Session 8: Statistical Analysis of Measurements

Jennifer Hay
Factory Application Engineer
Nano-Scale Sciences Division
Agilent Technologies

jenny.hay@agilent.com

To view previous sessions:
https://agilenteseminar.webex.com/agilenteseminar/onstage/g.php?p=117&t=m

Drawing conclusions

We perform 10 indentations on each of two materials.
Material A: H = 4.91±0.23
Material B: H = 5.17±0.21

Is this difference in hardness significant or not?

Designing experiments

We have changed our film-deposition process and wish to know if that change has caused the hardness to change by more than 2%. Indentation tests on our films typically have a standard deviation of 5%.

How many indentations must we do on each film (old and new) in order to determine whether the hardness of the new film differs by more than 2% from that of the old film?

History of the Student’s t-test (From Wikipedia)

The t-statistic was introduced in 1908 by William Sealy Gosset, a chemist working for the Guinness brewery in Dublin, Ireland ("Student" was his pen name)… Gosset devised the *t*-test as a cheap way to monitor the quality of stout. He published the test in *Biometrika* in 1908, but was forced to use a pen name by his employer, who regarded the fact that they were using statistics as a trade secret. In fact, Gosset's identity was known to fellow statisticians.

Student-t test - Overview

- Student-t test is a statistical test used to determine, to a reasonable degree of confidence, whether two observation sets obtain from different populations .
- Null hypothesis: The two observations sets obtain from the same population.
- If the difference between the two averages is sufficiently large, relative to measurement scatter, we reject the null hypothesis and conclude that the two measurement sets obtain from **different** underlying populations.
- **Exemplary conclusion**: With 95% confidence, we conclude that material A is significantly harder than material B.

Assumptions

- Populations are normally distributed. (An “observation set” is a limited, random sampling of a “population”.)
- Observation sets are independent (not paired). (An example of paired sets would be test scores of a group of students before studying and test scores for that same group of students after studying.)
- Second population may have a mean that is either higher or lower than first population. (Two-tailed t-test, not one-tailed t-test. One-tailed t-test is used when you start by assuming that the controlled variable MUST cause the second population to have a mean which is greater or less than the first; i.e. studying MUST cause higher grades.)

Assumptions

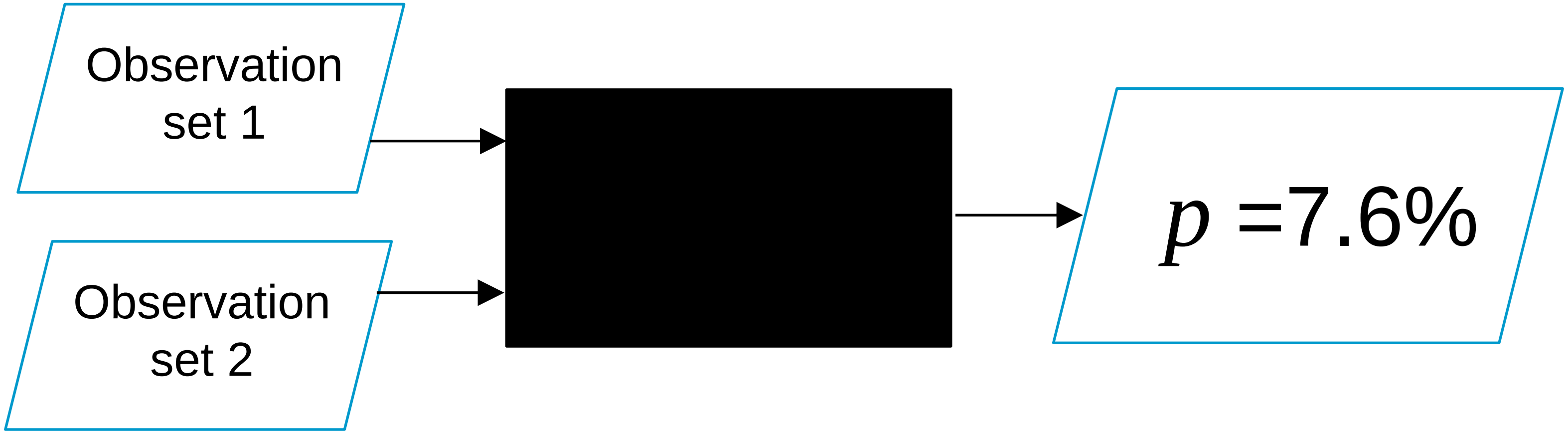
- Populations are normally distributed. (An “observation set” is a limited representative of “population”.)
- Observation sets are independent (not paired). (An example of paired sets would be test scores of a group of students before studying and test scores for that same group of students after studying.)
- Second population may have a mean that is either higher or lower than first population (**Two-tailed t-test**).

Step-by-step procedure for using Student’s t-test to determine if a measured difference is signficant

1. Decide on a confidence level.

What confidence level?

The student’s t-test involves calculating the statistical probability (*p*) that the two measurement sets could have come from the same population.

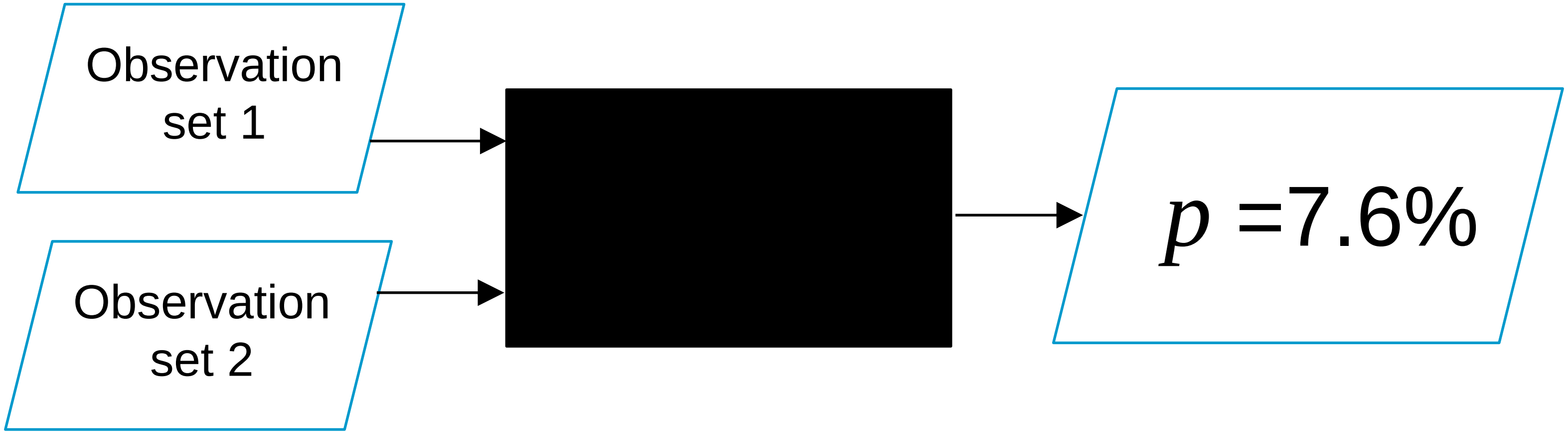


Observation set 1

Observation set 2

p =7.6%

What confidence level?

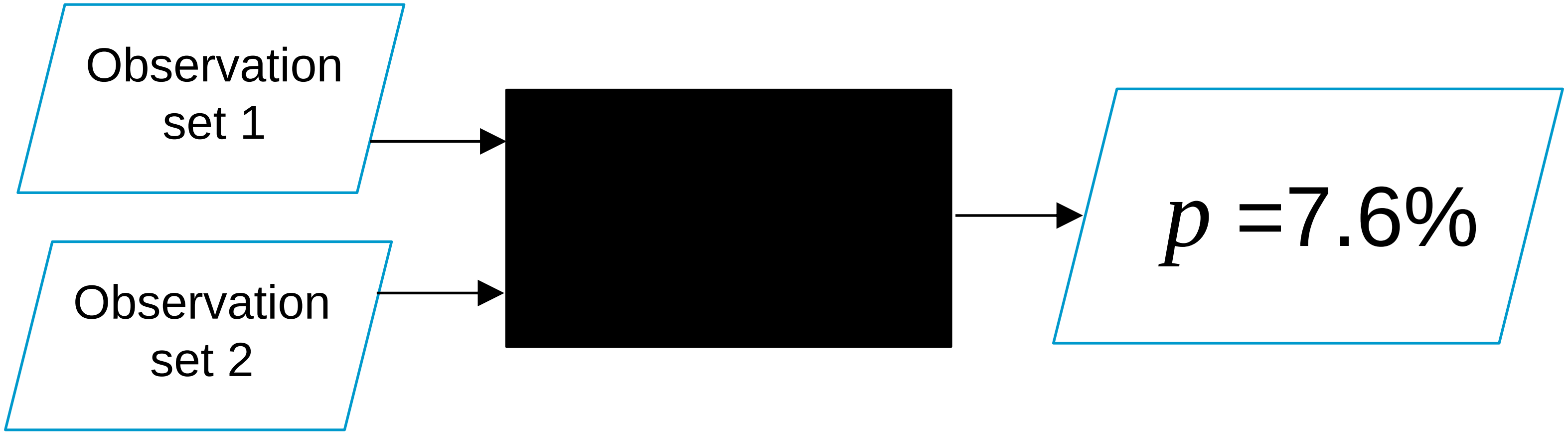


Observation set 1

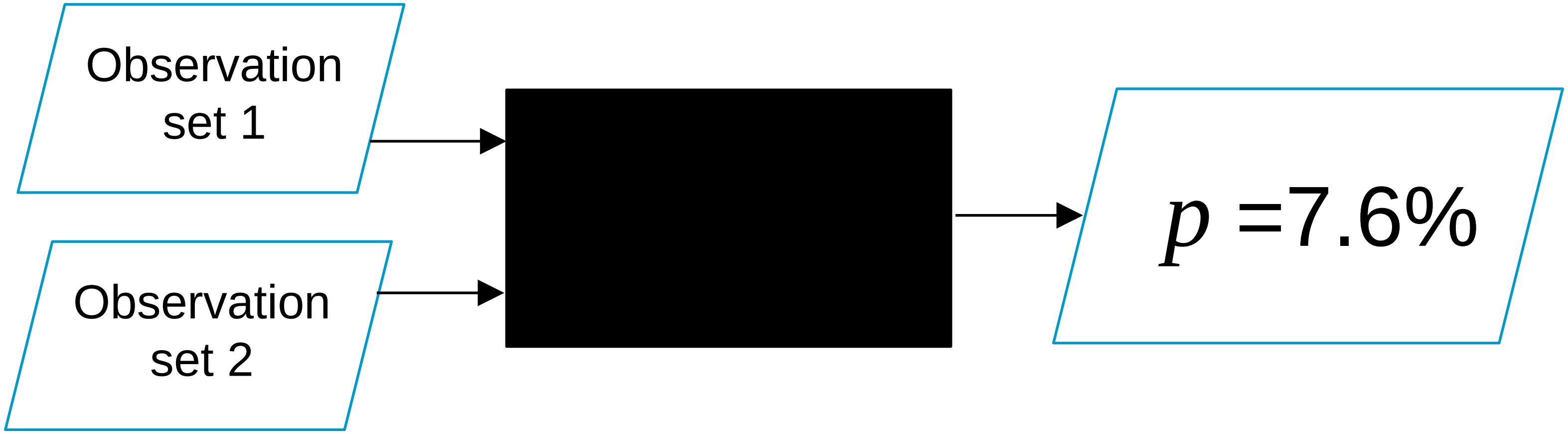
Observation set 2

p =7.6%

What confidence level?



What confidence level?

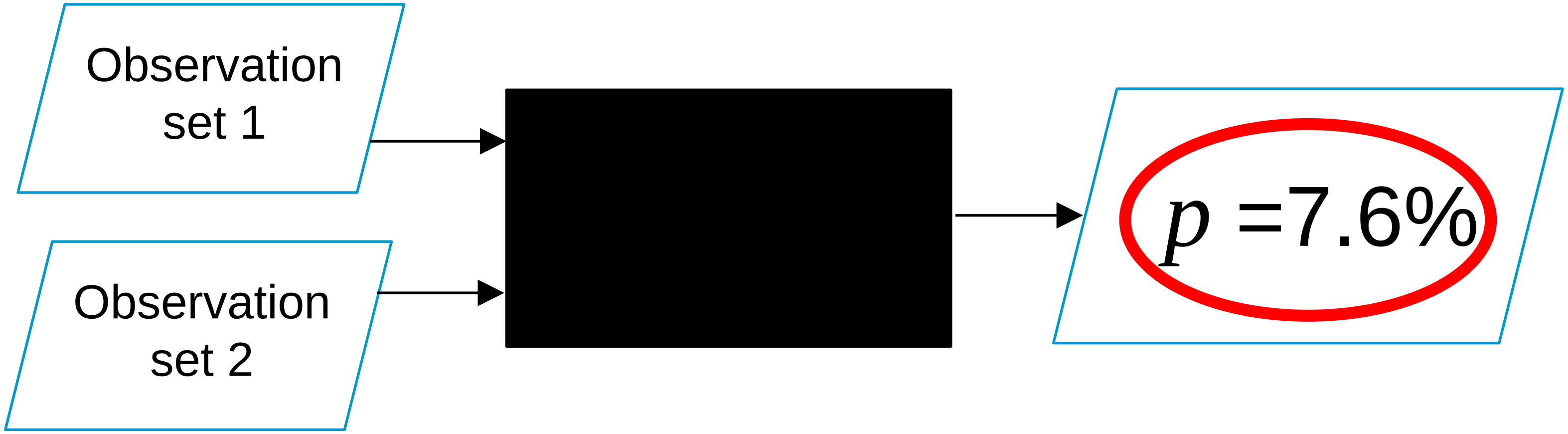


Observation set 1

Observation set 2

p =7.6%

What confidence level?



Observation set 1

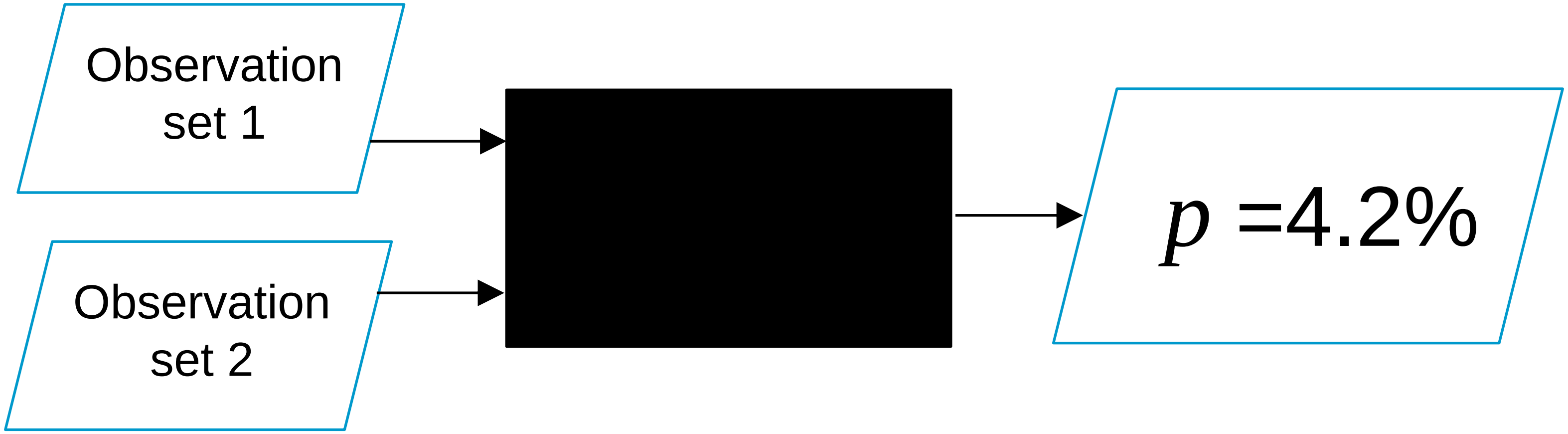
Observation set 2

p =7.6%

The experimenter must define what threshold probability is to be accepted as significant (confidence level). This is a personal judgment in the light of tradition.

Aside: The selection of confidence level is one example by which Michael Polanyi contradicts the notion that “objectivity” is synonymous with “impersonal” in science (*Personal Knowledge*).

Is this significant?



Observation set 1

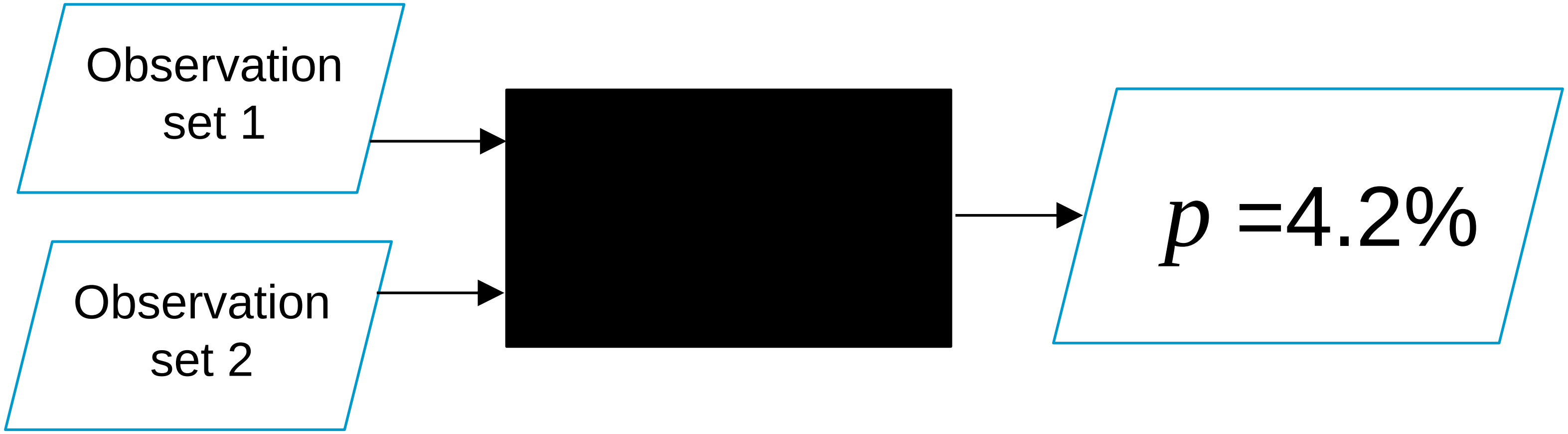
Observation set 2

p =4.2%

In the biological/medical communities, the tradition is to consider any probability less than 5% as significant, so if the observations are made in those fields, then yes.

But this tradition (of $p < 0.05$) derives largely from the relative difficulty of making experimental observations in biology.

Probability depends on number of observations



Say (for a moment) that the two observation sets derive from two populations that really are different.

Then the probability returned by the t-test will decrease as the number of observations increases.

If observations are “cheap” (as they are with indentation), there is no reason to settle for $p < 0.05$ for significance.

Threshold on probability of common population ➡ Confidence level

Probability threshold	Confidence level
5%	95%
1%	99%
0.1%	99.9%

Threshold on probability of common population ➡ Confidence level

Probability threshold	Confidence level
5%	95%
1%	99%
0.1%	99.9%

Student's t-test, step-by-step

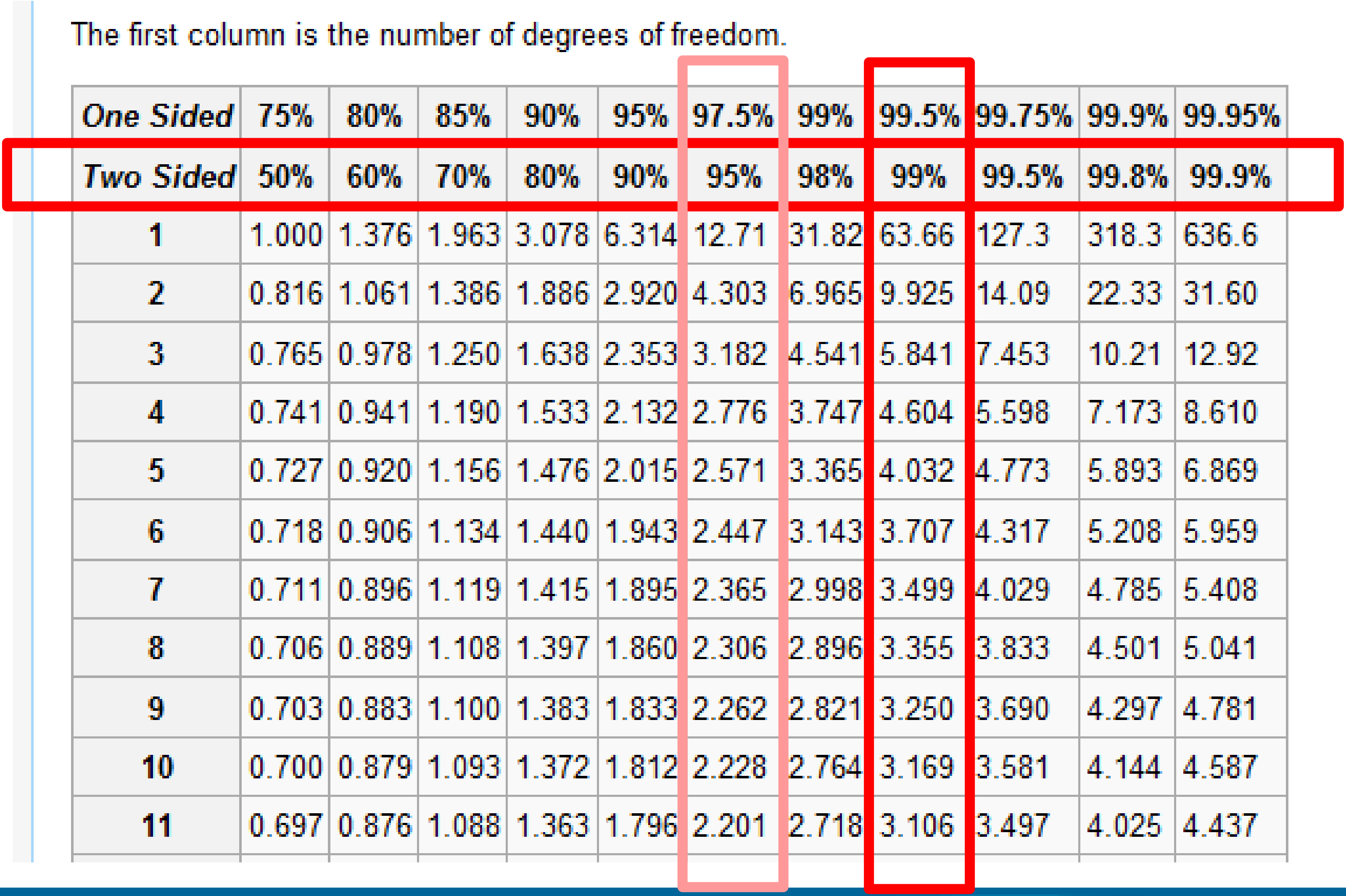
1. Decide on a confidence level (JLH recommends 99% for indentation testing, but 95% is also common).
2. Calculate average and standard deviation for each observation set (X_1, σ_1; X_2, σ_2).
3. Calculate degrees of freedom, df, where N_1 and N_2 are the number of observations in sets 1 and 2, respectively:

$$df = N_1 + N_2 - 2$$

Student's t-test, step-by-step

4. Look up Z_{crit} in a *t*-distribution table by looking for the value of df in the first column and then finding the corresponding value under the column for a two-tailed comparison at 99% (or other desired) confidence.

t-distribution table
(http://en.wikipedia.org/wiki/Student's_t-distribution)



The first column is the number of degrees of freedom.

One Sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two Sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437

Student’s t-test, step-by-step

4. Look up Z_{crit} in a *t*-distribution table by looking for the value of df in the first column and then finding the corresponding value under the column for a two-tailed comparison at 99% (or other desired) confidence.

Student’s t-test, step-by-step

4. Look up Z_{crit} in a t-distribution table by looking for the value of df in the first column and then finding the corresponding value under the column for a two-tailed comparison at 99% (or other desired) confidence.

5. Calculate the value of the test statistic, Z, as:

$$Z = \frac{\left| X_1 - X_2 \right|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

Student’s t-test, step-by-step

4. Look up Z_{crit} in a t-distribution table by looking for the value of df in the first column and then finding the corresponding value under the column for a two-tailed comparison at 99% (or other desired) confidence.

5. Calculate the value of the test statistic, Z, as:

$$Z = \frac{\left| X_1 - X_2 \right|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

6. Compare Z and Z_{crit}. If Z is greater, reject the null hypothesis and conclude that the observation sets are significantly different (drawn from different populations).

Example (H_A = 4.91±0.23; H_B = 5.17±0.21; N = 10)

We perform 10 indentations on each of two materials.
Material A: H = 4.91±0.23
Material B: H = 5.17±0.21

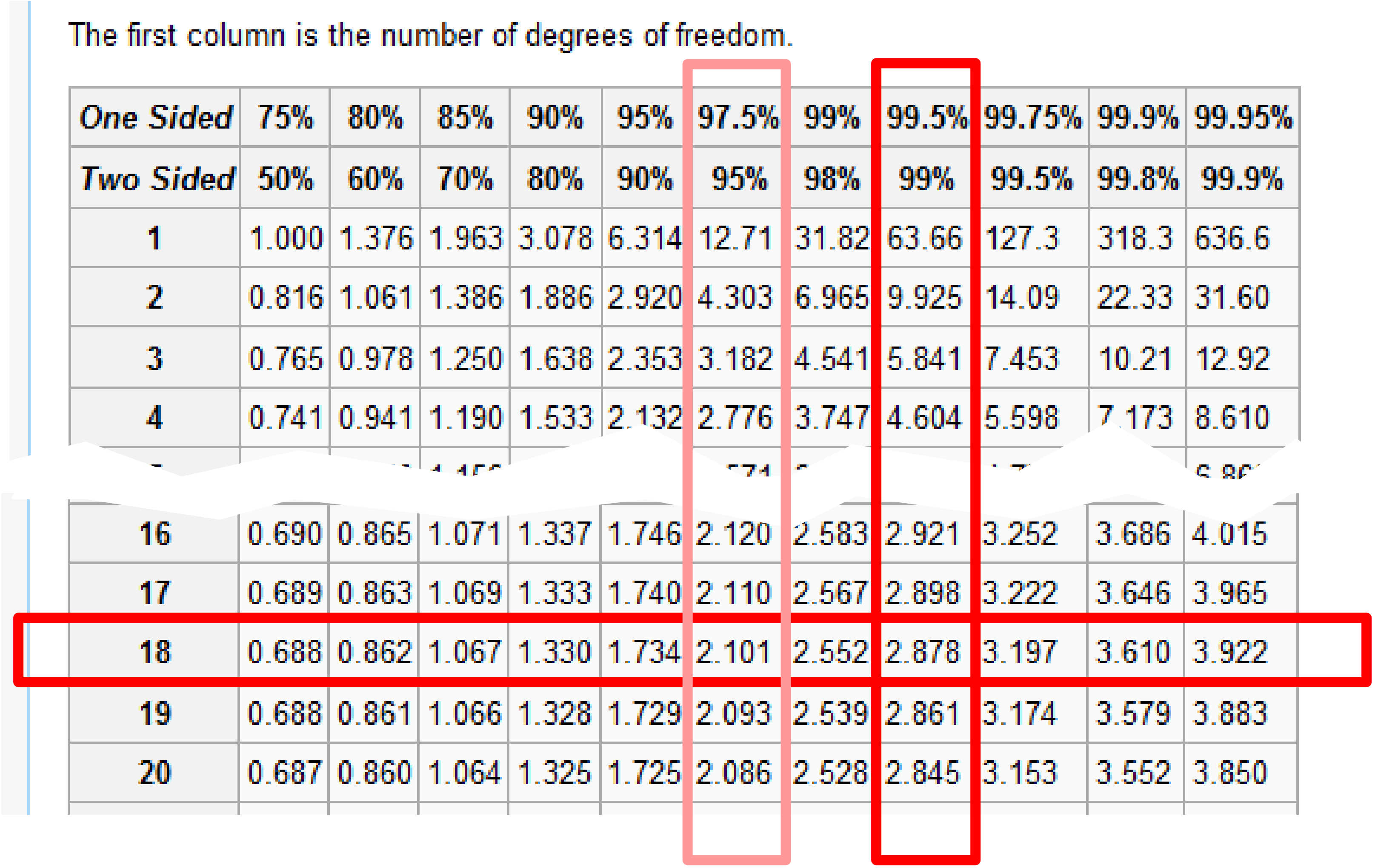
Is this difference in hardness significant or not?

Example (H_A = 4.91±0.23; H_B = 5.17±0.21; N = 10)

1. Decide on a confidence level: **99%**.
2. Calculate average and standard deviation for each observation set (X_1=4.91, σ_1=0.23; X_2=5.17, σ_2=0.21).
3. Calculate degrees of freedom: df = 10+10-2 = 18.
4. Look up Z_{crit}

t-distribution table

(http://en.wikipedia.org/wiki/Student's_t-distribution)



The first column is the number of degrees of freedom.

One Sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two Sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850

Example (H_A = 4.91±0.23; H_B = 5.17±0.21; N = 10)

1. Decide on a confidence level: **99%**.
2. Calculate average and standard deviation for each observation set (X_1=4.91, σ_1=0.23; X_2=5.17, σ_2=0.21).
3. Calculate degrees of freedom: df = 10+10-2 = 18.
4. Look up Z_{crit} : 2.878 (Z_{crit} = 2.101 at 95% confidence).

Example (H_A = 4.91±0.23; H_B = 5.17±0.21; N = 10)

1. Decide on a confidence level: **99%**.
2. Calculate average and standard deviation for each observation set (X_1=4.91, σ_1=0.23; X_2=5.17, σ_2=0.21).
3. Calculate degrees of freedom: df = 10+10-2 = 18.
4. Look up Z_{crit}: 2.878 (Z_{crit} = 2.101 at 95% confidence).
5. Calculate the value of the test statistic:

$$Z = \frac{|4.91 - 5.17|}{\sqrt{\frac{(0.23)^2}{10} + \frac{(0.21)^2}{10}}} = 2.64$$

Example (H_A = 4.91±0.23; H_B = 5.17±0.21; N = 10)

1. Decide on a confidence level: **99%**.
2. Calculate average and standard deviation for each observation set (X_1=4.91, σ_1=0.23; X_2=5.17, σ_2=0.21).
3. Calculate degrees of freedom: df = 10+10-2 = 18.
4. Look up Z_{crit} : 2.878 (Z_{crit} = 2.101 at 95% confidence).
5. Calculate the value of the test statistic: Z = 2.64.
6. Compare Z and Z_{crit}.

Example (H_A = 4.91±0.23; H_B = 5.17±0.21; N = 10)

1. Decide on a confidence level: **99%**.
2. Calculate average and standard deviation for each observation set (X_1=4.91, σ_1=0.23; X_2=5.17, σ_2=0.21).
3. Calculate degrees of freedom: df = 10+10-2 = 18.
4. Look up Z_{crit} : 2.878 (Z_{crit} = 2.101 at 95% confidence).
5. Calculate the value of the test statistic: Z = 2.64.
6. Compare Z and Z_{crit}.

Conclusion: At the level of 99% confidence, the two observation sets are not sigificantly different. (But at 95% they are.)

What's N got to do with it?

$$Z = \frac{\left|4.91 - 5.17\right|}{\sqrt{\frac{(0.23)^2}{10} + \frac{(0.21)^2}{10}}} = 2.64$$

What’s N got to do with it?

$$Z = \frac{|4.91 - 5.17|}{\sqrt{\frac{(0.23)^2}{\cancel{10} \searrow 20} + \frac{(0.21)^2}{\cancel{10} \searrow 20}}} = 2.64$$

What’s N got to do with it?

$$Z = \frac{\left|4.91 - 5.17\right|}{\sqrt{\frac{(0.23)^2}{\cancelto{20}{10}} + \frac{(0.21)^2}{\cancelto{20}{10}}}} = \cancelto{3.37}{2.64}$$

What's N got to do with it?

$$Z = \frac{|4.91-5.17|}{\sqrt{\frac{(0.23)^2}{\underset{\searrow\;20}{10}} + \frac{(0.21)^2}{\underset{\searrow\;20}{10}}}} = \underset{\searrow\;3.37 > 2.88}{2.64}$$

Whereas we could NOT conclude this difference to be significant at 99% confidence with only 10 indents, we CAN conclude it to be signficant at 99% confidence with 20 indents!!

A word of warning

- The student's t-test can discern significant difference between two observation sets.
- IT DOES NOT discern the <u>cause</u> of that difference for you.
- If the experiment is well designed, then the difference (if there is one) can be credited to the influence of <u>a single</u> independent parameter (e.g. tempering time).
- The well designed experiment minimizes the influence of all other physical variables (sample temperature, frame stiffness, indenter shape).

Using the student’s t-test to estimate the number of observations needed to conclude significant difference at a particular confidence level

Solving Student's t-test for N

$$\frac{\left|X_1 - X_2\right|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} > Z_{crit}$$

Set $N = N_1 = N_2$

$$\frac{\left|X_1 - X_2\right|}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{N}}} > Z_{crit}$$

Isolate variables which depend on N

$$\frac{N}{Z_{crit}^2} > \frac{\sigma_1^2 + \sigma_2^2}{\left(X_1 - X_2\right)^2}$$

Simplifications

$$\frac{N}{Z_{crit}^2} > \frac{\sigma_1^2 + \sigma_2^2}{\left(X_1 - X_2\right)^2}$$

$$X_2 = F \cdot X_1,\ F < 1$$

$$\sigma_1 = q \cdot X_1$$

$$\sigma_2 = q \cdot X_2 = q \cdot F \cdot X_1$$

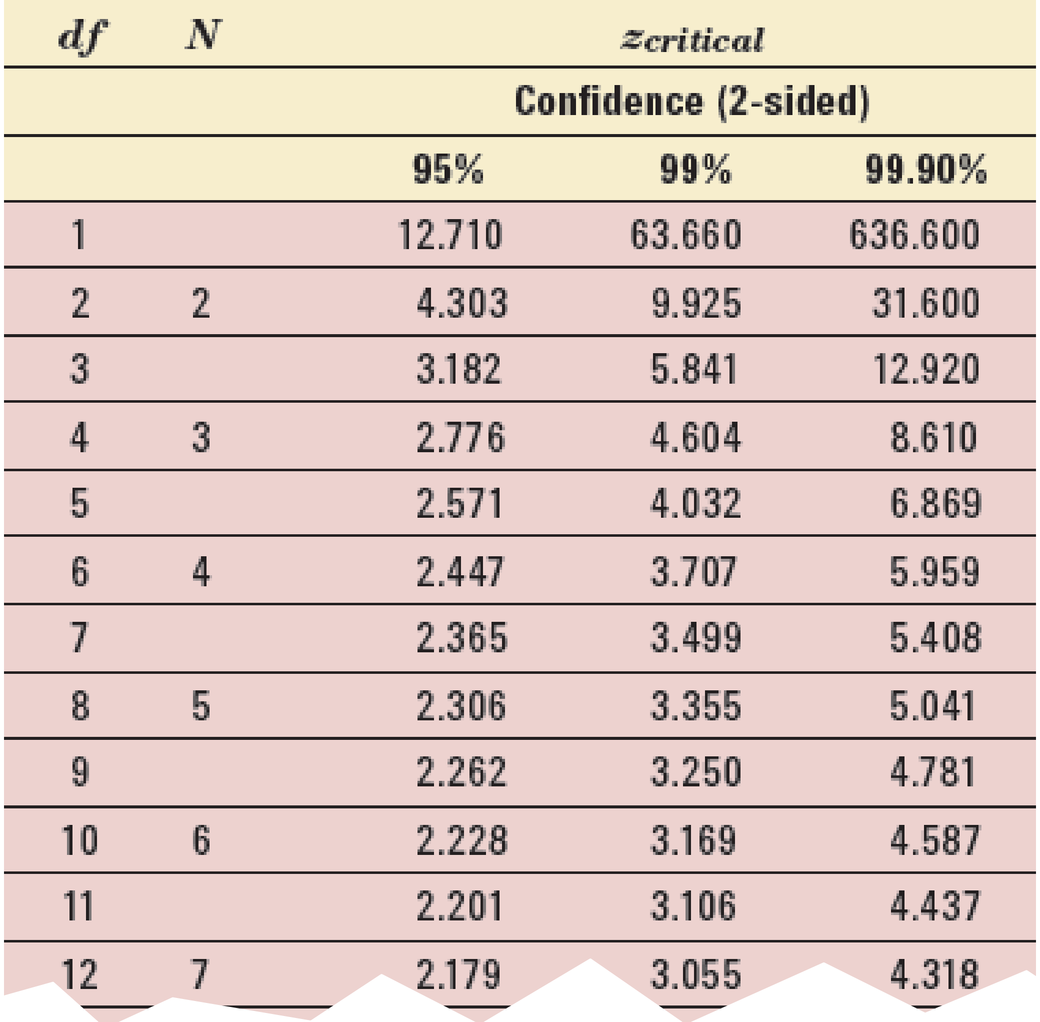
$$\frac{N}{Z_{crit}^2} > \frac{q^2 X_1^2 + F^2 q^2 X_1^2}{X_1^2 \left(1 - F\right)^2}$$

$$\frac{N}{Z_{crit}^2} > \frac{q^2 \left(1 + F^2\right)}{\left(1 - F\right)^2}$$

Z_{crit} = 2.58 for $N = \infty$ (99%);

Z_{crit} = 3.36 for $N = 5$ (99%)

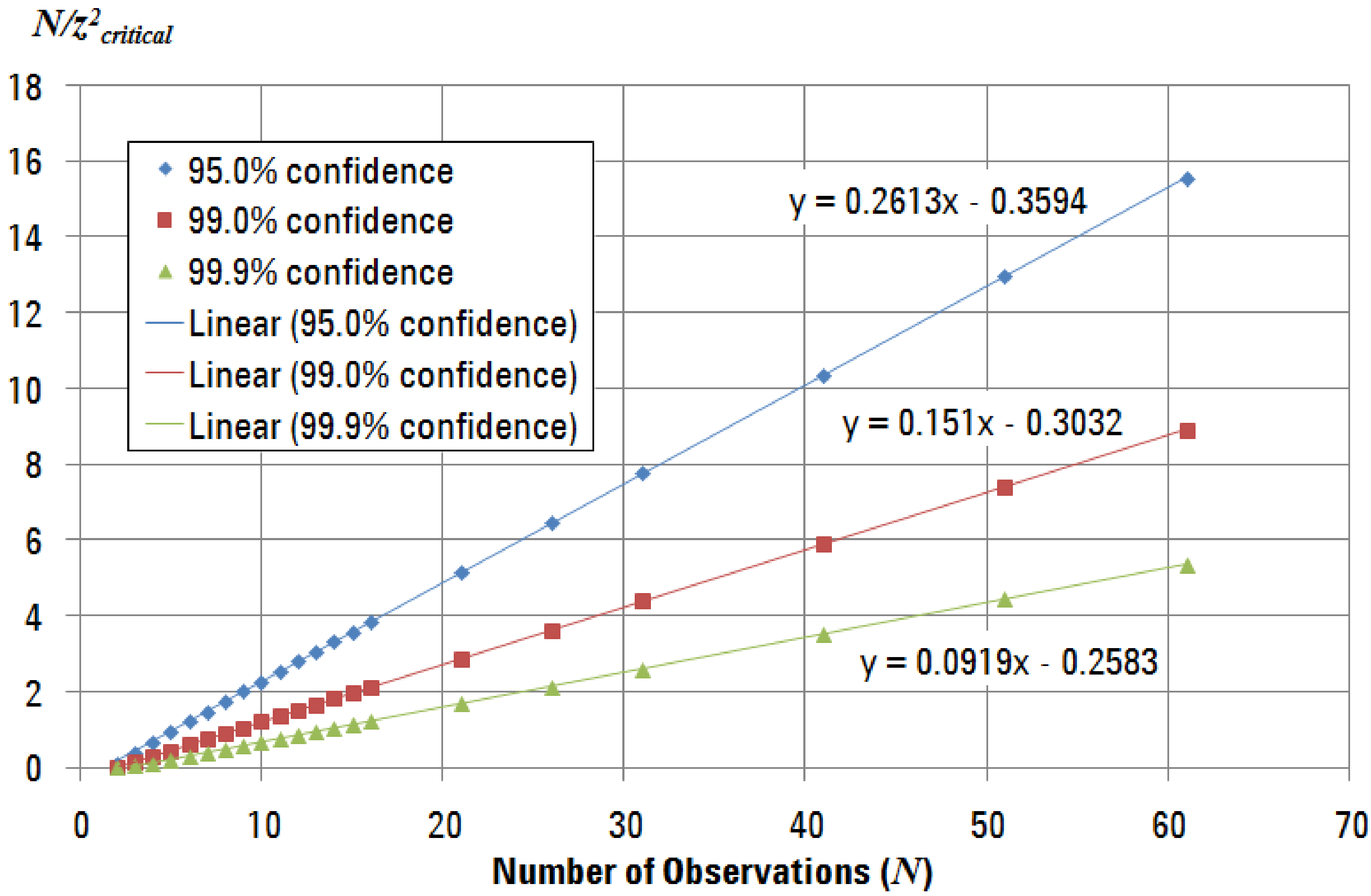
N/Z^2_{crit} as a function of N



df	N	$z_{critical}$		
		Confidence (2-sided)		
		95%	99%	99.90%
1		12.710	63.660	636.600
2	2	4.303	9.925	31.600
3		3.182	5.841	12.920
4	3	2.776	4.604	8.610
5		2.571	4.032	6.869
6	4	2.447	3.707	5.959
7		2.365	3.499	5.408
8	5	2.306	3.355	5.041
9		2.262	3.250	4.781
10	6	2.228	3.169	4.587
11		2.201	3.106	4.437
12	7	2.179	3.055	4.318

Calculate a column of N/Z^2_{crit} and plot as a function of N.

N/Z^2_{crit} as a function of N



Solving Student's t-test for N

$$\frac{N}{Z_{crit}^{2}} > \frac{q^{2}\left(1+F^{2}\right)}{\left(1-F\right)^{2}}$$

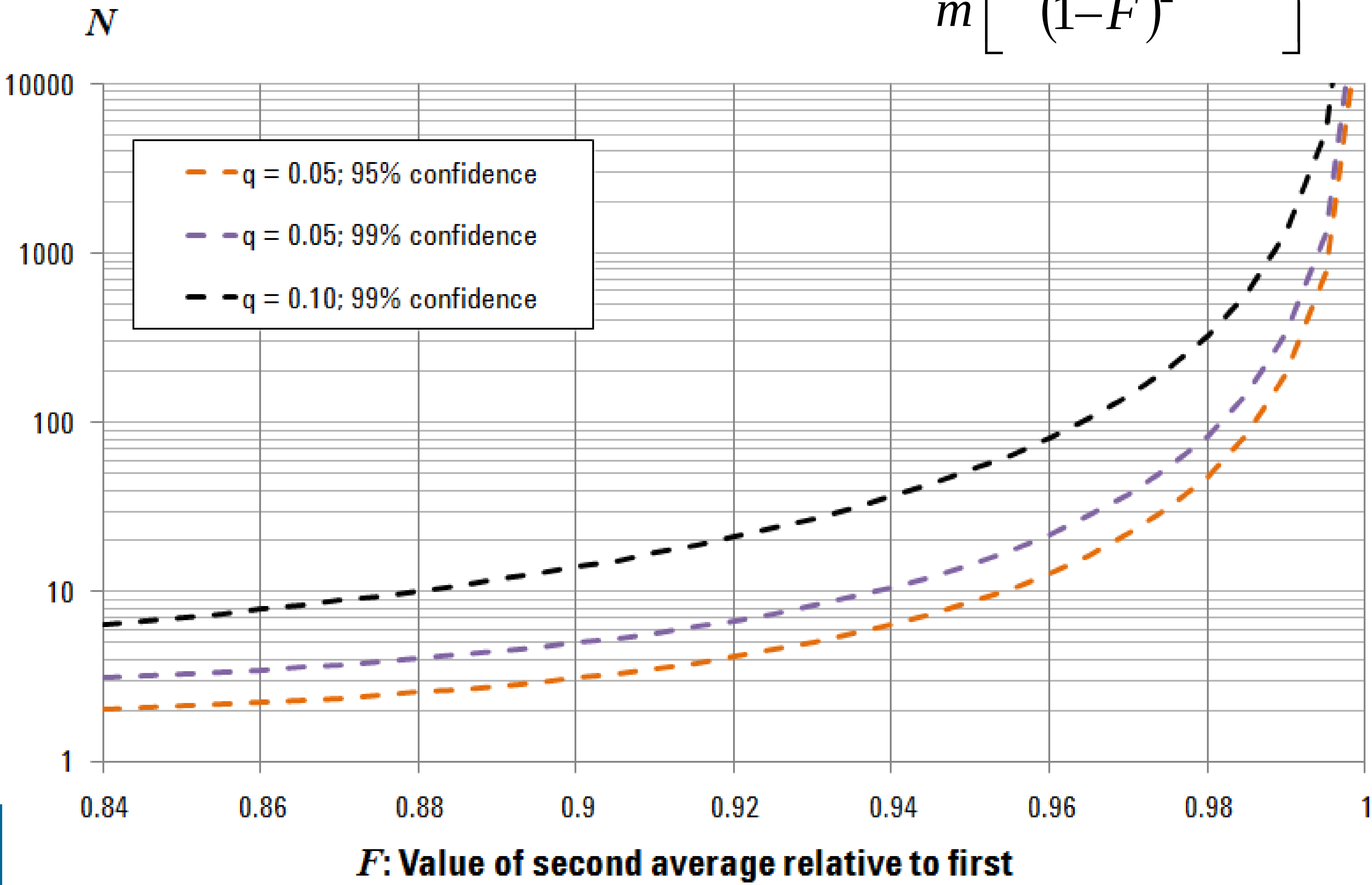
$$mN + b > \frac{q^{2}\left(1+F^{2}\right)}{\left(1-F\right)^{2}}$$

$$N > \frac{1}{m}\left[\frac{q^{2}\left(1+F^{2}\right)}{\left(1-F\right)^{2}} - b\right]$$

Confidence Level	m	b
95.0%	0.2613	-0.3594
99.0%	0.1510	-0.3032
99.9%	0.0919	-0.2583

Functionality

$$N > \frac{1}{m}\left[\frac{q^2\left(1+F^2\right)}{\left(1-F\right)^2} - b\right]$$



Example

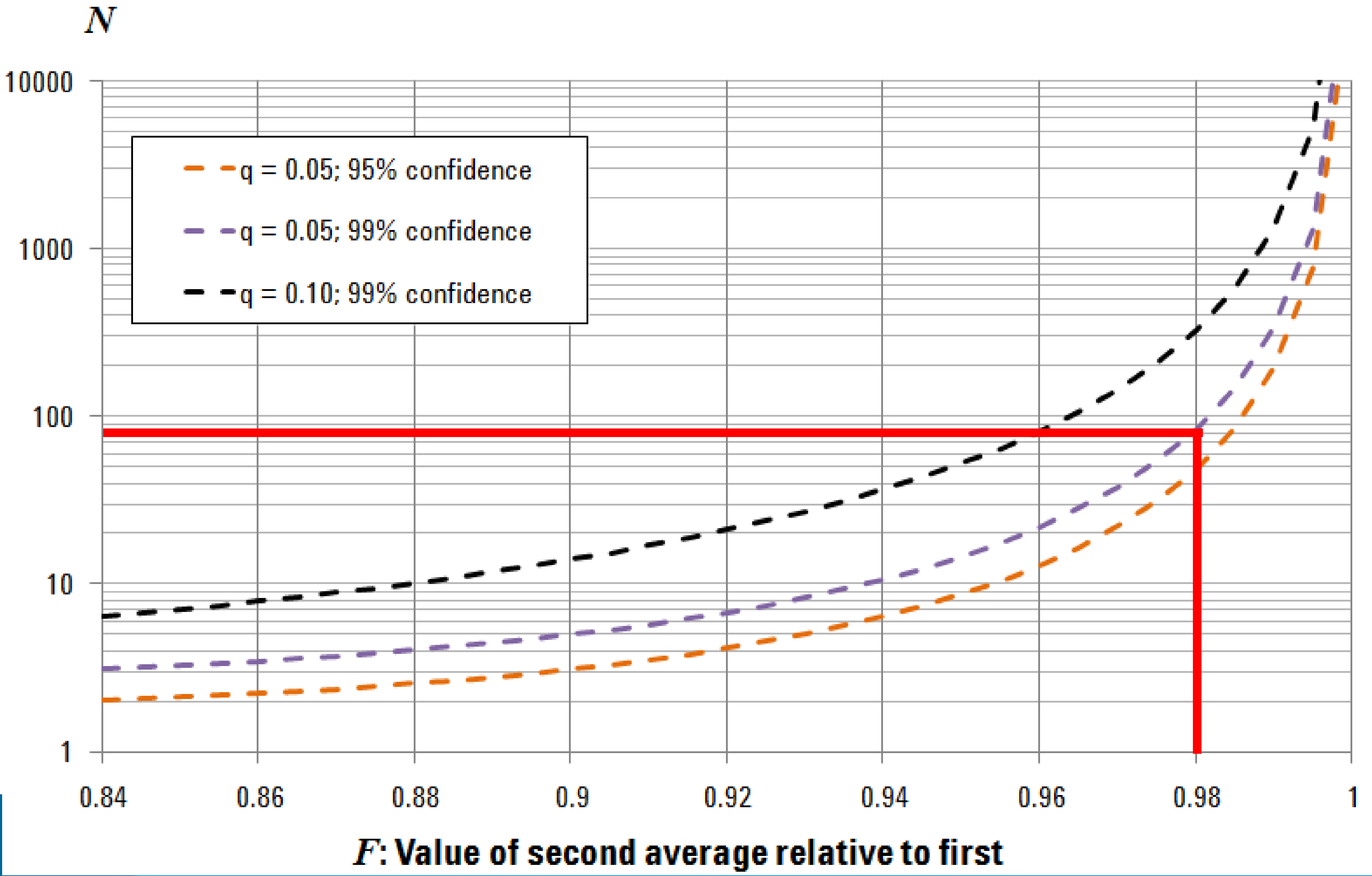
$$N > \frac{1}{m}\left[\frac{q^2\left(1+F^2\right)}{\left(1-F\right)^2}-b\right]$$

We have changed our film-deposition process and wish to know if that change has caused the hardness to change by more than 2%. Indentation tests on our films typically have a standard deviation of 5%.

How many indentations must we do on each film (old and new) in order to determine whether the hardness of the new film differs by more than 2% from that of the old film?

$$N > \frac{1}{0.151}\left[\frac{(0.05)^2\left(1+0.98^2\right)}{\left(1-0.98\right)^2}-0.3\right] \Rightarrow N > 79.1 \rightarrow 80$$

Example: Detect a 2% change with 5% std. dev.



Summary

- The Student’s t-test provides a means for discerning whether the difference between two observation sets is significant.
- The experimentalist must select the threshold probability (of a common population) below which the null hypothesis will be rejected. This is done with consideration for tradition and the cost of individual observations.
- With some simplification, the Student’s t-test can be used to predict the number of observations needed to discern a particular difference.

Session 9: Theory of indentation creep

Wednesday, June 12, 2013, 11:00 (New York)

Abstract

In addition to hardness and Young's modulus, instrumented indentation can be used to characterize creep in metals. This is because hardness is a manifestation of the yield stress of the metal. Under conditions of creep, the yield stress depends on temperature and strain rate. As a manifestation of yield stress, hardness is not a constant, but instead depends on temperature and strain rate just as yield stress does. By quantifying these dependencies, instrumented indentation can be used to determine the stress exponent and activation energy for creep.

To register:

https://agilenteseminar.webex.com/agilenteseminar/onstage/g.php?p=117&t=m

Suggested reading for next session

- Lucas, B. N., and W. C. Oliver. "Indentation Power-Law Creep of High-Purity Indium." *Metallurgical and Materials Transactions A-Physical Metallurgy and Materials Science* 30.3 (1999): 601-610.
- Agilent application note: “In Situ Young’s Modulus and Strain-Rate Sensitivity of Lead-Free SAC 105 Solder” (http://cp.literature.agilent.com/litweb/pdf/5991-2144EN.pdf)

Upcoming workshops

- **June 2**: Theory & Practice of Instrumented Indentation Testing Course, Sunday, 1:00-5:00p.m., The Westin Lombard Yorktown Center, Lombard, Illinois. A short course offered in conjunction with the SEM 2013 Annual Conference & Exposition on Experimental and Applied Mechanics (http://sem.org/CONF-AC-TOP.asp)
- **June 25-26**: Nano Measure 2013—A symposium for sensing and understanding nano-scale phenomena, University of Warsaw, Poland.

Thank you!