Analysis of Covariance

Reference: Section 15.3 of Montgomery (7e) Maghsoodloo

ANCOVA is often said to be the marriage between ANOVA and Regression, where the response variable, y, is modeled versus one treatment (or more), but as y is measured, one or more uncontrollable variable(s) that are correlated with y must also be measured that may impact the measurement on y. Montgomery pp. 590-606 (pp. 574-590 of the 6th edition) provides a good example on page 591, where the experimental interest is to determine if the breaking strength of fiber, y, is different from 3 different machines (treatments), but y is also correlated with the diameter or thickness, x, of the fiber (i.e., larger diameters generally lead to larger breaking strength and hence a positive correlation). In this example, y is measured in lbs and x in units of inches$\times 10^{-3}$. An uncontrollable variable, x, is also called a nuisance, a covariate (Minitab's terminology), an auxiliary, or a concomitant variable. The data with 5 paired observations per treatment are displayed atop page 591 in Table 15.10 of Montgomery's 7th edition (p. 575 of the 6th Ed.), where there are a = 3 machines (or treatments).

The covariance model with one factor (τ) and one covariate, x, is given by

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \overline{x}_{..}) + \in_{ij} \quad (69)$$

In order to compute the Regression SS's, we must 1st obtain the LSEs (least-squares estimates) of the parameters μ, τ_i, and β. The LSF for the model (69) is given by

$$L(\mu, \tau_i, \beta) = \sum_{i=1}^{a}\sum_{j=1}^{n} \in_{ij}^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}[y_{ij} - \mu - \tau_i - \beta(x_{ij} - \overline{x}_{..})]^2 \quad (70)$$

where "*a*" stands for the number of treatments. So in our example, the value of *a* = 3 machines and n = 5 so that N = *a*$\times$n = 15 total observations.

Partially differentiating the LSF in (70) wrt μ, τ_i (i = 1, 2, ..., a), β and requiring them to be zero yields a set of 1 + a +1 = a + 2 normal equations with a + 2, $(\hat{\mu}, \hat{\tau}_1, \hat{\tau}_2, ..., \hat{\tau}_a, \hat{\beta})$, unknowns.

$$N\hat{\mu} + n\sum_{i=1}^{a}\hat{\tau}_i = y_{..} \quad (71a)$$

$$n\hat{\mu} + n\hat{\tau}_i + \hat{\beta}(x_{i.} - n\,\overline{x}_{..}) = y_{i.} \quad (71b)$$

$$\sum_{i=1}^{a}\hat{\tau}_i(x_{i.} - n\,\overline{x}_{..}) + \hat{\beta}\,S_{xx} = S_{xy} \quad (71c)$$

where $S_{xx} = \sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \overline{x}_{..})^2$ and $S_{xy} = \sum_{i=1}^{a}\sum_{j=1}^{n}[(x_{ij} - \overline{x}_{..})y_{ij}] = \sum_{i=1}^{a}\sum_{j=1}^{n}[(x_{ij} - \overline{x}_{..})(y_{ij} - \overline{y}_{..})]$. The above heterogeneous system of normal equations has no unique solution unless we impose the constraint $\sum_{i=1}^{a}\hat{\tau}_i = 0$. Note that the sum of Eq. (71b) from 1 to "*a*" adds to the 1st normal equation, and hence there are only a + 1 independent equations but with *a* + 2 unknowns leading to no unique solutions (or zero determinant). The constraint $\sum_{i=1}^{a}\hat{\tau}_i = 0$ imposed in equation (71a) leads to $\hat{\mu} = \overline{y}_{..}$. Substituting $\hat{\mu} = \overline{y}_{..}$ into the normal Eqs. (71b) and dividing by n yields $\overline{y}_{..} + \hat{\tau}_i + \hat{\beta}(\overline{x}_{i.} - \overline{x}_{..}) = \overline{y}_{i.} \rightarrow$

$$\hat{\tau}_i = (\overline{y}_{i.} - \overline{y}_{..}) - \hat{\beta}(\overline{x}_{i.} - \overline{x}_{..}) = [\overline{y}_{i.} - \hat{\beta}(\overline{x}_{i.} - \overline{x}_{..})] - \overline{y}_{..} \quad (72)$$

Eq. (72) tells a remarkable story! Recall that W/O the regression of y on x, the LS estimate of the ith treatment effect is simply $\hat{\tau}_i = (\overline{y}_{i.} - \overline{y}_{..})$ in a one–way ANOVA model. Eq. (72) basically gives the same LS estimate of τ_i except that it also corrects $\hat{\tau}_i$ (also $\overline{y}_{i.}$) for the regression of y on x, (i.e., for the slope of y on x). The larger the slope estimator $\hat{\beta}$, the greater is the amount of adjustment for regression, i.e., Adj($\overline{y}_{i.}$) = $\overline{y}_{i.} - \hat{\beta}(\overline{x}_{i.} - \overline{x}_{..})$.

So far we have obtained the point estimates of μ and τ_i (i = 1, 2, ..., a); we next substitute these (*a*+1) estimates into equation (71c) in order to obtain the point estimator $\hat{\beta}$.

$$\sum_{i=1}^{a}\{[(\overline{y}_{i.} - \overline{y}_{..}) - \hat{\beta}(\overline{x}_{i.} - \overline{x}_{..})](x_{i.} - n\,\overline{x}_{..})\} + \hat{\beta}S_{xx} = S_{xy} \longrightarrow$$

$$\sum_{i=1}^{a}\{n[(\overline{y}_{i.} - \overline{y}_{..}) - \hat{\beta}(\overline{x}_{i.} - \overline{x}_{..})](\overline{x}_{i.} - \overline{x}_{..})\} + \hat{\beta}\,S_{xx} = S_{xy} \rightarrow$$

$\hat{\beta}$ [$S_{xx} - n \sum_{i=1}^{a} (\bar{x}_{i.} - \bar{x}_{..})^2$] = $S_{xy} - n \sum_{i=1}^{a} (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..})$ $\longrightarrow$

$\hat{\beta}\,[S_{xx} - T_{xx}] = S_{xy} - T_{xy} \longrightarrow \hat{\beta} = \dfrac{S_{xy} - T_{xy}}{S_{xx} - T_{xx}} = \dfrac{E_{xy}}{E_{xx}}$ (73a)

In Eq. (73a), $T_{xx} = \sum_{i=1}^{a}\sum_{j=1}^{n} (\bar{x}_{i.} - \bar{x}_{..})^2 = n\sum_{i=1}^{a} (\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^{a} \dfrac{x_{i.}^2}{n} - N(\bar{x}_{..})^2$ and $T_{xy} = n\sum_{i=1}^{a} (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}) = \sum_{i=1}^{a} x_{i.}(\bar{y}_{i.} - \bar{y}_{..}) = \sum_{i=1}^{a} \dfrac{x_{i.}\, y_{i.}}{n} - (x_{..})(y_{..})/N$. Further, the symbol T stands for treatment, E for error and S for Total, when computing SS's in ANCOVA. We now use Eq. (73a) to compute $\hat{\beta}$ for the data of Table 15.10 on page 591 of Montgomery, where these computations are also carried out in Example 15.5 on pp. 595-597 of Montgomery (pp. 578-583 of 6th edition). You may verify that N =15, $x_{..} = \sum_{i,j} x_{ij} = 362$, $y_{..} = \sum_{i,j} y_{ij} = 603$, $USS_x = \sum_{i,j} x_{ij}^2 = 8998$, $USS_y = \sum_{i,j} y_{ij}^2 = 24{,}587$, and $\sum_{i=1}^{3}\sum_{j=1}^{5} x_{ij} y_{ij} = 14{,}835$. Then, $S_{xy} = 36\times20 + 41\times25 + 39\times24 + \ldots + 32\times15 - 362\times603/15 = 282.60$, $T_{xy} = (207\times126 + 216\times130 + 180\times106)/5 - 362\times603/15 = 96.00$, $S_{xx} = 8998 - 362^2/15 = 261.7333333$, $T_{xx} = (126^2 + 130^2 + 106^2)/5 - 362^2/15 = 66.1333333333333 \rightarrow$

$$\hat{\beta} = \frac{282.60 - 96.00}{261.733\bar{3} - 66.133\bar{3}} = \frac{186.6}{195.60} = \frac{E_{xy}}{E_{xx}} = 0.9540. \qquad (73b)$$

Finally, we use the estimates $\hat{\mu} = \bar{y}_{..} = 40.20$ and $\hat{\beta} = 0.9540$ in Eq. (72) to obtain the LSEs of τ_i's.

$\hat{\tau}_1$ = (41.4 – 40.20) – 0.9540 (25.2 – 24.1333333) = 0.18240

$\hat{\tau}_2$ = (43.2 – 40.20) – 0.9540 (26.0 – 24.1333333) = 1.21920

$\hat{\tau}_3$ = (36.0 – 40.20) – 0.9540 (21.2 – 24.1333333) = –1.4016

As expected, the constraint $\sum_{i=1}^{3} \hat{\tau}_i \equiv 0$ is precisely satisfied.

To compute the $SS_{Reg}(\hat{\mu}, \hat{\tau}_i, \hat{\beta})$, I need to remind you as to what exactly happens when y is regressed only on 3 independent variables (i.e., no treatment effects) as shown below.

$$y_i = b_0 + \beta_1(x_{i1} - \overline{x}_1) + \beta_2(x_{i2} - \overline{x}_2) + \beta_3(x_{i3} - \overline{x}_3) + \epsilon_i \quad (74)$$

where the point LSE of b_0 is $\hat{b}_0 = \overline{y}$.

As always, the 1st step is to obtain the LSEs of β_i (i = 1, 2, 3) by obtaining the LS normal equations, which will be of the following form:

$$\begin{aligned} \hat{\beta}_1 S_{11} + \hat{\beta}_2 S_{12} + \hat{\beta}_3 S_{13} &= S_{1y} \\ \hat{\beta}_1 S_{12} + \hat{\beta}_2 S_{22} + \hat{\beta}_3 S_{23} &= S_{2y} \\ \hat{\beta}_1 S_{13} + \hat{\beta}_2 S_{23} + \hat{\beta}_3 S_{33} &= S_{3y} \end{aligned} \quad (75)$$

Recall from regression theory that $SS_{Reg}(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = \sum_{j=1}^{3} \hat{\beta}_j S_{jy}$, which shows that in order to compute SS_{Reg}, we must multiply each point estimate by the corresponding term on the RHS of the LSNEQs (75) and then sum the resulting products.

Bearing the above procedure in mind, we can compute SS_{Reg}(due to $\hat{\mu}, \hat{\tau}_i, \hat{\beta}$) by multiplying each estimate by the corresponding term on the RHS of equations (71 a, b, c), and then adding the 3 terms. That is,

SS_{Reg}(due to $\hat{\mu}, \hat{\tau}_i, \hat{\beta}$) = $\hat{\mu} y_{..} + \sum \hat{\tau}_i y_{i.} + \hat{\beta} S_{xy}$

$$= CF + \sum_{i=1}^{a} [(\overline{y}_{i.} - \overline{y}_{..}) - \hat{\beta}(\overline{x}_{i.} - \overline{x}_{..})] y_{i.} + \hat{\beta} S_{xy}$$

$$= CF + T_{yy} - \hat{\beta} T_{xy} + \hat{\beta} S_{xy} = CF + T_{yy} + \hat{\beta}(S_{xy} - T_{xy})$$

$$SS_{Reg}(\hat{\mu}, \hat{\tau}_i, \hat{\beta}) = CF + T_{yy} + \hat{\beta} E_{xy} = CF + T_{yy} + E_{xy}^2 / E_{xx} \quad (76)$$

where I have made use of the fact that $\hat{\beta} = \dfrac{E_{xy}}{E_{xx}}$ from Eq. (73a). Eq. (76) describes that SS(Model) consists of 2 terms because by definition SS(Model) = $SS_{Reg}(\hat{\tau}_i, \hat{\beta} | \hat{\mu}) = SS_{Reg}(\hat{\mu}, \hat{\tau}_i, \hat{\beta}) - SS_{Reg}(\hat{\mu}) = T_{yy} + E_{xy}^2 / E_{xx}$, where $SS_{Reg}(\hat{\mu})$ = CF. The 1st model term is the 2nd term on the RHS of (76) giving the amount of variation explained by treatments, namely SS(Treatments) = T_{yy} = $\sum_{i=1}^{a} \frac{y_{i.}^2}{n}$ – CF, and the 2nd model term [the last on the RHS of (76)] describes the variation due to the slope, namely $\hat{\beta} E_{xy} = E_{xy}^2/E_{xx}$. Therefore, SS(Residuals) = SS_{Total} – SS(Model) = $S_{yy} - T_{yy} - E_{xy}^2/E_{xx} = E_{yy} - E_{xy}^2/E_{xx}$, where S_{yy} = USS – CF = 24,587 – 603^2/15 = 346.40 (with 14 *df*), T_{yy} = $(207^2 + 216^2 + 180^2)/5$ – CF = 140.40 (with 2 *df*), and SS(Slope) = SS(Diameter|Machines) = $E_{xy}^2/E_{xx} = 186.6^2 / 195.60$ = 178.01411 (with 1 *df*). Hence, SS(Model) = 318.41411 (with 3 *df*), and SS(Residuals) = 346.40 – 318.41411 = 27.9859 (with 11 *df*). It is paramount that you note the pattern for computing the error SS in an ANCOVA model as compared to an ANOVA model. In a one-way ANOVA model, SS(Error) = $E_{yy} = S_{yy} - T_{yy}$ = 346.40 – 140.40 = 206.00, while in an ANCOVA model SS(Residuals) = $E_{yy} - E_{xy}^2 / E_{xx}$, i.e., in the ANCOVA model the Error SS is also adjusted for the regression of y on x, and as a consequence the Adj $SS_{Res} = E_{yy} - E_{xy}^2/E_{xx}$ = 206.00 – 178.01411 = 27.9859 has one less *df* than the corresponding ANOVA model.

To test for treatment effects under model (69), we have to hypothesize that H_0: $\tau_i = 0$ so that our ANCOVA model (69) reduces to

$$y_{ij} = \mu + \beta(x_{ij} - \overline{x}_{..}) + \epsilon_{ij} \qquad (77)$$

Again we have to obtain the LSE's of the two parameters μ and β, which is done by minimizing the reduced LSF

$$L(\mu, \beta) = \sum_{i=1}^{a}\sum_{j=1}^{n} \epsilon_{ij}^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}[y_{ij} - \mu - \beta(x_{ij} - \overline{x}_{..})]^2 \qquad (78)$$

Setting $\frac{\partial L}{\partial \mu}$ to zero yields $\hat{\mu} = \overline{y}_{..} = y_{..}/N$ as before. Setting $\frac{\partial L}{\partial \beta}$ equal to zero leads to the following normal equation.

$$\sum_{i=1}^{a}\sum_{j=1}^{n}[y_{ij} - \overline{y}_{..} - \beta(x_{ij} - \overline{x}_{..})][-(x_{ij} - \overline{x}_{..})] = 0 \longrightarrow \hat{\beta} = S_{xy} / S_{xx}.$$

Therefore, for the reduced model (77), the $SS_{Reg}(\hat{\mu}, \hat{\beta}) = \overline{y}_{..} y_{..} + \hat{\beta} S_{xy}$. Then it follows that the net contribution of τ is given by

$SS_{Reg}(\tau| \mu, \beta) = SS_{Reg}(\text{due to } \hat{\mu}, \hat{\tau}_i, \text{and } \hat{\beta}) - SS_{Reg}(\text{due to } \hat{\mu} \text{ and } \hat{\beta})$

$= CF + T_{yy} + E_{xy}^2/E_{xx} - CF - \hat{\beta} S_{xy} = (S_{yy} - E_{yy}) + E_{xy}^2/E_{xx} - (S_{xy}/S_{xx}) S_{xy}$

$= (S_{yy} - S_{xy}^2/S_{xx}) - (E_{yy} - E_{xy}^2/E_{xx}) = \text{Adj } SS_{Total} - \text{Adj } SS(\text{Error})$.

For the example under consideration the adjusted for regression SS are

Adj $SS_{Total} = S_{yy} - S_{xy}^2/S_{xx} = (25587 - 603^2/15) - 282.60^2/261.73333$

$= 346.40 - 305.13026 = 41.26974$ (with 13 *df*)

Adj $SS_{Treatments}$ = Adj SS_{Total} – Adj SS(Error) = 41.26974 – 27.9859

= 13.28384 (with 13 – 11 = 2 *df*).

All the above calculations are summarized in the ANCOVA Table below. Further, SS(Model) = SS(Diameter) + SS(Machines|x) = S_{xy}^2/S_{xx} + Adj $SS_{Treatments}$ = 305.13026 + 13.28384 = 318.41411, as before. Note that an incorrect analysis of variance gives MS(Treatments) = $T_{yy}/2$ =140.4/2 = 70.10, MS(Error) = $E_{yy}/12$ = 206/12 = 17.16667, which leads to F_0 = 4.0835

The ANCOVA Table for the Example 15.5 of Montgomery

Source of Variation	Sums of Squares and Products x	xy	y	Adjusted for Regression SS	df	MS	F_0	*P-value*
Total	261.733	282.60	346.40	41.270	13			
Machines	66.133	96.00	140.40					
Error	195.60	186.60	206.00	27.986	11	2.5442		$\hat{\alpha}$
Adjusted	Treatments			13.284	2	6.642	2.611	0.1181

($p = \hat{\alpha}$ = *P-value* = 0.0444) and leads to an erroneous decision regarding H_0: $\tau_i = 0$ for i = 1, 2, 3. That is, if we disregard the correlation between y and x and ignore the nuisance variable (or covariate) x , we would then come to the wrong decision that there exists significant differences amongst the three machines.

In order to test for the significance of slope, we must hypothesize that $\beta = 0$ and compute SS($\hat{\beta}\,|\hat{\mu},\hat{\tau}$) = SS($\hat{\mu},\hat{\tau},\hat{\beta}$) – SS($\hat{\mu},\hat{\tau}$). Putting β = 0 into model (69) reduces that model to a one-factor ANOVA with SS($\hat{\mu},\hat{\tau}$) = CF + T_{yy}. Hence, SS($\hat{\beta}\,|\hat{\mu},\hat{\tau}$) = (CF + T_{yy} + E_{xy}^2/E_{xx}) – (CF + T_{yy}) = E_{xy}^2/E_{xx}. Thus, the statistic for testing H_0: β = 0 is $F_{1,\nu_2} = \dfrac{E_{xy}^2/E_{xx}}{\text{MS(RES)}}$ =178.01411/ 2.5442 = 69.9694, where ν_2 is the *df* of SS_{RES} equaling 11 for this example. Finally, ANCOVA assumes that treatments have no significant effect on the covariate x. This assumption can be tested using ordinary ANOVA on x_{ij}'s. The F-statistic is F_0 = MS(Diameter)/ $MSRES_x$ = $\dfrac{61.133\overline{3}/2}{195.6000/12}$ = 2.02863, with P-value = 0.174204.

Confidence Intervals For μ_i Using The Adjusted Treatment Means

Eq. (72) clearly shows that $\hat{\tau}_i = (\bar{y}_{i.} - \bar{y}_{..}) - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})$, or $\hat{\tau}_i = [\bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})] - \bar{y}_{..}$ = Adj $\bar{y}_{i.} - \bar{y}_{..}$. This last relationship shows that the adjusted means for regression are given by

$$\text{Adj } \bar{y}_{i.} = \bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) = \text{UnAdj Mean – Adj Due to Regression} \qquad (79)$$

From Eq. (79), Adj $\bar{y}_{1.}$ = 41.40 – 0.9540 (25.20 – 24.13333) = 40.3824; similarly, Adj $\bar{y}_{2.}$ = 41.4192 and Adj $\bar{y}_{3.}$ = 38.7984. In order to obtain the 95% CIs For μ_i using Adj $\bar{y}_{i.}$, it is necessary 1st to compute the variance of Adj $\bar{y}_{i.}$ so that its *se* can be determined. Applying the variance operator V to equation (79) yields

$$V(\text{Adj } \bar{y}_{i.}) = V[\bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})] = V(\bar{y}_{i.}) + (\bar{x}_{i.} - \bar{x}_{..})^2 V(\hat{\beta}), \qquad (80)$$

where we are using the fact that in classical regression the x's are fixed, and the mean and slope are independent rvs. Next we compute the V($\hat{\beta}$).

$$V(\hat{\beta}) = V\left(\frac{E_{xy}}{E_{xx}}\right) = \frac{1}{E_{xx}^2} V[\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})]$$

$$= \frac{1}{E_{xx}^2} V \sum_{i=1}^{a}\sum_{j=1}^{n}[(x_{ij} - \bar{x}_{i.})\, y_{ij}] = \frac{1}{E_{xx}^2} \sum_{i=1}^{a}\sum_{j=1}^{n}[(x_{ij} - \bar{x}_{i.})^2\, V(y_{ij})]$$

$$= \frac{1}{E_{xx}^2}\sum_{i=1}^{a}\sum_{j=1}^{n}[(x_{ij}-\overline{x}_{i.})^2\sigma_{\in}^2] = \frac{1}{E_{xx}^2}(E_{xx}\sigma_{\in}^2) = \sigma_{\in}^2/E_{xx} \ . \qquad (81)$$

Thus, the $se(\hat{\beta}) = \sqrt{MS_{RES}/E_{xx}}$, and the HCIL(β) = $t_{0.025,n-1-a} \times se(\hat{\beta})$.

Inserting (81) into (80) yields:

$$V(\text{Adj } \overline{y}_{i.}) = \frac{\sigma_{\in}^2}{n} + (\overline{x}_{i.} - \overline{x}_{..})^2 \frac{\sigma_{\in}^2}{E_{xx}} = [\frac{1}{n} + \frac{(\overline{x}_{i.}-\overline{x}_{..})^2}{E_{xx}}]\sigma_{\in}^2 \qquad (82)$$

For the example under consideration, Eq. (81) yields $\hat{V}(\hat{\beta})$ = 2.5442/195.60 = 0.013007;

then from Eq. (82) we obtain: $V(\text{Adj } \overline{y}_{i.}) = [\frac{1}{5} + \frac{(\overline{x}_{i.}-24.1333\overline{3})^2}{195.60}]\sigma_{\in}^2 \rightarrow se(\text{Adj } \overline{y}_{i.})$ =

$\sqrt{[\frac{1}{5}+\frac{(\overline{x}_{i.}-24.13333)^2}{195.60}]\times 2.5442}$. For μ_1, this last equation results in HCIL =

$(t_{0.025;11}) \times se(\text{Adj } \overline{y}_{1.})$ = 2.201×0.72363 = 1.59271 → 38.7897 ≤ μ_1 ≤ 41.9751; similarly, 39.7807 ≤ μ_2 ≤ 43.0577. Note that these 2 CIs are not disjoint, which may imply that the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu$ cannot be rejected at the 5% level.

<u>Exercise 30.</u> **(a)** Obtain the CI for μ_3 of Example 15.5 on pp. 595-602 Montgomery (pp. 578-586 of 6e).

Residuals For an ANCOVA Model

Recall that by definition, a residual is the difference between the actual observed response, y_{ij}, and the corresponding fitted value from the model, which is always denoted by $\hat{y}_{ij}$, i.e., $e_{ij} = y_{ij} - \hat{y}_{ij}$. We now use the ANCOVA model (69) 1st to compute the fitted values.

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}(x_{ij} - \overline{x}_{..}) + \hat{\in}_{ij}$$

$$= \overline{y}_{..} + (\overline{y}_{i.} - \overline{y}_{..}) - \hat{\beta}(\overline{x}_{i.} - \overline{x}_{..}) + \hat{\beta}(x_{ij} - \overline{x}_{..}) + 0 \quad \rightarrow$$

$$\hat{y}_{ij} = \overline{y}_{i.} + \hat{\beta}(x_{ij} - \overline{x}_{i.}) \quad \rightarrow \quad e_{ij} = y_{ij} - \overline{y}_{i.} - \hat{\beta}(x_{ij} - \overline{x}_{i.}) \qquad (83)$$

As an example, the value of the 1st residual for the Example 15.5 of Montgomery on p. 595 is e_{11} = 36 – 41.40 – 0.9540(20 – 25.20) = –0.4392, etc.

<u>Exercise 30(b).</u> Compute the values of e_{23} and e_{35} of Example 15.5.

(c) Show that $V(e_{ij}) = [\frac{n-1}{n} - \frac{(x_{ij}-\overline{x}_{i.})^2}{E_{xx}}]\sigma^2_{\in}$. **(d)** Use your results in part (c) to compute the Studentized residual r_{23} of the Example 15.5. Use Minitab to check your answers.

(e) Work problems 15.15 & 15.16 on page 610 of Montgomery (p. 594 of the 6th edition and use Minitab to check your answers. Further, obtain the CI for β and use it to test H_0: $\beta = 0$ at the 5% level. **(f)** Test the null hypothesis that Glue formulation has no significant impact on the covariate x = “App. Thickness”.

I used Minitab’s GLM to obtain the ANCOVA Table and the corresponding Studentized residuals, r_{ij}, for the Example 15-5 of Montgomery. Minitab also has options for nearly everything.

Randomized Complete Blocks in an ANCOVA Model

Consider the following ANCOVA model with blocking effect B_j:

$$y_{ij} = \mu + \tau_i + B_j + \beta(x_{ij} - \overline{x}_{..}) + \in_{ij},\ i = 1, 2, ..., a;\ j = 1, 2, ..., b, \qquad (84)$$

where N = $a \times b$. I will summarize the step-by-step procedure below and in an exercise I will ask you to derive the slope estimator $\hat{\beta}$.

<u>**Step 1.**</u> Adjust the Error SS for blocks and then follow by adjusting for regression as shown below.

(1) Block adjustment: $E_{yy} = S_{yy} - T_{yy} - B_{yy}$, $E_{xx} = S_{xx} - T_{xx} - B_{xx}$

$E_{xy} = S_{xy} - T_{xy} - B_{xy} \longrightarrow \hat{\beta} = E_{xy} / E_{xx}$

(2) Regression adjustment: Adj $SS_{Error} = E_{yy} - E^2_{xy}/E_{xx}$.

<u>**Step 2.**</u> Adjust the Total SS’s for blocks and then for regression.

Block adjustment: $S'_{yy} = S_{yy} - B_{yy}$, $S'_{xx} = S_{xx} - B_{xx}$, $S'_{xy} = S_{xy} - B_{xy}$.

Regression adjustment: Adj $SS_{Total} = S'_{yy} - (S'_{xy})^2/ S'_{xx}$.

<u>**Step 3.**</u> Compute Adj $SS_{Treatments}$ by subtraction, i.e, Adj $SS_{Treatments}$ = Adj SS_{Total} – Adj SS_{Error}.

<u>**Step 4.**</u> Compute Adj $MS_{Treatments}$ = Adj $SS_{Treatments}/(a-1)$; Compute the *df* of the Adj $SS_{Error} = (N-1) - (a-1) - (b-1) - 1 = ab - a - b$. Next compute the Adj MS_{Error} = Adj SS_{Error} /(N

– *a* – *b*); Now form the F statistic as F_0 = Adj $MS_{Treatments}$ /Adj MS_{Error} with $\nu_1 = a - 1$ and $\nu_2 = ab - a - b$ degrees of freedom and compare against $F_{0.05;\nu_1,\nu_2}$.

Finally, it can be shown that in a randomized complete block ANCOVA model the residuals are given by $e_{ij} = (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) - \hat{\beta}(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$.

Exercise 31. Write the "2 + *a* +*b*" LSN equations for the model (84) and use them to show that $\hat{\beta} = E_{xy} / E_{xx}$ as indicated in **Step 1** above, where $E_{xy} = S_{xy} - T_{xy} - B_{xy}$ and $E_{xx} = S_{xx} - T_{xx} - B_{xx}$. Bear in mind that you have to impose 2 constraints to obtain a unique solution, namely, $\sum_{i=1}^{a} \tau_i = \sum_{j=1}^{b} B_j = 0$.

The covariance analysis can be generalized to two or more auxiliary variables (or covariates), and two or more factors. For example, if the relationship between y and x is nonlinear, then the covariance block model with two covariates is given by

$$y_{ij} = \mu + \tau_i + B_j + \beta_1(x_{ij} - \bar{x}_{..}) + \beta_{11}(x_{ij}^2 - \overline{x_{..}^2}) + \in_{ij}$$

where $\in_{ij}$ is the output error. For a 2-covariate model with block effect, we have

$$y_{ij} = \mu + \tau_i + B_j + \beta_1(x_{ij} - \bar{x}_{..}) + \beta_2(u_{ij} - \bar{u}_{..}) + \in_{ij}$$

It can be shown that the LSEs of β_1 and β_2 are given by

$$\hat{\beta}_1 = \frac{\begin{vmatrix} E_{xu} & E_{xy} \\ E_{uu} & E_{uy} \end{vmatrix}}{\begin{vmatrix} E_{xu} & E_{xx} \\ E_{uu} & E_{xu} \end{vmatrix}} \quad \text{and} \quad \hat{\beta}_2 = \frac{\begin{vmatrix} E_{xx} & E_{xy} \\ E_{xu} & E_{uy} \end{vmatrix}}{\begin{vmatrix} E_{xu} & E_{xx} \\ E_{uu} & E_{xu} \end{vmatrix}}.$$

For more details of the above formulas, the reader is referred to an Auburn University Master's Thesis (June 1993) by Hung-Hsiang (Kevin) Hsu directed by this author.

Exercise 30(g). Repeat the ANCOVA analysis of Problem 15.15 on page 610 (p. 594 of the 6th edition) assuming that there were five randomly selected operators used as blocks. First use Minitab's GLM to obtain the correct answers and then show all calculation on a MS Excel file.