



COLUMBIA
BUSINESS
SCHOOL

Estimation and Confidence Intervals

Fall 2001
B6014: Managerial Statistics

Professor Paul Glasserman
403 Uris Hall

Properties of Point Estimates

1. We have already encountered two **point estimators**: the sample mean \bar{X} is an estimator of the population mean μ , and the sample proportion \hat{p} is an estimator of the population proportion p . These are called point estimates in contrast to **interval estimates**. A point estimate is a single number (such as \bar{X} or \hat{p}) whereas an interval estimate gives a **range** likely to contain the true value of the unknown parameter.
2. What makes one estimator better than another? Is there some sense in which \bar{X} and \hat{p} are the best possible estimates of μ and p , given the available data? We briefly discuss the relevant considerations in addressing these questions.
3. There are two distinct considerations to keep in mind in evaluating an estimator: **bias** and variability. Variability is measured by the standard error of the estimator, and we have encountered this already a couple of times. We have not explicitly discussed bias previously, so we focus on it now.
4. In everyday usage, *bias* has many meanings, but in statistics it has just one: the bias of an estimator is the difference between its expected value and the quantity being estimated:

$$\text{Bias}(\text{estimator}) = E[\text{estimator}] - \text{quantity being estimated.}$$

An estimator is called **unbiased** if this difference is zero; i.e., if

$$E[\text{estimator}] = \text{quantity being estimated.}$$

5. The intuitive content of this is that an unbiased estimator is one that does not systematically overestimate or underestimate the target quantity.

- We have already met two unbiased estimators: the sample mean \bar{X} is an unbiased estimator of the population mean μ because $E[\bar{X}] = \mu$; the sample proportion \hat{p} is an unbiased estimator of the population proportion p because $E[\hat{p}] = p$. Each of these estimators is correct “on average,” in the sense that neither systematically overestimates or underestimates.
- We have discussed estimation of a population mean μ but not estimation of a population variance σ^2 . We estimate σ^2 from a random sample X_1, \dots, X_n using the **sample variance**

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (1)$$

It is a (non-obvious) mathematical fact that $E[s^2] = \sigma^2$; i.e., the sample variance is unbiased. That’s why we divide by $n-1$; if we divided by n , we would get a slightly smaller estimate and indeed one that is biased low. Notice, however, that even if we divided by n , the bias would vanish as n becomes large because $(n-1)/n$ approaches 1 as n increases.

- The **sample standard deviation**

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

is not an unbiased estimator of the population standard deviation σ . It is biased low, because $E[s] < \sigma$ (this isn’t obvious, but it’s true). However, the bias vanishes as the sample size increases, so we use this estimator all the time.

- The preceding example shows that a biased estimator is not necessarily a bad estimator, so long as the bias is small and disappears as the sample size increases.

Confidence Intervals for the Population Mean

- We have seen that the sample mean \bar{X} is an unbiased estimator of the population mean μ . This means that \bar{X} is accurate, on average; but of course for any particular data set X_1, \dots, X_n , the sample mean may be higher or lower than the true value μ . The purpose of a **confidence interval** is to supplement the point estimate \bar{X} with information about the uncertainty in this estimate.
- A confidence interval for μ takes the form $\bar{X} \pm \Delta$, with the range Δ chosen so that there is, e.g., a 95% probability that the error $\bar{X} - \mu$ is less than $\pm\Delta$; i.e.,

$$P(-\Delta < \bar{X} - \mu < \Delta) = .95.$$

We know that $\bar{X} - \mu$ is approximately normal with mean 0 and standard deviation σ/\sqrt{n} . So, to capture the middle 95% of the distribution we should set

$$\Delta = 1.96 \frac{\sigma}{\sqrt{n}}.$$

We can now say that we are 95% confident that the true mean lies in the interval

$$\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right).$$

This is a **95% confidence interval** for μ .

3. In deriving this interval, we assumed that \bar{X} is approximately normal $N(\mu, \sigma^2/n)$. This is valid if the underlying population (from which X_1, \dots, X_n are drawn) is normal or else if the sample size is sufficiently large (say $n \geq 30$).

4. Example: Apple Tree Supermarkets is considering opening a new store at a certain location, but wants to know if average weekly sales will reach \$250,000. Apple Tree estimates weekly gross sales at nearby stores by sending field workers to collect observations. The field workers collect 40 samples and arrive at a sample mean of \$263,590. Suppose the standard deviation of these samples is \$42,000. Find a 90% confidence interval for μ .

If we ask for a 90% confidence interval, we should use 1.645 instead of 1.96 as the multiplier. This interval is thus given by

$$(263590 - 1.645(42000/\sqrt{40}), 263590 + 1.645(42000/\sqrt{40}))$$

which works out to be $263,590 \pm 10,924$. Since \$250,000 falls outside the confidence interval, Apple Tree can be reasonably confident that their minimum cut-off is met.

5. Statistical interpretation of a confidence interval: Suppose we repeated this sampling experiment 100 times; that is, we collected 100 different sets of data, each set consisting of 40 observations. Suppose that we computed a confidence interval based on each of the 100 data sets. On average, we would expect 90 of the confidence intervals to include the true mean μ ; we would expect that 10 would not. The figure 90 comes from the fact that we chose a 90% confidence interval.
6. More generally, we can choose whatever **confidence level** we want. The convention is to specify the confidence level as $1 - \alpha$, where α is typically 0.1, 0.05 or 0.01. These three α values correspond to confidence levels 90%, 95% and 99%. (α is the Greek letter alpha.)
7. Definition: For any α between 0 and 1, we define z_α to be the point on the z -axis such that the area to the right of z_α under the standard normal curve is α ; i.e.,

$$P(Z > z_\alpha) = \alpha.$$

See Figure 1.

8. Examples: If $\alpha = .05$, then $z_\alpha = 1.645$; if $\alpha = .025$, then $z_\alpha = 1.96$; if $\alpha = .01$, then $z_\alpha = 2.33$. These values are found by looking up $1 - \alpha$ in the body of the normal table.
9. We can now generalize our earlier 95% confidence interval to any level of confidence $1 - \alpha$:

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

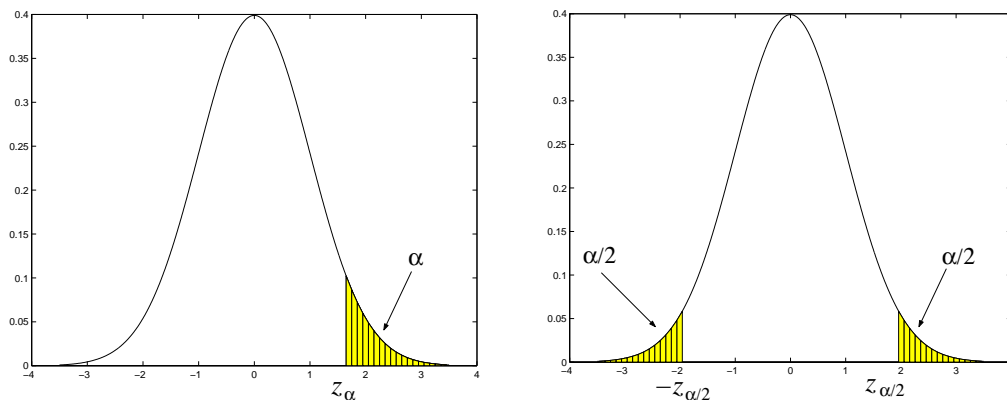


Figure 1: The area to the right of z_α is α . For example, $z_{.05}$ is 1.645. The area outside $\pm z_{\alpha/2}$ is $\alpha/2 + \alpha/2 = \alpha$. For example, $z_{.025} = 1.96$ so the area to the right of 1.96 is 0.025, the area to the left of -1.96 is also 0.025, and the area outside ± 1.96 is .05.

10. Why $z_{\alpha/2}$, rather than z_α ? We want to make sure that the total area outside the interval is α . This means that $\alpha/2$ should be to the left of the interval and $\alpha/2$ should be to the right. In the special case of a 90% confidence interval, $\alpha = 0.1$, so $\alpha/2 = 0.05$, and $z_{.05}$ is indeed 1.645.
11. Here is a table of commonly used confidence levels and their multipliers:

Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
90%	0.10	0.05	1.645
95%	0.05	0.025	1.96
99%	0.01	0.005	2.58

Unknown Variance, Large Sample

1. The expression given above for a $(1 - \alpha)$ confidence interval assumes that we know σ^2 , the population variance. In practice, if we don't know μ we generally don't know σ either. So, this formula is not quite ready for immediate use.
2. When we don't know σ , we replace it with an estimate. In equation (1) we have an estimate for the population variance. We get an estimate of the population standard deviation by taking the square root:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

This is the **sample standard deviation**. A somewhat more convenient formula for computation is

$$s = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)}.$$

In Excel, this formula is evaluated by the function STDEV.

3. As the sample size n increases, s approaches the true value σ . So, for a large sample size ($n \geq 30$, say),

$$\frac{\bar{X} - \mu}{s/\sqrt{n}}$$

is approximately normally distributed $N(0, 1)$. Arguing as before, we find that

$$\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

provides an approximate $1 - \alpha$ confidence level for μ . We say approximate because we have s rather than σ .

4. Example: Consider the results of the Apple Tree survey given above, but suppose now that σ is unknown. Suppose the sample standard deviation of the 40 observations collected is found to be 38900. Then we arrive at the following (approximate) 90% confidence interval for μ :

$$(263590 - 1.645(38900/\sqrt{40}), 263590 + 1.645(38900/\sqrt{40})).$$

The sample mean, the sample size, and the $z_{\alpha/2}$ are just as before; the only difference is that we have replaced the true standard deviation with an estimate.

5. Example: Suppose that Apple Tree also wants to estimate the number of customers per day at nearby stores. Suppose that based on 50 observations they obtain a sample mean of 876 and a sample standard deviation of 139. Construct a 95% confidence interval for the mean number of customers per day.
6. Solution: The confidence level is now 95%. This means that $\alpha = .05$ and $\alpha/2 = .025$. So we would look in the normal table to find $1 - .025 = .9750$. This corresponds to a z value of 1.96; i.e., $z_{\alpha/2} = z_{.025} = 1.96$. The rest is easy; our confidence interval is

$$\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = 876 \pm (1.96)(139/\sqrt{50}).$$

7. The expression

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad z_{\alpha/2} \frac{s}{\sqrt{n}}$$

is called the **halfwidth** of the confidence interval or the **margin of error**. The halfwidth is a measure of precision; the tighter the interval, the more precise our estimate. Not surprisingly, the halfwidth

- decreases as the sample size increases;
- increases as the population standard deviation increases;
- increases as the confidence level increases (higher confidence requires larger $z_{\alpha/2}$).

Unknown Variance, Normal Data

1. The confidence interval given above (with σ replaced by s) is based on the approximation

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \approx N(0, 1).$$

In statistical terminology, we have approximated the **sampling distribution** of the ratio on the left by the standard normal. To the extent that we can improve on this approximation, we can get more reliable confidence intervals.

2. The **exact** sampling distribution is known when the underlying data X_1, X_2, \dots, X_n comes from a normal population $N(\mu, \sigma^2)$. Recall that in this case, the distribution

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1),$$

is exact, **regardless of the sample size** n . However, we are no longer dividing by a constant σ ; instead, we are dividing by the random variable s . This changes the distribution. As a result, the sample statistic

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

is not (exactly) normally distributed; rather, it follows what is known as the **Student t -distribution**.

3. The t -distribution looks a bit like the normal distribution, but it has heavier tails, reflecting the added variability that comes from estimating σ rather than knowing it.
4. “Student” refers to the pseudonym of W.S. Gosset, not to you. Gosset identified this distribution while working on quality control problems at the Guinness brewery in Dublin.
5. Once we know the exact sampling distribution of $(\bar{X} - \mu)/(s/\sqrt{n})$, we can use this knowledge to get exact confidence intervals. To do so, we need to know something about the t -distribution. It has a parameter called the **degrees of freedom** and usually denoted by ν , the Greek letter nu. (This is a parameter in the same way that μ and σ^2 are parameters of the normal.) Just as Z represents a standard normal random variable, the symbol t_ν represents a t random variable with ν degrees of freedom (df).
6. If X_1, \dots, X_n is a random sample from $N(\mu, \sigma^2)$, then the random variable

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

has the t -distribution with $n - 1$ degrees of freedom, t_{n-1} . Thus, the df is one less than the sample size.

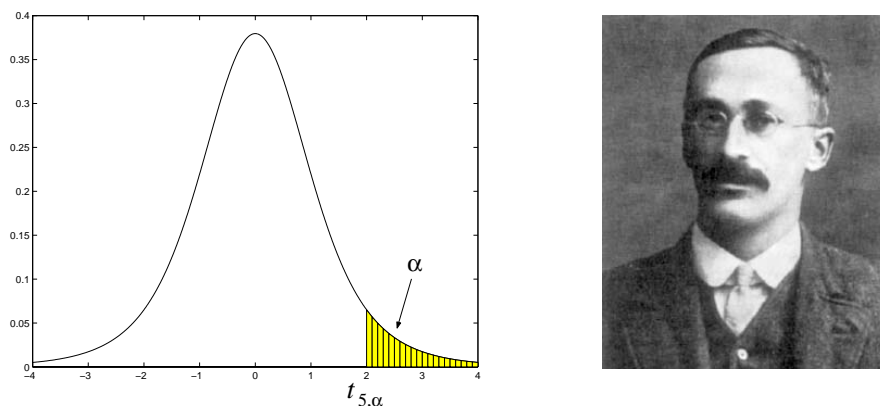


Figure 2: The area to the right of $t_{\nu,\alpha}$ under the t_ν density is α . The figure shows t_5 , the t distribution with five degrees of freedom. At $\alpha = 0.05$, $t_{5,.05} = 2.015$ whereas $z_{.05} = 1.645$. The tails of this distribution are noticeably heavier than those of the standard normal; see Figure 1. Right panel shows a picture of “Student” (W.S. Gosset).

7. Just as we defined the cutoff z_α by the condition $P(Z > z_\alpha) = \alpha$, we define a cutoff $t_{\nu,\alpha}$ for the t distribution by the condition

$$P(t_\nu > t_{\nu,\alpha}) = \alpha.$$

In other words, the area to the right of $t_{\nu,\alpha}$ under the t_ν density is α . See Figure 2.

Note that t_ν is a random variable but $t_{\nu,\alpha}$ is a number (a constant).

8. Values of $t_{\nu,\alpha}$ are tabulated in a t table. See the last page of these notes. Different rows correspond to different df; different columns correspond to different α 's.
9. Example: $t_{20,.05} = 1.725$; we find this by looking across the row labeled 20 and under the column labeled 0.05. This tells us that the area to the right of 1.725 under the t_{20} distribution is 0.05. Similarly, $t_{20,.025} = 2.086$ so the area to the right of 2.086 is 0.025.
10. We are now ready to use the t -distribution for a confidence interval. Suppose, once again, that the underlying data X_1, \dots, X_n is a random sample from a normal population. Then

$$\bar{X} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

is a $1 - \alpha$ confidence interval for the mean. Notice that all we have done is replace $z_{\alpha/2}$ with $t_{n-1,\alpha/2}$.

11. For any sample size n and any confidence level $1 - \alpha$, we have $t_{n-1,\alpha/2} > z_{\alpha/2}$. Consequently, intervals based on the t distribution are always wider than those based on the standard normal.
12. As the sample size increases, the df increases. As the df increases, the t distribution becomes the normal distribution. For example, we know that $z_{.05} = 1.645$. Look down the .05 column of the t table. $t_{\nu,.05}$ approaches 1.645 as ν increases.

13. Strictly speaking, intervals based on the t distribution are valid only if the underlying population is normal (or at least close to normal). If the underlying population is decidedly non-normal, the only way we can get a confidence interval is if the sample size is sufficiently large that \bar{X} is approximately normal (by the central limit theorem).
14. The main benefit of t -based intervals is that they account for the additional uncertainty that comes from using the sample standard deviation s in place of the (unknown) population standard deviation σ .
15. When to use t , when to use z ? Strictly speaking, the conditions are as follows:
 - z -based confidence intervals are valid if we have a large sample;
 - t -based confidence intervals are valid if we have a sample from a normal distribution with an unknown variance.

Notice that both of these may apply: we could have a large sample from a normal population with unknown variance. In this case, both are valid and give nearly the same results. (The t and z multipliers are close when the df is large.) As a practical matter, t -based confidence intervals are often used even if the underlying population is not known to be normal. This is conservative in the sense that t -based intervals are always wider than z -based intervals.

Confidence Intervals for a Proportion

1. Example: To determine whether the location for a new store is a “family” neighborhood, Apple Tree Supermarkets wants to determine the proportion of apartments in the vicinity that have 3 bedrooms or more. Field workers randomly sample 100 apartments. Suppose they find that 26.4% have 3 bedrooms or more. We know that our best guess of the (unknown) underlying population proportion is the sample proportion $\hat{p} = .264$. Can we supplement this point estimate with a confidence interval?
2. We know that if the true population proportion is p and the sample size is large, then

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right).$$

As a result,

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0, 1).$$

So a $1 - \alpha$ confidence interval is provided by

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}.$$

3. The problem, of course, is that we don't know p , so we don't know $\sqrt{p(1-p)/n}$. The best we can do is replace p by our estimate \hat{p} . This gives us the approximate confidence interval

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

4. Returning to the Apple Tree example, a 95% confidence interval for the proportion is

$$0.264 \pm (1.96) \sqrt{\frac{(.264)(1-.264)}{100}}.$$

As before, the 1.96 comes from the fact that $P(Z < 1.96) = .975$, so $P(-1.96 < Z < 1.96) = .95$.

Comparing Two Means

- Often, we are interested in comparing two means, rather than just evaluating them separately:
 - Is the expected return from a managed mutual fund greater than the expected return from an index fund?
 - Are purchases of frozen meals greater among regular television viewers than among non-viewers?
 - Does playing music in the office increase worker productivity?
 - Do students with technical backgrounds score higher on statistics exams?
- In each case, the natural way to address the question is to estimate the two means in question and determine which is greater. But this is not enough to draw reliable conclusions. Because of variability in sampling, one sample mean may come out larger than another even if the corresponding population mean is smaller. To indicate whether our comparison is reliable, we need a **confidence interval for a difference of two population means**.
- We consider two settings: **matched pairs** and **independent samples**. The latter is easy to understand; it applies whenever the samples collected from the two populations are independent. The former applies when each observation in one population is paired to one in the other so that independence fails.
- Examples of matched pairs:
 - If we compare two mutual funds over 30 periods of 6 months each, then the performance of the two funds within a single period is generally not independent, because both funds will be affected by the same world events. If we let X_1, \dots, X_{30} be the returns on one fund and Y_1, \dots, Y_{30} the returns on another, then X_i and Y_i are generally not independent. However, the differences $X_1 - Y_1, \dots, X_{30} - Y_{30}$ should be close to independent.

- Consider an experiment in which a software firm plays music in its offices for one week. Suppose there are 30 programmers in the firm. Let X_i be the number of lines of code produced by the i -th programmer in a week without music and let Y_i be the number of lines for the same programmer in a week with music. We cannot assume that X_i and Y_i are independent. A programmer who is exceptionally fast without music will probably continue to be fast with music. We might, however, assume that the changes $X_1 - Y_1, \dots, X_{30} - Y_{30}$ are independent; i.e., the effect of music on one programmer should not depend on the effects on other programmer.
5. Getting a confidence interval for the matched case is easy. Let X_1, \dots, X_n be a random sample from a population with mean μ_X and let Y_1, \dots, Y_n be matched samples from a population with mean μ_Y . We want a confidence interval for $\mu_X - \mu_Y$. To get this, we simply compute the paired differences $X_1 - Y_1, \dots, X_n - Y_n$. This gives us a new random sample of size n . Our point estimator $\mu_X - \mu_Y$ is the average difference

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i) = \bar{X} - \bar{Y}.$$

Similarly, the sample standard deviation of the differences is

$$s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - Y_i - \bar{d})^2}.$$

If the sample size is large, a $1 - \alpha$ confidence interval for $\mu_X - \mu_Y$ is

$$\bar{d} \pm z_{\alpha/2} \frac{s_d}{\sqrt{n}}.$$

If the underlying populations are normal, then regardless of the sample size an exact confidence interval is

$$\bar{d} \pm t_{n-1, \alpha/2} \frac{s_d}{\sqrt{n}}.$$

6. Example: Suppose we monitor 30 computer programmers with and without music. Suppose the sample average number of lines of code produced without music is 195 and that with music the sample average is 185. Does music affect the programmers? Suppose the sample standard deviation of the differences s_d is 32. Then a 95% confidence interval for the difference is

$$(195 - 185) \pm (1.96) \frac{32}{\sqrt{30}} = 10 \pm 11.45.$$

Since this interval includes zero, we cannot be 95% confident that productivity is affected by the music.

7. We now turn to the case of comparing independent samples. Let X_1, \dots, X_{n_1} be a random sample from a population with mean μ_X and variance σ_X^2 . Define Y_1, \dots, Y_{n_2} and μ_Y and σ_Y^2 analogously. We do not assume that the sample sizes n_1 and n_2 are equal. Let \bar{X} and \bar{Y} be the corresponding sample means.

8. We know that $E[\bar{X} - \bar{Y}] = E[\bar{X}] - E[\bar{Y}] = \mu_X - \mu_Y$. Thus, the difference of the sample means is an unbiased estimator of the difference of the population means. To supplement this estimator with a confidence interval, we need the standard error of this estimator. We know that

$$\text{Var}[\bar{X} - \bar{Y}] = \text{Var}[\bar{X}] + \text{Var}[\bar{Y}],$$

by the assumption of independence. This, in turn, is given by

$$\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}.$$

So, the standard error of $\bar{X} - \bar{Y}$ is

$$\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}.$$

9. With this information in place, the confidence interval for the difference follows the usual pattern. If the sample size is sufficiently large that the normal approximation is valid, a $1 - \alpha$ confidence interval is

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}.$$

10. Of course, if we don't know the standard deviations σ_X and σ_Y , we replace them with estimates s_X and s_Y .
11. If the underlying populations are normal and we use sample standard deviations, we can improve the accuracy of our intervals by using the t distribution. The only slightly tricky step is calculating the degrees of freedom in this setting. The rule is this: Calculate the quantity

$$\frac{(s_X^2/n_1 + s_Y^2/n_2)^2}{(s_X^2/n_1)^2/(n_1 - 1) + (s_Y^2/n_2)^2/(n_2 - 1)}.$$

Round it off to the nearest integer \bar{n} . Then \bar{n} is the appropriate df. The confidence interval becomes

$$(\bar{X} - \bar{Y}) \pm t_{\bar{n}, \alpha/2} \sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}.$$

(We will not cover this case and it will not be on the final exam. Excel does this calculation automatically; see the Excel help on t-tests.)

Comparing Two Proportions

1. We have seen how to get confidence intervals for the difference of two means. We now turn to confidence intervals for the difference of two proportions.

2. Examples:

- 100 people participate in a study to evaluate a new anti-dandruff shampoo. Fifty are given the new formula and fifty are given an existing product. Suppose 38 show improvement with the new formula and only 30 show improvement with the existing product. How much better is the new formula?
 - One vendor delivers 5 shipments out of 33 late. Another vendor delivers 7 out of 52 late. How much more reliable is the second vendor?
3. The general setting is this: Consider two proportions p_X and p_Y , as in the examples above. Suppose we observe n_X outcomes of the first type and n_Y outcomes of the second type to get sample proportions \hat{p}_X and \hat{p}_Y . Our point estimate of the difference $p_X - p_Y$ is the $\hat{p}_X - \hat{p}_Y$. This estimate is unbiased because \hat{p}_X and \hat{p}_Y are unbiased. The variance of $\hat{p}_X - \hat{p}_Y$ is

$$\frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y},$$

the sum of the variances. So, the standard error is

$$\sqrt{\frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}}.$$

Since we don't know p_X and p_Y , we replace them with their estimates \hat{p}_X and \hat{p}_Y . The result is the confidence interval

$$(\hat{p}_X - \hat{p}_Y) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}},$$

valid for large samples.

4. You can easily substitute the corresponding values from the shampoo and vendor examples to compute confidence intervals for those differences. In the shampoo example, the sample proportions are $38/50 = .76$ and $30/50 = .60$. A 95% confidence interval for the difference is therefore

$$(.76 - .60) \pm \sqrt{\frac{.76(1-.76)}{50} + \frac{.60(1-.60)}{50}}$$

which is

$$.16 \pm .18.$$

Summary of Confidence Intervals

1. Population mean, large sample:

$$\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

2. Population mean, normal data with unknown variance:

$$\bar{X} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$$

3. Difference of two means, independent samples:

$$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

4. Difference of two means, matched pairs:

$$\bar{X} - \bar{Y} \pm z_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

$$\bar{X} - \bar{Y} \pm t_{n-1, \alpha/2} \frac{s_d}{\sqrt{n}}$$

5. Population proportion, large sample:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

6. Difference of two population proportions, independent large samples:

$$\hat{p}_X - \hat{p}_Y \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_X(1 - \hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1 - \hat{p}_Y)}{n_Y}}$$

Cumulative Probabilities for the Standard Normal Distribution.
This table gives probabilities to the left of given z values for the standard normal distribution.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Percentiles of the t Distribution
This table gives values with specified probabilities to the right of them for the specified df

Degrees of freedom	0.10	0.05	0.025	0.01
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602
16	1.337	1.746	2.120	2.583
17	1.333	1.740	2.110	2.567
18	1.330	1.734	2.101	2.552
19	1.328	1.729	2.093	2.539
20	1.325	1.725	2.086	2.528
21	1.323	1.721	2.080	2.518
22	1.321	1.717	2.074	2.508
23	1.319	1.714	2.069	2.500
24	1.318	1.711	2.064	2.492
25	1.316	1.708	2.060	2.485
26	1.315	1.706	2.056	2.479
27	1.314	1.703	2.052	2.473
28	1.313	1.701	2.048	2.467
29	1.311	1.699	2.045	2.462
30	1.310	1.697	2.042	2.457
40	1.303	1.684	2.021	2.423
60	1.296	1.671	2.000	2.390
120	1.289	1.658	1.980	2.358
z, alpha	1.282	1.645	1.960	2.326