Hypothesis Testing Procedures

Objectives

- Define null and research hypothesis, test statistic, level of significance and decision rule
- Understand Type I and Type II errors
- Differentiate hypothesis testing procedures based on type of outcome variable and number of samples

Hypothesis Testing

- Research hypothesis is generated about unknown population parameter
- Sample data are analyzed and determined to support or refute the research hypothesis

Hypothesis Testing Procedures

Step 1

Null hypothesis (H_0):

No difference, no change

Research hypothesis (H_1):

What investigator

believes to be true

Hypothesis Testing Procedures

Step 2

Collect sample data and determine whether sample data support research hypothesis or not.

For example, in test for μ, evaluate $\bar{x}$.

Hypothesis Testing Procedures

Step 3

- Set up decision rule to decide when to believe null versus research hypothesis
- Depends on level of significance, $\alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true})$

Hypothesis Testing Procedures

Steps 4 and 5

- Summarize sample information in test statistic (e.g., Z value)
- Draw conclusion by comparing test statistic to decision rule. Provide final assessment as to whether H_1 is likely true given the observed data.

P-values

- P-values represent the exact significance of the data
- Estimate p-values when rejecting H_0 to summarize significance of the data (can approximate with statistical tables, can get exact value with statistical computing package)
- P-value is the smallest α where we still reject H_0

Hypothesis Testing Procedures

1. Set up null and research hypotheses, select α
2. Select test statistic
4. Set up decision rule
5. Compute test statistic
6. Draw conclusion & summarize significance

Errors in Hypothesis Tests

Conclusion of Statistical Test

	Do Not Reject H_0	Reject H_0
H_0 true	Correct	Type I error
H_0 false	Type II error	Correct

Hypothesis Testing for μ

- Continuous outcome
- 1 Sample

H_0: $\mu=\mu_0$

H_1: $\mu>\mu_0$, $\mu<\mu_0$, $\mu\neq\mu_0$

Test Statistic

$n \geq 30$

$$Z = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

(Find critical value in Table 3C, Table 4C)

$n<30$

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

Example 7.2.

Hypothesis Testing for μ

The National Center for Health Statistics (NCHS) reports the mean total cholesterol for adults is 203. Is the mean total cholesterol in Framingham Heart Study participants significantly different?

In 3310 participants the mean is 200.3 with a standard deviation of 36.8.

Hypothesis Testing for μ

1. H_0: $\mu=203$

   H_1: $\mu\neq 203$ $\alpha=0.05$

2. Test statistic

$$Z = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

3. Decision rule

   Reject H_0 if $z \geq 1.96$ or if $z \leq -1.96$

Hypothesis Testing for μ

4. Compute test statistic

$$Z = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} = \frac{200.3 - 203}{36.8/\sqrt{3310}} = -4.22$$

5. Conclusion. Reject H_0 because $-4.22 \le -1.96$. We have statistically significant evidence at $\alpha=0.05$ to show that the mean total cholesterol is different in the Framingham Heart Study participants.

Hypothesis Testing for μ

Significance of the findings. Z = -4.22.

Table 1C. Critical Values for Two-Sided Tests

α	Z
0.20	1.282
0.10	1.645
0.05	1.960
0.010	2.576
0.001	3.291
0.0001	3.819

$p<0.0001$.

Hypothesis Testing for p

- Dichotomous outcome
- 1 Sample

H_0: $p=p_0$

H_1: $p>p_0$, $p<p_0$, $p\neq p_0$

Test Statistic

$$\min[np_0, n(1-p_0)] \geq 5$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

(Find critical value in Table 1C)

Example 7.4.

Hypothesis Testing for p

The NCHS reports that the prevalence of cigarette smoking among adults in 2002 is 21.1%. Is the prevalence of smoking lower among participants in the Framingham Heart Study?

In 3536 participants, 482 reported smoking.

Hypothesis Testing for p

1. H_0: $p=0.211$

   H_1: $p<0.211$ $\alpha=0.05$

2. Test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

3. Decision rule

   Reject H_0 if $z \leq -1.645$

Hypothesis Testing for p

4. Compute test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.136 - 0.211}{\sqrt{\frac{0.211(1-0.211)}{3536}}} = -10.93$$

5. Conclusion. Reject H_0 because $-10.93 \le -1.645$. We have statistically significant evidence at $\alpha=0.05$ to show that the prevalence of smoking is lower among the Framingham Heart Study participants. ($p<0.0001$)

Hypothesis Testing for Discrete Outcomes*

Discrete (ordinal or categorical) outcome

- 1 Sample

H_0: $p_1=p_{10}$, $p_2=p_{20}, \ldots, p_k=p_{k0}$

H_1: H_0 is false

Test Statistic

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

(Find critical value in Table 3)

* χ^2 goodness-of-fit test

Example 7.6.

χ^2 goodness-of-fit test

A university survey reveals that 60% of students get no regular exercise, 25% exercise sporadically and 15% exercise regularly. The university institutes a health promotion campaign and re-evaluates exercise one year later.

None Sporadic

Regular

χ^2 goodness-of-fit test

1. H_0: p_1=0.60, p_2=0.25, p_3=0.15

   H_1: H_0 is false α=0.05

2. Test statistic $\chi^2 = \sum \frac{(O - E)^2}{E}$

3. Decision rule df=k-1=3-1=2

   Reject H_0 if $\chi^2 \geq 5.99$

χ^2 goodness-of-fit test

4. Compute test statistic

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

	None	Sporadic	Regular	Total
No. students (O)	255	125	90	470
Expected (E)	282	117.5	70.5	470
$(O-E)^2/E$	2.59	0.48	5.39	

$\chi^2 = 8.46$

Example 7.6.

χ^2 goodness-of-fit test

5. Conclusion. Reject H_0 because $8.46 \geq 5.99$. We have statistically significant evidence at $\alpha=0.05$ to show that the distribution of exercise is not 60%, 25%, 15%.

Using Table 3, the p-value is $p<0.005$.

Hypothesis Testing for ($\mu_1 - \mu_2$)

- Continuous outcome
- 2 Independent Sample

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$, $\mu_1 < \mu_2$, $\mu_1 \neq \mu_2$

Test Statistic

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$n1 \geq 30$ and

$n_2 \geq 30$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

(Find critical value in Table 1C, Table 2)

$n_1 < 30$ or

$n_2 < 30$

Pooled Estimate of Common Standard Deviation, Sp

- Previous formulas assume equal variances ($\sigma_1^2=\sigma_2^2$)
- If $0.5 \leq s_1^2/s_2^2 \leq 2$, assumption is reasonable

$$Sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Example 7.9.

Hypothesis Testing for $(\mu_1 - \mu_2)$

A clinical trial is run to assess the effectiveness of a new drug in lowering cholesterol. Patients are randomized to receive the new drug or placebo and total cholesterol is measured after 6 weeks on the assigned treatment.

Is there evidence of a statistically significant reduction in cholesterol for patients on the new drug?

Hypothesis Testing for $(\mu_1 - \mu_2)$

	Sample Size	Mean	Std Dev
New Drug	15	195.9	28.7
Placebo	15	227.4	30.3

Example 7.2. Hypothesis Testing for $(\mu_1 - \mu_2)$

1. $H_0: \mu_1 = \mu_2$

   $H_1: \mu_1 < \mu_2$ $\alpha = 0.05$

2. Test statistic

$$t = \frac{\bar{X}_1 - \bar{X}_2}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

3. Decision rule, $df = n_1 + n_2 - 2 = 28$

   Reject H_0 if $t \leq -1.701$

Assess Equality of Variances

- Ratio of sample variances: $28.7^2/30.3^2 = 0.90$

$$Sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$Sp = \sqrt{\frac{(15-1)28.7^2 + (15-1)30.3^2}{15+15-2}} = \sqrt{870.89} = 29.5$$

Example 7.2. Hypothesis Testing for $(\mu_1 - \mu_2)$

4. Compute test statistic

$$t = \frac{\bar{X}_1 - \bar{X}_2}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{195.9 - 227.4}{29.5\sqrt{\frac{1}{15} + \frac{1}{15}}} = -2.92$$

5. Conclusion. Reject H_0 because $-2.92 \leq -1.701$. We have statistically significant evidence at $\alpha = 0.05$ to show that the mean cholesterol level is lower in patients on treatment as compared to placebo. $(p < 0.005)$

Hypothesis Testing for μ_d

- Continuous outcome
- 2 Matched/Paired Sample

H_0: $\mu_d=0$

H_1: $\mu_d>0$, $\mu_d<0$, $\mu_d\neq0$

Test Statistic

n≥30

$$Z = \frac{\overline{X}_d - \mu_d}{s_d / \sqrt{n}}$$

n<30

$$t = \frac{\overline{X}_d - \mu_d}{s_d / \sqrt{n}}$$

(Find critical value in Table 1C, Table 2)

Example 7.10. Hypothesis Testing for μ_d

Is there a statistically significant difference in mean systolic blood pressures (SBPs) measured at exams 6 and 7 (approximately 4 years apart) in the Framingham Offspring Study?

Among n=15 randomly selected participants, the mean difference was -5.3 units and the standard deviation was 12.8 units. Differences were computed by subtracting the exam 6 value from the exam 7 value.

Example 7.10. Hypothesis Testing for μ_d

1. H_0: $\mu_d=0$

   H_1: $\mu_d \neq 0$ $\alpha=0.05$

2. Test statistic

$$t = \frac{\overline{X}_d - \mu_d}{s_d / \sqrt{n}}$$

3. Decision rule, df=n-1=14

   Reject H_0 if $t \geq 2.145$ or if $z \leq -2.145$

Example 7.10. Hypothesis Testing for μ_d

4. Compute test statistic

$$t = \frac{\overline{X}_d - \mu_d}{s_d / \sqrt{n}} = \frac{-5.3 - 0}{12.8/\sqrt{15}} = -1.60$$

5. Conclusion. Do not reject H_0 because $-2.145 < -1.60 < 2.145$. We do not have statistically significant evidence at $\alpha=0.05$ to show that there is a difference in systolic blood pressures over time.

Hypothesis Testing for $(p_1 - p_2)$

- Dichotomous outcome
- 2 Independent Sample

H_0: $p_1 = p_2$

H_1: $p_1 > p_2$, $p_1 < p_2$, $p_1 \neq p_2$

$$\min[n_1\hat{p}_1, n_1(1-\hat{p}_1), n_2\hat{p}_2, n_2(1-\hat{p}_2)] \geq 5$$

Test Statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(Find critical value

in Table 1C)

Example 7.12. Hypothesis Testing for (p_1-p_2)

Is the prevalence of CVD different in smokers as compared to nonsmokers in the Framingham Offspring Study?

	Free of CVD	History of CVD	Total
Nonsmoker	2757	298	3055
Current smoker	663	81	744
Total	3420	379	3799

Example 7.12. Hypothesis Testing for (p_1-p_2)

1. $H_0: p_1=p_2$

   $H_1: p_1 \neq p_2$ $\alpha=0.05$

2. Test statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

3. Decision rule

   Reject H_0 if $Z \leq -1.96$ or if $Z \geq 1.96$

Example 7.12. Hypothesis Testing for $(p_1 - p_2)$

4. Compute test statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{p}_1 = \frac{81}{744} = 0.1089, \hat{p}_2 = \frac{298}{3055} = 0.0975$$

$$\hat{p} = \frac{81 + 298}{744 + 3055} = 0.0988$$

$$Z = \frac{0.1089 - 0.0975}{\sqrt{0.0988(1-0.0988)\left(\frac{1}{744} + \frac{1}{3055}\right)}} = 0.927$$

Example 7.12: Hypothesis Testing for (p_1-p_2)

5. Conclusion. Do not reject H_0 because $-1.96 < 0.927 < 1.96$. We do not have statistically significant evidence at $\alpha=0.05$ to show that there is a difference in prevalent CVD between smokers and nonsmokers.

Hypothesis Testing for More than 2 Means*

- Continuous outcome
- k Independent Samples, k > 2

$H_0: \mu_1=\mu_2=\mu_3 \ldots =\mu_k$

H_1: Means are not all equal

Test Statistic

$$F = \frac{\Sigma n_j(\overline{X}_j - \overline{X})^2/(k-1)}{\Sigma\Sigma(X - \overline{X}_j)^2/(N-k)}$$

(Find critical value in Table 4)

*Analysis of Variance

ANOVA Table

Source of Variation	Sums of Squares	df	Mean Squares	F
Between Treatments	$SSB = \Sigma\, n_j (\bar{X}_j - \bar{X})^2$	k-1	SSB/k-1	MSB/ MSE
Error	$SSE = \Sigma\Sigma (X - \bar{X}_j)^2$	N-k	SSE/N-k	
Total	$SST = \Sigma\Sigma (X - \bar{X})^2$	N-1		

Example[7.14] ANOVA

Is there a significant difference in mean weight loss among 4 different diet programs? (Data are pounds lost over 8 weeks)

Low-Cal	Low-Fat	Low-Carb	Control
8	2	3	2
9	4	5	2
6	3	4	-1
7	5	2	0
3	1	3	3

Example 7.14. ANOVA

1. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

   H_1: Means are not all equal , $\alpha = 0.05$

2. Test statistic

$$F = \frac{\Sigma n_j (\overline{X}_j - \overline{X})^2 / (k-1)}{\Sigma\Sigma (X - \overline{X}_j)^2 / (N-k)}$$

Example 7.14. ANOVA

3. Decision rule

$df_1=k-1=4-1=3$

$df_2=N-k=20-4=16$

Reject H_0 if $F \geq 3.24$

Example 7.14. ANOVA

Summary Statistics on Weight Loss by Treatment

	Low-Cal	Low-Fat	Low-Carb	Control
N	5	5	5	5
Mean	6.6	3.0	3.4	1.2

Overall Mean = 3.6

Example 7.14. ANOVA

$$SSB = \Sigma\, n_j (\overline{X}_j - \overline{X})^2$$

$=5(6.6-3.6)^2+5(3.0-3.6)^2+5(3.4-3.6)^2+5(1.2-3.6)^2$

$= 75.8$

Example 7.14. ANOVA

$$SSE = \Sigma \Sigma (X - \overline{X}_j)^2$$

Low-Cal	(X-6.6)	$(X-6.6)^2$
8	1.4	2.0
9	2.4	5.8
6	-0.6	0.4
7	0.4	0.2
3	-3.6	13.0
Total	0	21.4

Example 7.14. ANOVA

$$SSE = \Sigma\,\Sigma\,(X - \overline{X}_j)^2$$

=21.4 + 10.0 + 5.4 + 10.6 = 47.4

Example 7.14. ANOVA

Source of Variation	Sums of Squares	df	Mean Squares	F
Between Treatments	75.8	3	25.3	8.43
Error	47.4	16	3.0	
Total	123.2	19		

Example 7.14. ANOVA

4. Compute test statistic

$F=8.43$

5. Conclusion. Reject H_0 because $8.43 \geq 3.24$. We have statistically significant evidence at $\alpha=0.05$ to show that there is a difference in mean weight loss among 4 different diet programs.

Hypothesis Testing for Discrete Outcomes*

- Discrete (ordinal or categorical) outcome
- 2 or More Samples

H_0: The distribution of the outcome is independent of the groups

H_1: H_0 is false

Test Statistic

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

(Find critical value in Table 3)

* χ^2 test of independence

Example 7.16.

χ^2 test of independence

Is there a relationship between students' living arrangement and exercise status?

	Exercise Status			
	None	Sporadic	Regular	Total
Dormitory	32	30	28	90
On-campus Apt	74	64	42	180
Off-campus Apt	110	25	15	150
At Home	39	6	5	50
Total	255	125	90	470

Example

χ^2 test of independence

1. H_0: Living arrangement and exercise status are independent

   H_1: H_0 is false $\alpha=0.05$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

2. Test statistic

3. Decision rule   df=(r-1)(c-1)=3(2)=6

   Reject H_0 if $\chi^2 > 12.59$

Example
χ^2 test of independence

4. Compute test statistic

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O = Observed frequency

E = Expected frequency

E = (row total)*(column total)/N

Example

χ^2 test of independence

4. Compute test statistic

Table entries are Observed (Expected) frequencies

	Exercise Status			
	None	Sporadic	Regular	Total
Dormitory	32	30	28	90
	(90*255/470=48.8)	(23.9)	(17.2)	
On-campus Apt	74	64	42	180
	(97.7)	(47.9)	(34.5)	
Off-campus Apt	110	25	15	150
	(81.4)	(39.9)	(28.7)	
At Home	39	6	5	50

Example

χ^2 test of independence

4. Compute test statistic

$$\chi^2 = \frac{(32-48.8)^2}{48.8} + \frac{(30-23.9)^2}{23.9} + \frac{(28-17.2)^2}{17.2} + \ldots + \frac{(5-9.6)^2}{9.6}$$

$$\chi^2 = 60.5$$

Example

χ^2 test of independence

5. Conclusion. Reject H_0 because $60.5 \geq 12.59$. We have statistically significant evidence at $\alpha=0.05$ to show that living arrangement and exercise status are not independent. ($P<0.005$)

Ch 6 g&m, Q 6,10,22,23.

6. The data below represent the systolic blood pressures (in mmHg) of 14 patients undergoing drug therapy for hypertension. Assuming normality of systolic blood pressures, on the basis of these data can you conclude that the mean is significantly less than 165 mmHg?

183 152 178 157 194 163 144

194 163 114 178 152 118 158

- The hypotheses are $H0$: $\mu \geq 165$ mmHg versus Ha: $\mu < 165$ mmHg.
- $t = -0.671$ and the P value = 0.2550.
- So accept $H0$; we cannot conclude the mean systolic blood pressure is significantly less than 165 mmHg.

10. Recently there have been concerns about the effects of phthalates on the development of the male reproductive system. Phthalates are common ingredients in many plastics. In a pilot study a researcher gave pregnant rats daily doses of 750 mg/kg of body weight of DEHP (di-2-ethylhexyl phthalate) throughout the period when their pups' sexual organs were developing. The newly born male rat pups were sacrificed and their seminal vesicles were dissected and weighed. Below are the weights for the eight males (in mg).

1710 1630 1580 1670 1350 1650 1600 1650

If untreated newborn males have a mean of 1700 mg, can you say that rats exposed to DHEP in utero have a significantly lower weight?

The hypotheses are $H0$: $\mu \geq 1700$ mg versus Ha: $\mu < 1700$ mg.

For a t test, the c.v. = −1.895.

$t = -2.44$. Since $-2.44 < -1.895$

reject $H0$. Exposure to DEHP significantly decreases seminal vesicle weight.

22. The hypotheses are $H0$: $\mu \leq 12.5$ yr versus Ha: $\mu > 12.5$ yr. Use a t test with $\alpha = 0.05$.

$X = 14.75$ yr, $n = 10$, and $s = 0.84$ yr.

c.v. = 1.833.

$t = X - \mu \; / s \sqrt{n} = 14.75 - 12.5 \; / \; 0.84 / \sqrt{10} = 2.25 \; / \; 0.27 =$ **8.33.**

- Since 8.33 > 1.833, reject $H0$. Menarche is significantly later in world-class swimmers.

23. Redo Problem 6 in Chapter 1 as a test of hypothesis question.

The hypotheses are $H0$: $\mu \leq 24$ hours versus Ha: $\mu > 24$ hours.

For a t test with $\alpha = 0.05$.

$X = 14.75$ yr, $n = 15$, and $s^2 = 0.849$ hr^2. c.v. = 1.761.

$t = X - \mu / s\sqrt{n} = 25.87 - 24 / 0.92/\sqrt{15} = 1.87 / 0.24 =$ **7.79.**

- Since 7.79 > 1.761, reject $H0$. The average day for bunkered people is significantly longer than 24 hours.