Solution: Sample Midterm

1.

(a) $E(Y_{ij}) = \mu, Var(Y_{ij}) = \sigma_1^2 + \sigma^2$.

(b) When $j \neq j'$

$$\begin{aligned}
Cov(Y_{ij}, Y_{ij'}) &= Cov(\mu + \tau_i + \varepsilon_{ij}, \mu + \tau_i + \varepsilon_{ij'}) \\
&= Var(\tau_i) = \sigma_1^2, \\
Corr(Y_{ij}, Y_{ij'}) &= \frac{Cov(Y_{ij}, Y_{ij'})}{\sqrt{Var(Y_{ij})Var(Y_{ij'})}} = \frac{\sigma_1^2}{\sqrt{(\sigma_1^2+\sigma^2)(\sigma_1^2+\sigma^2)}} = \frac{\sigma_1^2}{\sigma_1^2+\sigma^2}.
\end{aligned}$$

(c) Note that $\bar{Y}_{\cdot\cdot} = \mu + \bar{\tau} + \bar{\varepsilon}_{\cdot\cdot}$ and hence

$$\begin{aligned}
E(\bar{Y}_{\cdot\cdot}) &= \mu, \\
Var(\bar{Y}_{\cdot\cdot}) &= Var(\bar{\tau}) + Var(\bar{\varepsilon}_{\cdot\cdot}) = \sigma_1^2/k + \sigma^2/(rk) = \frac{r\sigma_1^2 + \sigma^2}{rk}.
\end{aligned}$$

(d) A fairly easy argument shows that

$$\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} = (\tau_i - \bar{\tau}) + (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}).$$

Since $\{\tau_i\}$ and $\{\varepsilon_{ij}\}$ are mutually uncorrelated, we have

$$\begin{aligned}
E(SSTR) &= r\sum E(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \\
&= r\sum E(\tau_i - \bar{\tau})^2 + r\sum E(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2 \\
&= r(k-1)\sigma_1^2 + r(k-1)\sigma^2/r = (k-1)(r\sigma_1^2 + \sigma^2)., \\
E(MSTR) &= \frac{E(SSTR)}{k-1} = r\sigma_1^2 + \sigma^2.
\end{aligned}$$

An unbiased estimator of $Var(\bar{Y}_{\cdot\cdot})$ is $s^2(\bar{Y}_{\cdot\cdot}) = MSTR/(rk)$.

2.

(a) Let Y_{ijk} be the k^{th} replication when factor B is at level j and factor A is at level i. Here numbers of levels of factors A and B are $a = 3$ and $b = 3$, respectively. The model is

$$Y_{ijk} = \mu_{\cdot\cdot} + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk},$$

where $\mu_{\cdot\cdot}$ is the overall mean, $\{\varepsilon_{ijk}\}$ are iid $N(0, \sigma^2)$, $\{\alpha_i\}$ are the main effects of factor A, $\{\beta_{j(i)}\}$ is the effect of factor B nested in A. The constraints are

$$\sum \alpha_i = 0 \text{ and, for each } i,\ \sum_j \beta_{j(i)} = 0\,.$$

(b) ANOVA table

Source	df	SS	MS	F	p-val
A	$a-1=2$	641.4344	320.7172	30.9074	<0.001
B(A)	$a(b-1)=6$	696.8066	116.1344	11.1918	Close to 0.001
Error	$(n-1)ab=9$	93.3900	10.3767		
Total	$nab-1=17$	1431.3767			

(c) Are regions different?

Test $H_0: \alpha_i = 0$ for all i against H_1 :not all α_i equal zero.

$F^* = 30.9074$, $F(0.95; 2, 9) = 4.26$. Since $F^* > F(0.95; 2, 9)$, we reject H_0.

P-value<0.001

Are there differences between the auditors in at least one of the regions?

Test $H_0: \beta_{j(i)} = 0$ for all j and i against H_1 :not all $\beta_{j(i)}$ equal zero.

$F^* = 11.1918$, $F(0.95; 6, 9) = 3.37$. Since $F^*(0.95; 6, 9)$, we reject H_0.

There are differences between the auditors in at least one of the regions.

The p-value is close to 0.001.

(d) Tests for the differences between the auditors at each of the regions are given below.

Region 1. $H_0: \beta_{j(1)} = 0$ for all j, H_1 : not all $\beta_{j(1)}$ equal zero.

$$\begin{aligned} SSB(A_1) &= 231.04, df = b-1 = 2, MSB(A_1) = SSB(A_1)/2 = 115.52, \\ F^* &= MSB(A_1)/MSE = 11.1326, \text{ df} = (2,9),\ F(0.95; 2, 9) = 4.26. \end{aligned}$$

Since $F^* > F(0.95; 2, 9)$, we reject H_0 and conclude that the auditor effect is present in region 1.

Region 2. $H_0: \beta_{j(2)} = 0$ for all j, H_1 : not all $\beta_{j(2)}$ equal zero.

$$\begin{aligned} SSB(A_2) &= 424.563, df = b-1 = 2, MSB(A_2) = SSB(A_2)/2 = 212.35, \\ F^* &= MSB(A_2)/MSE = 20.4614, \text{ df} = (2,9), F(0.95; 2, 9) = 4.26. \end{aligned}$$

Since $F^* > F(0.95; 2, 9)$, we reject H_0 and conclude that the auditor effect is present in region 2.

Region 3. $H_0: \beta_{j(3)} = 0$ for all j, H_1 : not all $\beta_{j(3)}$ equal zero.

$$\begin{aligned} SSB(A_3) &= 41.123, df = b-1 = 2, MSB(A_3) = SSB(A_3)/2 = 20.5615, \\ F^* &= MSB(A_3)/MSE = 1.9818. \text{ df} = (2,9), F(0.95; 2, 9) = 4.26. \end{aligned}$$

Since $F^* < F(0.95; 2, 9)$, we cannot reject H_0.

3.

(a) This is a repeated measures design. Let Y_{ij} be the response of the i^{th} student in j^{th} week. The model is

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij}, j = 1, \ldots, r = 3, i = 1, \ldots, s = 10,$$

where $\mu_{..}$ is the overall mean, $\{\rho_i\}$ are iid $N(0, \sigma_\rho^2)$, $\{\tau_j\}$ are weak effects with $\sum \tau_j = 0$, $\{\varepsilon_{ij}\}$ are iid $N(0, \sigma^2)$, and $\{\rho_i\}$ and $\{\varepsilon_{ij}\}$ are independent.

(b) ANOVA table

Source	df	SS	MS	F	p-val
Subject	$s - 1 = 9$	654.3	72.7	4.7899	between 0.001 and 0.005
Week	$r - 1 = 2$	204.8	102.4	6.7467	<0.001
Error	$(s-1)(r-1) = 18$	273.2	15.1778		
Total	$sr - 1 = 29$	1132.2			

(c) Estimate of σ_ρ^2

$$s_\rho^2 = \frac{MSS - MSE}{r} = \frac{72.7 - 15.1778}{3} = 19.1741.$$

Proportion of variability in stress due to subjects is $\sigma_\rho^2/(\sigma_\rho^2 + \sigma^2)$ and its estimate is

$$\frac{s_\rho^2}{s_\rho^2 + MSE} = \frac{19.1741}{19.1741 + 15.1778} = 0.5582.$$

(d) Estimate of $\tau_j - \tau_{j'}$, $j \neq j'$, is $\bar{Y}_{.j} - \bar{Y}_{.j'}$, $Var(\bar{Y}_{.j} - \bar{Y}_{.j'}) = (2/s)\sigma^2$ and hence

$$s^2(\bar{Y}_{.j} - \bar{Y}_{.j'}) = (2/10)(15.1778) = 3.03556,\ s(\bar{Y}_{.j} - \bar{Y}_{.j'}) = 1.7423.$$

Tukey Multiplier is

$$T = (1/\sqrt{2})q(0.95; 3, 18) = (1/\sqrt{2})(3.6093) = 2.5527.$$

Simultaneous 95% confidence intervals are

$$\begin{aligned}
\tau_1 - \tau_2 &: \bar{Y}_{.1} - \bar{Y}_{.2} \pm (2.5527)(1.7423), \text{ i.e., } (-7.65, 1.25),\\
\tau_1 - \tau_3 &: \bar{Y}_{.1} - \bar{Y}_{.3} \pm (2.5527)(1.7423), \text{ i.e., } (-1.25, 7.65),\\
\tau_2 - \tau_3 &: \bar{Y}_{.2} - \bar{Y}_{.3} \pm (2.5527)(1.7423), \text{ i.e., } (1.95, 10.85).
\end{aligned}$$

Only the last interval excludes zero. Thus we conclude that the mean stress levels during and one week after are different. There is no evidence to conclude that there is difference in mean stress levels before and during the exam, and before and after the exams.

4.

(a) Number of levels of factors A, B and C are $a = 3, b = 2, c = 2$. The number of replications is $n = 2$.

ANOVA table

Source	df	SS	MS	F	p-val
A	$a - 1 = 2$	254.75	127.375	179.8244	< 0.001
B	$b - 1 = 1$	43.375	43.375	61.2356	< 0.001
C	$c - 1 = 1$	22.042	22.042	31.1183	$< 0,001$
AB	$(a - 1)(b - 1) = 2$	5.250	2.625	3.7059	between 0.05 and 0.1
AC	$(a - 1)(c - 1) = 2$	0.583	0.2915	0.4115	> 0.5
BC	$(b - 1)(c - 1) = 1$	1.042	1.042	1.4711	between 0.1 and 0.5
ABC	$(a - 1)(b - 1)(c - 1) = 2$	1.083	0.5415	0.7645	Close to 0.5
Error	$(n - 1)abc = 12$	8.500	0.70833		
Total	$nabc - 1 = 23$	336.625			

(b) Based on the ANOVA table above, we may drop the interactions AB, AC, BC and ABC.

(c) If we drop the two-factor and three factor interactions, then the model is

$$Y_{ijkm} = \mu_{...} + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijkm}, \; m = 1, 2, k = 1, 2, j = 1, 2, i = 1, 2, 3,$$

where $\mu_{...}$ is the overall mean, $\{\alpha_i\}$ are the main effects of factor A, $\{\beta_j\}$ are the main effects of factor B, $\{\gamma_k\}$ are the main effects of factor C, and $\{\varepsilon_{ijkm}\}$ are iid $N(0, \sigma^2)$.

(d) The ANOVA table for model in part (c) is

Source	df	SS	MS	F	p-val
A	$a - 1 = 2$	254.75	127.375	146.0486	< 0.001
B	$b - 1 = 1$	43.375	43.375	50.0745	< 0.001
C	$c - 1 = 1$	22.042	22.042	25.4465	$< 0,001$
Error	19	16.458	0.86621		
Total	$nabc - 1 = 23$	336.625			

Note that the SSE for the model in part(c) (call it SSE_{new}) is

$$SSE_{new} = SSAB + SSAC + SSBC + SSABC + SSE,$$

where SSE here refers to the residual sum of squares for the saturated model (from part (b)). The degrees of freedom also are added up and hence

$$df(SSE_{new}) = 2 + 2 + 1 + 2 + 12 = 19.$$

Mathematically, we may also think of $df(SSE_{new})$ as $n_T = nabc$ minus the number of (mean) parameters in the model in part (c). Thus

$$df(SSE_{new}) = nabc - (a + b + c - 2) = 24 - 5 = 19.$$