Definition

- A hypothesis is a tentative, yet testable statement which you expect to find in your empirical data.
- (Empirical data is data that is acquired by observation, experience, or experimentation)

1

Hypothesis Testing

- Tests a claim about a parameter using evidence data in a sample
- Null and Alternative Hypotheses
- Test Statistic
- *P*-Value
- Significance Level
- One-Sample *z* Test
- Power and Sample Size

2

Terms

- **Population** ≡ all possible values
- **Sample** ≡ a portion of the population
- **Statistical inference** ≡ generalizing from a sample to a population with calculated degree of certainty
- Two forms of statistical inference
  - **Hypothesis testing**
  - **Estimation**
- **Parameter** ≡ a characteristic of population, e.g., population mean μ
- **Statistic** ≡ calculated from data in the sample, e.g., sample mean ($\bar{x}$)

3

Type I and Type II Errors

	Accept null	Reject null
Null is true	Correct- no error	Type I error
Null is false	Type II error	Correct- no error

4

Definition

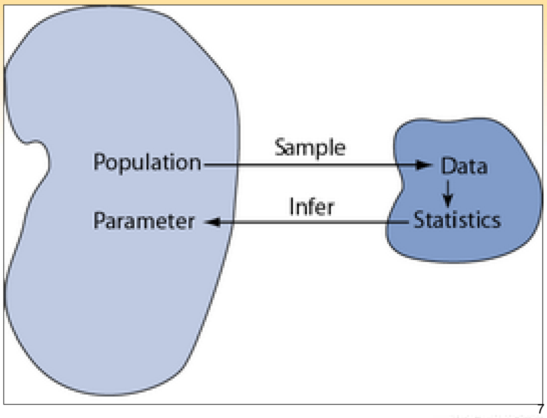
- Significance level (α) is the probability of rejecting the null hypothesis in a statistical test when it is true.
- It is the probability of Type I error.

5

Difference between Parameters and Statistics

	Parameters	Statistics
Source	Population	Sample
Notation	Greek (e.g., μ)	Roman (e.g., *xbar*)
Vary	No	Yes
Calculated	No	Yes

6



7

Null and Alternative Hypotheses

- Convert the research question to null and alternative hypotheses
- The null hypothesis (H_0) is a claim of "no difference in the population"
- The alternative hypothesis (H_a) claims "H_0 is false"
- Collect data and seek evidence against H0 as a way of bolstering H_a (deduction)

8

General Procedure for Hypothesis Test

1. Formulate H_0 (null hypothesis) and H_1 (alternative hypothesis)
2. Select appropriate test
3. Choose level of significance
4. Calculate the test statistic (SPSS)
5. Determine the probability associated with the statistic (p-value)
6. Determine the critical value of the test statistic

9

6. Compare with the level of significance, α

7. Determine if the critical value falls in the rejection region. (check tables)

8. Reject or do not reject H_0

9. Draw a conclusion

10

Example

- If students are given adequate training on library skills, the time spent to search for information will be reduced.
- Is it testable?
- Is the relationship statistically significant?

11

- **Data of student information search time, without prior training:**
  - mean (μ) time of 170 seconds.
  - standard deviation (σ) was 40 seconds.

We test whether after given training, the mean time taken in the population now differs.

12

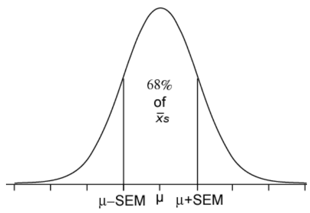
- **Null hypothesis** H_0: μ = 170 ("no difference")
- The **alternative hypothesis** can be either
  - H_a: μ < 170 (**one-sided test**) or
  - H_a: μ ≠ 170 (**two-sided test**)

13

Sampling Distributions of a Mean, $\overline{x}$ are Normally distributed

$$\overline{x} \sim N\left(\mu, SE_{\overline{x}}\right)$$

$$\text{where } SE_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$



14

Test Statistic, z_{stat}

Statistics of a one-sample test of a mean when σ is known. Use this statistic to test the hypothesis:

$$z_{stat} = \frac{\overline{x} - \mu_0}{SE_{\overline{x}}}$$

$$\text{where } \mu_0 \equiv \text{population mean assuming } H_0 \text{ is true}$$

$$\text{and } SE_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

15

Example 1: Using test statistic, z_{stat}

Example: Right-tailed test (https://onlinecourses.science.psu.edu/statprogram/print/book/export/html/139)

An engineer measured the hardness of 25 pieces of iron pieces. The resulting data were:

170	167	174	179	179
156	163	156	187	156
183	179	174	179	170
156	187	179	183	174
187	167	159	170	179

The engineer hypothesized that the mean hardness of *all* such iron pieces is greater than 170. Therefore, he was interested in testing the hypotheses:

$H_0 : \mu = 170$
$H_A : \mu > 170$

The engineer entered his data into Minitab and conducted the "one-sample *t*-test". He obtained the following output:

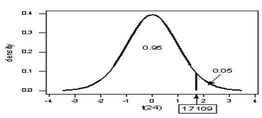
One-Sample T: Brinell

```
Test of mu = 170 vs mu > 170

Variable   N   Mean    StDev  SE Mean    T      P
Brinell   25  172.52  10.31     2.06   1.22  0.117
```

The mean hardness of the n = 25 pieces of ductile iron was 172.52 with a standard deviation of 10.31. (The standard error of the mean "SE Mean", calculated by dividing the standard deviation 10.31 by the square root of n = 25, is 2.06). The test statistic, z_{stat} is 1.22, and the *P*-value is 0.117.

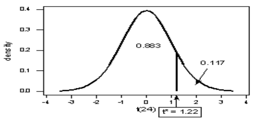
At significance level, α = 0.05, reject the null hypothesis if the test statistic, z_{stat} is greater than the critical value, 1.7109 (determined using statistical software or a *t*-table):



Since the test statistic, z_{stat} = 1.22, is not greater than 1.7109, the engineer fails to reject the null hypothesis. That is, the test statistic does not fall in the "critical region."
There is insufficient evidence, at the α = 0.05 level, to conclude that the mean hardness of all such iron pieces is greater than 170.

Using p-value

Determine the area under a $z_{n-1} = z_{24}$ curve and to the *right* of the test statistic $t^* = 1.22$:



In the output above, the *P*-value is 0.117.
Since the *P*-value, 0.117, is greater than α = 0.05, the engineer fails to reject the null hypothesis.
There is insufficient evidence, at the α = 0.05 level, to conclude that the mean hardness of all the iron pieces is greater than 170.

Example 2: test statistic, z_{stat}

- Suppose, $\mu_0 = 170$ and $\sigma = 40$
- Take a random sample of $n = 64$. Therefore

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{64}} = 5$$

- If we found a sample mean of 173, then

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{173 - 170}{5} = 0.60$$

20

Example 3: Zstat

If we found a sample mean of 185, then

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{185 - 170}{5} = 3.00$$

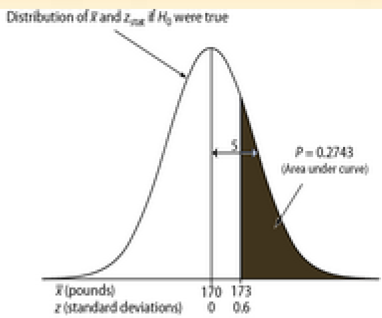
21

P-value

- The *P*-value answer the question: What is the probability of the observed test statistic or one more extreme **when H_0 *is true*?**
- This corresponds to the AUC in the tail of the Standard Normal distribution beyond the z_{stat}.
- Convert *z* statistics to *P*-value :
  - For H_a: $\mu > \mu_0 \Rightarrow P = \Pr(Z > z_{stat})$ = right-tail beyond z_{stat}
  - For H_a: $\mu < \mu_0 \Rightarrow P = \Pr(Z < z_{stat})$ = left tail beyond z_{stat}
  - For H_a: $\mu \neq \mu_0 \Rightarrow P = 2 \times$ one-tailed *P*-value
- Use Table B or software to find these probabilities (next two slides).

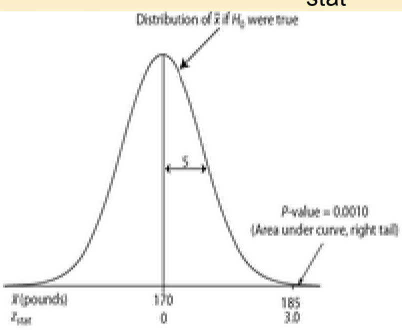
22

One-sided *P*-value for z_{stat} of 0.6



23

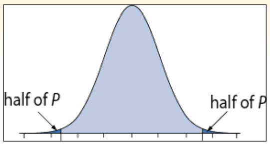
One-sided *P*-value for z_{stat} of 3.0



24

Two-Sided *P*-Value

- One-sided $H_a \Rightarrow$ AUC in tail beyond z_{stat}
- Two-sided $H_a \Rightarrow$ consider potential deviations in both directions $\Rightarrow$ double the one-sided *P*-value



Examples: If one-sided $P = 0.0010$, then two-sided $P = 2 \times 0.0010 = 0.0020$. If one-sided $P = 0.2743$, then two-sided $P = 2 \times 0.2743 = 0.5486$.

25

Interpretation

- *P*-value answer the question: What is the probability of the observed test statistic … **when H_0 *is true*?**
- Thus, smaller and smaller *P*-values provide stronger and stronger evidence against H_0
- Small *P*-value $\Rightarrow$ strong evidence

26

Interpretation

Conventions*

$P > 0.10 \Rightarrow$ non-significant evidence against H_0

$0.05 < P \le 0.10 \Rightarrow$ marginally significant evidence

$0.01 < P \le 0.05 \Rightarrow$ significant evidence against H_0

$P \le 0.01 \Rightarrow$ highly significant evidence against H_0

Examples

$P = .27 \Rightarrow$ non-significant evidence against H_0

$P = .01 \Rightarrow$ highly significant evidence against H_0

*** It is *unwise* to draw firm borders for "significance"**

27

α-Level (Used in some situations)

- Let α ≡ probability of erroneously rejecting H_0
- Set α threshold (e.g., let α = .10, .05, *or whatever*)
- Reject H_0 when $P \le \alpha$
- Retain H_0 when $P > \alpha$
- Example: Set α = .10. Find $P = 0.27 \Rightarrow$ retain H_0
- Example: Set α = .01. Find $P = .001 \Rightarrow$ reject H_0

28

(Summary) One-Sample *z* Test

A. Hypothesis statements

H_0: $\mu = \mu_0$ vs.

H_a: $\mu \ne \mu_0$ (two-sided) or

H_a: $\mu < \mu_0$ (left-sided) or

H_a: $\mu > \mu_0$ (right-sided)

B. Test statistic

$$z_{stat} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} \text{ where } SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

C. P-value: convert z_{stat} to P value

D. Significance statement

29

Conditions for z test

- σ known (not from data)
- Population approximately Normal or large sample (central limit theorem)
- SRS (or facsimile)
- Data valid

30

Worked-out Example 2

- Let X represent Weschler Adult Intelligence scores (WAIS)
- Typically, X ~ N(100, 15)
- Take SRS of n = 9 from College ABC student population
- Data ⇒ {116, 128, 125, 119, 89, 99, 105, 116, 118}
- Calculate: x-bar = 112.8
- Does sample mean provide strong evidence that population mean μ > 100?

31

A. Hypotheses:

H_0: μ = 100 versus

H_a: μ > 100 (one-sided)

H_a: μ ≠ 100 (two-sided)

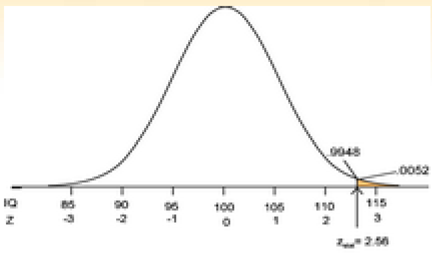
B. Test statistic:

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{9}} = 5$$

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}} = \frac{112.8 - 100}{5} = 2.56$$

32

C. *P*-value: P = Pr(Z ≥ 2.56) = 0.0052

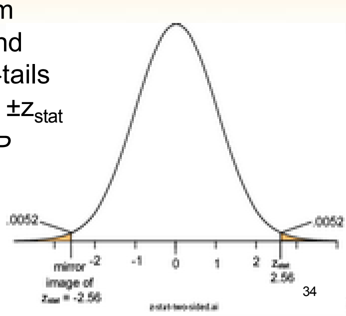


P =.0052 ⇒ it is unlikely the sample came from this null distribution ⇒ strong evidence against H_0

33

Two-Sided *P*-value: College ABC

- H_a: μ ≠100
- Considers random deviations "up" and "down" from μ_0 ⇒tails above and below ±z_{stat}
- Thus, two-sided P = 2 × 0.0052 = 0.0104



34

Power and Sample Size

Two types of decision errors:

Type I error = erroneous rejection of true H_0

Type II error = erroneous retention of false H_0

Decision	Truth: H_0 true	H_0 false
Retain H_0	Correct retention	Type II error
Reject H_0	Type I error	Correct rejection

α ≡ probability of a Type I error

β ≡ Probability of a Type II error

35

Power

- β ≡ probability of a Type II error
  β = Pr(retain H_0 | H_0 false)
  (the "|" is read as "given")
- 1 – β = "Power" ≡ probability of avoiding a Type II error
  1– β = Pr(reject H_0 | H_0 false)

36

Power of a z test

$$1-\beta = \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{|\mu_0 - \mu_a|\sqrt{n}}{\sigma} \right)$$

where

- $\Phi(z)$ represent the cumulative probability of Standard Normal Z
- μ_0 represent the population mean under the null hypothesis
- μ_a represents the population mean under the alternative hypothesis

37

Calculating Power: Example

A study of n = 16 retains H_0: μ = 170 at α = 0.05 (two-sided); σ is 40. What was the power of test's conditions to identify a population mean of 190?

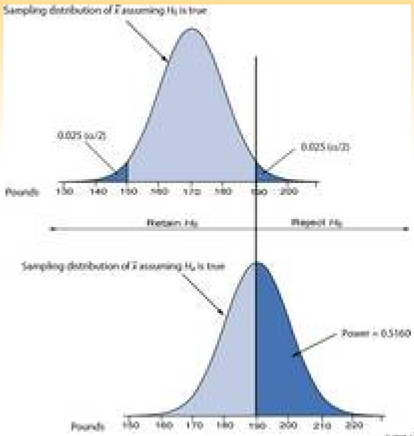
$$1-\beta = \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{|\mu_0 - \mu_a|\sqrt{n}}{\sigma} \right)$$
$$= \Phi\left(-1.96 + \frac{|170-190|\sqrt{16}}{40} \right)$$
$$= \Phi(0.04)$$
$$= 0.5160$$

38

Reasoning Behind Power

- Competing sampling distributions
  - Top curve (next page) assumes H_0 is true
  - Bottom curve assumes H_a is true
  - α is set to 0.05 (two-sided)
- We will reject H_0 when a sample mean exceeds 189.6 (right tail, top curve)
- The probability of getting a value greater than 189.6 on the bottom curve is 0.5160, corresponding to the power of the test

39



40

Sample Size Requirements

Sample size for one-sample z test:

$$n = \frac{\sigma^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}} \right)^2}{\Delta^2}$$

where

- $1-\beta \equiv$ desired power
- $\alpha \equiv$ desired significance level (two-sided)
- $\sigma \equiv$ population standard deviation
- $\Delta = \mu_0 - \mu_a \equiv$ the **difference worth detecting**

41

Example: Sample Size Requirement

How large a sample is needed for a one-sample z test with 90% power and α = 0.05 (two-tailed) when σ = 40? Let H_0: μ = 170 and H_a: μ = 190 (thus, $\Delta = \mu_0 - \mu_a$ = 170 – 190 = −20)

$$n = \frac{\sigma^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}} \right)^2}{\Delta^2} = \frac{40^2(1.28+1.96)^2}{-20^2} = 41.99$$

Round up to 42 to ensure adequate power.

42

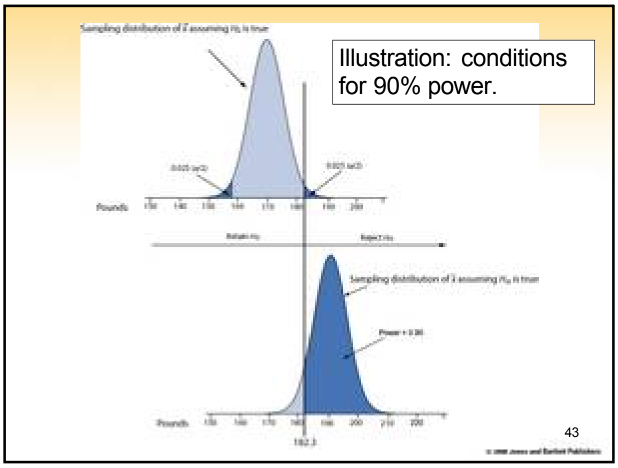


Illustration: conditions for 90% power.
Pounds
Pounds
182.3
43