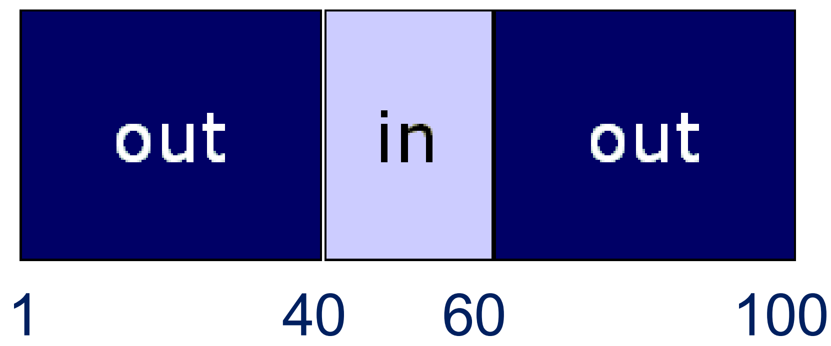
Other Population Distributions

- What about populations that are not normal?
- What does the distribution of sample means look like if the samples come from a *uniform* distribution?
- Does sample size matter?
- To investigate this, let us look at a fictitious data set.
- Remember that, in a uniform or rectangular distribution, each outcome has the same chance (or probability) of occurring.

The University of Hell

The University of Hell charges $1000 per application.

Applicants are selected by spinning a chocolate wheel containing the numbers from 1 to 100. If your number is in the range 41-60, you are admitted.



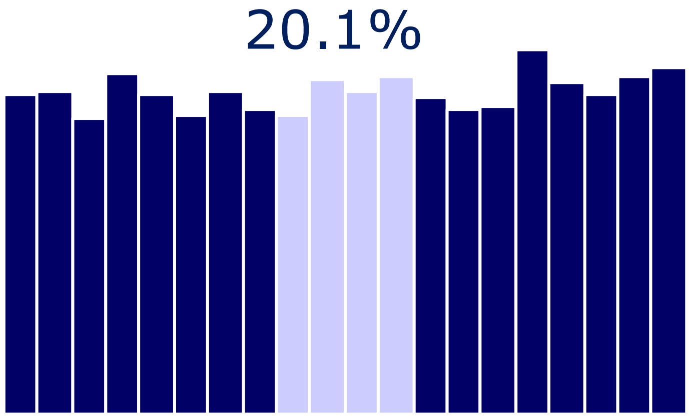
So the chance of an applicant being admitted is 1 in 5: if there are 5000 applicants, approximately 1000 should be admitted.

Improving Your Prospects

- Suppose the University allows applicants to pay $2000 for *two* spins of the chocolate wheel.
- For admission purposes, the two results are *averaged*.
- So if spin 1 gives 36, and spin 2 gives 65, the average is (36+65)/2 = 50.5, and you are admitted;
- BUT if spin 1 gives 45, and spin 2 gives 83, the average is (45+83)/2 = 64, and you are not admitted.
- *How does this affect the chance of gaining admission?*
- Simulation may be used to answer this question.

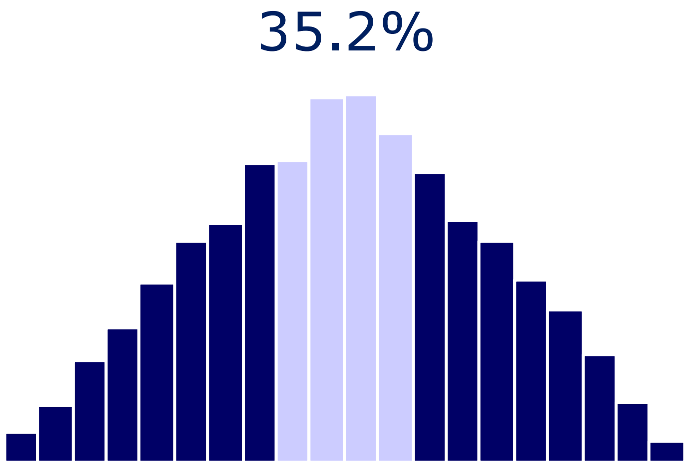
Simulation Studies

5000 individual spins



result:
1004 (20.1%)
admitted

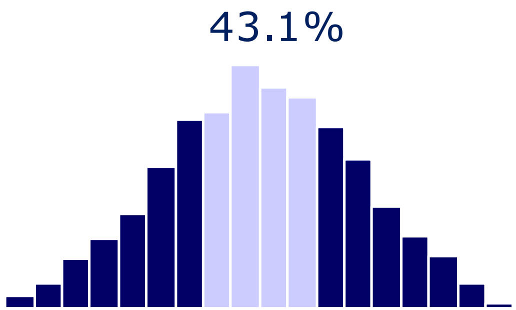
5000 double spins (averages of samples of size n = 2)



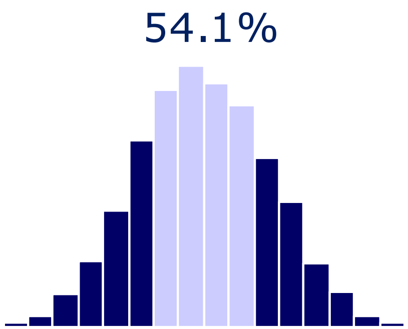
result:
1758 (35.2%)
admitted

Simulation Studies: Distributions of Averages When the Population is Uniform

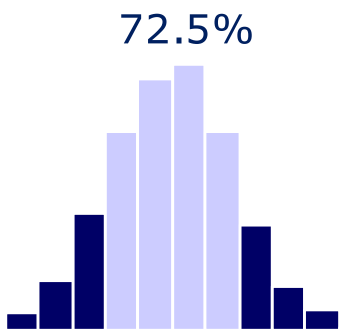
5000 triple spins
(average of samples of size n = 3)
43.1% admitted



5000 quintuple spins
(average of samples of size n = 5)
54.1% admitted



5000 decuple spins
(average of samples of size n = 10)
72.5% admitted



What Did We Expect?
Plots of the Distributions

Uniform
(U)

n = 1

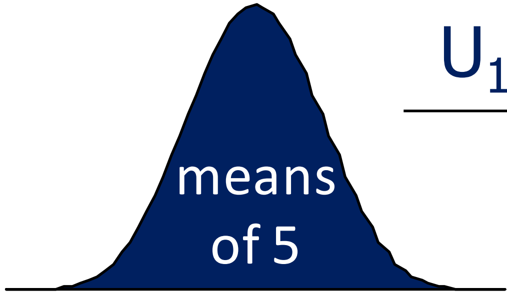
$$\frac{U_1+U_2}{2}$$

n = 2

$$\frac{U_1+U_2+U_3}{3}$$

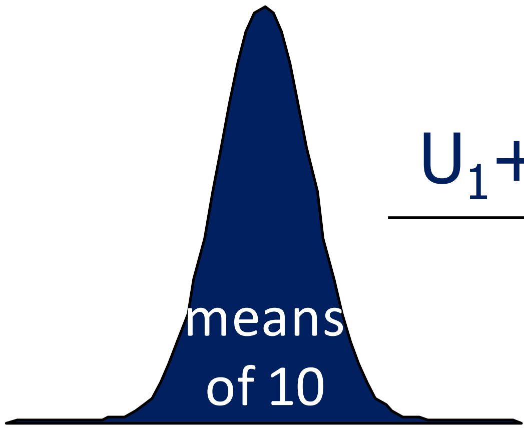
n = 3

n = 5



$$\frac{U_1+U_2+\ldots+U_5}{5}$$

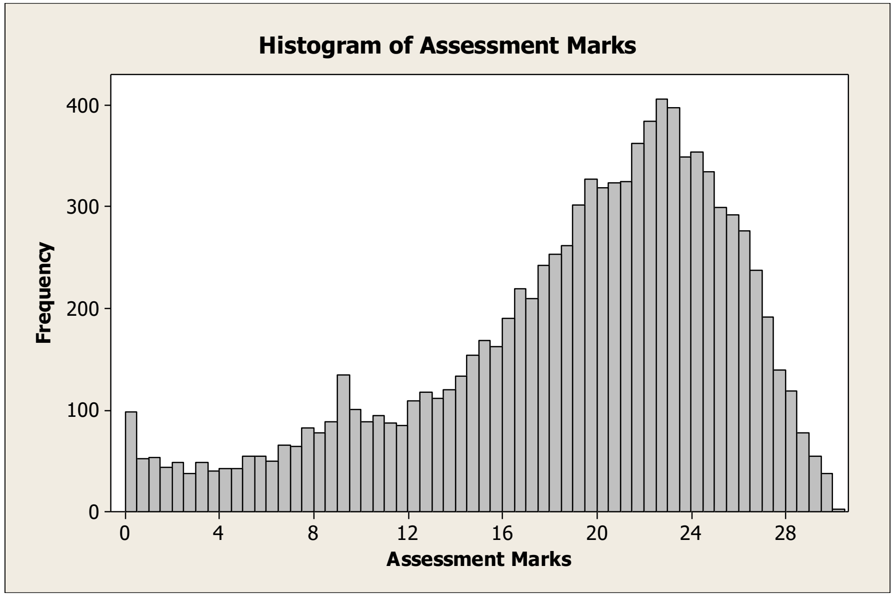
n = 10



$$\frac{U_1+U_2+\ldots+U_{10}}{10}$$

A Further Illustration

Students in an Introductory Statistics unit are awarded assessment marks for homework, assignments and online quizzes throughout the semester. These marks are found to follow a non-symmetric distribution with mean, $\mu = 19$ marks (out of a possible 30) and standard deviation, $\sigma = 6.7$ marks. The histogram below shows the distribution of assessment marks for the last 10000 students who completed the unit.



If random samples of the same size are selected from this distribution, what will the distribution of the sample means look like?

Simulation Studies: Distributions of Averages When the Population is Not Symmetric

In order to investigate this, we performed some simulations.

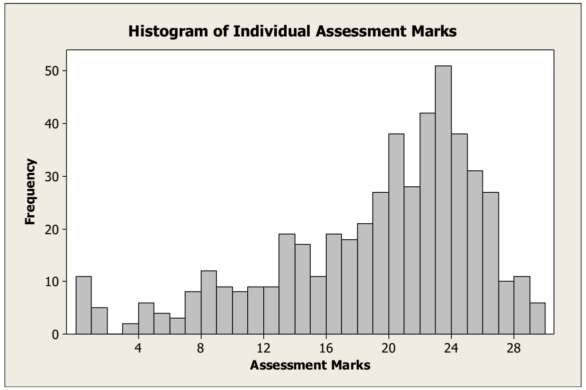
Simulation 1: 500 students were randomly selected and the **individual assessment mark** for each of these 500 students was recorded.

Simulation 2: 500 samples of size n = 4 students were randomly selected and the **sample mean** of each of these 500 samples was recorded.

Simulation 3: 500 samples of size n = 25 students were randomly selected and the **sample mean** of each of these 500 samples was recorded.

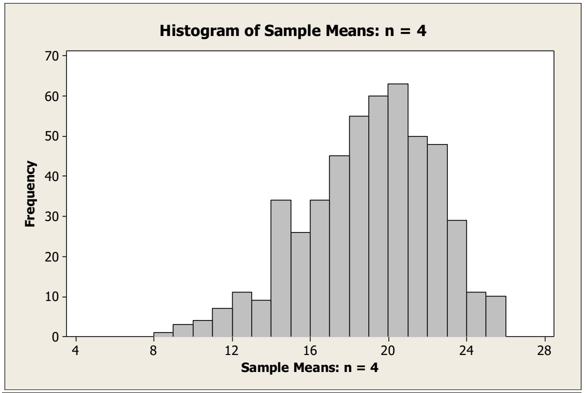
Simulation Studies: Distributions of Averages When the Population is Not Symmetric

500 individual marks



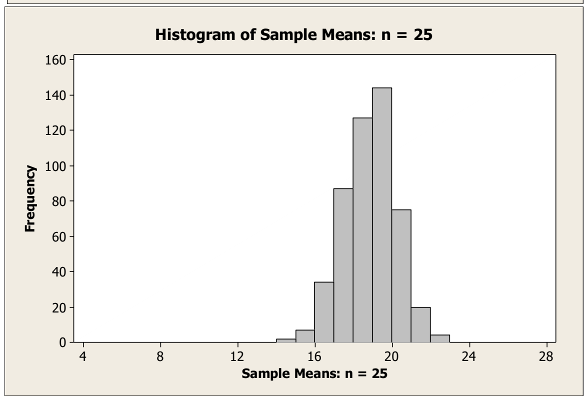
Simulation 1:
Individual assessment marks for 500 randomly selected students
Mean of individual marks = 19.15
Standard deviation = 6.69

500 sample means: n = 4



Simulation 2:
Sample means from 500 randomly selected samples of size 4
Mean of sample means = 19.01
Standard deviation = 3.34

500 sample means: n = 25



Simulation 3:
Sample means from 500 randomly selected samples of size 25
Mean of sample means = 18.87
Standard deviation = 1.32

Sampling Distributions when the Population is Not Normal

In the simulation studies where the population was not normally distributed (uniformly distributed numbers on a chocolate wheel and assessment marks from a non-symmetric distribution) we observed the distribution of *sample means*, in repeated sampling. In all of these simulations we found that, **as the sample size, n, increased**:

- the centre of the distribution of sample means remained close to the population mean.
- the dispersion (spread) of the distribution of sample means decreased.
- the shape of the distribution of sample means became closer to a *normal distribution*.

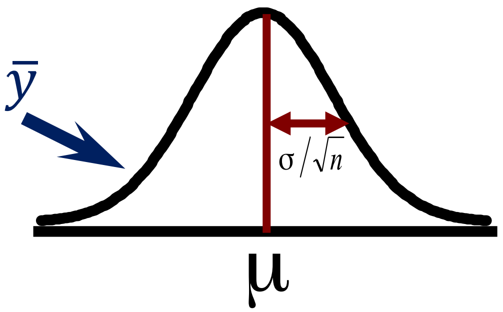
The Central Limit Theorem

The *normal* distribution is the limiting distribution of the *average* of several (n) independent components, as n increases. This result is called the *central limit theorem* and it states that:

If we take random samples of the same size n from a population which is not normally distributed, then the sample means will follow a normal distribution provided n, the sample size, is large enough.

$$\bar{y} = \frac{1}{n}(y_1 + y_2 + \cdots + y_n)$$

Distribution of $\bar{y}$
(n large)



What is “Large Enough”?

The closer the original population is to a normal distribution, the smaller the sample size required for the CLT effect to apply.

If the original population is not too far from normal, an n of 25 will be ‘large enough’ to assume an approximate normal distribution for sample means.

If the original population is a normal distribution, then sample means will follow an exact normal distribution, regardless of sample size (n).

Behaviour of Sample Means

- We've seen that sample means have a standard deviation $\sigma/\sqrt{n}$, which we call the *standard error* of the sample means, denoted by $se_{\bar{y}}$ or $\sigma_{\bar{y}}$.
- We have also seen that, if we take many samples of size n, the average of the sample means will be close to μ, the population mean.
- What is more, the distribution of sample means, where samples may be taken from almost any parent population, will be approximately normal provided the sample size is large.
- We will now combine our understanding of the Central Limit Theorem, the behaviour of averages, standard errors and the normal tables to find probabilities relating to sample means.

Example 1: Finding a Probability for an Average

What is the probability that the **mean** IQ in a class of 30 school students will be greater than 110?

Let Y = IQ score of a school student.

Then $\bar{y}$ = average IQ of 30 school students.

We know that Y is normally distributed with a mean of $\mu = 100$ and standard deviation $\sigma = 15$.

We also know that $\bar{y}$ is from a normal distribution with $\mu = 100$ and standard error $\sigma_{\bar{y}} = 15/\sqrt{30} = 2.739$.

Question???

What is the **probability** that

the ***average* IQ** of a class of 30

school students will be

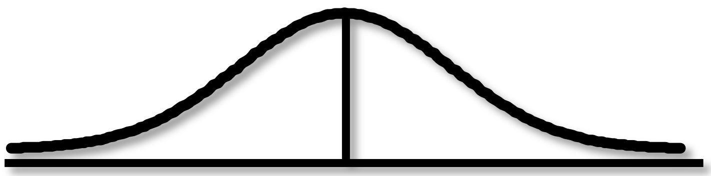
greater than 110?

Example 1: Distributions

Let Y = IQ score of a school student.

Distribution of IQ scores:

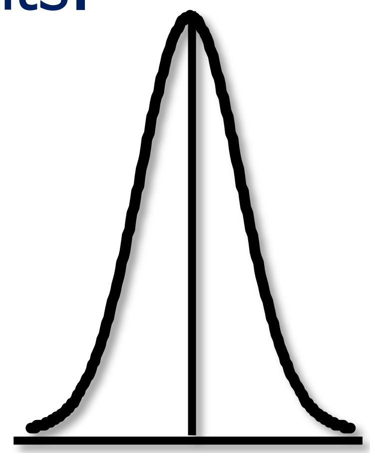
shape	centre	spread
normal	$\mu = 100$	$\sigma = 15$



Then $\bar{y}$ = *average* IQ of 30 school students.

Distribution of average IQ scores:

shape	centre	spread
normal	$\mu = 100$	$\sigma_Y = 15/\sqrt{30}$ $= 2.739$



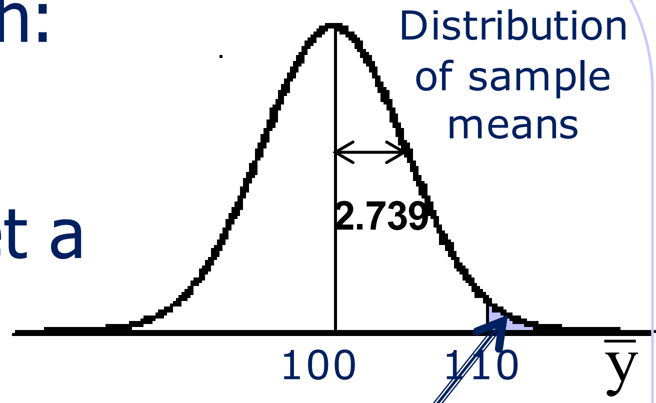
Standardising

Individuals scores (y) $$z = \frac{y-\mu}{\sigma}$$

Sample means $(\bar{y})$ $$z = \frac{\bar{y}-\mu}{\sigma/\sqrt{n}}$$

Example 1 continued…..

Sample means for samples of size 30 would follow a normal distribution with:
μ = 100 and $se_{\bar{y}}$ = 2.739.



What is the probability that we will get a sample mean of at least 110?

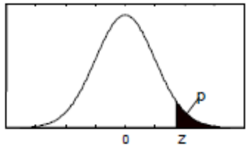
This probability is represented by the shaded area.
We need to calculate a z-score:

$$z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{110 - 100}{15/\sqrt{30}} = 3.65$$

From the z-table, the tabulated area = 0.00013

So the probability of obtaining a mean IQ of at least 110, from a sample of size 30, is 0.00013.

z-table



|z|	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
4.0	.000032									

Standard Normal Probabilities: Single tail areas corresponding to z-values for the standardised normal curve

Example 2: Finding a Probability for an Average

The weekly income of first year full time students at MU has a mean of \$200 and standard deviation of \$240. The tutor of one class has said that she will treat each student to a drink in the bar if the average income of the class is less than \$100 per week. What is the probability that the tutor of the class will have to buy each of her 36 students a drink?

Let Y = weekly income of full time students.

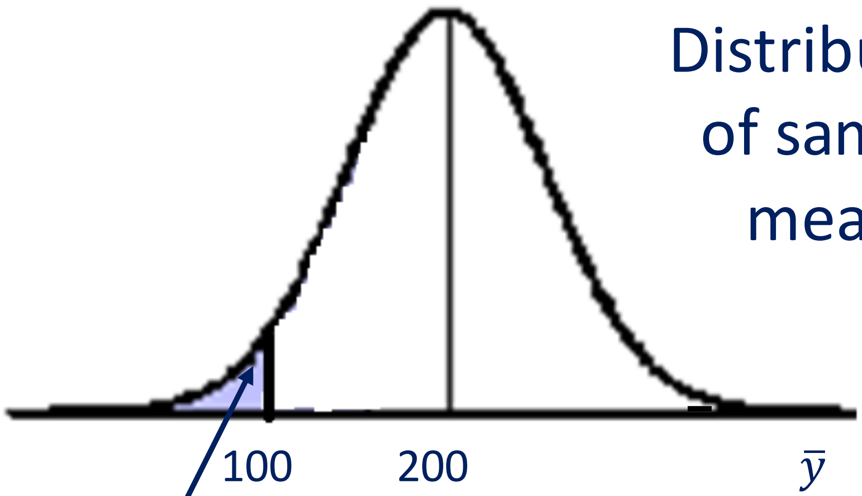
Then $\bar{y}$= average weekly income of 36 full time students.

We are not told here that Y is normally distributed. However, a sample of size 36 is 'large enough' for the CLT to apply. As long as incomes are not extremely skewed we can assume that $\bar{y}$ is drawn from a normal distribution. The distribution of sample means will have a mean of \$200 and standard error of $240/\sqrt{36}$ = \$40.

Y may not be normally distributed, so an understanding of the CLT is needed

Example 2 continued

μ	= 200
σ	= 240
n	= 36
$se_{\bar{y}}$	240/√36
	= 40

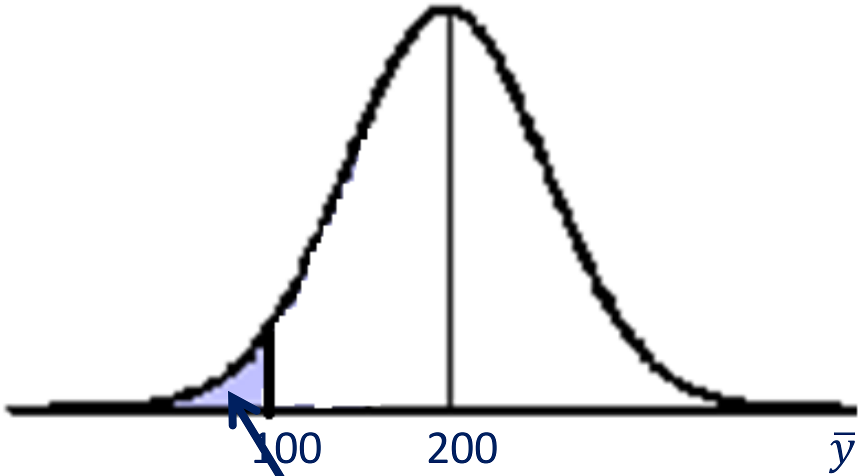


Distribution of sample means

Sample means are from an approximately normal distribution with μ = 200 and $se_{\bar{y}}$= 40.

What is the probability that we will get a sample with a mean which is less than $100?

Example 2 continued



This probability is represented by the shaded area. We need to calculate a z-score:

$$z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}} = \frac{100 - 200}{240 / \sqrt{36}} = -2.5$$

Tabulated area = probability = 0.0062.

So the probability that the tutor will have to buy a drink for each of her 36 students is 0.0062.

Quiz 3

A speed camera on a highway in the Northern Territory records the speed of passing cars. The average speed is found to be 118 kilometres per hour with a standard deviation of 8 kilometres per hour.

a. Random samples of 40 cars are selected and the mean speed for each sample is recorded. Consider the distribution of these sample means:

   i. What is the population mean, μ, of these sample means?

   ii. What is the standard error, $se_{\bar{y}}$, of these sample means?

   iii. Describe the shape of the sampling distribution of $\bar{y}$.

b. Find the probability that the average speed of a randomly selected sample of 40 cars exceeds 120kph.

Solution to Quiz 3

The behaviour of sample proportions

Applying the Central Limit Theorem to Sample Proportions

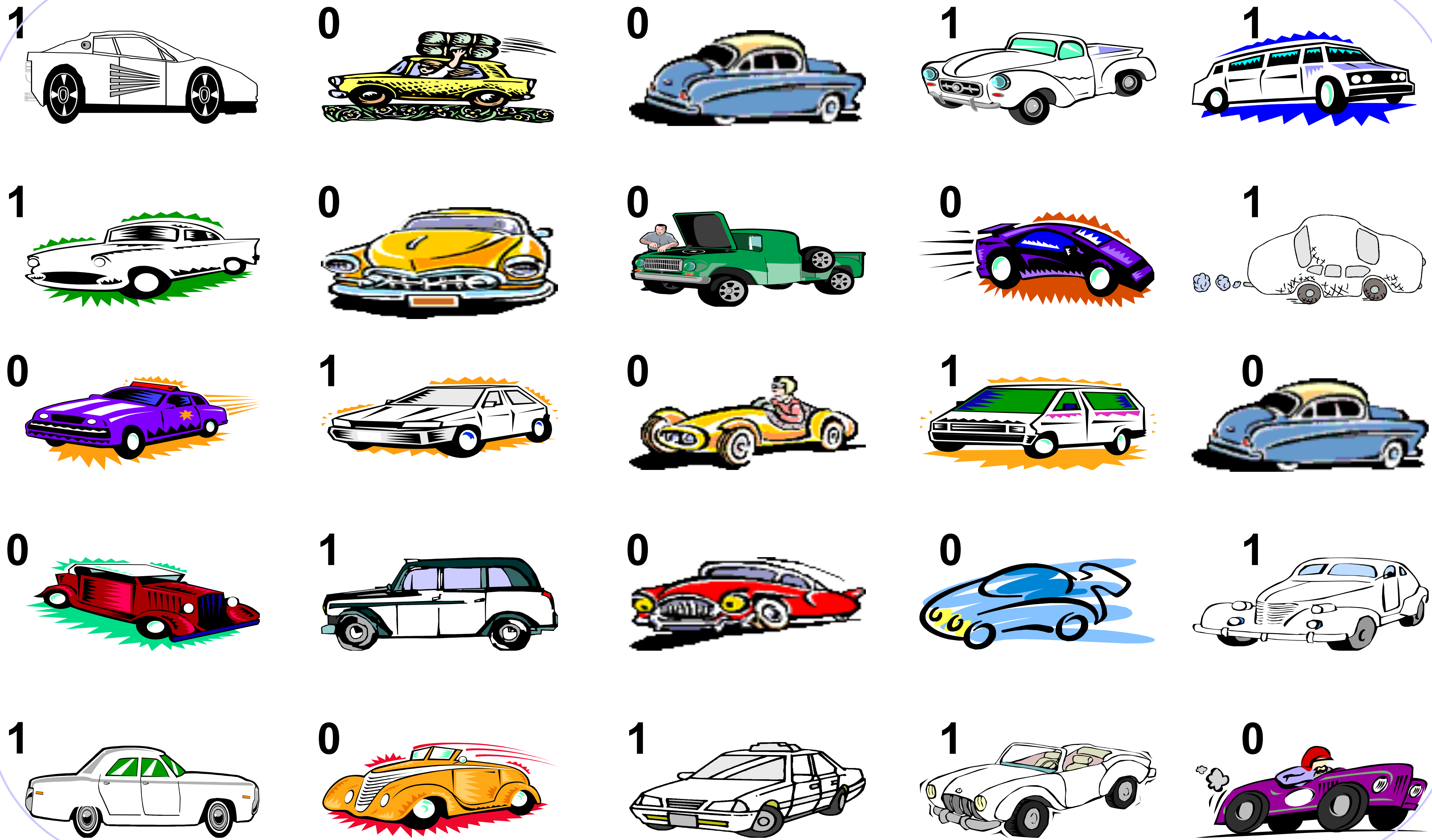
A proportion is just a special case of an average, with dichotomous components (ie. taking values 0 or 1)

So it follows that in repeated sampling, sample proportions should follow an approximately normal distribution, provided n is large enough.

Note that:

- π represents the population proportion
- p represents the sample proportion

What Proportion of These Cars are White?



Calculating a Sample Proportion

Note that the white cars have been given a code of 1 whilst cars of any other colour have been coded 0.

So, to find the sample proportion of white cars we can **add all the codes and divide by the total number of cars in the sample (ie. we find the average).**

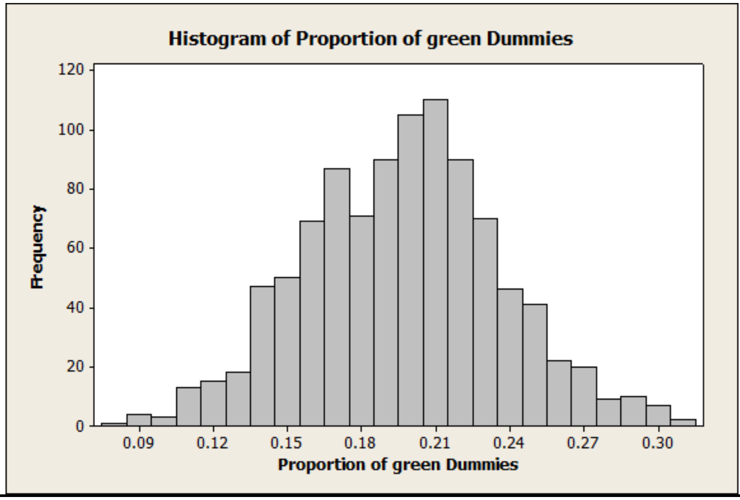
Add the codes of all cars: $1 + 0 + 0 + \ldots + 1 + 0 = 12$

Total number of cars in the sample: $n = 25$

So, $\mathbf{p = 12/25 = 0.48}$

ie. **48% of the cars in the sample are white.**

A Simulation to Illustrate the Central Limit Theorem Applied to Sample Proportions



Descriptive Statistics: Prop. of green Dummies

Variable	Total Count	Mean	Minimum	Median	Maximum
Prop. of green Dummies	1000	0.19665	0.08000	0.20000	0.31000

Dummies are a new item of confectionary that come in colours **red**, **blue**, **yellow**, **green** and **black**. **Each packet contains exactly 100 Dummies** and **should contain similar proportions (ie 20%) of each colour**. To check the proportions of **green** dummies, we buy 1000 packets and record the proportion of **green** in each packet. The output above shows that the proportion of **green** was centred around 0.2 (the population proportion) and has an approximately normal distribution.

The Central Limit Theorem

Remember that the Central Limit Theorem for sample means applies if the sample size, n, is large enough. **In the context of sample proportions, n is large enough if:**

$$n\pi \geq 5 \text{ and } n(1-\pi) \geq 5$$

For the Dummies, $\pi = 0.2$ (the population proportion of green dummies) and n = 100 (ie. 100 dummies in each packet).

So 100*0.2 = 20 and 100 (1 – 0.2) = 80.

Since both values are ≥ 5, we can assume from the CLT, the sample proportions will follow an approximately normal distribution which will be centred around π, the population proportion. The standard error of sample proportions is:

$$\sigma_p = se_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Sample Statistics → estimate → Population Parameters

$Mean$	$\bar{y}$	$\longrightarrow$	μ
$Median$	$\tilde{y}$	$\longrightarrow$	$\tilde{\mu}$
$Std.dev$	s	$\longrightarrow$	σ
$Std.dev$	s^2	$\longrightarrow$	σ^2
$se\ (mean)$			${}^{\sigma}/_{\sqrt{n}}$
$Proportion$	p	$\longrightarrow$	π
$se\ (p)$			$\sqrt{\frac{\pi(1-\pi)}{n}}$

Calculating Probabilities for Sample Proportions

Suppose we are told that **40%** of new cars sold in Australia are white. **Is it unusual** to find 12 *or more* white cars in a random sample of 25 cars?

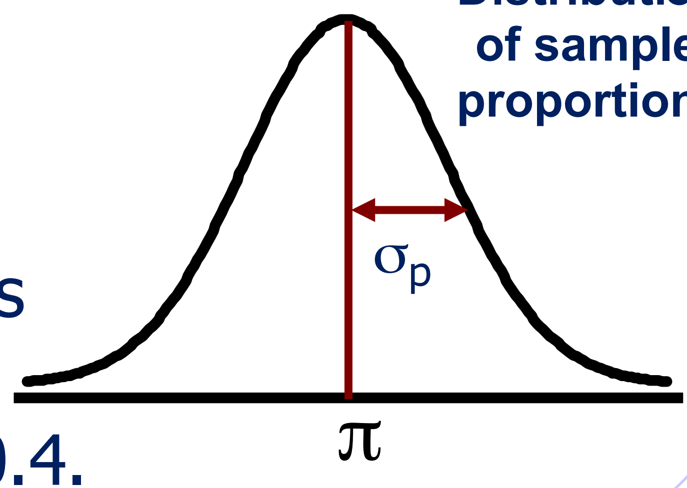
π = 0.4 = population proportion of white cars

Let p = proportion of white cars in a randomly selected sample of size n = 25

$n\pi = 25 \times 0.4 = 10$ and

$n(1 - \pi) = 25 \times 0.6 = 15$

Since both are ≥ 5, we can assume that proportions, from samples of this size, are approximately normally distributed and centred around $\pi = 0.4$.



Standardising: z-scores

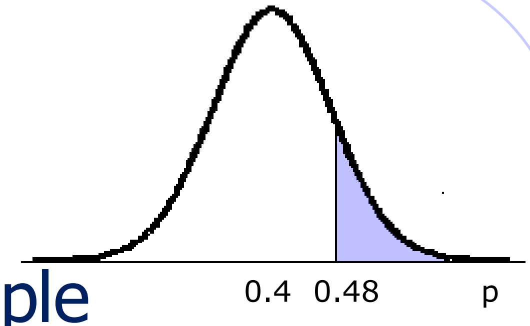
Individuals scores (y) $$z = \frac{y - \mu}{\sigma}$$

Sample means $(\bar{y})$ $$z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}}$$

Sample proportions (p) $$z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}}$$

Probability for a Sample Proportion

$$\pi = 0.4 \; and \; se_p = \sqrt{\frac{0.4(1-0.4)}{25}} = 0.098$$



What is the probability that we'll get a sample proportion of at least 12/25 = 0.48?

We'll calculate a z-score corresponding to p = 0.48:

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.48 - 0.4}{0.098} = 0.82$$

tabulated area = 0.2061.
So the probability of having at least 12 white cars in a random sample of 25 cars is 0.2061.

Quiz 4

According to figures released in 2007 by the Australian Bureau of Statistics following a national survey of the Australian population's mental health and wellbeing, young people carry the greatest burden of mental illness. The onset of mental illness is typically around mid-to-late adolescence. Australian youth (aged 18 to 24 years) have a higher prevalence of mental illness than any other age group.

It was found that 26% of young Australians experience a mental illness every year. The two most common mental illnesses in young Australians are anxiety disorders and depressive disorders.

Find the probability that less than ten out of a random sample of one hundred 18 to 24 year old Australians experience a mental illness in a year.

Solution to Quiz 4

Homework Questions

Homework Question 1

Sheila's doctor is concerned that she may suffer from gestational diabetes (high blood glucose levels during pregnancy). There is variation in both the actual glucose level and blood test that measures the level. A patient is classified as having gestational diabetes if the glucose level is above 140 milligrams per decilitre (mg/dl) one hour after a sugary drink is ingested. Sheila's measured glucose level one hour after ingesting the sugary drink varies according to the normal distribution with μ = 125 mg/dl and σ = 10 mg/dl.

a. If a single glucose measurement is made, what is the probability that Sheila is diagnosed as having gestational diabetes?

b. If measurements are made instead on three separate days and the mean result is compared with the criterion 140 mg/dl, what is the probability that Sheila is diagnosed as having gestational diabetes?

Source: Moore, McCabe and Craig (2012) *Introduction to the practice of statistics (seventh edition), Freeman (adapted)*

Solution to Homework Question 1

Homework Question 2

A recent study by Allstate Insurance Co. Finds that 82% of teenagers have used mobile phones whilst driving. Suppose a random sample of 100 teen drivers is taken.

a. What is the probability that the sample proportion is less than 0.80?

b. What is the probability that the sample proportion is within ± 0.02 of the population proportion?

Source: Jaggia , S. and Kelly , A. (2013) *Business statistics: Communicating with numbers McGraw-Hill Irwin (adapted*)

Solution to Homework Question 2

Lecture 5 Summary

- Averages are less variable than individual measurements.
- When individual measurements are not normally distributed, averages of *n* components have distributions that are closer to a normal distribution than the distribution of the individual measurements.
- The normal distribution is the *limiting distribution of an average*, as n increases (*Central Limit Theorem*).
- The standard deviation of averages (called the *standard error*) is given by $\sigma/\sqrt{n}$ (denoted by $se_{\bar{y}}$ or $\sigma_{\bar{y}}$) where σ is the standard deviation in the population of individual components.
- Sample proportions have a standard error, se_p or $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ where π is the population proportion.
- The Central Limit Theorem applies to the distribution of proportions.

  ie. sample proportions follow an approximately normal distribution as long as $n\pi$ and $n(1-\pi)$ are both ≥ 5.

Textbook References

Further information on the topics discussed in this lecture can be found in:

Modern Statistics: An Introduction
by Don McNeil and Jenny Middledorp
(ISBN 9781486007011).

- Chapter 5: Pages 92 – 112