IX. One-Sample Hypothesis Testing

A. A Few Important Definitions

1. **Hypothesis Testing - Scientific method-based means for using sample data to evaluate conjectures about a population.**

2. **Null Hypothesis - Statement of the conjectured value(s) for the parameter that includes (but is not necessarily limited to) equality between the conjectured value and the tested parameter. Usually denoted**

   H_0: parameter ($\leq$ = $\geq$) hypothesized value

   This is equivalent to a claim that the difference between the observations and the hypothesized value are due to *random variation*.

3. **Alternative Hypothesis - Statement of the conjectured value(s) for the parameter that is mutually exclusive and collectively exhaustive with respect to the Null Hypothesis (and so includes the < and/or > relationship between conjectured value and the tested parameter. Usually denoted**

   H_a: parameter ($<$ $\neq$ $>$) hypothesized value.

   This is equivalent to a claim that the difference between the observations and the hypothesized value are systematic (i.e., due to *something other than random variation*.

   Note that our conclusion is stated with regards to the null hypothesis, which can either be i) rejected or ii) not rejected – that is, we *never accept* the null hypothesis.

4. **Critical Region - Area containing all possible estimated values of the parameter that will result in rejection of the null hypothesis.**

5. **Critical Value(s) – The value(s) that divide the critical region(s) from the 'do not reject' region**

6. **Test Statistic – Sample-based value that will be compared to the critical region(s) to decide whether to reject or not reject the null hypothesis. The generic form of the test statistic is**

$$\text{Test Statistic} = \frac{\text{Observed Value of the Test Statistic} - \text{Hypothesized Value of the Test Statistic}}{\text{Standard Error of the Test Statistic}}$$

7. **Decision Rule – Statement that specifies values of the test statistic that will result in rejection or non-rejection of the null hypothesis**

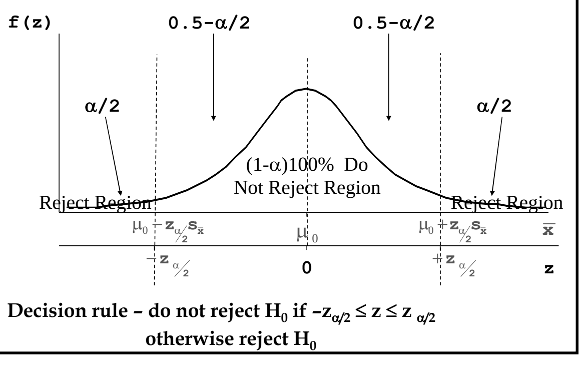
8. **Two-Tailed Hypothesis Test - Evaluation of a conjecture for which sample results that are *either sufficiently less than or greater than the conjectured value of the parameter* will result in rejection of the null hypothesis, i.e., for a null hypothesis that only includes equality between the conjectured value and the tested parameter.**

9. **One-Tailed Hypothesis Test - Evaluation of a conjecture for which sample results that are *only sufficiently less than or greater than the conjectured value of the parameter* will result in rejection of the null hypothesis, i.e., for a null hypothesis that includes an inequality between the conjectured value and the tested parameter.**

10. **Upper-Tailed Test – Hypothesis test for which sample results that are *only sufficiently greater than the conjectured value of the parameter* will result in rejection of the null hypothesis)**

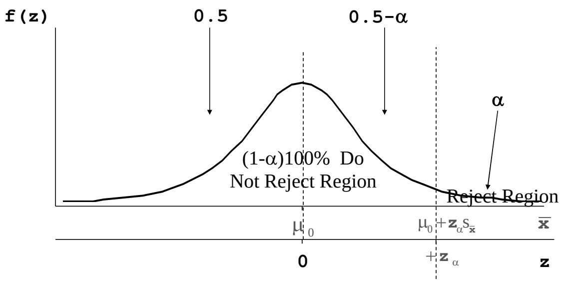
11. **Lower-tailed Test – Hypothesis test for which sample results that are *only sufficiently less than the conjectured value of the parameter* will result in rejection of the null hypothesis**

12. **Level of Significance – The probability of rejecting the null hypothesis at equality *when it is actually true*. For a two-tailed test:**



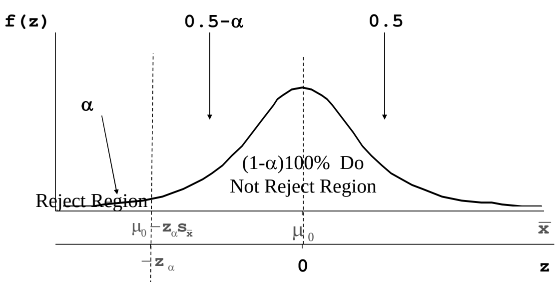
Decision rule – do not reject H_0 if $-z_{\alpha/2} \leq z \leq z_{\alpha/2}$
otherwise reject H_0

For an upper-tailed test:



Decision rule – do not reject H_0 if $z \le z_\alpha$
otherwise reject H_0

For a lower-tailed test:



Decision rule – do not reject H_0 if $-z_\alpha \le z$
otherwise reject H_0

13. Type I Error - Rejection of a true null hypothesis. The probability of this occurrence (given that the null hypothesis is true) is denoted as α.

14. Type II Error – Non-rejection of a false null hypothesis. The probability of this occurrence (given that the null hypothesis is false in some specific manner) is denoted as β.

Conclusion	State of Nature: H_0 is True	State of Nature: H_0 is False
Do Not Reject H_0	Correct Decision Probability=$1-\alpha$	Type II Error Probability=β
Reject H_0	Type I Error Probability=α	Correct Decision Probability=$1-\beta$

A note about Type I and Type II Errors - The probability of a Type I Error is actually a conditional probability

$$P[\text{Reject } H_0 \mid H_0 \text{ is true}] = \alpha$$

…so we also have that

$$P[\text{Do Not Reject } H_0 \mid H_0 \text{ is true}] = 1- \alpha$$

These is only one condition under which Type I can occur – *the null hypothesis must be true*!

Note that, for this null hypothesis, we could write

$$P[\text{Reject } H_0 \mid \mu = \mu_0] = \alpha$$

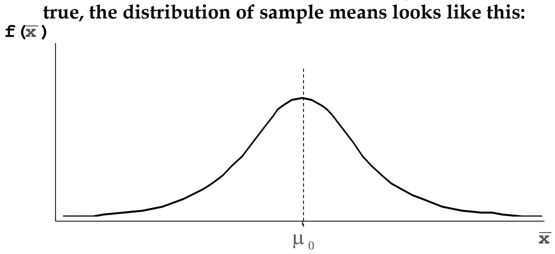
…so we also have that

$$P[\text{Do Not Reject } H_0 \mid \mu = \mu_0] = 1- \alpha$$

If we are testing the null hypothesis

$$H_0: \mu = \mu_0$$

then, under the condition that the null hypothesis is true, the distribution of sample means looks like this:

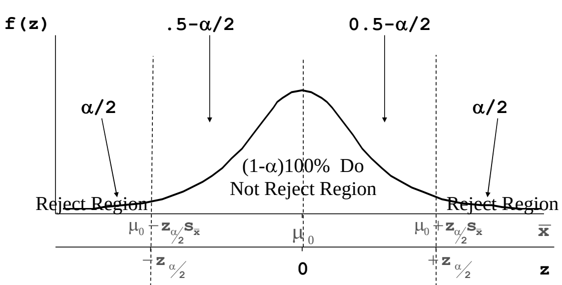


provided the sample means are normally distributed (by virtue of the Central Limit Theorem or a normally distributed parent population).

Under the condition that the null hypothesis

$$H_0: \mu = \mu_0$$

is true, the area in the reject regions is equal to α, i.e.



and the area in the do not reject region is equal to $1 - \alpha$

The probability of a Type II Error is also a conditional probability that is frequently written

P[Do Not Reject H_0 | H_0 is false] = β

…so we also have that

P[Reject H_0 | H_0 is false] = 1- β

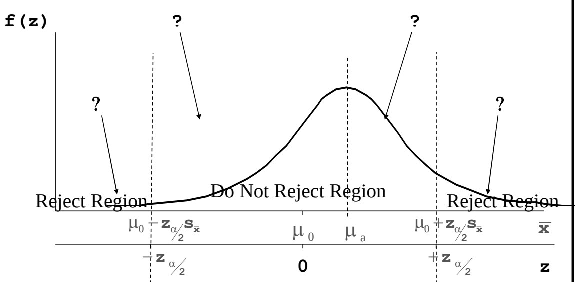
However, Type II Error is nonrejection of a false null hypothesis – ***there are an infinite number of ways in which this condition can exist*****!**

Thus, we usually discuss Type II Error in terms of a specific manner in which the null hypothesis is false, i.e., what is the probability of Type II Error if the true value of the mean is $\mu_a \neq \mu_0$?

Under the condition that the null hypothesis

H_0: $\mu = \mu_0$

is false because $\mu=\mu_a\neq\mu_0$, the area in the reject regions is

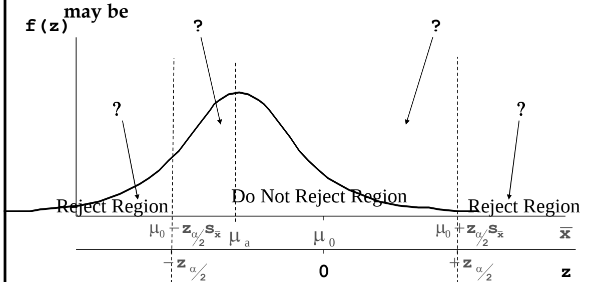


Note that the area of the ***do not reject*** **region is now β**

If the null hypothesis

H_0: $\mu = \mu_0$

is false in some other way, the area in the reject regions may be



The area of the ***do not reject*** **region has changed!**

Thus, the probability of Type II Error depends on the *manner in which the null hypothesis is false*. For a given manner in which the null hypothesis is false (i.e., the true value of the mean is $\mu_a \neq \mu_0$), the probability of a Type II Error is

$$P[\text{Do Not Reject } H_0: \mu = \mu_0 \mid \mu = \mu_a \neq \mu_0] = \beta$$

…so we also have that

$$P[\text{Reject } H_0: \mu = \mu_0 \mid \mu = \mu_a \neq \mu_0] = 1 - \beta$$

Again, there are an infinite number of ways the null hypothesis can be false, and so an infinite number of ways to make Type II Error!

Finally, note that $P[\text{Reject } H_0: \mu = \mu_0 \mid \mu = \mu_a \neq \mu_0] = 1 - \beta$ is referred to as the *power of the hypothesis test* when $\mu = \mu_a \neq \mu_0$ and that this probability can be calculated.

B. Steps in Hypothesis Testing

1. State the Null & Alternative Hypotheses
2. Select the Appropriate Test Statistic
3. State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule
4. Calculate the Test Statistic from the sample data
5. Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis. Interpret your results.

C. Large Sample ($n \geq 30$) Tests About a Single Mean (Critical Value Approach)

- Null and Alternative Hypotheses

$$H_0: \mu \ (\leq = \geq) \ \mu_0 \qquad H_a: \mu \ (< \neq >) \ \mu_0$$

- Appropriate Test Statistic

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

*Note that if population standard deviation σ is known, it can be directly substituted for the sample standard s in this formula.

An Example - The superintendent of a printing plant has selected a random sample of 100 rolls of paper from a large shipment. The average length of the sample rolls is 516 feet, with a variance of 2704 feet. The superintendent wants you to test that the mean length of the rolls is 525 feet at a significance level of 0.05.

1. **State the Null and Alternative Hypotheses**

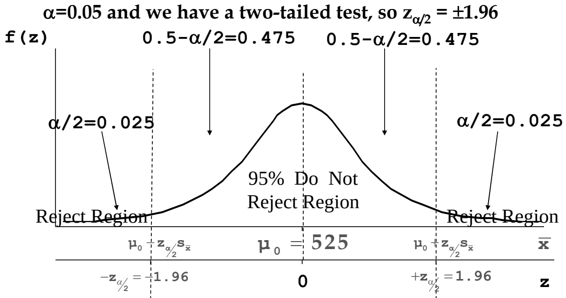
$$H_0: \mu = 525$$
$$H_a: \mu \neq 525$$

2. **Select the Appropriate Test Statistic – n = 100 ≥ 30, so use**

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

3. **State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule**

α=0.05 and we have a two-tailed test, so $z_{\alpha/2} = \pm 1.96$



Decision rule – do not reject H_0 if $-1.96 \leq z \leq 1.96$ otherwise reject H_0

4. **Calculate the Test Statistic**

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{516 - 525}{52/\sqrt{100}} = -1.73077$$

(note that the *sample variance* is 2704!)

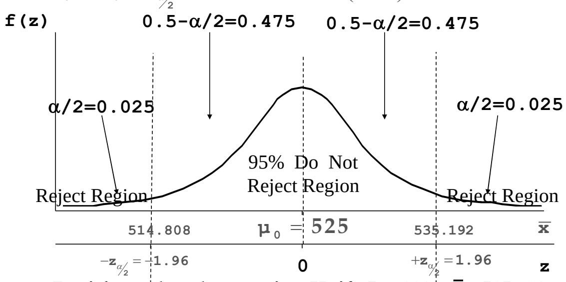
5. **Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis**

$-z_{\alpha/2} \leq z \leq z_{\alpha/2}$, i.e., $-1.96 \leq -1.73077 \leq 1.96$

so do not reject H_0. The sample evidence does not support the claim that the mean length of the rolls differs from 525 feet.

Note that we could compute the critical values in terms of the random variable X (feet of paper on a roll). We have:

$$\bar{x}_L = \mu_0 - z_{\alpha/2} s_{\bar{x}} = 525 - 1.96(5.2) = 514.808$$
$$\bar{x}_U = \mu_0 + z_{\alpha/2} s_{\bar{x}} = 525 + 1.96(5.2) = 535.192$$



Decision rule – do not reject H_0 if $-514.808 \le \bar{x} \le 535.192$ otherwise reject H_0

Another Example - The superintendent of a printing plant has selected a random sample of 100 rolls of paper from a large shipment. The average length of the sample rolls is 516 feet, with a variance of 2704 feet. The supplier has guaranteed that the mean length of the rolls is at least 525 feet. The superintendent needs to determine if the mean length of the rolls is at least 525 feet at a significance level of 0.05.

1. **State the Null and Alternative Hypotheses**

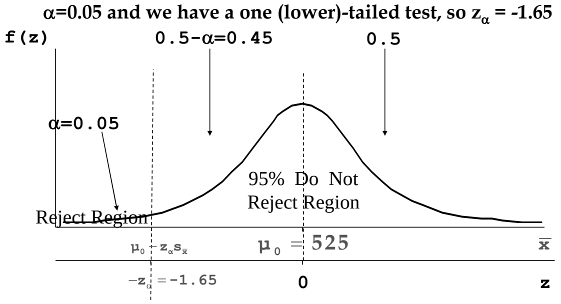
$$H_0: \mu \ge 525$$
$$H_a: \mu < 525$$

2. **Select the Appropriate Test Statistic – $n = 100 \ge 30$, so use**

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

3. **State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule**

α=0.05 and we have a one (lower)-tailed test, so $z_\alpha = -1.65$



Decision rule – do not reject H_0 if $-1.65 \le z$ otherwise reject H_0

4. Calculate the Test Statistic

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{516 - 525}{52/\sqrt{100}} = -1.73077$$

5. Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis

$-z_\alpha > z$, i.e., $-1.65 > -1.73077$

so reject H_0. The sample evidence does not support the claim that the mean length of the rolls is at least 525 feet.

D. Small Sample (n<30) Tests About a Single Mean – Normal Parent Population (Critical Value Approach)

- Null and Alternative Hypotheses

H_0: μ ($\leq = \geq$) μ_0 $\qquad$ H_a: μ ($< \neq >$) μ_0

- Appropriate Test Statistic

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$df = n - 1$$

An Example – A furniture store has taken a random sample of twenty-five customers. The mean age of the respondents is 38 and the standard deviation is 3.5. Previous samples have shown the age of the store's customers to be approximately normal. Test the claim that the mean age of all consumers of a furniture store is under 40 years, at the $\alpha = 0.10$ level of significance.

1. State the Null and Alternative Hypotheses

H_0: $\mu \geq 40$

H_a: $\mu < 40$

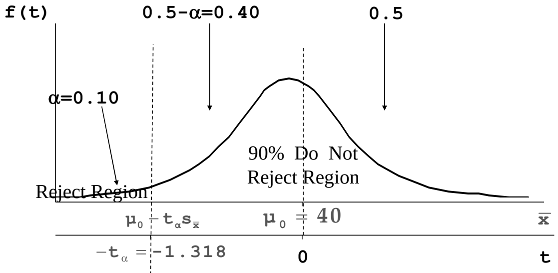
2. Select the Appropriate Test Statistic – n = 25 < 30 and x (age of customers) is normally distributed, so use

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$df = n - 1 = 25 - 1 = 24$$

3. State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule

α=0.10 and we have a one (lower)-tailed test with 24 degrees of freedom, , so t_α = -1.318



Decision rule – do not reject H_0 if $-1.318 \le t$ otherwise reject H_0

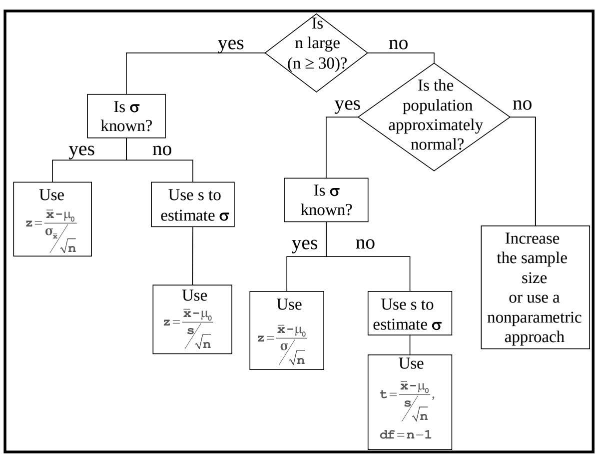
4. Calculate the Test Statistic

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{38 - 40}{3.5/\sqrt{25}} = -2.857$$

5. Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis

$-t_\alpha > t$, i.e., $-1.318 > -2.857$

so reject H_0. The sample evidence does not refute the claim that the mean age of the furniture store's customers is under 40.



E. Tests About a Single Proportion ($np_0 \geq 5$, $n(1-p_0) \geq 5$) (Critical Value Approach)

- **Null and Alternative Hypotheses**

H_0: p (≤ = ≥) p_0 H_a: p (< ≠ >) p_0

- **Appropriate Test Statistic**

$$z=\frac{\bar{p}-p_0}{\sigma_{\bar{p}}}=\frac{\bar{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

An Example – The Gibby Glass company is concerned that public awareness of their brand of kitchenware is low. They feel that, in order to be successful, over 35% of consumers should be aware of their brand. They conduct a survey of 144 randomly selected consumers and find that 54 of the respondents are aware of their brand. Test the hypothesis that brand awareness for Gibby Glass kitchenware is sufficient at the α=.01 level of significance.

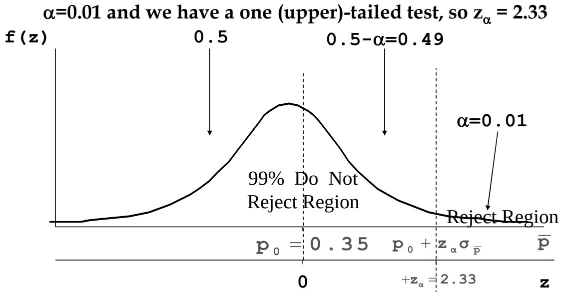
1. **State the Null and Alternative Hypotheses**

H_0: $p \leq 0.35$
H_a: $p > 0.35$

2. **Select the Appropriate Test Statistic – $np_0 \geq 5$, $n(1-p_0) \geq 5$, so use**

$$z=\frac{\bar{p}-p_0}{\sigma_{\bar{p}}}=\frac{\bar{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

3. **State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule**
   α=0.01 and we have a one (upper)-tailed test, so z_α = 2.33



Decision rule – do not reject H_0 if $z \leq 2.33$
otherwise reject H_0

4. Calculate the Test Statistic

$$z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}} = \frac{54/144 - 0.35}{\sqrt{\frac{0.35(1-0.35)}{144}}} = 0.629$$

5. Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis

$z \leq z_\alpha$, i.e., $0.629 \leq 2.33$

so do not reject H_0. The sample evidence does not support the claim that proportion of customers aware of the Gibby Glass brand of kitchenware exceeds 0.35.

F. p-Values – Under the condition that the *null hypothesis is true*, the p-value is the probability of obtaining a sample result that is at least as extreme as what we actually observed in our sample. The p-value is the smallest significance level at which the null hypothesis would not be rejected, and is sometimes referred to as the observed level of significance.

Finding the p-Value

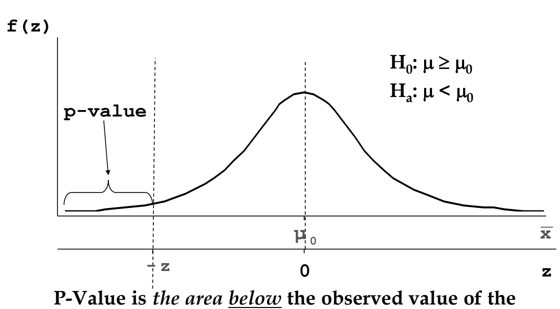
find the value in the appropriate (z or t) table that corresponds to the calculated value of the test statistic and

i) report this value (the cumulative probability) if the test is lower-tailed

ii) report (1.000 - this value) if the test is upper tailed

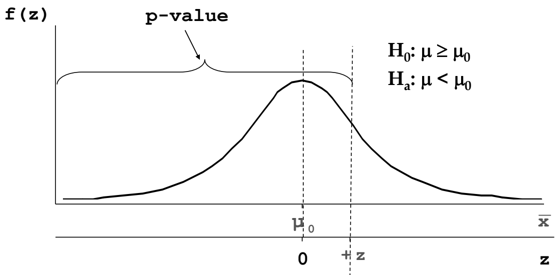
iii) report double the smaller of the two values given above if the hypothesis is two tailed

Finding p-values – Lower tailed tests: find the cumulative probability associated with the calculated value of the test statistic:



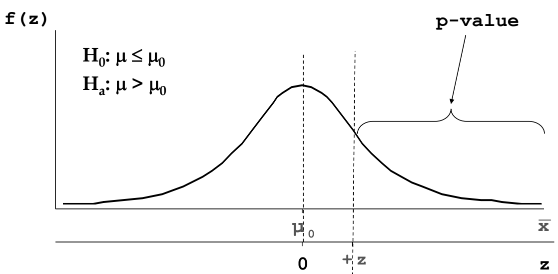
P-Value is *the area* <u>below</u> the observed value of the test statistic

Finding p-values – Lower tailed tests: find the cumulative probability associated with the calculated value of the test statistic:



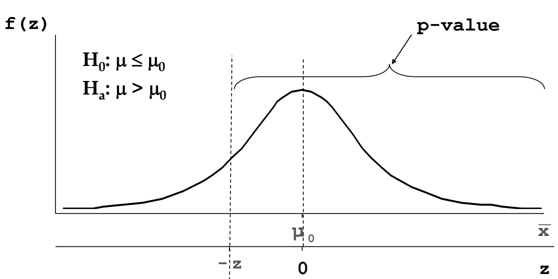
P-Value is *the area <u>below</u>* the observed value of the test statistic

Finding p-values – Upper tailed tests: find 1 - the cumulative probability associated with the calculated value of the test statistic:



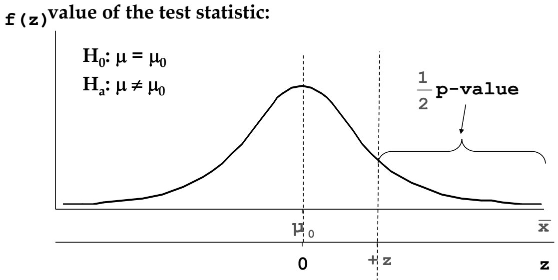
P-Value is *the area <u>above</u>* the observed value of the test statistic

Finding p-values – Upper tailed tests: find 1 - the cumulative probability associated with the calculated value of the test statistic:



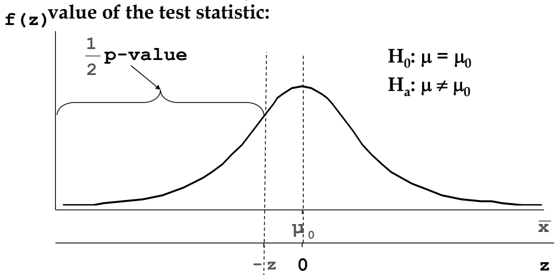
P-Value is *the area <u>above</u>* the observed value of the test statistic

Finding p-values – Two tailed tests: find double the minimum of the cumulative probability associated with the calculated value of the test statistic and 1 - the cumulative probability associated with the calculated value of the test statistic:



P-Value is *two times the area* <u>*more extreme*</u> than the observed value of the test statistic

Finding p-values – Two tailed tests: find double the minimum of the cumulative probability associated with the calculated value of the test statistic and 1 - the cumulative probability associated with the calculated value of the test statistic:



P-Value is *two times the area* <u>*more extreme*</u> than the observed value of the test statistic

How do you use p-values?

If the p value is less than the significance level α, we reject the null hypothesis!

Why are p-values preferred?

a. **they allow anyone to select their own significance level α**

b. **they provide a measure of the strength of the evidence the sample data provides against the null hypothesis (smaller p-value – stronger evidence against H_0)**

c. **they are usually reported by computer packages**

How do you find p-values with Excel?

If you are using the standard normal (z) distribution, use the paste function

=NORMSDIST(z)

to find the cumulative probability under the standard normal curve for the selected value of z (the same value provided in the ASW standard normal table)

For example,

=NORMSDIST(1.96)

yields 0.975002105, and

=NORMSDIST(-1.96)

yields 0.024997895.

If you are using the Student's t distribution, use the paste function

=TDIST(t, df, number of tails)

this will provide the probability of being in the tail(s) for a given value of t with a given number of degrees of freedom.

For example,

=TDIST(1.372,10,1)

yields 0.100027671, and

=NORMSDIST(1.372,10,2)

yields 0.200055342.

Note – TDIST() only works for positive values of t!

An Example - The superintendent of a printing plant has selected a random sample of 100 rolls of paper from a large shipment. The average length of the sample rolls is 516 feet, with a variance of 2704 feet. The superintendent wants you to test that the mean length of the rolls is 525 feet at a significance level of 0.05.

1. State the Null and Alternative Hypotheses

H_0: $\mu = 525$

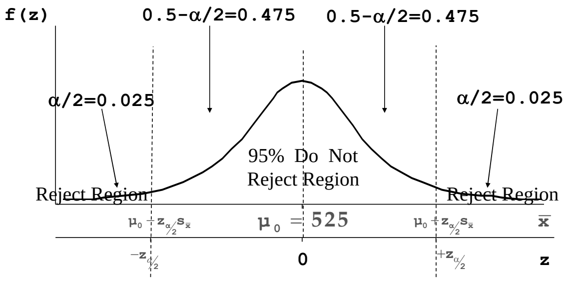
H_a: $\mu \neq 525$

2. Select the Appropriate Test Statistic – n = 100 ≥ 30, so use

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

3. **State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule**

α=0.05 and we have a two-tailed test, so



Decision rule – do not reject H_0 if the p-value ≥ α (= 0.05) otherwise reject H_0

4. **Calculate the Test Statistic**

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{516 - 525}{52/\sqrt{100}} = -1.73077$$

Since this is a two tailed z test with sample results in the lower tail, by Excel the p-value is

=2*NORMSDIST(-1.73077)

which yields 0.0834927945899173.

The standard normal tables yield approximately

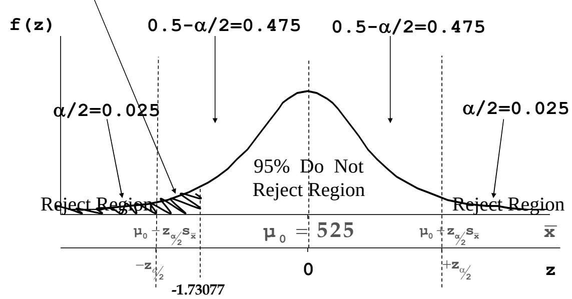
between
2*0.0409 = 0.0818
and
2*0.0418 = 0.0836

This slight imprecision occurs because we can use a more precise value of z with Excel than we can with the tables.

The Cumulative Standard Normal Distribution (Appendix B, Table 1)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Because this is a two-tailed test, the p-value is twice this area



We can see that these sample results are in the *do not reject H_0* region.

5. **Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis**

p-value ≥ α, i.e., 0.0835 ≥ 0.05

so do not reject H_0. The sample evidence does not support the claim that the mean length of the rolls differs from 525 feet (note we would reach a similar conclusion using the Standard Normal Table to approximate the p-value for this test).

Another Example - The superintendent of a printing plant has selected a random sample of 100 rolls of paper from a large shipment. The average length of the sample rolls is 516 feet, with a variance of 2704 feet. The supplier has guaranteed that the mean length of the rolls is at least 525 feet. The superintendent needs to determine if the mean length of the rolls is at least 525 feet at a significance level of 0.05.

1. **State the Null and Alternative Hypotheses**

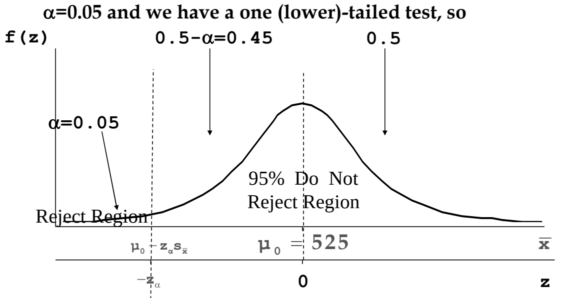
$$H_0: \mu \geq 525$$
$$H_a: \mu < 525$$

2. **Select the Appropriate Test Statistic – n = 100 ≥ 30, so use**

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

3. **State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule**

α=0.05 and we have a one (lower)-tailed test, so



Decision rule – do not reject H_0 if the p-value ≥ α (= 0.05) otherwise reject H_0

4. **Calculate the Test Statistic**

$$z = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{516 - 525}{52/\sqrt{100}} = -1.73077$$

Since this is a lower-tailed z test, by Excel the p-value is

=NORMSDIST(-1.73077)

which yields 0.0417463972949587.

The standard normal tables yield approximately

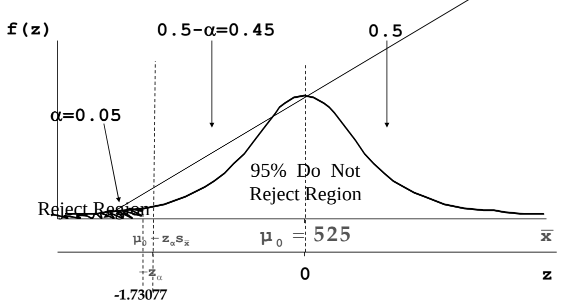
between 0.0409 and 0.0418

This slight imprecision occurs because we can use a more precise value of z with Excel than we can with the tables.

The Cumulative Standard Normal Distribution (Appendix B, Table 1)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Because this is a lower-tailed test, the p-value is this area



We can see that these sample results are in the *critical/ reject H_0* region.

5. Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis

p-value ≥ α, i.e., 0.0417 < 0.05

so reject H_0. The sample evidence does not support the claim that the mean length of the rolls is at least 525 feet (again note the conclusion would be the same if we used the approximation from the Standard Normal Table).

Yet Another Example – A furniture store has taken a random sample of twenty-five customers. The mean age of the respondents is 38 and the standard deviation is 3.5. Previous samples have shown the age of the store's customers to be approximately normal. Test the claim that the mean age of all consumers of a furniture store is under 40 years, at the α = 0.10 level of significance.

1. **State the Null and Alternative Hypotheses**

$$H_0: \mu \geq 40$$
$$H_a: \mu < 40$$

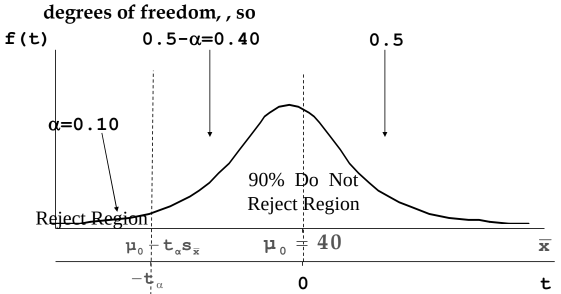
2. **Select the Appropriate Test Statistic – n = 25 < 30 and x (age of customers) is normally distributed, so use**

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$df = n - 1 = 25 - 1 = 24$$

3. **State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule**

α=0.10 and we have a one (lower)-tailed test with 24 degrees of freedom, , so



Decision rule – do not reject H_0 if the p-value ≥ α (= 0.10) otherwise reject H_0

4. Calculate the Test Statistic

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{38-40}{3.5 / \sqrt{25}} = -2.857$$

Since this is a lower-tailed t test with 24 degrees of freedom, by Excel the p-value is

=TDIST(2.857, 24, 1)

which yields 0.00434696931822273.

The t Distribution table (at 24 degrees of freedom) yield approximately

between 0.005 and 0.001 (closer to 0.005)

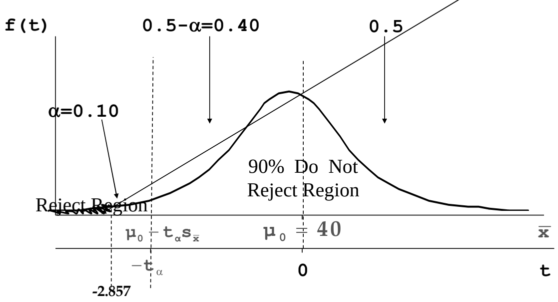
This slight imprecision occurs because we can use a more precise value of t with Excel than we can with the tables and the tables have a very limited number of columns (areas in the tail).

The t Distribution (Appendix B, Table 2)

degrees of freedom	Area in Upper Tail					
	0.10	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
∞	1.282	1.645	1.960	2.326	2.576	3.090

24 degrees of freedom

Because this is a lower-tailed test, the p-value is this area



We can see that these sample results are in the *critical/ reject H_0* region.

---

5. **Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis**

p-value < α, i.e., 0.0043 < 0.10

so reject H_0. The sample evidence does not refute the claim that the mean age of the furniture store's customers is under 40.

A Final Example – The Gibby Glass company is concerned that public awareness of their brand of kitchenware is low. They feel that, in order to be successful, over 35% of consumers should be aware of their brand. They conduct a survey of 144 randomly selected consumers and find that 54 of the respondents are aware of their brand. Test the hypothesis that brand awareness for Gibby Glass kitchenware is sufficient at the α=.01 level of significance.

1. **State the Null and Alternative Hypotheses**

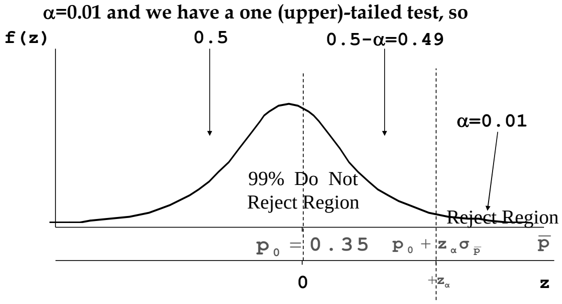
$$H_0: p \le 0.35$$
$$H_a: p > 0.35$$

2. **Select the Appropriate Test Statistic – $np_0 \ge 5$, $n(1-p_0) \ge 5$, so use**

$$z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}} = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

3. State the Desired Level of Significance α, Find the Critical Value(s) and State the Decision Rule

α=0.01 and we have a one (upper)-tailed test, so



Decision rule – do not reject H_0 if the p-value ≥ α (= 0.01) otherwise reject H_0

4. Calculate the Test Statistic

$$z = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}} = \frac{54/144 - 0.35}{\sqrt{\frac{0.35(1-0.35)}{144}}} = 0.629$$

Since this is a lower-tailed z test, by Excel the p-value is

=1-NORMSDIST(0.629)

which yields 0.264674528106677.

The standard normal tables yield approximately

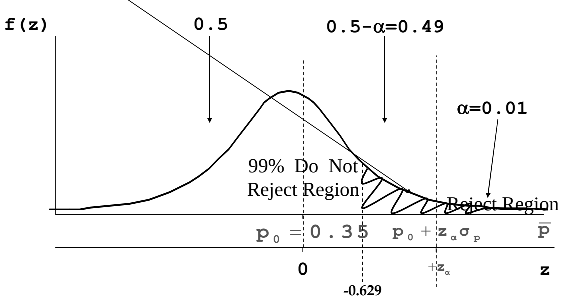
between
1.0000 – 0.7324 = 0.2676
and
1.0000 - 0.7357 = 0.2643

This slight imprecision occurs because we can use a more precise value of z with Excel than we can with the tables.

The Cumulative Standard Normal Distribution (Appendix B, Table 1)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	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	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	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	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Because this is an upper-tailed test, the p-value is this area



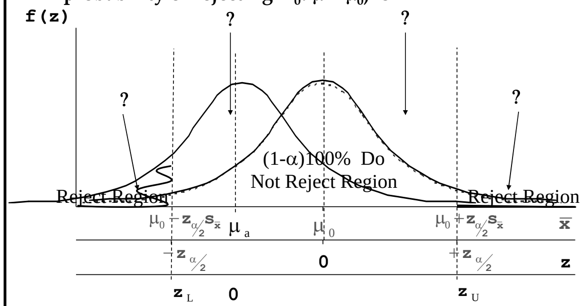
We can see that these sample results are in the *do not reject H_0* region.

5. Use the Decision Rule to Evaluate the Test Statistic and Decide Whether to Reject or Not Reject the Null Hypothesis

p-value ≥ α, i.e., 0.2647 ≥ 0.01

so do not reject H_0. The sample evidence does not support the claim that proportion of customers aware of the Gibby Glass brand of kitchenware exceeds 0.35.

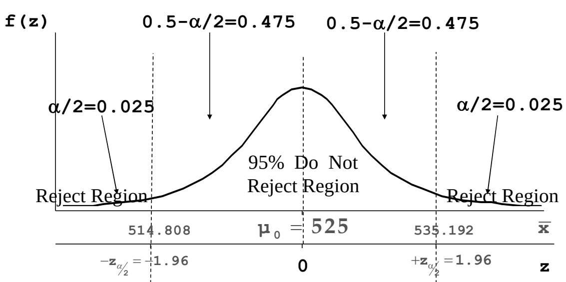
G. The Power of a Hypothesis Test – Under the condition that *the null hypothesis is false because the true value of the mean is* $\mu_a \neq \mu_0$, the power of the test (i.e., the probability of rejecting H_0: $\mu = \mu_0$) is



An Example - The superintendent of a printing plant has selected a random sample of 100 rolls of paper from a large shipment. Historically, the variance of the sample rolls has been 2704. The superintendent wants you to test that the mean length of the rolls is 525 feet at a significance level of 0.05. What is the power of this hypothesis test if the true population mean is $\mu = \mu_a$ = 516?

Let's review our critical values and decision rules from when we previously worked on this problem:

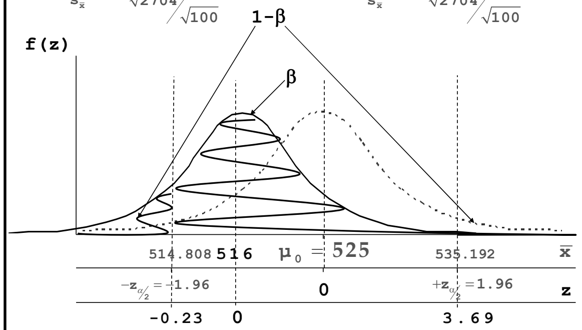
α=0.05 and we have a two-tailed test, so $z_{\alpha/2} = \pm 1.96$



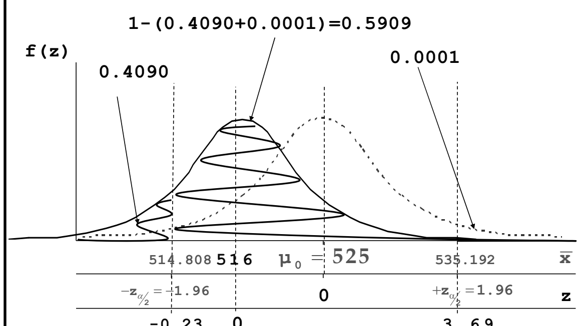
Decision rule – do not reject H_0 if –514.808 ≤ $\bar{x}$ ≤ 535.192
otherwise reject H_0

We have

$$z_L = \frac{\bar{x}_L - \mu_a}{s_{\bar{x}}} = \frac{514.808-516}{\sqrt{2704}/\sqrt{100}} = -2.29 \qquad z_U = \frac{\bar{x}_L - \mu_a}{s_{\bar{x}}} = \frac{535.192-516}{\sqrt{2704}/\sqrt{100}} = 8.41$$



Using the standard normal distribution table, we find



So the probability of rejecting H0: m = 525 if the true mean is $\mu = \mu_a = 516$ is 0.5909!

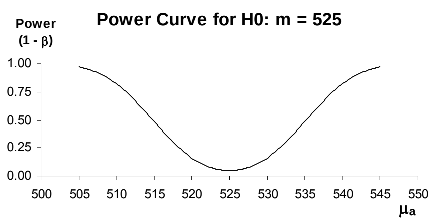
Note:

- **this approach can be applied to one-tailed (upper or lower) tailed tests**
- **α and β have an inverse (but nonlinear) relationship**
- **The power $(1 - \alpha)$ of a test is often plotted against corresponding values of μ_a - this is called a *Power Curve*.**

An Example - The superintendent of a printing plant has selected a random sample of 100 rolls of paper from a large shipment. Historically, the variance of the sample rolls has been 2704. The superintendent wants you to test that the mean length of the rolls is 525 feet at a significance level of 0.05. Create a power curve for the appropriate null hypothesis (H_0: $\mu = 525$).

We repeatedly select values for μ_a and calculate the probability of type II error β as before and use these results to calculate the associated power ($1 - \beta$)

μ_a	$1 - \beta$
505.0	0.970365
507.5	0.920052
510.0	0.822427
512.5	0.671443
515.0	0.485339
517.5	0.302683
520.0	0.160778
522.5	0.076873
525.0	0.050000
527.5	0.076873
530.0	0.160778
532.5	0.302683
535.0	0.485339
537.5	0.671443
540.0	0.822427
542.5	0.920052
545.0	0.970365



This curve enables us to asses how sensitive our hypothesis test is to an incorrect null hypothesis (note that flatter curves suggest a less-sensitive test – why?)

H. Deciding on the Null Hypothesis – Sometimes the null hypothesis is not explicitly specified or implied by circumstances. In such situations we may have latitude in selecting the null hypothesis to test.

Consider the following two circumstances:

i. **The FDA is considering whether to approve a new drug – if the drug is successful it will cure liver cancer, and its only side effect is that it may cause excess nasal discharge (i.e., a *runny nose*)**

ii. **The FDA is considering whether to approve a new drug – if the drug is successful it will relieve excess nasal discharge, and its only side effect is that it may cause liver cancer**

In the first case (if the drug is successful it will cure liver cancer, and its only side effect is possible excess nasal discharge), the two errors we can make when testing the drug's effectiveness are:

- **to find the drug is effective when in fact it is ineffective**
- **To find the drug is ineffective when in fact it is effective**

About which of these potential errors should we be more concerned? Why?

If we find the drug to be effective when in fact it is ineffective, we

- **do not cure liver cancer (no change to the patient's prognosis), and**
- **cause liver cancer patients to possibly suffer from excess nasal discharge**

On the other hand, if we find the drug to be ineffective when in fact it is effective, we

- **miss the opportunity to cure liver cancer, and**
- **avoid potentially causing liver cancer patients to suffer from excess nasal discharge**

Now can you see which of these potential errors is potentially more serious?

Since we can control the probability of Type I error but not the probability of Type II error, we should formulate our null hypothesis so that the more serious error is Type I error.

In this situation we use:

H_0: the drug is effective
H_a: the drug is ineffective

Now the more serious error (we find the drug to be ineffective when in fact it is effective) is the Type I error (rejecting the null hypothesis when it is actually true).

We can control for this possible outcome by setting the significance level α relatively low!

In the second case (if the drug is successful it will relieve excess nasal discharge, and its only side effect is possible liver cancer), the two errors we can make when testing the drug's effectiveness are again:

- **to find the drug is effective when in fact it is ineffective**
- **To find the drug is ineffective when in fact it is effective**

About which of these potential errors should we now be more concerned? Why?

If we find the drug to be effective when in fact it is ineffective, we

- **do not relieve excess nasal discharge (no change to the patient's prognosis), and**
- **possibly cause the patients with excess nasal discharge to develop liver cancer**

On the other hand, if we find the drug to be ineffective when in fact it is effective, we

- **miss the opportunity to relieve excess nasal discharge, and**
- **avoid potentially causing patients with excess nasal discharge to develop liver cancer**

Now can you see which of these potential errors is potentially more serious?

Since we can control the probability of Type I error but not the probability of Type II error, we should formulate our null hypothesis so that the more serious error is Type I error.

In this situation we use:

H_0: the drug is ineffective
H_a: the drug is effective

Now the more serious error (we find the drug to be effective when in fact it is ineffective) is the Type I error (rejecting the null hypothesis when it is actually true).

We can control for this possible outcome by setting the significance level α relatively low!