

Chapter 7

Inference for a Mean or Median

7.1 Introduction

There are many situations when we might wish to make inferences about the location of the “center” of the population distribution of a quantitative variable. We will consider methods for making inferences about a population mean or a population median, which are the two most commonly used measures of the center of the distribution of a quantitative variable, in this chapter.

In Chapter 5 we considered inferences about the distribution of a dichotomous variable. Since the distribution of a dichotomous variable is completely determined by the corresponding population success proportion, we found that the sampling distribution of the sample proportion \hat{p} was determined by the sampling method. In general, the distribution of a quantitative variable is not completely determined by a single parameter. Therefore, before we can make inferences about the distribution of a quantitative variable we need to make some assumptions about the distribution of the variable.

A probability model for the distribution of a discrete variable X is a theoretical relative frequency distribution which specifies the probabilities (theoretical relative frequencies) with which each of the possible values of X will occur. In contrast to a relative frequency distribution, which indicates the relative frequencies with which the possible values of X occur in a sample, a probability model or probability distribution specifies the probabilities with which the possible values of X will occur when we observe a single value of X . That is, if we imagine choosing a single value of X at random from all of the possible values of X , then the probability model specifies the probability with which each possible value will be observed. We can represent a discrete probability distribution graphically via a probability histogram (theoretical relative frequency histogram) which is simply a histogram based on the probabilities specified by the probability model.

A probability model for the distribution of a continuous variable X can be represented by a density curve. A **density curve** is a nonnegative curve for which the area under the curve (over the x -axis) is one. We can think of the density curve as a smooth version of a probability histogram with the rectangles of the histogram replaced by a smooth curve indicating where the tops of the rectangles would be. With a continuous variable X it does not make sense to talk about the probability that X would take on a particular value, after all if we defined positive probabilities for the infinite collection (continuum) of possible values of X these probabilities could not add up to one. It does, however, make sense to talk about the probability that X will take on a value in a specified interval or range of values. Given two constants $a < b$ the probability that X takes on a value in the

interval from a to b , denoted by $P(a \leq X \leq b)$, is equal to the area under the density curve over the interval from a to b on the x -axis. Areas of this sort based on the density curve give the probabilities which a single value of X , chosen at random from the infinite population of possible values of X , will satisfy.

Given a probability model for the distribution of a continuous variable X , *i.e.*, given a density curve for the distribution of the continuous variable X , we can define population parameters which characterize relevant aspects of the distribution. For example, we can define the population mean μ as the balance point of the unit mass bounded by the density curve and the number line. We can also think of the population mean as the weighted average of the infinite collection of possible values of X with weights determined by the density curve. We can similarly define the population median M as the point on the number line where a vertical line would divide the area under the density curve into two equal areas (each of size one-half).

7.2 Inference for a population mean

7.2a Introduction

In this section we will consider inference for the mean μ of the population distribution of a continuous variable X . The basic problem we will consider is that of using a random sample of values of the continuous variable X to estimate the corresponding population mean or to test a hypothesis about this population mean.

Before we go further it is instructive to consider some situations where inference about a population mean could be used and the way in which we might interpret a population mean.

In some applications the population mean represents an actual physical constant. Let μ denote the true value of the physical constant (such as the speed of light) which we wish to estimate. Suppose that an experiment has been devised to produce a measurement X of the physical constant μ . A probability model for the distribution of X provides a model for the behavior of an observed value of X by specifying probabilities which X must satisfy. Thus, the probability model provides an explanation of the variability in X as an estimate of μ . It would be unreasonable to expect an observed value of X to be exactly equal to μ ; however, if the experiment is carefully planned and executed it would be reasonable to expect the average value of X based on a long series of replications of the experiment to be equal to μ . If this is the case, the population mean of the probability model for X will be equal to the physical constant μ and the standard deviation σ of the probability model will serve as a useful quantification of the variability of the measurement process.

When interest centers on an actual, physical population of units the population mean is the average value of the variable of interest corresponding to all of the units in the

population. Imagine a large population of units, *e.g.*, a population of humans or animals. Let the continuous variable X denote a characteristic of a unit, *e.g.*, X might be some measurement of the size of a unit. For concreteness, let X denote the height of an adult human male and consider the population of all adult human males in the United Kingdom. A probability model for the distribution of X provides a model for the behavior of an observed value of the height X of an adult male selected at random from this population. In this situation a probability model explains the variability among the heights of the adult males in this population. Let μ denote the population mean height of all adult males in the United Kingdom, *i.e.*, let μ be the average height we would get if we averaged the heights of all the adult males in this population. In this context the population mean height is obviously not the “true height” of each of the adult males; however, we can think of the height X of a particular adult male as being equal to the population mean height μ plus or minus an adjustment for this particular male which is due to hereditary, environmental, and other factors. The standard deviation σ of the probability model serves to quantify the variability among this population of heights.

In many applications interest centers on the population mean difference between two paired values. For example, consider a population of individuals with high cholesterol levels and a drug designed to reduce cholesterol levels. Let X_1 denote the cholesterol level of an individual before taking the drug, let X_2 denote this same individual’s cholesterol level after being treated with the drug, and let $D = X_1 - X_2$ denote the difference between the two cholesterol levels (the decrease in cholesterol level). A probability model for the distribution of D provides a model for the behavior of an observed value of the difference D for an individual selected at random from this population. In this situation a probability model explains the variability among the differences in cholesterol level due to treatment with the drug for the individuals in this population. The corresponding population mean difference μ is the average difference (decrease) in cholesterol level that we would observe if all of the individuals in this population were treated with this drug. The standard deviation σ of the probability model serves to quantify the variability among the differences for the individuals in this population.

We can envision a probability model for the distribution of a quantitative variable X in terms of a box model. If X has a finite number of possible values, then a probability model specifies the probabilities with which these possible values will occur. If the balls in a box are labeled with the possible values of X and the proportion of balls with each label (value of X) in the box is equal to the probability for that value specified by the probability model, then, according to the probability model, observing a value of X is equivalent to selecting a single ball at random from this box and observing the label on the ball. For a continuous variable X observing the value of X is like selecting a ball at random from a box containing an infinite collection of suitably labeled balls.

Given a probability model for the distribution of X , a collection of n values of X is said to form a random sample of size n if it satisfies the two properties given below.

1. Each value of the variable X that we observe can be viewed as a single value chosen at random from a (usually infinite) population of values which are distributed according to the specified probability model for the distribution of X .
2. The observed values of X are independent. That is, knowing the value of one or more of the observed values of X has no effect on the probabilities associated with other values of X .

In other words, in terms of the box model for the probability distribution of X a random sample of n values of X can be viewed as a collection of n labels corresponding to a simple random sample selected with replacement from a box of suitably labeled balls.

Given a random sample of values of X it seems obvious that the sample mean \bar{X} is an appropriate estimate of the corresponding population mean μ . The sampling distribution of \bar{X} , which describes the sample to sample variability in \bar{X} , serves as the starting point for our study of the behavior of the sample mean \bar{X} as an estimator of the population mean μ . The exact form of the sampling distribution of \bar{X} depends on the form of the distribution of X . However, the two important properties of the sampling distribution of the sample mean given below are valid regardless of the exact form of the distribution of X .

Let \bar{X} denote the sample mean of a random sample of size n from a population (distribution) with population mean μ and population standard deviation σ . The sampling distribution of the sample mean \bar{X} has the following characteristics.

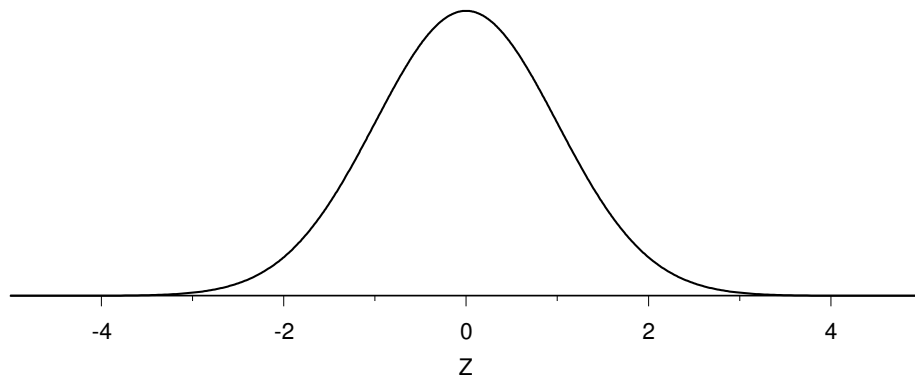
1. The mean of the sampling distribution of \bar{X} is the corresponding population mean μ . This indicates that the sample mean \bar{X} is unbiased as an estimator of the population mean μ . Recall that saying that a statistic is unbiased means that, even though the statistic will overestimate the parameter for some samples and will underestimate the parameter in other samples, it will do so in such a way that, in the long run, the values of the statistic will average to give the correct value of the parameter.
2. The population standard error of the sample mean \bar{X} (the standard deviation of the sampling distribution of \bar{X}) is $\text{S.E.}(\bar{X}) = \sigma/\sqrt{n}$. That is, the standard deviation of the sampling distribution of \bar{X} is equal to the standard deviation of the distribution of X divided by the square root of the sample size. Notice that this implies that the sample mean is less variable than a single observation as an estimator of μ ; and that if μ and σ are held constant, then the variability in \bar{X} as an estimator of μ decreases as the sample size increases reflecting the fact that a larger sample provides more information than a smaller sample.

Since the form of the sampling distribution of \bar{X} depends on the form of the distribution of X , we will need to make some assumptions about the distribution of X before we can proceed with our discussion of inference for the population mean μ . These assumptions correspond to the choice of a probability model (density curve) to represent the distribution of X . There is an infinite collection of probability models to choose from but we will restrict our attention to a single probability model, the normal probability model, which is appropriate for many situations when the distribution of X is symmetric and mound shaped. This does not imply that all, or even most, distributions of continuous variables are normal distributions. Some of the reasons that we will use the normal distribution as a probability model are: (1) the theory needed for inference has been worked out for the normal model; (2) there are many situations where a normal distribution provides a reasonable model for the distribution of a quantitative variable; (3) even though the inferential methods we discuss are based on the assumption that the distribution of the variable is exactly a normal distribution, it is known that these inferential methods actually perform reasonably well provided the true distribution of the variable is “reasonably similar to a normal distribution”; and, (4) it is often possible to transform or redefine a variable so that its distribution is reasonably modeled by a normal distribution.

7.2b The normal distribution

The normal distribution with mean μ and standard deviation σ can be characterized by its density curve. The density curve for the normal distribution with mean μ and standard deviation σ is the familiar bell shaped curve. The standard normal density curve, which has mean $\mu = 0$ and standard deviation $\sigma = 1$, is shown in Figure 1.

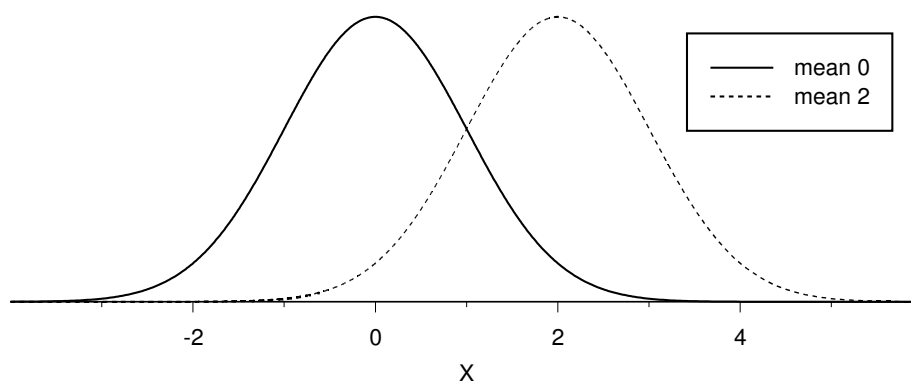
Figure 1. The standard normal density curve.



The normal distribution with mean μ and its density curve are symmetric around μ , *i.e.*, if we draw a vertical line through μ , then the two sides of the density curve are mirror images of each other. Therefore the mean of a normal distribution μ is also the median of the normal distribution. The mean μ locates the normal distribution on the number line so

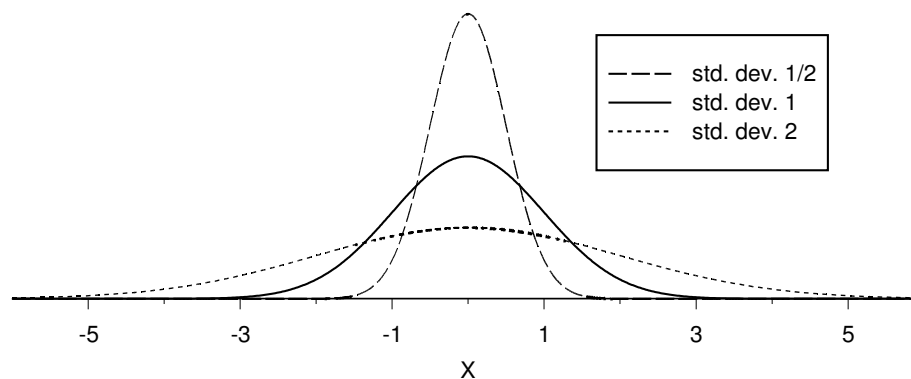
that if we hold σ constant and change the mean μ , the normal distribution is simply shifted along the number line until it is centered at the new mean. In other words, holding σ fixed and changing μ simply relocates the density curve on the number line; it has no effect on the shape of the curve. Figure 2 provides the density curves for normal distributions with respective means $\mu = 0$ and $\mu = 2$ and common standard deviation $\sigma = 1$.

Figure 2. Normal distributions with common standard deviation one and means of zero and two.



The standard deviation σ indicates the amount of variability in the normal distribution. If we hold μ fixed and increase the value of σ , then the normal density curve becomes flatter, while retaining its bell-shape, indicating that there is more variability in the distribution. Similarly, if we hold μ fixed and decrease the value of σ , then the normal density curve becomes more peaked around the mean μ , while retaining its bell-shape, indicating that there is less variability in the distribution. Normal distributions with mean $\mu = 0$ and respective standard deviations $\sigma = .5$, $\sigma = 1$, and $\sigma = 2$ are plotted in Figure 3.

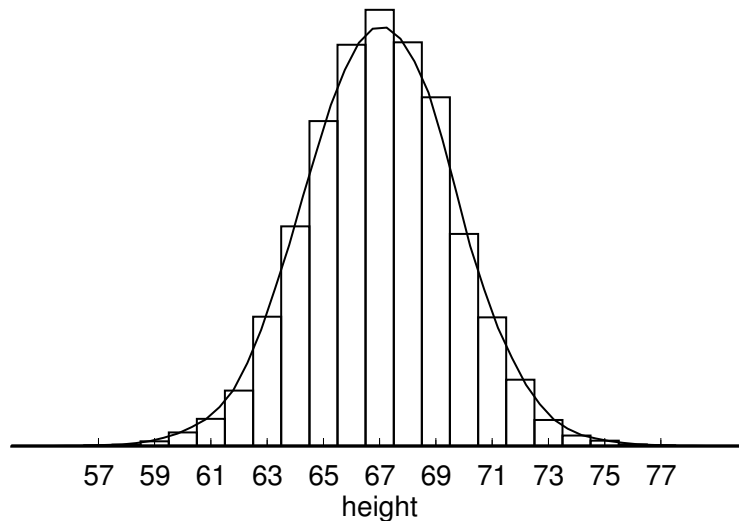
Figure 3. Normal distributions with common mean zero and standard deviations one-half, one, and two.



Example. Heights of adult males. The 8585 heights (in inches) of adult males born in the United Kingdom (including the whole of Ireland) which are summarized in

Table 8 of Section 3.3 provide a good illustration of the fact that normal distributions often provide very good models for populations of physical measurements, such as heights or weights, of individuals. Figure 4 provides a histogram for this height distribution and the density curve for a normal distribution chosen to model these data. You can see that the normal distribution provides a very reasonable model for the heights of adult males born in the United Kingdom.

Figure 4. Histogram and normal density curve for the UK height example.



Computer programs and many calculators can be used to compute normal probabilities or equivalently to compute areas under the normal density curve. These probabilities can also be calculated using tables of standard normal distribution probabilities such as Table 1. Recall that the standard normal distribution is the normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The relationship between the standard normal variable Z and the normal variable X , which has mean μ and standard deviation σ , is

$$Z = \frac{X - \mu}{\sigma} \text{ or equivalently } X = \mu + Z\sigma.$$

This relationship implies that a probability statement about the normal variable X can be re-expressed as a probability statement about the standard normal variable Z by re-expressing the statement in terms of standard deviation units from the mean. Given two constants $a < b$, observing a value of X between a and b (observing $a \leq X \leq b$) is equivalent to observing a value of $Z = (X - \mu)/\sigma$ between $(a - \mu)/\sigma$ and $(b - \mu)/\sigma$ (observing $(a - \mu)/\sigma \leq (X - \mu)/\sigma \leq (b - \mu)/\sigma$). Furthermore, $Z = (X - \mu)/\sigma$ behaves in accordance with the standard normal distribution so that the probability of observing a value of X between a and b , denoted by $P(a \leq X \leq b)$, is equal to the probability that the standard normal variable Z takes on a value between $(a - \mu)/\sigma$ and $(b - \mu)/\sigma$, *i.e.*,

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right).$$

In terms of areas this probability equality says that the area under the normal density curve with mean μ and standard deviation σ over the interval from a to b is equal to the area under the standard normal density curve over the interval from $(a - \mu)/\sigma$ to $(b - \mu)/\sigma$. Similarly, given constants $c < d$, we have the analogous result that

$$P(c \leq Z \leq d) = P(\mu + c\sigma \leq X \leq \mu + d\sigma).$$

Table 1 provides cumulative standard normal probabilities of the form $P(Z \leq a)$ for values of a (Z in the table) between 0 and 3.69. Computer programs usually produce cumulative probabilities like these. To use these cumulative probabilities to compute a probability of the form $P(a \leq Z \leq b)$ note that

$$P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$$

and note that the symmetry of the normal distribution implies that

$$P(Z \leq -a) = P(Z \geq a) = 1 - P(Z \leq a).$$

Calculators will usually provide probabilities of the form $P(a \leq Z \leq b)$ directly.

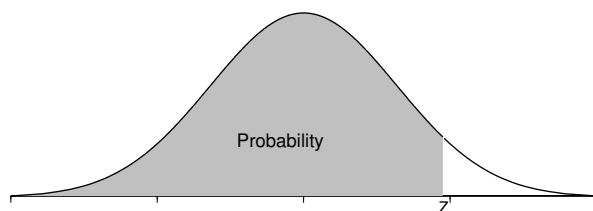


Table 1. Cumulative normal probabilities.
(Areas under the standard normal curve to the left of Z .)

Z	Second decimal place in Z									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389

continued on next page

7.2c Sampling from a normal population

Strictly speaking, the inferential methods based on the Student's t distribution described in Sections 7.2d and 7.2e are only appropriate when the data constitute a random sample from a normal population. However, these methods are known to be generally reasonable even when the underlying population is not exactly a normal population, provided the underlying population distribution is reasonably symmetric and the true density curve has a more or less normal (bell-shaped) appearance. We cannot be sure that an underlying population is normal; however, we can use descriptive methods to look for evidence of possible nonnormality, provided the sample size is reasonably large. The most easily detected and serious evidence of nonnormality you should look for is evidence of extreme skewness or evidence of extreme outlying observations. If there is evidence of extreme skewness or extreme outlying observations, then the inferential methods based on the Student's t distribution should not be used. An alternate approach to inference (for a population median) which may be used when the Student's t methods are inappropriate is discussed in Section 7.3.

Figure 5. Stem and leaf histograms for eight random samples of size 10 from a standard normal distribution.

In these stem and leaf histograms the decimal point is between the stem and the leaves.

(A)	(B)	(C)	(D)
-2 41	-2 4	-2 1	-2
-1	-1 3	-1 55	-1 20
-0 20	-0 51	-0 853	-0 6442
0 345569	0 33	0 28	0 145
1	1 112	1 14	1 4
2	2 0	2	2
(E)	(F)	(G)	(H)
-2 62	-2	-2	-2
-1	-1 3	-1	-1 63
-0 42	-0 51	-0 721	-0 77654
0 66	0 11234	0 69	0 68
1 34	1 39	1 00357	1 2
2 24	2	2	2

Table 2. Five number summaries for the eight random samples of size 10 from Figure 5.

	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
min	-2.38	-2.38	-2.14	-1.21	-2.57	-1.29	-.65	-1.62
Q_1	-.13	-.37	-1.29	-.58	-.38	-.08	.08	-.71
median	.34	.30	-.43	-.28	.57	.14	.95	-.56
Q_3	.54	1.14	.67	.32	1.40	.33	1.25	.36
max	.89	2.01	1.35	1.44	2.40	1.93	1.66	1.21

Using a sample to determine whether the underlying population is normal requires some practice. Some indication of the sorts of samples which may arise when the underlying population is normal is provided by the stem and leaf histograms and five number summaries given in Figures 5 and 6 and Tables 2 and 3 for several computer generated random samples from a standard normal distribution. The eight random samples of size 10 of Figure 5 and Table 2 indicate what may happen when a small sample is taken from a population which is normal. Based on these examples it is clear that we should not necessarily view slight skewness (as in A, B, D, F, G, and H) or mild outliers (as in A and E) as evidence of nonnormality. The eight random samples of size 50 of Figure 6 and Table 3 indicate that with a reasonably large sample we can expect to see a reasonably symmetric distribution; but, we may see a few mild outliers as in A, C, and F.

Figure 6. Stem and leaf histograms for eight random samples of size 50 from a standard normal distribution.

In these stem and leaf histograms the decimal point is between the stem and the leaves.

(A)	(B)	(C)	(D)
-3	-3	-3 2	-3
-2	-2	-2	-2 86
-2	-2 10	-2 21	-2 10
-1 85	-1 6	-1 66	-1 998
-1 4332110	-1 3100	-1 4322110	-1 4443222
-0 9876665	-0 8877655	-0 97655	-0 9876655
-0 4442211110	-0 42221000	-0 443221111	-0 4421111
0 113344	0 011222333	0 03344	0 011122334
0 555667778	0 55556689	0 566677789	0 566789
1 0134	1 01223	1 011224	1 022
1 5778	1 579	1 55	1 558
2	2 01	2 00	2 1
2 7	2 7	2	2
(E)	(F)	(G)	(H)
-2	-2 5	-2	-2
-2	-2	-2 0	-2
-1	-1	-1 7666	-1 9666
-1 4320	-1 333111	-1 222110	-1 33220
-0 9875555555	-0 98876655	-0 98766555	-0 87666
-0 32	-0 443333211	-0 4443221	-0 433322211000
0 01223333444	0 0233	0 122223	0 222233334
0 556678889999	0 566666777789	0 6678899	0 5578889
1 0002344	1 022444	1 011	1 11244
1 5567	1 6	1 555699	1 56
2	2 124	2 14	2 4

Table 3. Five number summaries for the eight random samples of size 50 from Figure 6.

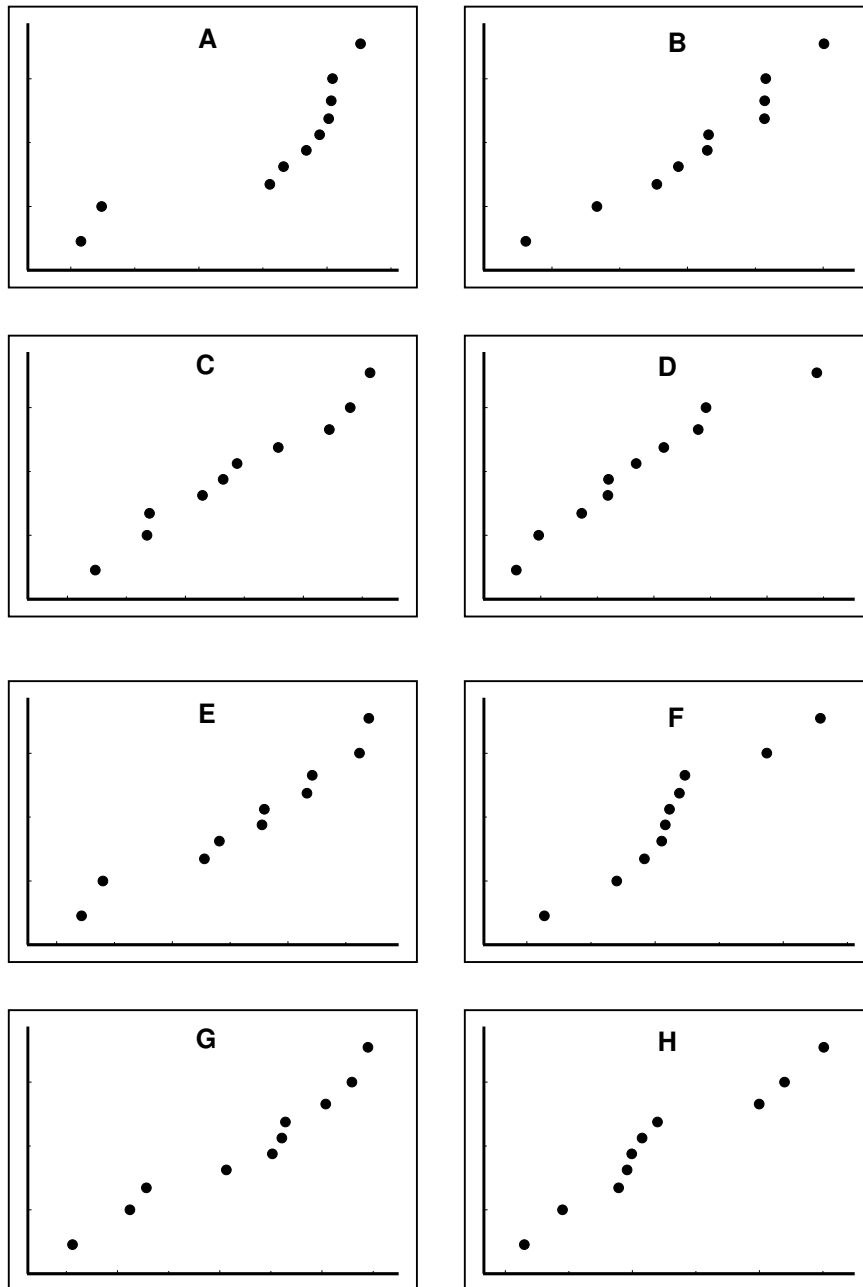
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
min	-1.77	-2.13	-3.20	-2.84	-1.39	-2.47	-1.98	-1.90
Q_1	-.64	-.55	-.91	-1.23	-.48	-.56	-.78	-.59
median	-.06	.11	-.09	-.11	.36	.10	-.13	-.02
Q_3	.67	.78	.72	.54	.92	.71	.87	.67
max	2.75	2.70	1.97	2.10	1.74	2.37	2.44	2.37

Normal probability plots. If you have access to a suitable calculator or computer program, you can construct a normal probability plot as an aid for assessing whether your data are consistent with a sample from a normal distribution. A normal probability plot provides a graphical comparison of the observed sample values with idealized sample values which we might expect to observe if the sample came from a normal distribution. The idealized sample values used in a normal probability plot are known as normal scores. Ideally, we would expect a random sample of size n from a normal distribution to partition the region under the normal density curve into $n + 1$ regions of equal area, with each of these areas being $1/(n + 1)$. The n normal scores, which constitute our idealized random sample, can thus be formed by determining the n values which would partition the area under the normal density curve into $n + 1$ regions each of area $1/(n + 1)$ as suggested above. Once these normal scores are obtained we can plot the ordered normal scores versus the ordered observed data values and examine this normal probability plot looking for evidence of systematic disagreement between the actual sample values and the expected normal scores. If the sample really was a sample from a normal distribution, then we would expect the normal probability plot to approximate a straight line. Therefore, a normal probability plot which differs greatly in appearance from a straight line provides evidence that the sample may not come from a normal distribution.

Figures 7 and 8 provide normal probability plots (normal score versus observed value) for the computer generated random samples from a standard normal distribution of Figures 5 and 6. Some representative examples of normal probability plots for actual data are provided in Figures 9, 10, and 11. (In some of these plots and subsequent normal probability plots the points have been subjected to small random shifts to better indicate points which are coincident or very close together.)

The normal probability plot in Figure 9 is for the height of adult males in the United Kingdom example of Section 7.2b. We noted that the histogram for the distribution of the height of adult males in the United Kingdom given in Figure 4 of this chapter is very well approximated by a normal distribution. The straight line nature of the normal probability plot of Figure 9 indicates that it is quite reasonable to model these heights as forming a random sample from a normal distribution. The normal probability plot in Figure 10 is for the cholesterol levels of the rural Guatemalans from the example in Section 3.1.

Figure 7. Normal probability plots for the for eight random samples of size 10 (from a standard normal distribution) of Figure 5.



The stem and leaf histogram of Figure 1 in Section 3.1 is reasonably symmetric and the fact that the normal probability plot in Figure 10 is reasonably linear indicates that it is reasonable to model the cholesterol levels of these rural Guatemalans as forming a sample from a normal distribution. The normal probability plot in Figure 11 is for the rainfall amounts for the 26 days when the cloud was unseeded in the cloud seeding example of

Section 4.3. The curvature in this normal probability plot indicates that it is not reasonable to model these rainfall amounts as forming a random sample from a normal distribution. The stem and leaf histogram for this example given in Figure 12 shows that this type of curvature in a normal probability plot corresponds to skewness to the right.

Figure 8. Normal probability plots for the for eight random samples of size 50 (from a standard normal distribution) of Figure 6.

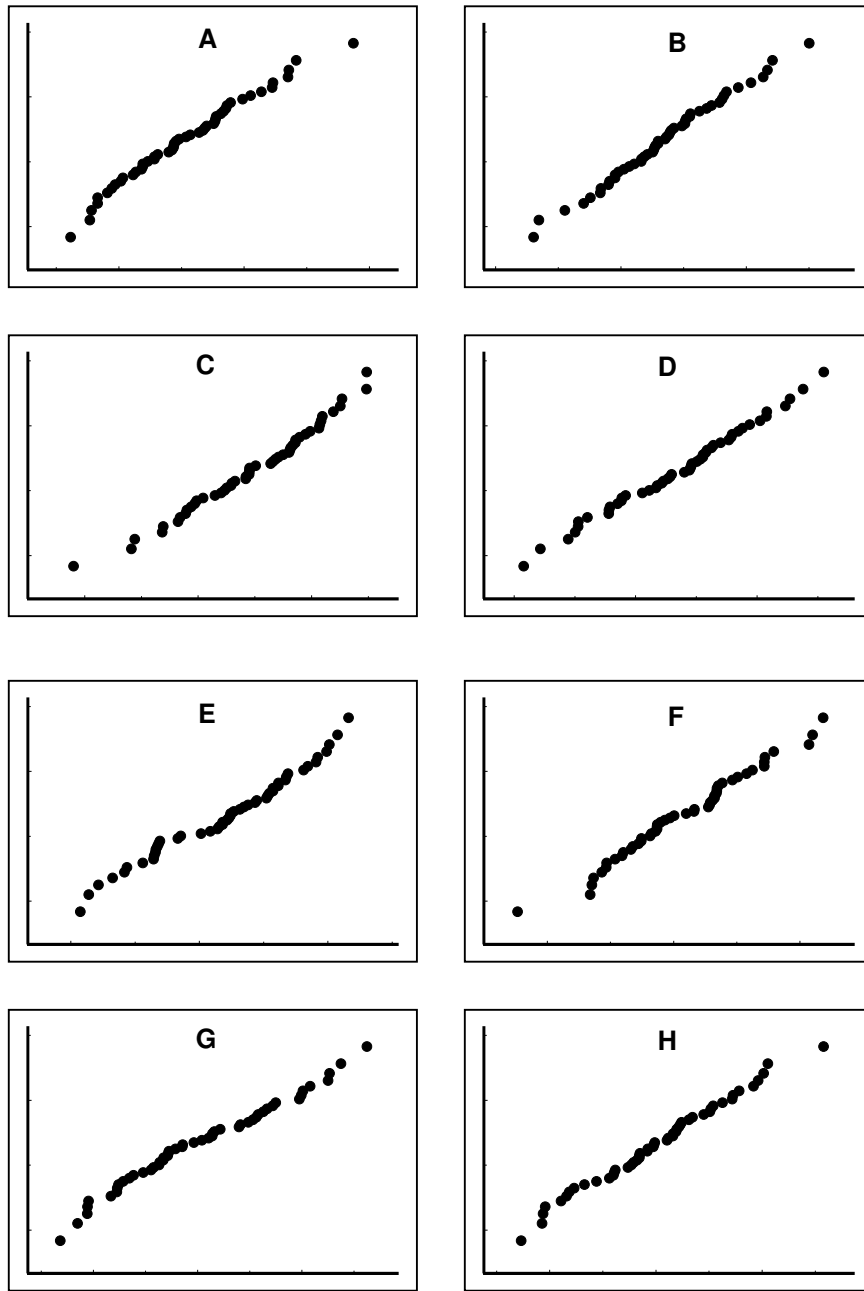
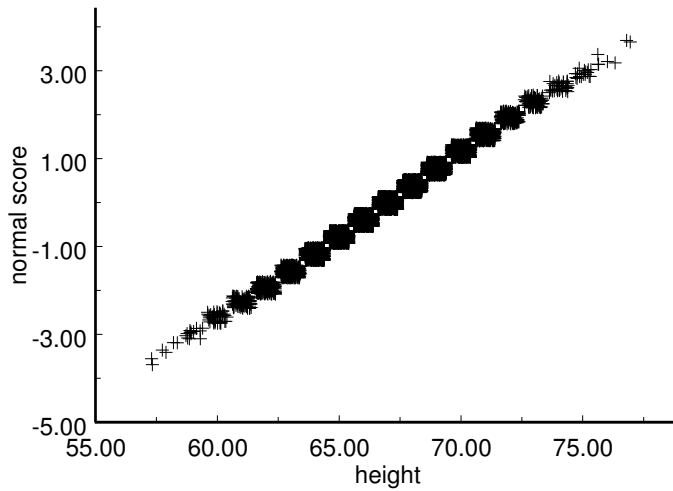
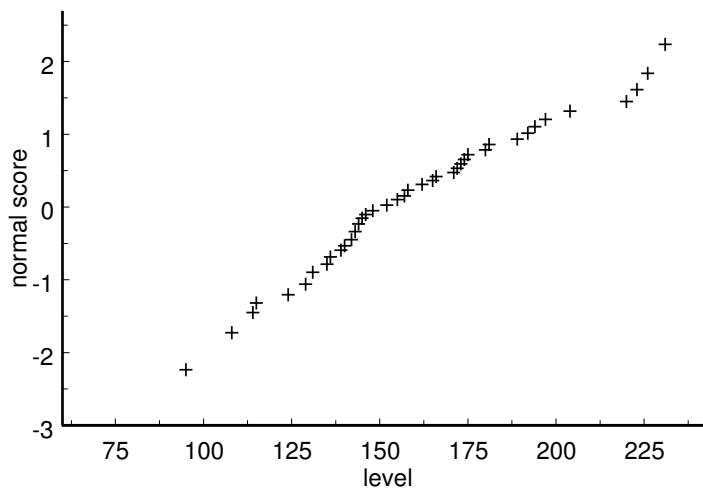


Figure 9. Normal probability plot for the UK height example.

**Figure 10. Normal probability plot for rural cholesterol levels.**

**Figure 11. Normal probability plot for unseeded rainfall.**

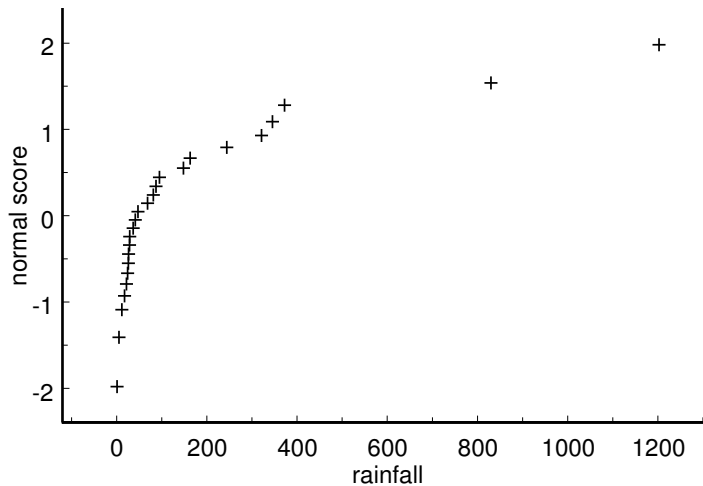


Figure 12. Stem and leaf histogram for unseeded rainfall.

In this stem and leaf histogram the stem represents hundreds and the leaf represents tens. (acrefeet)

stem	leaf
0	000112222223446889
1	56
2	4
3	247
4	
5	
6	
7	
8	3
9	
10	
11	
12	0

7.2d Estimating a normal population mean

The sample mean \bar{X} provides a single number estimate of the population mean μ . We can think of \bar{X} as our “best guess” of the value of μ . As noted above we know that \bar{X} is unbiased as an estimator of μ ; therefore, on the average in the long run (under repeated sampling) we know that \bar{X} provides a good estimate of the unknown mean μ . Since this unbiasedness does not guarantee that the observed value of \bar{X} , based on a single sample, will be close to the true, unknown value of μ , it would be useful to have a confidence interval estimate of μ .

Recall that when the sample mean corresponds to a random sample from a population distribution with population mean μ and population standard deviation σ the sampling distribution of \bar{X} has mean μ and standard deviation $\text{S.E.}(\bar{X}) = \sigma/\sqrt{n}$ (the population standard error of \bar{X}). When the underlying population distribution is normal we can say more about the form of this sampling distribution. If the random sample from which the sample mean \bar{X} is computed is a random sample from a normal population with population mean μ and population standard deviation σ , then the sampling distribution of \bar{X} is a normal distribution with population mean μ and population standard deviation $\text{S.E.}(\bar{X}) = \sigma/\sqrt{n}$. Thus, under these assumptions, the quantity

$$Z = \frac{\bar{X} - \mu}{\text{S.E.}(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

behaves in accordance with the standard normal distribution.

We know that a standard normal variable Z will take on a value between -1.96 and 1.96 with probability $.95$ ($P(-1.96 \leq Z \leq 1.96) = .95$). This implies that

$$P\left(-1.96 \leq \frac{\bar{X} - \mu}{\text{S.E.}(\bar{X})} \leq 1.96\right) = .95$$

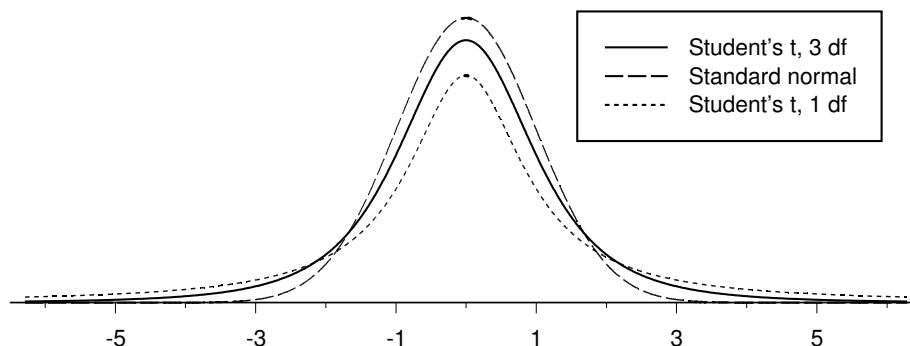
which is equivalent to

$$P(\bar{X} - 1.96\text{S.E.}(\bar{X}) \leq \mu \leq \bar{X} + 1.96\text{S.E.}(\bar{X})) = .95.$$

This probability statement says that 95% of the time we will observe a value of \bar{X} such that the population mean μ will be between $\bar{X} - 1.96\text{S.E.}(\bar{X})$ and $\bar{X} + 1.96\text{S.E.}(\bar{X})$. Unfortunately, we cannot use this interval as a confidence interval for μ , since the population standard error $\text{S.E.}(\bar{X}) = \sigma/\sqrt{n}$ depends on the unknown population standard deviation σ and thus is not computable. We can avoid this difficulty by replacing the unknown population standard error σ/\sqrt{n} by the sample standard error $\widehat{\text{S.E.}}(\bar{X}) = S/\sqrt{n}$, where S is the sample standard deviation, and basing our confidence interval estimate on the quantity

$$T = \frac{\bar{X} - \mu}{\widehat{\text{S.E.}}(\bar{X})} = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

If the sample mean \bar{X} and the sample standard deviation S are computed from a random sample of size n from a normal population with population mean μ and population standard deviation σ , then the quantity T defined above follows the **Student's t distribution with $n - 1$ degrees of freedom**. The Student's t distribution with $n - 1$ degrees of freedom is symmetric about zero and has a density curve very similar to that of the standard normal distribution. The main difference between these two distributions is that the Student's t distribution has “heavier” tails than the standard normal distribution. That is, the tails of the Student's t density curve approach the x -axis more slowly than do the tails of the standard normal density curve. As the sample size (and the degrees of freedom) increases the Student's t distribution becomes more similar to the standard normal distribution. In fact, the standard normal distribution is the limiting version of the Student's t distribution in the sense that the Student's t density curve approaches the standard normal density curve when the degrees of freedom increases without bound. The relationship between Student's t distributions and the standard normal distribution is indicated by the plots in Figure 13.

Figure 13. Student's t distributions with 1 and 3 degrees of freedom and standard normal distribution.

Given a constant k such that $P(-k \leq T \leq k) = .95$, where T denotes a Student's t variable with $n - 1$ degrees of freedom, we have

$$P\left(-k \leq \frac{\bar{X} - \mu}{\widehat{\text{S.E.}}(\bar{X})} \leq k\right) = .95$$

which is equivalent to

$$P\left(\bar{X} - k\widehat{\text{S.E.}}(\bar{X}) \leq \mu \leq \bar{X} + k\widehat{\text{S.E.}}(\bar{X})\right) = .95.$$

The quantity

$$\text{M.E.}(\bar{X}) = k\widehat{\text{S.E.}}(\bar{X}) = \frac{kS}{\sqrt{n}}$$

is the 95% **margin of error** of \bar{X} . The preceding probability statement says that 95% of the time we will observe values of \bar{X} and S such that the population mean μ will be between $\bar{X} - \text{M.E.}(\bar{X})$ and $\bar{X} + \text{M.E.}(\bar{X})$. Thus, the interval from $\bar{X} - \text{M.E.}(\bar{X})$ to $\bar{X} + \text{M.E.}(\bar{X})$ is a 95% confidence interval estimate for μ . To compute this confidence interval we need to determine the value of the appropriate margin of error multiplier k . This multiplier depends on the size of the sample so that there is a different multiplier for each sample size. The symmetry of the Student's t distribution and the definition of k imply that k is the 97.5 percentile of the Student's t distribution with $n - 1$ degrees of freedom. The 95% margin of error multipliers (k 's) based on the Student's t distribution, for several choices of the degrees of freedom (d.f.), are given in Table 4.

This confidence interval estimate may be reported using a statement such as: We are 95% confident that the population mean μ is between $\bar{X} - \text{M.E.}(\bar{X})$ and $\bar{X} + \text{M.E.}(\bar{X})$. Notice that it is the sample mean \bar{X} and the margin of error $\text{M.E.}(\bar{X})$ that vary from sample to sample. The population mean μ is a fixed, unknown parameter that does not

vary. Therefore, the 95% confidence level applies to the method used to generate the confidence interval estimate. That is, the method (obtain a simple random sample and compute the numbers $\bar{X} - \text{M.E.}(\bar{X})$ and $\bar{X} + \text{M.E.}(\bar{X})$ forming the limits of the confidence interval) is such that 95% of the time it will yield a pair of confidence interval limits which bracket the population mean μ .

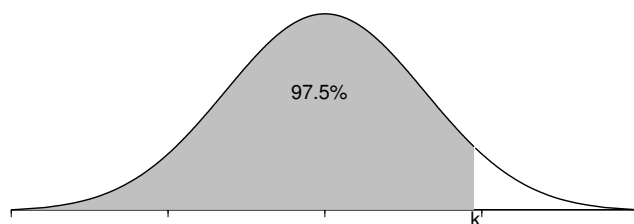


Table 4. 97.5 percentiles of the Student's t distribution.

$d.f.$	k	$d.f.$	k	$d.f.$	k	$d.f.$	k
1	12.706	14	2.145	27	2.052	40	2.021
2	4.303	15	2.131	28	2.048	45	2.014
3	3.182	16	2.120	29	2.045	50	2.008
4	2.776	17	2.110	30	2.042	55	2.004
5	2.571	18	2.101	31	2.040	60	2.000
6	2.447	19	2.093	32	2.037	65	1.997
7	2.365	20	2.086	33	2.034	70	1.994
8	2.306	21	2.080	34	2.032	75	1.992
9	2.262	22	2.074	35	2.030	80	1.989
10	2.228	23	2.069	36	2.028	85	1.988
11	2.201	24	2.064	37	2.026	90	1.986
12	2.179	25	2.060	38	2.024	95	1.985
13	2.160	26	2.056	39	2.023	100	1.982

Example. Newcomb's measurements of the speed of light. In 1882 Simon Newcomb conducted an investigation to measure the speed of light. The essence of Newcomb's experiment was to determine the time it took for light to travel from Fort Myer on the west bank of the Potomac river to a fixed mirror at the foot of the Washington monument 3721 meters away and back. (More details about this and similar examples can be found in Stigler (1977), Do robust estimators work with real data? (with discussion), *Annals of Statistics*, **5**, 1055–1098.) Data from 64 replications of this experiment are provided in Table 5. For ease of handling the times in Table 5 are simplified. The values given in Table 5 are times expressed as billionths of a second in excess of 24.8 millionths of a second, *i.e.*, if a time value from Table 5 is multiplied by 10^{-3} and added to 24.8, the result is the time which Newcomb observed measured in millionths of a second. For example, the first observation in Table 5 is 28 which corresponds to an observed time of 24.828 millionths of a second.

Table 5. Newcomb's time data.

28	22	36	26	28	28	26	24	32	30	27	24	33	21	36	32
31	25	24	25	28	36	27	32	34	30	25	26	26	25	23	21
30	33	29	27	29	28	22	26	27	16	31	29	36	32	28	40
19	37	23	32	29	24	25	27	24	16	29	20	28	27	39	23

Figure 14. Stem and leaf histogram for Newcomb's time data.

In this stem and leaf histogram the stem represents tens of billionths of a second and the leaf represents billionths of a second.

stem	leaf
1	66
1	9
2	011
2	22333
2	4444455555
2	66666777777
2	88888899999
3	00011
3	2222233
3	4
3	66667
3	9
4	0

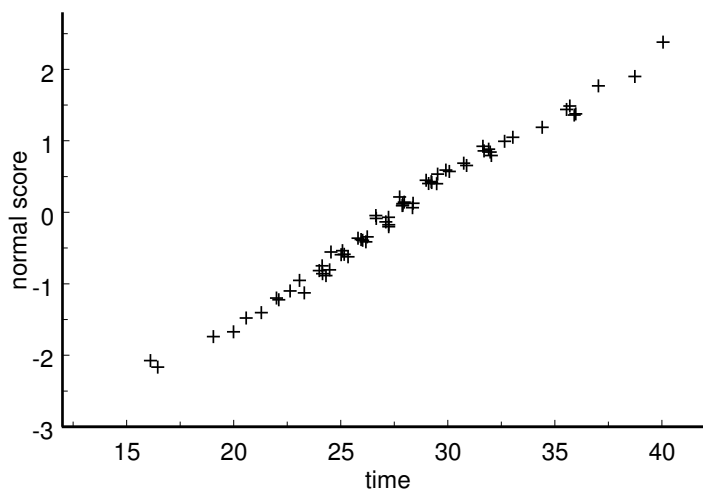
Figure 15. Normal probability plot for Newcomb's time data.

Table 6. Summary statistics for Newcomb's time data.

minimum	16.0	Q_1 - minimum	8.5
Q_1	24.5	median - Q_1	3.0
median	27.5	Q_3 - median	3.5
Q_3	31.0	maximum - Q_3	9.0
maximum	40.0		
mean	27.7500	range	24.0
standard deviation	5.0834	IQ range	6.5

Summary statistics for these times are given in Table 6, a stem and leaf histogram is provided in Figure 14, and a normal probability plot is given in Figure 15. Based on the summary statistics, the stem and leaf histogram, and the normal probability plot it seems reasonable to assume that these data form a random sample from a normal population. More specifically, it seems reasonable to assume that these 64 measurements are independent realizations of a normally distributed variable with population mean μ . Actually, the assumption of independence is somewhat questionable, since repetitions of the experiment conducted at nearly the same point in time might exhibit some dependence. We will hope that any such dependence is minor and treat these observations as if they were independent. The population mean μ can be thought of as the “true” time it would take light to make the “trip” as indicated for this experiment if this time could be measured very precisely. We can also view the population mean μ as the long run average of the times which would be obtained if this experiment was repeated over and over. We must keep in mind the fact that Newcomb's method of measurement may introduce a systematic bias which would make the “true” time μ differ from the actual time it would take for light to travel a distance of $7442 = 2 \times 3721$ meters.

The sample mean for these $n = 64$ observations is 27.7500, the sample standard deviation is 5.0834, the standard error of the sample mean is $5.0834/\sqrt{64} = .6354$, and the multiplier for the 95% margin of error, based on the Student's t distribution with $n - 1 = 63$ degrees of freedom, is $k = 1.9983$. Thus the 95% margin of error of the sample mean is $(1.9983)(.6354) = 1.2697$ and the limits for the 95% confidence interval estimate of μ are $27.7500 - 1.2697 = 26.4803$ and $27.7500 + 1.2697 = 29.0197$. Therefore, we are 95% confident that the population mean time for Newcomb's experiment is between 26.4803 and 29.0197 billionths of a second. If we re-express this in terms of the time required for light to travel through the experimental set-up, we find that we are 95% confident that the population mean time measured in this way is between 24.8264803 and 24.8290197 millionths of a second.

For the preceding analysis we have removed two “outliers” from Newcomb's data. The two unusual values are -44 and -2, which correspond to measured times of 24.756 and

24.798 millionths of a second. If these values are added to the stem and leaf histogram of Figure 14, they are clearly inconsistent with the other 64 data values. It seems reasonable to conjecture that something must have gone wrong with the experiment when these unusually small values were obtained. Newcomb chose to omit the observation of -44 and retain the observation of -2. If we consider the 65 observations including the -2, we find that the sample mean is reduced to 27.2923, the sample standard deviation is increased to 6.2493, and the standard error of the mean becomes .7752. Notice that, as we would expect, the presence of this outlier reduces the sample mean (moves the mean toward the outlier) and increases the sample standard deviation (increases the variability in the data). With $n = 65$ observations the multiplier for the 95% margin of error, based on the Student's t distribution with $n - 1 = 64$ degrees of freedom, is $k = 1.9977$, which gives a margin of error of $(1.9977)(.7752) = 1.5486$. Hence, when the unusually small value -2 is included we are 95% confident that the population mean time for Newcomb's experiment is between 25.7437 and 28.8409 billionths of a second. Re-expressing this in terms of the time required for light to travel through the experimental set-up, we find that we are 95% confident that the population mean time measured in this way is between 24.8257437 and 24.8288409 millionths of a second. Including the outlier, -2, has the effect of shifting the confidence interval to the left and making it longer.

Example. Heights of husbands and wives. The data used in this example are part of data set 231 in Hand, Daly, Lunn, McConway, and Ostrowski (1994), *A Handbook of Small Data Sets*, Chapman and Hall, London. The original source is Marsh (1988), *Exploring Data*, Cambridge, UK. A random sample of $n = 169$ married couples was selected from the 1980 OPCS study of the heights and weights of the adult population of Great Britain. The data consist of paired heights (in mm) for these husbands and wives. A few of these paired heights are given in Table 7 and all of the differences (husband's height minus wife's height) are given in Table 8.

Table 7. Husband and wife height data (partial).

couple	husband's height	wife's height	difference
1	1786	1590	196
2	1754	1660	94
3	1755	1590	165
4	1725	1550	175
5	1796	1550	246
...
169	1641	1570	71

Table 8. Husband and wife height differences.

-96	-65	-46	-37	-30	-30	-21	-12	0	0
1	9	13	14	15	19	34	35	35	36
39	40	50	55	56	59	60	60	65	65
66	68	70	70	71	71	73	74	75	75
79	79	81	83	84	84	84	85	85	88
90	94	94	95	96	100	103	103	105	110
110	113	113	114	115	118	119	120	120	123
123	123	123	125	125	125	125	125	128	128
128	129	130	130	130	133	134	135	135	135
135	139	140	141	141	144	144	145	145	145
145	150	151	155	155	155	159	159	160	160
160	160	161	161	164	165	165	165	166	169
170	174	175	175	178	180	181	183	183	188
189	190	190	191	193	194	195	195	195	196
196	196	204	205	209	210	215	219	225	228
228	229	233	235	236	239	241	243	244	246
250	255	258	259	271	276	281	295	303	

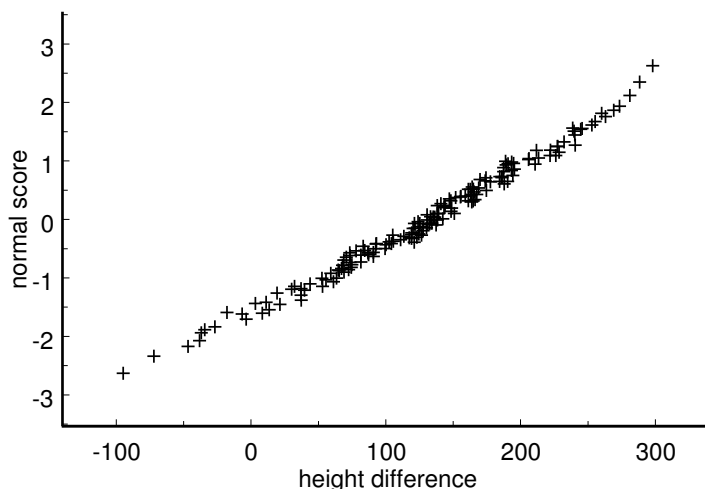
Table 9. Summary statistics for height differences.

minimum	-96	Q_1 - minimum	177
Q_1	81	median - Q_1	49
median	130	Q_3 - median	51
Q_3	181	maximum - Q_3	122
maximum	303		
mean	129.8225	range	399.0
standard deviation	76.0211	IQ range	100.0

Figure 16. Stem and leaf histogram for height differences.

In this stem and leaf histogram the stem represents hundreds and the leaf tens (mm).

stem	leaf
-0	96
-0	4333210
0	0001111333334
0	555566666677777777778888888899999
1	000011111111222222222222223333333333444444444
1	555555566666666666667777788888899999999999
2	000111222233334444
2	55557789
3	0

Figure 17. Normal probability plot for height differences.

Summary statistics for these differences are given in Table 9, a stem and leaf histogram is provided in Figure 16, and a normal probability plot is given in Figure 17. Based on the summary statistics, the stem and leaf histogram, and the normal probability plot it seems reasonable to assume that these differences form a random sample from a normal population of differences. The population mean difference μ_D is the average difference in height corresponding to the population of all married couples in Great Britain in 1980. (Technically, we should restrict this mean to those married couples included in the 1980 COPS study from which the sample was taken.)

The sample mean for these $n = 169$ differences is 129.8225 mm, the sample standard deviation is 76.0211 mm, the standard error of the sample mean is $76.0211/\sqrt{169} = 5.8478$, and the multiplier for the 95% margin of error, based on the Student's t distribution with $n - 1 = 168$ degrees of freedom, is $k = 1.9742$. Thus the 95% margin of error of the sample mean is $(1.9742)(5.8478) = 11.5447$ and the limits for the 95% confidence interval estimate of μ_D are $129.8225 - 11.5447 = 118.2778$ and $129.8225 + 11.5447 = 141.3672$. Therefore, we are 95% confident that for the population of all married couples in Great Britain in 1980, on average, the husband's height exceeds the wife's height by at least 118.2778 mm and perhaps as much as 141.3672 mm.

Remark regarding directional confidence bounds. *The use of one of the confidence limits of a 90% confidence interval as a 95% confidence bound discussed in Section 5.4 can also be used in the present context. That is, we can find an upper or lower 95% confidence bound for μ by selecting the appropriate confidence limit from a 90% confidence interval estimate of μ .*

7.2e Tests of hypotheses about a normal population mean

The hypothesis testing procedures discussed in this section are based on the fact that, when $\mu = \mu_0$, the **Student's t test statistic**

$$T_{calc} = \frac{\bar{X} - \mu_0}{\widehat{\text{S.E.}}(\bar{X})} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

follows the Student's t distribution with $n - 1$ degrees of freedom. Recall that, technically, this result requires that the sample from which the sample mean \bar{X} and the sample standard deviation S are computed forms a random sample from a normal distribution with population mean μ . However, these methods are known to be generally reasonable even when the underlying population is not exactly a normal population, provided the underlying population distribution is reasonably symmetric and the true density curve has a more or less normal (bell-shaped) appearance. An alternate approach to inference (for a population median) which may be used when the Student's t methods are inappropriate is discussed in Section 7.3.

Recall that a hypothesis (statistical hypothesis) is a conjecture about the nature of the population. When we considered hypotheses about dichotomous populations we noted that the population was completely determined by the population success proportion p . In the present context of sampling from a normal population two parameters, the mean and the standard deviation, must be specified to completely determine the normal population. A hypothesis about the value of the population mean μ of a normal distribution specifies where the center of this normal distribution, μ , is located on the number line but places no restriction on the population standard deviation.

Even though the logic behind a hypothesis test for a population mean μ is the same as the logic behind a hypothesis test for a population proportion p , we will introduce hypothesis testing for a mean in the context of a simple hypothetical example.

Example. Strength of bricks. Consider a brick manufacturer that has produced a large batch of bricks and wants to determine whether these bricks are suitable for a particular construction project. The specifications for this project require bricks with a mean compressive strength that exceeds 3200 psi. In order to assess the suitability of this batch of bricks the manufacturer will obtain a simple random sample of bricks from this batch and measure their compressive strength. In this example a single brick is a unit, the entire large batch of bricks is the population, and a suitable variable X is the compressive strength of an individual brick (in psi). We will assume that the distribution of X is reasonably modeled by a normal distribution with population mean μ and population standard deviation σ . In this example the population mean μ represents the mean

compressive strength for all of the bricks in this large batch and the population standard deviation quantifies the variability from brick to brick in (measured) compressive strength.

The manufacturer does not want to use these bricks for this construction project unless there is sufficient evidence to claim that the population mean compressive strength for this batch exceeds 3200 psi. Thus, the question of interest here is: “Is there sufficient evidence to justify using this batch of bricks for this project?” In terms of the population mean μ the research hypothesis is $H_1 : \mu > 3200$ (the mean compressive strength for this batch of bricks exceeds 3200 psi); and the null hypothesis is $H_0 : \mu \leq 3200$ (the mean compressive strength for this batch of bricks does not exceed 3200 psi). In other words, the manufacturer will tentatively assume that these bricks are not suitable for this project and will check to see if there is sufficient evidence against this tentative assumption to justify the conclusion that these bricks are suitable for the project.

A test of the null hypothesis $H_0 : \mu \leq 3200$ versus the research hypothesis $H_1 : \mu > 3200$ begins by tentatively assuming that the mean compressive strength for this batch of bricks is no larger than 3200 psi. Under this tentative assumption it would be surprising to observe a sample mean compressive strength \bar{X} that was much larger than 3200. Thus the test should reject $H_0 : \mu \leq 3200$ in favor of $H_1 : \mu > 3200$ if the observed value of \bar{X} is sufficiently large relative to 3200. In order to determine whether \bar{X} is large relative to 3200 we need an estimate of the sample to sample variability in \bar{X} . The sample standard error of the sample mean $\widehat{S.E.}(\bar{X}) = S/\sqrt{n}$, where S denotes the sample standard deviation, provides a suitable measure of the sample to sample variability in \bar{X} . Our conclusion will hinge on deciding whether the observed value of the Student’s t test statistic

$$T_{calc} = \frac{\bar{X} - 3200}{\widehat{S.E.}(\bar{X})} = \frac{\bar{X} - 3200}{S/\sqrt{n}}$$

is far enough above zero to make $\mu > 3200$ more tenable than $\mu \leq 3200$. We will base this decision on the probability of observing a value of T as large or larger than the actual calculated value T_{calc} of T , under the assumption that $\mu \leq 3200$. This probability (computed assuming that $\mu = 3200$) is the P -value of the test. We will use the fact that, when $\mu = 3200$, the Student’s t statistic T follows the Student’s t distribution with $n - 1$ degrees of freedom to calculate the P -value.

First suppose that a simple random sample of $n = 100$ bricks is selected from the large batch. Further suppose that the sample mean compressive strength of these 100 bricks is $\bar{X} = 3481$ psi and the sample standard deviation is $S = 1118.38$. In this case we know that the mean compressive strength of the bricks in the sample $\bar{X} = 3481$ exceeds 3200 and we need to decide whether this suggests that the mean compressive strength of all of the bricks in the batch μ exceeds 3200. In this case the sample standard error of \bar{X} is $\widehat{S.E.}(\bar{X}) = 1118.38/\sqrt{100} = 111.838$. Thus \bar{X} exceeds 3200 by 281 psi which is

$281/111.838 = 2.5126$ standard error units, *i.e.*, $T_{calc} = 2.5126$. In this case the P -value .0068 is the probability of observing a calculated T value as large or larger than 2.5126 when $\mu \leq 3200$. That is, if $\mu \leq 3200$ and we perform this experiment with $n = 100$, then we would expect to see a value of \bar{X} that is 2.5126 standard error units above the hypothesized value 3200 (2.5126 times $\widehat{S.E.}(\bar{X})$ psi above 3200) about 0.68% of the time. Therefore, observing values of \bar{X} and S in a sample of size $n = 100$ which yield a value of T_{calc} as large or larger than 2.5126 would be very surprising if the true mean compressive strength for the batch μ was not larger than 3200 and we have sufficient evidence to reject the null hypothesis $H_0 : \mu \leq 3200$ in favor of the research hypothesis $H_1 : \mu > 3200$. In the case we would conclude that there is sufficient evidence to contend that the mean compressive strength for this batch of bricks exceeds 3200 psi, *i.e.*, these bricks are suitable for the construction project.

Now suppose that a simple random sample of $n = 100$ bricks is selected from the large batch and the sample mean compressive strength of these 100 bricks is $\bar{X} = 3481$ psi as before. However, suppose that in this case the sample standard deviation is $S = 2329.3$. As before we know that the mean compressive strength of the bricks in the sample $\bar{X} = 3481$ exceeds 3200 and we need to decide whether this suggests that the mean compressive strength of all of the bricks in the batch μ exceeds 3200. In this case the sample standard error of \bar{X} is $\widehat{S.E.}(\bar{X}) = 2329.3/\sqrt{100} = 232.93$. Thus \bar{X} exceeds 3200 by 281 psi which is $281/232.93 = 1.2064$ standard error units, *i.e.*, $T_{calc} = 1.2064$. In this case the P -value .1153 is the probability of observing a calculated T value as large or larger than 1.2064 when $\mu \leq 3200$. That is, if $\mu \leq 3200$ and we perform this experiment with $n = 100$, then we would expect to see a value of \bar{X} that is 1.2064 standard error units above the hypothesized value 3200 (1.2064 times $\widehat{S.E.}(\bar{X})$ psi above 3200) about 11.53% of the time. Therefore, observing values of \bar{X} and S in a sample of size $n = 100$ which yield a value of T_{calc} as large or larger than 1.2064 would not be very surprising if the true mean compressive strength for the batch μ was not larger than 3200 and we do not have sufficient evidence to reject the null hypothesis $H_0 : \mu \leq 3200$ in favor of the research hypothesis $H_1 : \mu > 3200$. In the case we would conclude that there is not sufficient evidence to contend that the mean compressive strength for this batch of bricks exceeds 3200 psi, *i.e.*, these bricks are not suitable for the construction project.

The research hypothesis in the brick example is a directional hypothesis of the form $H_1 : \mu > \mu_0$, where $\mu_0 = 3200$. We will now discuss the details of a hypothesis test for a directional research hypothesis of this form. For the test procedure to be valid the specified value μ_0 and the direction of the research hypothesis must be motivated from subject matter knowledge before looking at the data that are to be used to perform the test.

Let μ_0 denote the hypothesized value which we wish to compare with μ . The research hypothesis states that μ is greater than μ_0 ; in symbols we will indicate this research hypothesis by writing

$$H_1 : \mu > \mu_0.$$

The null hypothesis is the negation of $H_1 : \mu > \mu_0$ which states that μ is not greater than μ_0 ; in symbols we will indicate this null hypothesis by writing

$$H_0 : \mu \leq \mu_0.$$

The research hypothesis $H_1 : \mu > \mu_0$ specifies that the normal distribution is one of the normal distributions for which the population mean μ is greater than μ_0 . The null hypothesis $H_0 : \mu \leq \mu_0$ specifies that the normal distribution is one of the normal distributions for which the population mean μ is at most μ_0 . The population standard deviation σ is not restricted by either hypothesis. Assuming that the population standard deviation is the same regardless of which hypothesis is true, this competing pair of hypotheses provides a decomposition of all possible normal distributions with this σ into the collection of normal distributions where $\mu > \mu_0$ and the research hypothesis is true and the collection of normal distributions where $\mu \leq \mu_0$ and the null hypothesis is true. Our goal is to use the data to decide which of these two collections of normal distributions contains the normal distribution we are actually sampling from.

Since a hypothesis test begins by tentatively assuming that the null hypothesis is true, we need to decide what constitutes evidence against the null hypothesis $H_0 : \mu \leq \mu_0$ and in favor of the research hypothesis $H_1 : \mu > \mu_0$. We will assume that the unknown population standard deviation σ is fixed regardless of the value of μ . The difference $\bar{X} - \mu_0$ between the sample mean \bar{X} and the hypothesized value μ_0 , expressed in standard error units, will be used to assess the strength of the evidence in favor of the research hypothesis. Generally, we would expect to observe larger values of \bar{X} more often when the research hypothesis $H_1 : \mu > \mu_0$ is true than when the null hypothesis $H_0 : \mu \leq \mu_0$ is true. In particular, we can view the observation of a value of \bar{X} that is sufficiently large relative to μ_0 as constituting evidence against the null hypothesis $H_0 : \mu \leq \mu_0$ and in favor of the research hypothesis $H_1 : \mu > \mu_0$. To decide whether the observed value of \bar{X} is “sufficiently large relative to μ_0 ” we need to take the variability in the data into account. We can do this by basing our decision on the corresponding Student’s t test statistic,

$$T_{calc} = \frac{\bar{X} - \mu_0}{\widehat{\text{S.E.}}(\bar{X})} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

instead of \bar{X} alone. Notice that values of T_{calc} which are large relative to zero correspond to values of \bar{X} which are large relative to μ_0 . Deciding whether the observed value of T_{calc}

is sufficiently large relative to zero to allow rejection of H_0 is based on the corresponding P -value, which is defined below.

The P -value for testing the null hypothesis $H_0 : \mu \leq \mu_0$ versus the research hypothesis $H_1 : \mu > \mu_0$ is the probability of observing a value of a Student's t variable T as large or larger than the calculated value T_{calc} that we actually do observe, *i.e.*, $P\text{-value} = P(T \geq T_{calc})$, where T denotes a Student's t variable with $n - 1$ degrees of freedom. This P -value is computed under the assumption that the research hypothesis $H_1 : \mu > \mu_0$ is false and the null hypothesis $H_0 : \mu \leq \mu_0$ is true. Because the null hypothesis only specifies that $\mu \leq \mu_0$, we need to choose a particular value of μ (that is no larger than μ_0) in order to compute the P -value. It is most appropriate to use $\mu = \mu_0$ for this computation. Using $\mu = \mu_0$, which defines the boundary between $\mu \leq \mu_0$, where the null hypothesis is true, and $\mu > \mu_0$, where the research hypothesis is true, provides some protection against incorrectly rejecting $H_0 : \mu \leq \mu_0$.

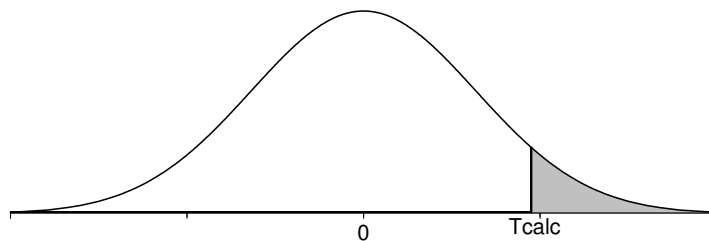
The steps for performing a hypothesis test for

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0$$

are summarized below.

1. Use a suitable calculator or computer program to find the P -value = $P(T \geq T_{calc})$, where T denotes a Student's t variable with $n - 1$ degrees of freedom and $T_{calc} = (\bar{X} - \mu_0) / \widehat{S.E.}(\bar{X})$ as described above. This P -value is the area to the right of T_{calc} under the density curve for the Student's t distribution with $n - 1$ degrees of freedom as indicated in Figure 18.

Figure 18. P -value for $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$.



- 2a. If the P -value is small enough (less than .05 for a test at the 5% level of significance), conclude that the data favor $H_1 : \mu > \mu_0$ over $H_0 : \mu \leq \mu_0$. That is, if the P -value is small enough, then there is sufficient evidence to conclude that the population mean μ is greater than μ_0 .
- 2b. If the P -value is not small enough (is not less than .05 for a test at the 5% level of significance), conclude that the data do not favor $H_1 : \mu > \mu_0$ over $H_0 : \mu \leq \mu_0$. That

is, if the P -value is not small enough, then there is not sufficient evidence to conclude that the population mean μ is greater than μ_0 .

The procedure for testing the null hypothesis $H_0 : \mu \leq \mu_0$ versus the research hypothesis $H_1 : \mu > \mu_0$ given above is readily modified for testing the null hypothesis $H_0 : \mu \geq \mu_0$ versus the research hypothesis $H_1 : \mu < \mu_0$. The essential modification is to change the direction of the inequality in the definition of the P -value. Consider a situation where the research hypothesis specifies that the population mean μ is less than the particular, hypothesized value μ_0 . For these hypotheses values of the sample mean \bar{X} that are sufficiently small relative to μ_0 provide evidence in favor of the research hypothesis $H_1 : \mu < \mu_0$ and against the null hypothesis $H_0 : \mu \geq \mu_0$. Therefore, the appropriate P -value is the probability of observing a value of a Student's t variable T as small or smaller than the value actually observed. As before, the P -value is computed under the assumption that $\mu = \mu_0$. The calculated t statistic T_{calc} is defined as before; however, in this situation the P -value is the area under the density curve of the Student's t distribution with $n - 1$ degrees of freedom to the left of T_{calc} , since values of \bar{X} that are small relative to μ_0 constitute evidence in favor of the research hypothesis.

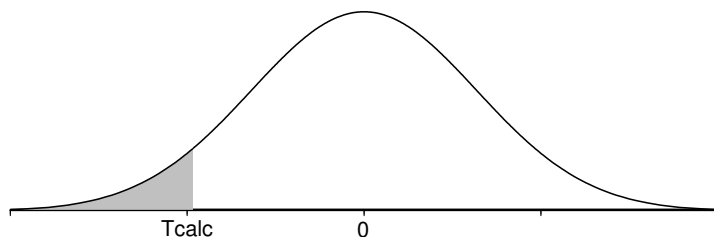
The steps for performing a hypothesis test for

$$H_0 : \mu \geq \mu_0 \text{ versus } H_1 : \mu < \mu_0$$

are summarized below.

1. Use a suitable calculator or computer program to find the P -value = $P(T \leq T_{calc})$, where T denotes a Student's t variable with $n - 1$ degrees of freedom and $T_{calc} = (\bar{X} - \mu_0) / \widehat{S.E.}(\bar{X})$ as described above. This P -value is the area to the left of T_{calc} under the density curve for the Student's t distribution with $n - 1$ degrees of freedom as indicated in Figure 19.

Figure 19. P -value for $H_0 : \mu \geq \mu_0$ versus $H_1 : \mu < \mu_0$.



- 2a. If the P -value is small enough (less than .05 for a test at the 5% level of significance), conclude that the data favor $H_1 : \mu < \mu_0$ over $H_0 : \mu \geq \mu_0$. That is, if the P -value is small enough, then there is sufficient evidence to conclude that the population mean μ is less than μ_0 .

- 2b. If the P -value is not small enough (is not less than .05 for a test at the 5% level of significance), conclude that the data do not favor $H_1 : \mu < \mu_0$ over $H_0 : \mu \geq \mu_0$. That is, if the P -value is not small enough, then there is not sufficient evidence to conclude that the population mean μ is less than μ_0 .

Example. Brain changes in response to experience. A study by Rosenzweig, Bennett, and Diamond, described in an article in *Scientific American* (1964), was conducted to examine the effects of psychological environment on the anatomy of the brain. The units for this study came from a strain of genetically pure rats. A pair of rats was selected at random from each of 12 litters of rats; one of these rats was placed in group A and the other in group B. Each animal in group A lived with eleven others in a large cage, furnished with playthings which were changed daily. Each animal in group B lived in isolation, with no toys. Both groups of rats were provided with as much food and drink as they desired. After a month, the rats were killed and dissected. One variable which was measured was the weight (in milligrams) of the cortex of the rat. The cortex is the “thinking” part of the brain. The question we wish to address here is whether there is evidence in favor of the contention that the cortex of a rat raised in the more stimulating environment of group A will tend to be larger than the cortex of a rat raised in the less stimulating environment of group B.

The researchers conducted this experiment five times. Data from one of these experiments are given in Table 10. There are sets of three values for each of twelve pairs of littermates in this table: the weight of the cortex of the rat in group A, the weight of the cortex of the rat in group B, and the difference between these two weights (A weight minus B weight); all of these values are measured in milligrams.

Table 10. Rat cortex weight data.

pair	group A	group B	difference	pair	group A	group B	difference
1	690	668	22	7	720	665	55
2	701	667	34	8	718	689	29
3	685	647	38	9	718	642	76
4	751	693	58	10	696	673	23
5	647	635	12	11	658	675	-17
6	647	644	3	12	680	641	39

Summary statistics for the differences in cortex weights (A weight minus B weight) are given in Table 11, a stem and leaf histogram is provided in Figure 20, and a normal probability plot is given in Figure 21. Based on the summary statistics and the stem and leaf histogram there is some evidence that this distribution is slightly skewed to the left; however, the normal probability plot is reasonably linear and it seems reasonable to

assume that these data form a random sample from a normal population. More specifically, it seems reasonable to assume that these 12 differences are independent realizations of a normally distributed variable with population mean μ_D . We can think of this population mean difference μ_D as the average difference that would be obtained if this experiment was conducted using all possible littermate pairs from this strain of genetically pure rats.

Table 11. Summary statistics for the rat cortex weight differences.

minimum	-17.0	Q_1 - minimum	34.0
Q_1	17.0	median - Q_1	14.5
median	31.5	Q_3 - median	15.5
Q_3	47.0	maximum - Q_3	29.0
maximum	76.0		
mean	31.0000	range	93
standard deviation	25.3162	IQ range	30

Figure 20. Stem and leaf histogram for the rat cortex weight difference data.

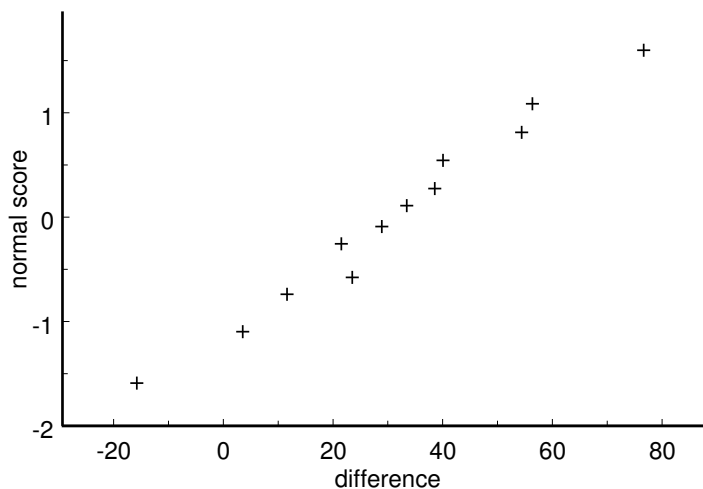
In this stem and leaf histogram the stem represents tens and the leaf represents ones. (milligrams)

stem	leaf
-1	7
-0	
0	3
1	2
2	239
3	489
4	
5	58
6	
7	6

For this example the research hypothesis can be formalized as $H_1 : \mu_D > 0$. This research hypothesis specifies that the population mean difference, μ_D , between the cortex weight of a stimulated rat (a rat raised in a stimulating environment like that of group A) and the cortex weight of a unstimulated rat (a rat raised in a non-stimulating environment like that of group B) exceeds zero, *i.e.*, for this population of pairs of rats, on average, the cortex weight of a stimulated rat would be higher than the cortex weight of an unstimulated rat. The observed value of the Student's t statistic is $T_{calc} = 4.2418$ with 11 degrees of freedom, which gives a P -value of .0007. Notice that this P -value is the probability that a Student's t variable with 11 degrees of freedom will be at least as large as $T_{calc} = 4.2418$

when $\mu_D = 0$. Since this P -value is quite small, there is very strong evidence that the population mean difference μ_D is greater than zero. Hence, the data do support the contention that for this population of pairs of rats, on average, the cortex weight of a stimulated rat would be higher than the cortex weight of an unstimulated rat.

Figure 21. Normal probability plot for weight differences.



To get a feel for the size of this population mean difference we can compute a 95% confidence interval estimate for μ_D . The sample mean difference is $\bar{D} = 31$, the sample standard error of the mean is $\widehat{S.E.}(\bar{D}) = 7.3082$, and the multiplier for the 95% margin of error, based on the Student's t distribution with 11 degrees of freedom, is $k = 2.201$. Thus the 95% margin of error for \bar{D} is $(2.201)(7.3082) = 16.0853$, and we are 95% confident that the population mean difference μ_D is between 14.9147 and 47.0853 milligrams. In other words, we are 95% confident that for this population of pairs of rats, on average, the cortex weight of the stimulated rat would exceed the cortex weight of the unstimulated rat by at least 14.9147 mg but by no more than 47.0853 mg.

The hypothesis tests we have discuss thus far are only appropriate when we have enough *a priori* information, *i.e.*, information that does not depend on the data to be used for the hypothesis test, to postulate that the population mean μ is on one side of a particular value μ_0 . That is, we have only considered situations where the research hypothesis is directional. There are situations when we will not have enough *a priori* information to allow us to choose the appropriate directional research hypothesis. Instead, we might only conjecture that the population mean μ is different from some particular value μ_0 . In a situation like this our research hypothesis specifies that the population mean μ is different from μ_0 , *i.e.*, $H_1 : \mu \neq \mu_0$.

To decide between the null hypothesis $H_0 : \mu = \mu_0$ and the research hypothesis $H_1 : \mu \neq \mu_0$, we need to decide whether the sample mean \bar{X} supports the null hypothesis

by being “close to μ_0 ”, or supports the research hypothesis by being “far away from μ_0 ”. In this situation the P -value is the probability that the sample mean \bar{X} would be as far or farther away from μ_0 in either direction as is the value that we actually observe. In other words, the P -value is the probability that the standardized distance from \bar{X} to μ_0 (the standardized absolute value of the difference between \bar{X} and μ_0) is as large or larger than the actual observed value of this standardized distance. As before, the P -value is computed under the assumption that the null hypothesis is true and $\mu = \mu_0$. In this situation the calculated t statistic T_{calc} is the absolute value of the t statistic that would be used for testing a directional hypothesis. That is, the calculated t statistic is

$$T_{calc} = \left| \frac{\bar{X} - \mu_0}{\widehat{\text{S.E.}}(\bar{X})} \right|.$$

In terms of this t statistic the P -value is the probability that the absolute value of a Student’s t variable with $n - 1$ degrees of freedom T would take on a value as large or larger than T_{calc} , assuming that $\mu = \mu_0$. This probability is the sum of the area under the appropriate Student’s t density curve to the left of $-T_{calc}$ and the area under this Student’s t density curve to the right of T_{calc} . We need to add these two areas (probabilities) since we are finding the probability that the sample mean \bar{X} would be as far or farther away from μ_0 in either direction as is the value that we actually observe, when $\mu = \mu_0$.

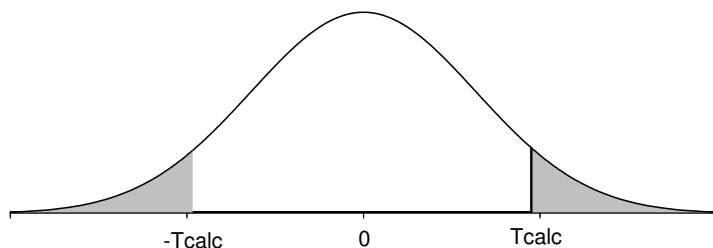
The steps for performing a hypothesis test for

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu \neq \mu_0$$

are summarized below.

1. Use a suitable calculator or computer program to find the P -value $= P(|T| \geq T_{calc}) = P(T \leq -T_{calc}) + P(T \geq T_{calc})$, where T denotes a Student’s t variable with $n - 1$ degrees of freedom and $T_{calc} = |\bar{X} - \mu_0|/\widehat{\text{S.E.}}(\bar{X})$ as described above. This P -value is the sum of the area to the left of $-T_{calc}$ and the area to the right of T_{calc} , where each area is that under the density curve for the Student’s t distribution with $n - 1$ degrees of freedom over the appropriate region on the x -axis as indicated in Figure 22.

Figure 22. P-value for $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.



- 2a. If the P -value is small enough (less than .05 for a test at the 5% level of significance), conclude that the data favor $H_1 : \mu \neq \mu_0$ over $H_0 : \mu = \mu_0$. That is, if the P -value is small enough, then there is sufficient evidence to conclude that the population mean μ is not equal to μ_0 .
- 2b. If the P -value is not small enough (is not less than .05 for a test at the 5% level of significance), conclude that the data do not favor $H_1 : \mu \neq \mu_0$ over $H_0 : \mu = \mu_0$. That is, if the P -value is not small enough, then there is not sufficient evidence to conclude that the population mean μ differs from μ_0 .

Example. Newcomb's measurements of the speed of light (revisited). The parameter we were estimating in our analysis of Newcomb's measurements of the speed of light was the population mean time μ for light to travel a distance of 7442 meters. Notice that the population mean time μ is actually defined by the process of measurement being used. That is, we can think of μ as the long run average time that we would observe if we replicated Newcomb's experiment a large times. We might wonder how this population mean time μ relates to the "true time" it would take for light to travel a distance of 7442 meters. We don't know exactly what this "true time" is, but we can use a generally accepted, modern measurement of the speed of light to obtain a hypothesized "true time." Stigler (*op. cit.*) used the modern estimate of the speed of light in a vacuum of 299,792.5 km/sec adjusted to give the speed of light in air and converted to a time as measured by Newcomb to obtain a hypothesized "true time" of 33.02. Therefore, our present goal is to determine how the population mean time μ relates to the "true time" $\mu_0 = 33.02$.

We do not have sufficient *a priori* information to specify a directional hypothesis; therefore, we will test the null hypothesis $H_0 : \mu = 33.02$ versus the research hypothesis $H_1 : \mu \neq 33.02$ to determine whether the population mean time μ is equal to or different from the hypothesized "true value" of 33.02. For Newcomb's $n = 64$ measurements the sample mean is $\bar{X} = 27.75$, the sample standard error of the mean is $\widehat{S.E.}(\bar{X}) = .6354$, and the calculated t statistic is

$$T_{calc} = \left| \frac{27.75 - 33.02}{.6354} \right| = 8.2940.$$

The P -value is less than .0001 indicating that observing a sample mean as far away from the hypothesized "true value" 33.02 (in either direction) as Newcomb did is extremely unlikely if in fact $\mu = 33.02$. We can conclude that the population mean μ corresponding to Newcomb's experiment is almost certainly not equal to the hypothesized value of 33.02. The confidence interval, (26.4803, 29.0197), we computed above suggests that the mean μ that Newcomb was estimating is less than the hypothesized value 33.02.

7.3 Inference for a population median

The inferential methods for a population mean discussed above require at least approximate normality of the population distribution of the variable of interest. In this section we will consider methods for making inferences about a population median which do not require the assumption of a particular form for the population distribution of the variable of interest. That is, the inferential methods considered in this section are applicable for any population distribution for a continuous variable regardless of the shape of the corresponding density curve.

We will begin by discussing a method for testing a hypothesis about a population median. The essence of this method is to re-express the hypothesis about the population median as a hypothesis about a related population proportion and to then use inferential methods for a population proportion to test the hypothesis about the population median.

For a continuous variable X we can think of the population median M as the point on the number line which divides the area under the corresponding population density curve into two equal areas (each of area one-half). The population median M is analogous to the sample median which divides the histogram into two equal areas. Notice that if we observe a single value of X , then the probability that we will observe a value larger than the population median M is $1/2$, *i.e.*, $P(X > M) = 1/2$, and similarly, the probability that we will observe a value smaller than M is $1/2$, *i.e.*, $P(X < M) = 1/2$.

Let M denote the population median of the distribution of the continuous variable X and consider a hypothesis relating the population median M to a particular, fixed value M_0 . We can dichotomize the population of values of X by thinking of the event “observe $X > M_0$ ” as a success and the event “observe $X < M_0$ ” as a failure. The population success proportion p corresponding to this dichotomization is $p = P(X > M_0)$, *i.e.*, p is the probability that a single value of X chosen from the infinite population of values of X will be larger than the hypothesized value M_0 . The corresponding population failure proportion is $1 - p = P(X < M_0)$.

The three possible relationships between the population median M and the particular value M_0 are readily re-expressed in terms of the corresponding population success probability $p = P(X > M_0)$. If the population median M exceeds the particular value M_0 , then, since the area or probability to the right of the population median M is $1/2$ and since M_0 is to the left of M on the number line, we must have $p = P(X > M_0) > 1/2 = P(X > M)$. Hence we see that $M > M_0$ is equivalent to $p > 1/2$. Similarly, if M is less than M_0 , then M_0 is to the right of M on the number line and we must have $p = P(X > M_0) < 1/2 = P(X > M)$; thus $M < M_0$ is equivalent to $p < 1/2$. Finally, if $M = M_0$, then we must have $p = P(X > M_0) = 1/2 = P(X > M)$; thus $M = M_0$ is equivalent to $p = 1/2$. These relationships allow us to re-express a

hypothesis relating the population median M to the particular value M_0 as a hypothesis about the corresponding population success probability $p = P(X > M_0)$.

For ease of reference the three standard pairs of null and research hypotheses about M are summarized below (recall that $p = P(X > M_0)$)

1. $H_0 : M \leq M_0$ versus $H_1 : M > M_0$ corresponds to $H_0 : p \leq .5$ versus $H_1 : p > .5$.
2. $H_0 : M \geq M_0$ versus $H_1 : M < M_0$ corresponds to $H_0 : p \geq .5$ versus $H_1 : p < .5$.
3. $H_0 : M = M_0$ versus $H_1 : M \neq M_0$ corresponds to $H_0 : p = .5$ versus $H_1 : p \neq .5$.

The dichotomy described above, where $X > M_0$ constitutes a success and $X < M_0$ constitutes a failure, does not allow for the possibility that the continuous variable X is exactly equal to M_0 . This is allowable, theoretically, when we are talking about the population distribution of X ; however, when we examine the data we may find that one or more of the observations are equal to the hypothesized value M_0 . The easiest solution to this potential difficulty is to remove any observations which are exactly equal to M_0 and adjust the sample size accordingly before we perform a hypothesis test. In other words, a hypothesis test for comparing the population median M to the hypothesized value M_0 is based on the observed proportion \hat{p} of successes (values of X which are greater than M_0) relative to n , where n is the number of observations which are not equal to M_0 . An alternative to discarding values exactly equal to M_0 is to classify half of these values as successes and half as failures and use the original sample size for n .

Example. Darwin's plant height comparison. Charles Darwin conducted an experiment to determine whether cross-fertilized plants tend to be more vigorous than self-fertilized plants. (The plants used in Darwin's study were young corn (*Zea mays*) plants.) (Darwin (1876), *The Effect of Cross- and Self-fertilization in the Vegetable Kingdom*, second edition, John Murray, London) Darwin selected several plants and fertilized several flowers on each plant; a number of flowers were cross-fertilized with pollen taken from distant plants and a number of flowers were self-fertilized with their own pollen. Seeds gathered from these flowers were allowed to ripen and were then placed in wet sand to germinate. Fifteen seedlings from cross-fertilized seeds were selected and fifteen seedlings from self-fertilized seeds were selected. These seedlings were paired (one cross-fertilized and one self-fertilized) and the two seedlings in each pair were planted on opposite sides of the same pot. After a fixed period of time the height of each plant (in inches) was recorded. The raw data and the associated differences are provided in Table 12. The distribution of

the differences is summarized in Table 13 and Figure 23 and a normal probability plot is given in Figure 24.

Table 12. Plant height data.

pair	plant height		difference
	cross-fertilized	self-fertilized	
1	23.500	17.375	6.125
2	12.000	20.375	-8.375
3	21.000	20.000	1.000
4	22.000	20.000	2.000
5	19.125	18.375	0.750
6	21.500	18.625	2.875
7	22.125	18.625	3.500
8	20.375	15.250	5.125
9	18.250	16.500	1.750
10	21.625	18.000	3.625
11	23.250	16.250	7.000
12	21.000	18.000	3.000
13	22.125	12.750	9.375
14	23.000	15.500	7.500
15	12.000	18.000	-6.000

If it is true that cross-fertilized plants tend to be more vigorous than self-fertilized plants, then we would expect a cross-fertilized plant to be taller than a self-fertilized plant of the same age. Therefore, we can formalize this theory as the research hypothesis that the population median of the difference between the height of a cross-fertilized plant and the height of a self-fertilized plant grown in the same pot M_D is greater than zero, *i.e.*, $H_1 : M_D > 0$. We can think of the population as consisting of all of the pairs of seedlings (one cross-fertilized and one self-fertilized) which could have been used in this experiment; and, we can think of the population median difference M_D as the median of the differences between the heights of these pairs of seedlings (height of the cross-fertilized plant minus height of the self-fertilized plant). Notice that this research hypothesis specifies that, for this population of potential pairs of plants, the cross-fertilized plant would be taller than the self-fertilized plant more than half of the time.

Table 13. Summary statistics for the plant height differences.

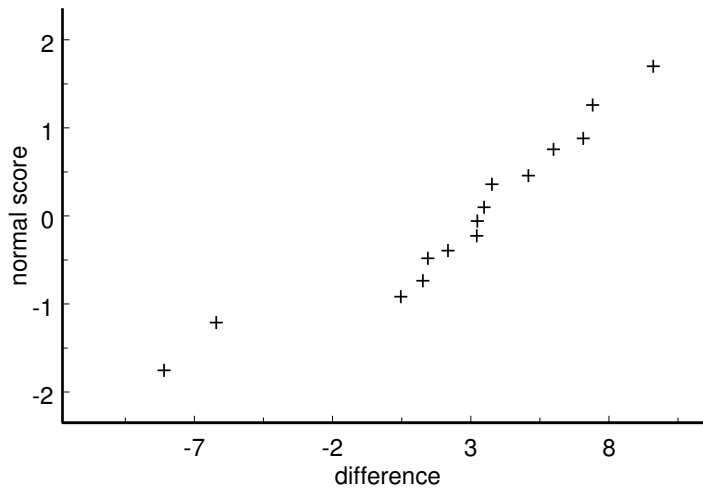
minimum	-8.375	Q_1 - minimum	9.375
Q_1	1.000	median - Q_1	4.000
median	3.000	Q_3 - median	3.125
Q_3	6.125	maximum - Q_3	3.250
maximum	9.375		

Figure 23. Stem and leaf histogram for the plant height difference data (rounded).

In this stem and leaf histogram the stem represents tens of inches and the leaf represents inches.

stem	leaf
-0	8
-0	6
-0	
-0	
-0	
0	01
0	2233
0	445
0	67
0	89

Figure 24. Normal probability plot for height differences.



The normality assumption is questionable in this example since the stem and leaf histogram and the normal probability plot show some evidence of extreme skewness to the left in this sample of differences.

Since all but two of the 15 plant height differences are positive, there appears to be strong evidence that the population median height difference is positive. For these data there are no zero differences so we use the actual sample size $n = 15$ to perform our test of $H_0 : M_D \leq 0$ versus $H_1 : M_D > 0$ ($H_0 : p \leq .5$ versus $H_1 : p > .5$, where $p = P(D > 0)$ denotes the proportion of this population of pairs of plants for which the cross-fertilized plant would be taller than the self-fertilized plant). Thirteen of the differences are positive (successes) which gives $\hat{p} = 13/15 = .8667$. The standard error of \hat{p} , assuming that $p = .5$,

is $S.E.(\hat{p}) = .1291$, the calculated Z -statistic is $Z_{calc} = 2.8402$, and the P -value is .0023. Therefore, there is very strong evidence that the population median height difference M_D is greater than zero which means that there is very strong evidence in support of the contention that, for this population of pairs of plants, the cross-fertilized plant would be taller than the self-fertilized plant more than half of the time.

We can construct a 95% confidence interval estimate for the population median M by finding the interval of values of M_0 for which a test at the 5% level of significance **does not** lead to the rejection of $H_0 : M = M_0$. Recall that $H_0 : M = M_0$ and $H_1 : M \neq M_0$ are equivalent to $H_0 : p = .5$ and $H_1 : p \neq .5$, where $p = P(X > M_0)$. A test at the 5% level of significance will fail to reject the null hypothesis $H_0 : M = M_0$ if

$$|\hat{p} - .5| \leq 1.96/(2\sqrt{n}),$$

where \hat{p} is the observed proportion of values of X which are greater than M_0 . The lower and upper limiting values of \hat{p} in this expression, denoted by \hat{p}_L and \hat{p}_U , are

$$\hat{p}_L = .5 - 1.96/(2\sqrt{n}) \quad \text{and} \quad \hat{p}_U = .5 + 1.96/(2\sqrt{n}).$$

The corresponding limits on the value of M_0 , denoted by M_L and M_U , are the \hat{p}_L 100 percentile and the \hat{p}_U 100 percentile of the observed values of X , respectively. We will adopt the rounding convention described below to avoid the need for averaging observed values of X when computing M_L and M_U . First convert \hat{p}_L and \hat{p}_U from proportions to counts by multiplying each by the sample size n . If the count $n\hat{p}_L$ is not a whole number, round it down to the next whole number. If the count $n\hat{p}_U$ is not a whole number, round it up to the next whole number. Finally, find the observed values of X which occur at these locations in an ordered listing of the observed values. As is true when counting to find a sample median or quartile be sure to list any repeated values as many times as they occur. The resulting values for M_L and M_U form the endpoints of our 95% confidence interval estimate for the population median M .

The procedure for calculating the 95% confidence interval estimate for a population median M is summarized below.

1. Arrange the data (observations) in increasing order from smallest (obs. no. 1) to largest (obs. no. n). Be sure to include all n values in this list, including repeats if there are any.
2. Compute the quantity $n\hat{p}_L$ and round it down to the nearest whole number if it is not a whole number. The observation at the location indicated by the rounded-down value in the ordered listing of the data is M_L .

3. Compute the quantity $n\hat{p}_U$ and round it up to the nearest whole number if it is not a whole number. The observation at the location indicated by the rounded-up value in the ordered listing of the data is M_U .
4. Conclude that we are 95% confident that the population median M is between M_L and M_U .

Example. Darwin's plant height comparison (revisited). For this example we have $1.96/(2\sqrt{15}) = .2530$, $\hat{p}_L = .5 - .2530 = .2470$, and $\hat{p}_U = .5 + .2530 = .7530$, giving $n\hat{p}_L = 3.705$ which we round down to 3 and $n\hat{p}_U = 11.295$ which we round up to 12. Observation number 3 is .75 and observation number 12 is 6.125. Therefore, we are 95% confident that the population median height difference for this population of pairs of plants is between .75 inches and 6.125 inches.

7.4 Summary

The majority of this chapter is devoted to inference for the population mean μ of the distribution of a continuous variable X . We began by discussing probability models for the distribution of a continuous variable (density curves) and then introduced the normal distribution which serves as the basis for our inferences about μ (based on the Student's t distribution).

Given data which form a random sample of size n from a population with population mean μ and population standard deviation σ , the sampling distribution of the sample mean \bar{X} has mean μ and the population standard error of \bar{X} is $\text{S.E.}(\bar{X}) = \sigma/n$. Thus the sample mean \bar{X} is unbiased as an estimator of the population mean μ and the variability in the sample mean \bar{X} as an estimator of the population mean μ can be quantified by this standard error. If we also assume that the population distribution of X is a normal distribution, *i.e.*, if we assume that the data form a random sample of size n from a normal distribution with population mean μ and population standard deviation σ , then the sampling distribution of \bar{X} is the normal distribution with population mean μ (the same as that of X) and standard deviation $\text{S.E.}(\bar{X})$.

Given data which form a random sample of size n from a normal distribution with population mean μ and population standard deviation σ and with sample mean \bar{X} and sample standard deviation $S_X = S$, the quantity

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows the Student's t distribution with $n - 1$ degrees of freedom. Therefore, if the assumption that the population distribution of X is normal is reasonable, then we can use the Student's t distribution to make inferences about the population mean μ . It is important to remember the normality assumption needed for the Student's t distribution

and to verify that this assumption is reasonable by examining the data for violations of this assumption. The Student's t methods work reasonably well provided the normality assumption is not totally unreasonable.

The interval from $\bar{X} - kS/\sqrt{n}$ to $\bar{X} + kS/\sqrt{n}$, where k denotes the 97.5 percentile of the Student's t distribution with $n - 1$ degrees of freedom, is a 95% confidence interval for μ . We can test a hypothesis relating μ to a specified value μ_0 by using the Student's t test statistic

$$T_{calc} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

to find the appropriate P -value. The P -value for $H_1 : \mu > \mu_0$ is the probability $P(T \geq T_{calc})$; the P -value for $H_1 : \mu < \mu_0$ is the probability $P(T \leq T_{calc})$; and, the P -value for $H_1 : \mu \neq \mu_0$ is the probability $P(|T| \geq |T_{calc}|)$, where T denotes a Student's t variable with $n - 1$ degrees of freedom, *i.e.*, these P -values are areas under the density curve of the Student's t distribution with $n - 1$ degrees of freedom.

The Student's t inferential methods for a population mean are based on the assumption that the underlying population distribution is reasonably modeled by a normal distribution. When this normality assumption is not tenable we need to consider a method of inference which is applicable under weaker assumptions. One approach to inference about the center of a distribution based on the population median is discussed in Section 7.4. This approach to inference about the population median does not require the assumption of a specific form for the underlying population distribution.

7.5 Exercises

Provide a complete analysis for the following example. Be sure to: define a relevant population mean μ ; setup and perform a relevant hypothesis test; and, find a confidence interval for μ . Be sure to include comments regarding the validity of the normality assumption for this example. Provide a complete summary of your findings in the context of the example.

1. An article by Rosner, Willett, and Spiegelman in *Statistics in Medicine* (1989) describes a study conducted to assess the make up of the diets of a population of women. The data used here are as reported in Ott and Longnecker (2002). A sample of $n = 168$ women was obtained and each of these women completed a food frequency questionnaire. The completed questionnaires were then used to determine the percentage of calories from fat in each woman's diet. The values of the variable X , the percentage of calories from fat in a woman's diet, are summarized in the stem and leaf histogram in Figure 25. For display purposes the data have been rounded to the nearest 1 percent. The actual data are given in Table 14. In 2002 the Food and Nutrition Board, a unit of the Institute of Medicine iom.edu, recommended that adults should restrict the percentage of calories from fat in their diet to the range from 20% to 35%.

If you wish to analyze these data without entering all $n = 168$ values in your computer or calculator you may use the facts that: the sample mean for these data is 36.91899 and the sample standard deviation is 6.72820.

Figure 25. Stem and leaf histogram for percentage of calories from fat.

In this stem and leaf histogram the stem represents tens and the leaf represents ones. For example, the smallest value is 16% and the largest is 58%.

stem	leaf
1	
1	6
2	014
2	55666677888999999
3	0000011111112222222222333333334444444444
3	5555555555556666666666667777778888888888999999999
4	00000000011111111122222222223344444444
4	555666667888
5	00134
5	68

Table 14. Percentage of calories from fat data.

15.92	20.22	20.80	24.06	24.98	25.07	25.53	26.16	26.18	26.27
26.79	27.29	27.54	28.16	28.35	28.76	29.07	29.34	29.34	29.45
29.46	29.55	29.72	29.94	29.99	30.50	30.61	30.71	30.96	30.99
31.02	31.10	31.27	31.61	31.71	31.75	31.97	32.24	32.26	32.28
32.43	32.73	32.86	32.96	33.08	33.11	33.13	33.23	33.32	33.51
33.71	33.72	33.86	33.87	33.95	34.11	34.17	34.25	34.31	34.51
34.86	34.87	34.87	34.89	35.09	35.18	35.18	35.20	35.29	35.32
35.49	35.50	35.56	35.67	35.69	35.71	35.74	35.77	36.07	36.07
36.30	36.32	36.32	36.37	36.58	37.04	37.10	37.14	37.34	37.39
37.47	37.56	37.85	37.88	37.89	38.02	38.04	38.10	38.19	38.21
38.36	38.41	38.45	38.58	38.81	38.88	38.89	38.97	39.13	39.22
39.25	39.40	39.80	39.84	39.95	40.05	40.12	40.25	40.29	40.46
40.48	40.69	41.12	41.15	41.22	41.29	41.30	41.32	41.41	41.42
41.44	41.53	41.64	41.69	41.78	41.89	42.12	42.17	42.20	42.37
42.60	42.98	43.57	43.74	43.79	44.27	44.28	44.33	44.39	44.66
45.06	45.28	45.51	45.82	45.83	46.22	46.38	46.97	47.63	47.83
48.29	49.65	49.86	50.72	53.08	54.05	55.54	57.85		

2. Cox and Snell (1981), *Applied Statistics*, Chapman and Hall, discuss a study conducted to examine changes in blood pressure due to the drug captopril. The original source is MacGregor, Markandu, Roulston, and Jones (1979), Essential hypertension: effect of an oral inhibitor of angiotension-converting enzyme, *British Medical Journal*, **2** 1106–1109. The data given in Table 15 are the blood pressures (in mm Hg) for 15 patients with moderate essential hypertension. The data consist of supine systolic and diastolic blood pressures measured immediately before and two hours after taking 25 mg of the drug captopril. Relevant differences are also provided.

Table 15. Blood pressure data.

patient	systolic			diastolic		
	before	after	difference	before	after	difference
1	210	201	9	130	125	5
2	169	165	4	122	121	1
3	187	166	21	124	121	3
4	160	157	3	104	106	-2
5	167	147	20	112	101	11
6	176	145	31	101	85	16
7	185	168	17	121	98	23
8	206	180	26	124	105	19
9	173	147	26	115	103	12
10	146	136	10	102	98	4
11	174	151	23	98	90	8
12	201	168	33	119	98	21
13	198	179	19	106	110	-4
14	148	129	19	107	103	4
15	154	131	23	100	82	18