

1. Rao 15.3

(a)

$$Y_{ij} = \mu + \alpha_i + B_j + E_{ij}$$

where α_i denote fixed effects for serum amylase determination method, B_j denote random subject effects, assumed iid $N(0, \sigma_B^2)$ and E_{ij} denote iid $N(0, \sigma^2)$ random errors, independent of B_j .

(b) ANOVA table

Source	DF	Sum of Squares	Mean Square	F	<i>p</i> -value
method	3	116979	38993	23.78	.0001
patient	5	2157788	431558	263.21	.0001
Residual	15	24594	1639.577778		

The *p*-value for the *F*-test of $H_0 : \alpha_i \equiv 0$ is less than .0001, providing strong evidence of a method effect.

(c) $\theta = \alpha_4 - \alpha_1$, the difference between the population means of methods four and one. A 95% confidence interval given by

$$\bar{y}_{4+} - \bar{y}_{1+} \pm t(.025, 15) \sqrt{MS(E) \frac{2}{6}}$$

which works out to $154 \pm 2.13(23.38)$ or (104, 204).

(d) No, the two sample mean amylase values differ significantly. In other words, the chances observing such a large difference between the two methods in this experiment, under the hypothesis that the methods are the same in the long run, are so small that we may rule out this hypothesis as implausible in light of these data.

(e) Simultaneous 95% upper bounds for $\mu + \alpha_1$ and $\mu + \alpha_2$ of the form $\bar{y}_{i+} + t(.05, 15) \sqrt{MS(E) \frac{1}{6}}$ under a model with fixed patient effects, or

$$\begin{aligned} i = 1 & \quad 700.3 + 2.49(16.5) = 741.5 \\ i = 3 & \quad 697 + 2.49(16.5) = 738.2 \end{aligned}$$

Under a model with random patient effects, the standard error of a sample method mean must be estimated with a linear combination of mean squares:

$$\begin{aligned} \bar{Y}_{i+} &= \mu + \alpha_i + \bar{B} + \bar{E}_{i+} \\ Var(\bar{Y}_{i+}) &= \frac{1}{b}(\sigma_B^2 + \sigma^2) \\ &= \frac{1}{ab}[EMS(B) + (a-1)MS(E)] \end{aligned}$$

This variance may be estimated by substitution of observed mean squares, but the distribution of the corresponding *t*-statistic requires Satterthwaite approximation. In this case, $\widehat{SE}(\bar{y}_{i+}) = 134.9$ on $\widehat{df} = 5.11$ leading to the simultaneous upper bounds

$$\begin{aligned} i = 1 & \quad 700.3 + 3.14(134.9) = 1123.3 \\ i = 3 & \quad 697 + 3.14(134.9) = 1119.9 \end{aligned}$$

I tricked SAS into obtaining this upper endpoints with the following code:

```

proc mixed;
  class patient method;
  model y=method/ddfm=satterth;
  random patient;
  lsmeans method/cl alpha=.025;
run;

```

- (f) One would not be able to control for the rather large patient-to-patient variability and would thus have much less power to detect method effects.

2. 15.11 (solution appears later in this document)

3. A person's blood-clotting ability is typically expressed in terms of a ...

- (a) Carry out a paired t -test of the hypothesis that prothrombin time is unaffected by aspirin. $\bar{D} = 0.108$, $s_D = 0.5$, $t = \bar{d}/(s_D/\sqrt{12}) = 0.74(NS)$
- (b) Carry out an F -test of the same hypothesis treating subjects as blocks in an analysis for a RCBD.

Source	df	SS	MS	F
Blocks	11	10.21	0.93	
Aspirin	1	0.07	0.07	0.55 (NS)
Error	11	1.41	0.13	
Total	23	11.70		

- (c) Show that, in general, the paired t -test is equivalent to the F -test for the RCBD with block size equal to 2. It suffices to show that $t^2 = F$ where $t = \bar{D}/(s_D/\sqrt{n})$ and $F = SS(trt)/MS(E)$.

$$\begin{aligned}
SS[Trt] &= \sum_i \sum_j (\bar{Y}_{i+} - \bar{Y}_{++})^2 \\
&= \frac{n}{2} \sum_i (\bar{Y}_{i+} - \bar{Y}_{++})^2 \\
&= n \sum_i \left(\bar{Y}_{i+} - \frac{1}{2}(\bar{Y}_{1+} + \bar{Y}_{2+}) \right)^2 \\
&= n \left(\frac{\bar{Y}_{1+} - \bar{Y}_{2+}}{2} \right)^2 + \frac{n}{2} \bar{D}^2 \\
SS[E] &= \sum_i \sum_j (Y_{ij} - \bar{Y}_{i+} - \bar{Y}_{+j} + \bar{Y}_{++})^2 \\
&= \sum_j \left[(Y_{1j} - \bar{Y}_{1+} - \frac{1}{2}(Y_{1j} + Y_{2j}) + \frac{1}{2}(\bar{Y}_{1+} + \bar{Y}_{2+}))^2 \right. \\
&\quad \left. + (Y_{2j} - \bar{Y}_{2+} - \frac{1}{2}(Y_{1j} + Y_{2j}) + \frac{1}{2}(\bar{Y}_{1+} + \bar{Y}_{2+}))^2 \right] \\
&= \sum_j 2 \left(\frac{Y_{1j} - Y_{2j}}{2} - \frac{\bar{Y}_{1+} - \bar{Y}_{2+}}{2} \right)^2 \\
&= \frac{1}{2} \sum_j (D_j - \bar{D})^2 \\
&= \frac{n-1}{2} S_D^2 \\
SS[Trt]/(SS[E]/(n-1)) &= \frac{n\bar{D}^2}{S_D^2} = t^2
\end{aligned}$$

(d) Consider the mixed model

$$Y_{ij} = \mu + \tau_i + B_j + E_{ij}$$

where $B_j \stackrel{iid}{\sim} N(0, \sigma_B^2)$ and $E_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ with $B \perp E$ for $i = 1, \dots, a$ and $j = 1, \dots, b$.

i. Show that $E[MS(block)] = \sigma^2 + a\sigma_B^2$

$$\begin{aligned} SS(Block) &= a \sum_j (\bar{Y}_{+j} - \bar{Y}_{++})^2 \\ &= a \sum_j (B_j - \bar{B} + \bar{E}_{+j} - \bar{E}_{++})^2 \\ &= a \left[\sum_j (B_j - \bar{B})^2 + \sum_j B_j (\bar{E}_{+j} - \bar{E}_{++}) + \sum_j (\bar{E}_{+j} - \bar{E}_{++})^2 \right] \\ E[MS(Block)] &= a\sigma_B^2 + 0 + a \frac{\sigma^2}{a} \\ E[MS(Block)] &= a\sigma_B^2 + \sigma^2 \end{aligned}$$

ii. Use this result to estimate the variance component for subject effects in a mixed model for the prothrombin data.

$$\hat{\sigma}_B^2 = \frac{MS[Block] - MS[E]}{a} = \frac{0.93 - 0.13}{2} = 0.4$$

iii. Report an estimate of the intra-subject correlation. Is the scatterplot above consistent with this estimate?

Yes:

$$\hat{\rho}_I = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_B^2 + \hat{\sigma}^2} = 0.4 / (0.4 + 0.13) = 0.75$$

which is the same as the correlation coefficient for *before* and *after* and the measurements are clearly positively correlated: those with slow times before tend to be slow after and there is subject-to-subject variability.

4. Ott and Longnecker 15.10 (p. 889) I chose to fit a mixed model with fixed effects for blend and random effects for driver and model, though a model with all fixed effects is also reasonable.

(a)

$$Y_{ijk} = \mu + D_i + C_j + \tau_k + E_{ij}$$

(Here τ_k is determined by the design.)

- (b) To get fixed effects of blend (the τ_k 's), I created three indicator variables **blend1**, **blend2** and **blend3** as follows:

$$\text{blend1}_j = \begin{cases} 1 & \text{if obs}_j \text{ is blend 1} \\ -1 & \text{if obs}_j \text{ is blend 4} \\ 0 & \text{else} \end{cases}$$

$$\text{blend2}_j = \begin{cases} 1 & \text{if obs}_j \text{ is blend 2} \\ -1 & \text{if obs}_j \text{ is blend 4} \\ 0 & \text{else} \end{cases}$$

$$\text{blend3}_j = \begin{cases} 1 & \text{if obs}_j \text{ is blend 3} \\ -1 & \text{if obs}_j \text{ is blend 4} \\ 0 & \text{else} \end{cases}$$

to get the fixed effects estimates below:

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	22.2438	3.9549	3	5.62	0.0111
blend1	0.2563	0.8990	6	0.29	0.7852
blend2	2.7313	0.8990	6	3.04	0.0229
blend3	-4.1938	0.8990	6	-4.66	0.0035

and the estimated variance components below:

Covariance Parameter Estimates	
Cov Parm	Estimate
cartype	61.8700
driver	-0.3833
Residual	4.3106

There is some evidence of blend differences ($p = 0.0151$)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	869.9756250	96.6639583	22.42	0.0006
Error	6	25.8637500	4.3106250		
Corrected Total	15	895.8393750			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Driver	3	8.3318750	2.7772917	0.64	0.6143
Model	3	755.3718750	251.7906250	58.41	<.0001
Blend	3	106.2718750	35.4239583	8.22	0.0151

5. Rao 15.11

(a) Model $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + B_k + E_{ijk}$ where

- $i = 1, 2$ indexes bed height
- $j = 1, 2, 3$ indexes row spacing
- $k = 1, 2, 3, 4$ indexes block

$B_k \stackrel{iid}{\sim} N(0, \sigma^2)$ and $E_{ijk} \stackrel{iid}{\sim} N(0, \sigma_B^2)$, B and E are independent.

(b) The ANOVA table produced by PROC MIXED is given below.

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
a	1	14.400556	14.400556	Var(Residual) + Q(a,a*b)
b	2	127.853333	63.926667	Var(Residual) + Q(b,a*b)
a*b	2	9.231111	4.615556	Var(Residual) + Q(a*b)
block	2	236.893333	118.446667	Var(Residual) + 6 Var(block)
Residual	10	44.586667	4.458667	Var(Residual)

Covariance Parameter Estimates

Cov Parm	Estimate
block	18.9980
Residual	4.4587

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
a	1	10	3.23	0.1025
b	2	10	14.34	0.0012
a*b	2	10	1.04	0.3903

(c) 95% confidence limits given below

Estimates						
Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha
a_hi - a_lo	-1.7889	0.9954	10	-1.80	0.1025	0.05
Label	Lower	Upper				
a_hi - a_lo	-4.0068	0.4290				

(d) The Tukey procedure indicates that differences involving the closest row spacing (0.25m) are significant (adjusted p -values all less than .01).

Least Squares Means						
Effect	b	Estimate	Standard Error	DF	t Value	Pr > t
b	0.25	22.7500	2.6600	10	8.55	<.0001
b	0.5	17.2167	2.6600	10	6.47	<.0001
b	1	16.9833	2.6600	10	6.38	<.0001

Differences of Least Squares Means							
Effect	b	_b	Estimate	Standard Error	DF	t Value	Pr > t
b	0.25	0.5	5.5333	1.2191	10	4.54	0.0011
b	0.25	1	5.7667	1.2191	10	4.73	0.0008
b	0.5	1	0.2333	1.2191	10	0.19	0.8520

Differences of Least Squares Means						
Effect	b	_b	Adjustment	Adj P		
b	0.25	0.5	Tukey-Kramer	0.0028		
b	0.25	1	Tukey-Kramer	0.0021		
b	0.5	1	Tukey-Kramer	0.9800		

6. Randomization exercise I used the command `sample(1:5)` two times to get the following two permutations of positive integers less than 6:

(5, 3, 2, 1, 4) and (2, 5, 4, 3, 1)

Using these to permute rows and columns respectively, I got the following design

1	2	3	4	5	→	2	3	4	5	1	→	3	1	5	4	2
5	1	2	3	4		4	5	1	2	3		5	3	2	1	4
4	5	1	2	3		5	1	2	3	4		1	4	3	2	5
3	4	5	1	2		1	2	3	4	5		2	5	4	3	1
2	3	4	5	1		3	4	5	1	2		4	2	1	5	3

7. Rao 14.1

- (a) Y_{ij} denotes percentage protein for seeds produced by i^{th} plant in j^{th} plot.

$$Y_{ij} = \mu + T_i + E_{ij}$$

where T_1, T_2, T_3 are i.i.d. $N(0, \sigma_T^2)$ random plant effects and E_{ij} are i.i.d. $N(0, \sigma^2)$ error terms

- (b) $H_0 : \sigma_T^2 = 0$
 (c) $H_0 : \mu = 40$
 (d) σ^2

8. Rao 14.2

- (a) (14.2a) μ is the overall mean serum cholesterol level for a randomly sampled subject. T_i is the random effect of subject i . E_{ij} is the random error of the j^{th} sample taken on the i^{th} subject.
 (b) (14.2c) σ_T^2 quantifies variance due to effect of randomly sampled subject. σ^2 quantifies variance due to effect randomly sampled test within subject. If $\sigma^2 = 2\sigma_T^2$, then the subject effect accounts for a third of total variability in cholesterol.

9. Rao 14.5 (I changed $y_{21} = 28.6$ to $y_{21} = 38.6$)

	Source	df	SS	MS	F	p-value	EMS
(a)	Plant	9	74	8.2	4.25	.0034	$\sigma^2 + 3\sigma_{plant}^2$
	Error	20	38.7	1.9			
	Total	20					

- (b) Yes, the plant-to-plant variability of protein is significant. A test that heritability is zero has p -value .0034.
 (c)

$$\begin{aligned} \hat{\sigma}^2 &= MS[E] \\ &= 1.9 \\ \hat{\sigma}_P^2 &= \frac{1}{3}(MS[Plant] - MS[E]) \\ &= (8.2 - 1.9)/3 \\ &= 2.1 \end{aligned}$$

(These two variance components are in the same ballpark, heritability is close to 50%.)

- (d)

$$\widehat{CV} = \frac{\sqrt{\hat{\sigma}_P^2 + \hat{\sigma}^2}}{|\bar{Y}_{++}|} = .05$$

- (e) Required critical values: $F(0.025, 9, 20) = 2.84$, $F(.975, 9, 20) = 0.27$
 Observed F -ratio: $F_{obs} = 4.25$
 Confidence interval:

$$\left(\frac{4.25 - 2.84}{4.25 + 2 * 2.84}, \frac{4.25 - 0.27}{4.25 + 2 * 0.27} \right)$$

or

$$(0.14, 0.83)$$

(This is an interval estimate of the proportion of variability in protein content ($\sigma_P^2 + \sigma^2$) that may be ascribed to plant genetics.)

- (f) A 95% confidence interval for μ is given by

$$\bar{Y}_{++} \pm t(.025, 9) \sqrt{\frac{MS(plant)}{40}}$$

or $39.4 \pm 2.26(0.52)$ or $(38.2, 40.6)$, an interval estimate which contains $\mu = 40$ as a plausible value, not inconsistent with the data. Alternatively, a lower confidence bound is given by $\bar{Y}_{++} - t(.05, 9)SE$, which is less than 40, leading to the same conclusion.

10. Rao 14.6

- (a) $\hat{\sigma}^2 = MS[E] = 3.94$ and $\hat{\sigma}_S^2 = \frac{MS[T] - MS[E]}{2-1} = 103.5$. $\hat{\sigma}_T^2$ is much bigger than $\hat{\sigma}^2$. (So, there is a big intrasubject correlation and not much error in the measurement of serum cholesterol relative to variability in population.)
- (b) A 95% confidence interval for ρ_I is given by ... (we skipped this material)
- (c) A 99% confidence interval for μ is given by

$$\bar{Y}_{++} \pm t(.025, 7) \sqrt{MS(Subj)/16} \quad \text{or} \quad 181.4 \pm 2.36 \sqrt{\frac{210.9}{16}} \quad \text{or} \quad (173, 190)$$