

EXAMPLE 20.1a The $n \times M$ Data Matrix for a Hypothetical Multiple Regression or Correlation ($n = 33$; $M = 5$)

j	Variable (i)				
	1 ($^{\circ}\text{C}$)	2 (cm)	3 (mm)	4 (min)	5 (ml)
1	6	9.9	5.7	1.6	2.12
2	1	9.3	6.4	3.0	3.39
3	-2	9.4	5.7	3.4	3.61
4	11	9.1	6.1	3.4	1.72
5	-1	6.9	6.0	3.0	1.80
6	2	9.3	5.7	4.4	3.21
7	5	7.9	5.9	2.2	2.59
8	1	7.4	6.2	2.2	3.25
9	1	7.3	5.5	1.9	2.86
10	3	8.8	5.2	0.2	2.32
11	11	9.8	5.7	4.2	1.57
12	9	10.5	6.1	2.4	1.50
13	5	9.1	6.4	3.4	2.69
14	-3	10.1	5.5	3.0	4.06
15	1	7.2	5.5	0.2	1.98
16	8	11.7	6.0	3.9	2.29
17	-2	8.7	5.5	2.2	3.55
18	3	7.6	6.2	4.4	3.31
19	6	8.6	5.9	0.2	1.83
20	10	10.9	5.6	2.4	1.69
21	4	7.6	5.8	2.4	2.42
22	5	7.3	5.8	4.4	2.98
23	5	9.2	5.2	1.6	1.84
24	3	7.0	6.0	1.9	2.48
25	8	7.2	5.5	1.6	2.83
26	8	7.0	6.4	4.1	2.41
27	6	8.8	6.2	1.9	1.78
28	6	10.1	5.4	2.2	2.22
29	3	12.1	5.4	4.1	2.72
30	5	7.7	6.2	1.6	2.36
31	1	7.8	6.8	2.4	2.81
32	8	11.5	6.2	1.9	1.64
33	10	10.4	6.4	2.2	1.82

EXAMPLE 20.1b A Matrix of Simple Correlation Coefficients, as It Might Appear as Computer Output (from the Data of Example 20.1a)

	1	2	3	4	5
1	1.00000	0.32872	0.16767	0.05191	-0.73081
2	0.32872	1.00000	-0.14550	0.18033	-0.21204
3	0.16767	-0.14550	1.00000	0.24134	-0.05541
4	0.05191	0.18033	0.24134	1.00000	0.31267
5	-0.73081	-0.21204	-0.05541	0.31267	1.00000

EXAMPLE 20.1c A Computer Fit of a Multiple-Regression Equation to the Data of Example 20.1a, Where Variable 5 Is the Dependent Variable

Regression model:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

For each i (where $i = 1, 2, 3, 4$),

$$H_0: \beta_i = 0$$

$$H_A: \beta_i \neq 0$$

Variable i	b_i	s_{b_i}	t	ν	b'_i
X_1	-0.12932	0.021287	-6.075	28	-0.73176
X_2	-0.018785	0.056278	-0.334	28	-0.41108
X_3	-0.046215	0.20727	-0.223	28	-0.26664
X_4	0.20876	0.067034	3.114	28	0.36451

Y intercept: $a = 2.9583$

EXAMPLE 20.1d A Computer Analysis of Variance for the Multiple Regression Data of Example 20.1a

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_A: \beta_1 \text{ and/or } \beta_2 \text{ and/or } \beta_3 \text{ and/or } \beta_4 \neq 0$$

Source of Variation	SS	DF	MS
Total	14.747	32	
Multiple regression	9.7174	4	2.4294
Residual	5.0299	28	0.17964

$F = 13.5$, with DF of 4 and 28

$F_{0.05(1),4,28} = 2.71$, so reject H_0 .

$$P \ll 0.0005 \quad [P = 2.99 \times 10^{-6}]$$

Coefficient of determination: $R^2 = 0.65893$

Adjusted coefficient of determination: $R_a^2 = 0.61021$

Multiple correlation coefficient: $R = 0.81175$

Standard error of estimate: $s_{Y \cdot 1,2,3,4} = 0.42384$

EXAMPLE 20.1e Backward Elimination of Variables in Multiple-Regression Analysis, Using the Data from Example 20.1a

As shown in Example 20.1c, the multiple regression analysis for the model $\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$ yields the following statistics:

Variable	b_i	s_{b_i}	t	ν
X_1	-0.12932	0.021287	-6.075	28
X_2	-0.018785	0.056278	-0.334	28
X_3	-0.046215	0.20727	-0.223	28
X_4	0.20876	0.067034	3.114	28
$a = 2.9583$				

The critical value for testing $H_0: \beta_j = 0$ against $H_A: \beta_j \neq 0$ is $t_{0.05(2),28} = 2.048$. Therefore, H_0 would be rejected for β_1 and β_4 , but not for β_2 or β_3 . Of the t tests for the latter two, the t for testing the significance of β_3 has the smaller absolute value. Therefore, $\beta_3 X_3$ is deleted from the model, leaving $\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4$. The data are then subjected to a multiple-regression analysis using this model with three independent variables, and the following statistics are obtained:

Variable	b_i	s_{b_i}	t	ν
X_1	-0.13047	0.020312	-6.423	29
X_2	-0.015424	0.053325	-0.289	29
X_4	0.20450	0.063203	3.236	29
$a = 2.6725$				

The critical value for testing the significance of these partial regression coefficients is $t_{0.05(2),29} = 2.045$. Therefore, $H_0: \beta_j = 0$ would be rejected for β_1 and for β_4 , but not for β_2 . Therefore, $\beta_2 X_2$ is deleted from the regression model, leaving $\hat{Y} = \alpha + \beta_1 X_1 + \beta_4 X_4$. The analysis of the data using this model, with two independent variables, yields the following statistics:

Variable	b_i	s_{b_i}	t	ν
X_1	-0.13238	0.018913	-6.999	30
X_4	0.20134	0.061291	3.285	30
$a = 2.5520$				

The critical value for testing $H_0: \beta_j = 0$ against $H_A: \beta_j \neq 0$ is $t_{0.05(2),30} = 2.042$. Therefore, both β_1 and β_4 are concluded to be different from zero, and $\hat{Y} = 2.552 - 0.132X_1 + 0.201X_4$ is the final model.

R

CODE

#ex20.1

x1=c(6,1,-2,11,-1,2,5,1,1,3,11,9,5,-3,1,8,-2,3,6,10,4,5,5,3,8,8,6,6,3,5,1,8,10)

x2=c(9.9,9.3,9.4,9.1,6.9,9.3,7.9,7.4,7.3,8.8,9.8,10.5,9.1,10.1,7.2,11.7,8.7,7.6,8.6,10.9,

	<pre> 7.6,7.3,9.2,7.7,7.2,7.8,8.10.1,12.1,7.7,7.8,11.5,10.4) x3=c(5.7,6.4,5.7,6.1,6.5,7.5,9.6,2.5,5.5,2.5,7.6,1.6,4.5,5.5,6.5,5.6,2.5,9.5,6.5,8.5,5.2 ,6.5,5.6,4.6,2.5,4.5,4.6,2.6,8.6,2.6,4) x4=c(1.6,3.0,3.4,3.4,3.0,4.4,2.2,2.2,1.9,0.2,4.2,2.4,3.4,3.0,2.3,9.2,2.4,4.0,2.2,4.2,4.4, 1.6,1.9,1.6,4.1,1.9,2.2,4.1,1.6,2.4,1.9,2.2) x5=c(2.12,3.39,3.61,1.72,1.8,3.21,2.59,3.25,2.86,2.32,1.57,1.5,2.69,4.06,1.98,2.29,3.55,3.31, 1.83 ,1.69,2.42,2.98,1.84,2.48,2.83,2.41,1.78,2.22,2.72,2.36,2.81,1.64,1.82) x=data.frame(x1,x2,x3,x4,x5) cor(x) model=lm(x5~x1+x2+x3+x4) summary(model) anova(model) rmodel=step(model, direction = "backward") </pre>
<p>OUTPU T</p>	<pre> > cor(x) x1 x2 x3 x4 x5 x1 1.00000000 0.3287174 0.16767376 0.05191106 -0.73080887 x2 0.32871738 1.0000000 -0.14549623 0.18032591 -0.21203749 x3 0.16767376 -0.1454962 1.00000000 0.24133905 -0.05540952 x4 0.05191106 0.1803259 0.24133905 1.00000000 0.31266707 x5 -0.73080887 -0.2120375 -0.05540952 0.31266707 1.00000000 > summary(model) Call: lm(formula = x5 ~ x1 + x2 + x3 + x4) Residuals: Min 1Q Median 3Q Max -1.5070 -0.1112 0.0401 0.1550 0.9617 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 2.95828 1.36361 2.169 0.03869 * x1 -0.12932 0.02129 -6.075 1.5e-06 *** x2 -0.01878 0.05628 -0.334 0.74102 x3 -0.04621 0.20727 -0.223 0.82518 x4 0.20876 0.06703 3.114 0.00423 ** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.4238 on 28 degrees of freedom Multiple R-squared: 0.6589, Adjusted R-squared: 0.6102 F-statistic: 13.52 on 4 and 28 DF, p-value: 2.948e-06 </pre>

```

> anova(model)
Analysis of Variance Table

Response: x5
      Df Sum Sq Mean Sq F value    Pr(>F)    
x1      1  7.8762   7.8762  43.8450 3.496e-07 ***
x2      1  0.0131   0.0131   0.0732 0.788785    
x3      1  0.0859   0.0859   0.4781 0.494963    
x4      1  1.7421   1.7421   9.6979 0.004227 ** 
Residuals 28  5.0299   0.1796
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> rmodel=step<del>(model</del>, direction = "backward")
Start:  AIC=-52.08
x5 ~ x1 + x2 + x3 + x4

      Df Sum of Sq    RSS    AIC
- x3    1    0.0089  5.0388 -54.018
- x2    1    0.0200  5.0499 -53.946
<none>                 5.0299 -52.077
- x4    1    1.7421  6.7720 -44.263
- x1    1    6.6298 11.6596 -26.332

Step:  AIC=-54.02
x5 ~ x1 + x2 + x4

      Df Sum of Sq    RSS    AIC
- x2    1    0.0145  5.0533 -55.923
<none>                 5.0388 -54.018
- x4    1    1.8191  6.8579 -45.847
- x1    1    7.1685 12.2073 -26.818

Step:  AIC=-55.92
x5 ~ x1 + x4

      Df Sum of Sq    RSS    AIC
<none>                 5.0533 -55.923
- x4    1    1.8177  6.8710 -47.784
- x1    1    8.2522 13.3055 -25.975

```

SAS

```

CODE  data ex20_1;
      input x1 x2 x3 x4 x5;
      cards;
6  9.9 5.7 1.6 2.12
1  9.3 6.4 3.0 3.39
-2 9.4 5.7 3.4 3.61
11 9.1 6.1 3.4 1.72
-1 6.9 6.0 3.0 1.80
2  9.3 5.7 4.4 3.21
5  7.9 5.9 2.2 2.59
1  7.4 6.2 2.2 3.25
1  7.3 5.5 1.9 2.86
3  8.8 5.2 0.2 2.32
11 9.8 5.7 4.2 1.57
9 10.5 6.1 2.4 1.50
5  9.1 6.4 3.4 2.69
-3 10.1 5.5 3.0 4.06
1  7.2 5.5 0.2 1.98
8 11.7 6.0 3.9 2.29

```

	<pre>-2 8.7 5.5 2.2 3.55 3 7.6 6.2 4.4 3.31 6 8.6 5.9 0.2 1.83 10 10.9 5.6 2.4 1.69 4 7.6 5.8 2.4 2.42 5 7.3 5.8 4.4 2.98 5 9.2 5.2 1.6 1.84 3 7.0 6.0 1.9 2.48 8 7.2 5.5 1.6 2.83 8 7.0 6.4 4.1 2.41 6 8.8 6.2 1.9 1.78 6 10.1 5.4 2.2 2.22 3 12.1 5.4 4.1 2.72 5 7.7 6.2 1.6 2.36 1 7.8 6.8 2.4 2.81 8 11.5 6.2 1.9 1.64 10 10.4 6.4 2.2 1.82 ;run; proc corr data=ex20_1; var x1 x2 x3 x4 x5; run; proc reg data=ex20_1; model x5=x1 x2 x3 x4 / selection=backward; run;</pre>																																				
OUTPUT	<table><tr><th colspan="6">Analysis of Variance</th></tr><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>Model</td><td>4</td><td>9.71735</td><td>2.42934</td><td>13.52</td><td><.0001</td></tr><tr><td>Error</td><td>28</td><td>5.02986</td><td>0.17964</td><td></td><td></td></tr><tr><td>Corrected Total</td><td>32</td><td>14.74721</td><td></td><td></td><td></td></tr></table>	Analysis of Variance						Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	Model	4	9.71735	2.42934	13.52	<.0001	Error	28	5.02986	0.17964			Corrected Total	32	14.74721									
Analysis of Variance																																					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F																																
Model	4	9.71735	2.42934	13.52	<.0001																																
Error	28	5.02986	0.17964																																		
Corrected Total	32	14.74721																																			
T	<table><tr><th>Variable</th><th>Parameter Estimate</th><th>Standard Error</th><th>Type II SS</th><th>F Value</th><th>Pr > F</th></tr><tr><td>Intercept</td><td>2.95828</td><td>1.36361</td><td>0.84547</td><td>4.71</td><td>0.0387</td></tr><tr><td>x1</td><td>-0.12932</td><td>0.02129</td><td>6.62976</td><td>36.91</td><td><.0001</td></tr><tr><td>x2</td><td>-0.01878</td><td>0.05628</td><td>0.02001</td><td>0.11</td><td>0.7410</td></tr><tr><td>x3</td><td>-0.04621</td><td>0.20727</td><td>0.00893</td><td>0.05</td><td>0.8252</td></tr><tr><td>x4</td><td>0.20876</td><td>0.06703</td><td>1.74211</td><td>9.70</td><td>0.0042</td></tr></table>	Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F	Intercept	2.95828	1.36361	0.84547	4.71	0.0387	x1	-0.12932	0.02129	6.62976	36.91	<.0001	x2	-0.01878	0.05628	0.02001	0.11	0.7410	x3	-0.04621	0.20727	0.00893	0.05	0.8252	x4	0.20876	0.06703	1.74211	9.70	0.0042
Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F																																
Intercept	2.95828	1.36361	0.84547	4.71	0.0387																																
x1	-0.12932	0.02129	6.62976	36.91	<.0001																																
x2	-0.01878	0.05628	0.02001	0.11	0.7410																																
x3	-0.04621	0.20727	0.00893	0.05	0.8252																																
x4	0.20876	0.06703	1.74211	9.70	0.0042																																

피어슨 상관 계수, N = 33
H0: Rho=0 가정하에서 Prob > |r|

	x1	x2	x3	x4	x5
x1	1.00000 0.0618	0.32872 0.0618	0.16767 0.3510	0.05191 0.7742	-0.73081 <.0001
x2	0.32872 0.0618	1.00000	-0.14550 0.4191	0.18033 0.3153	-0.21204 0.2362
x3	0.16767 0.3510	-0.14550 0.4191	1.00000	0.24134 0.1760	-0.05541 0.7594
x4	0.05191 0.7742	0.18033 0.3153	0.24134 0.1760	1.00000	0.31267 0.0765
x5	-0.73081 <.0001	-0.21204 0.2362	-0.05541 0.7594	0.31267 0.0765	1.00000

Backward Elimination: Step 1

Variable x3 Removed: R-Square = 0.6583 and C(p) = 3.0497

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	9.70842	3.23614	18.63	<.0001
Error	29	5.03879	0.17375		
Corrected Total	32	14.74721			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.67252	0.45798	5.91674	34.05	<.0001
x1	-0.13047	0.02031	7.16849	41.26	<.0001
x2	-0.01542	0.05332	0.01454	0.08	0.7745
x4	0.20450	0.06320	1.81907	10.47	0.0030

Backward Elimination: Step 2

Variable x2 Removed: R-Square = 0.6573 and C(p) = 1.1306

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	9.69388	4.84694	28.77	<.0001
Error	30	5.05332	0.16844		
Corrected Total	32	14.74721			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.55204	0.18741	31.23603	185.44	<.0001
x1	-0.13238	0.01891	8.25219	48.99	<.0001
x4	0.20134	0.06129	1.81767	10.79	0.0026

Bounds on condition number: 1.0027, 4.0108

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination							
Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x3	3	0.0006	0.6583	3.0497	0.05	0.8252
2	x2	2	0.0010	0.6573	1.1306	0.08	0.7745

결과해
석

5번 변수를 종속변수로 하고 나머지 1,2,3,4번 변수를 독립변수로 하는 다중선형회귀모형을 적합시켰다. 회귀식은 $y=2.9583-0.12932x_1-0.018785x_2-0.046215x_3+0.20876x_4$ 로 적합된다. 회귀분석 자료에 대해 분산분석을 실시한 결과 p-value가 매우 유의한 값이 나와 $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ 라는 귀무가설을 기각하여 적어도 하나의 회귀계수는 0이 아니라고 판단할 수 있다. 또한 결정계수는 0.65893으로 총 변동이 해당 회귀식에 의해 설명되는 부분이 약 66%정도인 것을 알 수 있다.

Example. 20.2

EXAMPLE 20.2 A Matrix of Partial Correlation Coefficients, as It Might Appear as Computer Output (from the Data of Example 20.1a)

	1	2	3	4	5
1	1.00000	0.19426	0.12716	0.33929	-0.75406
2	0.19426	1.00000	-0.26977	0.23500	-0.06296
3	0.12716	-0.26977	1.00000	0.26630	-0.04210
4	0.33929	0.23500	0.26630	1.00000	0.50720
5	-0.75406	-0.06296	-0.04210	0.50720	1.00000

R

CODE	<pre>#ex20.2 x=data.frame(x1,x2,x3,x4,x5) library(ppcor) pcor(x)</pre>
OUTPUT	<pre>> pcor(x) \$estimate x1 x2 x3 x4 x5 x1 1.0000000 0.19426472 0.12716260 0.3392934 -0.75406153 x2 0.1942647 1.00000000 -0.26976905 0.2349965 -0.06295562 x3 0.1271626 -0.26976905 1.00000000 0.2662979 -0.04209985 x4 0.3392934 0.23499646 0.26629789 1.0000000 0.50720111 x5 -0.7540615 -0.06295562 -0.04209985 0.5072011 1.00000000</pre>

SAS

CODE	<pre>proc corr data=ex20_1; var x1 x2; partial x3 x4 x5; run; proc corr data=ex20_1; var x1 x3; partial x2 x4 x5; run; proc corr data=ex20_1; var x1 x4; partial x2 x3 x5; run; proc corr data=ex20_1; var x1 x5; partial x3 x4 x2; run; proc corr data=ex20_1; var x2 x3; partial x1 x4 x5; run; proc corr data=ex20_1; var x2 x4; partial x1 x3 x5; run; proc corr data=ex20_1; var x2 x5; partial x1 x3 x4; run; proc corr data=ex20_1;</pre>
------	--

	<pre>var x3 x4; partial x1 x2 x5; run; proc corr data=ex20_1; var x3 x5; partial x1 x2 x4; run; proc corr data=ex20_1; var x4 x5; partial x1 x2 x3; run;</pre>																																																												
OUTPUT	<table><tr><td colspan="3">피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r </td></tr><tr><td></td><td>x1</td><td>x2</td></tr><tr><td>x1</td><td>1.00000</td><td>0.19426 0.3036</td></tr><tr><td>x2</td><td>0.19426 0.3036</td><td>1.00000</td></tr><tr><td colspan="3">피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r </td></tr><tr><td></td><td>x1</td><td>x3</td></tr><tr><td>x1</td><td>1.00000</td><td>0.12716 0.5031</td></tr><tr><td>x3</td><td>0.12716 0.5031</td><td>1.00000</td></tr><tr><td colspan="3">피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r </td></tr><tr><td></td><td>x1</td><td>x4</td></tr><tr><td>x1</td><td>1.00000</td><td>0.33929 0.0666</td></tr><tr><td>x4</td><td>0.33929 0.0666</td><td>1.00000</td></tr><tr><td colspan="3">피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r </td></tr><tr><td></td><td>x1</td><td>x5</td></tr><tr><td>x1</td><td>1.00000</td><td>-0.75406 <.0001</td></tr><tr><td>x5</td><td>-0.75406 <.0001</td><td>1.00000</td></tr><tr><td colspan="3">피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r </td></tr><tr><td></td><td>x2</td><td>x3</td></tr><tr><td>x2</td><td>1.00000</td><td>-0.26977 0.1494</td></tr><tr><td>x3</td><td>-0.26977 0.1494</td><td>1.00000</td></tr></table>	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r				x1	x2	x1	1.00000	0.19426 0.3036	x2	0.19426 0.3036	1.00000	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r				x1	x3	x1	1.00000	0.12716 0.5031	x3	0.12716 0.5031	1.00000	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r				x1	x4	x1	1.00000	0.33929 0.0666	x4	0.33929 0.0666	1.00000	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r				x1	x5	x1	1.00000	-0.75406 <.0001	x5	-0.75406 <.0001	1.00000	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r				x2	x3	x2	1.00000	-0.26977 0.1494	x3	-0.26977 0.1494	1.00000
피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r																																																													
	x1	x2																																																											
x1	1.00000	0.19426 0.3036																																																											
x2	0.19426 0.3036	1.00000																																																											
피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r																																																													
	x1	x3																																																											
x1	1.00000	0.12716 0.5031																																																											
x3	0.12716 0.5031	1.00000																																																											
피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r																																																													
	x1	x4																																																											
x1	1.00000	0.33929 0.0666																																																											
x4	0.33929 0.0666	1.00000																																																											
피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r																																																													
	x1	x5																																																											
x1	1.00000	-0.75406 <.0001																																																											
x5	-0.75406 <.0001	1.00000																																																											
피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r																																																													
	x2	x3																																																											
x2	1.00000	-0.26977 0.1494																																																											
x3	-0.26977 0.1494	1.00000																																																											

	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r 		
		x2	x4
	x2	1.00000 0.2113	0.23500 0.2113
	x4	0.23500 0.2113	1.00000
	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r 		
		x2	x5
	x2	1.00000	-0.06296 0.7410
	x5	-0.06296 0.7410	1.00000
	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r 		
		x3	x4
	x3	1.00000	0.26630 0.1549
	x4	0.26630 0.1549	1.00000
	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r 		
		x3	x5
	x3	1.00000	-0.04210 0.8252
	x5	-0.04210 0.8252	1.00000
	피어슨 부분 상관 계수, N = 33 H0: 부분 Rho=0 가정하에서 Prob > r 		
		x4	x5
	x4	1.00000	0.50720 0.0042
	x5	0.50720 0.0042	1.00000
결과해석	변수 1,2,3,4,5에 대해 나머지 3개의 변수의 영향을 보정한 부분상관계수를 구하였다. 일반적으로 구한 상관계수와 비교했을 때 각 변수들끼리 상관계수가 증가한 것도 있고 감소한 것도 있는 것을 확인할 수 있다.		

Example. 20.3

EXAMPLE 20.3 The Standard Error of a Predicted Y

For the equation $\hat{Y} = 2.552 - 0.132X_1 + 0.201X_2$, derived from the data of Example 20.1a, where X_1 is the variable in column 1 of the data matrix, X_2 is the variable in column 4, and Y is the variable in column 5, we obtain the following quantities needed to solve Equation 20.37:

$$\begin{aligned}s_{Y.1,2}^2 &= 0.16844, \quad n = 33, \quad \bar{X} = 4.4546, \\ \bar{X}_2 &= 2.5424, \quad \sum x_1^2 = 472.18, \quad \sum x_2^2 = 44.961, \\ d_{11} &= 1.0027, \quad d_{12} = -0.052051, \quad d_{21} = -0.052051, \quad d_{22} = 1.0027.\end{aligned}$$

By employing Equation 20.9, each d_{ik} is converted to a c_{ik} , resulting in

$$c_{11} = 0.0021236, \quad c_{12} = -0.00035724, \quad c_{21} = -0.00035724, \quad c_{22} = 0.022302.$$

What is the mean population value of Y at $X_1 = 7^\circ\text{C}$ and $X_4 = 2.0\text{ min}$?

$$\hat{Y} = 2.552 - (0.132)(7) + (0.201)(2.0) = 2.030\text{ ml}$$

What is the standard error of the mean population value of Y at $X_1 = 7^\circ\text{C}$ and $X_4 = 2.0\text{ min}$? [Equation 20.37 is used.]

$$\begin{aligned}s_Y^2 &= 0.16844 \left[\frac{1}{33} + (0.0021236)(7 - 4.4546)^2 \right. \\ &\quad + (-0.00035724)(7 - 4.4546)(2.0 - 2.5424) \\ &\quad + (-0.00035724)(2.0 - 2.5424)(7 - 4.4546) \\ &\quad \left. + (0.022302)(2.0 - 2.5424)^2 \right] \\ &= 0.16844 \left(\frac{1}{33} + 0.0213066 \right) \\ &= 0.008693\text{ ml}^2 \\ s_{\hat{Y}} &= \sqrt{0.008693\text{ ml}^2} = 0.093\text{ ml}\end{aligned}$$

As $t_{0.05(2),30} = 2.042$, the 95% prediction interval for the predicted Y is $2.030 \pm (2.042)(0.093)\text{ ml} = 2.030 \pm 0.190\text{ ml}$.

What is the predicted value of one additional Y value taken from the population at $X_1 = 7^\circ\text{C}$ and $X_4 = 2.0\text{ min}$?

$$\hat{Y} = 2.552 - (0.132)(7) + (0.201)(2.0) = 2.030\text{ ml}$$

What is the standard error of the predicted value of one additional Y value taken from the population at $X_1 = 7^\circ\text{C}$ and $X_4 = 2.0\text{ min}$? [Equation 20.39 is used.]

$$\begin{aligned}s_{\hat{Y}} &= \sqrt{0.16844 \left[1 + \frac{1}{33} + 0.0213066 \right]} \\ &= 0.421\text{ ml}\end{aligned}$$

	<p>As $t_{0.05(2),30} = 2.042$, the 95% prediction interval for the preceding predicted \hat{Y} is $2.03 \pm (2.042)(0.421)$ ml = 2.03 ± 0.86 ml.</p> <p>What is the predicted value of the mean of 10 additional values of Y taken from the population at $X_1 = 7^\circ\text{C}$ and $X_4 = 2.0$ min?</p> $\hat{Y} = 2.552 - (0.132)(7) + (0.201)(2.0) = 2.030 \text{ ml}$ <p>What is the standard error of the predicted value of the mean of 10 additional values of Y taken from the population at $X_1 = 7^\circ\text{C}$ and $X_4 = 2.0$ min? [Equation 20.40 is used.]</p> $s_{\hat{Y}} = \sqrt{0.16844 \left[\frac{1}{10} + \frac{1}{33} + 0.0213066 \right]} = 0.16 \text{ ml}$ <p>As $t_{0.05(2),30} = 2.042$, the 95% prediction interval for the predicted Y is $2.03 \pm (2.042)(0.16)$ ml = 2.03 ± 0.33 ml.</p>
--	---

R

CODE	<pre>#ex20.3 rx=data.frame(x5,x1,x4) newdata=data.frame(x1=7, x4=2) mdl=lm(x5~x1+x4) summary(mdl) predict(mdl, newdata, interval="predict")</pre>
OUTPUT	<pre>> summary(mdl) Call: lm(formula = x5 ~ x1 + x4) Residuals: Min 1Q Median 3Q Max -1.48843 -0.09287 0.03683 0.14880 1.01484 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 2.55204 0.18741 13.618 2.22e-14 *** x1 -0.13238 0.01891 -6.999 8.89e-08 *** x4 0.20134 0.06129 3.285 0.0026 ** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.4104 on 30 degrees of freedom Multiple R-squared: 0.6573, Adjusted R-squared: 0.6345 F-statistic: 28.77 on 2 and 30 DF, p-value: 1.054e-07 > predict(mdl, newdata, interval="predict") fit lwr upr 1 2.028069 1.168523 2.887615</pre>

SAS

CODE	<pre>data ex20_3; set ex20_1; keep x1 x4 x5; run; data newdata;</pre>
------	---

	<pre> input x1 x4; cards; 7 2 ;run; data pred; set ex20_3 newdata; run; proc reg data=pred; model x5=x1 x4 /p cli clm; run; </pre>							
OUTPUT	34	.	2.0281	0.0932	1.8377	2.2185	1.1685	2.8876
결과해 석	<p>예제 20.1의 자료에서 5번째 변수를 종속변수로, 1, 4번째 변수들을 독립변수로 하는 다중선형회귀모형을 적합시켰다. 이 때 모집단에서 $x_1=7$, $x_4=2$ 인 하나의 자료가 추가하였을 때의 y값에 대한 예측구간을 구하였더니 $[1.168, 2.888]$임을 확인할 수 있었다.</p>							

Example. 20.4	
<p>EXAMPLE 20.4 Comparing Multiple Regressions</p> <p>Let us consider three multiple regressions, each fitted to a different sample of data, and each containing the same dependent variable and the same four independent variables. (Therefore, $m = 4$ and $k = 3$.) The residual sums of squares from each of the regressions are 437.8824, 449.2417, and 411.3548, respectively.</p> <p>If the residual degrees of freedom for each of the regressions are 41, 32, and 38, respectively (that is, the three sample sizes were 46, 37, and 43, respectively), then the pooled residual sum of squares, SS_p, is 1298.4789, and the pooled residual degrees of freedom, DF_p, is 111.</p> <p>Then, we combine the 126 data from all three samples and fit to these data a multiple regression having the same variables as the three individual regressions fitted previously. From this multiple regression let us say we have a total residual sum of squares, SS_t, of 1577.3106. The total residual degrees of freedom, DF_t, is 121.</p> <p>Then we test H_0: All three sample regression functions estimate the same population regression, against H_A: All three sample regression functions do not estimate the same population regression:</p> $F = \frac{\frac{SS_t - SS_p}{(m + 1)(k - 1)}}{\frac{SS_p}{DF_p}}$ $= \frac{\frac{1577.3106 - 1298.4789}{(5)(2)}}{\frac{1298.4789}{111}}$ $= 2.38.$ <p>The degrees of freedom associated with F are 10 and 111.</p> <p>Since $F_{0.05(1),10,111} \approx 1.93$, reject H_0.</p> <p style="text-align: center;">$0.01 < P < 0.025 \quad [P = 0.013]$</p>	

R						
CODE	<pre>#ex20.4 m=4 k=3 ssep=1298.4789 sset=1577.3106 v1=41 v2=32 v3=38 vp=111 vt=121 nof=(sset-ssep)/((m+1)*(k-1)) denof=ssep/vp f=nof/denof fp=qf(0.05,10,111, lower.tail = F) if(f>fp){ answer= "reject H0" }else{ answer= "cannot reject H0" } pv= round(pf(f,10,111, lower.tail = F),5) print(c(answer, paste("p-value=", pv)))</pre>					
OUTPUT	<pre>> print(c(answer, paste("p-value=", pv))) [1] "reject H0" "p-value= 0.01341"</pre>					
SAS						
CODE	<pre>proc iml; m=4; k=3; ssep=1298.4789; sset=1577.3106; vp=111; vt=121; nof=(sset-ssep)/((m+1)*(k-1)); denof=ssep/vp; f=nof/denof; fp= finv(0.95,10,111); if f > fp then answer= "reject H0"; else answer= "cannot reject H0"; pv= 1-cdf('F', f, 10, 111); print answer pv; run; quit;</pre>					
OUTPUT	<table><tr><th>answer</th><th>pv</th></tr><tr><td>reject H0</td><td>0.0134116</td></tr></table>	answer	pv	reject H0	0.0134116	
answer	pv					
reject H0	0.0134116					
결과해석	각각의 같은 독립변수, 종속변수에서 서로 다른 표본을 3개 뽑아 각각 회귀식에 적합하였다. 그 다음 세 개의 표본을 모두 합쳐 하나의 자료로 만든 후 해당 자					

	<p>료에 대해 회귀식에 적합한 후 세 개의 회귀식이 통합된 회귀식과 같은지 확인해 보려고 한다.</p> <p>귀무가설: 세 개의 회귀식은 모두 같은 하나의 모회귀식을 추정한다.</p> <p>대립가설: 세 회귀식 모두가 하나의 모회귀식을 추정하는 것은 아니다.</p> <p>검정 결과 $P\text{-value}=0.013$ 정도로 유의하여 귀무가설을 기각한다. 따라서 세 회귀식 모두가 하나의 모회귀식을 추정한다고 볼 수는 없다.</p>
--	--

EXAMPLE 20.5 Kendall's Coefficient of Concordance

H_0 : In the sampled population, there is no association among the three variables (wing, tail, and bill lengths).

H_0 : In the sampled population, there is a relationship among wing, tail, and bill lengths.

Birds (i)	Wing Length (cm)		Tail Length (cm)		Bill Length (mm)		Sums of ranks (R_i)
	Data	Ranks	Data	Ranks	Data	Ranks	
1	10.4	4	7.4	5	17	5.5	14.5
2	10.8	8.5	7.6	7	17	5.5	21
3	11.1	10	7.9	11	20	9.5	30.5
4	10.2	1.5	7.2	2.5	14.5	2	6
5	10.3	3	7.4	5	15.5	3	11
6	10.2	1.5	7.1	1	13	1	3.5
7	10.7	7	7.4	5	19.5	8	20
8	10.5	5	7.2	2.5	16	4	11.5
9	10.8	8.5	7.8	9.5	21	11	29
10	11.2	11	7.7	8	20	9.5	28.5
11	10.6	6	7.8	9.5	18	7	22.5
12	11.4	12	8.3	12	22	12	36

$$M = 3$$

$$n = 12$$

Without correction for ties:

$$W = \frac{\sum R_i^2 - \frac{(\sum R_i)^2}{n}}{M^2(n^3 - n)}$$

$$= \frac{(14.5^2 + 21^2 + 30.5^2 + \dots + 36^2) - \frac{(14.5 + 21 + 30.5 + \dots + 36)^2}{12}}{\frac{3^2(12^3 - 12)}{12}}$$

$$= \frac{5738.5 - \frac{(234)^2}{12}}{\frac{15444}{12}}$$

$$= \frac{1175.5}{1287} = 0.913$$

$$\chi_r^2 = M(n - 1)W$$

$$= (3)(12 - 1)(0.913)$$

$$= 30.129$$

<p>From Appendix Table B.14, $(\chi^2_{\tau})_{0.05,3,12} = 6.167$. Reject H_0: $P \ll 0.001$.</p> <p>Incorporating the correction for ties:</p> <p>In group 1 (wing length): there are 2 data tied at 10.2 cm (i.e., $t_1 = 2$); there are 2 data tied at 10.8 cm (i.e., $t_2 = 2$).</p> <p>In group 2 (tail length): there are 2 data tied at 7.2 cm (i.e., $t_3 = 2$); there are 3 data tied at 7.4 cm (i.e., $t_4 = 3$); there are 2 data tied at 7.8 cm (i.e., $t_5 = 2$).</p> <p>In group 3 (bill length): there are 2 data tied at 17 mm (i.e., $t_6 = 2$); there are 2 data tied at 20 mm (i.e., $t_7 = 2$).</p> <p>Considering all seven groups of ties,</p> $\sum t = \sum_{i=1}^7 (t_i^3 - t_i)$ $= (2^3 - 2) + (2^3 - 2) + (2^3 - 2) + (3^3 - 3) + (2^3 - 2) + (2^3 - 2) + (2^3 - 2) = 60$ <p>and</p> $W_c = \frac{1175.5}{\frac{15444 - 3(60)}{12}} = \frac{1175.5}{1272} = 0.924.$ <p>Then, to test the significance of W_c:</p> $(\chi^2_{\tau})_c = M(n - 1)W_c$ $= (3)(12 - 1)(0.924) = 30.492.$ <p>For these data, the same conclusion is reached with W_c as with W, namely: Reject H_0; and $P \ll 0.001$.</p>	
R	
CODE	<pre>#ex20.5 wing=c(10.4,10.8,11.1,10.2,10.3,10.2,10.7,10.5,10.8,11.2,10.6,11.4) tail=c(7.4,7.6,7.9,7.2,7.4,7.1,7.4,7.2,7.8,7.7,7.8,8.3) bill=c(17,17,20,14.5,15.5,13,19.5,16,21,20,18,22) x=data.frame(wing, tail, bill) library(irr) kendall(x) kendall(x, correct=T)</pre>

<p>OUTPUT T</p>	<pre> > kendall(x) kendall's coefficient of concordance w Subjects = 12 Raters = 3 W = 0.913 Chisq(11) = 30.1 p-value = 0.00151 Coefficient may be incorrect due to ties > kendall(x, correct=T) kendall's coefficient of concordance wt Subjects = 12 Raters = 3 wt = 0.924 Chisq(11) = 30.5 p-value = 0.00132 </pre>
SAS	
<p>CODE</p>	<pre> data ex20_5; do s=1 to 12; do r= "wing", "tail", "bill"; input y @@; output; end; end; cards; 10.4 7.4 17 10.8 7.6 17 11.1 7.9 20 10.2 7.2 14.5 10.3 7.4 15.5 10.2 7.1 13 10.7 7.4 19.5 10.5 7.2 16 10.8 7.8 21 11.2 7.7 20 10.6 7.8 18 11.4 8.3 22 ; run; %macro magree(data=_last_, items=, raters=, response=, stat=BOTH); %let error=0; %if %upcase(&stat) ne KAPPA and %upcase(&stat) ne KENDALL and %upcase(&stat) ne BOTH %then %do; %put ERROR: STAT= must be set to KAPPA, KENDALL, or BOTH.; %goto exit; %end; options nonotes; *title2 "MAGREE macro"; %if &data=_last_ %then %let data=&syslast; /* Get type of response variable so can format later. ===== =====*/ %let dsid=%sysfunc(open(&data)); %if &dsid %then %do; %let ynum=%sysfunc(varnum(&dsid,&response)); %let ytype=%sysfunc(vartype(&dsid,&ynum)); </pre>

```

%let rc=%sysfunc(close(&dsid));
%end;
%else %do;
  %put WARNING: Could not check type of &response variable.;
  %put %str(          ) Continuing assuming it is numeric.;
%end;

/* Verify that all subjects have the same number of ratings. Note
 * that the raters variable must use the same values to identify
the
 * raters in all subjects even if different raters rate the
subjects.
 * But the number of raters must be the same across all subjects.
 * Create response by item summary table needed to compute kappas
and
 * get numbers of subjects, raters, and response categories.

=====
=====*/
data _nomiss; set &data; if &response ne " "; run;

proc freq data=_nomiss noprint;
  table &items*&raters / sparse out=_balance;
  table &response*&items / out=_ycnts(drop=percent);
  table &items / out=_n;
  table &raters / out=_m;
  table &response / out=_k(drop=count percent);
run;

data _null_;
  set _balance nobs=_nobs;
  if count ne 1 then call symput('error',1);
  run;
%if &error %then %do;
  %put ERROR: Each rater must rate each subject exactly once.;
  %goto exit;
%end;

data _null_;
  set _n nobs=n;
  set _m nobs=m;
  set _k nobs=k;
  call symput('m',left(m));
  call symput('n',left(n));
  call symput('k',left(k));
  run;

/***** Compute kappa
*****/
%if %upcase(&stat)=KAPPA or %upcase(&stat)=BOTH %then %do;

/* Create coded values (1,2,3,...) for response categories to use
as
 * indices.

=====
=====*/
data _ycnts;
  set _ycnts;

```

```

by &response;
if first.&response then _code+1;
run;

/* Compute kappa statistics for each category and overall
=====
=====*/
data _kappas;
set _ycnts end=eof;
array kapnum {&k} knum1-ksnum&k;
array catsum {&k} sum1-sum&k;
array kapj {&k} kappal-kappa&k; /* V6 limit: 999 categories */
array pb {&k} pbar1-pbar&k;
array zkapj {&k} zkjl-zkj&k;
array prkapj {&k} prkjl-prkj&k;
kapnum[_code] + count*(&m - count);
catsum[_code] + count;
if eof then do;
  knum=0; kden=0; pqqp=0;
  nmm = &n*&m*(&m-1);
  sekapj=sqrt(2/nmm);
  do j=1 to &k;
    pb{j} = catsum{j}/(&m*&n);
    pq = pb{j}*(1-pb{j});
    kapj{j} = 1 - (kapnum{j}/(nmm*pq));
    zkapj{j} = kapj{j}/sekapj;
    prkapj{j} = 1-probnorm(zkapj{j});
    knum = knum + pq*kapj{j};
    kden = kden + pq;
    pqqp = pqqp + pq*((1-pb{j})-pb{j});
  end;
  kappa=ksnum/kden;
  sekap = (sqrt(2)/(kden*sqrt(nmm)))*sqrt(kden**2-pqqp);
  zkap=kappa/sekap; prkap=1-probnorm(zkap);
  keep &response kapp stderr z prob;
  do i=1 to &k;
    set _k;
    kapp=kapj{i}; stderr=sekapj; z=zkapj{i}; prob=prkapj{i};
  output;
  end;
  &response=.; kapp=kappa; stderr=sekap; z=zkap; prob=prkap;
  output;
end;
run;

/* Print kappa statistics
=====
=====*/
%let fexist=0;
data _null_;
  if ("%ytype"="N" and cexist("work.formats._yfmt.format") ne 1)
or
  ("%ytype"="C" and cexist("work.formats._yfmt.formatc") ne 1)
then
  call symput("fexist","1");
run;
%if &fexist %then %do;
proc format;

```

	<pre> value %if &ytype=C %then \$_yfmt; %else _yfmt; .="Overall"; run; %end; proc print noobs label; format prob pvalue. &response %if &ytype=C %then \$_yfmt.; %else _yfmt.; ; label kapp="Kappa" stderr="Standard Error" prob="Prob>Z"; title2 "Kappa statistics for nominal response"; run; %end; /***** Compute Kendall's Coefficient of Concordance, W *****/ %if %upcase(&stat)=KENDALL or %upcase(&stat)=BOTH %then %do; %if &ytype=C %then %do; %put %str(NOTE: Kendall%'s Coefficient of Concordance requires a); %put %str() numeric, ordinal response.; %goto exit; %end; /* Rank the data using average ties. If data are already ranked, this * won't change anything. ===== =====*/ proc sort data=&data out=_sortr; by &raters; run; proc rank data=_sortr out=_ranked; by &raters; var &response; run; /* R-square from one-way ANOVA is Kendall's W. ===== =====*/ proc anova data=_ranked outstat=_anova noprint; class &items; model &response = &items; run; /* Compute F statistic and p-value for testing W=0. Ties are handled correctly. ===== =====*/ data _w; set _anova; retain SSsubj; if _n_=2 then do; w=ss/(ss+SSsubj); f=(&m-1)*w/(1-w); numdf=&n-1-2/&m; </pre>
--	---

	<pre>dendf=(&m-1)*numdf; prob=1-probf(f,numdf,dendf); keep w f numdf dendf prob; output; end; SSsubj=ss; run; /* Print Kendall's coefficient of concordance and test ===== =====*/ proc print noobs label; var w f numdf dendf prob; format prob pvalue.; label w="Coeff of Concordance" numdf="Num DF" dendf="Denom DF" prob="Prob>F"; title2 "Kendall's Coefficient of Concordance for ordinal response"; run; %end; %exit: options notes; title2; %mend magree; %magree(data=ex20_5, items=s, raters=r, response=y,stat=KENDALL)</pre>										
OUTPUT	<p>Kendall's Coefficient of Concordance for ordinal response</p> <table><tr><th>Coeff of Concordance</th><th>F</th><th>Num DF</th><th>Denom DF</th><th>Prob>F</th></tr><tr><td>0.92414</td><td>24.3627</td><td>10.3333</td><td>20.6667</td><td><.0001</td></tr></table>	Coeff of Concordance	F	Num DF	Denom DF	Prob>F	0.92414	24.3627	10.3333	20.6667	<.0001
Coeff of Concordance	F	Num DF	Denom DF	Prob>F							
0.92414	24.3627	10.3333	20.6667	<.0001							
결과해 석	<p>12마리의 새들에 따른 날개, 꼬리, 부리 길이 세가지 변수가 서로 관련이 있는지 Kendall의 일치도 계수를 통해 확인해 보려고 한다.</p> <p>귀무가설: 12마리의 새들의 날개, 꼬리, 부리 길이의 관련성이 서로 일치하지 않는다.</p> <p>대립가설: 12마리의 새들의 날개, 꼬리, 부리 길이의 관련성이 서로 일치한다.</p> <p>검정결과 일치도 계수는 0.924로 높게 나왔으며 p-value 또한 매우 유의하게 나와 귀무가설을 기각한다. 따라서 12마리의 새들의 날개, 꼬리, 부리 길이의 연관성이 있다고 판단할 수 있다.</p>										

Example. 20.6

EXAMPLE 20.6 Kendall's Coefficient of Concordance Used to Assess Agreement

Each of three girls ranked her taste preference for each of six flavors of ice cream (chocolate-chip, chocolate, spumoni, vanilla, butter-pecan, Neapolitan.)

H_0 : There is no agreement in flavor preference.

H_A : There is agreement in flavor preference.

Girl	Flavors (<i>i</i>)						
	CC	C	S	V	BP	N	
1	5	1	3	2	4	6	
2	6	2	3	1	5	4	
3	6	3	2	1	4	5	
Rank sum (R_i)	17	6	8	4	13	15	$\sum R_i = 63$

$$M = 3$$

$$n = 6$$

$$W = \frac{\sum R_i^2 - \frac{(\sum R_i)^2}{n}}{\frac{M^2(n^3 - n)}{12}}$$

$$= \frac{17^2 + 6^2 + 8^2 + 4^2 + 13^2 + 15^2 - \frac{63^2}{6}}{\frac{3^2(6^3 - 6)}{12}} = \frac{137.50}{157.50} = 0.873$$

$$\chi_r^2 = M(n - 1)W = (3)(6 - 1)(0.873) = 13.095$$

Using Appendix Table B.14, $(\chi_r^2)_{0.05,3,6} = 7.000$. Therefore, reject H_0 . The conclusion is that there is agreement in flavor preferences.

$$P < 0.001$$

R

CODE

```
#ex20.6
girl1=c(5,1,3,2,4,6)
girl2=c(6,2,3,1,5,4)
girl3=c(6,3,2,1,4,5)
x=data.frame(girl1,girl2,girl3)
kendall(x)
```

OUTPUT

```
> kendall(x)
kendall's coefficient of concordance w

Subjects = 6
Raters = 3
w = 0.873

chisq(5) = 13.1
p-value = 0.0225
```

SAS

	$C_T = \frac{1}{M^2(n - S_1)} \left(\sum_{i=1}^n R_i^2 - M^2 n \right)$ $= \frac{1}{3^2(6 - 2.450)} [0.701^2 + 4.850^2 + 3.350^2 + 6.350^2 + 1.601^2 + 1.151^2 - (3^2)(6)]$ $= 0.03130[79.4469 - 54] = 0.03130(25.4469) = 0.7965$ $\chi_T^2 = 3(6 - 1)C_T$ $= (15)(0.7965) = 11.948$ $\nu = n - 1 = 5$ $\chi_{0.05,5}^2 = 11.070$ <p>Reject H_0.</p> $0.025 < P < 0.05 \quad [P = 0.036]$	
R		
CODE	<pre>#ex20.7 ##### library(coin) cscores.default(girl1,type = "Savage",int=F) g1=c(0.367,2.450,0.950,1.450,0.617,0.167) g2=c(0.167,1.450,0.950,2.450,0.367,0.617) g3=c(0.167,0.950,1.450,2.450,0.617,0.367) x=data.frame(g1,g2,g3) kendall(x) ##### r=g1+g2+g3 m=3 c=1/(m^2*(length(r)-2.450))*(sum(r^2)-m^2*length(r)) chi=m*(length(r)-1)*c chip=qchisq(0.05,5, lower.tail = F) if(chi>chip){ answer= "reject H0" }else{ answer= "cannot reject H0" } pv= round(pchisq(chi,5, lower.tail = F),5) print(c(answer, paste("p-value=", pv)))</pre>	

가	Savage score	가
.	rank	.
.		.

OUTPUT	<pre>> cscores.default(girl1, [1] -0.4500000 0.8333333 attr(,"scores") [1] "Savage" 원자료가 없어서 savage sco > kendall(x) kendall's coefficient of concordance W Subjects = 6 Raters = 3 W = 0.873 chisq(5) = 13.1 p-value = 0.0225 일반 Kendall's W로 구했을 때는 값에 차이가 있다. > print(c(answer, paste("p-value=", pv))) [1] "reject H0" "p-value= 0.03552"</pre>			
SAS				
CODE	<pre>proc iml; g1={0.367,2.450,0.950,1.450,0.617,0.167}; g2={0.167,1.450,0.950,2.450,0.367,0.617}; g3={0.167,0.950,1.450,2.450,0.617,0.367}; r= g1+g2+g3; m=3; n=nrow(r); c=1/(m**2*(n-2.450))*(sum(r#r)-m**2*n); chi=m*(n-1)*c; chip=cinv(0.95,n-1); if chi > chip then answer= "reject H0"; else answer= "cannot reject H0"; pv= 1-cdf('chisquared',chi,n-1); print chi chip answer pv; run;quit;</pre>			
OUTPUT	chi	chip	answer	pv
	11.946903	11.070498	reject H0	0.0355226
결과해석	<p>3명의 소녀의 6가지 아이스크림 맛의 선호하는 정도에 따른 순위가 일치하는지 확인해 보려고 한다. 따라서 여기서는 6가지 맛에 대한 단순한 선호 순위가 아닌 선호하는 순위차이의 간격까지 고려한 Savage score를 이용하여 Kendall의 일치도 계수를 구하였다.</p> <p>귀무가설: 3명의 소녀의 아이스크림 맛을 좋아하는 정도를 고려한 선호 순위가 일치하지 않는다.</p> <p>대립가설: 3명의 소녀의 아이스크림 맛을 좋아하는 정도를 고려한 선호 순위가 일치한다.</p> <p>검정결과 p-value=0.036정도로 유의수준 0.05하에서 귀무가설을 기각할 충분한 근거가 된다. 따라서 3명의 소녀의 아이스크림 맛을 좋아하는 정도를 고려한 선호 순위가 일치한다고 결론을 내릴 수 있다.</p>			