

**Stat 213: Intro to Statistics 10**  
**Testing**

H. Kim

Fall 2007

## statistical inference

- Hypothesis testing
  - To assess the evidence provided by the data in favor of some claim about the population parameter
  - Making a decision about the value of parameter based on some preconceived idea about what its value might be

step 1: the null hypothesis,  $H_0$

- The null hypothesis, denoted by  $H_0$ 
  - the statement being tested and we want to find evidence against it
  - the null hypothesis is a statement of **no effect** or **no difference**
  - eg.  $H_0 : \mu = \mu_0$

step 2: the alternative hypothesis,  $H_a$

- The alternative hypothesis, denoted by  $H_a$ 
  - the statement that we hope/subject instead of  $H_0$
  - **one-tailed** alternative: eg.  $H_a : \mu > \mu_0$  or  $H_a : \mu < \mu_0$
  - **two-tailed** alternative: eg.  $H_a : \mu \neq \mu_0$

eg. The mean score of the first year students at some school on their stat exam is 20. This year a special preparation course is offered and all the 49 students planning to take the exam enrolled in this course. The mean of their score is 22.2 and standard deviation is 5. Is this an indication that the mean score of those who take the course is higher than 20 ?

–  $H_0 : \mu = 20$

$H_a : \mu > 20$

### step 3: a test statistic

- The decision to reject or accept  $H_0$  is based on a test statistic which is computed from the sample:

– if  $H_0 : \mu = \mu_0$ ,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

– if  $H_0 : p = p_0$ ,

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0, 1)$$

eg. For  $H_0 : \mu = 20$  v.s.  $H_a : \mu > 20$

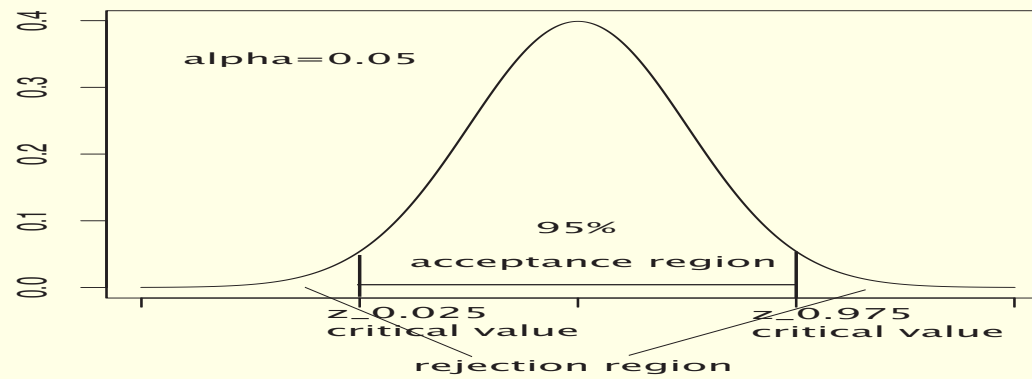
– test statistic :  $Z_{cal} = \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}} = \frac{22.2 - 20}{5/\sqrt{49}} = 3.080$

### step 4: decision: $\alpha$

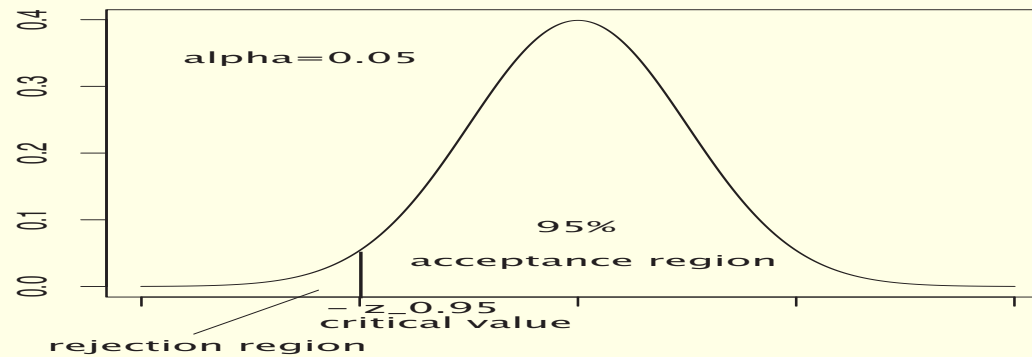
- The **significance level** ( $\alpha$ ) of a test is preassigned:  
 $\alpha = 0.01, 0.05, 0.1$
- Rejection region
  - the set of values of the test statistic for which  $H_0$  is rejected
- Acceptance region
  - the set of values of the test statistic for which  $H_0$  is accepted

for two-tail test

$$H_a : \mu \neq \mu_0$$

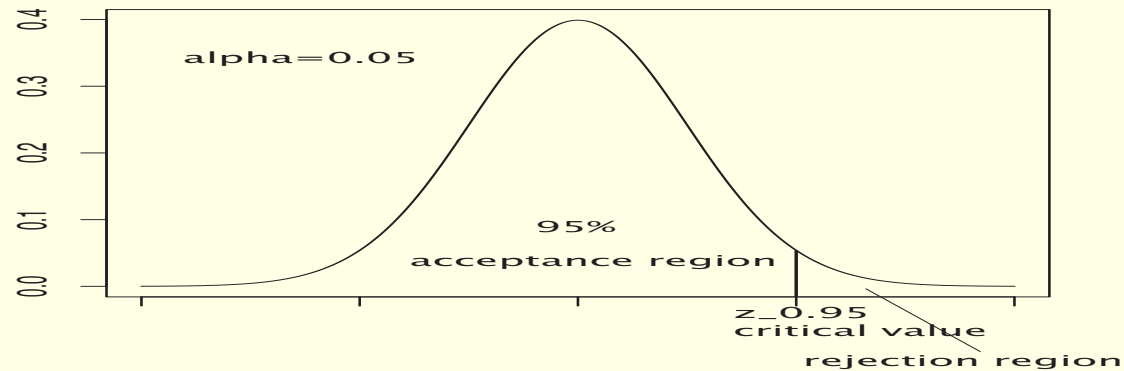


for one-tail test:  $H_a : \mu < \mu_0$



for one-tail test

$$H_a : \mu > \mu_0$$



eg. For  $H_0 : \mu = 20$  v.s.  $H_a : \mu > 20$

- test statistic :  $Z_{cal} = \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}} = \frac{22.2 - 20}{5/\sqrt{49}} = 3.080$
- $\alpha = 0.05$  and  $z_{0.95} = 1.645$
- the calculated  $z$ -value is in rejection region since  $Z_{cal} = 3.080 > z_{0.95} = 1.645$



#### step 4: decision: $p$ -value

- The  $p$ -value measures the **strength of the evidence against  $H_0$**  and it is the probability that the observed statistic value could occur if the null hypothesis  $H_0$  were correct
- The **data is rare** when the null hypothesis  $H_0$  is true ?
  - $p$ -value is high  $\approx$  the data is not rare when  $H_0$  is true  $\implies$  fail to reject  $H_0$
  - $p$ -value is low  $\approx$  the data are rare when  $H_0$  is true  $\implies$  reject  $H_0$

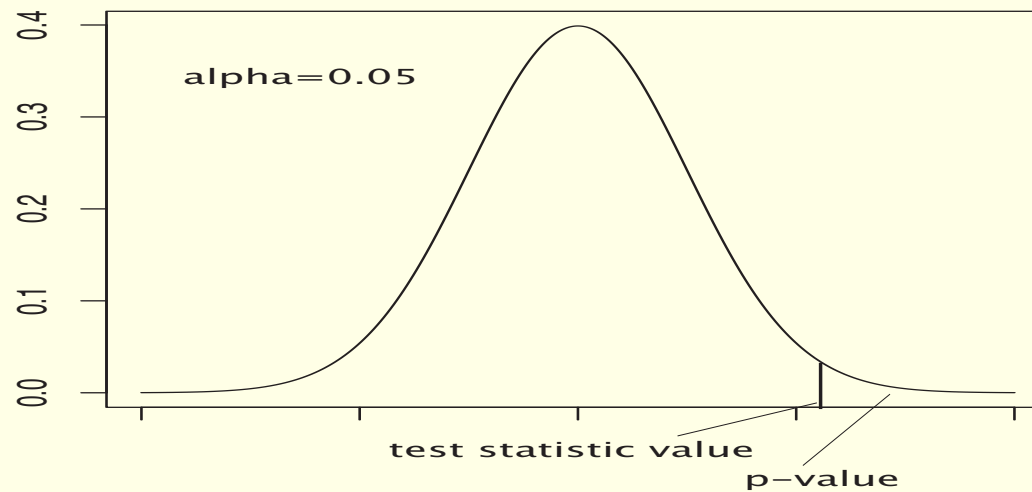
### step 4: decision: $p$ -value

- $p$ -value =  $P$ (observing a test statistic at least as extreme as the one computed |  $H_0$ : true)
  - for one-tail test:  $p$ -value =  $P(Z > Z_{cal} | H_0: \text{true})$  or  $P(Z < Z_{cal} | H_0: \text{true}) = P(Z > Z_{cal} | H_0: \text{true})$
  - for two-tail test:  $p$ -value =  $2P(Z > Z_{cal} | H_0: \text{true})$

eg.  $p$ -value =  $P(Z > 3.080 | H_0 : \text{true})$

$$= P(Z > 3.08) = 1 - P(Z < 3.08) = 0.001$$

decision:  $p$ -value



- If the  $p$ -value  $<$  the preassigned significance level  $\alpha$ , reject  $H_0$  and conclude that the results are statistically significant at level  $\alpha$ .

eg.  $p$ -value = 0.001  $<$   $\alpha$  = 0.05

### step 5: conclusion

- For the example, reject  $H_0$ , conclude that the mean score of those who take the course is higher than 20.
- Do not reject  $H_0$ , there is no sufficient evidence to declare a difference in the average ...

### summary: test

- Construct hypotheses:  $H_0$  and  $H_a$  (one-tailed or two-tailed test).
- Decide the preassigned  $\alpha$  value.
- Find a test statistic, and its sampling distribution (CLT).
- Find rejection regions (for one-tailed or two-tailed test) or calculate  $p$ -value.
- Make decisions (comparing the statistic value based on the sample or  $p$ -value with the critical value  $\alpha$  )
- Conclusion

### example

- Assume the data are normally distributed with  $\sigma = 6$ ,  $\bar{X} = 22.2$ , and  $n = 43$ . Test  $H_0 : \mu = 20$  v.s.  $H_a : \mu > 20$ .

### example

- Body temperatures of 64 intertidal crabs placed in air at  $24.5^{\circ}\text{C}$  yielded a mean of  $24.71^{\circ}\text{C}$  with standard deviation of  $1.34^{\circ}\text{C}$ . Test at  $\alpha = 0.05$  if the mean body temperature for this species of crab differs from the ambient air temperature. And find the  $p$ -value.

### example

- According to a recent report, the proportion of remarried couples who divorce within three years of remarriage is 30%. A random sample of 150 remarried couples contained 61 couples who had divorced within 3 years. At 5% significant level, test the hypothesis that  $p$  is different from 30%.



## decision error

- Decision error (type I error and type II error )
  - type I error
    - \* the risk you are **willing to take of making an incorrect** decision
    - \* this occurs if you reject  $H_0$  when in fact it is true
  - type II error
    - \* this occurs if you accept  $H_0$  when in fact it is false
- The **power** of the test a particular point of the alternative
  - correctly reject  $H_0$  when  $H_0$  is false

decision box

	$H_0$ is false	$H_0$ is true
reject $H_0$	correct decision <b>power</b>	incorrect decision type I error ( $\alpha$ ) <small>(significance level, test size)</small>
fail to reject $H_0$	incorrect decision type II error ( $\beta$ )	correct decision

- The goodness of a statistical test of hypothesis is measured by  $\alpha$  and  $\beta$ .
- We would like  $\alpha$  to be small and power  $(1 - \beta)$  to be large.

## example

- $H_0$ : the student has learned only 60% of the material
  - type I error: conclude the means are different when the means are actually no different passing a student who in fact learned less than 60% of the material
  - type II error: conclude the means are same when the means are actually different failing a student who knew enough to pass

## reducing errors

- If we reduce type I error, we automatically must increase type II error, but there is a way to reduce both: make both curves (under  $H_0$  and  $H_a$ ) narrower, then the power of the test will increase.
- Reduce the standard deviation by increasing the **sample size**