COMP 7036
Applied Research Methods in Software Development

Borna Noureddin, Ph.D.
British Columbia Institute of Technology

More significance testing

Moving from the Z to T

When you know the standard deviation of the population use a Z-Test.

The test statistic is z.

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Use the normal distribution to calculate your p-value.

When you don't know the standard deviation of the population use a T-Test.

The test statistic is t.

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

Use a t-distribution with (n-1) degrees of freedom to calculate your p-value.

Checking Assumptions

CAUTION!

The t-test is NOT always appropriate!

In order to appropriately use the t-test:

The population distribution must be normal

OR

The sample size must be large enough

Checking Assumptions

How large a sample is large enough?

Sample Size	Check
n<15	In order to use t-procedures, the sample distribution should be pretty normal. There should not be any outliers and very little skewness.
15<n<40	In order to use t-procedures, there should not be any large outliers and there should not be extreme skewness.
n>40	The sample size is large enough to use t-procedures regardless.

Example of a T-test

A manufacturer of a special bolt requires that this type of bolt have a mean shearing strength in excess of 110 lb. To determine if the manufacturer's bolts meet the required standards a sample of 25 bolts was obtained and tested. The sample mean was 112.7 lb and the sample standard deviation was 9.62 lb. Use this information to perform an appropriate hypothesis test with a significance level of 0.05.

Example of a T-test

μ = the mean shearing strength of this specific type of bolt

The hypotheses to be tested are

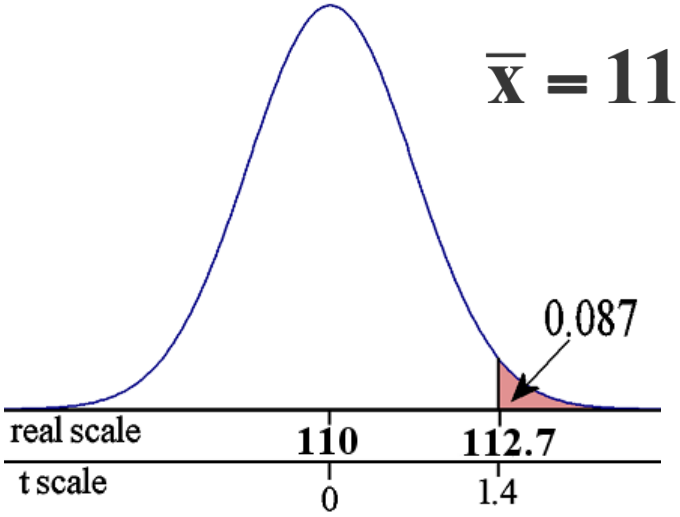
$$H_0: \mu = 110 \text{ lb}$$

$$H_a: \mu > 110 \text{ lb}$$

The significance level to be used for the test is $\alpha = 0.05$.

The test statistic is $$t = \frac{\bar{x} - 110}{s/\sqrt{n}}$$

Example of a T-test



$$\bar{x} = 112.7,\ s = 9.62,\ n = 25,\ df = 24$$

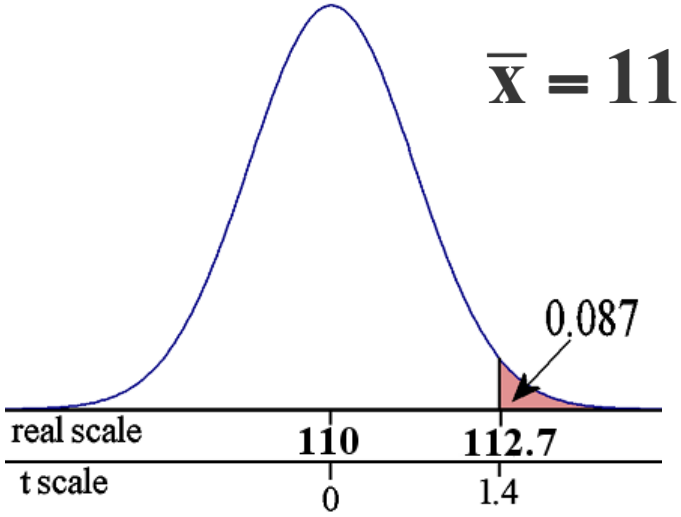
$$\text{P-value} = P\left(t > \frac{112.7 - 110}{9.62/\sqrt{25}} \right)$$

$$= P(t > 1.4) = 0.087$$

***t* Table**

cum. prob	*t* .50	*t* .75	*t* .80	*t* .85	*t* .90	*t* .95	*t* .975	*t* .99	*t* .995	*t* .999	*t* .9995
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745

Example of a T-test



$$\bar{x} = 112.7,\ s = 9.62,\ n = 25,\ df = 24$$

$$\text{P-value} = P\left(t > \frac{112.7 - 110}{9.62/\sqrt{25}} \right)$$

$$= P(t > 1.4) = 0.087$$

Because *p*-value = 0.087 > 0.05 = α, we fail to reject H_0.

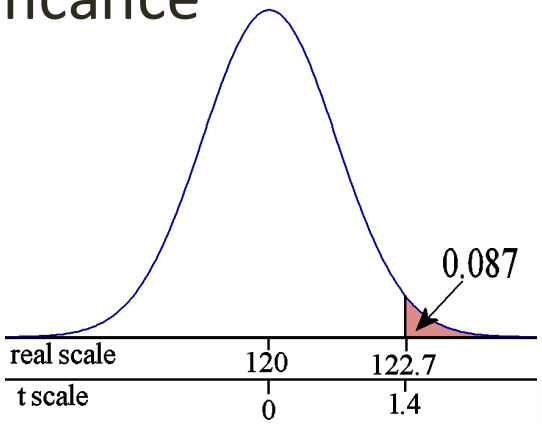
At a level of significance of 0.05, there is insufficient evidence to conclude that the mean shearing strength of this brand of bolt exceeds 110 lbs.

Example of a T-test

What if significance level was 0.10 instead of 0.05?

p-value = 0.087 < 0.10 (α)

We can reject H_0 at 0.10 level of significance



Now we can conclude:

At the 0.10 level of significance there is sufficient evidence to conclude that the mean shearing strength of this brand of bolt exceeds 120 lbs.

Different conclusions?

Many people are bothered by the fact that different choices of α lead to different conclusions.

This is natural of a process where you control the probability of being wrong.

Remember, when you select the level of significance, you choose your willingness to accept a certain chance of a Type I error.

Example #2 Gold

A jeweler is planning on manufacturing gold charms. His design calls for a particular piece to contain 0.08 ounces of gold. The jeweler would like to know if the pieces that he makes contain (on the average) 0.08 ounces of gold. To test to see if the pieces contain 0.08 ounces of gold, he made a sample of 16 of these particular pieces and obtained the following data.

0.0773	0.0779	0.0756	0.0792	0.0777
0.0713	0.0818	0.0802	0.0802	0.0785
0.0764	0.0786	0.0776	0.0793	0.0755
0.0806				

Use a level of significance of 0.01 to perform an appropriate hypothesis test.

Example #2 Gold

1. The population characteristic being studied is μ = true mean gold content for this particular type of charm.
2. Null hypothesis: $H_0: \mu = 0.08$ oz
3. Alternate hypothesis: $H_a: \mu \neq 0.08$ oz
4. Significance level: $\alpha = 0.01$
5. Test statistic:

$$t = \frac{\bar{x} - \text{hypothesized mean}}{s/\sqrt{n}} = \frac{\bar{x} - 0.08}{s/\sqrt{n}}$$

Example #2 Gold

Computations:

$$n = 16, \bar{x} = 0.077891, s = 0.0025143$$

$$t = \frac{0.077891 - 0.08}{0.0025143/\sqrt{16}} = -3.2$$

p-value: Two tailed test!

Look up table for t curves, t = 3.2, *df* = 15 we get 0.003

p-value = 2(0.003) = 0.006

***t* Table**

cum. prob	*t* .50	*t* .75	*t* .80	*t* .85	*t* .90	*t* .95	*t* .975	*t* .99	*t* .995	*t* .999	*t* .9995
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073

Example #2 Gold

Conclusion:

Since *p*-value = 0.006 ≤ 0.01, we reject H_0

At the 0.01 level of significance there is sufficient evidence that the true mean gold content of this type of charm is not 0.08 ounces.

Actually when rejecting a null hypothesis for the ≠ alternative, a one tailed claim is supported. In this case, at the 0.01 level of significance, there is convincing evidence that the true mean gold content of this type of charm is less than 0.08 ounces.

Matched Pairs t-Test Example

A weight reduction center advertises that participants in its program lose an average of at least 5 pounds during the first week of the participation. Because of numerous complaints, the state's consumer protection agency doubts this claim. To test the claim at the 0.05 level of significance, 12 participants were randomly selected. Their initial weights and their weights after 1 week in the program appear on the next slide. Set up and perform an appropriate hypothesis test.

Matched Pairs t-Test Example

Member	Initial Weight	One Week Weight
1	195	195
2	153	151
3	174	170
4	125	123
5	149	144
6	152	149
7	135	131
8	143	147
9	139	138
10	198	192
11	215	211
12	153	152

Matched Pairs t-Test Example

Member	Initial Weight	One Week Weight	Difference Initial -1week
1	195	195	0
2	153	151	2
3	174	170	4
4	125	123	2
5	149	144	5
6	152	149	3
7	135	131	4
8	143	147	-4
9	139	138	1
10	198	192	6
11	215	211	4
12	153	152	1

This last column is the data we are after!

Matched Pairs t-Test Example

μ_d = mean of the individual weight changes (initial weight–weight after one week)

This is equivalent to the difference of means:

$$\mu_d = \mu_1 - \mu_2 = \mu_{\text{initial weight}} - \mu_{\text{1 week weight}}$$

H_0: $\mu_d = 5$

H_a: $\mu_d < 5$

Significance level: $\alpha = 0.05$

Test statistic:

$$t = \frac{\bar{x}_d - \text{hypothesized value}}{s_d / \sqrt{n}} = \frac{\bar{x}_d - 5}{s_d / \sqrt{n}}$$

Matched Pairs t-Test Example

Calculations:

$$n = 12, \bar{x}_d = 2.333, s_d = 2.674$$

$$t = \frac{\bar{x}_d - 5}{s_d/\sqrt{n}} = \frac{2.333 - 5}{2.674/\sqrt{12}} = -3.45$$

p-value = 0.0027

t Table

cum. prob	t .50	t .75	t .80	t .85	t .90	t .95	t .975	t .99	t .995	t .999	t .9995
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437

Conclusions: *p*-value = 0.0027 < 0.05 so we reject H_0

There is strong evidence that the mean weight loss for those who took the program for one week is less than 5 pounds.