

Biostatistics

Chapter 5 Hypothesis Testing : One or Two samples

Jing Li

jing.li@sjtu.edu.cn

<http://cbb.sjtu.edu.cn/~jingli/courses/2017fall/bi372/>

Dept of Bioinformatics & Biostatistics, SJTU



Review Questions (10 min)

- A study is conducted concerning the blood pressure of 60 year old women with glaucoma (青光眼). In the study 200 60-year old women with glaucoma are randomly selected and the sample mean systolic blood pressure is 140 mm Hg and the sample standard deviation is 25 mm Hg.
 - Calculate a 95% confidence interval for the true mean systolic blood pressure among the population of 60 year old women with glaucoma.
 - Suppose the study above was based on 100 women instead of 200 but the sample mean (140) and standard deviation (25) are the same. Recalculate the 95% confidence interval. Does the interval get wider or narrower? Why?



Review lecture 4

1. Make Assumptions and meet test requirements.
2. State the null hypothesis.
3. Select the sampling distribution and establish the critical region.
4. Compute the test statistic.
5. Make a decision and interpret results.



Topics

- Z-test
 - Testing single population mean
 - Testing sample proportions



Significant Differences

- Hypothesis testing is designed to detect *significant differences*: differences that did *not* occur by random chance.
- In the “one sample” case: we compare a random sample (from a large group) to a population.
- We compare a sample statistic to a population parameter to see if there is a *significant difference*.



Our Problem:

- The education department at a university has been accused of “**grade inflation**” so education majors have much higher GPAs than students in general.
- GPAs of all education majors should be compared with the GPAs of all students.
 - There are 1000s of education majors, far too many to interview.
 - How can this be investigated without interviewing all education majors?



What we know:

- The average GPA for **all** students is 2.70. This value is a **parameter**.

$$\mu = 2.70$$

- To the right is the statistical information for a random sample of education majors:

$\bar{X} =$	3.00
$s =$	0.70
$n =$	117



Questions to ask:

- Is there a difference between the parameter (2.70) and the statistic (3.00)?
- Could the observed difference have been caused by random chance?
- Is the difference real (significant)?



Two Possibilities:

1. The sample mean (3.00) is the same as the pop. mean (2.70).
 - **The difference is trivial and caused by random chance.**

2. The difference is real (significant).
 - **Education majors are different from all students.**



The Null and Alternative Hypotheses:

1. Null Hypothesis (H_0)

- The difference is caused by random chance.
- The H_0 always states there is “no significant difference.” In this case, we mean that there is no significant difference between the population mean and the sample mean.

2. Alternative hypothesis (H_1)

- “The difference is real”.
- (H_1) always contradicts the H_0 .

One (and only one) of these explanations *must* be true.

Which one?

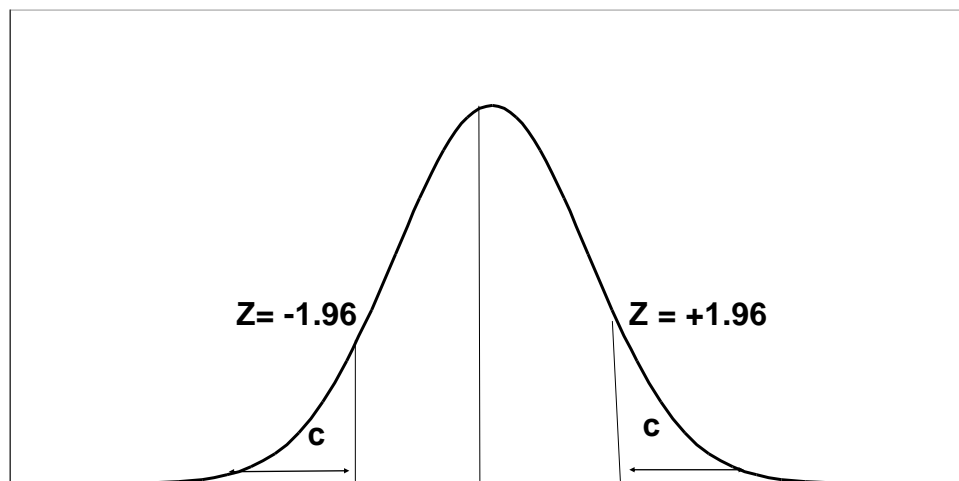


Test the Hypotheses

- Use the .05 value as a guideline to identify differences that would be rare or extremely unlikely if H_0 is true. This “alpha” value delineates the “region of rejection.”
- Use the Z score formula for single samples and determine the probability of getting the observed difference.
- If the probability is less than .05, the calculated or “observed” Z score will be beyond ± 1.96 (the “critical” Z score).



Two-tailed Hypothesis Test:



When $\alpha = .05$, then .025 of the area is distributed on either side of the curve in area **(C)**

The .95 in the **middle section** represents **no significant difference** between the population and the sample mean.

The cut-off between the middle section and +/- .025 is represented by a **Z-value of +/- 1.96**.



Testing Hypotheses: Using The Five Step Model...

1. Make Assumptions and meet test requirements.
2. State the null hypothesis.
3. Select the sampling distribution and establish the critical region.
4. Compute the test statistic.
5. Make a decision and interpret results.



Step I: Make Assumptions and Meet Test Requirements

- Random sampling
 - Hypothesis testing assumes samples were selected using random sampling.
 - In this case, the sample of 117 cases was randomly selected from all education majors.
- Level of Measurement is Numerical
 - the mean is an appropriate statistic.
- Sampling Distribution is normal in shape
 - This is a “large” sample ($n \geq 100$).



Step 2 State the Null Hypothesis

- $H_0: \mu = 2.7$ (in other words, $H_0: \bar{X} = \mu$)
 - You can also state H_0 : No difference between the sample mean and the population parameter
 - (In other words, the sample mean of 3.0 really the same as the population mean of 2.7 – the difference is not real but is due to chance.)
 - The sample of 117 comes from a population that has a GPA of 2.7.
 - The difference between 2.7 and 3.0 is trivial and caused by random chance.



Step 2 (cont.) State the Alternate Hypothesis

- $H_1: \mu \neq 2.7$ (or, $H_0: \bar{X} \neq \mu$)
 - Or H_1 : There is a difference between the sample mean and the population parameter
 - The sample of 117 comes a population that *does not* have a GPA of 2.7. In reality, it comes from a different population.
 - The difference between 2.7 and 3.0 reflects an actual difference between education majors and other students.
 - Note that we are testing whether the population the sample comes from is from a different population or is the same as the general student population.



Step 3 Select Sampling Distribution and Establish the Critical Region

- Sampling Distribution= Z
 - Alpha (α) = .05
 - α is the indicator of “rare” events.
 - Any difference with a probability less than α is rare and will cause us to reject the H_0 .



Step 3 (cont.) Select Sampling Distribution and Establish the Critical Region

- Critical Region begins at $Z = \pm 1.96$
 - This is the critical Z score associated with $\alpha = .05$, two-tailed test.
 - If the obtained Z score falls in the Critical Region, or “the region of rejection,” then we would reject the H_0 .



Step 4: Use Formula to Compute the Test Statistic

When σ is known:

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

When the Population σ is not known, large samples ≥ 100 , use the following formula:

$$Z = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$



Test the Hypotheses

$$Z = \frac{3.0 - 2.7}{\frac{.7}{\sqrt{117}}} = 4.66$$

- Substituting the values into the formula, we calculate a Z score of 4.66.

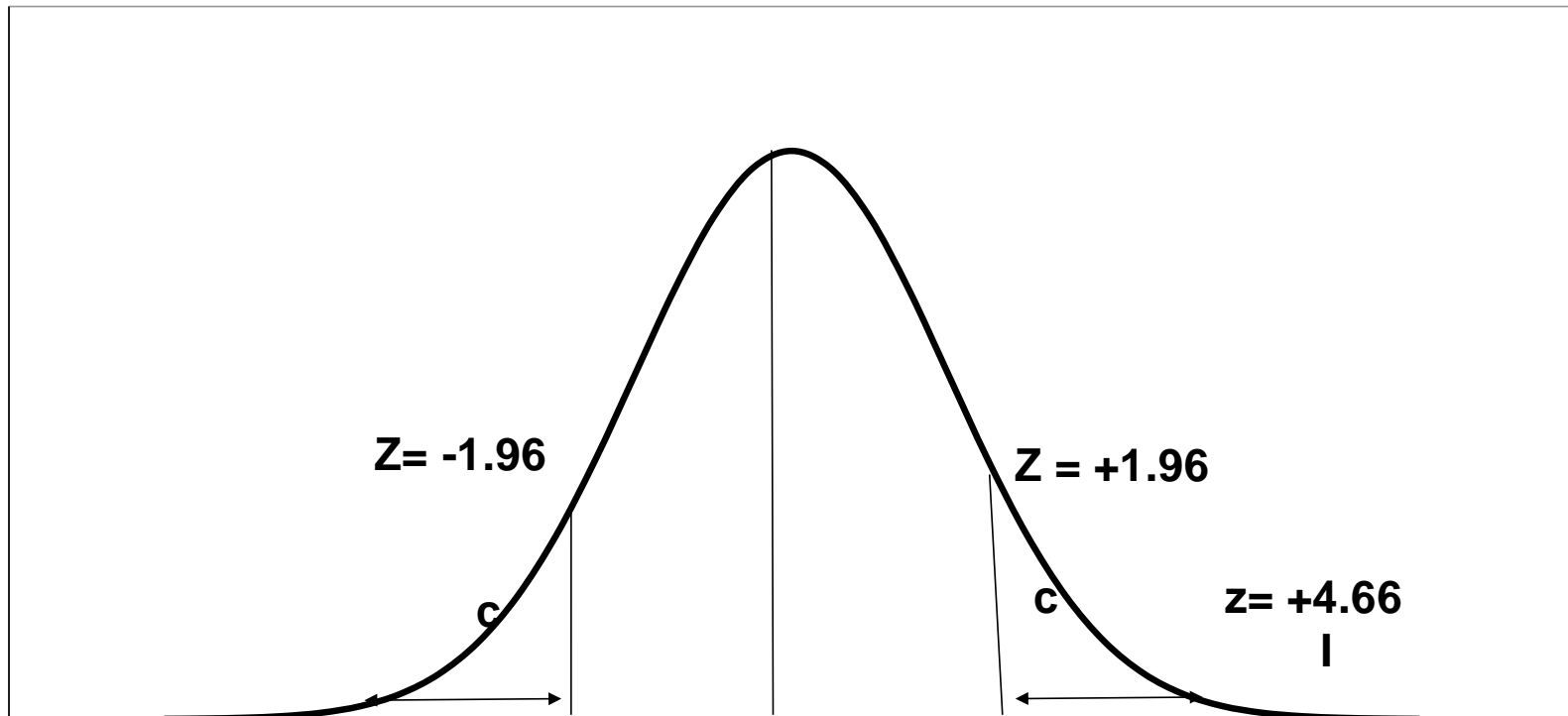


Step 5 Make a Decision and Interpret Results

- The obtained Z score fell in the Critical Region, so we *reject* the H_0 .
 - **If the H_0 were true, a sample outcome of 3.00 would be unlikely.**
 - **Therefore, the H_0 is false and must be rejected.**
- Education majors have a GPA that is significantly different from the general student body ($Z = 4.66$, $\alpha = .05$).*
- *Note: Always report significant statistics.



Looking at the curve:
(Area C = Critical Region when $\alpha=.05$)



Summary:

- The GPA of education majors is *significantly* different from the GPA of the general student body.
- In hypothesis testing, we try to identify statistically significant differences that did not occur by random chance.
- In this example, the difference between the parameter 2.70 and the statistic 3.00 was large and unlikely ($p < .05$) to have occurred by random chance.
- We rejected the H_0 and concluded that the difference was significant.
- It is very likely that Education majors have GPAs higher than the general student body



Rule of Thumb:

- If the test statistic is in the Critical Region ($\alpha=.05$, beyond ± 1.96):
 - ***Reject*** the H_0 . The difference is significant.
- If the test statistic is not in the Critical Region (at $\alpha=.05$, between $+1.96$ and -1.96):
 - ***Fail to reject*** the H_0 . The difference is not significant.



Using the Student's t Distribution for **Small Samples** (One Sample T-Test)

- When the sample size is small (approximately < 100) then the Student's t distribution should be used.
- The test statistic is known as “ t ”.
- The curve of the t distribution is flatter than that of the Z distribution but as the sample size increases, the t -curve starts to resemble the Z -curve.



Topics

- Z-test
 - Testing single population mean
 - Testing sample proportions



Testing Sample Proportions:

- When your variable is at the nominal (or ordinal) level the one sample z-test for proportions should be used.
- If the data are in % format, convert to a proportion first.
- The method is the same as the one sample Z-test for means (see above)



One-sample test of proportions

Examples:

A survey claims that 9 out of 10 doctors recommend aspirin for their patients with headaches. To test this claim, a random sample of 100 doctors is obtained. Of these 100 doctors, 82 indicate that they recommend aspirin. Is this claim accurate? Use $\alpha = 0.05$.

- Hypothesis Test
 - critical value method
 - p-value method



One-sample test of proportions

Hypothesis Test:

Specify the null and alternative

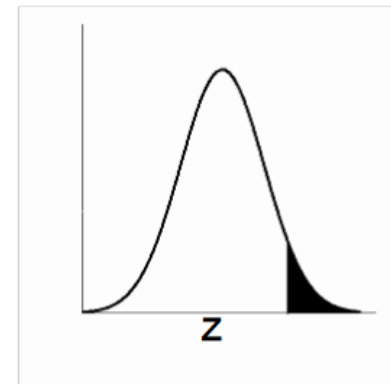
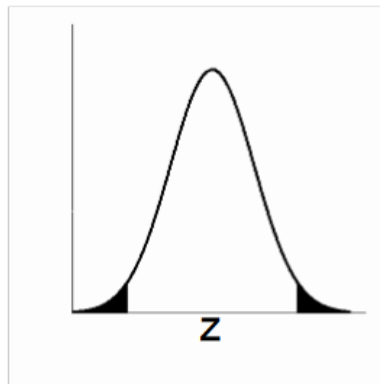
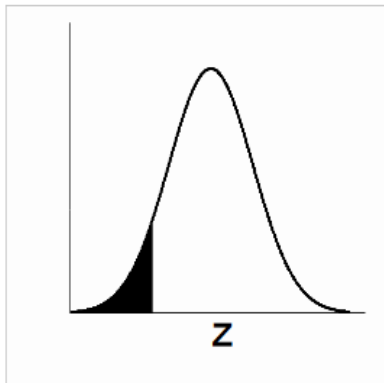
Define rejection region for the test statistic z

$$H_0: p = p_0$$

$$H_A: p < p_0$$

$$H_A: p \neq p_0$$

$$H_A: p > p_0$$



One-sample test of proportions

Test Statistic:
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

If H_0 is true, z comes from standard normal distribution

Critical value method:

Choose z^* from normal table to define rejection region

If z falls in rejection region, reject H_0

P-value method:

Use the normal table to determine probability of a value at least as extreme as z coming from standard normal distribution. Report this p-value and your conclusion.



One-sample test of proportions

1. Define Null and Alternative Hypotheses

$$H_0; p = .90$$

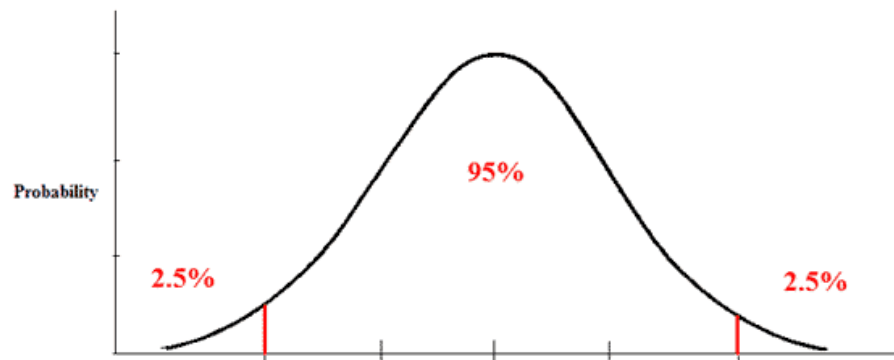
$$H_1; p \neq .90$$

2. State Alpha

Alpha = 0.05

3. State Decision Rule

Using an alpha of 0.05 with a two-tailed test, we would expect our distribution to look something like this:



4. Calculate Test Statistic

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

$$\hat{p} = .82$$

$$p_0 = .90$$

$$n = 100$$

$$z_0 = \frac{.82 - .90}{\sqrt{\frac{.90(1 - .90)}{100}}} = \frac{-0.08}{0.03} = -2.667$$

5. State Results

$$z = -2.667$$

Result: Reject the null hypothesis.

6. State Conclusion

The claim that 9 out of 10 doctors recommend aspirin for their patients is not accurate, $z = -2.667$, $p < 0.05$.



Two-sample test of proportions

- In two independent populations, individuals can be classified into one of two categories. You draw samples from both populations to make inference about whether the category proportions are the same in both populations



Two-sample test of proportions

- Z-test (recall lecture 4)

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

Total number of successes
in both samples

$$SE_{D_p} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$



Practice

“Drug X is administered to 100 patients with a particular disease. 50 improve. Test whether this drug is better than drug Y, which is known to produce improvement in 45% of patients.”

- Hypothesis Test
 - critical value method
 - p-value method

One-sample test of proportions



Practice

- Time magazine reported the result of a telephone poll of 800 adult Americans. The question posed of the Americans who were surveyed was: "Should the federal tax on cigarettes be raised to pay for health care reform?" The results of the survey were:

Non-Smokers	Smokers
$n_1 = 605$	$n_2 = 195$
$y_1 = 351$ said "yes"	$y_2 = 41$ said "yes"
$\hat{p}_1 = \frac{351}{605} = 0.58$	$\hat{p}_2 = \frac{41}{195} = 0.21$

Is there sufficient evidence at the $\alpha = 0.05$ level, say, to conclude that the two populations — smokers and non-smokers — differ significantly with respect to their opinions?



Topics

- Z-test
 - Testing single population mean
 - Testing sample proportions
- One sample T-test
- Comparison of Two means t-test
 - Paired t-test
 - Independent samples



Recall

- When the sample size is small (approximately < 100) then the **Student's t distribution** should be used.
- The test statistic is known as “ t ”.
- The curve of the t distribution is flatter than that of the Z distribution but as the sample size increases, the t -curve starts to resemble the Z -curve.



Degrees of Freedom

- The curve of the t distribution varies with sample size (the smaller the size, the flatter the curve)
- In using the t-table, we use “degrees of freedom” based on the sample size.
- For a one-sample test, $df = n - 1$.
- When looking at the table, find the t-value for the appropriate $df = n - 1$. This will be the cutoff point for your critical region.



Formula for one sample t-test:

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$



Example

- A random sample of 26 sociology graduates scored 458 on the GRE advanced sociology test with a standard deviation of 20. Is this significantly different from the population average ($\mu = 440$)?



Solution (using five step model)

- Step 1: Make Assumptions and Meet Test Requirements:
 - 1. Random sample
 - 2. Level of measurement is numerical
 - 3. The sample is small (<100)



Solution (cont.)

Step 2: State the null and alternate hypotheses.

$$H_0: \mu = 440 \text{ (or } H_0: \bar{X} = \mu)$$

$$H_1: \mu \neq 440$$



Solution (cont.)

- Step 3: Select Sampling Distribution and Establish the Critical Region
 1. Small sample, I-R level, so use t distribution.
 2. Alpha (α) = .05
 3. Degrees of Freedom = $n-1 = 26-1 = 25$
 4. Critical $t = \pm 2.060$



Solution (cont.)

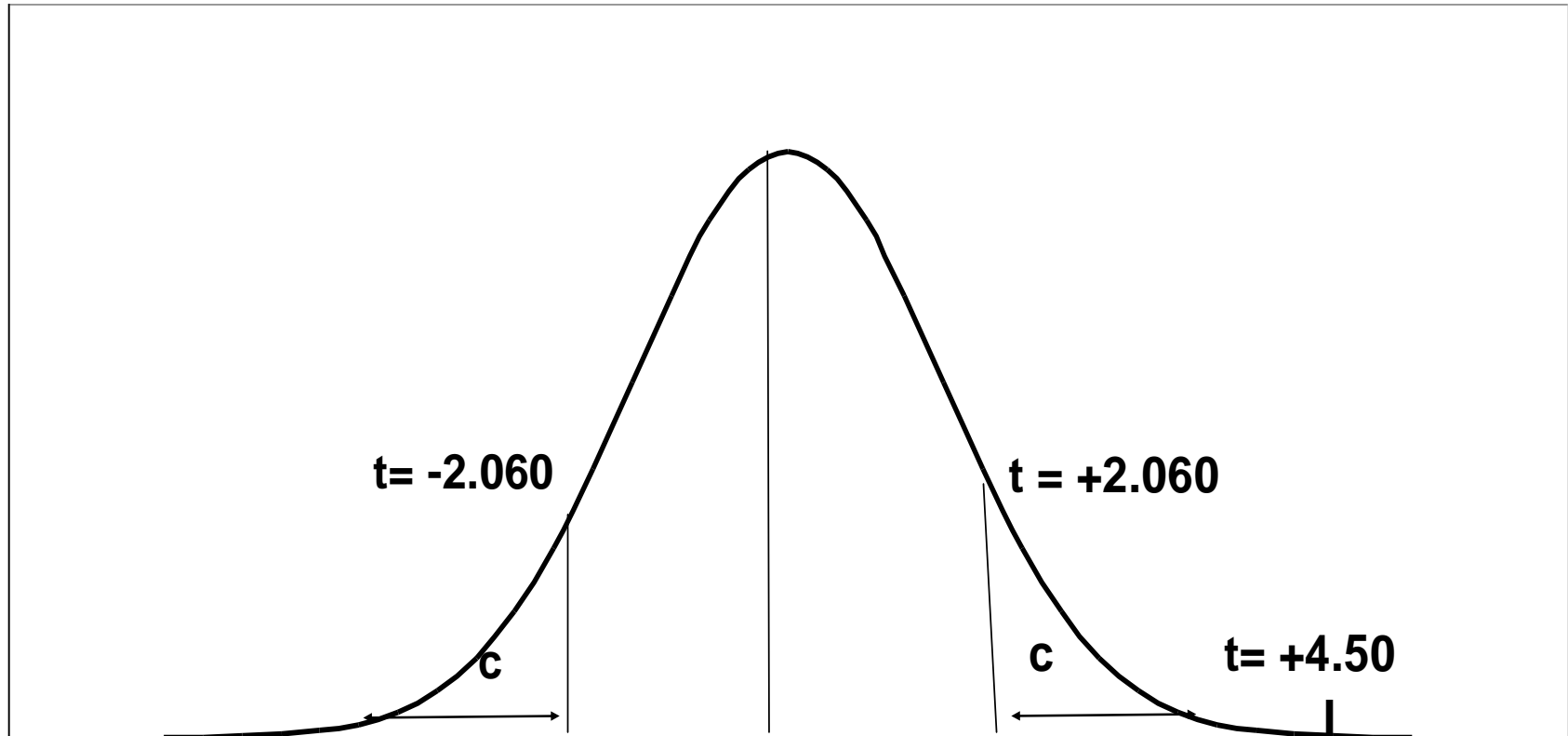
- Step 4: Use Formula to Compute the Test Statistic

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}} = \frac{458 - 440}{20 / \sqrt{26}} = 4.5$$



Looking at the curve for the t distribution

Alpha (α) = .05



Step 5 Make a Decision and Interpret Results

- The obtained t score fell in the Critical Region, so we *reject* the H_0 (t (obtained) $>$ t (critical))
 - **If the H_0 were true, a sample outcome of 458 would be unlikely.**
 - **Therefore, the H_0 is false and must be rejected.**
- Sociology graduates have a GRE score that is significantly different from the general student body ($t = 4.5$, $df = 25$, $\alpha = .05$).

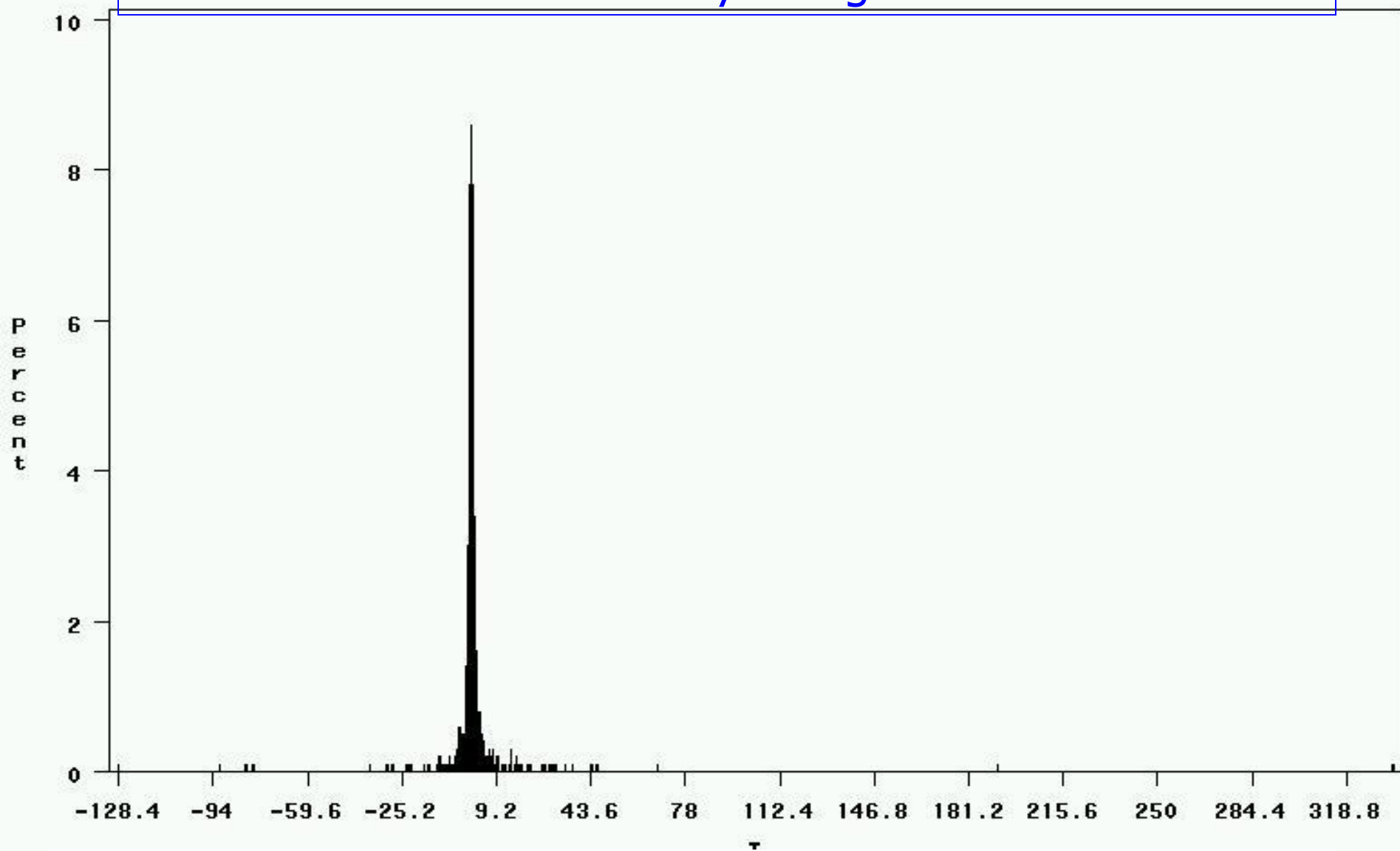


What is a T-distribution?

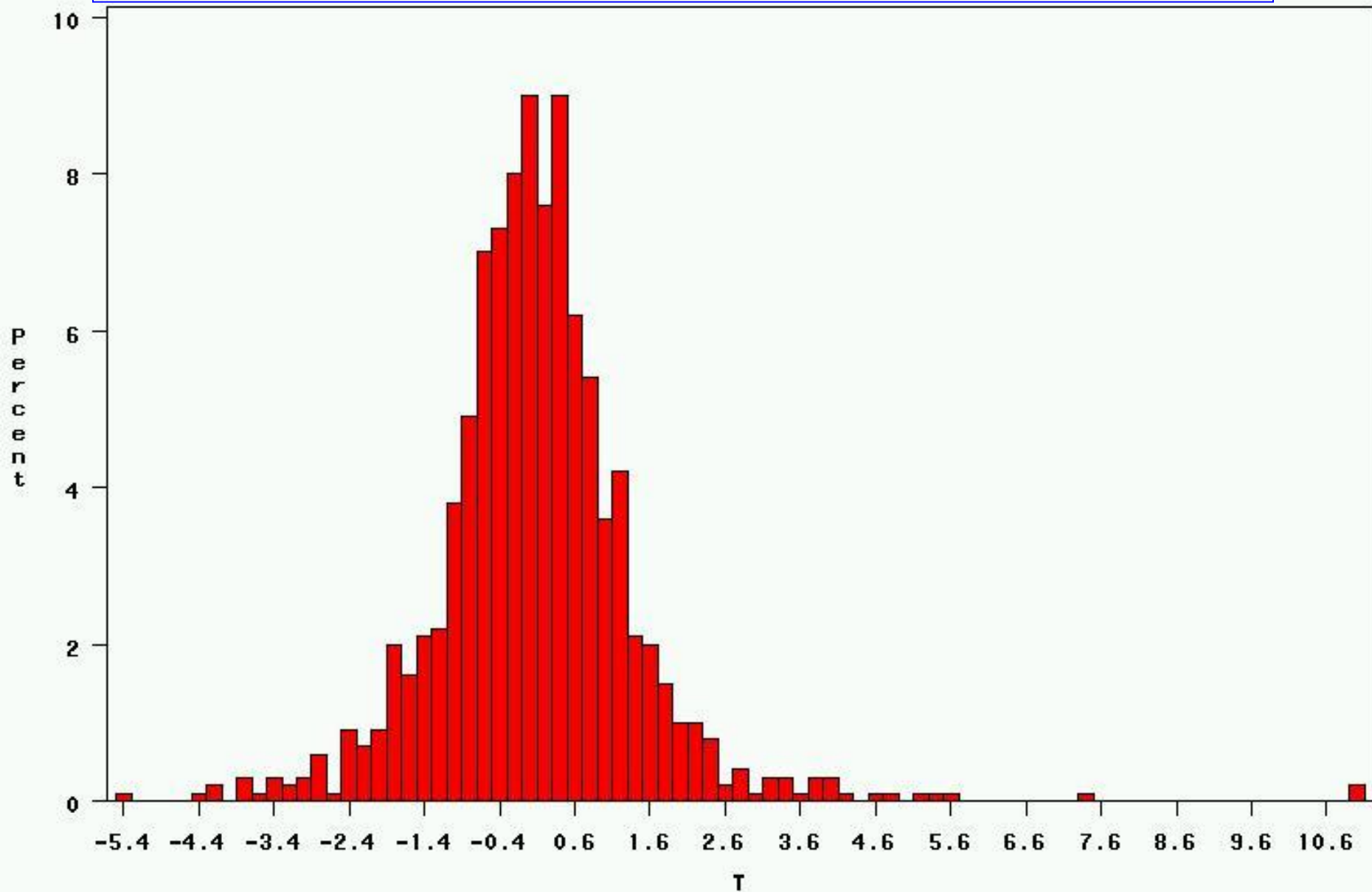
- A t-distribution is like a Z distribution, except has slightly fatter tails to reflect the uncertainty added by estimating σ .
- The bigger the sample size (i.e., the bigger the sample size used to estimate σ), then the closer t becomes to Z.
- If $n > 100$, t approaches Z.



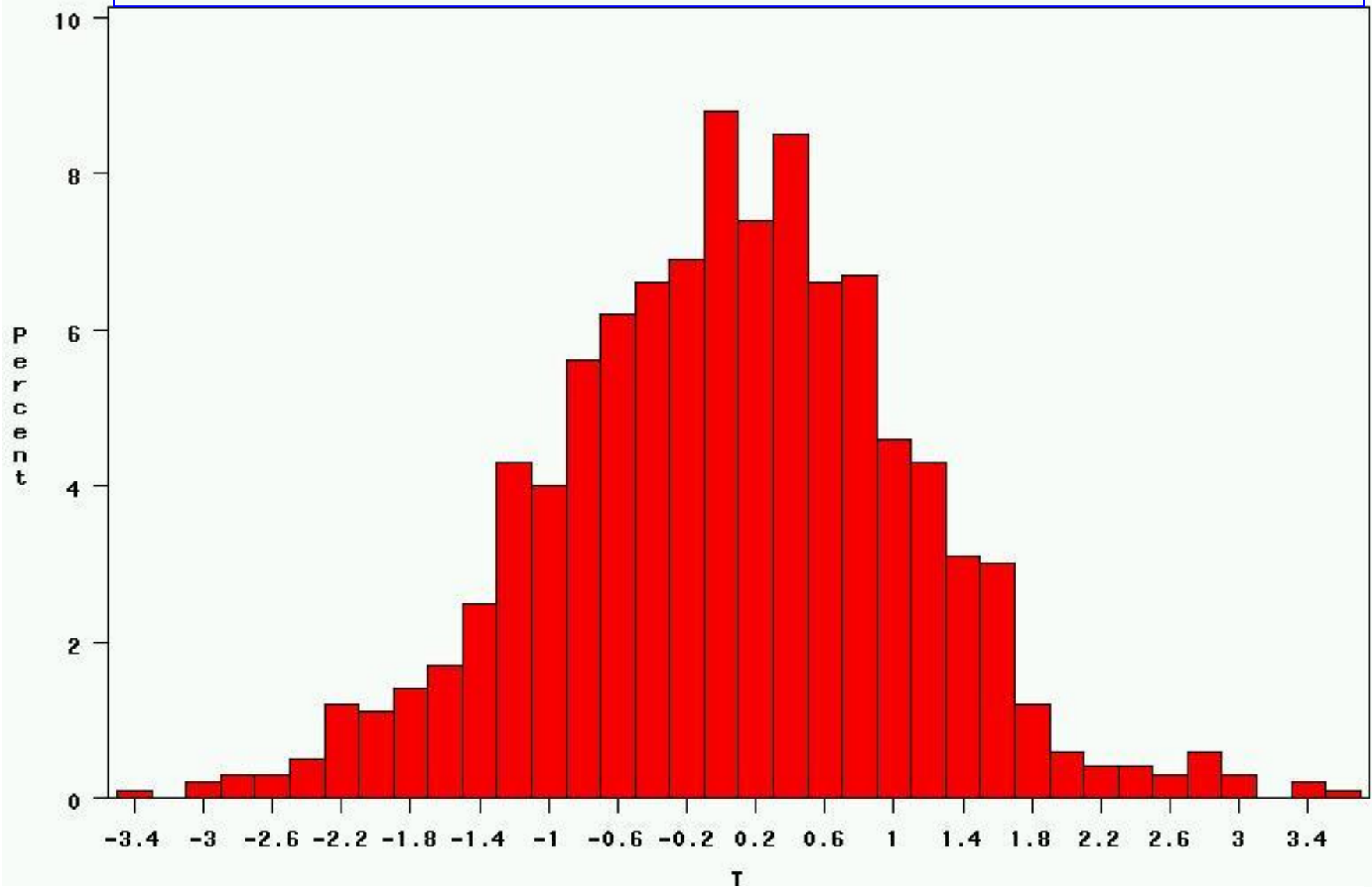
T-distribution with only 1 degree of freedom.



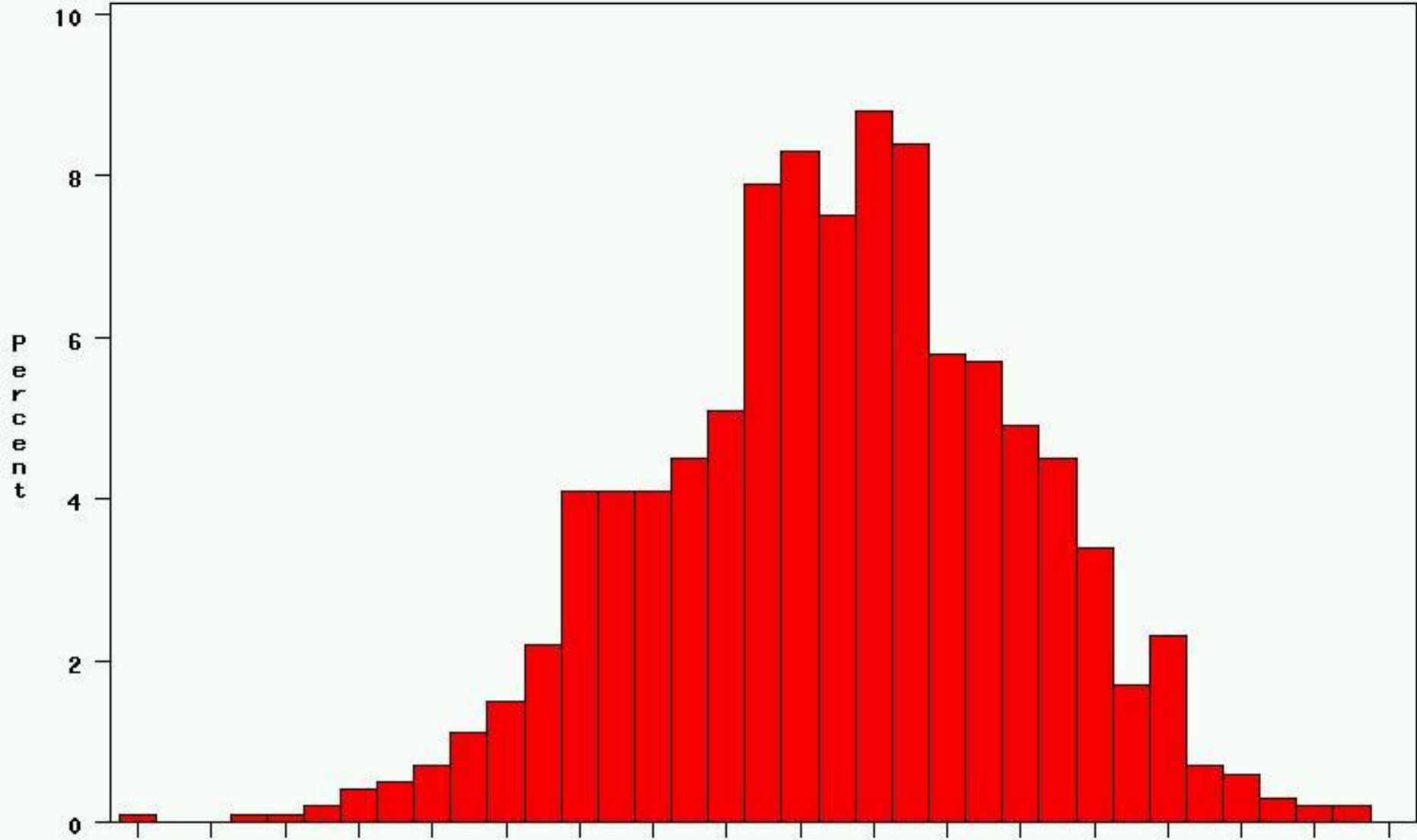
T-distribution with 4 degrees of freedom.



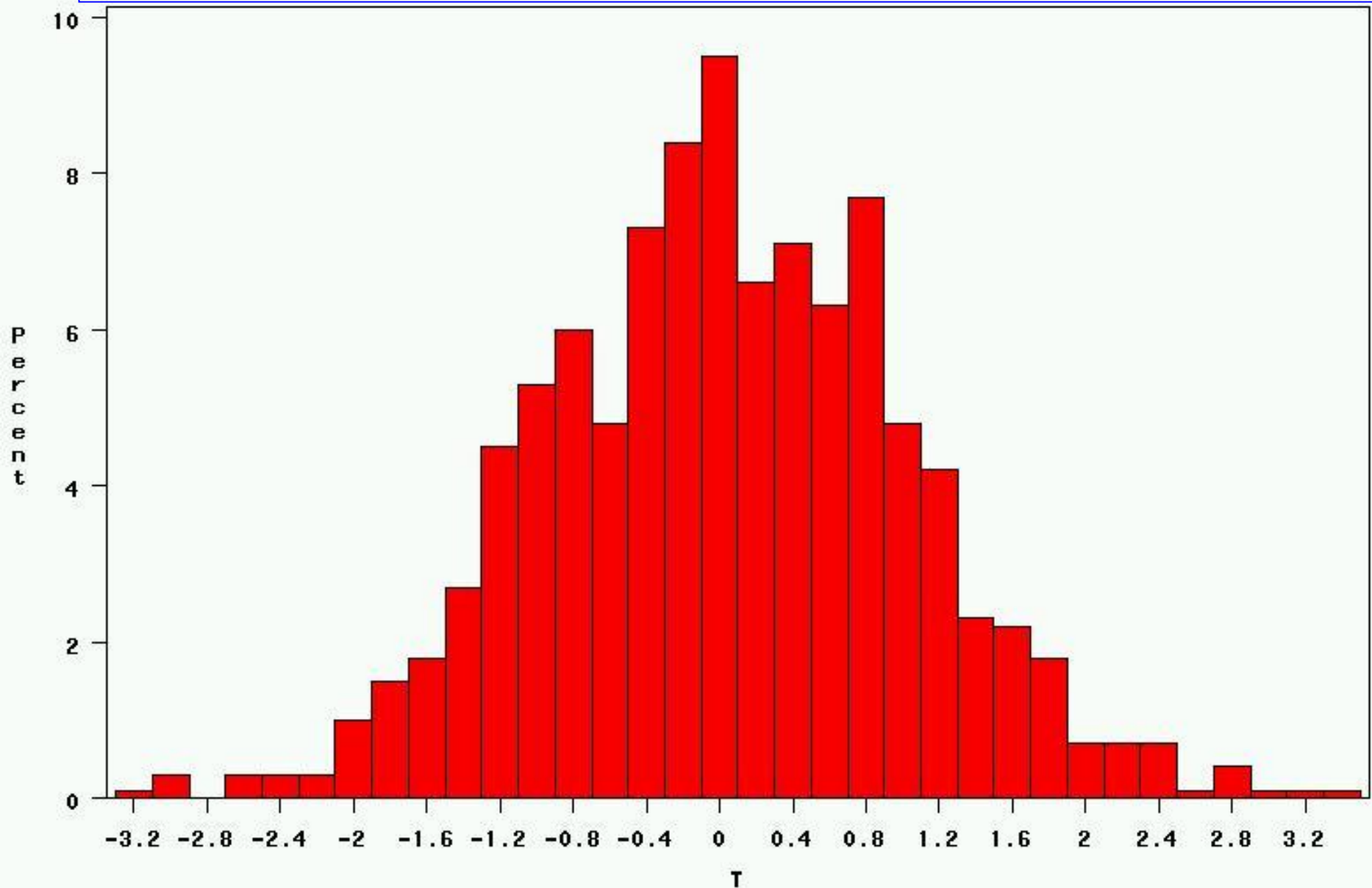
T-distribution with 9 degrees of freedom.



T-distribution with 29 degrees of freedom

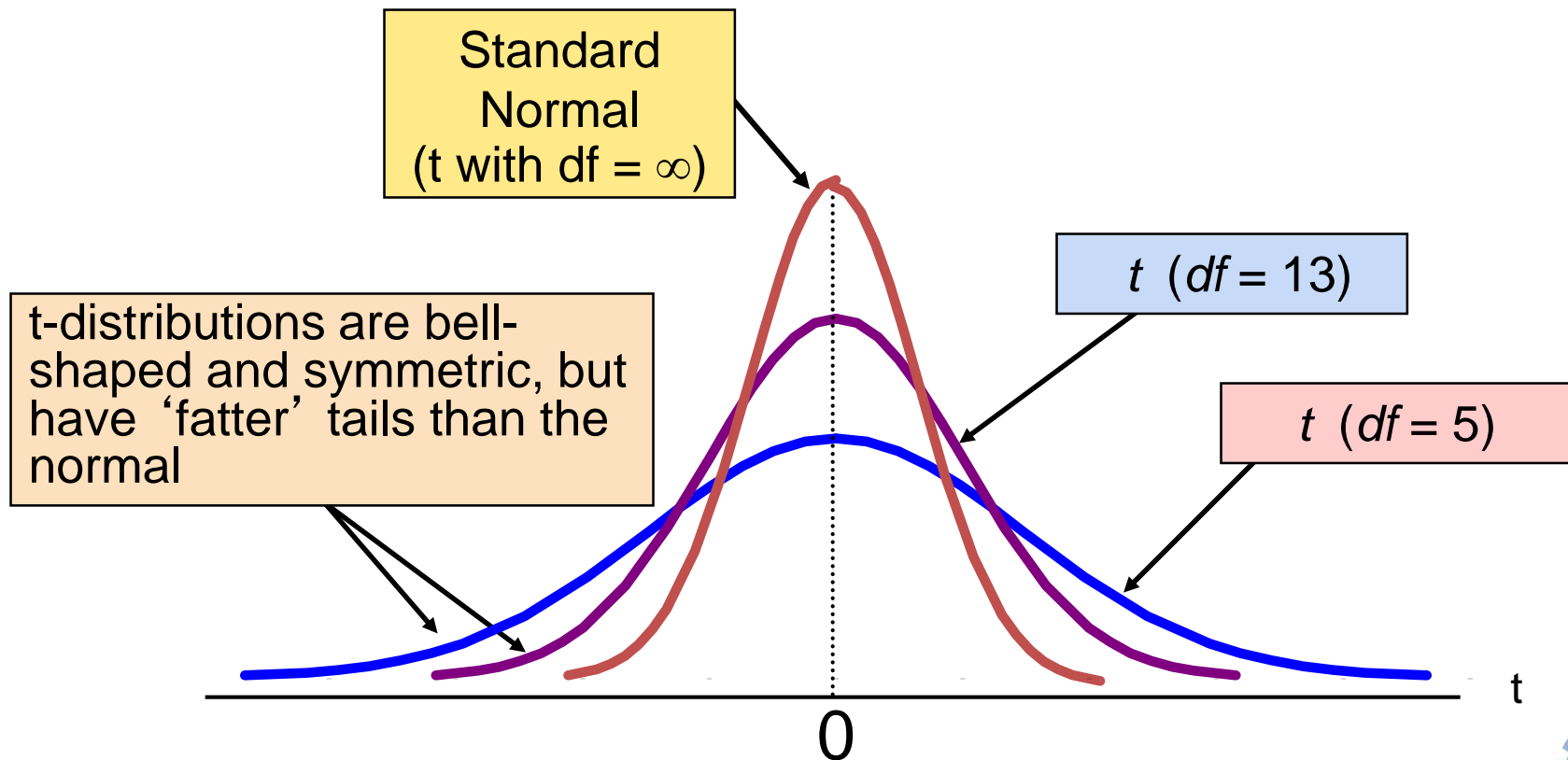


T-distribution with 99 degrees of freedom. Looks a lot like Z!!



Student's t Distribution

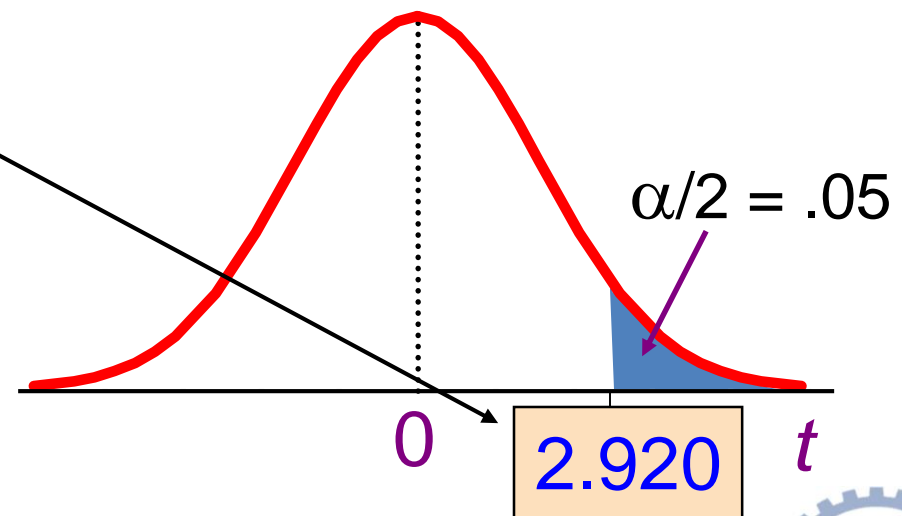
Note: $t \rightarrow Z$ as n increases



Student's t Table

Upper Tail Area			
df	.25	.10	.05
1	1.000	3.078	6.314
2	0.817	1.886	2.920
3	0.765	1.638	2.353

Let: $n = 3$
 $df = n - 1 = 2$
 $\alpha = .10$
 $\alpha/2 = .05$



The body of the table contains t values, not probabilities



T distribution Values

With comparison to the Z value

<u>Confidence Level</u>	<u>t (10 d.f.)</u>	<u>t (20 d.f.)</u>	<u>t (30 d.f.)</u>	<u>Z</u>
.80	1.372	1.325	1.310	1.28
.90	1.812	1.725	1.697	1.64
.95	2.228	2.086	2.042	1.96
.99	3.169	2.845	2.750	2.58

Note: $t \rightarrow Z$ as n increases



Main Considerations in Hypothesis Testing:

- Sample size
 - **Use Z for large samples, t for small (<100)**
- There are two other choices to be made:
 - **One-tailed or two-tailed test**
 - “Is there a difference?” = 2-tailed test
 - “Is the difference less than or greater than?” = 1-tailed test
- Alpha (α) level
 - **.05, .01, or .001? ($\alpha=.05$ is most common)**

