Lecture 27: Chapter 10, Sections 2-3 Inference for Quantitative Variable Hypothesis Test with *t*

- Compare *z* and *t*; *t* Test with Software
- How Large is “Large” *t*?
- *t* Test with Small *n*
- What Leads to Rejecting Ho; Errors, Multiple Tests
- Relating Confidence Interval and Test Results

Looking Back: *Review*

- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-4)
  - Displaying and Summarizing (Lectures 5-12)
  - Probability (discussed in Lectures 13-20)
  - Statistical Inference
    - 1 categorical (discussed in Lectures 21-23)
    - 1 quantitative: z CI, z test, t CI, t test
    - categorical and quantitative
    - 2 categorical
    - 2 quantitative

Standardizing Sample Mean to t *(Review)*

For random sample of size n from population with mean μ, standard deviation σ, sample mean $\bar{X}$ has

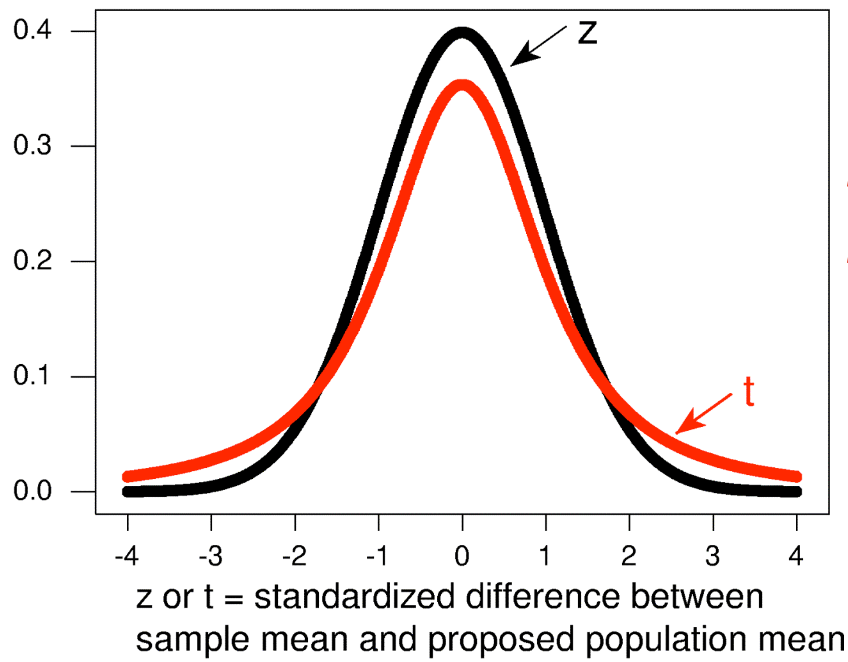
- mean μ
- s.d. $\frac{\sigma}{\sqrt{n}}$ (may have to substitute s for σ)
- shape approximately normal for large enough n

→For σ unknown and n small, $\frac{\bar{x}-\mu}{s/\sqrt{n}} = t$

- t (like z) centered at 0, symmetric, bell-shaped
- t has n-1 df (spread depends on n)

Inference Based on z or t *(Review)*

By Hand	σ known	σ unknown
small sample ($n < 30$)	$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = z$	$\frac{\bar{x}-\mu}{s/\sqrt{n}} = t$
large sample ($n \geq 30$)	$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = z$	$\frac{\bar{x}-\mu}{s/\sqrt{n}} \approx z$*



With software, simply use t if sigma is unknown.

Distribution of t is "heavy tailed" for small n.

Comparing *z* and *t* Distributions

How different are the *z* and *t* distributions?

Unless *n* is very small, distributions are similar; cut-offs for various tail probabilities quite close.

Compared values of *z* and *t* (df=18).

Example: *t Test (with Software)*

- **Background**: Random sample of shoe sizes for 9 college males: 11.5, 12.0, 11.0, 15.0, 11.5, 10.0, 9.0, 10.0, 11.0
- **Question:** Can we believe mean shoe size of all college males is 11?
- **Response:** Use software: enter values, specify proposed mean 11 and "not-equal" alternative.

```
One-Sample T: Shoe
Test of mu = 11 vs mu not = 11
Variable           N       Mean      StDev   SE Mean
Shoe               9     11.222      1.698     0.566
Variable              95.0% CI              T       P
Shoe           (   9.917,  12.527)     0.39   0.705
```

Note: small sample is OK because shoe sizes are normal.

Is *t* large? *P*-value small?

Believe population mean=11?

How Large is “Large” for t Statistic

Excerpts from t table→

- May call values near **2** borderline for df>10
- May call values near **3** borderline for df< 5

Confidence Level

	90%	95%	98%	99%
z (infinite n)	1.645	1.960 or 2	2.326	2.576
t: df = 19 (n = 20)	1.73	2.09	2.54	2.86
t: df = 11 (n = 12)	1.80	2.20	2.72	3.11
t: df = 3 (n = 4)	2.35	3.18	4.54	5.84

Use of *t* with Very Small Samples

Can assume shape of $\bar{X}$ for random samples of **any** size *n* is approximately normal if graph of sample data appears normal.

Normal population → $\dfrac{\bar{x}-\mu}{s/\sqrt{n}}$ is exactly *t*

Example: *t Test with Small n*

- **Background**: Random sample of 4 Math SATs (570, 580, 640, 760) have mean 637.5, s.d. 87.3.
- **Question:** Do they represent population with mean greater than 500? (Use cut-off alpha=0.05.)
- **Response:** *n* is small but *t* procedure is OK because SATs are normal:

```
One-Sample T: MathSAT
Test of mu = 500 vs mu > 500
Variable           N       Mean      StDev    SE Mean
MathSAT            4      637.5       87.3       43.7
Variable       95.0% Lower Bound         T          P
MathSAT                   534.7       3.15      0.026
```

P-value =

Using cutoff 0.05, small enough to reject H_0 ?

Conclude population mean > 500?

Practice: 10.20 p.490

Example: *t Test with Small n, 2-Sided Alternative*

- **Background**: Random sample of 4 Math SATs (570, 580, 640, 760) have mean 637.5, s.d. 87.3.
- **Question:** Do they represent population with mean different from 500? (Use cut-off alpha=0.05.)
- **Response:** Now use ≠ alternative:

```
One-Sample T: MathSAT
Test of mu = 500 vs mu not = 500
Variable          N       Mean      StDev    SE Mean
MathSAT           4      637.5       87.3       43.7
Variable              95.0% CI               T        P
MathSAT         (   498.6,   776.4)       3.15    0.051
```

P-value =

Using cutoff 0.05, small enough to reject H_0?

Conclude population mean ≠ 500?

A Closer Look: t near 3 can be considered borderline for very small n.

One-sided vs. Two-sided Results

- Tested $H_o : \mu = 500$ vs. $H_a : \mu > 500$
  P-value=0.026→rejected H_0
- Tested $H_o : \mu = 500$ vs. $H_a : \mu \neq 500$
  P-value=0.051→ did not reject H_0

Suspecting mean > 500 got us significance

Example: *Concerns about 2-Sided Test*

- **Background**: Random sample of 4 Math SATs (570, 580, 640, 760) have mean 637.5, s.d. 87.3. The *t* test failed to reject $H_0: \mu = 500$ vs. 2-sided H_a because *P*-value=0.051.
- **Question:** Should we believe 500 is a plausible value for the population mean?
- **Response:** Several concerns:
  - If these were students admitted to university, should have used “>” alternative.
  - *n*=4 very small→vulnerable to Type _____ Error
  - MUST we stick to 0.05 as cut-off for small *P*-value?
  - Maybe could have found out σ and done ____ test instead.
  - Does μ=500 seem plausible when smallest value is 570?

Factors That Lead to Rejecting H_0

Statistically significant data produce P-value small enough to reject H_0. t plays a role:

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

Reject H_0 if P-value small; if $|t|$ large; if…

- Sample mean far from μ_o
- Sample size n large
- Standard deviation s small

Factors That Lead to *Not* Rejecting H_0

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

Can't reject H_0 if P-value not small; if $|t|$ not large; if…

- Sample mean close to μ_0
- Sample size n small
- Standard deviation s large

Types I and II Error

- Small *n* can lead to Type II Error (Fail to reject false H_0) *(Sampled only 4 SATs.)*
- Multiple tests can lead to Type I Error (Reject true H_0)…

Example: *Multiple Tests*

- **Background**: Suppose all Verbal SATs have mean 500. Sample *n*=20 scores each in 100 schools, each time test $H_o: \mu = 500$ vs. $H_a: \mu < 500$.
- **Question:** If we reject H_0 in 4 of those schools, can we conclude that mean Verbal SAT in those 4 schools is significantly lower than 500?
- **Response:** If we set 0.05 as cut-off for small *P*-value then long-run probability of committing Type I Error (rejecting true H_0) is

  Even if all 100 schools actually have mean 500, by chance alone some samples will produce a sample mean low enough to reject H_0 % of the time.

Practice: 10.26 p.500

Confidence Interval and Hypothesis Test Results

- Confidence Interval: range of plausible values
- Hypothesis Test: decides if a value is plausible

Informally,

- If μ_o is in confidence interval, don't reject $H_o : \mu = \mu_o$
- If μ_o is outside confidence interval, reject $H_o : \mu = \mu_o$

Example: *Relating Confidence Interval to Test*

- **Background**: Consider these confidence intervals:
  - 95% CI for pop mean earnings (3171, 4381)
  - 95% CI for pop mean shoe size (9.9, 12.5)
  - 95% CI for pop mean Math SAT (498.6, 776.4)
- **Question:** What to conclude about hypotheses...?
  - $H_o : \mu = 5000$ vs. $H_a : \mu < 5000$
  - $H_o : \mu = 11$ vs. $H_a : \mu \neq 11$
  - $H_o : \mu = 500$ vs. $H_a : \mu \neq 500$
- **Response:** Check if proposed mean is in interval:
  - Reject H_0 ?
  - Reject H_0 ?
  - Reject H_0 ?

Practice: 10.27c,e p.500

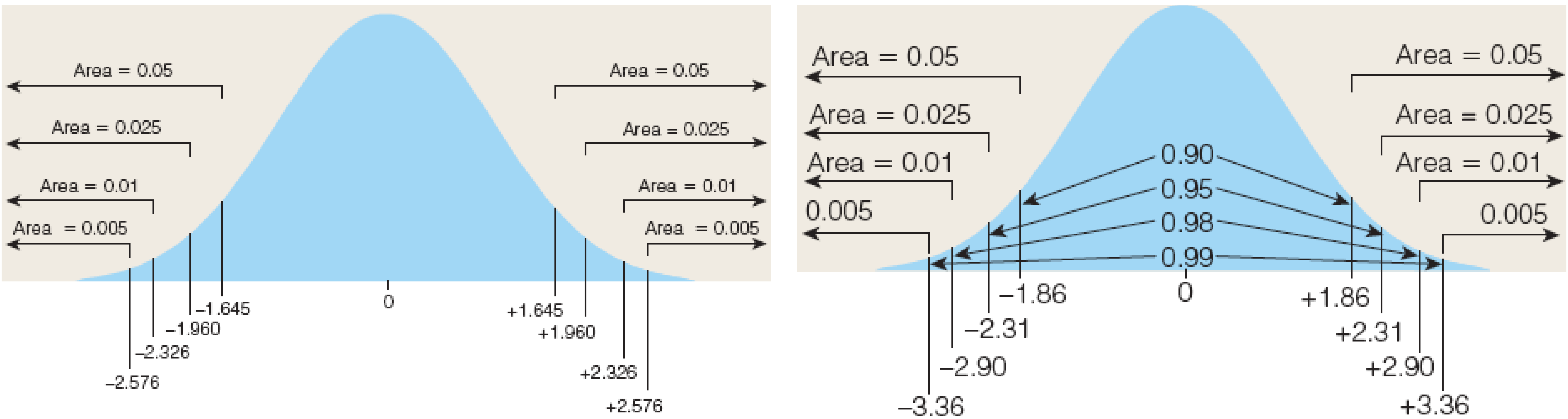
Examples: *Reviewing z and t Tests (#1-#4)*

- **Background**: Sample mean and standard deviation of amount students spent on textbooks in a semester is being used to test if the mean for all students exceeds \$500. The null hypothesis will be rejected if the *P*-value is less than 0.01.) We want to draw conclusions about mean credits taken by all students at a particular college.

Looking Back: *If the sample is biased, or n is too small to guarantee $\bar{X}$ to be approximately normal, neither z nor t is appropriate. Otherwise, use **z** if population standard deviation is known or n is large. Use **t** if population standard deviation is unknown and n is small.*

Example: *Reviewing z and t Tests (#1)*

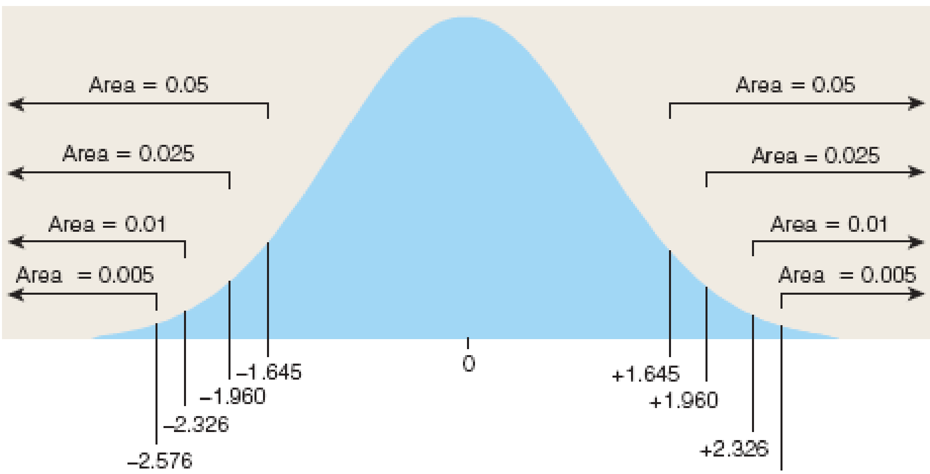
- **Background**: Sample mean and s.d. of textbook costs are used to test if $\mu > 500$ (reject H_0 if P-value < 0.01). Refer to *z* (on left) or *t* for 8 df (on right) or neither.

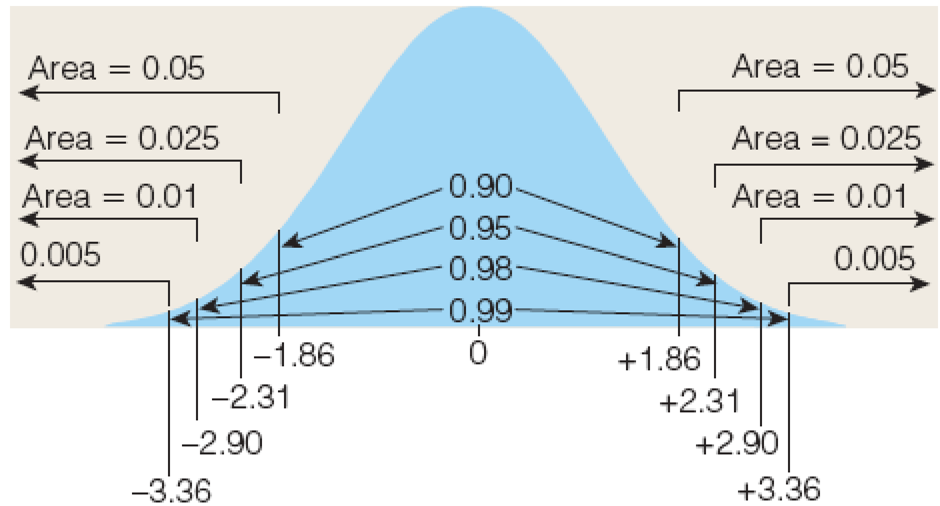


- **Question:** What do we conclude if a **representative** sample of **9** students have t=+2.5? There is an **outlier** in the data set.
- **Response:**

Example: *Reviewing z and t Tests (#2)*

- **Background**: Sample mean and s.d. of textbook costs are used to test if $\mu > 500$ (reject H_0 if P-value < 0.01). Refer to z (on left) or t for 8 df (on right) or neither.

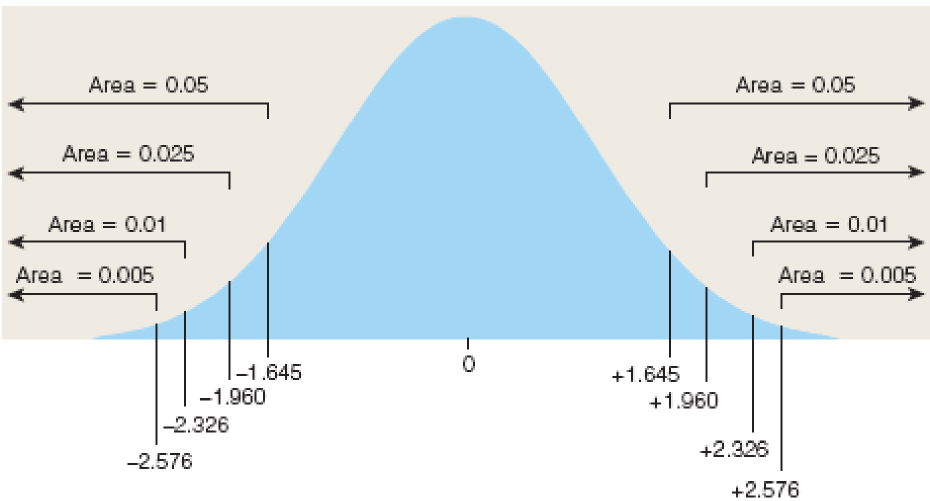


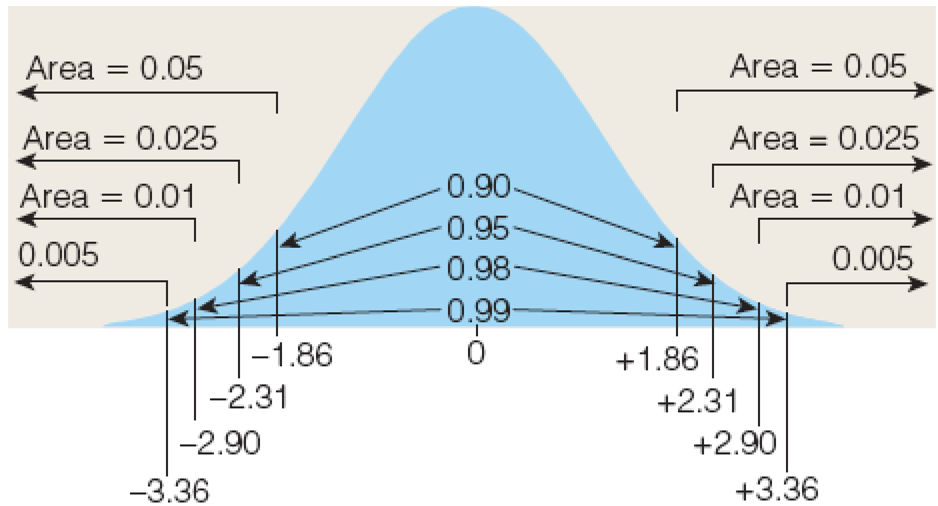


- **Question:** What do we conclude if a **representative** sample of **9** students have t=+2.5? The data set appears **normal**.
- **Response:**

Example: *Reviewing z and t Tests (#3)*

- **Background**: Sample mean and s.d. of textbook costs are used to test if $\mu > 500$ (reject H_0 if P-value < 0.01). Refer to *z* (on left) or *t* for 8 df (on right) or neither.

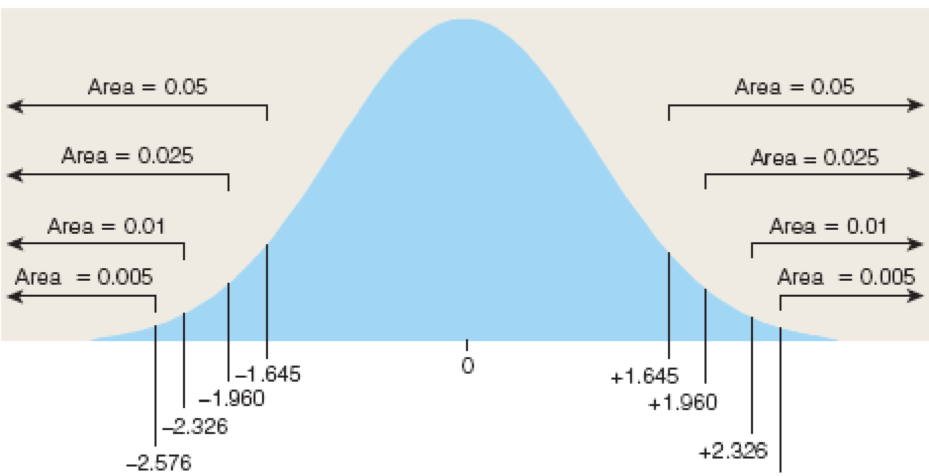


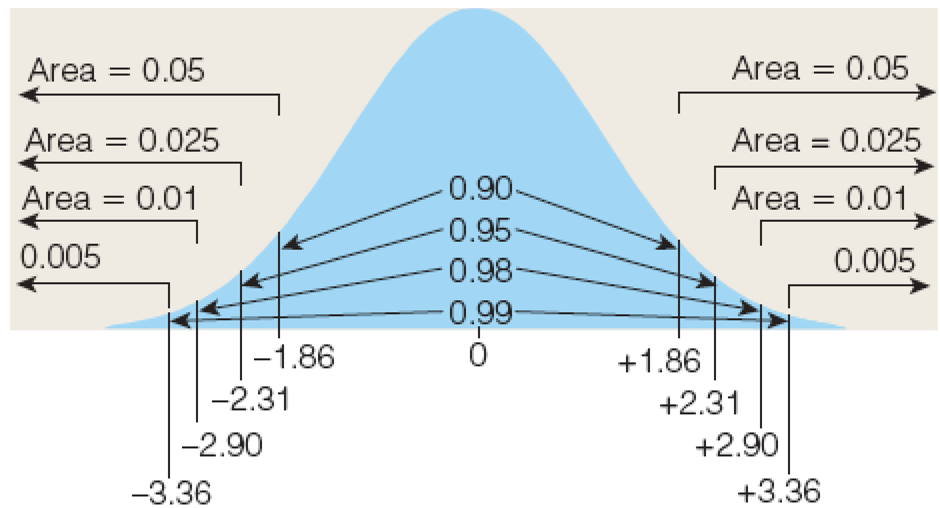


- **Question:** What do we conclude if a **representative** sample of **90** students have *t*=+2.5? There is an outlier in the data set.
- **Response:**

Example: *Reviewing z and t Tests (#4)*

- **Background**: Sample mean and s.d. of textbook costs are used to test if $\mu > 500$ (reject H_0 if P-value < 0.01). Refer to z (on left) or t for 8 df (on right) or neither.





- **Question:** What do we conclude if a sample of **90 biology majors** have t=+2.5? The data set appears **normal**.
- **Response:**

Lecture Summary
(Inference for Means: t Hypothesis Test)

- Comparing *z* and *t* distributions
- *t* test with software
- How large is "large" *t*?
- *t* test with small *n* (one-sided or two-sided alternative)
- Factors that lead to rejecting null hypothesis
- Type I or II Error; multiple tests
- Relating confidence interval and test results
- Examples for review