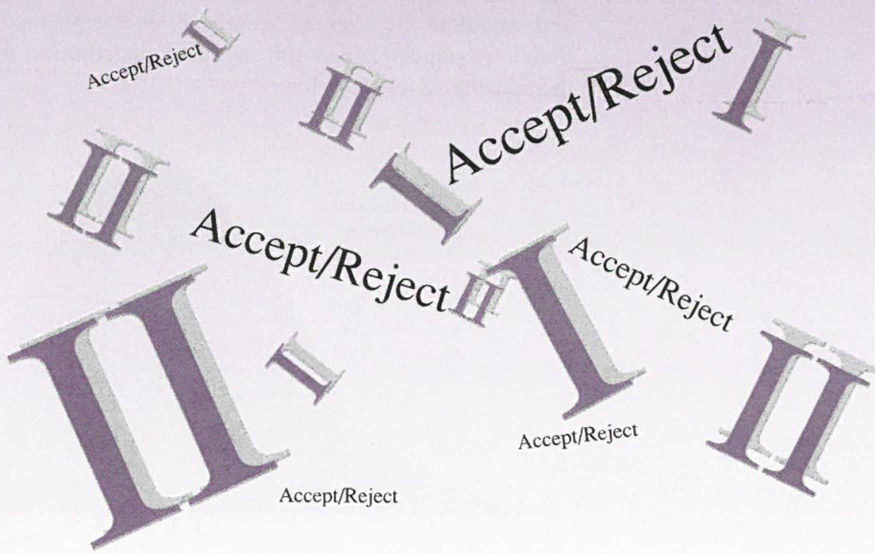
CHAPTER 6

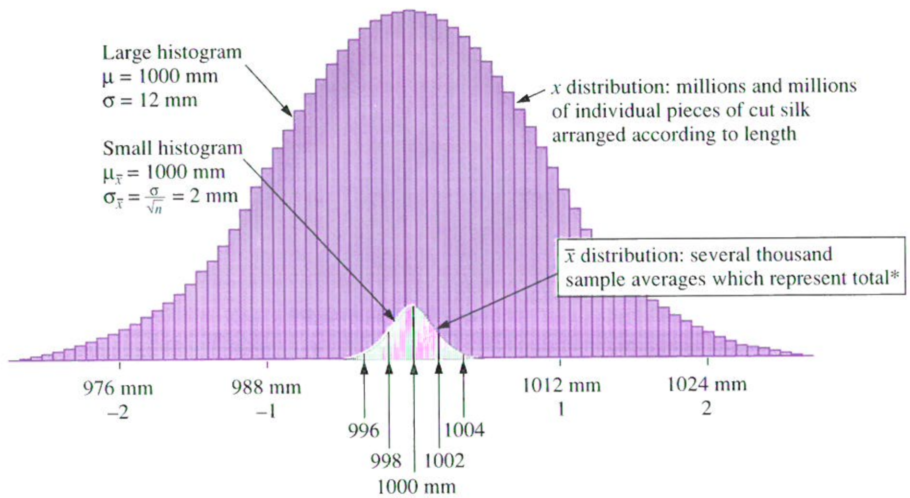
Introduction to Hypothesis Testing



6.1 Basic Concepts of Hypothesis Testing

For a complete understanding of **hypothesis testing,** one must first understand three important concepts in statistics, namely, **accept/reject** decision making, the **Type I error,** and the **Type II error.** Let's see how these concepts interrelate using chapter 5's cutting machine problem. ▼

Suppose a machine in a dress factory cuts pieces of silk material to an average length of $\mu = 1000$ mm with standard deviation $\sigma = 12$ mm. If we were to continually take random samples, with 36 pieces of cut material in each sample, and calculate the average length ($\bar{x}$) of the 36 pieces in each of these samples, then experience tells us the resulting distribution of $\bar{x}$'s would take the form of the *small histogram* below.



Now let us suppose the machine is turned off because it is late Friday afternoon and all the workers are leaving the dress factory for the weekend. The weekend passes and on Monday morning a sleepy-eyed operator starts up the cutting machine in preparation for the week's operation. The dial on the machine is set to cut at 1000 mm. We know from our experiment the prior week that *if* the machine is operating properly, the machine will be cutting pieces to an average length of 1000 mm, although some pieces will be shorter and some longer in accordance with the *large histogram* above.

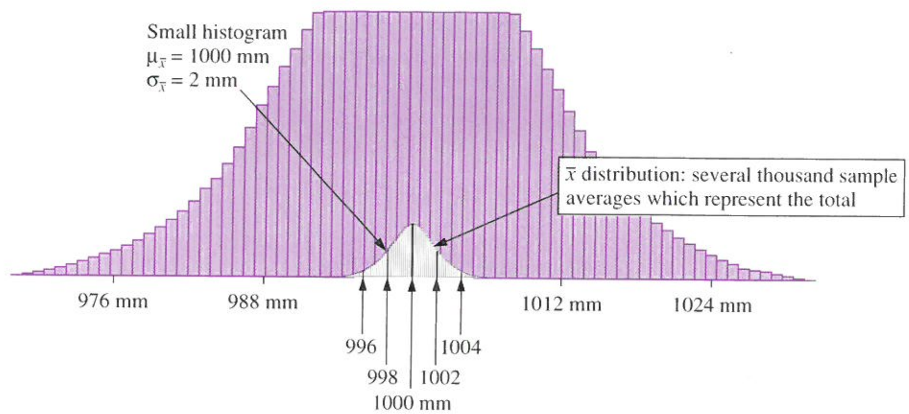
But how can we be sure that in our absence someone has not tampered with the machine, that in closing down or starting up the machine critical parts have

*As stated in Section 5.1: through computer simulation, fifteen thousand samples of size $n = 36$ can be chosen, sample averages calculated and these values organized into a histogram to represent the total. Refer to chapter 5 endnote 2 for detailed discussion of sampling distributions.

not vibrated loose, or that our sleepy-eyed operator has not accidentally moved the dial off its setting. The answer is: until we take some cuts, we usually have no way of knowing.

Certainly, the length of the first two or three pieces of material off the cutting machine will be measured. This acts as a quick and simple check for any gross malfunctioning. However, after you measure the first few pieces, let the machine run for a period to stabilize. Then take a true random sample, say for instance of 36 pieces. Measure the length of each piece in the sample and calculate $\bar{x}$, the average length.

With this completed, we refer to the small histogram reprinted below for your reference.



The small histogram shows us where sample averages ($\bar{x}$'s) should fall on a properly operating machine. Just about every sample average falls between 994 mm and 1006 mm. Certainly, on Monday morning, if you were to calculate an $\bar{x}$ *outside* this 994 mm to 1006 mm range, then most likely the machine is *not* operating properly, *not* cutting on average at 1000 mm. On the other hand, if you were to calculate a sample average between 998 mm and 1002 mm, you would feel the machine was probably cutting okay, since this is where you would expect to find most of the sample averages ($\bar{x}$'s) on a properly operating machine.

However, this leaves a gigantic borderline. What if you were to get a sample average of exactly 994 mm? Or 996 mm? Or 998 mm? At what point do we say the machine is cutting okay, or *not* cutting okay? Fortunately industry and research have been grappling with this question for decades and, from vast experience, have come up with some guidelines. One guideline establishes the middle 95% of the $\bar{x}$'s on a properly operating machine as a gauge in a cold-hearted *accept/reject* decision.

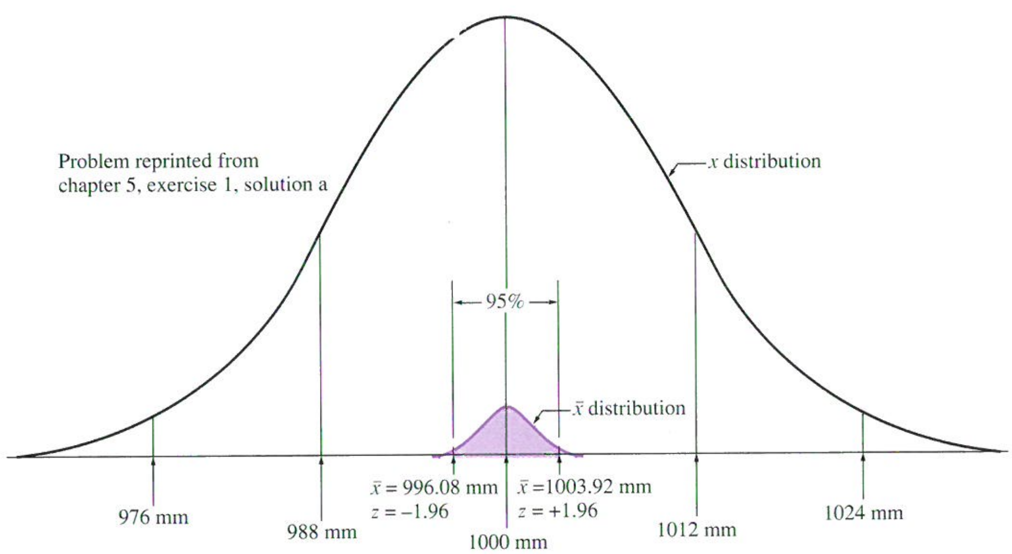
Accept/Reject Decision Making

Accept/reject decision making fundamentally proceeds in three steps, as follows:

1. The first step in an accept/reject decision is to set up some initial assumption. Although many initial assumptions are possible, in this case, the preferred initial assumption is $\mu = 1000$ mm since this is where we suspect (and hope) the machine is cutting on average. So, as first step, we state:

$$\text{Initial assumption: } \mu = 1000 \text{ mm}$$

2. The second step in an accept/reject decision is to establish some guideline for accepting or rejecting your initial assumption. In this case, we will choose the often used middle 95% of the $\bar{x}$'s guideline, as follows:



As calculated in a prior example, the middle 95% of the $\bar{x}$'s on a properly operating machine would be expected to fall between 996.08 mm and 1003.92 mm, so your accept/reject decision would be as follows:

Accept $\mu = 1000$ mm
if your sample $\bar{x}$ is between 996.08 mm and 1003.92 mm

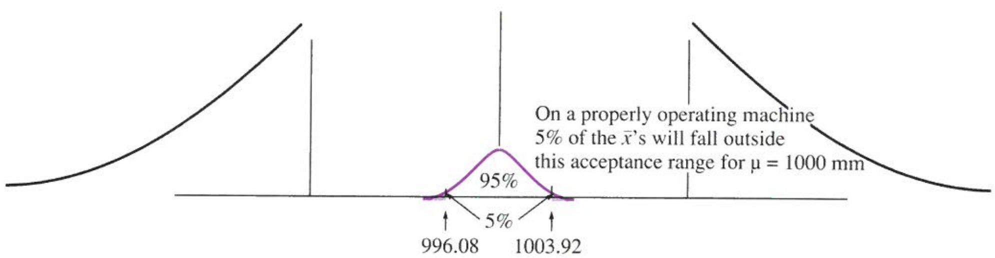
Reject $\mu = 1000$ mm
if your sample $\bar{x}$ is *outside* this range

3. The third step is a decision. If your sample average ($\bar{x}$) is between 996.08 mm and 1003.92 mm, you accept the machine as cutting properly at $\mu = 1000$ mm. If your sample $\bar{x}$ is outside this range, you assume the machine is cutting *improperly,* that is, *not* at $\mu = 1000$ mm. It's a simple *accept* $\mu = 1000$ mm or *reject* $\mu = 1000$ mm decision, no maybes, no in-betweens. You accept $\mu = 1000$ mm or reject $\mu = 1000$ mm. Period.

But aren't accept/reject decisions risky? Accept/reject decisions if properly thought out are one of the most powerful and efficient devices in statistical research, however, yes, they do come with risk. One of the risks is called the Type I error.

Type I Error

Let's continue with the cutting machine problem. If we adopt the middle 95% of the $\bar{x}$'s as our gauge for accepting or rejecting whether a machine is operating properly, then we must remember 5% of the $\bar{x}$'s will fall outside this 996.08 mm to 1003.92 mm range (100% minus 95%) when the machine is operating properly, as follows:



That means on a properly operating machine, 5% of the time you will reject the machine as operating properly. In other words, there is a 5% probability of rejecting a properly operating machine when, in fact, we should not reject it.

Type I error: rejecting an initial assumption in error.

In this case, our initial assumption is that the machine is operating properly at $\mu = 1000$ mm. However, on a properly operating machine, 5% of the time you will reject this initial assumption in error. This is called a *Type I error* and its probability, in this case, is 5%. The probability of a

Type I error is denoted by the symbol, α (alpha), traditionally labeled the **level of significance** and written in decimal form as follows:

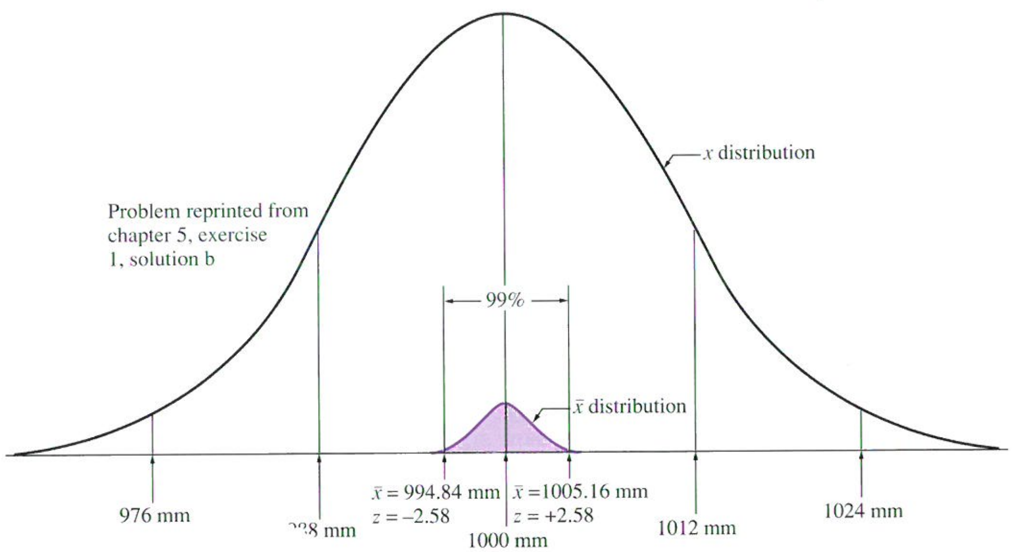
$$\text{Level of significance, } \alpha = .05$$

Unfortunately we must live with some Type I error risk for the convenience and efficiency of an accept/reject decision.

Hold it a minute! What is he jabbering about? We don't have to live with this 5% risk, you might say. Why not establish a middle 99% of the $\bar{x}$'s for accepting the machine as operating properly. This lowers the Type I error risk to 1% (99% + 1% = 100%). In other words, only 1% of the $\bar{x}$'s on a properly operating machine will fall outside the 99% range. Written in statistical terms, you would say

$$\text{Why not impose a level of significance, } \alpha = .01?$$

The level of significance $\alpha = .01$ is also often used in industry and research. But let's see what happens when this is done.



In this case, the middle 99% of the $\bar{x}$'s on a properly operating machine would be expected to fall between 994.84 mm and 1005.16 mm (as calculated in a prior exercise, noted above). Notice, if we lower the risk of a Type I error from 5% to 1%, we ''open up'' the range of sample averages where we would consider the machine operating properly, as follows:

$\alpha = .05$ Machine cutting okay if $\bar{x}$ is between 996.08 mm and 1003.92 mm
$\alpha = .01$ Machine cutting okay if $\bar{x}$ is between 994.84 mm and 1005.16 mm

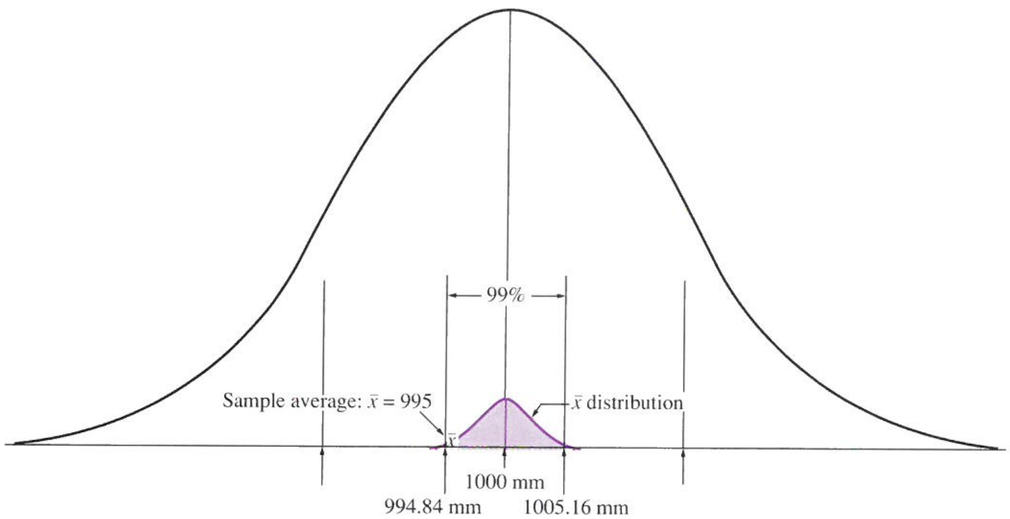
By opening up the range for accepting the machine as operating properly, we reduce the risk of a Type I error from 5% to 1%. However, this leaves us more vulnerable to another form of risk, called the Type II error.

Type II Error

Okay, we have just decided to reduce our chances of a Type I error by opening up the range of sample averages where we would consider the machine as operating properly. We have adopted the following guideline.

$\alpha = .01$ Any sample average ($\bar{x}$) between 994.84 mm and 1005.16 mm is assumed to come from a properly operating machine, that is, a machine cutting okay at $\mu = 1000$ mm.

The situation now arises: what if we turn on the machine Monday morning, sample 36 pieces as usual, and obtain an $\bar{x}$ of say, 995.00 mm? You must conclude: machine okay, probably cutting at $\mu = 1000$ mm, since $\bar{x} = 995.00$ mm falls between 994.84 mm and 1005.16 mm, as follows:



Now here's the problem. A sample average $\bar{x} = 995.00$ mm is also "typical" for a machine cutting at, say for instance, $\mu = 995$ mm. How do we know a pin is not stuck in the machine so that the machine is now, in reality, cutting on average to 995 mm? Or to 994 mm? Or to 996 mm? In other words, *Has there been a shift?* The answer to this question is: we don't know. Based on a sample average $\bar{x} = 995$ mm you accepted the fact of the machine cutting at $\mu = 1000$ mm. However if the machine in reality has actually shifted to $\mu = 995$ mm (or to any other value), we have just committed a Type II error.

Type II error: accepting an initial assumption in error.

Again, our initial assumption is that the machine is operating properly at $\mu = 1000$ mm. If we accept $\mu = 1000$ mm, when in reality, $\mu \neq 1000$ mm, we have committed a Type II error. The probability of committing a Type II error is denoted by the symbol, β (beta).

Note, had we adopted the middle 95% of the $\bar{x}$'s for accepting a machine as operating properly ($\alpha = .05$) by accepting $\mu = 1000$ mm if $\bar{x}$ falls between 996.08 mm and 1003.92 mm, then a sample average ($\bar{x}$) of 995.00 mm would have alerted us to a possible malfunction since $\bar{x} = 995$ mm falls *outside* the 996.08 to 1003.92 range. In other words, had we accepted our original 5% Type I error risk, the Type II error above would not have occurred. Generally, decreasing the probability of a Type I error merely increases the probability of a Type II error.

To summarize

Type I error:	rejecting an initial assumption in error.
Type II error:	accepting an initial assumption in error.

Remember: decreasing the Type I error by imposing a lower α, say for instance going from .05 to .01, merely increases your Type II error risk.

Power

Statisticians will often evaluate a statistical test in terms of its **power.** Power simply means the probability of making the correct decision by avoiding a Type II error. If you calculate a Type II error risk of 10% the *power* of the test is 90%. If your Type II error risk is 30%, your power is 70%. Either you make the Type II error or you make the correct decision. The sum of the two must equal 100%, that is, β + Power = 100% *or* Power = 100% − β. Written in decimal form, it is expressed as Power = 1 − β. Remember: β is the probability of making a Type II error.

Precise calculations of the Type II error and power will be demonstrated in the following two examples. To summarize:

> **Power**
> Probability of making a correct decision by avoiding a Type II error
>
> $$\text{Power} = 100\% - \beta = 1 - \beta$$

6.2 Applications

Accept/reject decision making is standard practice in statistical testing. However, accept/reject decisions do come with risk. Four examples that demonstrate this risk are presented.

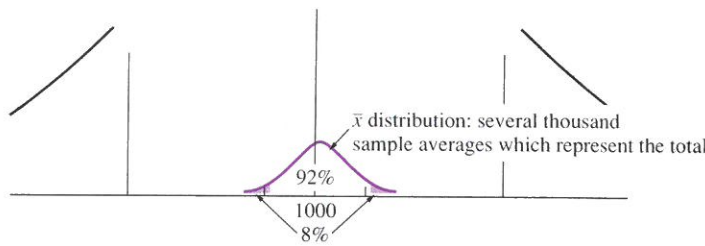
Example

In the cutting machine problem, suppose we establish the middle 92% of the $\bar{x}$'s as our *cutoffs* for accepting $\mu = 1000$ mm. Assuming $n = 36$ and $\sigma = 12$ mm,

a. What is the probability of a Type I error?
b. Between what $\bar{x}$ values would you accept the machine cutting at $\mu = 1000$ mm?
c. Explain briefly how one might commit a Type I error.

Solution to (a)

The probability of a Type I error is simply 8% (100% − 92% = 8%). Written in statistical terms, you would state $\alpha = .08$. In other words, on a properly operating machine, 8% of the $\bar{x}$'s would fall *outside* your acceptance range for $\mu = 1000$ mm, as shown:



Solution to (b)

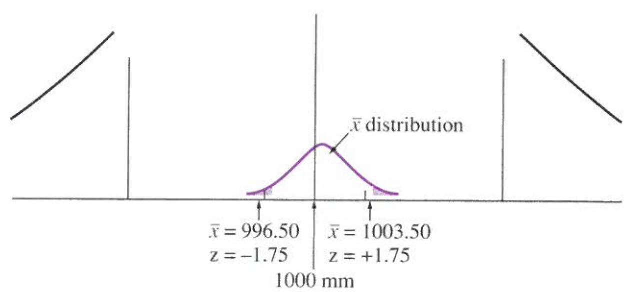
This is a typical "working backward (given the area, find z)" problem for the normal curve, only now we are dealing with the normal curve of the $\bar{x}$ distribution. Since we know the area between the critical cutoffs is 92%, we merely look up the corresponding z scores. Doing this, we get $z = -1.75$ and $z = +1.75$. (Remember the table reads half the normal curve, so we must look up an area of 46% ($\frac{1}{2}$ of 92%) or in decimal form .4600.)

Using $z = -1.75$ and $z = +1.75$, we calculate the values at the cutoffs as follows:

$$\sigma_{\bar{x}} = \frac{12}{\sqrt{36}} = 2$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad -1.75 = \frac{\bar{x} - 1000}{2} \qquad \bar{x} = 996.50 \text{ mm}$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad +1.75 = \frac{\bar{x} - 1000}{2} \qquad \bar{x} = 1003.50 \text{ mm}$$



So, you would accept the machine as cutting properly, cutting at $\mu = 1000$ mm, if you obtained a sample average between $\bar{x} = 996.50$ mm and $\bar{x} = 1003.50$ mm.

Solution to (c)

The following might be a Type I error scenario: You random sample 36 pieces and your sample average falls *outside* this 996.50 mm to 1003.50 mm range, say for instance you obtain a sample average of $\bar{x} = 994$ mm. Based on this you reject the machine as operating properly and shut down production.

If indeed the machine is *not* operating properly, not cutting on average to $\mu = 1000$ mm, then you have made no error. Your decision was correct. However, if the machine is okay, cutting properly at $\mu = 1000$ mm, and you happened to have sampled one of those rare 8% occurrences, then you have made a Type I error. ■

Example

In the cutting machine problem, suppose we arbitrarily establish $\bar{x} = 997$ mm to $\bar{x} = 1003$ mm as our cutoffs for accepting $\mu = 1000$ mm. Assuming $n = 36$ and $\sigma = 12$ mm,

a. What is the probability of a Type I error?
b. What is the probability of a Type II error if the machine shifts and is now cutting at $\mu = 995$ mm?
c. What is the power of the test in part b?

Solution to (a)

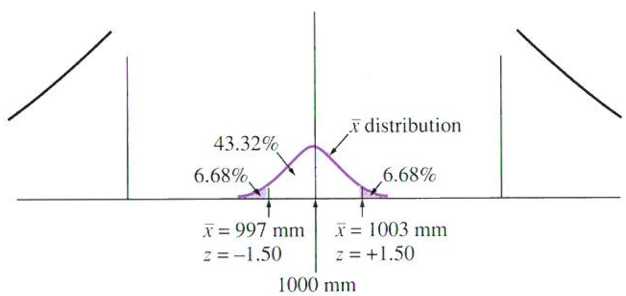
The percentage of $\bar{x}$'s on a properly operating machine that fall *outside* the 997 mm to 1003 mm range is your Type I error risk. It is represented by the shaded regions in the diagram below. Calculating the percentage of data in the shaded regions we get

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{36}} = 2 \text{ mm} \qquad z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$z = \frac{997 - 1000}{2} \qquad z = \frac{1003 - 1000}{2}$$

$$z = -1.50 \qquad z = +1.50$$

Looking up $z = 1.50$, we get 43.32%, with 6.68% in the tail.

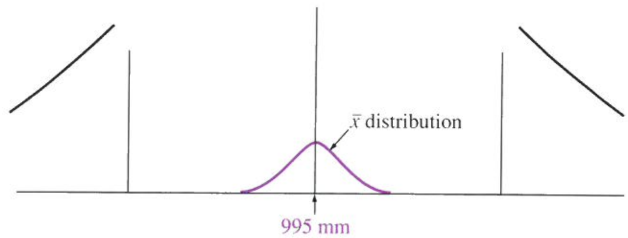


The probability of a Type I error is 13.36% (6.68% + 6.68%). In other words, on a properly operating machine the percentage of $\bar{x}$'s you would expect to fall *outside* the 997 mm to 1003 mm range is 13.36%. Written in decimal form (.1336) and rounded to two decimal places, you would say:

$$\alpha = .13$$

Solution to (b)

If a machine is indeed cutting at $\mu = 995$ mm, the sample averages ($\bar{x}$'s) now would cluster about $\mu = 995$ mm, as follows:



You would commit a Type II error if you accepted (in error) the machine cutting at $\mu = 1000$ mm. The only way you would accept a machine cutting at $\mu = 1000$ mm is if you took a random sample of 36 pieces and your sample average ($\bar{x}$) fell between 997 mm and 1003 mm. So, to calculate β, the probability of a Type II error, you must calculate the percentage of $\bar{x}$'s that would fall between 997 mm and 1003 mm from a machine cutting at $\mu = 995$ mm.

Remember: the machine is cutting at $\mu = 995$ mm but you are unaware of this. The only information available to you is your one sample average, $\bar{x}$.

To find the probability of a Type II error in this problem, we calculate the percentage of $\bar{x}$'s that we would expect to fall between 997 mm and 1003 mm, represented by the shaded region in the diagram below:

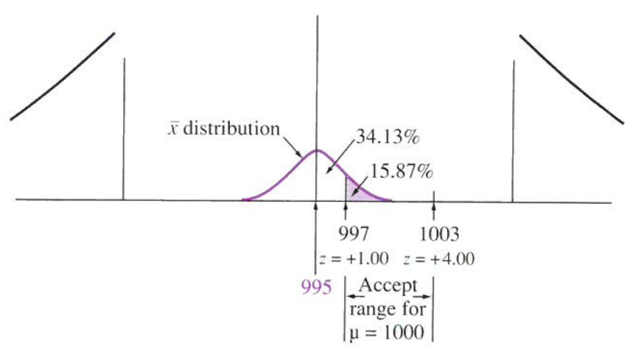
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{36}} = 2 \text{ mm*}$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$z = \frac{997 - 995}{2} \qquad z = \frac{1003 - 995}{2}$$

$$z = +1.00 \qquad z = +4.00$$

(disregard since negligible data exists in this region)



*Technical note: μ and σ are independent. A shift in μ will generally not affect σ. See endnote 10 in chapter 10 for further reading.

Looking up $z = 1.00$, we get 34.13%, thus our Type II error risk is 15.87% (50% minus 34.13%). Since 15.87% of the $\bar{x}$'s coming from a machine cutting at $\mu =$ 995 mm would be expected to fall between 997 mm and 1003 mm, this 15.87% is the risk you will obtain one of these $\bar{x}$'s and thus conclude (erroneously) the machine was cutting at $\mu = 1000$ mm. So, your risk of a Type II error when μ shifts to 995 mm is 15.87%. Written in decimal form (.1587) and rounded to two decimal places, you would say:

$\beta = .16$ which is the probability you will accept your initial assumption, $\mu = 1000$ mm, in error. This is your Type II error risk for a shift to $\mu = 995$ mm.

Solution to (c)

Power is the probability of making a correct decision by avoiding a Type II error. For this test, Power = 84% (100% – 16%). Written in decimal form,

$$\text{Power} = .84 \qquad (\text{Note: Power} = 1 - \beta = 1.00 - .16 = .84)$$

Explained another way, since 84.13% of the $\bar{x}$'s (50% + 34.13%) will be lower than 997 mm, as shown on the previous diagram, you have an 84.13% chance your sample average will be less than 997 mm. In that case, you would reject $\mu = 1000$ mm, which would be the correct decision. In other words, we have an 84.13% chance of making the correct decision (rejecting $\mu = 1000$ mm) and a 15.87% chance of making the wrong decision (accepting $\mu = 1000$ mm). Power is the probability of making the correct decision in this situation and β the probability of making the wrong decision. ■

Example

The National Institutes of Health agreed to supply active disease viruses, such as polio and AIDS, to research firms for the purpose of experimentation. A process is set up to automatically fill millions of small test tubes to an average of 9.00 ml of disease virus with standard deviation .35 ml.

With sample sizes of $n = 49$ test tubes, it was calculated that 99% of the sample averages ($\bar{x}$'s) would fall between 8.87 ml and 9.13 ml (chapter 5, section 5.2, second example). If we use this 8.87 ml to 9.13 ml range of sample averages as our criterion to accept $\mu = 9.00$ ml,

a. What is the probability of a Type I error?
b. What is the probability of a Type II error if the process shifts to $\mu = 9.20$ ml?
c. What is the power of the test in part b?

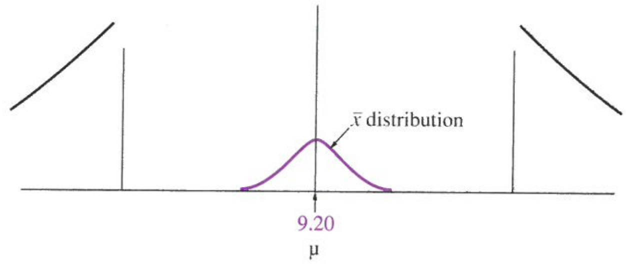
Solution to (a)

Since 99% of the $\bar{x}$'s fall between 8.87 ml and 9.13 ml, 1% of the $\bar{x}$'s will fall outside this range. In other words, there is a 1% chance you will obtain an $\bar{x}$ *outside* this 8.87 ml to 9.13 ml range when the process is operating properly. So your probability of a Type I error is 1%. Written in decimal form you would say:

$\alpha = .01$ which is the probability you would reject your initial assumption ($\mu = 9.00$ ml) in error. This 1% is your Type I error risk.

Solution to (b)

The test tubes are now filling on average at $\mu = 9.20$ ml, so your sample averages ($\bar{x}$'s) would now cluster about 9.20 ml, as follows:



You would commit a Type II error if you accepted the process filling at $\mu = 9.00$ ml. The only way you would accept the process filling at $\mu = 9.00$ ml is if you took a random sample of 49 test tubes and your average ($\bar{x}$) fell between 8.87 ml and 9.13 ml. So, to calculate β, the probability of a Type II error, you must calculate the percentage of $\bar{x}$'s that would fall between 8.87 ml and 9.13 ml from a process filling at $\mu = 9.20$ ml.

Remember: the process is filling at $\mu = 9.20$ ml but you are unaware of this. The only information available to you is your one sample average, $\bar{x}$.

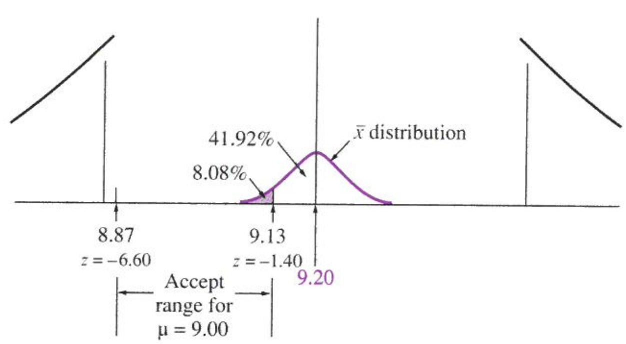
To find the probability of a Type II error in this problem, we calculate the percentage of $\bar{x}$'s that we would expect to fall between 8.87 ml and 9.13 ml, represented by the shaded region in the diagram that follows:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.35}{\sqrt{49}} = .05 \text{ ml}$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad z = \frac{8.87 - 9.20}{.05} \qquad z = -6.60$$

(disregard since negligible data exists in this region)

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad z = \frac{9.13 - 9.20}{.05} \qquad z = -1.40$$



Looking up $z = -1.40$, we get 41.92%, thus our Type II error risk is 8.08%. Since 8.08% (50.00% minus 41.92%) of the $\bar{x}$'s coming from a process filling at $\mu = 9.20$ ml would be expected to fall between 8.87 ml and 9.13 ml, this 8.08% is the risk you will obtain one of these $\bar{x}$'s and thus conclude (erroneously) the process was filling at $\mu = 9.00$ ml. So, your risk of a Type II error when μ shifts to 9.20 ml is 8.08%. Written in decimal form (.0808) and rounded to two decimal places, you would say:

$\beta = .08$ which is the probability you will accept your initial assumption, $\mu = 9.00$ ml, in error; this is your Type II error risk for a shift to $\mu = 9.20$ ml.

Solution to (c)

Since power is the probability of making the correct decision by avoiding a Type II error, we get 100% − 8.08% = 91.92%

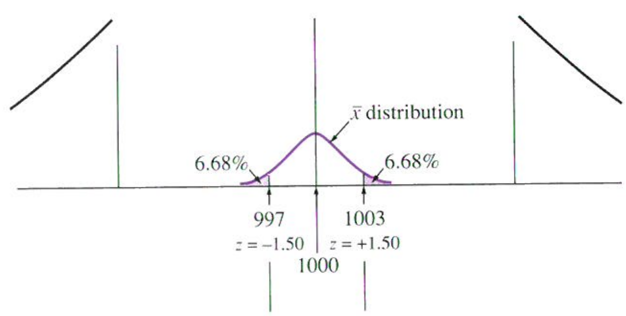
$$\text{Power} = 91.92\% \text{ or } .92$$

Put another way: since 91.92% (50% plus 41.92%) of the $\bar{x}$'s will be greater than 9.13 ml, you have a 91.92% chance your sample average will be greater than 9.13 ml. In that case, you would reject $\mu = 9.00$ ml, which would be the correct decision. In other words, we have a 91.92% chance of making the correct decision (rejecting $\mu = 9.00$ ml) and a 8.08% chance of making the wrong decision (accepting $\mu = 9.00$ ml). ■

6.3 Controlling Error

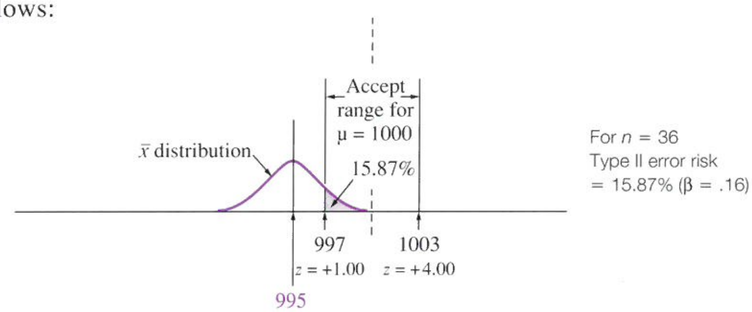
Although a number of techniques can be used to decrease error risk, perhaps the most broadly preferred is increasing n, your sample size. This is best explained through practical example, as follows.

In our cutting machine problem (section 6.2, second example), we arbitrarily established $\bar{x} = 997$ mm to $\bar{x} = 1003$ mm as our cutoffs for accepting $\mu = 1000$ mm. Using $n = 36$ and $\sigma = 12$ mm, we calculated the probability of a Type I error to be 13.36% (6.68% + 6.68%), as follows:



For $n = 36$
Type I error risk
= 13.36% (σ = .13)

For the probability of a Type II error if μ shifts to 995 mm, we calculated 15.87%, as follows:



Now here's a problem: 13.36% is too high for a Type I error risk. It means, 13 times out of 100 you will reject a properly operating machine as malfunctioning. Checking out a properly operating machine for a malfunction that doesn't exist can be time consuming and expensive. Well, then, how do we lower this Type I error risk? There are a number of ways, but each come with drawbacks. We will discuss three.

In production or quality control situations, as in this example, most likely a **back up sample** would be taken, that is, a second random sample drawn from the machine's output. However, as mentioned above, this can be time consuming and expensive. Moreover, the use of back up samples in most industry and research situations is just not practical. Most statistical studies are exceedingly expensive (marketing studies often cost hundreds of thousands of dollars), exceedingly time consuming (scientific studies can easily range two to ten years), and exceedingly enervating.

A second approach is to arbitrarily set a lower Type I error risk, say from 13.36% to 1%, but as we discussed at length in prior sections, arbitrarily lowering your Type I error risk merely increases your Type II error risk (in fact, in this case lowering the Type I error risk from $\alpha = 13.36\%$ to $\alpha = 1\%$ would merely increase the β error, the Type II error risk, from 15.87% to over 53%) and a large Type II error risk means if the machine actually shifts, there is a high probability you won't be able to detect it.

A third approach, and perhaps the most preferred, is to increase your sample size. Let's see what happens when we increase our sample size to $n = 144$.

Example

In our cutting machine problem, suppose we arbitrarily establish $\bar{x} = 997$ mm to $\bar{x} = 1003$ mm as our cutoffs for accepting $\mu = 1000$ mm, only this time we increase our sample size to $n = 144$. (Assume $\sigma = 12$ mm.)

a. Calculate the probability of making a Type I error.
b. Calculate the probability of making a Type II error if μ shifts to 995 mm.
c. Compare these results to the results when $n = 36$.

Solution to (a)

Since n, your sample size, has increased from 36 pieces of cut material to 144 pieces of cut material, the sample averages ($\bar{x}$'s) now cluster much closer to $\mu = 1000$ mm. In fact, calculating $\sigma_{\bar{x}}$ we get

$$\sigma_{\bar{x}} = \frac{12}{\sqrt{144}} = \frac{12}{12} = 1 \text{ mm}$$

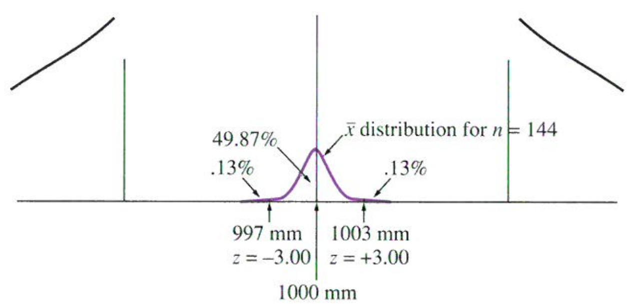
The percentage of $\bar{x}$'s on a properly operating machine that fall outside the 997 mm to 1003 mm range is shown by the shaded regions in the following diagram. This is your Type I error risk. Notice how the shaded area can barely be seen. This is because, now, the $\bar{x}$'s have clustered much closer to $\mu = 1000$ mm. Calculating the percentage of data in the shaded region,

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$z = \frac{997 - 1000}{1} \qquad z = \frac{1003 - 1000}{1}$$

$$z = -3.00 \qquad z = +3.00$$

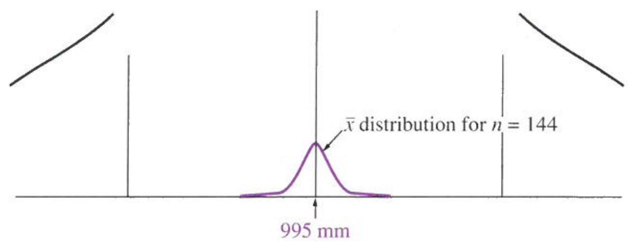
Looking up $z = 3.00$, we get 49.87% with .13% in the tail.



On a properly operating machine, the percentage of $\bar{x}$'s you would expect to fall outside the 997 mm to 1003 mm range is .26% (.13% + .13%), which is much less than 1%. This is your Type I error risk. Written in decimal form (.0026) and rounded to three decimal places, you would write $\alpha = .003$. In other words, there is less than 3 chances in 1000 of making a Type I error. Small indeed.

Solution to (b)

If the machine is indeed cutting at $\mu = 995$ mm, the sample averages ($\bar{x}$'s) would now cluster about $\mu = 995$ mm, as indicated in the following figure.



You would commit a Type II error if you accepted (in error, of course) the machine cutting at $\mu = 1000$ mm. The only way you would accept the machine cutting at $\mu = 1000$ mm is if you took a random sample of 144 pieces and your sample average ($\bar{x}$) fell between 997 mm and 1003 mm. So, to calculate β, the probability of a Type II error, you must calculate the percentage of $\bar{x}$'s that would fall between 997 mm and 1003 mm from a machine operating at $\mu = 995$ mm.

Remember: the machine is cutting at $\mu = 995$ mm but you are unaware of this. The only information available to you is your one sample average, $\bar{x}$. To find the probability of a Type II error in this problem, we calculate the percentage of $\bar{x}$'s that we would expect to fall between 997 mm and 1003 mm, represented by the shaded region in the diagram below:

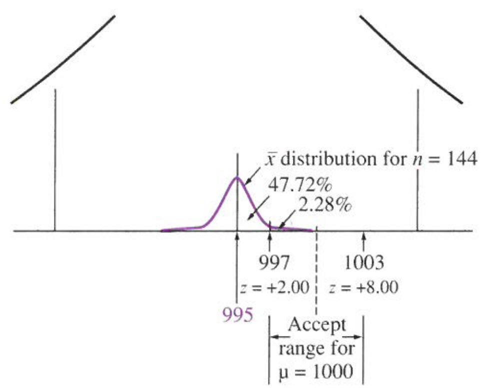
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{144}} = \frac{12}{12} = 1 \text{ mm}$$

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \qquad z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$z = \frac{997 - 995}{1} \qquad z = \frac{1003 - 995}{1}$$

$$z = +2.00 \qquad z = +8.00$$

(disregard since negligible data exists in this region)



For n = 144
Type II error
= 2.28% (β = .02)

Looking up $z = +2.00$, we get 47.72%, thus our Type II error risk is 2.28%. Since 2.28% (50% − 47.72%) of the $\bar{x}$'s coming from a machine cutting at $\mu = 995$ mm would be expected to fall between $\bar{x} = 997$ mm and $\bar{x} = 1003$ mm, this is the risk you will obtain one of these $\bar{x}$'s and thus conclude (erroneously) the machine was cutting at $\mu = 1000$ mm. So, your risk of a Type II error when μ shifts to 995 mm is 2.28%. Written in decimal form (.0228) and rounded to two decimal places, you would say

$$\beta = .02$$

Solution to (c) The following is a comparison of results:

For $n = 36$	For $n = 144$
$\alpha = 13.36\%$	$\alpha =$ less than 1%
$\beta = 15.87\%$ (for μ shifting to 995 mm)	$\beta = 2.28\%$ (for μ shifting to 995 mm)

When we increase our sample size from 36 pieces to 144 pieces, note the formidable drop in risk. The Type I error risk (α) drops from 13.36% to less than 1%. The Type II error risk (β) drops from 15.87% to 2.28%. Of course, in practical terms, increasing your sample size will add cost and inconvenience, but usually these are small prices to pay for the added protection. ■

In conclusion, controlling errors should be thought out at the initial stages of planning a statistical study. Although other methods are available, the preferred method for lowering Type I and Type II error risks is by increasing your sample size.

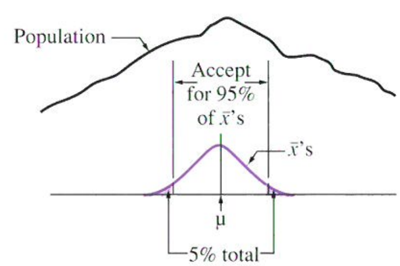
Summary

Three fundamental concepts were presented in this chapter as follows.

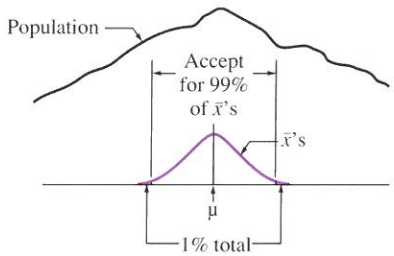
1. **Accept/reject decision making:** Accept/reject decision making proceeds as follows. The first step is to set up some initial assumption, say for instance that a population mean, μ, equals some specific value. Then we establish guidelines for accepting and rejecting this initial assumption. The final step is to accept or reject based on sample results.

Two popular guidelines for accept/reject decision making are as follows.

95% Guideline — We calculate where the middle 95% of the $\bar{x}$'s would be expected to fall if the initial assumption were true. If our sample $\bar{x}$ falls in this 95% range, we accept our initial assumption, otherwise reject (see sketch).



99% Guideline — We calculate where the middle 99% of the $\bar{x}$'s would be expected to fall if the initial assumption were true. If our sample $\bar{x}$ falls in this range, we accept our initial assumption, otherwise reject (see sketch).



Accept/reject decision making is an efficient procedure well-suited for the cost-conscious needs of research and business, however it does come with risks, discussed as follows.

2. **Type I error (the α risk):** One risk is the probability of rejecting your initial assumption in error. For instance, say we establish the 95% guideline for accepting some initial assumption. Thus, if the initial assumption were true, 95% of the $\bar{x}$'s would fall in the accept zone and indeed we would make the correct decision (by accepting the initial assumption). However, this also implies that 5% of the $\bar{x}$'s will fall *outside* the accept zone even though the initial assumption is true, and we will reject the initial assumption in error.

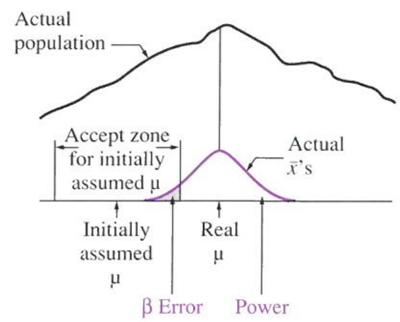
   This 5% is called the level of significance or Type I error risk for this experiment and is usually denoted as $\alpha = .05$.

3. **Type II error (the β risk):** Another risk is the probability of *accepting* the initial assumption in error. For instance, say we establish a 95% guideline for accepting some initial assumption, however this initial assumption is *not* true. Since we usually have no way of knowing the initial assumption is not true prior to sampling, we set up an accept zone and proceed with the experiment.

   Now, by chance, we may actually get sample $\bar{x}$'s falling in this accept zone even though the initial assumption is incorrect.

The probability that your sample $\bar{x}$ will fall in this accept zone when the initial assumption is incorrect is called a Type II error risk and its probability will vary depending on several factors.

Power: Essentially power phrases the Type II error risk in positive terms. For instance, say under a particular set of conditions, we calculate the Type II error risk to be 12%, then the power of the experiment is 88% (100% minus 12%), meaning 88% of the time you will reject the initial assumption correctly and only 12% of the time will you commit the Type II error by accepting it erroneously (see sketch for detailed explanation of this example).



β Error: Say for instance, 12% of the actual $\bar{x}$'s fall in the accept zone for some initially assumed μ, thus 12% of the time you will accept the initially assumed μ in error (not realizing the real μ is in a different position).

Power: For this example the power is 88% (100% minus 12%), meaning 88% of the time the sample $\bar{x}$'s will fall outside the accept zone for the initially assumed μ and you will *reject* the initially assumed μ (which is the correct decision).

Controlling error in experiments: Generally in an experiment, we wish both risks (Type I and Type II) to be as low as possible. However, decreasing the Type I risk (say for instance, by lowering the α level of the experiment from .05 to .01) merely increases your Type II error risk.

Perhaps the best approach to reducing errors is to increase your sample size substantially, which lowers your Type II error risk, thus allowing for greater flexibility in setting a Type I error risk (α level) for your experiment. The negative side of increasing the sample size is that it may be costly, time-consuming, or in some cases not feasible.

Another approach is to conduct a second study, however this is often exceedingly costly, time consuming and may arouse questions as to why you didn't plan the initial study more carefully (say for instance, by using a larger sample size). And in many cases conducting a second study is simply not feasible.

Exercises

Note that full answers for exercises 1–5 and abbreviated answers for odd-numbered exercises thereafter are provided in the Answer Key.

6.1 In the cutting machine problem, suppose we establish the middle 98% of the $\bar{x}$'s as our guideline for accepting $\mu = 1000$ mm. Assuming $n = 36$ and $\sigma = 12$ mm,

a. What is the probability of a Type I error?
b. Between what $\bar{x}$ values would you accept the machine cutting at $\mu = 1000$ mm?
c. Explain briefly how one might commit a Type I error.

6.2 Referring to exercise 6.1,

a. What is the probability of a Type II error if the machine shifts to $\mu = 995$ mm?
b. Compare Type I and Type II error risks calculated in this question with those of the second example of section 6.2 where we calculated the Type I and II error risks to be 13.36% and 15.87%, respectively. What principle concerning Type I and Type II errors is demonstrated?

6.3 In the cutting machine problem, for $\mu = 1000$ mm, $\sigma = 12$ mm, and $n = 36$,

a. If you wish to establish a Type I error risk of 10%, find the $\bar{x}$ cutoffs for accepting $\mu = 1000$ mm.
b. Calculate the probability of a Type II error for a shift to $\mu = 1004$ mm.
c. What is the power of the test in part b?

6.4 In the National Institutes of Health problem, suppose we arbitrarily establish $\bar{x} = 8.90$ ml to $\bar{x} = 9.10$ ml as our cutoffs for accepting test tubes filling on average to $\mu = 9.00$ ml. Assuming $n = 49$ and $\sigma = .35$ ml,

a. What is the probability of making a Type I error?
b. Using this example, briefly define a Type I error and discuss the consequences of making a Type I error.
c. What is the probability of a Type II error for a shift to $\mu = 9.14$ ml?
d. Using this example, briefly define a Type II error and discuss the consequences of making a Type II error.

6.5 Referring to exercise 6.4,

a. What is the probability that when the process is "in control" (filling properly), you will believe the process malfunctioning?
b. What is the probability that when the process goes "out of control" (filling improperly), say filling at $\mu = 9.14$ ml, you will believe the process is filling correctly?

6.6 In the National Institutes of Health problem, for $\mu = 9.00$ ml, $\sigma = .35$ ml, and $n = 49$,

a. Establish a Type I error risk of 1% and find the $\bar{x}$ cutoffs for accepting $\mu = 9.00$ ml.
b. Calculate the probability of a Type II error for a shift to $\mu = 8.85$ ml.
c. What is the power of the test in part b?

6.7 In the cutting machine problem, for $\mu = 1000$ mm and $\sigma = 12$ mm, suppose we establish $\bar{x} = 997$ mm to $\bar{x} = 1003$ mm as our cutoffs for accepting $\mu = 1000$ mm, calculate your Type I error risk and your Type II error risk (for a shift to $\mu = 995$ mm) for,

a. $n = 30$.
b. $n = 100$.
c. Compare the results in parts a and b.

6.8 Brell shampoo, an ''in-house'' brand, is marketed through a large national chain of convenience stores. This chain also carries other national brands of shampoo. Brell's in-house market share is $\mu = 24.0$ (meaning: on average 24.0% of the shampoo sold in these stores is Brell) with standard deviation 3.2.

Suppose we arbitrarily establish $\bar{x} = 23.3$ to $\bar{x} = 24.7$ as our cutoffs for accepting $\mu = 24.0$; assuming sample size $n = 75$,

a. What is the probability of a Type I error?
b. What is the probability of a Type II error for a shift to $\mu = 23.0$?
c. What is the power of the test in part b?

6.9 Referring to exercise 6.8,

a. Suppose we establish $\bar{x} = 23.1$ to $\bar{x} = 24.9$ as our cutoffs for accepting $\mu = 24.0$, what effect would this have on our Type I and Type II error risks and on power?
b. Recommend a way to decrease both your Type I and Type II error risks.

6.10 In the horror film moviegoer problem, suppose we arbitrarily establish $\bar{x} = 17.0$ to $\bar{x} = 17.8$ years old as our cutoffs for accepting the average age of $\mu = 17.4$ years old. Assuming $\sigma = 2.7$ years, calculate the risks for a Type I error and for a Type II error (assuming a shift to $\mu = 16.8$ years old) for,

a. $n = 45$.
b. $n = 250$.
c. Compare the results in parts a and b.