Analysis of Covariance (ANOCOVA)
(or ANCOVA)

Stat 566
4-23-15

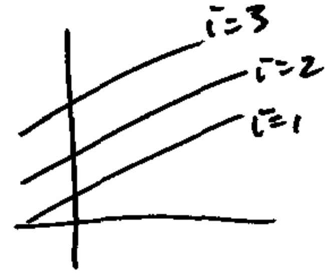
①

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \qquad i = 1,\dots,a \quad j = 1,\dots,n$$

$$\sum_{i=1}^{a} \tau_i = 0 \qquad E[\varepsilon_{ij}] \equiv 0$$

$$V[\varepsilon_{ij}] \equiv \sigma^2$$

ε_{ij} indep $\forall\, i,j$



Notation: $S_{yy} = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j y_{ij}^2 - an\bar{y}_{..}^2$ ②

$$S_{xx} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2$$

$$S_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..})$$

$$T_{yy} = n \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$T_{xx} = n \sum_i (\bar{x}_{i.} - \bar{x}_{..})^2$$

$$T_{xy} = n \sum_i (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..})$$

$$E_{yy} = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$ ③

$$E_{xx} = \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2$$

$$E_{xy} = \sum_i \sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})$$

Goal: Estimate all of the parameters + then test

$$H_0: \tau_i \equiv 0$$

$$SSE = \sum_i \sum_j \left[y_{ij} - \mu - \tau_i - \beta(x_{ij} - \bar{x}_{..}) \right]^2$$

$$\frac{\partial SSE}{\partial \mu} = \sum_i \sum_j \cancel{2} \left[y_{ij} - \mu - \tau_i - \beta(x_{ij} - \bar{x}_{..}) \right](\cancel{-1}) \overset{set}{=} 0$$

$$y_{..} - an\mu - \beta\cancel{(x_{..} - an\bar{x}_{..})} = 0$$ ④

$$\hat{\mu} = \frac{y_{..}}{an} = \bar{y}_{..}$$

$$\frac{\partial SSE}{\partial \tau_i} = \sum_j \cancel{2} \left[y_{ij} - \mu - \tau_i - \beta(x_{ij} - \bar{x}_{..}) \right](\cancel{-1}) \overset{set}{=} 0$$

$$y_{i.} - n\mu - n\tau_i - \beta(x_{i.} - n\bar{x}_{..}) = 0$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})$$

$$\frac{\partial SSE}{\partial \beta} = \sum_i \sum_j \cancel{2} \left[y_{ij} - \mu - \tau_i - \beta(x_{ij} - \bar{x}_{..}) \right](\cancel{-1})(x_{ij} - \bar{x}_{..}) \overset{set}{=} 0$$

$$\sum_i \sum_j (y_{ij} - \hat{\mu})(x_{ij} - \bar{x}_{..}) - \sum_i \sum_j \hat{\tau}_i (x_{ij} - \bar{x}_{..}) - \hat{\beta} \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = 0$$

(5)

$$S_{xy} - \sum_i \sum_j \hat{\tau}_i (x_{ij} - \bar{x}_{..}) - \hat{\beta} S_{xx} = 0$$

$$S_{xy} - \sum_i \sum_j \left(\bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})\right)(x_{ij} - \bar{x}_{..}) - \hat{\beta} S_{xx} = 0$$

$$S_{xy} - \sum_i \left[\left(\bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})\right) \underbrace{\sum_j (x_{ij} - \bar{x}_{..})}_{\substack{x_{i.} - n\bar{x}_{..} \\ n(\bar{x}_{i.} - \bar{x}_{..})}} \right] - \hat{\beta} S_{xx} = 0$$

$$S_{xy} - n \sum_i \left[(\bar{y}_{i.} - \bar{y}_{..})(\bar{x}_{i.} - \bar{x}_{..}) - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})^2 \right] - \hat{\beta} S_{xx} = 0$$

$$S_{xy} - T_{xy} + \hat{\beta} T_{xx} - \hat{\beta} S_{xx} = 0$$

(6)

$$\hat{\beta} = \frac{S_{xy} - T_{xy}}{S_{xx} - T_{xx}} = \frac{E_{xy}}{E_{xx}}$$

Results: $\hat{\mu} = \bar{y}_{..}$, $\hat{\beta} = \frac{E_{xy}}{E_{xx}}$, $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})$

To test $H_0: \tau_i \equiv 0$, use the additional sum of squares F test, which involves SSE and df_E from

Full: $y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}$

Reduced: $y_{ij} = \mu + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}$

(7)

Full model:

$$SSE = \sum_i \sum_j \Big[y_{ij} - \bar{y}_{..} - \big(\bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \big) - \hat{\beta}(x_{ij} - \bar{x}_{..}) \Big]^2$$

$$= \sum_i \sum_j \Big[y_{ij} - \bar{y}_{i.} - \hat{\beta}(x_{ij} - \bar{x}_{i.}) \Big]^2$$

$$= E_{yy} + \hat{\beta}^2 E_{xx} - 2\hat{\beta} E_{xy}$$

(8)

$$= E_{yy} + \left(\frac{E_{xy}}{E_{xx}}\right)^2 E_{xx} - 2\left(\frac{E_{xy}}{E_{xx}}\right) E_{xy}$$

$$= E_{yy} - \frac{E_{xy}^2}{E_{xx}}$$

Full model ANOVA

Source	SS	df	
Model	$T_{yy} + \frac{E_{xy}^2}{E_{xx}}$	a	$(a-1)+1$
Error	$E_{yy} - \frac{E_{xy}^2}{E_{xx}}$	$an-1-a$	
Total	S_{yy}	$an-1$	

9

Reduced model:

$$y_{ij} = \underbrace{(\mu - \beta \bar{x}_{..})}_{\beta_0} + \underbrace{\beta}_{\beta_1} x_{ij} + \varepsilon_{ij}$$

From Stat 564, $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

$$\hat{\beta}_0 = \bar{y}_{..} - \hat{\beta}_1 \bar{x}_{..} \overset{set}{=} \hat{\mu} - \hat{\beta}_1 \bar{x}_{..}$$

$$\hat{\mu} = \bar{y}_{..}$$

Also, $SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$ $\quad df_E = an - 2$

10

Recall:

$$F = \frac{\left(\frac{SSE_{red} - SSE_{full}}{dfE_{red} - dfE_{full}}\right)}{MSE_{full}}$$

$$= \frac{\left(\frac{S_{yy} - \frac{S_{xy}^2}{S_{xx}} - \left(E_{yy} - \frac{E_{xy}^2}{E_{xx}}\right)}{an-2 - (an-1-a)}\right)}{MSE_{full}}$$

(11)

$$= \left(\frac{T_{yy} - \frac{S_{xy}^2}{S_{xx}} + \frac{E_{xy}^2}{E_{xx}}}{a-1} \right) \Big/ MSE_{full}$$

Compare this to an F critical value,
using $a-1$ and $an-1-a$ df

ANCOVA table

Source	SS	df
Reg	S_{xy}^2/S_{xx}	1
TRT	$T_{yy} - \frac{S_{xy}^2}{S_{xx}} + \frac{E_{xy}^2}{E_{xx}}$	$a-1$
ERR	$E_{yy} - \frac{E_{xy}^2}{E_{xx}}$	$an-1-a$
TOT	S_{yy}	$an-1$

(12)

* Type I SS are sequential

Type III SS are each conditioned upon all other model terms.

Hw#4 due 4/30

14.20, 14.21

14.14. Suppose that in Problem 14.13 a large number of power settings could have been used and that the two selected for the experiment were chosen randomly. Obtain the expected mean squares for this situation assuming the restricted form of the mixed model and modify the previous analysis appropriately.

14.15. Reanalyze the experiment in Problem 14.14 assuming the unrestricted form of the mixed model. You may use a computer software package to do this. Comment on any differences between the restricted and unrestricted model analysis and conclusions.

14.16. A structural engineer is studying the strength of aluminum alloy purchased from three vendors. Each vendor submits the alloy in standard-sized bars of 1.0, 1.5, or 2.0 inches. The processing of different sizes of bar stock from a common ingot involves different forging techniques, and so this factor may be important. Furthermore, the bar stock is forged from ingots made in different heats. Each vendor submits two test specimens of each size bar stock from three heats. The resulting strength data is shown in Table P14.2. Analyze the data, assuming that vendors and bar size are fixed and heats are random. Use the restricted form of the mixed model.

■ **TABLE P14.2**
Strength Data in Problem P14.16

	Vendor 1			Vendor 2		
Heat	**1**	**2**	**3**	**1**	**2**	**3**
Bar size:						
1 in.	1.230	1.346	1.235	1.301	1.346	1.315
	1.259	1.400	1.206	1.263	1.392	1.320
$1\frac{1}{2}$ in.	1.316	1.329	1.250	1.274	1.384	1.346
	1.300	1.362	1.239	1.268	1.375	1.357
2 in.	1.287	1.346	1.273	1.247	1.362	1.336
	1.292	1.382	1.215	1.215	1.328	1.342

	Vendor 3		
Heat	**1**	**2**	**3**
Bar size:			
1 in.	1.247	1.275	1.324
	1.296	1.268	1.315
$1\frac{1}{2}$ in.	1.273	1.260	1.392
	1.264	1.265	1.364
2 in.	1.301	1.280	1.319
	1.262	1.271	1.323

14.17. Rework Problem 14.16 using the unrestricted form of the mixed model. You may use a computer software package to do this. Comment on any differences between the restricted and unrestricted model analysis and conclusions.

14.18. Suppose that in Problem 14.16 the bar stock may be purchased in many sizes and that the three sizes actually used in the experiment were selected randomly. Obtain the expected mean squares for this situation and modify the previous analysis appropriately. Use the restricted form of the mixed model.

14.19. Steel is normalized by heating above the critical temperature, soaking, and then air cooling. This process increases the strength of the steel, refines the grain, and homogenizes the structure. An experiment is performed to determine the effect of temperature and heat treatment time on the strength of normalized steel. Two temperatures and three times are selected. The experiment is performed by heating the oven to a randomly selected temperature and inserting three specimens. After 10 minutes one specimen is removed, after 20 minutes the second is removed, and after 30 minutes the final specimen is removed. Then the temperature is changed to the other level and the process is repeated. Four shifts are required to collect the data, which are shown below. Analyze the data and draw conclusions, assuming both factors are fixed.

		Temperature (°F)	
Shift	**Time (min)**	**1500**	**1600**
	10	63	89
1	20	54	91
	30	61	62
	10	50	80
2	20	52	72
	30	59	69
	10	48	73
3	20	74	81
	30	71	69
	10	54	88
4	20	48	92
	30	59	64

14.20. An experiment is designed to study pigment dispersion in paint. Four different mixes of a particular pigment are studied. The procedure consists of preparing a particular mix and then applying that mix to a panel by three application methods (brushing, spraying, and rolling). The response measured is the percentage reflectance of pigment. Three days are required to run the experiment, and the data obtained follow. Analyze the data and draw conclusions, assuming that mixes and application methods are fixed.

Day	Application Method	Mix 1	2	3	4
	1	64.5	66.3	74.1	66.5
1	2	68.3	69.5	73.8	70.0
	3	70.3	73.1	78.0	72.3
	1	65.2	65.0	73.8	64.8
2	2	69.2	70.3	74.5	68.3
	3	71.2	72.8	79.1	71.5
	1	66.2	66.5	72.3	67.7
3	2	69.0	69.0	75.4	68.6
	3	70.8	74.2	80.1	72.4

14.21. Repeat Problem 14.20, assuming that the mixes are random and the application methods are fixed.

14.22. Consider the split-split-plot design described in Example 14.4. Suppose that this experiment is conducted as described and that the data shown in Table P14.3 are obtained. Analyze the data and draw conclusions.

14.23. Rework Problem 14.22, assuming that the technicians are chosen at random. Use the restricted form of the mixed model.

14.24. Suppose that in Problem 14.22 four technicians had been used. Assuming that all the factors are fixed, how many blocks should be run to obtain an adequate number of degrees of freedom on the test for differences among technicians?

■ **TABLE P14.3**
The Absorption Time Experiment

Replicates (or Blocks)	Dosage Strengths	Technician 1: 1	2	3	Technician 2: 1	2	3	Technician 3: 1	2	3
	Wall Thickness									
	1	95	71	108	96	70	108	95	70	100
1	2	104	82	115	99	84	100	102	81	106
	3	101	85	117	95	83	105	105	84	113
	4	108	85	116	97	85	109	107	87	115
	1	95	78	110	100	72	104	92	69	101
2	2	106	84	109	101	79	102	100	76	104
	3	103	86	116	99	80	108	101	80	109
	4	109	84	110	112	86	109	108	86	113
	1	96	70	107	94	66	100	90	73	98
3	2	105	81	106	100	84	101	97	75	100
	3	106	88	112	104	87	109	100	82	104
	4	113	90	117	121	90	117	110	91	112
	1	90	68	109	98	68	106	98	72	101
4	2	100	84	112	102	81	103	102	78	105
	3	102	85	115	100	85	110	105	80	110
	4	114	88	118	118	85	116	110	95	120

14.25. Consider the experiment described in Example 14.4. Demonstrate how the order in which the treatment combinations are run would be determined if this experiment were run as (a) a split-split-plot, (b) a split-plot, (c) a factorial design in a randomized block, and (d) a completely randomized factorial design.

14.26. An article in *Quality Engineering* ("Quality Quandries: Two-Level Factorials Run as Split-Plot Experiments," Bisgaard et al., Vol. 8, No. 4, pp. 705–708, 1996) describes a 2^5 factorial experiment in a plasma process focused on making paper more susceptible to ink. Four of the factors (A–D) are difficult to change from run to run, so the experimenters set up the reactor at the eight sets of conditions specified by the low and high levels of those factors, and then processed the two paper types (factor E) together. The placement of the paper specimens in