Chapter 6

Statistical inference for the population mean

6.1 Introduction

In Chapter 3 we concentrated on pure description of data, although we recognised that this might prompt us to ask pertinent questions about the population from which the sample was drawn. What exactly does the sample, often a tiny subset, tell us of the population? We can never observe the whole population, even if it is finite, except at enormous expense, and so the population mean and variance (or indeed any aspect of the population distribution) can never be known exactly. We call these unknown quantities *parameters* and use Greek letters to denote them:

- μ ("mu") is the symbol commonly used for the population mean, and
- σ ("sigma") is commonly used for the population standard deviation.

Hopefully (and if we have a representative sample), the sample mean ($\bar{x}$) will be quite close to the true population mean μ; likewise, the sample standard deviation (s) will be a good estimator for σ. **In this section, we concentrate on $\bar{x}$ as an estimator for μ.**

Before we can use our sample of n observations we must ask the question: Is $\bar{x}$ a "good" estimate of μ? How do we infer (find something out about) the unknown μ using $\bar{x}$? So long as the sample size n is fairly large, we can hope that $\bar{x}$ is close to μ. But how close is it? To answer this question, we must make some plausible assumptions about the population. But first, let's consider the following example.

Example: *The Vintage Clothing Co.*

The *Vintage Clothing Co.* are a large retailer of bespoke and retro clothing. They have 1,000 branches across the U.K., and all of their branches are open on Sundays. However, they are considering whether or not it is worthwhile staying open on Sundays. The table overleaf shows the number of transactions at each of their shops on Sunday 24th February 2013.

Suppose the marketing department of *The Vintage Clothing Co.* are interested in the average number of transactions across all their stores on Sunday 24th February 2013. Can we work this out *exactly*?

The answer is "yes", as we have data from every single branch in the table overleaf. Actually, this table shows we have taken a *census* – every single branch has been asked to provide us with data. So in this case, it is possible to work out the *population mean* μ:

$$\mu = \frac{282 + 258 + 399 + 271 + \ldots + 426 + 477}{1000} = 320 \text{ transactions.}$$

Now, let's suppose the company don't have the time/resources to take a census. In fact, a week before the 24th February, just five stores are selected at random, and we work out the mean number of transactions using the data from these five stores only. Let's suppose the top left–hand block in the table are stores 1–100, the next block along are stores 101–200, etc. We put the numbers 1–1000 into a bag and draw, without replacement, 5 numbers at random:

Store	No. of transactions, X
637	$x_1 = 374$
327	$x_2 = 452$
849	$x_3 = 271$
666	$x_4 = 419$
680	$x_5 = 643$

Let's suppose this is the only information we have. It is no longer possible to work out the true population mean, as we don't have information from every single shop; we can now only work out the *sample mean* $\bar{x}$:

$$\bar{x} = \frac{374 + 452 + 271 + 419 + 643}{5} = 431.8 \approx 432 \text{ transactions.}$$

Obviously, the marketing team are not just interested in what goes on in these five shops; however, this is the only information they have, and so they use this information to draw conclusions about all 1,000 shops as a whole. This is known as the process of *statistical inference* – we are trying to *infer* things about the population, based on the limited information in our sample. Hopefully, provided we don't have a *biased sample*, $\bar{x}$ will do a good job at estimating μ. Has it done a good job here?

282	258	399	271	343	285	247	513	171	123	168	327	430	240	410	341	90	512	245	336
290	263	446	185	330	111	243	376	139	351	311	389	546	321	393	487	287	514	149	315
264	320	217	257	349	640	97	298	393	454	363	354	360	326	199	502	154	273	213	413
293	407	362	270	344	263	290	263	50	253	345	581	229	264	304	394	246	235	417	452
499	276	412	323	310	177	248	178	409	275	278	307	495	515	232	432	577	269	370	248
339	404	371	262	336	218	274	483	211	245	316	381	432	233	223	447	412	250	262	337
202	133	356	408	224	379	197	278	235	509	171	232	429	315	326	602	63	290	230	121
242	389	219	206	393	437	306	152	294	271	230	398	346	344	379	347	468	300	325	237
305	174	291	261	214	532	335	63	100	357	190	347	208	420	322	463	203	216	356	504
389	236	445	378	255	301	308	150	289	453	464	273	211	450	222	250	214	259	296	356
320	420	357	160	372	99	316	218	248	322	145	399	433	393	403	361	241	234	388	255
261	279	369	342	168	322	304	254	99	503	303	212	105	166	257	422	460	331	288	410
346	370	235	355	65	340	420	338	568	644	164	288	319	159	324	208	452	297	305	259
268	340	305	361	319	519	293	380	286	431	402	329	363	330	612	248	302	592	589	349
446	588	304	454	164	240	293	478	540	339	245	257	222	471	469	273	244	126	174	183
277	216	555	401	380	338	212	476	77	363	140	451	329	66	217	461	435	380	314	324
522	111	119	316	116	471	142	336	277	101	518	264	226	256	539	324	320	292	476	324
333	332	404	362	202	204	341	80	333	267	439	136	343	389	244	370	268	362	317	400
372	595	314	182	470	192	555	374	368	192	225	321	435	403	316	312	307	368	236	452
192	63	407	125	253	89	70	186	491	342	122	367	106	334	161	177	180	355	356	317
454	122	286	39	361	262	316	272	285	201	191	162	229	334	278	231	154	290	277	392
644	297	398	118	246	148	478	167	337	344	395	334	255	401	504	304	408	204	673	126
192	507	41	457	405	306	282	446	195	512	252	510	557	191	321	404	542	438	291	449
377	240	441	308	346	265	375	332	580	130	353	426	95	588	332	109	467	333	388	309
263	529	172	529	315	257	481	260	297	382	438	64	226	185	369	275	320	126	321	375
190	340	337	224	363	212	371	229	175	388	332	315	389	452	266	393	323	253	280	420
219	400	378	241	616	551	359	489	314	450	645	224	320	405	182	251	370	341	318	232
240	471	293	240	184	296	617	565	206	147	169	401	140	462	389	310	262	334	263	269
323	351	187	544	387	425	353	175	378	484	205	295	413	189	559	251	480	283	262	304
213	574	579	325	246	206	419	306	471	264	270	300	278	131	561	328	440	514	280	391
281	403	256	348	183	161	444	482	338	268	313	252	179	414	444	266	400	435	433	506
203	269	450	322	459	183	212	242	144	406	401	174	605	270	487	494	235	316	368	319
260	254	157	377	145	284	401	220	452	59	335	467	251	192	371	298	317	382	363	397
282	303	328	378	363	636	374	143	495	239	423	496	411	462	282	411	203	395	590	388
278	272	417	666	233	316	287	268	186	247	339	397	276	291	324	81	271	399	129	325
247	152	315	224	130	323	352	276	398	338	231	258	310	421	215	85	237	356	439	348
507	277	240	188	321	419	370	374	211	224	340	264	441	226	563	279	297	114	546	277
281	196	498	375	348	234	469	103	324	643	315	293	444	109	408	100	477	293	138	206
485	279	494	513	97	293	669	312	425	70	181	210	241	187	448	55	253	564	404	382
31	496	234	200	411	386	218	382	483	405	435	414	379	360	194	291	393	247	314	285
204	188	444	416	106	485	276	250	248	200	352	463	251	197	197	456	293	333	373	240
295	297	271	141	319	256	197	110	338	237	249	291	393	437	432	274	202	182	176	212
482	96	272	296	323	289	285	160	203	336	217	321	202	266	253	436	390	259	596	383
236	291	226	250	270	439	360	310	326	415	447	336	354	273	243	390	213	318	346	599
637	255	61	393	324	492	484	259	271	150	550	185	224	352	387	441	232	261	313	410
246	529	97	448	369	199	140	498	287	293	258	431	267	396	217	340	278	297	387	281
162	237	305	239	246	412	632	385	342	340	673	414	298	383	152	438	408	452	492	603
439	223	404	466	380	214	155	410	291	234	248	325	391	338	416	262	361	358	484	129
152	363	90	383	365	500	362	190	343	138	233	179	200	476	128	308	221	649	278	152
525	275	355	585	394	183	488	323	312	595	257	434	160	375	478	353	239	331	426	477

Table 6.1: The number of transaction at each branch of *The Vintage Clothing Co.*

This begs the question: "*How accurate can our sample mean be in estimating the population mean?*". Let's take another random sample by drawing another five numbers from the bag, at random:

Store	No. of transactions, X
558	$x_1 = 253$
428	$x_2 = 446$
903	$x_3 = 251$
364	$x_4 = 256$
14	$x_5 = 185$

This gives

$$\bar{x} = \frac{253 + 446 + 251 + 256 + 185}{5} = 278.2 \approx 278 \text{ transactions.}$$

This sample mean is much closer to the population mean, but still not very close. Also, it is quite different from the mean of the previous sample. You could repeat this procedure yourself (filling in the table below) to select three more random samples of size 5, and calculate the sample means. How close are these sample means to the correct population value $\mu = 320$ transactions? In the space below, the u's are random numbers obtained using the *random number generator* button `Ran` on your calculator – I will show you how to do this in class – I wouldn't expect you to draw numbers from a bag!

Your random sample 1

$u =$	$store =$	$x_1 =$
$u =$	$store =$	$x_2 =$
$u =$	$store =$	$x_3 =$
$u =$	$store =$	$x_4 =$
$u =$	$store =$	$x_5 =$
		$\bar{x} =$

Your random sample 2

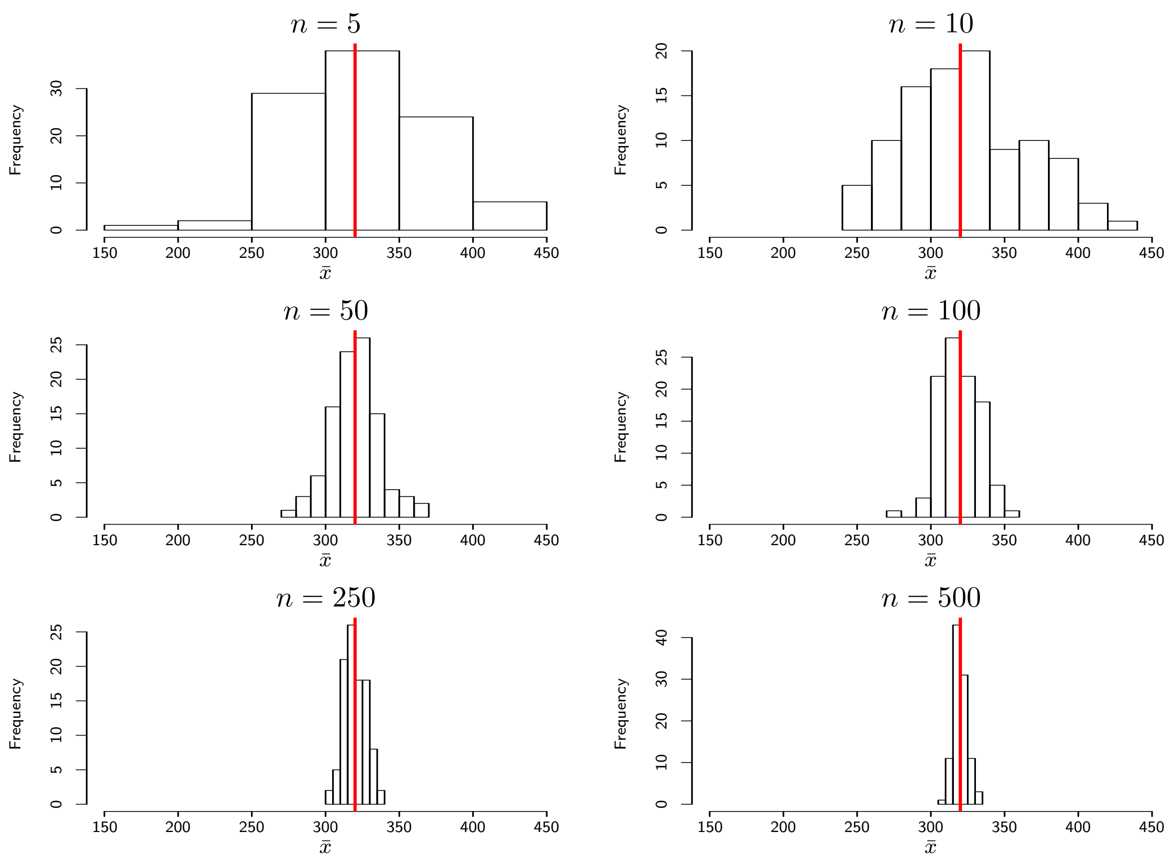
$u =$	$store =$	$x_1 =$
$u =$	$store =$	$x_2 =$
$u =$	$store =$	$x_3 =$
$u =$	$store =$	$x_4 =$
$u =$	$store =$	$x_5 =$
		$\bar{x} =$

Your random sample 3

$u =$	$store =$	$x_1 =$
$u =$	$store =$	$x_2 =$
$u =$	$store =$	$x_3 =$
$u =$	$store =$	$x_4 =$
$u =$	$store =$	$x_5 =$
		$\bar{x} =$

...Comments...

In fact, we could take many samples, and it's very likely that we'll get a different value for $\bar{x}$ each time; it's also very *unlikely* that any of our $\bar{x}$'s will be exactly the same as the true population mean μ. The graph below shows histograms of $\bar{x}$'s taken from 100 samples from our population of 1,000 stores. In the top–left graph, we have taken samples of size $n = 5$, like we did in the last couple of pages of these notes. In the other graphs, moving from left to right, we have increased this to 100 samples of size $n = 10$, $n = 50$, $n = 100$, $n = 250$ and $n = 500$. The vertical line indicates the true population mean $\mu = 320$.



You should notice two things:

1. The distribution of $\bar{x}$'s, in all plots, looks like a Normal distribution (bell–shaped curve; see Chapter 5);

2. As we increase the sample size (n), the distribution for $\bar{x}$ gets more and more concentrated – around the true population value $\mu = 320$!

In fact, what we can see in action in this graph is known as the *central limit theorem.* This is a very powerful result in Statistics which tells us about the distribution of the sample mean $\bar{x}$. We now state this formally.

The Central Limit Theorem

Suppose $x_1, x_2, \ldots, x_n$ are a random sample from *any* population, with mean μ and variance σ^2. If n is large, then

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{approximately;}$$

if $x_1, x_2, \ldots, x_n$ come from a Normal distribution themselves, then this result holds for *any* n.

This means that if we were to take many samples of size n, and for each sample calculate the mean $\bar{x}$, then our histogram of $\bar{x}$'s will *always* be Normally distributed around the true population mean μ (if n is large; however, this result is true for *any* n if our random sample is Normally distributed). What's more, we also know about the *variability* of $\bar{x}$. If we know the population variance σ^2, then the variance of $\bar{x}$ is σ^2/n, giving a standard deviation for $\bar{x}$ of $\sigma/\sqrt{n}$. We call this quantity the *standard error.* We will now use this result to form *confidence intervals* for the population mean μ.

6.2 Confidence Intervals

The values we calculate for sample means and variances are *point estimates*; they are single values based on a limited sample of the whole population. Suppose that we wish to estimate the mean μ of a population. The natural estimate for μ is the sample mean $\bar{x}$. However, as we have seen, $\bar{x}$ is never exactly equal to μ; all we really hope is that $\bar{x}$ will be close to μ. One way of improving our inference is to construct *interval estimates*, more commonly known as *confidence intervals.* We simply place an interval over the point estimate for μ which allows us to say (with a certain level of confidence) within what range the population mean lies. The calculation of these intervals depends on the size of our sample (n), the level of confidence we choose, and whether or not the population variance (σ^2) is known.

6.2.1 Case 1: Known variance σ^2

We know from the results above that, if our random sample is drawn from a Normal distribution, or if n is large (i.e. $n \geq 30$), then

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

If we initially assume we know the population variance σ^2, we can "standardise" $\bar{x}$ using "slide-squash"; i.e.

$$Z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}}.$$

Recall that the standard Normal distribution is $Z \sim N(0, 1)$, i.e. Z has zero mean and variance (and so standard deviation) 1; also recall that approximately 95% of the standard normal distribution lies between -1.96 and 1.96, i.e.

$$\Pr(-1.96 < Z < 1.96) = 0.95.$$

This is easier to see if we draw a picture:

Since we know that $Z = \dfrac{\bar{x} - \mu}{\sqrt{\sigma^2/n}}$, we can write this as

$$\Pr\left(-1.96 < \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} < 1.96\right) = 0.95;$$

rearranging for μ gives us an expression for the *95% confidence interval for* μ:

$$\left(\bar{x} - 1.96\sqrt{\sigma^2/n} \quad , \quad \bar{x} + 1.96\sqrt{\sigma^2/n}\right);$$

thus, we can say that the two values $\bar{x} - 1.96\sqrt{\sigma^2/n}$ and $\bar{x} + 1.96\sqrt{\sigma^2/n}$ are the lower and upper bounds (respectively) of the (95%) confidence interval.

We often write this more simply as

$$\bar{x} \pm 1.96\sqrt{\sigma^2/n}.$$

Going back to *The Vintage Clothing Co.* example, this means that if we were to take 100 samples and for each one calculate a 95% confidence interval (using the formula above), then about 95 of these confidence intervals would "capture" the true population value $\mu = 320$.

Example 6.1

Geordie Sparkz are an electrical company based in Newcastle producing circuitboards for large plasma televisions. One of their machines punches tiny holes in these curcuitboards that should be 0.5mm in diameter. A sample of 30 circuitboards off the production line is inspected; the average diameter of the holes produced by this machine, for this sample, is 0.54mm.
Assuming the machine is set to ensure a standard deviation of $\sigma = 0.12$mm, calculate the 95% confidence interval for the population mean diameter of holes produced by this machine. Do you think there is a real problem with this machine?

6.2.2 Case 2: Unknown variance σ^2

If the population variance σ^2 is unknown, we can no longer use the Normal distribution and instead have to use the t–distribution to calculate confidence intervals. We have seen that when our random sample follows a Normal distribution, or indeed any distribution (if the sample size is large), then the sample mean $\bar{x} \sim N(\mu, \sigma^2/n)$. From this, it follows that

$$Z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}},$$

where Z is the standard Normal distribution, i.e. $Z \sim N(0, 1)$. However, if the population variance is unknown, then the quantity

$$T = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}$$

does *not* have a $N(0, 1)$ distribution (note that the *population* variance σ^2 in Z has been replaced with the *sample* variance s^2 in T). Instead it has a Student's t–distribution. This distribution is similar to the $N(0, 1)$ distribution in that it is symmetrical and bell–shaped, but it is more heavily tailed to allow for greater uncertainty in $\bar{x}$ since the true variability is now unknown. Its exact shape is determined by one parameter called the "degrees of freedom". The table overleaf gives critical values of Student's t–distribution with various degrees of freedom. These numbers depend on two quantities: ν, the degrees of freedom, and p, a probability.

The expression for the confidence interval in this case is similar to the case where σ^2 is known:

$$\bar{x} \pm t_p\sqrt{s^2/n},$$

where σ^2 has been replaced with the sample variance s^2 and t_p is the appropriate value from the t–distribution tables. But how do we find this value?

Tables of critical values for Student's *t* distribution

The table contains values of t for which $\Pr(|T| > t) = p$, where T has a t–distribution with ν degrees of freedom.

		p				
		50%	20%	10%	5%	1%
	1	1.00	3.078	6.314	12.706	63.657
	2	0.816	1.886	2.920	4.303	9.925
	3	0.765	1.638	2.353	3.182	5.841
	4	0.741	1.533	2.132	2.776	4.604
	5	0.727	1.476	2.015	2.571	4.032
	6	0.718	1.440	1.943	2.447	3.707
	7	0.711	1.415	1.895	2.365	3.449
ν	8	0.706	1.397	1.860	2.306	3.355
	9	0.703	1.383	1.833	2.262	3.250
	10	0.700	1.372	1.812	2.228	3.169
	11	0.697	1.363	1.796	2.201	3.106
	12	0.695	1.356	1.782	2.179	3.055
	13	0.694	1.350	1.771	2.160	3.012
	14	0.692	1.345	1.761	2.145	2.977
	15	0.691	1.341	1.753	2.131	2.947
	16	0.690	1.337	1.746	2.120	2.921
	17	0.689	1.333	1.740	2.110	2.898
	18	0.688	1.330	1.734	2.101	2.878
	19	0.688	1.328	1.729	2.093	2.861
	20	0.687	1.325	1.725	2.086	2.845
	21	0.686	1.323	1.721	2.080	2.831
	22	0.686	1.321	1.717	2.074	2.819
	23	0.685	1.319	1.714	2.069	2.807
	24	0.685	1.318	1.711	2.064	2.797
	25	0.684	1.316	1.708	2.060	2.787
	26	0.684	1.315	1.706	2.056	2.779
	27	0.684	1.314	1.703	2.052	2.771
	28	0.683	1.313	1.701	2.048	2.763
	29	0.683	1.311	1.699	2.045	2.756
	⋮	⋮	⋮	⋮	⋮	⋮
	∞	0.674	1.282	1.645	1.960	2.576

First, we need to find p. If we are looking for the 95% confidence interval, we are looking for the value of p which satisfies the equation

$$\begin{aligned} 100(1-p)\% &= 95\%, \quad \text{i.e.} \\ p &= 0.05. \end{aligned}$$

We would look up the value in t tables in the p column, or in this case the 5% column.

We also need to know which row to look in. The rows are given as the degrees of freedom, ν, where $\nu = n - 1$. Hence, if our sample was of size $n = 10$ and we were looking for the 95% confidence interval, we would look in the $\nu = 9$ row and the $p = 5\%$ column to give us a value of **2.262** to use in our calculation.

Example 6.2

A credit card company wants to determine the mean income of its card holders. It also wants to find out if there are any differences in mean income between males and females. A random sample of 225 male card holders and 190 female card holders was drawn, and the following results obtained:

	Mean	Standard deviation
Males	£16 450	£3675
Females	£13 220	£3050

Calculate 95% confidence intervals for the mean income for males and females. Is there any evidence to suggest that, on average, males' and females' incomes differ? If so, describe this difference.

✎ *...95% confidence interval for male income...*

95% confidence interval for female income

Again, the true population variance, σ^2, is unknown, so we can't use the approach of section 7.2.1, and so again we use the t–distribution:

$$\bar{x} \pm t_p \times \sqrt{s^2/n}.$$

Now,

$$\begin{aligned} \bar{x} &= 13220, \\ s^2 &= 3050^2 \\ &= 9302500, \qquad \text{and} \\ n &= 190. \end{aligned}$$

Again, since the sample size is large, we use the ∞ row of the t tables to obtain the value of t_p, and so the 95% confidence interval for μ is found as

$$\begin{aligned} 13220 &\pm 1.96 \times \sqrt{9302500/190}, \qquad \text{i.e.} \\ 13220 &\pm 1.96 \times 221.27, \qquad \text{i.e.} \\ 13220 &\pm 433.69. \end{aligned}$$

So, the 95% confidence interval is (£12786.31, £13653.69).

Since the 95% confidence intervals for males and females *do not overlap*, there *is* evidence to suggest that males' and females' incomes, on average, are different. Further, it appears that male card holders earn more than women.

6.2.3 Confidence intervals: a general approach

In this section, we summarise the general procedure for calculating a confidence interval for the population mean μ.

Case 1: Known population variance σ^2

(i) Look out for "...the population variance/standard deviation is...", "the process variance/standard deviation is...", "$\sigma^2 = ...$"

(ii) Calculate the sample mean $\bar{x}$ from the data;

(iii) Calculate your interval! For example,

$$\bar{x} \pm z \times \sqrt{\sigma^2/n},$$

where $z = 1.96$ for a 95% confidence interval. For a 90%/99% confidence interval, $z = 1.64/2.58$ respectively.

Case 2: Unknown population variance σ^2

(i) Look out for "...the sample variance/standard deviation is...", "$s^2 = ...$"

(ii) Calculate the sample mean $\bar{x}$ *and* the sample variance s^2 from the data;

(iii) For a $100(1-p)\%$ confidence interval, look up the value of t under column p, row ν of the t tables, remembering that $\nu = n-1$. Note that, for a 90% confidence interval, $p = 10\%$, for a 95% confidence interval, $p = 5\%$ and for a 99% confidence interval, $p = 1\%$;

(iv) Calculate your interval, using

$$\bar{x} \pm t_p \times \sqrt{s^2/n}.$$

6.3 Hypothesis tests for one mean

We have seen that confidence intervals can be used to make inferences about population parameters. Sometimes, you may be asked to assess whether or not a parameter takes a specific value. For example, whether the population mean $\mu = 5$. One way of re-expressing this question is to ask whether the parameter value is plausible in light of the data. A simple check to see whether the value is contained in a 95% confidence interval will provide an answer. An alternative method, called a *hypothesis test*, is available. It is used extensively in reporting experimental results.

A hypothesis test is a rule for establishing whether or not a set of data is consistent with a hypothesis about a parameter of interest. The *null hypothesis* is a statement that a parameter has a certain value, and is usually written as H_0. For example, $H_0 : \mu = 2.7$, or $H_0 : \sigma^2 = 12.4$. If the null hypothesis is not true, what alternatives are there? Usually, the *alternative hypothesis* is written as H_1. Examples include $H_1 : \mu \neq 2.7$ and $H_1 : \sigma^2 \neq 12.4$.

Based on the information we have in our sample, we'd like to go with either the null hypothesis or the alternative hypothesis. We might use our sample as evidence to suggest that, for example, the population mean could well be equal to 2.7; alternatively, the sample might give evidence to the contrary and suggest that the population mean is not equal to 2.7, or for that matter the sample might suggest that the population mean is less than 2.7!

Suppose you are going on holiday to Sicily in March; a friend tells you that in March, Sicily has an average of 10 hours sunshine a day. On the first three days of your holiday there are 7, 8 and 9 hours of sunshine respectively. You consider that this is evidence that your friend is wrong. Thus, the null hypothesis would state that the average sunshine hours per day is 10 (as suggested by your friend), and your alternative hypothesis might state that the average sunshine hours per day *is less than* 10. You might be tempted to go with the alternative hypothesis. However, this sample of three days could be a fluke result – you might have chosen the most miserable period in March for years for your holiday. *Your test results are not conclusive; they only give you evidence for or against a particular belief.*

6.3.1 General methodology for hypothesis testing

All hypothesis tests follow the same basic methodology, although the actual calculations may vary depending on the data available.

1. **State the null hypothesis (H_0)**
   We use a hypothesis test to throw light on whether or not this statement is true. For example, you might ask "is the population mean equal to 10?", or "are the two population means equal?" (see next week for this!); such hypotheses are

expressed in the following way:

$$\begin{aligned} H_0 &: \mu = 10, && \text{and} \\ H_0 &: \mu_1 = \mu_2; && \text{or maybe} \\ H_0 &: \mu = c, \end{aligned}$$

Where c could be any constant.

2. **State the alternative hypothesis (H_1)**
   This is the conclusion to be reached if the null hypothesis is rejected. For example, "the population mean does not equal 10", or even "the population mean is less than 10"; to test for two different populations, we might say "the two population means are different" (again, see next weekfor this!). Remember, we can never reject the null hypothesis with certainty; the most we can say is that there is evidence against the null hypothesis, and so evidence in favour of the alternative. Such alternative hypotheses are expressed in the following way:

$$\begin{aligned} H_1 &: \mu \neq 10 && \text{or maybe} \\ H_1 &: \mu < 10. \end{aligned}$$

   To test for two different populations (see next week!), we might use

$$H_1 \; : \; \mu_1 \neq \mu_2.$$

3. **Calculate the test statistic**
   The value calculated from the sample which is used to perform the test is called the *test statistic*. It usually has a similar nature to the population value mentioned in the null hypothesis.

4. **Find the p–value of the test**
   The probability that such an extreme test statistic occurs, *assuming that H_0 is true*, is called the *p–value*, and can be found by comparing the test statistic to values from statistical tables (the tables used will depend on the nature of the test) or using a computer package such as `Minitab` (see the next computer practical).

5. **Reach a conclusion**
   A small p–value suggests that our test statistic is unlikely to occur if H_0 is true, and so we reject H_0 in favour of the alternative H_1. But what constitutes a "small" p–value? One commonly used yardstick, or *significance level*, is 5%, or 0.05, though others can be used. The *smaller* the p–value, the *more* evidence there is to reject the null hypothesis; conversely, the *larger* the p–value, the *less* evidence we have to reject H_0 and so in this case we are more likely to retain the null hypothesis. Table 6.2 gives some guidelines on how to interpret your p–value.

The above five steps are universal to all hypothesis testing.

p–value	Interpretation
p is bigger than 10%	no evidence against the null hypothesis: stick with H_0
p lies between 5% and 10%	*slight* evidence against H_0, but not enough to reject it
p lies between 1% and 5%	moderate evidence against H_0: reject it, and go with H_1
p is smaller than 1%	strong evidence against H_0: reject it, and go with H_1

Table 6.2: Conventional interpretation of p–values

6.3.2 Case 1: Known population variance σ^2: Example

A chain of shops believes that the average size of transactions is £130, and the population variance is known to be £900. The takings of one branch were analysed and it was found that the mean transaction size was £123 over the 100 transactions in one day. Based on this sample, test the null hypothesis that the true mean is equal to £130.

Since σ^2 is known (we are given that $\sigma^2 = 900$), this corresponds to case 1: population variance known (think back to confidence intervals). We now proceed with the five steps outlined in the previous section.

Steps 1 and 2 *(hypotheses)*
Here, we state our null and alternative hypotheses. The null hypothesis is given in the question – i.e.

$$H_0 \;:\; \mu = £130.$$

We could test against a general alternative, i.e.

$$H_1 \;:\; \mu \neq £130.$$

Step 3 *(calculating the test statistic)*
When σ^2 is known, we use following test statistic

$$\begin{aligned} z &= \frac{|\bar{x} - \mu|}{\sqrt{\sigma^2/n}}, \qquad \text{i.e.} \\ z &= \frac{|123 - 130|}{\sqrt{900/100}} \\ &= \frac{7}{\sqrt{9}} \\ &= 2.33. \end{aligned}$$

Step 4 *(finding the p–value)*
Recall the Central Limit Theorem from Section 6.1; this tells us that the quantity

$$\frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}}$$

follows a standard Normal distribution. Thus, the value we obtain from our test statistic formula above will be from the positive half of the standard Normal distribution. We can therefore compare our test statistic to critical values from the standard Normal distribution to find our p–value, or at least a range for our p–value. Remember, this is the probability of observing our data, or anything more extreme than this, if the null hypothesis is true; thus, the smaller the p–value, the more evidence there is *against* H_0.

Our alternative hypothesis is *two–tailed* (i.e. $\neq$ rather than $<$ or $>$), and so our values are (see table overleaf):

Significance level	10%	5%	1%
Critical value	1.645	1.96	2.576

Our test statistic $z = 2.33$ lies between the critical values of 1.96 and 2.576, and so our p–value lies between 1% and 5%. We can see this more clearly on a diagram:

Step 5 *(conclusion)*
Using Table 6.2 to interpret our p–value, we see that there is moderate evidence against H_0. Thus, we should reject H_0 in favour of the alternative hypothesis H_1; it appears that the population mean transaction size *is not equal to* £130.

Tables for hypothesis tests

Tabulated values of z for which $\Pr(Z > z) = p$, where Z has a standard normal distribution:

One–tailed test	10%	5%	2.5%	1%	0.5%
Two–tailed test	20%	10%	5%	2%	1%
Critical value	1.282	1.645	1.96	2.326	2.576

Tabulated values of t for which $\Pr(|T| > t) = p$, where T has a t–distribution with ν degrees of freedom:

	One–tailed test	10%	5%	2.5%	1%	0.5%
	Two–tailed test	20%	10%	5%	2%	1%
	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.449
ν	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	⋮	⋮	⋮	⋮	⋮	⋮
	∞	1.282	1.645	1.960	2.326	2.576

Alternatively, since our sample mean $\bar{x} = £123$ is smaller than the proposed value of £130, we could have set up a *one–tailed* alternative hypothesis in step 2, i.e. we could have tested

$$H_0 \; : \; \mu = £130 \qquad \text{against}$$
$$H_1 \; : \; \mu < £130.$$

This is now a one–tailed test and the critical values from the table on the previous page are:

Significance level	10%	5%	1%
Critical value	1.282	1.645	2.326

The test statistic is (as before) 2.33, which now lies "to the right" of the last critical value in the table (2.326). Thus, our p–value is now smaller than 1%, and so, using table 6.2, we see that in this more specific test there is *strong* evidence against H_0. Again, this can be seen more clearly with a diagram:

Notice that this one–tailed test is more specific than the two–tailed test previously carried out. If you're not sure whether you should perform a one–tailed test or a two–tailed test, (i.e. there might not be much difference between the proposed mean and the sample mean), it's usually safer to test against the more general two–tailed alternative.

6.3.3 Case 2: Unknown population variance σ^2: Example

The batteries for a fire alarm system are required to last for 20000 hours before they need replacing. 16 batteries were tested; they were found to have an average life of 19500 hours and a standard deviation of 1200 hours. Perform a hypothesis test to see if the batteries do, on average, last for 20000 hours.

Steps 1 and 2 *(hypotheses)*
Using a one–tailed test, our null and alternative hypotheses are:

$$\begin{aligned} H_0 &: \mu = 20000 \qquad \text{versus} \\ H_1 &: \mu < 20000. \end{aligned}$$

We use a one–tailed test because we are interested in whether the batteries are effective or not; there is no problem if they last longer than 20000 hours.

Step 3 *(calculating the test statistic)*
Unlike the previous example, the population variance σ^2 is unknown (i.e. the question does not say "the population variance is ...", or "the population standard deviation is ...", for example). However, the *sample* standard deviation is given, based on a sample of size 16, and so we need to use a slightly different test statistic. In fact, we do what we did when we were constructing confidence intervals – i.e. we replace σ^2 with s^2 and then use tables of values from Student's t distribution instead of the standard Normal distribution. Thus, the test statistic is given by

$$\begin{aligned} t &= \frac{|\bar{x} - \mu|}{\sqrt{s^2/n}} \\ &= \frac{|19500 - 20000|}{\sqrt{1200^2/16}} \\ &= \frac{500}{\sqrt{1440000/16}} \\ &= 1.667. \end{aligned}$$

Step 4 *(finding the p–value)*
Since σ^2 is unknown, we use t–distribution tables to obtain a range for our p–value. The degrees of freedom, $\nu = n - 1 = 16 - 1 = 15$, and under a one–tailed test this gives the following critical values:

Significance level	10%	5%	1%
Critical value	1.341	1.753	2.602

Our test statistic of $t = 1.667$ lies between the critical values of 1.341 and 1.753, and so the corresponding p–value lies between 5% and 10%.

Step 5 *(conclusion)*
Using table 6.2 to interpret our p–value, we see that there is only *slight* evidence against the null hypothesis and certainly not enough grounds to reject it, so we retain H_0. There is insufficient evidence to suggest there is a problem with these batteries.

6.4 Testing two means

Recall that, in the test for one mean, there were two cases: population variance (σ^2) known and population variance unknown. Similarly, when comparing two means, we can consider a test where both population variances are known and both are unknown.

6.4.1 Both population variances (σ_1^2 and σ_2^2) known

1. **State the null hypothesis**
   This time, the null hypothesis is
   $$H_0 \quad : \quad \mu_1 = \mu_2,$$
   i.e. the two population means are equal.

2. **State the alternative hypothesis**
   We usually test against the (two–tailed) alternative:
   $$H_1 \quad : \quad \mu_1 \neq \mu_2,$$
   i.e. the population means *are not* equal. However, if we have reason to believe that one population mean is larger (or smaller!) than the other, we might want to use the (one–tailed) alternatives:
   $$\begin{aligned} H_1 \quad &: \quad \mu_1 > \mu_2, \qquad \text{or} \\ H_1 \quad &: \quad \mu_1 < \mu_2. \end{aligned}$$

3. **Calculate the test statistic**
   The test statistic for a two–sample test (when both population variances are known) is
   $$z \quad = \quad \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$
   where $\bar{x}_1$, $\bar{x}_2$, n_1 and n_2 are the means and sample sizes of samples 1 and 2 (respectively), and σ_1^2 and σ_2^2 are the corresponding population variances.

4. **Find the p–value of the test**
   As in the tests for one mean, we use statistical tables to obtain our p–value, or rather a *range* for our p–value. Since, in this case, both population variances are known, we refer to standard normal tables (page 156). As before, the critical values depending on whether we are testing against a one– or two–tailed alternative hypothesis.

5. **Reach a conclusion**
   Use table 6.2 to form a conclusion (i.e. retain or reject the null hypothesis), remembering to also word your conclusions in plain English and in the context of the question posed.

Example 6.3

Before a training session for call centre employees a sample of 50 calls to the call centre had an average duration of 5 minutes, whereas after the training session a sample of 45 calls had an average duration of 4.5 minutes. The population variance is known to have been 1.5 minutes before the course and 2 minutes afterwards. Has the course been effective?

6.4.2 σ_1^2 and σ_2^2 unknown

In the more likely situation where the population variances are unknown, we use the test statistic

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where s is a "pooled standard deviation", and is found as

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

Also, as with the test for one mean when the population variance was unknown, we need to use t–tables (page 156) to obtain our p–value for the test. The degrees of freedom is now found as $\nu = n_1 + n_2 - 2$. Apart from these changes, the hypothesis test follows the same format as that for which both population variances are known.

Example 6.4

A company is interested in knowing if two branches have the same level of average transactions. The company sample a small number of transactions and calculates the following statistics:

Shop 1	$\bar{x}_1 = 130$	$s_1^2 = 700$	$n_1 = 12$
Shop 2	$\bar{x}_2 = 120$	$s_2^2 = 800$	$n_2 = 15$

Test whether or not the two branches have (on average) the same level of transactions.

Steps 1 and 2 *(hypotheses)*
Our null and alternative hypotheses are:

$$H_0 : \mu_1 = \mu_2 \quad \text{versus}$$
$$H_1 : \mu_1 \neq \mu_2.$$

Step 3 *(calculating the test statistic)*
Since both population variances are unknown (only the *sample* values are given), the test statistic is

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}};$$

thus, we first need to obtain the pooled standard deviation s.

... calculation of pooled standard deviation and test statistic in Example 6.4...

Step 4 *(finding the p–value)*
Since both population variances are unknown, we use t–tables to obtain our critical value. The degrees of freedom, $\nu = n_1 + n_2 - 2$, i.e. $\nu = 12 + 15 - 2 = 25$. Under a two–tailed test, and using t–tables on page 156, we get the following critical values:

Significance level	10%	5%	1%
Critical value	1.708	2.060	2.787

Our test statistic $t = 0.939$ lies to the left of the first critical value, and so our p–value is bigger than 10%.

Step 5 *(conclusion)*
Using table 6.2, we see that, since our p–value is larger than 10%, we have no evidence to reject the null hypothesis. Thus, we retain H_0 and conclude that there is no significant difference between the average level of transactions at the two shops.

6.5 Chapter 6 practice questions

1. A company packs sacks of flour. The variance of the filling process is 100g. A sample of 50 bags is taken and weighed and the resulting sample mean is 750g. Compute a 95% and 99% confidence interval for the mean weight of a bag of flour.

2. A company manufactures bolts. A sample of 100 bolts is taken and measured and their average length is calculated as 98mm; the variance is found to be 50mm. What is the 95% confidence interval for the mean length of bolts? If the bolts are designed to be 100mm long, is the process satisfactory?

3. A sample of 12 students is taken and their mean IQ calculated as 110. The sample variance is 220. What is the 95% and 99% confidence intervals for the population value based on this sample? What do you notice about the calculated interval as the confidence level increases? Do either of these two confidence intervals contain the known population mean IQ of 100?

4. A machine for filling cans of Coke has a process variance of 400ml. A sample of 100 cans is taken and it is found that the average contents are 240ml. Is this consistent with the cans containing the stated weight of 250ml? Use a 95% confidence interval to help you here.

5. Return to the scenario in question 4. Perform an appropriate hypothesis test to see if the data are consistent with the cans containing the stated volume of 250ml. Use a two–tailed test.

6. A chain of record shops believes that its Northumberland Street store (Shop 1) is more successful than its Metro Centre branch (Shop 2). The management take a random sample of daily takings and obtains the following summary statistics:

Shop 1	$\bar{x}_1 = £15000$	$s_1^2 = 100,000$	$n_1 = 18$
Shop 2	$\bar{x}_2 = £14500$	$s_2^2 = 95,000$	$n_2 = 12$

Is the management's belief correct?

7. (a) An airline company, EasyAir, advertises 5 hour flights from Glasgow to Cairo. Another company flying the same route, RyanJet, suspects EasyAir of false advertising, and samples 20 of EasyAir's flights between Glasgow and Cairo. The mean flight time from this sample is 5 hours and 10 minutes with a standard deviation of 20 minutes. Are RyanJet's suspicions supported?

(b) RyanJet claim that their Glasgow to Cairo flights are, on average, faster than EasyAir's flights. A sample of 23 RyanJet flights between Glasgow and Cairo has a mean flight time of 5 hours and 4 minutes with a standard deviation of 22 minutes. Are RyanJet's flights on this route shorter than EasyAir's?

8. The mean age of students attending a college is 19.5 years. The ages of a sample of 7 students who have been offered college accommodation are given below. Test the hypothesis that younger students are favoured.

17.9 18.2 19.1 20.3 17.8 17.4 17.8

9. During the streaking phase of the 1970's, a psychological test designed to determine extroversion was applied to a group of 19 admitted male streakers and a control group of 19 male non–streakers. The results were

streakers	non–streakers
$\bar{x}_1 = 15.26$	$\bar{x}_2 = 13.90$
$s_1^2 = 2.62$	$s_2^2 = 4.11$

Are streakers more extrovert than non–streakers?

10. In the comparison of two kinds of paint, a consumer testing service finds that eight one–gallon cans of "Wilko's Best", and ten one–gallon cans of "Dulor", cover the following areas (in square feet):

"Wilko's Best"	542	546	550	548	540	537	548	553		
"Dulor"	521	532	498	512	551	540	500	513	520	523

The population standard deviation coverage for "Wilko's Best" and "Dulor" is 31 square feet and 26 square feet respectively. Perform a test to find out whether or not there is any difference in mean coverage between the two brands.