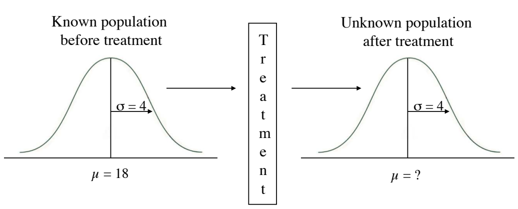
Chapter 8: Introduction to Hypothesis Testing

Hypothesis Testing

- An inferential procedure that uses sample data to evaluate the credibility of a hypothesis about a population.

2



3

Hypothesis Testing

- Step 1: State hypothesis
- Step 2: Set criteria for decision
- Step 3: Collect sample data
- Step 4: Evaluate null hypothesis
- Step 5: Conclusion

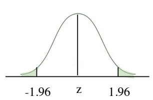
$\mu = 18$ g $\sigma = 4$ g

Step 1: H_o: $\mu_{\text{Weight of rats of alcoholic mothers}} = 18$ g (There is no effect of alcohol on the average birth weight)

H_1: $\mu \neq 18$ g (There is an effect…)

$\propto = 0.05$

Step 2: Set criteria



Critical Region
$z > 1.96$
or
$z < -1.96$

Step 3: n = 25 rats $\overline{X} = 15.5$ g $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{25}} = \frac{4}{5} = 0.8$

$$Z_{Obt} = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{15.5 - 18}{0.8} = -3.125$$

Step 4: Reject H_o because Z_{obt} of -3.125 is in the critical region.

Step 5: Conclusion

		Actual Situation	
		No Effect, H_o True	Effect Exists, H_o False
Experimenter's Decision	Reject H_o	Type I Error	Decision Correct
	Retain H_o	Decision Correct	Type II Error

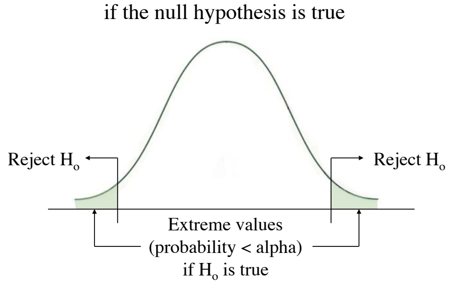
		Actual Situation	
		Did Not Commit Crime	Committed Crime
Jury's Verdict	Guilty	Type I Error	Verdict Correct
	Innocent	Verdict Correct	Type II Error

7

		Actual Situation	
		Coin O.K. (Fair)	Coin Fixed (Cheating)
Your Decision	Coin Fixed (Cheating)	Type I Error	Correct Decision
	Coin O.K. (Fair)	Correct Decision	Type II Error

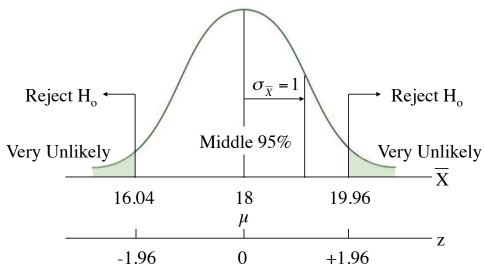
8

The distribution of sample means
(all possible experimental outcomes)
if the null hypothesis is true



Possible, but "very unlikely", outcomes

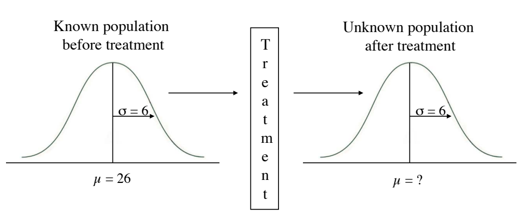
9



10

- We know from national health statistics that the average weight for 2-year olds is μ = 26 lbs. and σ = 6. A researcher is interested in the effect of early handling on body weight/ growth.
- A sample of 16 infants is taken. Each parent is instructed in how to provide additional handling. When the infants are 2-years old, each child is weighed, $\overline{X} = 31$ *pounds*.

11



12

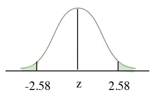
$\mu = 26$ lbs. $\sigma = 4$ lbs.

Step 1: H_o: $\mu_{\text{Handled infants}} = 26$ lbs. (There is no effect of extra handling on average weight of 2 year olds)

H_1: $\mu \neq 26$ lbs. (There is an effect of…)

$\propto = 0.01$

Step 2: Set criteria



Critical Region

$z > 2.58$
or
$z < -2.58$

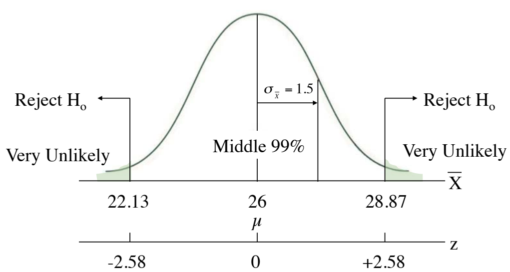
Step 3: n = 16 infants $\overline{X} = 31$ pounds $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{16}} = \frac{6}{4} = 1.5$

$$Z_{Obt} = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{31 - 26}{1.5} = +3.33$$

Step 4: Reject H_o because Z_{obt} of 3.33 is in the critical region.

Step 5: Handling infants more during infancy significantly changes their weight, $z = 3.3$, $p < 0.01$.

13



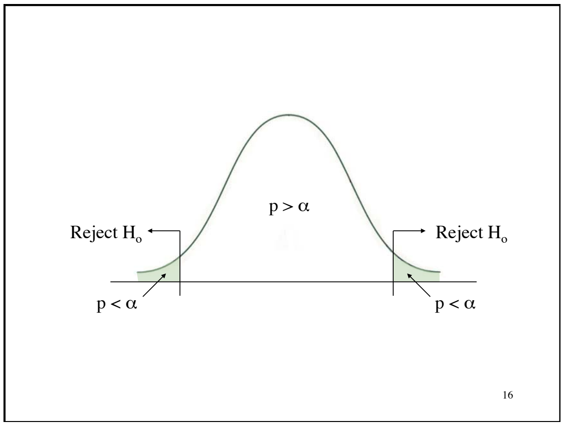
14

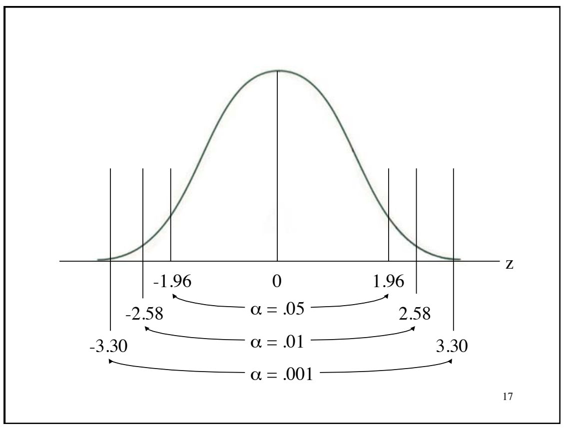
1. State hypothesis and set alpha level
2. Locate critical region
   - e.g. $z > | 1.96 |$
     $z > 1.96$ or $z < -1.96$
3. Obtain sample data and compute test statistic

$$e.g. \quad z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

4. Make a decision about the H_o
5. State the conclusion

15

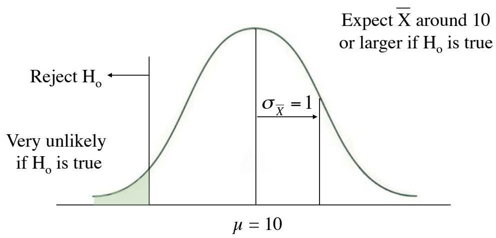




Assumptions for Hypothesis Tests with z-scores:

1. Random sampling
2. Value of σ unchanged by treatment
3. Sampling distribution normal
4. Independent observations

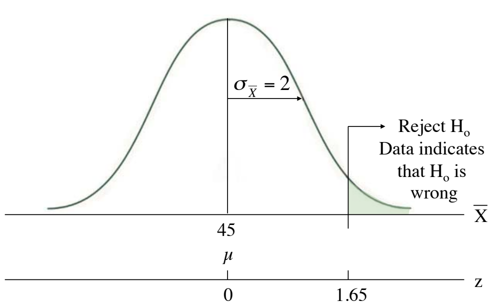
18



19

- A researcher wants to assess the "miraculous" claims of improvement made in a TV ad about a phonetic reading instruction program or package. We know that scores on a standardized reading test for 9-year olds form a normal distribution with $\mu = 45$ and $\sigma = 10$. A random sample of n = 25 8-year olds is given the reading program package for a year. At age 9, the sample is given the standardized reading test.

20



21

Two-tailed vs. One-tailed Tests

1. In general, use two-tailed test
2. Journals generally require two-tailed tests
3. Testing H_o not H_1
4. Downside of one-tailed tests: what if you get a large effect in the unpredicted direction? Must retain the H_o

22

Error and Power

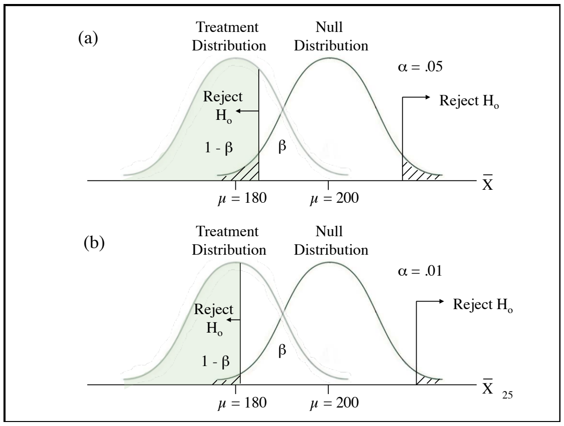
- Type I error = α
  - Probability of a false alarm
- Type II error = β
  - Probability of missing an effect when H_o is really false
- Power = $1 - \beta$
  - Probability of correctly detecting effect when H_o is really false

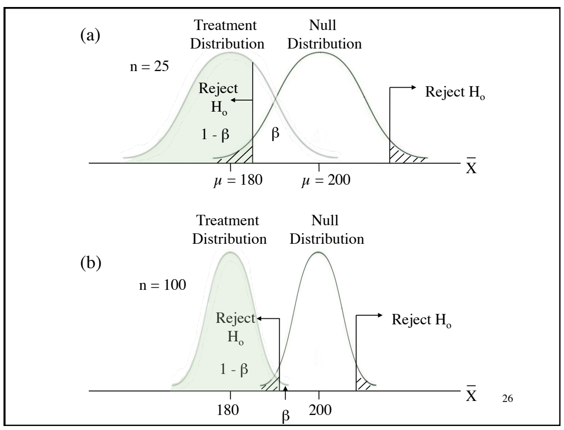
23

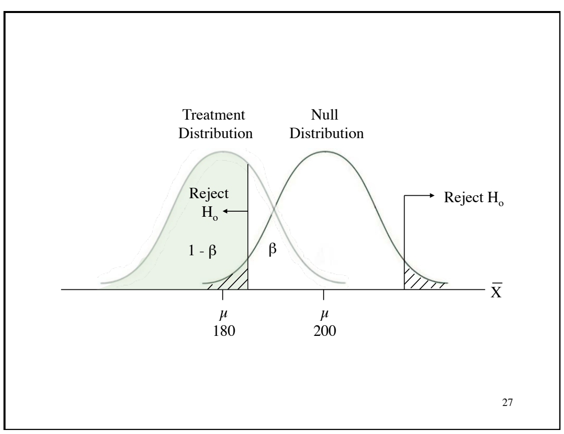
Factors Influencing Power ($1 - \beta$)

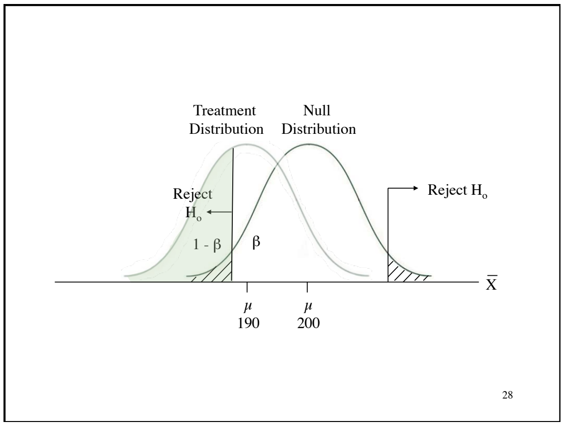
1. Sample size
2. One-tailed versus two-tailed test
3. Criterion (α level)
4. Size of treatment effect
5. Design of study

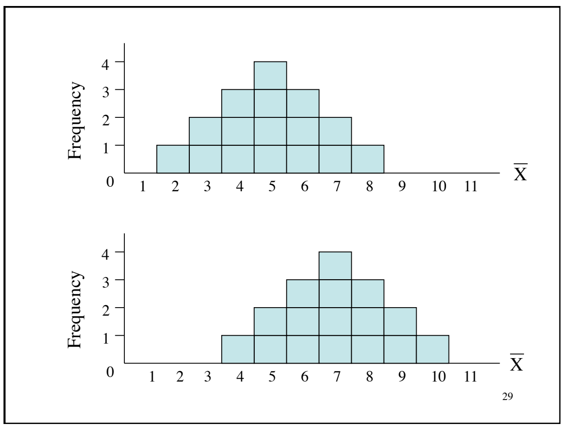
24


(a)
Treatment Distribution
Null Distribution
α = .05
Reject H_o
Reject H_o
1 - β
β
μ = 180
μ = 200
$\overline{X}$
(b)
Treatment Distribution
Null Distribution
α = .01
Reject H_o
Reject H_o
1 - β
β
μ = 180
μ = 200
$\overline{X}$
25


(a)
Treatment Distribution
Null Distribution
n = 25
Reject H_o
Reject H_o
1 - β
β
μ = 180
μ = 200
$\overline{X}$
(b)
Treatment Distribution
Null Distribution
n = 100
Reject H_o
Reject H_o
1 - β
180
β
200
$\overline{X}$
26


Treatment Distribution
Null Distribution
Reject H_o
Reject H_o
1 - β
β
μ 180
μ 200
$\overline{X}$
27





Are birth weights for babies of mothers who smoked during pregnancy significantly different?

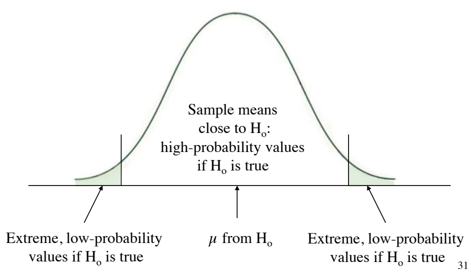
μ = 2.9 kg σ = 2.9 kg

Random Sample: n = 14

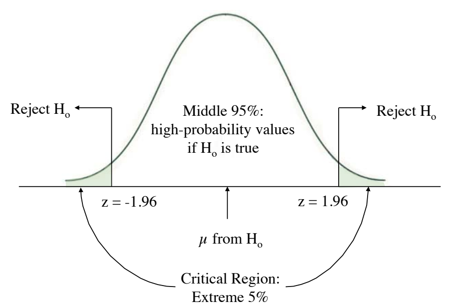
2.3, 2.0, 2.2, 2.8, 3.2, 2.2, 2.5,
2.4, 2.4, 2.1, 2.3, 2.6, 2.0, 2.3

30

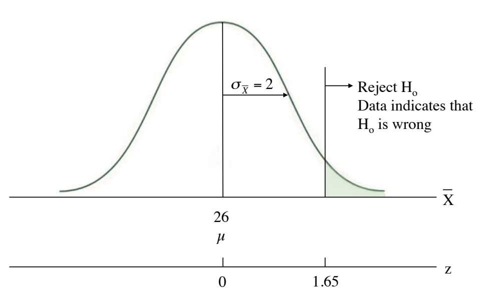
The distribution of sample means if the null hypothesis is true
(all the possible outcomes)



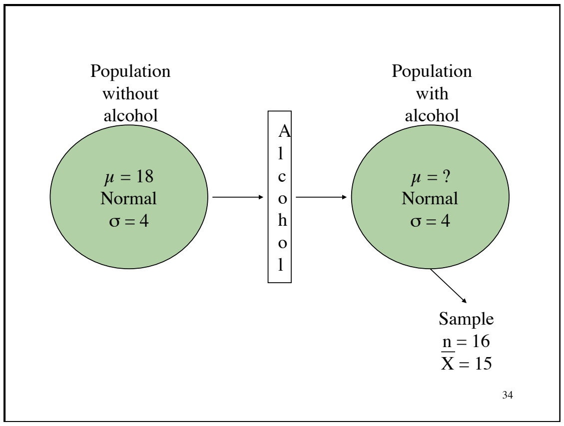
31

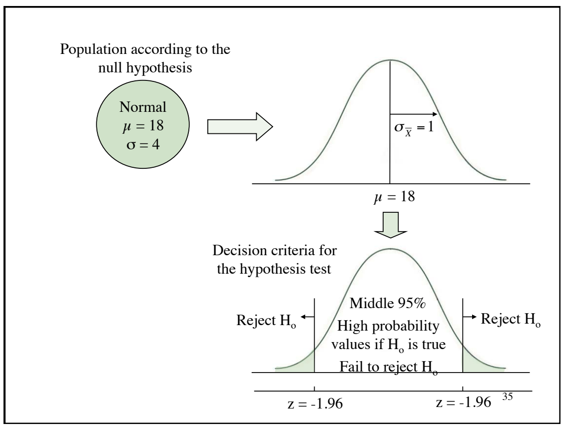


32



33


Population without alcohol
μ = 18
Normal
σ = 4
Alcohol
Population with alcohol
μ = ?
Normal
σ = 4
Sample
n = 16
X = 15
34


Population according to the null hypothesis
Normal
μ = 18
σ = 4
σX̄ = 1
μ = 18
Decision criteria for the hypothesis test
Reject Ho
Middle 95%
High probability values if Ho is true
Fail to reject Ho
Reject Ho
z = -1.96
z = -1.96
35