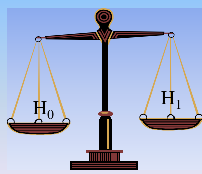
STATS 101
Introductory Statistics

TESTING STATISTICAL HYPOTHESES



EXAMPLE 1: SOUP TO GO sells chicken soup in to-go cups. Their flyer claims that each cup contains 25 ounces of soup. One customer thinks that the to-go cups only have 15 ounces of chicken soup.. The problem is to 'TEST' which of the two claims is correct. To test the claims, the weights of an independent random sample of 4 to-go cups are measured; the weights of the sample (in ounce) are given below.

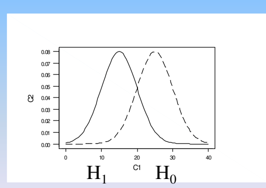
27.4480, 21.5236, 28.0366, 26.3915.

Based on this sample of size n=4, we test

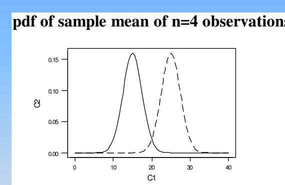
NULL HYPOTHESIS $H_0: \mu = 25$
ALTERNATIVE $H_1: \mu = 15$

Assume that weight of to-go cups has a normal distribution with population sd σ = 5.

The problem is to decide if one observation is coming from the probability distribution on the right (dashed lines) or the one on the left (solid line). Since there is quite a bit of overlap in the two distributions, it would not be an easy decision.

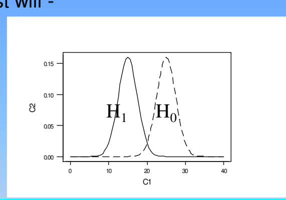


The overlap between the distributions is reduced if we make this decision on the basis of the sample mean of 4 observations which has a normal distribution with mean μ and sd = σ/√n = 5/√4 = 5/2 = 2.5



The mean and sd of the sample of size 4 are:
$\bar{x} = 25.85, \ s = 2.96$

Since, in this example, H_1 falls to the left of H_0, small values of the sample mean indicate that H_0 is false: in other words, a reasonable test will - $\bar{x}$



Reject H_0 if $\bar{x}$ is 'too small', or
$\bar{x} - \mu_0$ is too small, i.e., $\bar{x} - \mu_0 \le C$
(here μ_0 is the hypothesized mean under H_0)

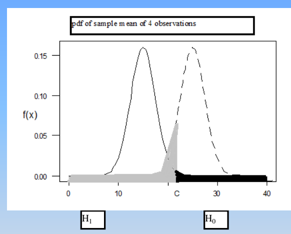
We will now have to decide 'How small is TOO SMALL'?, i.e, we will have to select a value for the CUT-OFF POINT C.

STATISTICIAN'S ACTION	TRUE STATE OF NATURE H0 is true	H0 is false
REJECT H0	Type I Error	No error
DO NOT REJECT H0	No error	Type II Error

P(Type I Error) = P(Reject H_0|H_0 is TRUE) = α

P(Type II Error) = P(Accept H_0|H_0 is FALSE) = β

The statistician would like to minimize both error probabilities.



α

β

Observe that α can be made small by decreasing C, but then β goes up.

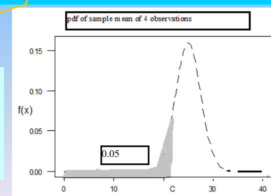
The following approach is used in testing hypotheses: fix α = say 0.05 and find the value of the cut-off point C, then calculate β.

Reject H_0 if

$\bar{x}_{obs} \le C$, or in terms of the standardized sample mean -

$$z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma / \sqrt{n}} \le -1.64$$

Distance between observed and hypothesized means, standardized by $sd(\bar{x}) = \sigma / \sqrt{n}$



Computation of P(Type II Error) :

First compute C:

$$\frac{C - 25}{2.5} = -1.64, \text{ or}$$

C = 25 - 2.5 x 1.64 = 25 - 4.1 = 20.9

and then

$$\beta = P(\bar{x} > 20.9 | H_1 \text{ is true})$$
$$= P(Z > \frac{20.9 - 15}{2.5}) = P(Z > 2.36) = 1 - 0.9909 = 0.0091$$

POWER OF THE TEST = 1 - P(Type II Error) = 0.9909

For data of example 1:

$$z_{obs} = \frac{\bar{x}_{obs} - 25}{2.5} = \frac{.85}{2.5} = .34 > -1.64$$

Hence the null hypothesis μ = 25 is NOT REJECTED.

Suppose now that the alternative hypothesis is changed to: $H_1 : \mu = 20$

How does this change the test of size .05, obtained earlier?

$\bar{x}_{obs} \le C$, or in terms of the standardized sample mean -

$$z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma / \sqrt{n}} = \frac{\bar{x}_{obs} - 25}{2.5} \le -1.64$$

OBSERVE that the cut-off point is determined from H_0 (which did not change), and therefore, the test of size 0.05 did not change.

The only change is in P(Type II Error) :

$$\frac{C - 25}{2.5} = -1.64, \text{ or}$$

C = 25 - 2.5 x 1.64 = 25 - 4.1 = 20.9

$$\beta = P(\bar{x} > 20.9 | H_1 \text{ is true})$$
$$= P(Z > \frac{20.9 - 20}{2.5}) = P(Z > 0.36) = 1 - 0.6406 = 0.3594$$

POWER OF THE TEST = 1 - P(Type II Error) = 0.6406

NOTE: In Example 1, population sd σ = 5 (given). In most real situations, σ is unknown, and is estimated by the sample sd s = 2.96. The test statistic changes from

$$z = \frac{(\bar{x}-\mu_0)}{\sigma/\sqrt{n}} \text{ to } t = \frac{(\bar{x}-\mu_0)}{\sigma/\sqrt{n}}$$

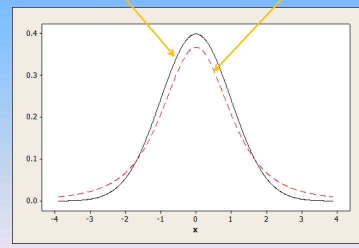
The null distribution of the test statistic changes from the standard normal (distribution of z) to the t-distribution with degrees of freedom df = n-1 (distribution of t), as shown on the next slide.

σ known, use null distribution of

$$z = \frac{(\bar{x}-\mu_0)}{\sigma/\sqrt{n}}$$

σ unknown, use null distribution of

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

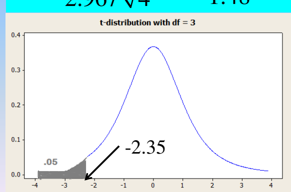


We will redo example 1 on slide 1, assuming that σ is not known.

The mean and sd of the sample of size 4 are:

$$\bar{x} = 25.85, \ s = 2.96$$

$$t_{obs} = \frac{\bar{x}-25}{2.96/\sqrt{4}} = \frac{25.85-25}{1.48} = 0.57$$



Since t_{obs} > -2.35 (value from t-table with df=3), we do not reject the null hypothesis (same result as on slide 9).

Remarks about setting up the null and alternative hypotheses

In testing hypotheses formulation:

•There will always be two hypotheses – a null hypothesis H_0 and an alternative hypothesis H_1

•H_0 and H_1 have no points in common, as H_1 is the complement of H_0

•The null hypothesis always will contain the equal sign.

•The problem description will always specify either the null or the alternative hypothesis.

Example 2: Motel 7 is trying to determine if the average length of stay of their customers is larger than 4 days or not. A random sample of n = 25 observations produced x_bar = 4.8, s = 2 (in days). Test the hypothesis that the mean is larger than 4,

Setting up the null and alternative hypotheses	$H_0: \mu \le 4, \ H_1: \mu > 4$

Reject H_0 if the *observed* t-statistic is 'large':

$$t_{obs} = \frac{\bar{x}_{obs}-\mu_0}{s/\sqrt{n}} > t_{n-1,1-\alpha}$$

For this example, n = 25, so df = n - 1 = 24.

For $\alpha = 0.05$, $t_{n-1,1-\alpha} = t_{24,0.95} = 1.7109$

For the given data: $t_{obs} = \frac{4.8-4}{2/\sqrt{25}} = 2 > 1.7109$

Therefore, we reject H_0.

The Significance Probability (P-Value)

A BETTER WAY OF TESTING HYPOTHESES IS BY COMPUTING THE P-VALUE.

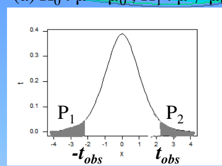
The Significance Probability (P-Value) is the probability under H_0 that the test statistic will be *more extreme* than its observed value (calculated from sample).

You can think of the P-value as EVIDENCE IN FAVOR OF H_0 ; if P-value is small, H_0 is rejected.

The null hypothesis H_0 is rejected if and only if P < α.

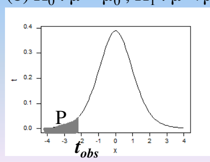
What values are more extreme than the observed value depends on the alternative hypothesis H_1.

(a) $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

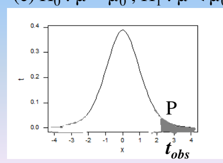


P = sum of two shaded areas = $P_1 + P_2$

(b) $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$



(c) $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$



We illustrate P-value calculation for data of Example 2.

Example 2: Motel 7 is trying to determine if the average length of stay of their customers is larger than 4 days or not. A random sample of n = 25 observations produced x_bar = 4.8, s = 2 (in days). Test the hypothesis that the mean is larger than 4, assuming the population to be approximately normal.

$H_0 : \mu \leq 4, \ H_1 : \mu > 4$

$$t_{obs} = \frac{4.8 - 4}{2/\sqrt{25}} = 2$$

$$P = P(t_{df=24} > 2)$$

In excel, type =tdist(2,24,1)*; this will give P = 0.02847 < .05, and as before(see slide 16), the null hypothesis is rejected.

*If tails = 1, TDIST is calculated as TDIST = P(X>x), where X is a random variable that follows the t-distribution. If tails = 2, TDIST is calculated as TDIST = P(|X| > x) = P(X > x or X < -x).

NOTE: EXCEL does not have a built-in 1-sample t-test function, 1-sample t-test must be run manually in EXEC, as follows:

1) Calculate $t_{obs} = \dfrac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$

2)Use EXCEL to calculate the P-value as follows:

= tdist(t_{obs}, df, 1) if 1-sided alternative

= tdist(t_{obs}, df, 2) if 2-sided alternative

3) Reject H_0 if $P < .05$ (if using α=.05)

TESTING STATISTICAL HYPOTHESES - TWO SAMPLE PROBLEM

Just as in the Confidence Interval Estimation approach, We have to consider two different ways of sampling.

Case 1: Given two independent random samples with descriptive statistics

sample size = n_1, sample mean = $\bar{x}_1$, sd = s_1

sample size = n_2, sample mean = $\bar{x}_2$, sd = s_2

Problem: To test

$$H_0 : \mu_1 = \mu_2 \quad vs.$$
$$(a) H_1 : \mu_1 \neq \mu_2 \quad or$$
$$(b) H_1 : \mu_1 < \mu_2 \quad or$$
$$(c) H_1 : \mu_1 > \mu_2$$

Assumptions: The two samples are independent of each other, and each sample comes from a normal population:

Sample 1 has mean μ_1 and sd σ_1 , Sample 2 has mean μ_2 and sd σ_2 .

The t-test formula depends on whether we have:

Case 1 (a) : The two population variances are equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Case 1 (b) : The two population variances are NOT equal.

The estimate of the difference on the two means μ_1 -μ_2 is $\bar{x}_1 - \bar{x}_2$ regardless of whether the variances are equal or not, but the estimate of $sd(\bar{x}_1 - \bar{x}_2)$ depends on whether we have Case 1 (a) or Case 1(b).

$$s^2_{pooled} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Case 1(a): Independent Samples, Equal Variances

The t-statistic is

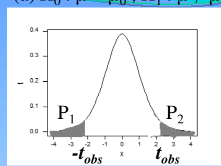
$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{s_{pooled}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}, \quad df = n_1 + n_2 - 2$$

The P-value, as mentioned earlier, depends on the alternative hypothesis, and is calculated as shown on the next slide. The EXCEL function tdist can be used to get the P-value. The null hypothesis is rejected if

P < α

The P-value comes from the t-distribution with df = $n_1 + n_2 - 2$

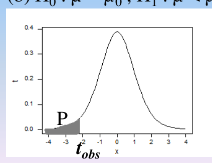
(a) $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$



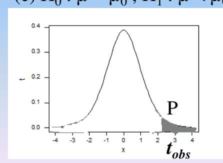
P = sum of two shaded areas = $P_1 + P_2$

t_{df} , df = $n_1 + n_2 - 2$

(b) $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$



(c) $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$



The 2-sample t-test can be run in EXCEL in two ways :

(a)By doing all calculations in EXCEL manually

(b)By using the EXCEL function ttest

We will demonstrate both methods; Examples 3 and 4 are solved manually in EXCEL, and Examples 5 and 6 are done using the EXCEL function ttest.

Example 3: The descriptives of scores on a test for two groups of students (Attenders=1, Non-attenders=2) are given below:

$$n_1 = 15, \bar{x}_1 = 4.75, s_1 = 1.0$$
$$n_2 = 22, \bar{x}_2 = 3.00, s_2 = 1.5$$

Test the hypothesis that, on the average, Attenders score higher than Non-attenders.

$$H_0 : \mu_1 - \mu_2 \leq 0, \quad H_1 : \mu_1 - \mu_2 > 0$$

Many books first test if the two variances are equal or not before testing the hypothesis about the means. We recommend running the test both ways, i.e., assuming variances to be equal, and then assuming variances to be unequal.

In most cases, conclusion would be the same; if not, go with the conclusion without assuming equal variances.

Example 3, Case 1a: Assume variances are equal.

$$s^2_{pooled} = \frac{14\times1^2 + 21\times1.5^2}{35} = 1.75$$

$$s_{pooled} = \sqrt{1.75} = 1.32$$

$$t_{obs} = \frac{4.75-3-0}{1.32\sqrt{\frac{1}{15}+\frac{1}{22}}} = \frac{1.75}{0.44} = 3.96$$

$$P = P(t_{35} > 3.96) = 0.000175$$

Hence we reject the null hypothesis and conclude that the attenders have a higher score on the average than the non-attenders.

Example 3, Case 1b: variances are not equal, Satterthwaite's t

Estimate of $\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 = 4.75 - 3 = 1.75$

$sd(estimate) = sd(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ is

estimated by $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$$\frac{s_1^2}{n_1} = \frac{1^2}{15} = .0667$$

$$\frac{s_2^2}{n_2} = \frac{1.5^2}{22} = .1022$$

$$sd(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{.0667 + .1022} = \sqrt{.1689} = .411$$

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = \frac{(.411)^2}{\frac{.0667^2}{14} + \frac{.1022^2}{21}} = \frac{.1689}{.0003 + .0005} = \frac{.1689}{.0008} = 211.12$$

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1.75}{.411} = 4.26$$

$$P = P(t_{211} > 4.26) = 0.000$$

Hence we reject the null hypothesis and conclude that the attenders have a higher score on the average than the non-attenders (same conclusion as in case 1a).

Example 4: The following table shows summary statistics for time (in seconds) taken by subjects to walk a short distance.

	n	mean	sd
Attenders	50	340	250
Nonattenders	40	45	25

Can we conclude on the basis of this data that the mean time taken by alcoholics is higher than that for nonalcoholics? Use $\alpha = 0.01$.

$$H_0 : \mu_1 - \mu_2 \leq 0$$
$$H\ : \mu_1 - \mu_2 > 0$$

We conclude on the basis of this data that the mean time taken by alcoholics is higher than that for nonalcoholics.

We will run the t-test both ways:

(a) Assuming equal variances,

(b) Assuming unequal variances.

(a) Example 4 data: assume equal variances

$$s^2_{pooled} = \frac{49\times 250^2 + 39\times 25^2}{88} = 35078.13$$

$$s_{pooled} = \sqrt{35078.13} = 187.29$$

$$t_{obs} = \frac{340-45}{187.29\sqrt{\frac{1}{50}+\frac{1}{40}}} = \frac{295}{30.73} = 7.43$$

$$df = 50+40-2 = 88$$

$$P = P(t_{88} > 7.43) = 3.40E-11$$

(b) Example 4 data: unequal variaces

Estimate of $\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 = 340 - 45 = 295$

$sd(estimate) = sd(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ is

estimated by $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 35.58$

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = 50.22$$

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{295}{35.58} = 8.29$$

$$P = P(t_{50} > 9.29) = 0.000$$

Reject null hypothesis since P < .01, just as in case (a).
In other words, the conclusion is same regardless of whether we have equal variances or not.

t-test for paired samples

Given: n pairs of observations $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

The observations (x,y) are either on the same subject (for example, a BEFORE and AFTER experiment) or taken at the same time, making x and y DEPENDENT rvs.

Assume: (x,y) ~ BIVARIATE NORMAL DISTRIBUTION in which case:

$$d = x - y \sim N(\mu_1 - \mu_2, \sigma_d^2)$$
$$H_0: \mu_1 - \mu_2 = 0$$

is tested by using the 1-sample t-test on d = x - y

Example 5:

2FastFoodSales.xlsx

The daily sales of two fast food chains at 14 intersections of a large city are given in the table on the right. Test if McB outsells DK.

McB	DK
1004.27	902.8
947.05	739.65
993.83	975.63
345.4	382.18
796.76	629.02
823.44	590.77
755.21	574.36
935.82	668.39
1114.31	773.89
1021.59	1121.81
996.96	1033.56
969.58	1127.74
1007.78	974.63
1090.06	1001.27

Problem: To test
$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 > 0$

$t_{obs} = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{93.3 - 0}{147/\sqrt{14}} = 2.38$, df = n-1 = 13

$P = tdist(2.38, .05, 1)$ [in excel, 1 is for 1-sided probability]

$P = .0167 < .05$ so we reject the null hypothesis and conclude that mean daily sale of McB is higher than that of DK.

Testing hypothesis for a single population proportion

Example 5: In a paired preference test of two brands of soda, A and B, , 670 subjects are given a cup each of soda A and soda B, and 423 subjects said that they preferred soda B over soda A. Can we conclude that more than 60% of the consumers would prefer soda B? Use test size of .05.

The problem inexample 5 can be formulated as testing
$H_0: p \le 0.60$
$H_1: p > 0.60$
where p = proportion of consumers who prefer soda B

X = #(successes) in n independent trials
X has the binomial distribution with n trials and success probability p , X ~ Bin(n,p)

In Example 5, n = 670, X = # of subjects who prefer soda B,
X has binomial distribution with n = 670, and
p = proportion in population who prefer soda B over soda A.

Observed value of X = 423.

Example 5 - contd

An estimate of p is

$\hat{p} = \frac{x}{n}$ which is approximately normal with

mean = p, sd = $\sqrt{\frac{p(1-p)}{n}}$

The hypothesis can be tested by the z-statistic (calculated assuming null is true, i.e., $p = p_0$

$$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

For data of Example 5:

$\hat{p} = 443/670 = 0.66, n = 670$

$$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.66-0.60}{\sqrt{\frac{0.6(1-0.6)}{670}}} = \frac{0.06}{0.0189} = 3.17$$

$P = P(Z > 3.17) = 1 - .9992 = .0008$

Since P < .05, null is rejected.

Testing hypothesis for testing equality of two population proportions

Example 6:

The "winner's curse" in auction bidding is the phenomenon of the highest bid price being above the expected value of the item being auctioned. . Let X represent the number of winning bids that exceed the expected price of the item. The following data was collected for two groups of bidders.

Compare the proportions of bidders in the two groups who fall prey to the winner's curse

Group	n	x
Super-experienced bidders	150	21
Less-experienced bidders	200	48

Testing equality of two population proportions

X_1 = #(successes) in n_1 independent trials ~ Bin(n_1,p_1)
X_2 = #(successes) in n_2 independent trials ~ Bin(n_2,p_2)

To test: $H_0: p_1 = p_2$ *or* $H_0: p_1 - p_2 = 0$

Test statistics is: (estimate – test value)/sd(estimate assuming H_0 is true)

$$z_{obs} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\frac{\overline{p}(1-\overline{p})}{n_1} + \frac{\overline{p}(1-\overline{p})}{n_2}}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\overline{p}(1-\overline{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$\overline{p}$ = estimate of common p if H_0 is true = $\frac{x_1 + x_2}{n_1 + n_2}$

Null distribution of test statistic Z is standard normal with 0 mean and sd = 1.

Example 6: Assuming H_0 to be true, estimate of common proportion is

$$\overline{p} = \frac{21+48}{150+200} = 0.1971$$

$$sd(\hat{p}_1 - \hat{p}_2) = \sqrt{\overline{p}(1-\overline{p})(\frac{1}{n_1} + \frac{1}{n_2})} = \sqrt{.1971 \times .8029 \times (\frac{1}{150} + \frac{1}{200})} = 0.043$$

$\hat{p}_1 = 21/150 = 0.14, \hat{p}_2 = 48/200 = 0.24$

$$z_{obs} = \frac{.14 - .24 - 0}{.043} = -2.33$$

$P = 2 \times P(Z \leq -2.33)$ since alternative is 2-sided

= 2 × .0099 [in EXCEL, type =normsdist(-2.33) in an empty cell]

= .0198 < .05, so reject H_0, conclude $p_1 \neq p_2$.

A closer look at each testing hypothesis problem (1-sample or 2-sample) will show you that the test statistic always has the form:

$$u_{obs} = \frac{(\hat{\theta} - \theta)}{sd(\hat{\theta})}$$

θ = parameter of interest

$\hat{\theta}$ = estimate of θ

$sd(\hat{\theta})$ = sd of $\hat{\theta}$ assuming null is H_0

The P-value is calculated from the null distribution of $\hat{\theta}$.