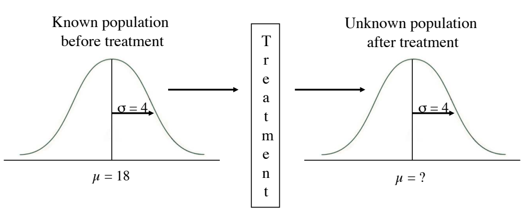
Chapter 8: Introduction to Hypothesis Testing

Hypothesis Testing

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

2



3

Hypothesis Testing

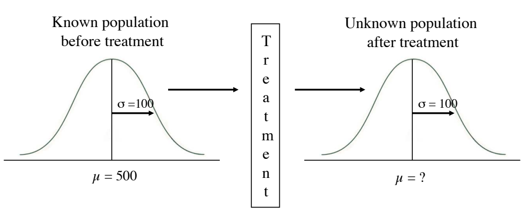
- Step 1: State hypothesis
- Step 2: Set criteria for decision
- Step 3: Collect sample data
- Step 4: Evaluate null hypothesis
- Step 5: Conclusion

4

- We know that the average GRE score for college seniors is μ = 500 and σ = 100. A researcher is interested in effect of the Kaplan course on GRE scores.
- A random sample of 100 college seniors is selected and take the Kaplan GRE Training course. Afterwards, each is given the GRE exam. Average scores after training are 525 for the sample of 100 students.

$\overline{X} = 525$

5



6

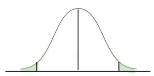
$\mu = 500$ g $\sigma = 100$ g

Step 1: H_o: $\mu_{\text{GRE after Kaplan Course}} = 500$ (There is no effect of the Kaplan training course on average GRE scores)

H_1: $\mu_{\text{GRE after Kaplan Course}} \neq 500$ (There is an effect…)

$\alpha = 0.05$

Step 2: Set criteria



Critical Region
z > 1.96
or
z < -1.96

Step 3: n = 100 $\overline{X} = 525$ $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{100}} = \frac{100}{10} = 10$

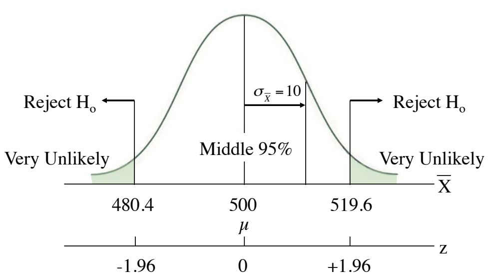
$$Z_{Obt} = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{525 - 500}{10} = \frac{25}{10} = 2.5$$

Step 4: Reject H_o because Z_{obt} of 2.5 is in the critical region.

Step 5: Conclusion. The Kaplan training course significantly increased GRE scores on average, $z = 2.5$, $p < .05$.

7

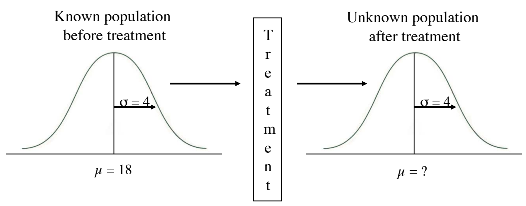
Evaluating Hypotheses – Sampling Distribution



8

What is the effect of alcohol on fetal development?

9



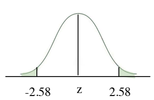
10

$\mu = 18$ g $\qquad \sigma = 4$ 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 of new born rat pups)

H_1: $\mu \neq 18$ g (There is an effect…) $\propto = 0.01$

Step 2: Set criteria



Critical Region
z > 2.58
or
z < -2.58

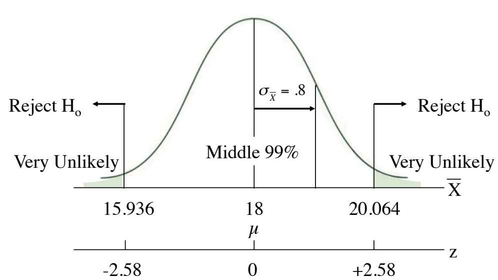
Step 3: n = 25 rats $\quad \overline{X} = 15.5$ g $\quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{25}} = \frac{4}{5} = 0.8$

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

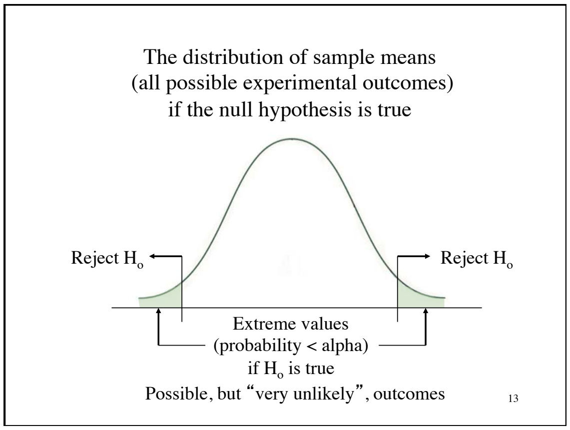
Step 4: Reject H_o because Z_{obt} of -3.125 is in the critical region.

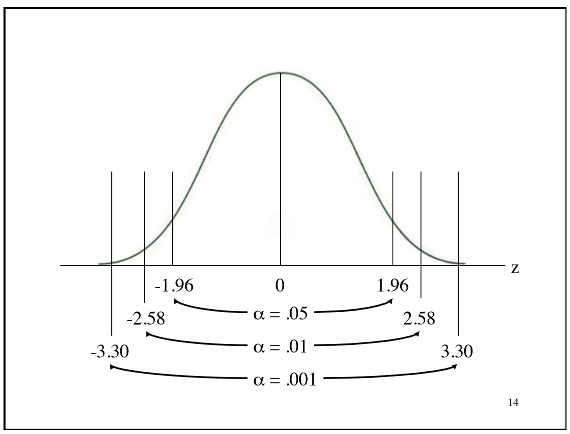
Step 5: Conclusion: Alcohol significantly reduced mean birth weight of newborn rat pups, $z = 3.125$, $p < .01$.

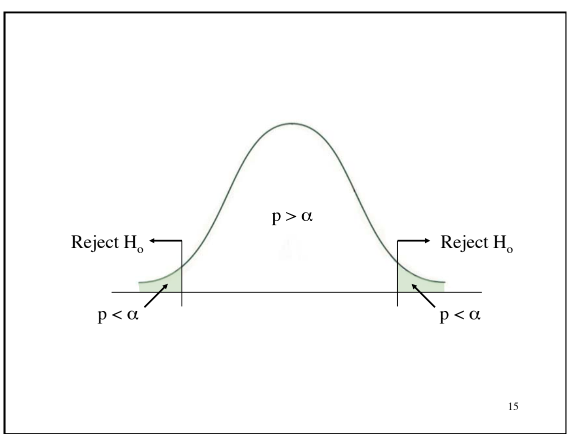
11



12


The distribution of sample means
(all possible experimental outcomes)
if the null hypothesis is true
Reject H_o
Reject H_o
Extreme values
(probability < alpha)
if H_o is true
Possible, but "very unlikely", outcomes
13


z
-1.96
0
1.96
-2.58
α = .05
2.58
-3.30
α = .01
3.30
α = .001
14


p > α
Reject H_o
Reject H_o
p < α
p < α
15

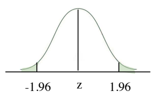
What if the sample mean does not indicate a large enough change to reject the null?

Step 1: H_o: $\mu_{\text{GRE after Kaplan Course}} = 500$ (There is no effect of the Kaplan training course on average GRE scores)

H_1: $\mu_{\text{GRE after Kaplan Course}} \neq 500$ (There is an effect…)

$\propto = 0.05$

Step 2: Set criteria



Critical Region
z > 1.96
or
z < -1.96

Step 3: n = 100 $\overline{X} = 515$ $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{100}{\sqrt{100}} = \dfrac{100}{10} = 10$

$$Z_{Obt} = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{515 - 500}{10} = \frac{15}{10} = 1.5$$

Step 4: Retain H_o because Z_{obt} of 1.5 is not in the critical region.

Step 5: Conclusion. There was no effect of the Kaplan training on GRE scores on average, $z = 1.5$, $p > .05$.

16

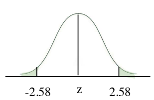
What if the sample mean is not in the critical region?

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 of new born rat pups)

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

$\propto = 0.01$

Step 2: Set criteria



Critical Region
z > 2.58
or
z < -2.58

Step 3: n = 25 rats $\overline{X} = 17$g $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{4}{\sqrt{25}} = \dfrac{4}{5} = 0.8$

$$Z_{Obt} = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{17 - 18}{0.8} = -1.25$$

Step 4: Retain H_o because Z_{obt} of -1.25 is not in the critical region.

Step 5: Conclusion: There was no effect of alcohol on the average birth weight of newborn rat pups, $z = 1.25$, $p > .01$.

17

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

18

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

19

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

20

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

21

Assumptions for Hypothesis Tests with z-scores:

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

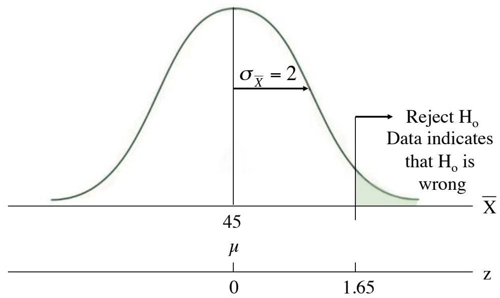
22

One-Tailed Hypothesis Tests

23

- 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.

24



25

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

26

Error and Power

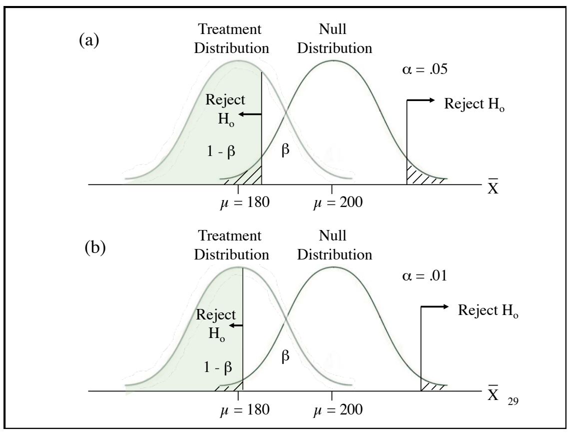
- Type I error = α
  - Probability of a false alarm
- Type II error = β
  - Probability of missing an effect when H_o is really false
- Power = 1 - β
  - Probability of correctly detecting effect when H_o is really false

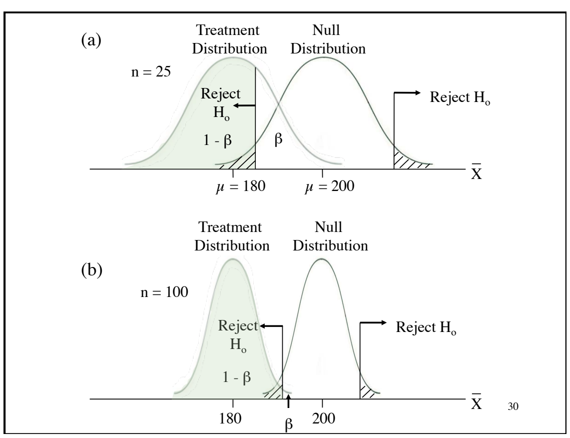
27

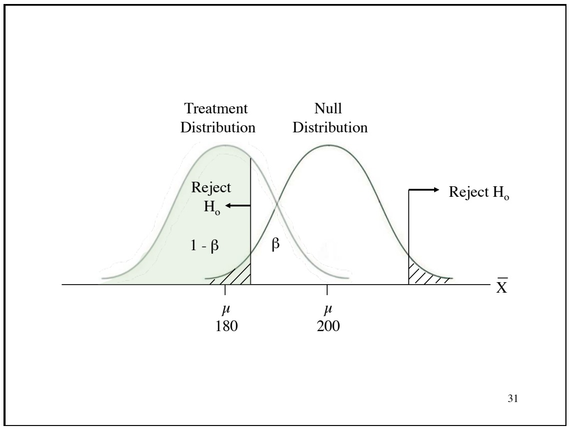
Factors Influencing Power (1 - β)

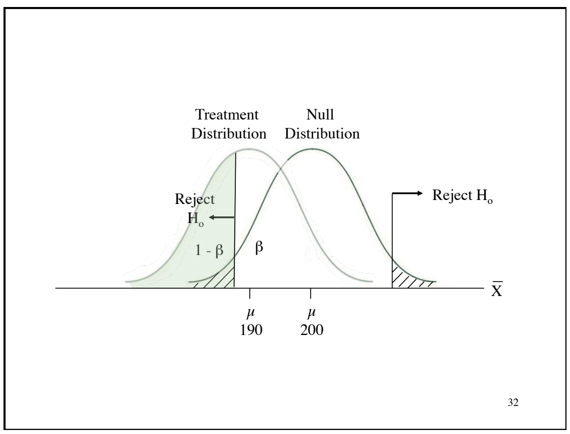
1. Sample size
2. One-tailed versus two-tailed test
3. Criterion (α level)
4. Size of treatment effect
5. Design of study

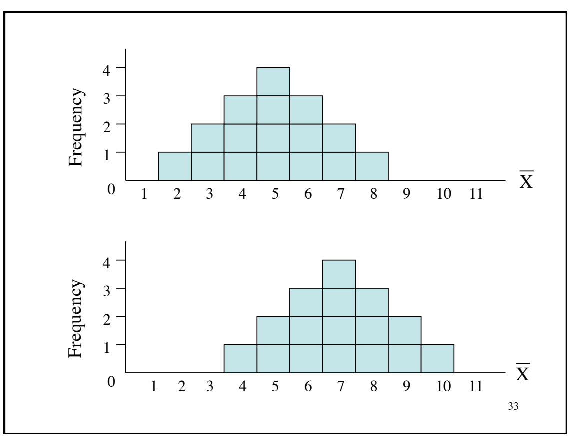
28






Treatment Distribution
Null Distribution
Reject H_o
Reject H_o
1 - β
β
X̄
μ 180
μ 200
31


Treatment Distribution
Null Distribution
Reject H_o
Reject H_o
1 - β
β
X̄
μ 190
μ 200
32


Frequency
0 1 2 3 4
1 2 3 4 5 6 7 8 9 10 11
X̄
Frequency
0 1 2 3 4
1 2 3 4 5 6 7 8 9 10 11
X̄
33

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

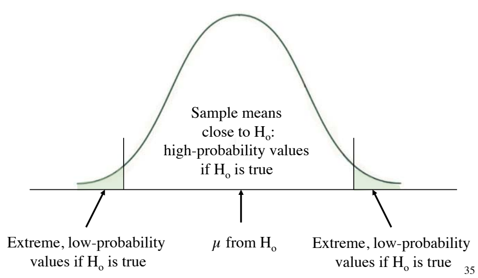
$\mu = 2.9$ kg $\sigma = 2.9$ kg

Random Sample: n = 14

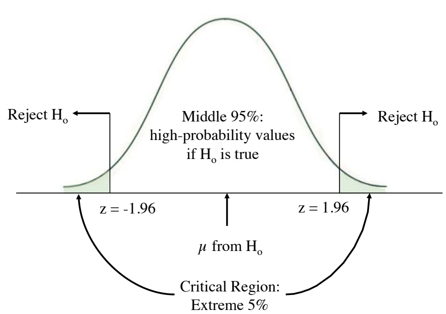
2.3, 2.0, 2.2, 2.8, 3.2, 2.2, 2.5,
2.4, 2.4, 2.1, 2.3, 2.6, 2.0, 2.3

34

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



35



36