6: Introduction to Hypothesis

The general idea of statistical inference is to use sample data to draw conclusions about the entire population.

Recall: Two Types of Inference

p. 5.1

- Estimation (prior chapter)
  - Confidence intervals to *estimate* population parameters
- Hypothesis testing (this chapter)
  - Test statistics to *judge* claim about population

The General Idea

p. 6.1

- Research question → hypothesis ("claim") → collect data → test claim
- Two Philosophies of Statistical Testing
  - Evidence-weighing ("significance testing")
  - Decision-making ("fixed-level testing")
- This chapter focuses on decision-making approach
  - Future chapters move toward evidence-weighing

Procedure Decision-Making

p. 6.1 - 6.2

(A) Convert research question to statistical hypotheses

Null (H_0): no difference in *populations*

Alternative (H_1): difference in *populations*

Assume H_0 is true until proved otherwise

(B) Set error threshold

(C) Calculate test statistic

(D)

Convert test statistic to probability

Reject or retain H_0

Potential Consequences of Decision

p. 6.3

Decision	H_0 true	H_0 false
Retain H_0	OK retention	Type II error
Reject H_0	Type I error	OK rejection

- Errors
  - Type I error ≡ false rejection of H_0
  - Type II error ≡ false retention of H_0
- Error probabilities
  - α ≡ Probability(Type I error)
  - β ≡ Probability(Type II error)
- Anti-error probabilities
  - $1-\alpha$ ≡ "confidence"
  - $1-\beta$ ≡ "power"

Z test, One-Sided Alternative

p. 6.4

- σ must be known to use *z* test
  - Otherwise, use *t* test
- Tests can be one-sided or two-sided
  - We start with one-sided tests
  - Which are simplier
  - But . . .
  - are less common in practice

One-Sample, *Z* Test

p. 6.4

(A)

- H_0: $\mu = \mu_0$, where μ_0 represents parameter if H_0 true ("null value")
- Alternative either

  H_1: $\mu < \mu_0$ (looking for values to the left of μ_0)

  H_1: $\mu > \mu_0$ (looking for values to the right of μ_0)

(B) Set α (Commonly .10, .05, .01)

- α set by researcher (*not* data)

(C) Test statistic

(D) Conclusion: *p* value and decision

Illustrative Example (IQ Scores)

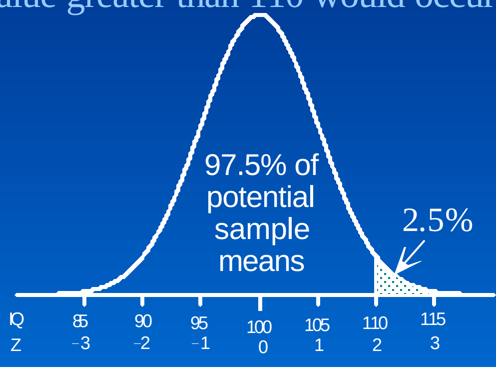
p. 6.6

- Welscher IQ scores are normally distributed with $\mu = 100$ and $\sigma = 15$
- Do children from a particular school (population) have *higher* than average IQ scores?
- Sample:
  - $n = 9$
  - sample mean = 112.8
  - SEM = 15 / $\sqrt{9}$ = 5

Illustrative Example: IFF H_0 *True*

(p. 6.6)

- If H_0 true, then
  - $\mu = 100$
  - Sample means would vary from sample-to-sample
  - SDM would be normal
- ∴ SDM would be normal, centered on 100 with SEM = 5
  - e.g., 97.5% of sample means will less than 110
  - Seeing a value greater than 110 would occur only 2.5%



Illustrative Example (cont.)

p. 6.6

(A) H_0: $\mu = 100$ vs. H_1: $\mu > 100$

(B) Let $\alpha = .025$ (or whatever)

(C) Test statistic: Formula 6.1

$$z_{\text{stat}} = \frac{\bar{x} - m_0}{SEM}$$

where:
"xbar" ≡ sample mean
μ_0 ≡ "null value"
SEM = $\sigma / \sqrt{n}$

$$z_{stat} = \frac{112.8 - 100}{5} = 2.56$$

Focus on z_{stat}

$$z_{stat} = \frac{\bar{x} - m_0}{SEM}$$

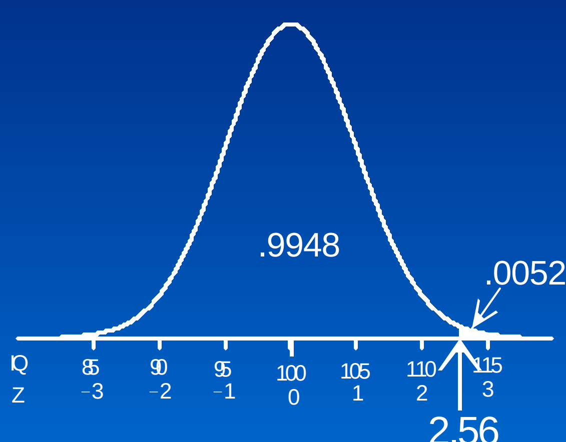
- Components of z_{stat}
  - Numerator = observed mean − expected mean
  - Denominator = standard error of mean
- z_{stat} = "standard deviations" above expectation if H_0 were true
- Convert z_{stat} to "*p* value"

Converting z_{stat} to Probability

p. 6.6

p value = probability test statistic is more extreme than observed *assuming* $_0$ true = area under curve beyond z_{stat}

Look-up probability with help of *z* table



Decision

p. 6.6

- If $p \leq \alpha \rightarrow$ reject H_0
- If $p > \alpha \rightarrow$ retain H_0
- Illustratrative example (IQ scores)
  - $\alpha = .025$ (set by researcher)
  - $p = .0052$ (calculated from data)
  - Since $p < \alpha \rightarrow$ reject H_0
  - $\therefore$ difference is "significant"

One-sample *Z* test: two-sided H_1

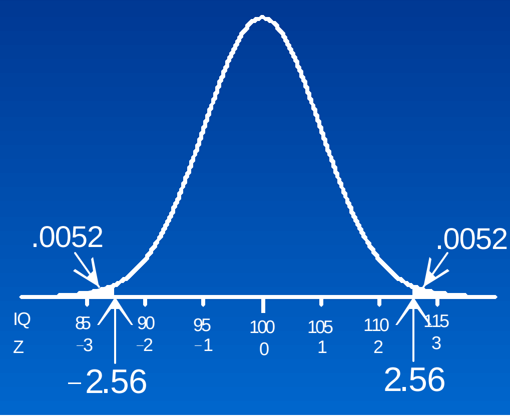
p. 6.4

- Similar to one-sided test but without prior specification of direction of difference
- More common than one-sided test
- Same four steps
  - A: H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$
  - B: Let α = something
  - C: Test stat = Formula 6.1
  - D: Two-sided p value = 2 × one-sided p value

Illustrative Example (IQ scores)

p. 6.7

- "Do IQs *differ*" (without specifying direction)
- Procedure
  - A: H_0: μ = 100 vs. $\boldsymbol{H_1}$: μ ≠ **100**
  - B: Let α = .025 (or whatever)
  - C: z_{stat} = 2.56 (same as before)
  - D: Two-sided $p = 2 \times .0052 = .0104$; reject H_0 since $p < \alpha$



One sample *t* test

p. 6.9 - 6.10

- Use *t* tests when σ is unknown (calculates *s*)
- Test can be one-sided or two-sided
- Illustrative example (%IDEAL)
  - Variable = % of ideal body weight
    - – e.g., value of 100 represents 100% of ideal body weight
  - Sample: 18 diabetics
  - Question: Is μ different than 100?

One-Sample *t* Test

p. 6.4

(A)

- H_0: $\mu = \mu_0$
- Either
  - H_1: $\mu < \mu_0$ (looking for values to the left of μ_0)
  - H_1: $\mu > \mu_0$ (looking for values to the right of μ_0)
  - H_1: $\mu \neq \mu_0$ (two-sided)

(B) Set α

(C) Test statistic (Formula 6.2) $t_{stat} = \dfrac{\bar{x} - m_0}{sem}$

- with df = $n-1$
- you lose 1 df b/c you have to estimate μ with xbar

(D)

Illustrative example %IDEAL (cont.)

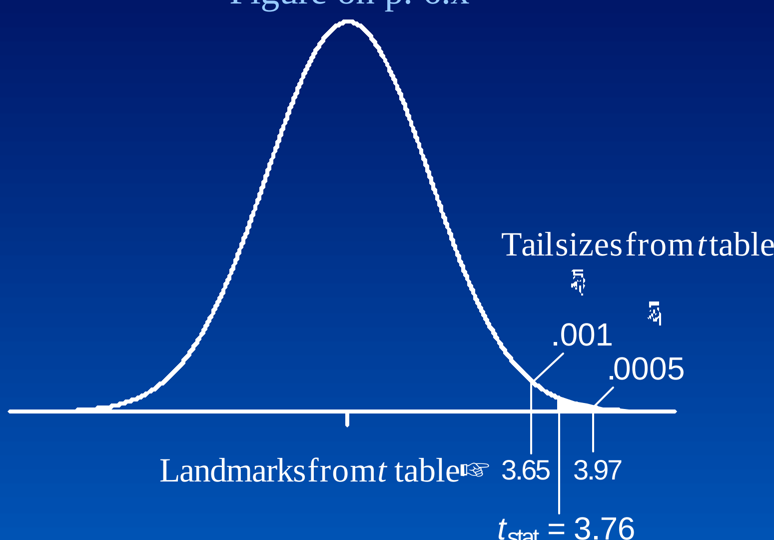
- H_0: $\mu = 100$ *vs.* H_1: $\mu \neq 100$ (two-sided)
- Let $\alpha = .05$ (or whatever)
- Test statistic
  - xbar = 112.778, s = 14.242, n = 18
  - sem = 14.242 / $\sqrt{18}$ = 3.40
  - t_{stat} = (112.778 − 100) / 3.40 = 3.76
  - df = 18 − 1 = 17

Convert t_{stat} to p value

- By computer – use SPSS or StaTable
- By hand – use t table "wedgie technique"
  - Put t_{stat} between two landmarks from t table
  - Determine tail areas of landmarks
  - p value less between these probabilities
  - Double probability if test is two-sided

Illustrative

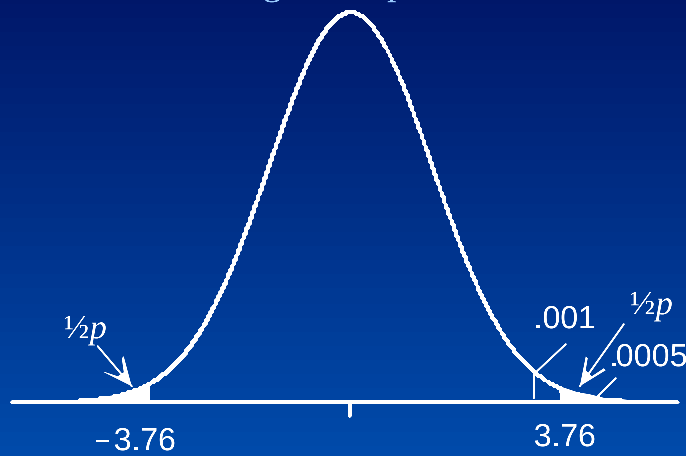
Figure on p. 6.x



∴ One tail less than .001 and more than .0005

Double tail areas for two-sided tests

Figure on p. 6.x



∴ Two tails less than .002 and more than .001

$.001 < p < .002$

Fallacies of Statistical Testing

p. 6.2

- Failure to reject H_0 = accept H_0
  - WRONG!
  - Failure to reject H_0 = insufficient evidence for rejection
- p value = probability H_0 is incorrect
  - WRONG!
  - p value = probability of data assuming H_0 is correct

Fallacies of Statistical Testing (cont.)

p. 6.2

- Statistical significance implies importance
  - WRONG!
  - WRONG!
  - WRONG!
  - Statisical significance does not address causality or size of effect