ANOVA

Chapter 13 from Walpole book

Assumptions and Hypotheses in One-Way ANOVA

It is assumed that the k populations are independent and normally distributed with means $\mu_1, \mu_2, \ldots, \mu_k$ and common variance σ^2. As indicated in Section 13.2, these assumptions are made more palatable by randomization. We wish to derive appropriate methods for testing the hypothesis

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_k,$$
$$H_1\colon\ \text{At least two of the means are not equal.}$$

Let y_{ij} denote the jth observation from the ith treatment and arrange the data as in Table 13.2. Here, $Y_{i.}$ is the total of all observations in the sample from the ith treatment, $\bar{y}_{i.}$ is the mean of all observations in the sample from the ith treatment, $Y_{..}$ is the total of all nk observations, and $\bar{y}_{..}$ is the mean of all nk observations.

Table 13.2: k Random Samples

Treatment:	**1**	**2**	$\cdots$	$\boldsymbol{i}$	$\cdots$	$\boldsymbol{k}$	
	y_{11}	y_{21}	$\cdots$	y_{i1}	$\cdots$	y_{k1}	
	y_{12}	y_{22}	$\cdots$	y_{i2}	$\cdots$	y_{k2}	
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	
	y_{1n}	y_{2n}	$\cdots$	y_{in}	$\cdots$	y_{kn}	
Total	$Y_{1.}$	$Y_{2.}$	$\cdots$	$Y_{i.}$	$\cdots$	$Y_{k.}$	$Y_{..}$
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\cdots$	$\bar{y}_{i.}$	$\cdots$	$\bar{y}_{k.}$	$\bar{y}_{..}$

Theorem 13.1: **Sum-of-Squares Identity**

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$$

Sum of Squares, Unequal Sample Sizes

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{..})^2,\ SSA = \sum_{i=1}^{k} n_i(\bar{y}_{i.}-\bar{y}_{..})^2,\ SSE = SST - SSA$$

Three Important Measures of Variability

$$\begin{aligned} SST &= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = \text{total sum of squares},\\ SSA &= n\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2 = \text{treatment sum of squares},\\ SSE &= \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 = \text{error sum of squares}. \end{aligned}$$

The sum-of-squares identity can then be represented symbolically by the equation

$$SST = SSA + SSE.$$

Neil. Weis book formula

Sum of squares	Defining formula	Computing formula
Total, SST	$\Sigma(x_i - \bar{x})^2$	$\Sigma x_i^2 - (\Sigma x_i)^2/n$
Treatment, $SSTR$	$\Sigma n_j(\bar{x}_j - \bar{x})^2$	$\Sigma(T_j^2/n_j) - (\Sigma x_i)^2/n$
Error, SSE	$\Sigma(n_j - 1)s_j^2$	$SST - SSTR$

$Where$

$$n = total\ no.of\ observations$$
$$n_j = size\ of\ jth\ sample$$
$$T_j = sum\ of\ jth\ sample$$

Table 13.3: Analysis of Variance for the One-Way ANOVA

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Computed f**
Treatments	SSA	$k-1$	$s_1^2 = \dfrac{SSA}{k-1}$	$\dfrac{s_1^2}{s^2}$
Error	SSE	$k(n-1)$	$s^2 = \dfrac{SSE}{k(n-1)}$	
Total	SST	$kn-1$		

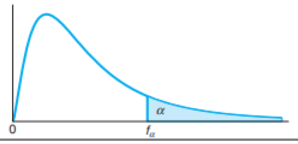

Table A.6 Critical Values of the *F*-Distribution

$f_{0.05}(v_1, v_2)$

v_2	v_1: 1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Table A.6 (continued) Critical Values of the *F*-Distribution

	$f_{0.05}(v_1, v_2)$									
	v_1									
v_2	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

*In Exercises **16.42–16.47**, we provide data from independent simple random samples from several populations. In each case,*

a. compute SST, SSTR, and SSE by using the computing formulas given in Formula 16.1 on page 726.

b. compare your results in part (a) for SSTR and SSE with those in Exercises 16.24–16.29, where you employed the defining formulas.

c. construct a one-way ANOVA table.

d. decide, at the 5% significance level, whether the data provide sufficient evidence to conclude that the means of the populations from which the samples were drawn are not all the same.

16.42

Sample 1	Sample 2	Sample 3
1	10	4
9	4	16
	8	10
	6	
	2	

16.43

Sample 1	Sample 2	Sample 3
8	2	4
4	1	3
6	3	6
		3

16.44

Sample 1	Sample 2	Sample 3	Sample 4
6	9	4	8
3	5	4	4
3	7	2	6
	8	2	
	6	3	

16.45

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
7	5	6	3	7
4	9	7	7	9
5	4	5	7	11
4		4	4	
		8	4	

16.46

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
4	8	9	4	3
2	5	6	0	6
3	5	9	2	9

16.47

Sample 1	Sample 2	Sample 3	Sample 4
11	9	16	5
6	2	10	1
7	4	10	3

16.47 (solution on excel)

1) $H_0: all\ four\ population\ means\ are\ equal$
2) $H_1: atleat\ two\ population\ means\ are\ not\ equal$
3) $\alpha = 0.05$
4) $Critical\ value\ and\ critical\ region$

$$F_{critical} = F_{0.05,v_1,v_2} = F_{0.05,3,8} = 4.07$$

$$critical\ region\ is\ F > 4.07$$

5) $Test\ Statistic: (F - test)$

Working below

	sample1	sample2	sample 3	sample4	total					
	11	9	16	5	41				grand mean	7
	6	2	10	1	19					
	7	4	10	3	24					
Total	24	15	36	9	84					
mean	8	5	12	3	28					
	SST							798-84^2/12		210
	121	81	256	25						
	36	4	100	1						
	49	16	100	9						
total	206	101	456	35	798					
							SST	210		
	SSA				total					
yi.^2	576	225	1296	81	2178					
	SSA	2178/3-84^2/12	138							
	SSE	210-138	72							

ANOVA Table

		sum of squares	d.f	mean ss	F-ratio
	SSA	138	3	46	5.11111111
	SSE	72	8	9	
Total	SST	210	11		

test statistic value		5.1111			critical value	4.07
		6) conclusion	reject H_0			