Summary of Chapter 9

1 Concepts

- **Chi-Square Tests**:

  Test statistic $\chi^2 = \sum\limits_{\text{all cells}} \frac{(f_o - f_e)^2}{f_e}$, where $f_e = \frac{n_{i+} n_{+j}}{n}$

  1. χ^2 **test for the difference between two proportions**:

  H_0: $p_1 = p_2$, H_1: $p_1 \neq p_2$ If $\chi^2 > \chi^2_1$, reject H_0.

  2. χ^2 **test for the difference in more than two proportions**:

  H_0: $p_1 = p_2 = \cdots = p_c$, H_1: Not all of the p_j are equal ($j = 1, 2, \cdots, c$)
  If $\chi^2 > \chi^2_{c-1}$, reject H_0.

  3. χ^2 **test for independence**:

  H_0: The two categorical variables are independent, H_1: The two categorical variables are dependent. If $\chi^2 > \chi^2_{(r-1)(c-1)}$, reject H_0.

- **Nonparametric Tests**:

1. Wilcoxon rank sum test for the difference between two medians

1) Small sample ($n_1 \le 10$, $n_2 \le 10$)

Test statistic: T_1 = sum of ranks from smaller sample

Two-Tail Test: H_0: $M_1 = M_2$, H_1: $M_1 \neq M_2$. If $T_1 < T_{1L}$ or $T_1 > T_{1U}$, reject H_0.

Upper-Tail Test: H_0: $M_1 \le M_2$, H_1: $M_1 > M_2$. If $T_1 > T_{1U}$, reject H_0.

Lower-Tail Test: H_0: $M_1 \ge M_2$, H_1: $M_1 < M_2$. If $T_1 < T_{1L}$, reject H_0.

2) Large sample ($n_1 \ge 10$ and/or $n_2 \ge 10$)

Test statistic: $Z = \frac{T_1 - \mu_{T_1}}{\sigma_{T_1}}$, where $\mu_{T_1} = \frac{n_1(n+1)}{2}$, $\sigma_{T_1} = \sqrt{\frac{n_1 n_2 (n+1)}{12}}$

Two-Tail Test: If $Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$, reject H_0.

Upper-Tail Test: If $Z > Z_\alpha$, reject H_0.

Lower-Tail Test: If $Z < -Z_\alpha$, reject H_0.

2. Kruskal-Wallis rank test for the difference in more than two medians

H_0: $M_1 = M_2 = \cdots = M_c$

H_1: Not all of the M_j are equal ($j = 1, 2, \cdots, c$)

Test statistic: $H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j} \right] - 3(n+1)$, where

n = sum of sample sizes in all samples c = number of samples

T_j = sum of ranks in the j-th sample n_j = size of the j-th sample

If $H > \chi^2_{c-1}$, reject H_0.

2 Examples

Example 1. More shoppers do their majority of grocery shopping on Saturday than any other day of the week. However, is the day of the week a person does the majority of grocery shopping dependent on age? A study cross-classified grocery shoppers by age and major shopping day ("Major Shopping by Day" *Progressive Grocer Annual Report*, April 30, 2002). The data were reported as percentages and no sample sizes were given.

	AGE		
MAJOR SHOPPING DAY	Under 35	35-54	Over 54
Saturday	24%	28%	12%
A day other than Saturday	76%	72%	88%

Assume that 200 shoppers for each age category were surveyed.

a. Is there any evidence of a significant difference among the age groups with respect to major shopping day? (Use $\alpha = 0.05$.)

b. Determine the p-value in (a) and interpret its meaning.

a. $H_0: \ p_1 = p_2 = p_3 \quad H_1:$ Not all p_j are equal $(j = 1, 2, 3)$.

Populations: 1 = Under 35; 2 = 35-54; 3 = Over 54.

$\alpha = 0.05$, $df = (2-1)(3-1) = 2$, then $\chi_2^2 = 5.9915$ (R command: qchisq(0.95,2)). So, we reject H_0 when $\chi^2 > 5.9915$.

We construct a 2×3 contingency table for the survey.

	AGE			
MAJOR SHOPPING DAY	Under 35	35-54	Over 54	Total
Saturday	48	56	24	128
A day other than Saturday	152	144	176	472
Total	200	200	200	600

$\bar{p} = 128/600 = 0.2133$, since $n_1 = n_2 = n_3 = 200$, then

f_e (Saturday) $= 42.667$, f_e (Other day) $= 157.333$.

The calculations of χ^2-test statistic are presented in the following table:

f_0	f_e	$(f_0 - f_e)$	$(f_0 - f_e)^2$	$(f_0 - f_e)^2/f_e$
48	42.667	5.333	28.44089	0.6665781
56	42.667	13.333	177.7689	4.166426
24	42.667	-18.667	348.4569	8.166895
152	157.333	-5.333	28.44089	0.180769
144	157.333	-13.333	177.7689	1.129889
176	157.333	18.667	348.4569	2.214773
				$\chi^2 = 16.5253$

Since $\chi^2 = 16.5253 > 5.9915$, so we reject H_0.

There is enough evidence to show that there is a significant relationship between age and major grocery shopping day.

b. p-value=$P(\chi^2 > 16.5253) = 0.0003$ (R command: 1-pchisq(16.5253,2)).

Under null hypothesis, the probability of obtaining a sample that gives rise to a test statistic that is equal to or more than 16.5253 is 0.0003.

Using R to conduct the chi-square test:

```
> shop<-matrix(data=c(0.24, 0.28, 0.12,

0.76, 0.72, 0.88)*200, nrow=2, byrow=T)

> shop
     [,1] [,2] [,3]
[1,]   48   56   24
[2,]  152  144  176

> chisq.test(shop)

        Pearson's Chi-squared test

data:  shop X-squared = 16.5254, df = 2,

p-value = 0.0002580
```

Note: as p-value = 0.0002580 < 0.05, H_0 is rejected.

Example 2. A vice president for marketing recently recruited 20 outstanding college graduates to be trained in sales and each, to one of two groups. A "traditional" method of training (*T*) is used in one group and an "experimental" method (*E*) is used in the other. After six months on the job, the vice president ranks the individuals on the basis of their performance from 1 (worst) to 20 (best) with the following results:

T	1	2	3	5	9	10	12	13	14	15
E	4	6	7	8	11	16	17	18	19	20

Is there evidence of a difference in performance based on the two methods? (Use $\alpha = 0.05$)

Solution: $H_0: \; M_1 = M_2 \quad H_1: \; M_1 \neq M_2$.

Populations: 1 = Traditional, 2 = Experimental.

$\alpha = 0.05$, $n_1 = n_2 = 10$, according to the Wilcoxon Rank Sum Test table (Table E.8), we reject H_0 when $T_1 < 78$ or $T_1 > 132$.

The test statistic is:

$$T_1 = 1 + 2 + 3 + 5 + 9 + 10 + 12 + 13 + 14 + 15 = 84.$$

Since $78 \leq 84 \leq 132$, we do not reject H_0.

There is not enough evidence to conclude that there is a difference in performance between traditional and the experimental training methods.

Using R to conduct the Wilcoxon rank sum test:

```
> T<-c(1,2,3,5,9,10,12,13,14,15)
> E<-c(4,6,7,8,11,16,17,18,19,20)
> wilcox.test(T, E)

        Wilcoxon rank sum test

data:  T and E
W = 29, p-value = 0.123 alternative hypothesis: true
location shift is not equal to 0
```

Note: as p-value > 0.05, we do not reject H_0.

Example 3. The following data represent the life time of four different alloys

Alloy			
1	2	3	4
999	1022	1026	974
1010	973	1008	1015
995	1023	1005	1009
998	1023	1007	1011
1001	996	981	995

At the 0.05 level of significance, is there evidence of a difference in the median lifetime of the four alloys?

Solution:

$H_0: \; M_1 = M_2 = M_3 = M_4 \qquad H_1$: Not all M_j are equal ($j = 1, 2, 3, 4$).

$\alpha = 0.05, \; df = c - 1 = 3, \; \chi^2_3 = 7.815$, so we reject H_0 when $H > 7.815$.

The ranks of the lifetime is listed below:

Alloy							
1		2		3		4	
lifetime	Rank	lifetime	Rank	lifetime	Rank	lifetime	Rank
999	8	1022	17	1026	20	974	2
1010	14	973	1	1008	12	1015	16
995	4.5	1023	18.5	1005	10	1009	13
998	7	1023	18.5	1007	11	1011	15
1001	9	996	6	981	3	995	4.5

Rank sums: $T_1 = 42.5,\ T_2 = 61,\ T_3 = 56,\ T_4 = 50.5$, and the test statistic is:

$$H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j}\right] - 3(n+1) = 1.077143 < 7.815.$$

We do not reject H_0, there is not enough evidence to conclude that there is a difference in the median lifetime of the four alloys.

Using R to conduct the Kruskal-Wallis Rank Sum Test:

```
> Alloy<-c(999,1022,1026,974,1010,973,1008,1015,995,1023,
+ 1005,1009,998,1023,1007,1011,1001,996,981,995)
> g<-factor(rep(1:4,5),labels=c("Alloy1","Alloy2",
"Alloy3","Alloy4"))
> kruskal.test(Alloy,g)

        Kruskal-Wallis rank sum test

data:  Alloy and g
Kruskal-Wallis chi-squared = 1.0788, df = 3,
p-value = 0.7822
```

Note: as p-value> 0.05, we do not reject H_0.