Hypothesis Testing

PSY 5101: Advanced Statistics for Psychological and Behavioral Research 1

Reminder of Inferential Statistics

- All inferential statistics have the following in common:
  - Use of some descriptive statistic
  - Use of probability
  - Potential for estimation
  - Sampling variability
  - Sampling distributions
  - Use of a theoretical distribution
  - Two hypotheses, two decisions, & two types of error

Hypothesis Testing: Introduction

- This is the last of the seven topics common to all inferential statistics, and so it integrates all of the other six.
- Hypothesis testing uses probability and the sampling distribution of a statistic to make decisions about a parameter
- Hypothesis testing is the process of testing tentative guesses about relationships between variables in populations
- These relationships between variables are evidenced in a statement (a hypothesis) about a population parameter

Hypothesis Testing: Examples

- IQ of deaf children example: Are the deaf children lower in IQ? Or are they average? If μ=100 and σ^2=225, is the $\overline{X}$=88.07 from the sample of N=59 deaf children significantly lower than 100?
- Compute $z_{\overline{X}} = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{N}}} = \frac{88.07 - 100}{\sqrt{\frac{225}{59}}} = \frac{-11.93}{1.95} = -6.11$
- Find $p(\overline{X} < 88.07) = p(z_{\overline{X}} < -6.11) = .00003$. It is unlikely the deaf children came from a population with μ=100 for the mean IQ. So we decide that μ<100 and that the deaf children have lower IQ scores than the general population
  - Remember, this is due to the fact that their language (ASL) is not English, so they score lower on the verbal part of the total IQ test

Hypothesis Testing: Examples

- Rat-shipment example:
  - Are the rats defective? Or are they OK? If μ=33 and σ^2=361, is the $\overline{X}$ =44.4 from the sample of N=25 rats significantly different from 33?
- Compute $z_{\overline{X}} = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{N}}} = \frac{44.4 - 33}{\sqrt{\frac{361}{25}}} = \frac{11.4}{3.8} = 3$
- Find $p(\overline{X} > 44.4) = p(z_{\overline{X}} > 3) = .0013$. It is unlikely the rats came from a population with μ=33 for the mean run time. So, we decide that $\mu \neq 33$ and that the rats are defective

Hypothesis Testing: Key Terms

- Test statistic: a statistic used only for the purpose of testing hypotheses (e.g., $z_{\overline{X}}$)
- Assumptions: conditions placed on a test statistic necessary for its valid use in hypothesis testing
  - For $z_{\overline{X}}$, the assumptions are that the population is normal in shape and that the observations are independent
- Null hypothesis: the hypothesis that we test (H_o)
- Alternative hypothesis: where we put what we believe (H_1)
- Both H_o and H_1 are stated in terms of a parameter

Hypothesis Testing: Key Terms

- Significance level: the standard for what we mean by a "small" probability in hypothesis testing (α)
- Directional and non-directional hypotheses
- One- and two-tailed tests, critical values, and rejection values
- Decision rules:
  - Critical value decision rules
  - p-value decision rules
    - **p-value** is the probability of obtaining a test statistic at least as extreme as the observed value given that the null hypothesis is true

Common Misunderstandings about p-values

- The p-value is not the probability that the null hypothesis is true
- The p-value is not the probability that a finding is merely a fluke
- The p-value is not the probability of falsely rejecting the null hypothesis
- The p-value is not the probability that a replicating study would yield a similar result
- (1 - p-value) is not the probability of the alternative hypothesis being true
- The p-value does not determine α
- The p-value is not an indicator of the size or importance of an effect

H_O and H_1

- Rat-shipment example:
  - We start with H_1. We believe that there is something wrong with the rats or that $\mu \neq 33$
    - So we have $H_1: \mu \neq 33$
  - Next, we state H_o. The null is always the opposite of the alternative. Within H_o and H_1, the set of potential values of the parameter to be tested usually contains all possible numbers. The null hypothesis usually has the "equals" in it
    - So we have $H_o: \mu = 33$
- IQ of deaf children example:
  - Again, we start with H_1. We believe that the deaf children will score lower on the IQ test because English is not their native language or that $\mu < 100$
    - So we have $H_1: \mu < 100$
  - Next, we state H_o
    - So we have $H_o: \mu \geq 100$

Significance Level

- The significance level is the small probability used in hypothesis testing to determine an unusual event that leads you to reject H_o
  - The significance level is symbolized by α (alpha)
  - The value of α is almost always set at α=.05
  - The value of α is chosen before data are collected
  - If H_o is rejected when α=.05, here are examples of what you say:
    - "The mean of the IQ of deaf children, $\overline{X}$ = 88.07, is significantly lower than 100, z = -6.11, p = .00003, one-tailed test"
    - "The mean of the run times, $\overline{X}$ = 44.4, is significantly different from 33, z = 3.00, p = .0013"

Directional and Non-Directional Hypotheses

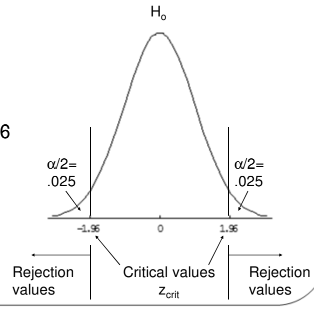
- Directional hypotheses specify a particular direction for values of the parameter
  - IQ of deaf children example:
    - H_o: $\mu \geq 100$
    - H_1: $\mu < 100$
- Non-directional hypotheses do not specify a particular direction for values of the parameter
  - Rat shipment example:
    - H_o: $\mu = 33$
    - H_1: $\mu \neq 33$
- Another example:
  - Suppose you believe that dancers are more introverted than other people. You have N=26 dancers and know that for this age group with your male/female ratio that μ=19.15 for introversion
    - H_o: $\mu \leq 19.15$
    - H_1: $\mu > 19.15$

One- and Two-Tailed Tests, Critical Values, and Rejection Values

- One- and two-tailed tests:
  - A one-tailed test is a statistical test that uses only one tail of the sampling distribution of the test statistic
  - A two-tailed test is a statistical test that uses two tails of the sampling distribution of the test statistic
- Critical values are values of the test statistic that cut off α or $\alpha/2$ in the tail(s) of the theoretical reference distribution
- Rejection values are the values of the test statistic that lead to rejection of H_o

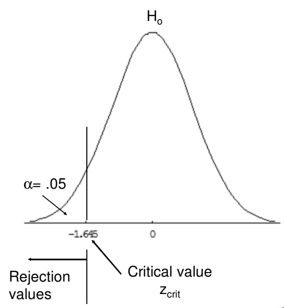
One- and Two-Tailed Tests, Critical Values, and Rejection Values

- Rat shipment example:
  - $H_o: \mu=33$
  - $H_1: \mu\neq33$
- Two-tailed test
- Critical values are $z_{crit}=-1.96$ and $z_{crit}= 1.96$
- Rejection values are $\leq$-1.96 and $\geq$1.96



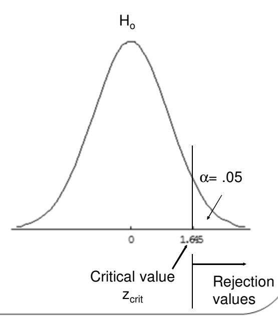
One- and Two-Tailed Tests, Critical Values, and Rejection Values

- IQ of deaf children example:
  - $H_o: \mu\geq100$
  - $H_1: \mu<100$
- One-tailed test
- Critical value is $z_{crit}=-1.645$
- Rejection values are $\leq$-1.645



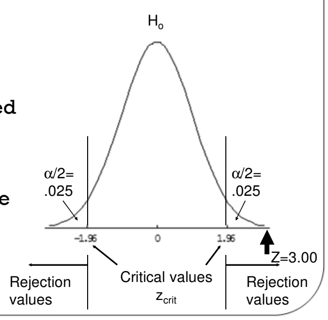
One- and Two-Tailed Tests, Critical Values, and Rejection Values

- Introversion of dancers example:
  - $H_o: \mu\leq19.15$
  - $H_1: \mu>19.15$
- One-tailed test
- Critical value is $z_{crit}=1.645$
- Rejection values are $\geq$1.645



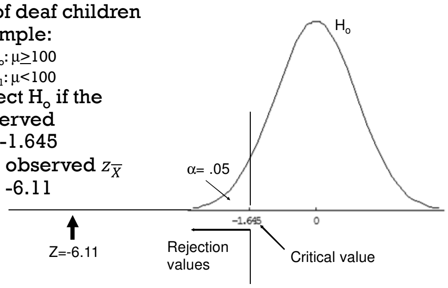
Critical Value Decision Rules

- Rat shipment example:
  - H_o: μ=33
  - H_1: μ≠33
- Reject H_o if the observed $z_{\overline{X}} \leq -1.96$ or if $z_{\overline{X}} \geq 1.96$
- The observed $z_{\overline{X}}$ was 3.00
- Reject H_o: μ=33 because 3.00>1.96 (observed $z_{\overline{X}} > \mathbf{z}_{crit}$)



Critical Value Decision Rules

- IQ of deaf children example:
  - H_o: μ≥100
  - H_1: μ<100
- Reject H_o if the observed $z_{\overline{X}} \leq -1.645$
- The observed $z_{\overline{X}}$ was -6.11



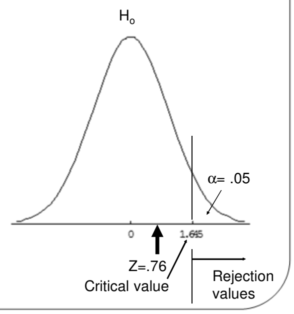
- Reject H_o: μ≥100 because -6.11<-1.645 (observed $z_{\overline{X}} < z_{crit}$)

Compute $z_{\overline{X}}$ for Introversion of Dancers

- Remember, you believe that dancers are more introverted than other people
- You have N=26 dancers and know that for this age group with your male/female ratio that μ=19.15. So you have H_o: μ≤19.15 and H_1: μ>19.15
- Introversion of dancers example: are the dancers higher in introversion? Or are they average? If μ=19.15 and σ^2=18.66, is the $\overline{X}$=19.79 from the sample of N=26 dancers significantly higher than 19.15?
- Compute $z_{\overline{X}} = \frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{N}}} = \frac{19.79 - 19.15}{\sqrt{\frac{18.66}{26}}} = .76$

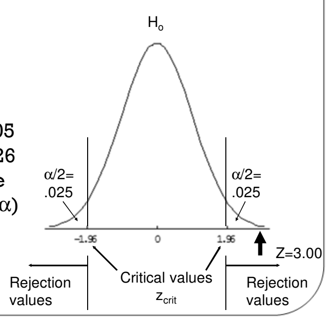
Critical Value Decision Rules

- Introversion of dancers example:
  - $H_o: \mu \leq 19.15$
  - $H_1: \mu > 19.15$
- Reject H_o if the observed $z_{\overline{X}} \geq 1.645$
- The observed $z_{\overline{X}}$ was .76
- Retain Ho: $\mu \leq 19.15$ because .76<1.645 (observed $z_{\overline{X}} < z_{crit}$)



p-Value Decision Rules

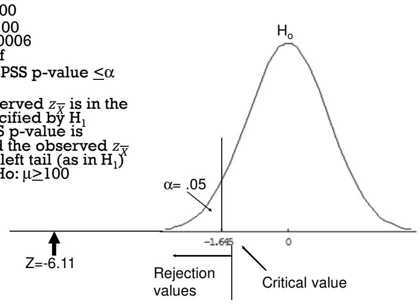
- Rat shipment example:
  - $H_o: \mu = 33$
  - $H_1: \mu \neq 33$
- Reject H_o if the SPSS (2-tailed) p-value is $\leq \alpha = .05$
- The SPSS p-value is .0026
- Reject H_o: $\mu = 33$ because .0026<.05 (p-value is < α)



p-Value Decision Rules

- IQ of deaf children:
  - Ho: $\mu \geq 100$
  - H1: $\mu < 100$

  SPSS p=.00006
- Reject H_o if
  - ½ the SPSS p-value $\leq \alpha$ ***and***
  - the observed $z_{\overline{X}}$ is in the tail specified by H_1
- ½ the SPSS p-value is .00003 and the observed $z_{\overline{X}}$ was in the left tail (as in H_1)
- So, reject Ho: $\mu \geq 100$



p-Value Decision Rules

- Introversion of dancers example:
  - $H_o: \mu \leq 19.15$
  - $H_1: \mu > 19.15$

  SPSS p=.4472
- Reject H_o if
  - ½ the SPSS p-value $\leq \alpha$

    and
  - the observed $z_{\overline{X}}$ is in the tail specified by H_1
- ½ the SPSS p-value is .2236 and the observed $z_{\overline{X}}$ was in the right tail (as in H_1)
- So, retain H_o: $\mu \leq 19.15$