Hypothesis Testing II

TESTS ON THE MEAN OF A NORMAL DISTRIBUTION, VARIANCE KNOWN

We consider hypothesis testing about the mean μ of a single normal population where the variance of the population σ^2 is known.

- **Suppose that** $X_1, X_2, \ldots, X_n$ **is a random sample from a normal distribution with unknown mean μ and known variance σ^2.**
- **We know the sample mean** $\bar{X}$ **is normally distributed with mean μ and variance σ^2/n.**

Hypothesis Tests on the Mean

Suppose that we wish to test the hypotheses

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu \neq \mu_0 \end{aligned}$$

Test Statistic

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

Distribution of Z when $H_0: \quad \mu = \mu_0$ is true

If the null hypothesis $H_0: \quad \mu = \mu_0$ is true, $E(\bar{X}) = \mu_0$, and it follows that the distribution of Z is the standard normal distribution. Consequently, if hypothesis $H_0: \quad \mu = \mu_0$ is true, the probability is 1-α that the test statistic Z falls between $-z_{\alpha/2}$ and $z_{\alpha/2}$, where $z_{\alpha/2}$ is the $100\alpha/2$ percentage point of the standard normal distribution.

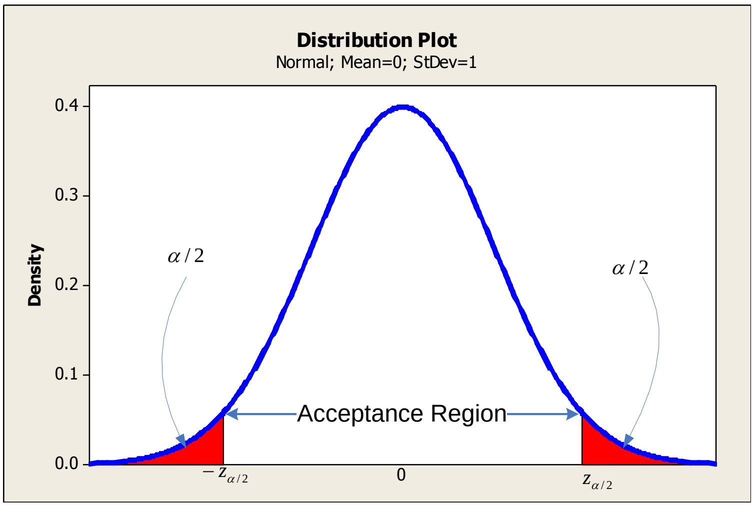
A sample producing a value of the test statistic that <u>falls in the tails</u> of the distribution of Z would be unusual if $H_0: \quad \mu = \mu_0$, therefore, it is an indication that H_0 is false . Thus we should reject H_0 if the observed value of the test statistic Z is either

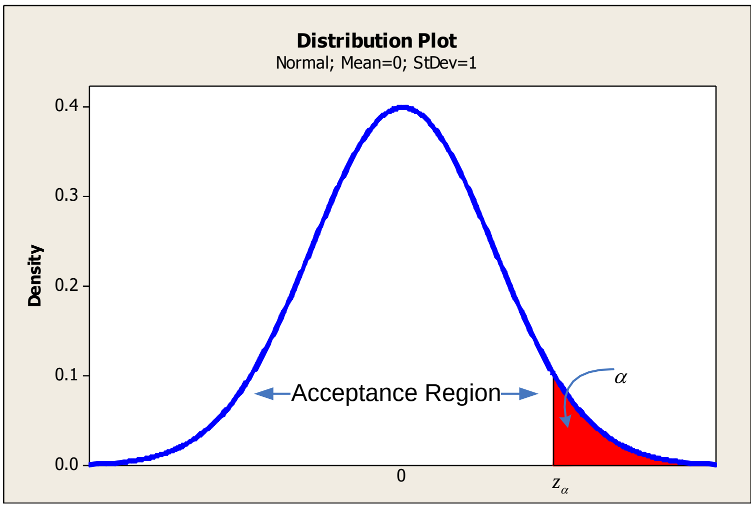
$$z > z_{\alpha/2} \ \textbf{or} \ \ z < -z_{\alpha/2}$$

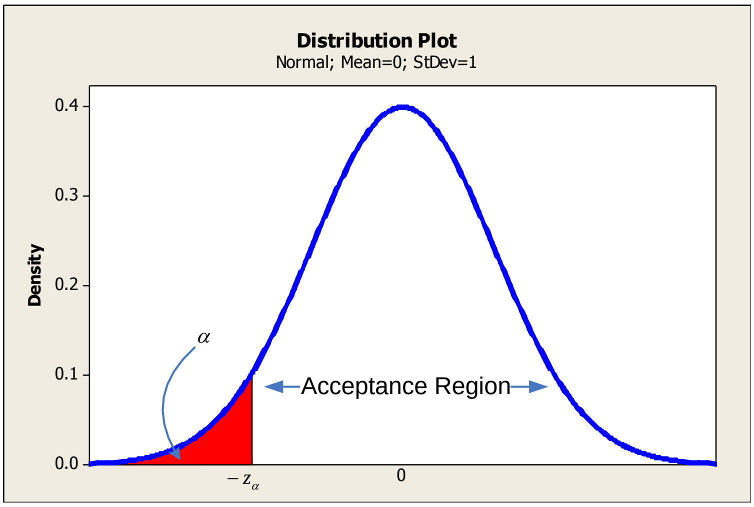
and we should fail to reject H_0 if

$$-z_{\alpha/2} \le z \le z_{\alpha/2}$$

Type-I error probability for this test procedure is α.


Distribution Plot
Normal; Mean=0; StDev=1
Density
0.4
0.3
0.2
0.1
0.0
$\alpha/2$
$\alpha/2$
Acceptance Region
$-z_{\alpha/2}$
0
$z_{\alpha/2}$


Distribution Plot
Normal; Mean=0; StDev=1
Density
0.4
0.3
0.2
0.1
0.0
α
Acceptance Region
0
z_{α}


Distribution Plot
Normal; Mean=0; StDev=1
Density
0.4
0.3
0.2
0.1
0.0
α
Acceptance Region
$-z_{\alpha}$
0

The Critical Value Approach:

The same critical region can be written in terms of the computed value of the sample mean $\bar{X}$. A procedure identical to the above is as follows;

Reject H_0 if $\bar{X} > a$ or $\bar{X} < b$

$a = \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}$ **and** $b = \mu_0 - z_{\alpha/2}\sigma/\sqrt{n}$

Example: **The daily yield for a local xxx plant has averaged 880 tons for the last several years. We know the standard deviation is σ=21. The quality control manager would like to know whether this average has changed in recent months. Randomly selected 50 days from the computer database averaged 871 tons. Test the appropriate hypothesis using α=0.05.**

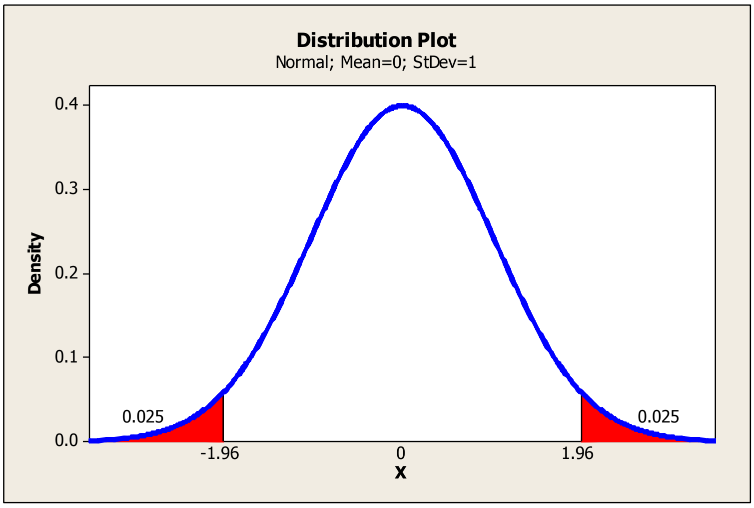
Null and alternate hypotheses:

$$H_0: \quad \mu = 880$$
$$H_1: \quad \mu \neq 880$$

Test Statistic:

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

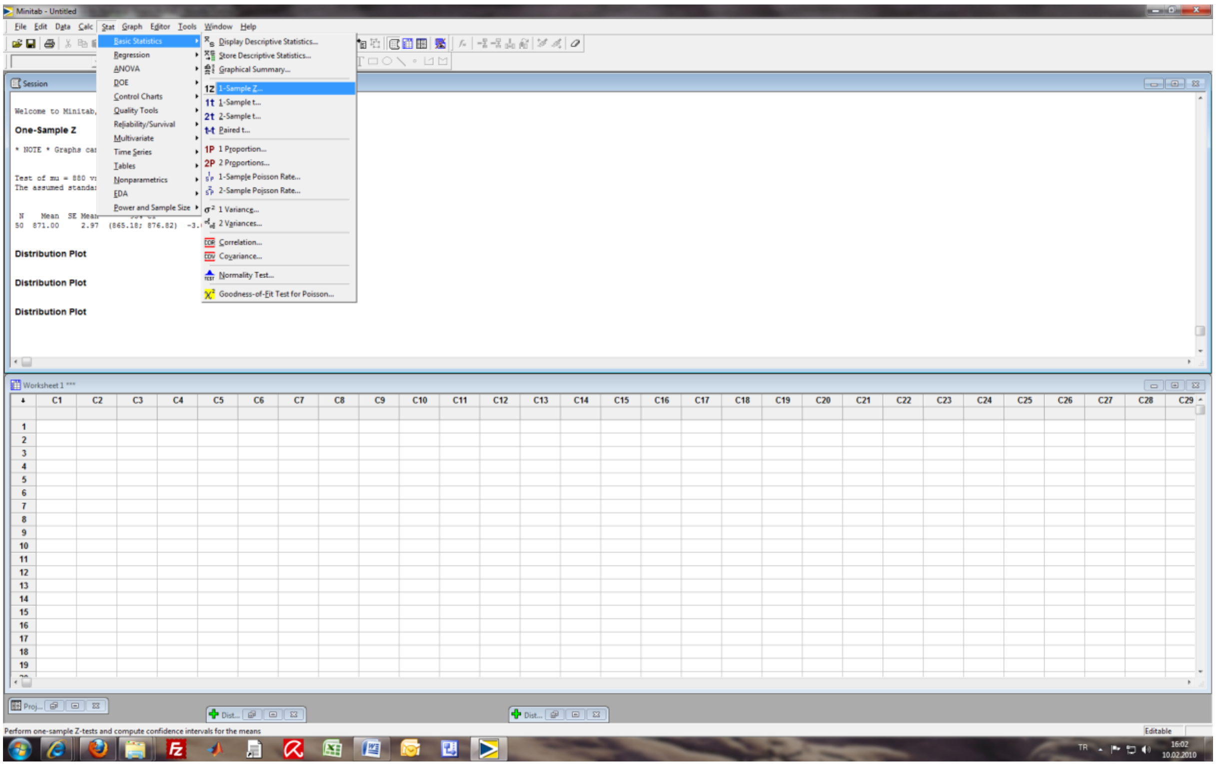
Rejection Region:

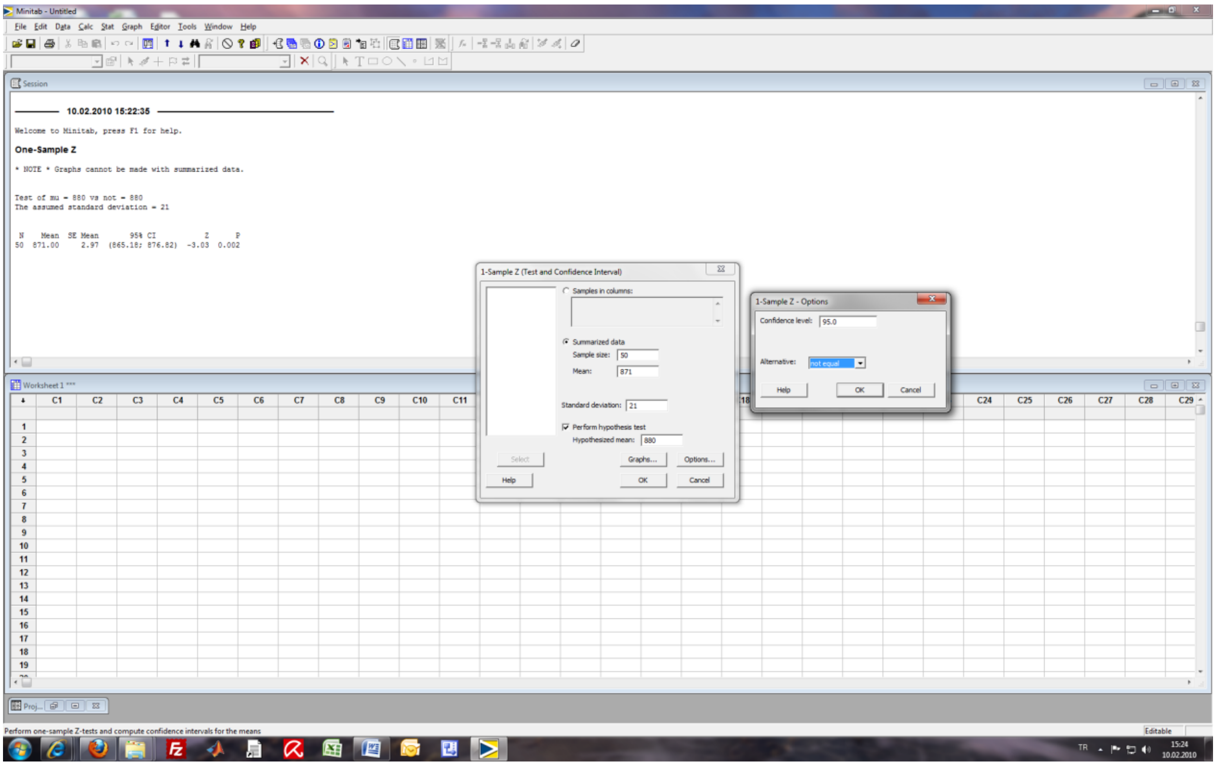


reject H_0 **if** $-z_{\alpha/2} \le z \le z_{\alpha/2}$**, that is** $-1.96 \le z \le 1.96$

Conclusion:

Since z=-3.03 and the calculated value of z falls in the rejection region, the manager can reject the null hypothesis that μ=880 tons and conclude that it has changed. The probability of rejecting H_0 when H_0 is true fairly small probability (0.05). Hence we are reasonably confident that the decision is correct.





One-Sample Z

```
Test of mu = 880 vs not = 880
The assumed standard deviation = 21

 N     Mean  SE Mean         95% CI              Z      P
50   871.00     2.97  (865.18; 876.82)  -3.03  0.002
```

?

Connection between Hypothesis Tests and Confidence Intervals

There is a <u>close relationship</u> between the test of a hypothesis about any parameter, say θ, and the confidence interval for θ. If [l,u] is a 100(1-α)% confidence interval for the parameter θ, the test of size α of the hypothesis

$$H_0: \quad \theta = \theta_0$$
$$H_1: \quad \theta \neq \theta_0$$

will lead to rejection of H_0 <u>if and only if</u> θ_0 is <u>not</u> in the a 100(1-α)% CI [l,u].

in our example

$$H_0: \quad \mu = 880$$
$$H_1: \quad \mu \neq 880$$

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

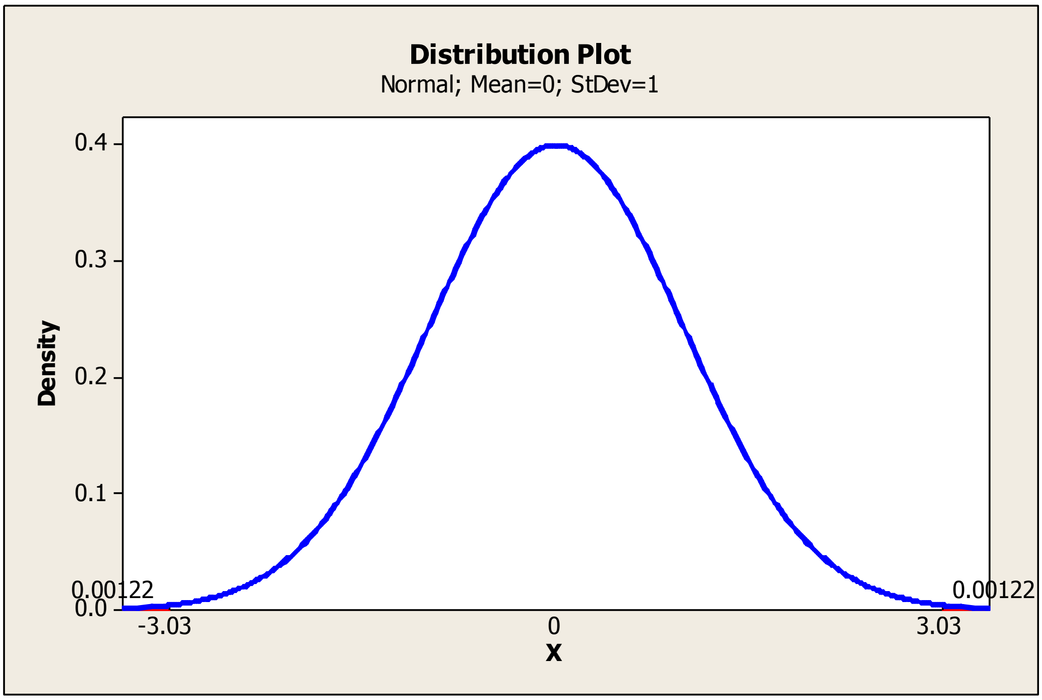
$$871 - 1.96\frac{21}{\sqrt{50}} \leq \mu \leq 871 + 1.96\frac{21}{\sqrt{50}}$$

$$865.18 \leq \mu \leq 876.82$$

We reject H_0 because 880 is not in the CI [865.18,876.82]

Calculating the p-value

The rejection region for this two-tailed test of hypothesis is found in both tails of the normal probability distribution. Since the observed value of the test statistic is Z=-3.03, the smallest region that we can use and still reject H_0 is $|Z| > 3.03$



$$p - value = |Z| > 3.03 =$$
$$P(z > 3.03) + P(z < -3.03) =$$
$$(1 - 0.9988) + 0.0012 = 0.0024$$

Alternative to the critical value approach

Definition**:*(The p-value approach)* The *p-value* or observed significance level of α for which H_0 can be rejected. It is the actual risk of committing a Type I error, if H_0 is rejected based on the observed value of the test statistic. The *p-value* measures the strength of the evidence against H_0.**

If you are reading a research report, how small should the p-value be before you decide to reject ***H_0* ?**

- **If the *p-value is less than 0.01*, H_0 is rejected. The results are *highly significant*.**
- **If the *p-value is between 0.01 and 0.05*, H_0 is rejected. The results are *statistically significant*.**
- **If the *p-value is between 0.05 and 0.1*, H_0 is usually not rejected. The results are only *tending toward statistically significance*.**
- **If the *p-value is greater than 0.1*, H_0 is not rejected. The results are *not statistically significant*.**

One-sided test-left tailed

$$H_0: \quad \mu = 880$$
$$H_1: \quad \mu < 880$$

One-Sample Z

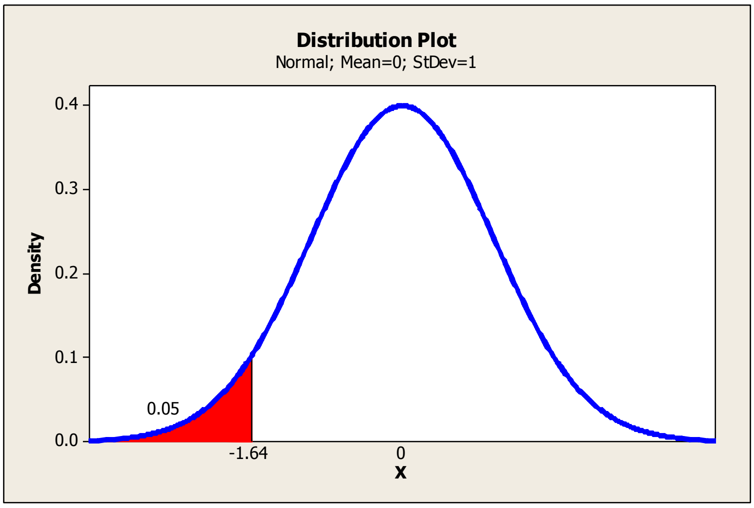
```
Test of mu = 880 vs < 880
The assumed standard deviation = 21


                         95% Upper
 N     Mean   SE Mean        Bound      Z      P
50   871.00      2.97       875.88  -3.03  0.001
```

In the left-tailed test with observed test statistic z=-3.03, the corresponding p-value is

$$p-value = P(z < -3.03) = 0.001$$

This probability is the p-value for the test.



reject H_0

One-sided test-right tailed

$$H_0: \quad \mu = 880$$
$$H_1: \quad \mu > 880$$

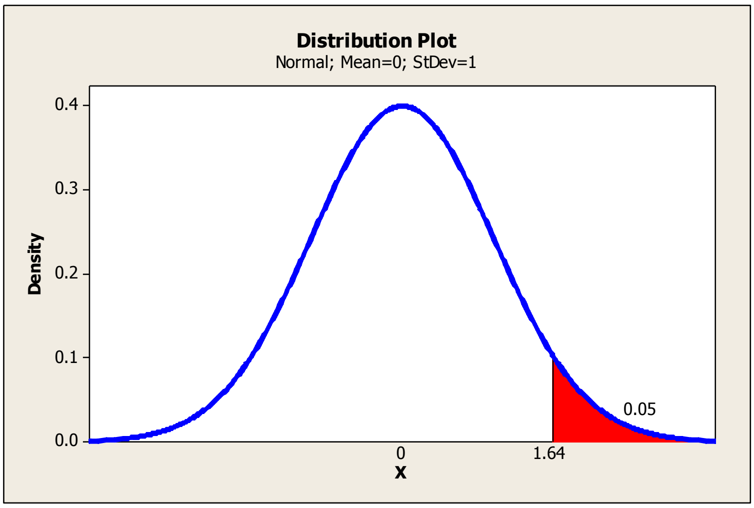
One-Sample Z

```
Test of mu = 880 vs > 880
The assumed standard deviation = 21
```

N	Mean	SE Mean	95% Lower Bound	Z	P
50	871.00	2.97	866.12	-3.03	0.999

In the right-tailed test with observed test statistic z=-3.03, the corresponding p-value is

$$p-value = P(z > -3.03) = 1 - 0.001 = 0.999$$



can not reject H_0

The Power
of a Statistical Test

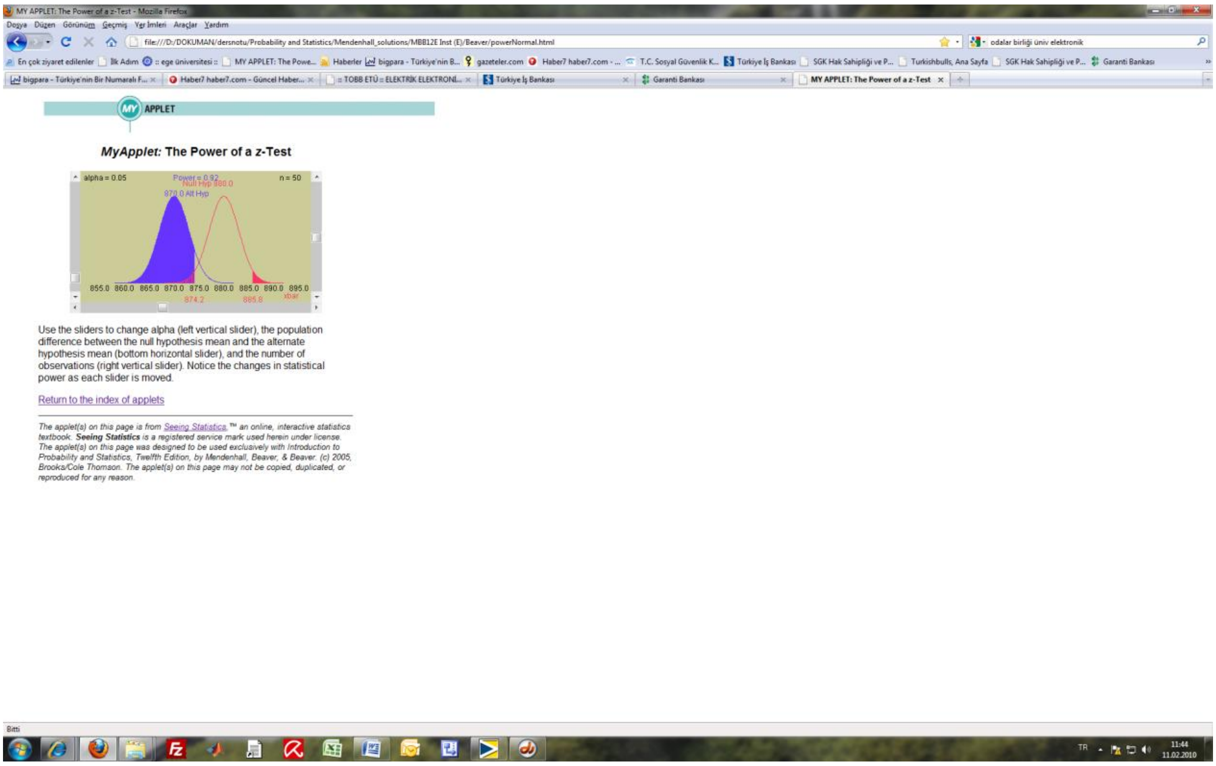
The power of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

The goodness of a statistical test is measured by the size of the two error rates:

- **α, the probability of rejecting H_0 when it is true,**
- **β, the probability of accepting H_0 when H_0 is false and H_a (H_1) is true.**

A "good" test is one for which both of these error rates are small.

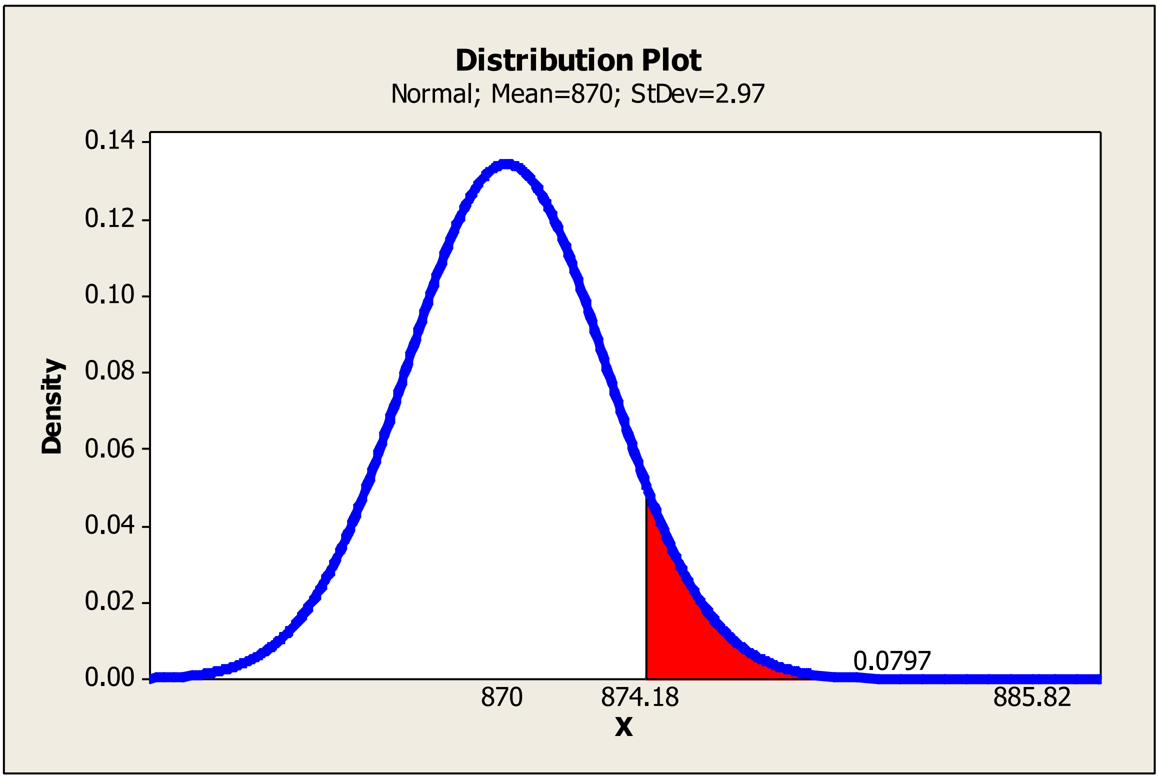
Calculate Type-two error β and the power of the test (1-β) <u>when μ is actually equal to</u> <u>870</u>.

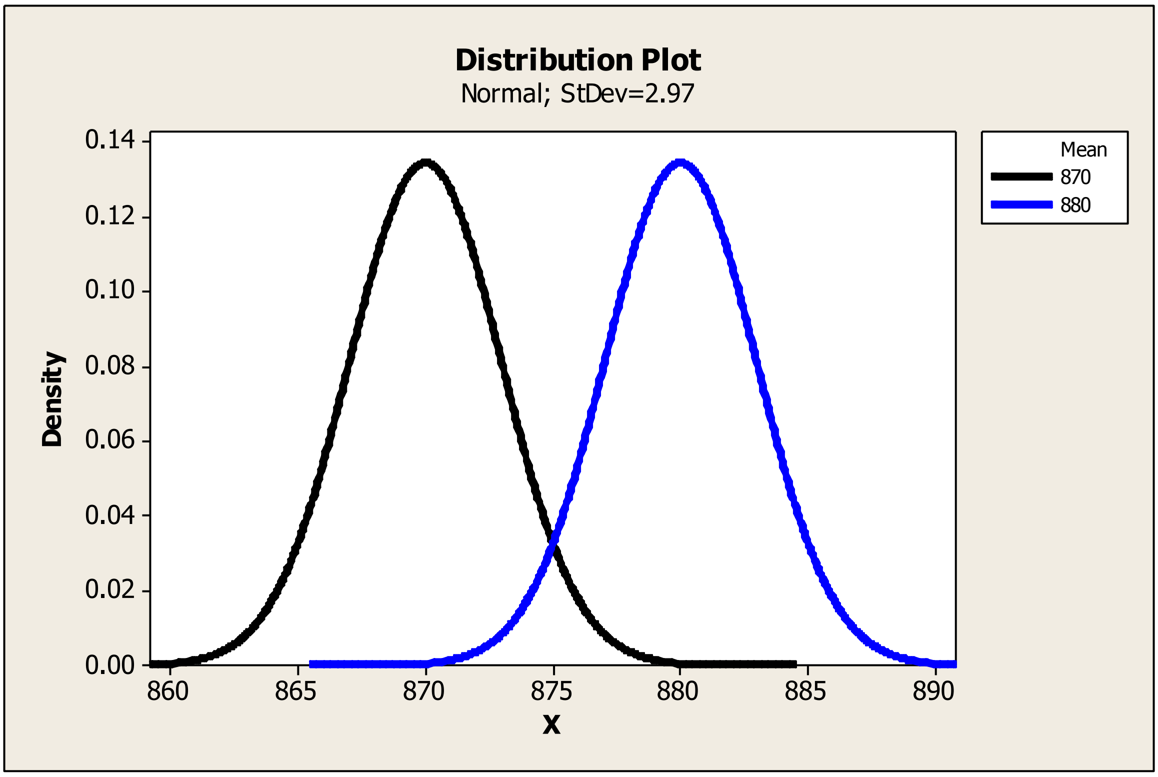


The acceptance region for the previous example is located in the interval $\mu_0 \pm 1.96\left(\dfrac{\sigma}{\sqrt{n}}\right)$**. Substituting numerical values, we get**

$$880 \pm 1.96\left(\frac{21}{\sqrt{50}}\right) \quad \textbf{or} \quad 874.18 \quad \mathit{to} \quad 885.82$$

The probability of accepting H_0 , given μ=870, is equal to the area under the sampling distribution for the test statistic $\bar{X}$ in the interval from 874.18 to 885.82. Since $\bar{X}$ is normally distributed with a mean of 870 and a standard error $SE = 21/\sqrt{50} = 2.97$, β is equal to the area under the normal curve with μ=870 located between 874.18 and 885.82.





Then

$$\begin{aligned}\beta &= P(accept\ \ H_0\ \ when\ \ \ \mu = 870)\\ &= P(874.18 < \bar{X} < 885.82)\\ &= P\left(\frac{874.18-870}{21/\sqrt{50}} < z < \frac{885.82-870}{21/\sqrt{50}}\right)\\ &= P(1.41 < z < 5.33) \cong P(z > 1.41) = 0.0797\end{aligned}$$

Hence, the power of the test is

$$1-\beta = 0.9203$$

The probability of correctly rejecting H_0, given that μ is really equal to 870, is 0.9203 or approximately 92 changes in 100.

Power in Minitab

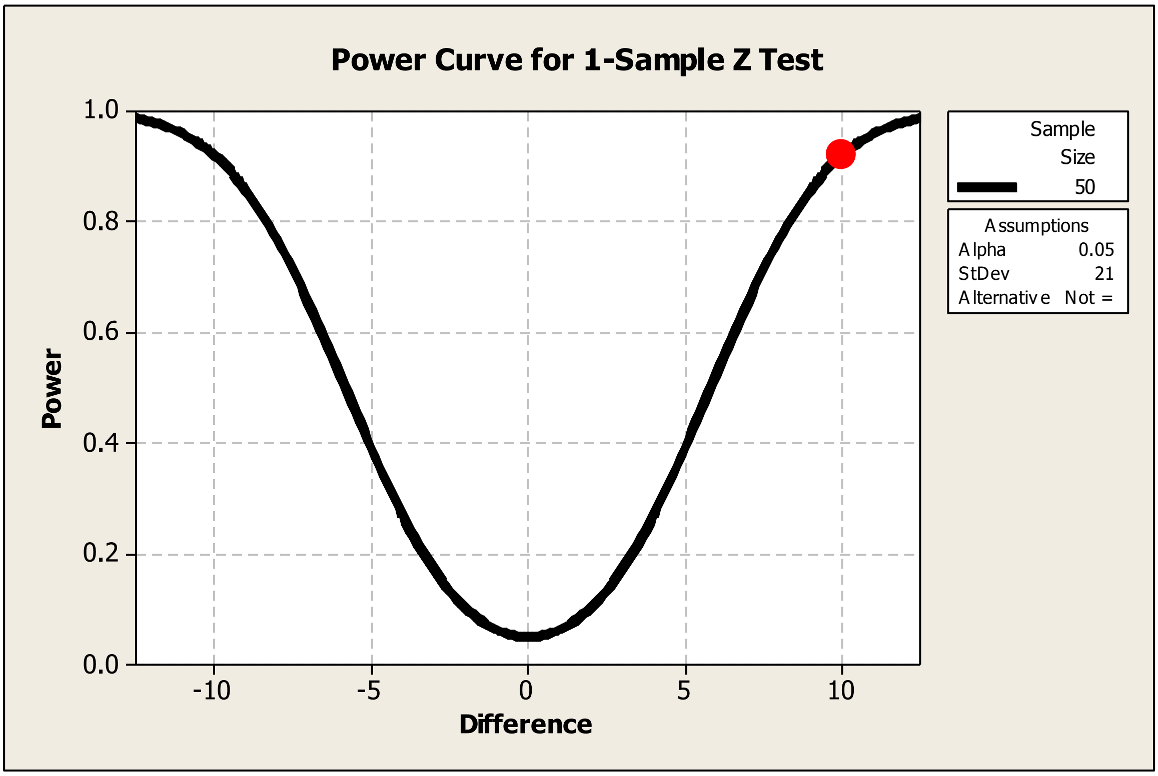
```
MTB > Power;
SUBC>   ZOne;
SUBC>     Sample 50;
SUBC>     Difference 10;
SUBC>     Sigma 21;
SUBC>   GPCurve.
```

Power and Sample Size

```
1-Sample Z Test

Testing mean = null (versus not =
null)
Calculating power for mean = null
+ difference
Alpha = 0.05  Assumed standard
deviation = 21

                  Sample
Difference     Size      Power
        10       50   0.920318
```



Power Curve (Minitab)

Difference	Sample Size	Power
10	50	0.920318

Power Computation by Simulation

```
MTB > random 10000 c1-c50;
SUBC> normal 870 21.
MTB > rmean c1-c50 c52
MTB > let c54=c52>874.18 and c52
<885.22
MTB > sum c54
```

Sum of C54

```
Sum of C54 = 804
MTB > let k1=sum(c54)/count(c54)
MTB > print k1
```

Data Display

```
K1     0.0804000 BETA ESTIMATION
```

```
MTB > let k2=1-k1
MTB > print k2
```

Data Display

```
K2     0.919600POWER ESTIMATION
```

Sample size effect to the power

Increasing Sample Size

Sample size n = 50 >>>> n= 100

The acceptance region for the previous example was located in the interval $\mu_0 \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$. Substituting numerical values before, we get

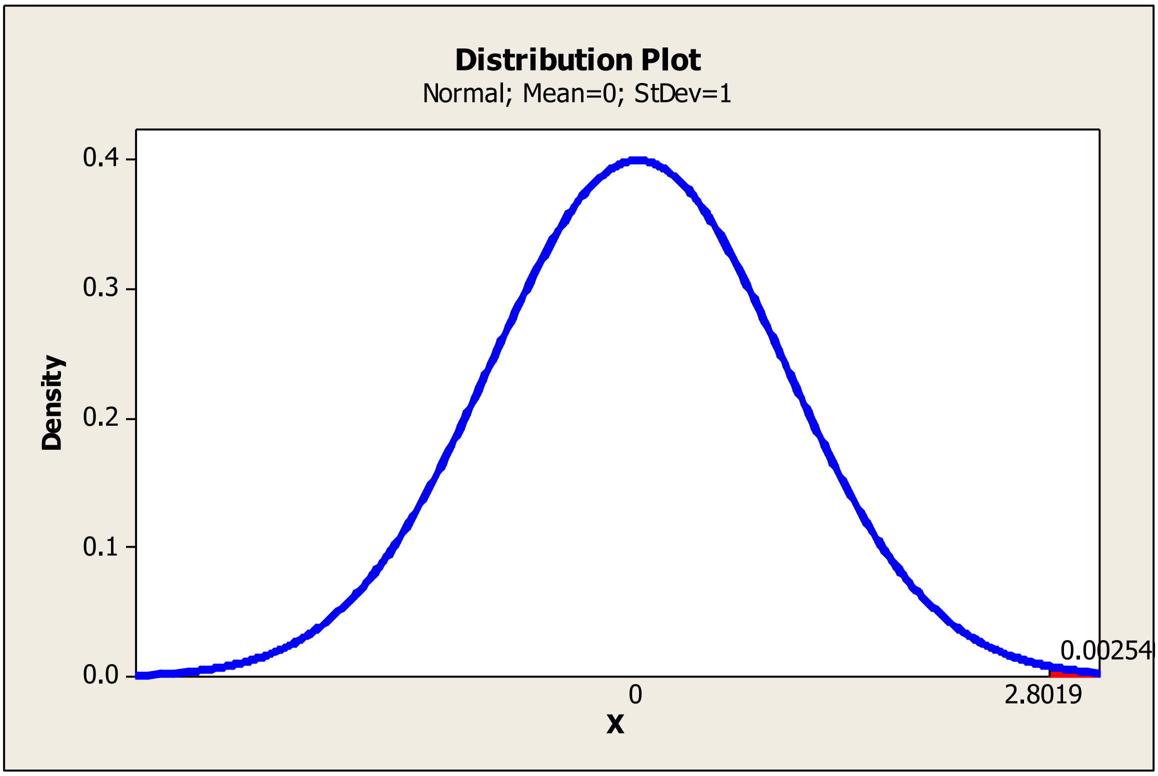
$$880 \pm 1.96\left(\frac{21}{\sqrt{50}}\right) \quad \textbf{or} \quad 874.18 \quad to \quad 885.82$$

But now for n=100, the new acceptance region becomes

$$880 \pm 1.96\left(\frac{21}{\sqrt{100}}\right) \quad \textbf{or} \quad 875.884 \quad to \quad 884.116$$

Then the new type two error

$$\beta = P(accept\ \ H_0\ \ when\ \ \ \mu = 870)$$
$$= P(875.884 < \overline{X} < 884.116)$$
$$= P\left(\frac{875.884-870}{21/\sqrt{100}} < z < \frac{884.116-870}{21/\sqrt{100}}\right)$$
$$= P(2.8019 < z < 6.72)$$
$$\cong P(z > 2.8019) = 0.00254$$



$$\beta = 0.00254$$

Hence, the power of the test is

$$1-\beta = 0.99746$$

Power Curve for 1-Sample Z Test

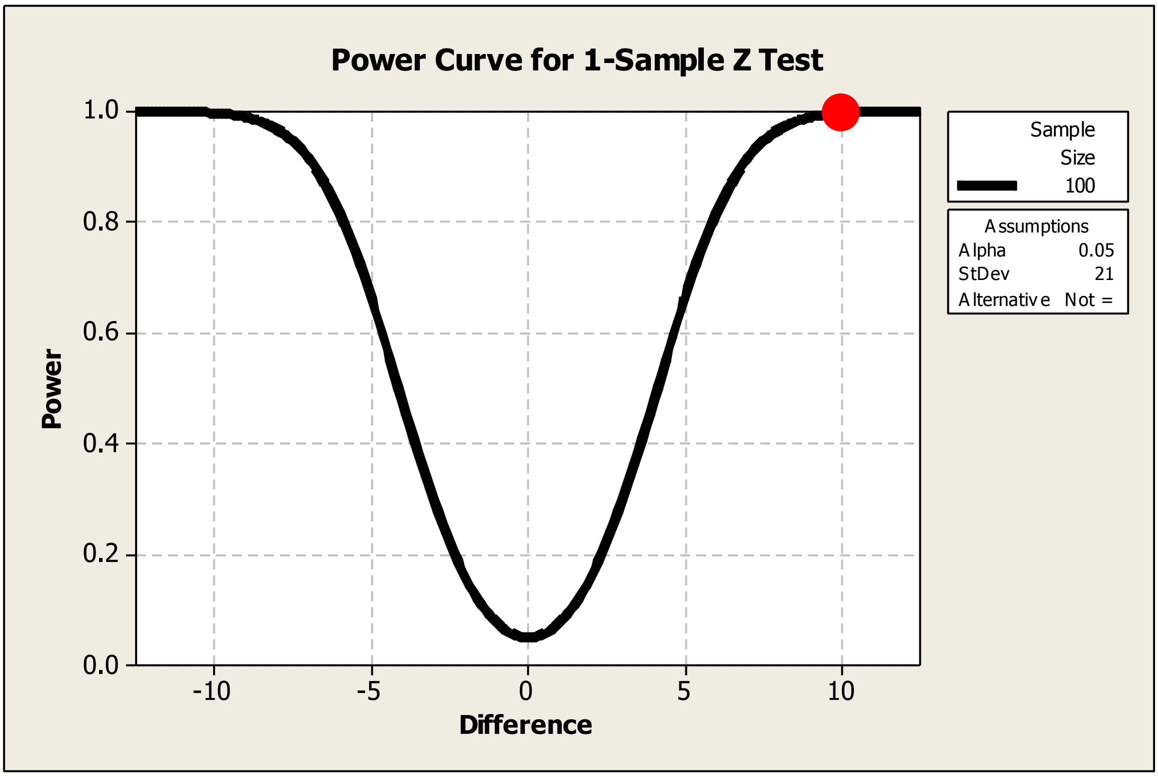
```
MTB > Power;
SUBC>   ZOne;
SUBC>     Sample 100;
SUBC>     Difference 10;
SUBC>     Sigma 21;
SUBC>   GPCurve.
```

Power and Sample Size

```
1-Sample Z Test

Testing mean = null (versus not =
null)
Calculating power for mean = null
+ difference
Alpha = 0.05  Assumed standard
deviation = 21


            Sample
Difference    Size     Power
        10     100  0.997460
```



Power Curve (Minitab)

	Sample	
Difference	Size	Power
10	100	0.997460

Increasing sample size provides more information about the population and therefore increases power.

Power Computation by Simulation (increasing sample size)

```
MTB > random 100000 c1-c100;
SUBC> normal 870 21.

MTB > rmean c1-c100 c102
MTB > let c104=c102>875.884 and
c102<884.116
MTB > sum c104
```

Sum of C104

```
Sum of C104 = 273

MTB > let k1=sum(c104)/count(c104)
MTB > print k1
```

Data Display

```
K1      0.00273000 (BETA ESTIMATION)

MTB > let k2=1-k1
MTB > print k2
```

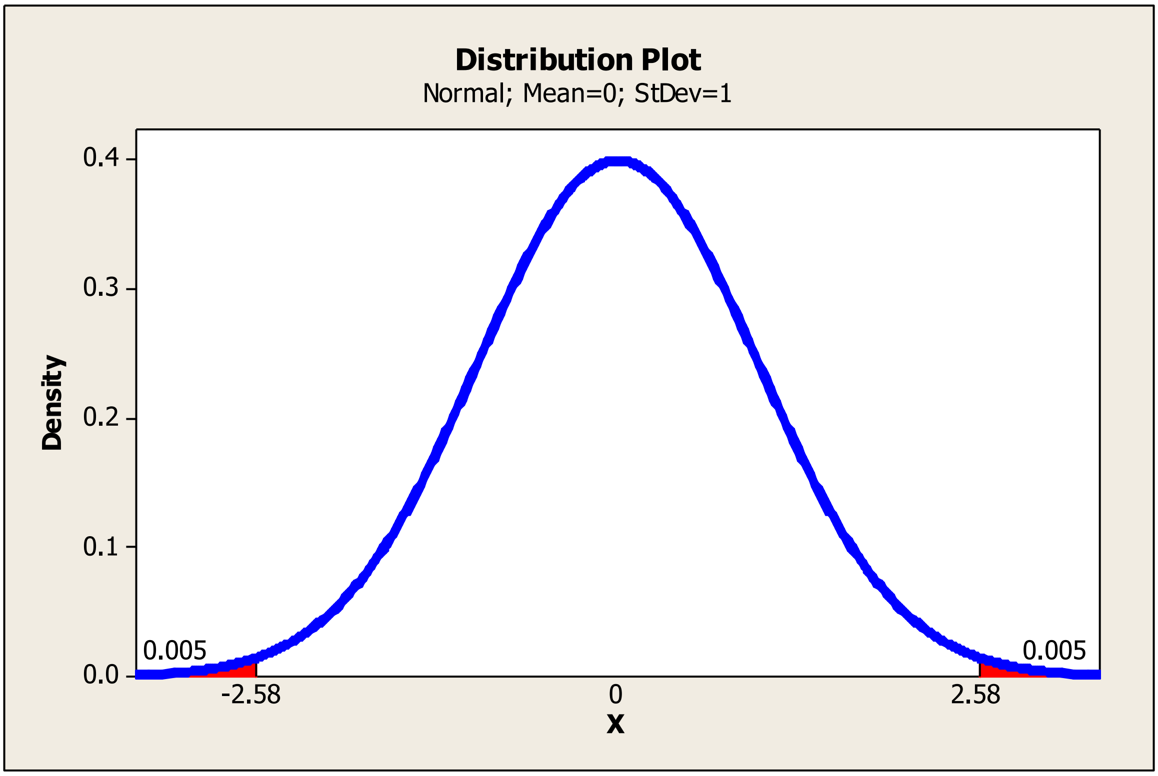
Data Display

```
K2      0.997270POWER ESTIMATION
```

Effect of α to the power

Effect of α

We assume α=0.01 instead of 0.05



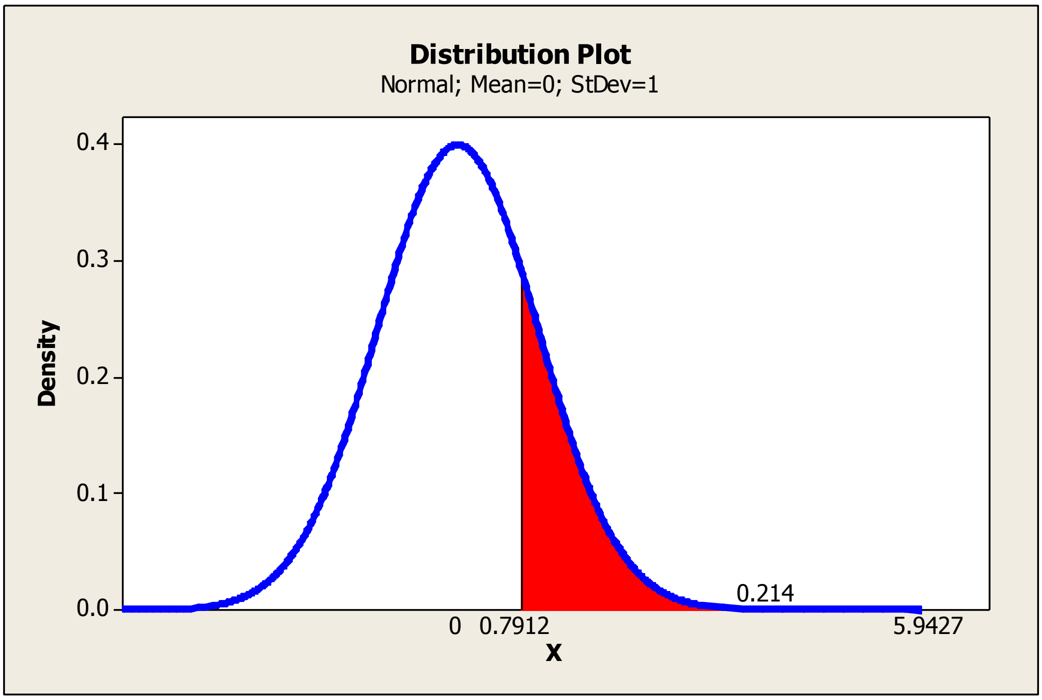
The acceptance region for the previous example is located in the interval $\mu_0 \pm 2.58\left(\frac{\sigma}{\sqrt{n}}\right)$. Substituting numerical values, we get

$$880 \pm 2.58\left(\frac{21}{\sqrt{50}}\right) \quad \textbf{or} \quad 872.35 \quad to \quad 887.65$$

```
N     Mean   SE Mean         99% CI
50   880.00      2.97   (872.35; 887.65)
```

Then

$$\begin{aligned}\beta &= P(accept \;\; H_0 \;\; when \;\; \mu = 870) \\ &= P(872.35 < \bar{X} < 887.65) \\ &= P\left(\frac{872.35-870}{21/\sqrt{50}} < z < \frac{887.65-870}{21/\sqrt{50}}\right) \\ &= P(0.7912 < z < 5.9427) \cong P(z > 0.79.12) = 0.214\end{aligned}$$



Hence, the power of the test is

$$1-\beta = 0.786$$

A small α value decreases power

Magnitude of population effect to the power

Calculate Type-two error β and the power of the test (1-β) when μ is actually equal to 875.

Difference 5(similar populations)

n=50 α=0.05

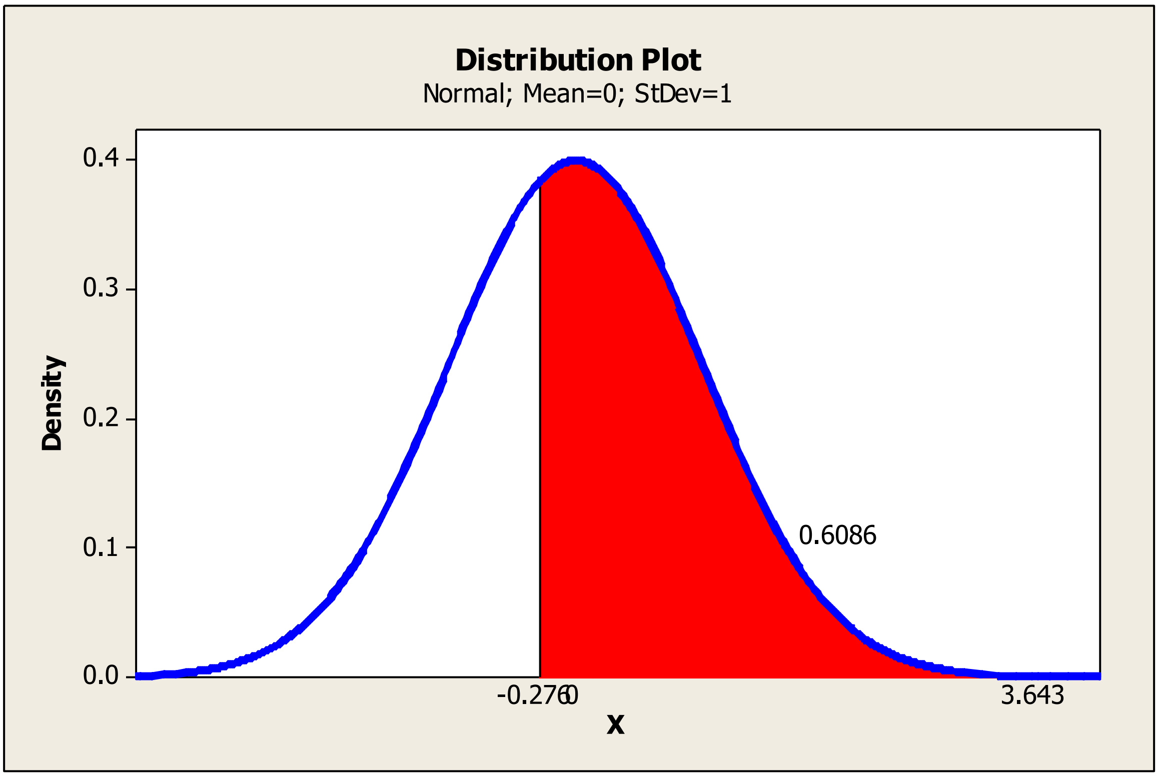
The acceptance region:

$$\mu_0 \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$880 \pm 1.96\left(\frac{21}{\sqrt{50}}\right)$$ **or** $$874.18 \quad \textit{to} \quad 885.82$$

Then

$$\beta = P(accept\ \ H_0\ \ when\ \ \ \mu = 875)$$
$$= P(874.18 < \overline{X} < 885.82)$$
$$= P\left(\frac{874.18 - 875}{21/\sqrt{50}} < z < \frac{885.82 - 875}{21/\sqrt{50}}\right)$$
$$= P(-0.276 < z < 3.643) = 0.6086$$



Hence, the power of the test is

$$1 - \beta = 1 - 0.6086 = 0.3914$$

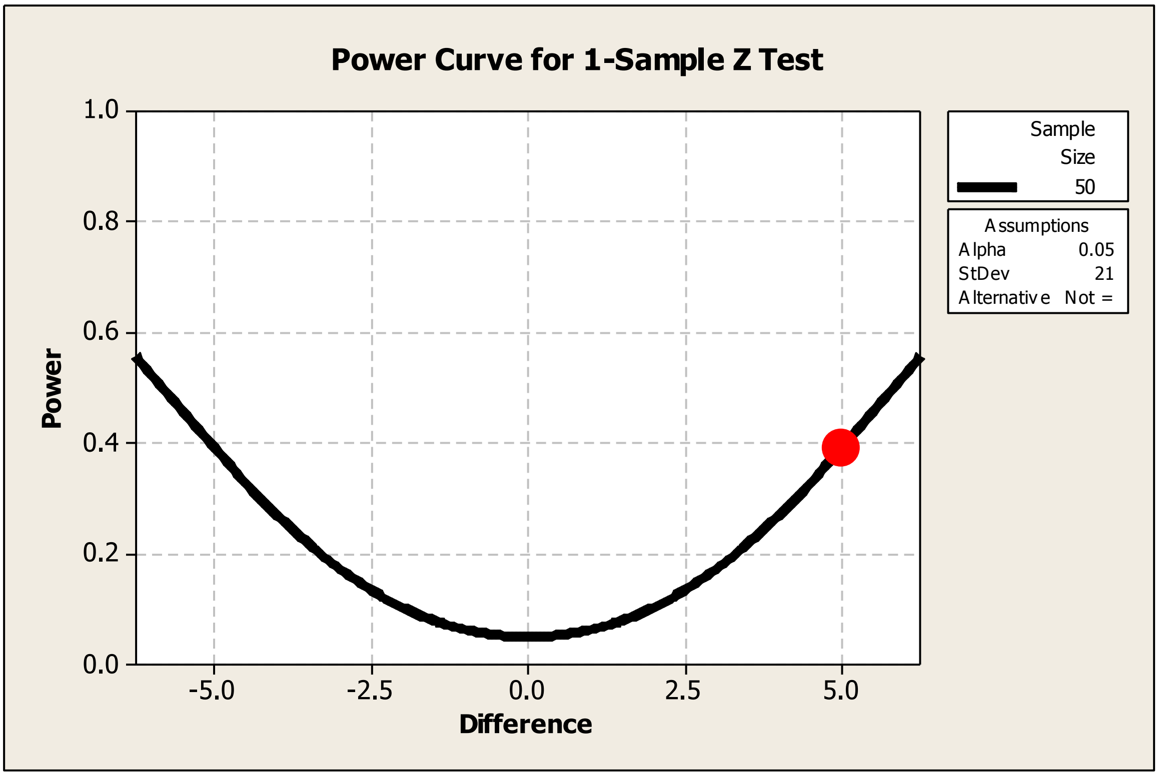
```
MTB > Power;
SUBC>   ZOne;
SUBC>     Sample 50;
SUBC>     Difference 5;
SUBC>     Sigma 21;
SUBC>   GPCurve.
```

Power and Sample Size

```
1-Sample Z Test

Testing mean = null (versus not =
null)
Calculating power for mean = null
+ difference
Alpha = 0.05  Assumed standard
deviation = 21

            Sample
Difference    Size     Power
         5      50  0.391264
```



Power Curve (Minitab)

Difference	Sample Size	Power
5	50	0.391264

The more similar populations are, the more difficult it is to detect a difference. Therefore, power decreases.

Power (Conclusions)

In a hypothesis test, the likelihood that you will find a significant effect or difference when one truly exists. Power is the probability that you will correctly reject the null hypothesis when it is false.

A number of factors affect power:

- ✓ **Increasing sample size provides more information about the population and therefore increases power.**
- ✓ **A large α value increases power because you are more likely to reject the null hypothesis with larger a values.**
- ✓ **When σ is small, it is easier to detect a difference, which increases power.**
- ✓ **Magnitude of population effect: The more similar populations are, the more difficult it is to detect a difference. Therefore, power decreases.**

You can calculate power before you collect data to ensure that your hypothesis test will detect significant differences or effects. For example, a pharmaceutical company wants to see how much power their hypothesis test has to detect differences among three different diabetes treatments. To increase power, they can increase the sample size to get more information about the population of diabetes patients using these medications. Also, they can try to decrease error variance by following good sampling practices.

You can also calculate power to understand the power of tests that you have already conducted. For example, an automobile parts manufacturer performs an experiment comparing the weight of two steel formulations, and the results are not statistically significant. Using Minitab, the manufacturer can calculate power based on the minimum difference that they would like to see. If the power to detect this difference is low, they may want modify the experimental design to increase the power and continue to evaluate the same problem. However, if the power is high, they may conclude that the two steel formulations are not different and discontinue further experimentation.

Power is calculated by 1-β, or 1 - Type II error (failing to reject the null hypothesis when it is false). As a (the level of significance) increases, the probability of a type II error (β) decreases. Therefore, as a increases, power also increases. Keep in mind that increasing a also increases **the probability of Type I error (rejecting the null hypothesis when it is true).**