$Effect^2 * n_i$

Note that the sum of $Effect^2 * n_i$ over the four sites is:

43.71 + 37.05 + 34.66 + 61.83 = 177.25,

which is SS(Among).

Module 29: Review of ANOVA

This module reviews how one way ANOVA relates to the two-sample t test and as well as the general approach to ANOVA

Relationship to t-test

Two Samples of serum uric acid values

Sample 1	Sample 2
1.2	1.7
0.8	1.5
1.1	2.0
0.7	2.1
0.9	1.1
1.1	0.9
1.5	2.2
0.8	1.8
1.6	1.3
0.9	1.5

Total

$n = 20$

$\overline{x} = 1.335$

Sum $= 26.7$

$\sum x^2 = 39.65$

$SS = 4.01$

$(26.7)^2/20 = 35.6445$

$s^2 = 4.00/19 = 0.21$

Within

Sample 1	Sample 2
$\overline{x}_1 = 1.06$	$\overline{x}_2 = 1.61$
$SS_1 = 0.824$	$SS_2 = 1.669$
$\sum x_1^2 = 12.06$	$\sum x_2^2 = 27.59$
$\sum x = 10.60$	$\sum x = 16.10$
$(10.60)^2/10 = 11.236$	$(16.10)^2/10 = 25.92$
$s_1^2 = 0.09$	$s_2^2 = 0.19$

$$SS_{within} = SS_1 + SS_2 = 0.824 + 1.669 = 2.49$$

Among

$$(10.60)^2/10 + (16.10)^2/10 - (26.7)^2/20 = 1.52$$

Independent Random Samples from

Populations of Serum Uric Acid Values

	Sample 1	Sample 2	Both
	1.2	1.7	
	0.8	1.5	
	1.1	2.0	
	0.7	2.1	
	0.9	1.1	
	1.1	0.9	
	1.5	2.2	
	0.8	1.8	
	1.6	1.3	
	0.9	1.5	
n	10	10	20
Σ	10.6	16.1	26.7
$\overline{X}$	1.06	1.61	$\overline{\overline{X}} = 1.33$
SS	0.82	1.67	4.01
S^2	0.09	0.19	0.21
S	0.30	0.43	0.46
$\overline{X} - \overline{\overline{X}}$	-0.27	0.28	

$SS(Within) = SS_1 + SS_2 = 0.82 + 1.67 = 2.49$

$SS(Among) = (10.6)^2/10 + (16.1)^2/10 - (26.7)^2/20 = 37.16 - 35.64 = 1.52$

$SS(Total) = SS(Within) + SS(Among) = 4.00$

1. *The hypothesis :* $H_0: \mu_1 = \mu_2$

vs $H_1: \mu_1 \neq \mu_2$

2. *The assumptions :* Independent random samples normal distributions

$\sigma_1^2 = \sigma_2^2$

3. *The α- level :* $\alpha = 0.05$

4. *The test statistics :* ANOVA

5. *Rejection region* : Reject H_0 if

$F = MS(Among)/MS(Within) > F_{0.95}(1,18) = 4.41$

Where $MS(Among) = SS(Among)/df(Among)$

$MS(Within) = SS(Within)/df(Within)$

6. *The test result* :

		ANOVA		
Source	Df	SS	MS	F
Among	1	1.52	1.52	10.86
Within	18	2.49	0.14	
Total	19	4.01		

7. *Reject H_0* : Because $F = 1.52/0.14 = 10.86 > 4.41$

One-Way ANOVA--Pulmonary Function Example

A calibration evaluation of four machines that measure pulmonary function yielded, with the four machines being located at four sites,

Machine /Site			
1	2	3	4
NC	Jackson	Minn	Balt
433	445	434	441
435	440	436	443
432	438	433	438
439	441	437	439
436		434	442
		438	444
		440	
		435	

The numbers each represent one replication and are a computer generated count equivalent to one liter. A difference of 1% or more is not acceptable.

Consider the following two questions:

- Is there evidence that the four machines are not equally calibrated?
- If each machine is to be recalibrated with a common syringe designed for this purpose, how many replications would be required?

The ANOVA below addresses Question 1

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, versus H_1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$.

ANOVA

Source	df	SS	MS	F
Among	3	177.25	59.08	9.22
Within	19	121.71	6.41	
Total	22	298.96		

- The outcome for the above test is that the null hypothesis is rejected, thus there is evidence that the four machines are not properly calibrated.
- Hence it is important to undertake a recalibration procedure and thus the reason for question 2 above. That is, if it is important to do the recalibration, it is necessary to decide how many replications are to be done.

Question 2

A 1% difference between machines not being tolerable implies that we would want to detect a difference of 1% of about 400 units or about 4 units.

From the ANOVA table above, we can get a pooled estimate of the variance among replications, which is the MS(Within) = 6.41 units.

An acceptable strategy is to ascertain the sample size required for a two-sample t-test. Hence, the formula for determining the sample size is

$$N = n_1 + n_2 = \frac{4\sigma^2 (Z_{1-\alpha/2} + Z_{1-\beta})^2}{(d = \mu_1 - \mu_2)^2},$$

Or

$$N = n_1 + n_2 = \frac{4(6.41)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(4)^2},$$

which can be calculated for varying levels of power and also, perhaps for varying levels of variance.

Sample Size Calculations

$\alpha = 0.05$

Variance = 6.41

	0.5	0.8	0.85	0.9	0.95
d	0	0.842	1.036	1.282	1.645
3.5	8.0	16.4	18.8	22.0	27.2
4.0	6.2	12.6	14.4	16.8	20.8
4.5	4.9	9.9	11.4	13.3	16.5
5.0	3.9	8.1	9.2	10.8	13.3

A Model

A model that characterizes the individual observations is:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

where

y_{ij} is the data point for the ith site and jth replication,

τ_i is the effect for site i,

ε_{ij} is the random effect.

	NC	Jackson	Minn	Balt
	433	445	434	441
	435	440	436	443
	432	438	433	438
	439	441	437	439
	436		434	442
			438	444
			440	
			435	
Total	2,175	1,764	3,487	2,647
n	5	4	8	6
Mean	435.0	441.0	435.9	441.2
Effect	-2.96	3.04	-2.08	3.21
Adj Effect	-0.64	0.53	-0.72	0.84
$Effect^2 * n$	43.71	37.05	34.66	61.83

Adjusted Effect

The adjusted effect is adjusted for the sample size by multiplying the unadjusted effect times n_i/N, which for NC would be

$n_{NC} = 5$

$N = 23$

Effect = - 2.96,

So,

adjusted effect = - 2.96*(5/23) = - 0.64