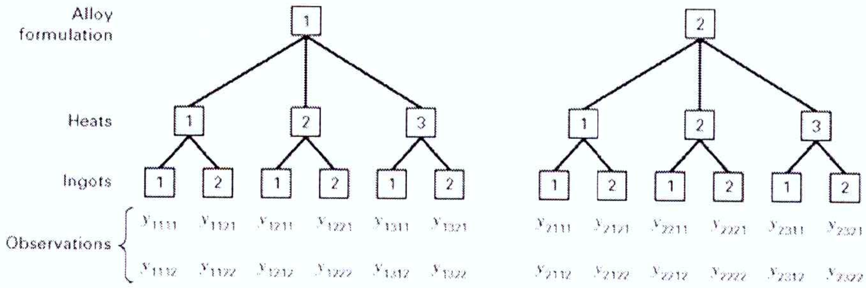
Stat 504: Final Exam (June 15. 2015) 학번: 이름: Solution.

1. Consider the three-stage nested design shown in the Figure below to investigate alloy hardness. A foundry wishes to investigate the hardness of **two** different formulations of metal alloy. **Three** heats of each alloy formulation are prepared, **two** ingots are selected at random from each heat for testing and **two** hardness measurements are made on each ingot.



The following table shows the part of ♠ANOVA table obtained using SAS/GLM procedure.

Source	DF	SS	
Alloy	?	315	1
Heat(Alloy)	?	6,453	2x2=4
Ingot(Alloy*Heat)	?	2,226	2x3x1=6
Error	?	2,141	12
Total	23		

(a) Specify the linear model under the assumption that only **Alloy and Heat are fixed effect**. Explain all the notations you used. Also state all the hypotheses(actually 3) you can test using the given information, and perform tests. Specify the F statistics, degree of freedom. What is your conclusion?

9

$$y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \eta_{ijk} + \varepsilon_{ijkl}$$

Alloy Heat Ingot Error

$\eta_{ijk} \sim N(0, \sigma_\eta^2)$

$\varepsilon_{ijkl} \sim N(0, \sigma^2)$.

0.85 H_0 : $\alpha_1 = \alpha_2$. (1, 6)

4.35 H_0 : ~~$\beta_1 = \beta_2 = \beta_3$~~ all $\beta_{j(i)}$ are the same. (4, 6)

2.08 H_0 : $\sigma_\eta^2 = 0$. (6, 12).

(b) Specify the linear model under the assumption that only **Alloy is a fixed effect**. Explain all the notations you used. Also state all the hypotheses(actually 3) you can test using the given information, and perform tests. Specify the F statistics, degree of freedom. What is your conclusion?

$\mathcal{J}$

$H_0: \alpha_1 = \alpha_2$ $(1, 4)$

$H_0: \sigma_\beta^2 = 0$ $(4, 6)$

$H_0: \sigma_\eta^2 = 0$ $(6, 12)$

$\beta_{ij} \sim N(0, \sigma_\beta^2)$

$\eta_{ijk} \sim N(0, \sigma_\eta^2)$

$\varepsilon_{ijk} \sim N(0, \sigma^2)$

(c) Specify the linear model under the assumption that all factors are random. Explain all the notations you used. Also state all the hypotheses(actually 3) you can test using the given information, and perform tests. Specify the F statistics, degree of freedom. Provide the ANOVA-based estimate of all variance components in your model.

$\mathcal{J}$

$H_0: \sigma_\alpha^2 = 0$

$H_0: \sigma_\beta^2 = 0$

$H_0: \sigma_\eta^2 = 0$

$\alpha_i \sim N(0, \sigma_\alpha^2)$

$\beta_{ij} \sim N(0, \sigma_\beta^2)$

$\eta_{ijk} \sim N(0, \sigma_\eta^2)$

$\varepsilon_{ijkl} \sim N(0, \sigma^2)$

	EMS
Alloy	$\sigma^2 + n\sigma_\eta^2 + cn\sigma_\beta^2 + bcn\sigma_\alpha^2$
Heat(Alloy)	$\sigma^2 + n\sigma_\eta^2 + cn\sigma_\beta^2$
Ingot(Alloy*Heat)	$\sigma^2 + n\sigma_\eta^2$
Error	σ^2

$\boxed{n = 2, \quad c = 2, \quad b = 3, \quad a = 2.}$

2. Consider an investigation to study the effects of two irrigation methods(factor A) and two fertilizers(factor B) on yield of crop, using 10 available fields. Researcher considered a split-plot design in which the layout is as below.

Irrigation Method	1					2				
Field	1	2	3	4	5	1	2	3	4	5
Fertilizer 1	43	40	31	27	36	63	52	45	47	54
2	48	43	36	30	39	70	53	48	51	57

For the analysis of data, researchers considered a mixed effects model, $i \rightarrow$

$$y_{ijk} = \mu + \underline{\alpha_i} + \eta_{ij} + \underline{\tau_k} + \underline{\gamma_{ik}} + \epsilon_{ijk},$$

where y_{ijk} is the yield from the j-th field which is assigned to irrigation method i at k-th fertilizer and α_i, τ_k and γ_{ik} are the fixed irrigation effect, fertilizer effect and interaction effect respectively. In the model, η_{ij} and error ϵ_{ijk} are independent random components such that

α γ_i j k

$$\eta_{ij} \sim NID(0, \sigma_\eta^2), \quad \epsilon_{ijk} \sim NID(0, \sigma_\epsilon^2).$$

(a) Under the restriction $\alpha_2 = \tau_2 = \gamma_{2k} = \gamma_{i2} = 0$ for all i and k, the mixed effect model can be written as a matrix form

$$\mathbf{y} = X\beta + Z\mathbf{u} + \epsilon.$$

Determine the first and last row of the matrix X and Z.

X

$$\begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{141} \\ y_{142} \\ y_{151} \\ y_{152} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \\ y_{241} \\ y_{242} \\ y_{251} \\ y_{252} \end{pmatrix} = \begin{pmatrix} \boxed{(1, 1, 1, 1)} \\ \vdots \\ \boxed{(1, 0, 0, 0)} \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \cancel{\alpha_2 = 0} \\ \tau_1 \\ \cancel{\tau_2 = 0} \\ \gamma_{11} \\ \cancel{\gamma_{12} = 0} \\ \cancel{\gamma_{21} = 0} \\ \cancel{\gamma_{22} = 0} \end{pmatrix} + \begin{pmatrix} \boxed{(1, 0 \cdots 0)} \\ \vdots \\ \boxed{(0, \cdots, 0, 1)} \end{pmatrix} \begin{pmatrix} \eta_{11} \\ \eta_{12} \\ \eta_{13} \\ \eta_{14} \\ \eta_{15} \\ \eta_{21} \\ \eta_{22} \\ \eta_{23} \\ \eta_{24} \\ \eta_{25} \end{pmatrix}$$

$n \times 1$ $n \times 4$ $n \times 10$

(b) According to the model fit by the researchers, what is the variance of a single response in terms of the variance components in the model?

5

$$V(y_{ijk}) = \sigma_\eta^2 + \sigma_\varepsilon^2 .$$

(c) According to the model fit by the researchers, what is the correlation between y_{111} and y_{112} in terms of the variance components in the model?

5

$$\mathrm{cov}(y_{111}, y_{112}) = \mathrm{cov}(\cancel{\mu} + \cancel{\alpha_1} + \eta_{11} + \cancel{\tau_1} + \cancel{\gamma_{11}} + \varepsilon_{111}, \cancel{\mu} + \cancel{\alpha_1} + \eta_{11} + \cancel{\tau_1} + \cancel{\gamma_{11}} + \varepsilon_{112})$$

$$= \sigma_\eta^2$$

$$\Rightarrow \mathrm{corr} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} .$$

(d) Under the given mixed model, construct a column of *source of variation* and *degree of freedom* of ANOVA table and specify the sources of error that should be used used to test the each three fixed effects.

10

Source	d.f
Method	1
Whole plot	$2(5-1) = 8$
Fertilizer	1
Fert * method	1
subplot	8
Total	19

3. An experimental study was made to investigate the effects of height of the shelf display(factor A: bottom, middle, top) and the width of the shelf display(factor B: regular, wide) on sales of this bakery's bread (Y, measures in cases) during the experimental period. Twelve supermarkets, similar in terms of sales volume and clientele, were utilized in this study. The six treatments were assigned randomly to two stores each. The results are presented as following.

	Factor B	
Factor A	$j=1$	$j=2$
$i=1$	47, 43	46, 40
$i=2$	62, 68	67, 71
$i=3$	41, 39	42, 46

Researcher considered the following two-way ANOVA model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim NID\left(0, \sigma^2\right)$$

and for all i and j, assume

$$\sum_{i=1}^{3} \alpha_i = 0, \quad \sum_{j=1}^{2} \beta_j = 0, \quad \sum_{i=1}^{3} \delta_{ij} = \sum_{j=1}^{2} \delta_{ij} = 0.$$

(a) Under the Model, derive the least square estimator of α_i, β_j and δ_{ij}.

Source	d.f
A	2
B	1
A*B	2
Error	6
Total	11

7

$$\hat{\alpha}_i = (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})$$

$$\hat{\beta}_j = (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot})$$

$$\hat{\delta}_{ij} = (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot})$$

$$\alpha_3 = 0 - \alpha_1 - \alpha_2$$

$$\beta_2 = -\beta_1$$

$$\delta_{i2} = -\delta_{i1}$$

$$\delta_{3j} = -\delta_{1j} - \delta_{2j}$$

i	j	$\boxed{x_1}$ α_1	$\boxed{x_2}$ α_2	$\boxed{x_3}$ β_1	$\boxed{x_1 x_3}$	$\boxed{x_2 x_3}$
1	1	1	0	1	1	0
2	1	0	1	1	0	1
3	1	-1	-1	1	-1	-1
1	2	1	0	-1	-1	0
2	2	0	1	-1	0	-1
3	2	-1	-1	-1	1	1

(b) Test $H_0 : \mu_{11} = (0.5)(\mu_{21} + \mu_{31})$, where $\mu_{ij} = \mu + \alpha_i + \beta_j + \delta_{ij}$. Just provide the formula for the test statistics and its distribution with degree of freedom.

8

Note that $E(\bar{Y}_{ij.}) = \mu_{ij}$. & $Var(\bar{Y}_{ij.})$

$$= Var(\mu + \alpha_i + \beta_j + \delta_{ij} + \bar{\varepsilon}_{ij.}) = Var(\bar{\varepsilon}_{ij.}) = \frac{\sigma^2}{2}$$

$$t = \frac{\bar{y}_{11.} - \frac{1}{2}(\bar{y}_{21.} + \bar{y}_{31.})}{\sqrt{MSE/2}} \sim t\,(6).$$

↳ d.f.

$\alpha_3 = -\alpha_1 - \alpha_2$

(c) Let x variables be defined as below.

10

$$x_1 = \begin{cases} 1 & \text{if case from level 1 for factor A,} \\ -1 & \text{if case from level 3 for factor A,} \\ 0 & \text{otherwise.} \end{cases}, \quad x_2 = \begin{cases} 1 & \text{if case from level 2 for factor A,} \\ -1 & \text{if case from level 3 for factor A,} \\ 0 & \text{otherwise.} \end{cases},$$

$$x_3 = \begin{cases} 1 & \text{if case from level 1 for factor B,} \\ -1 & \text{if case from level 2 for factor B.} \end{cases}$$

$\boxed{\text{Hint}}$:

Rewrite the two-way ANOVA model using $x_{1,ijk}$, $x_{2,ijk}$ and $x_{3,ijk}$ and parameters described in the two-way ANOVA model ($\mu, \alpha_i, \beta_j, \delta_{ij}, \epsilon_{ijk}$), where $x_{l,ijk}$ is the value of x_l for the observation ijk.

✓

π_i

A	B	x_1	x_2	x_3	μ	α_1	α_2	α_3	β_1	β_2	δ_{11}	δ_{12}	δ_{13}	δ_{21}	δ_{22}
1	1	1	0	1	1	1	0	~~0~~ −1	1	0	1	0	0	0	0 0
1	2	1	0	−1	1	1	0	−1	0	1		1			
2	1	0	1	1	1	0	1	−1	1	0					
2	2	0	1	−1	1	0	1	−1	0	1					
3	1	−1	−1	1	1	0	0	0	1	0					
3	2	−1	−1	−1	1	0	0	0	0	1					1

$$y_{ijk} = \mu + \alpha_1 X_{ijk1} + \alpha_2 X_{ijk2} + \beta_1 x_{ijk3} + \delta_{11} X_{13} + \delta_{21} X_{23} + \varepsilon$$

4. The relationship between the does of a drug that increases blood pressure and the actual amount of increase in mean diastolic pressure was investigated in a laboratory experiment. **Twelve** rabbits received in random order **six** different does levels (0.1, 0.3, 0.5, 1.0, 1.5, 3.0) of the drug, with a suitable interval between each drug administration. The increase in blood pressure was used as the response variable. Let

$$\mathbf{Y}_i = (y_{i1},\ y_{i2},\ y_{i3},\ y_{i4},\ y_{i5},\ y_{i6})'$$

denote the vector of values of increase in blood pressure provided by the i-th rabbit.

(a) One possible model for these data is

$$y_{ij} = \mu + \boxed{\eta_i} + \alpha_j + \boxed{\epsilon_{ij}},\quad i = 1, 2, \cdots, 12,\quad j = 1, 2, \cdots, 6$$

5

where α_j is a fixed does effects, $\eta_i \sim NID(0, \sigma_\eta^2)$, $\epsilon_{ij} \sim NID(0, \sigma^2)$ and any η_i is independent of any ϵ_{ij}. What is the distribution of $\mathbf{Y}_i$, the vector of six responses from a single rabbit, for this model?

$Y_i \sim$ Multivariate Normal

$$E(Y_i) = \begin{pmatrix} \mu+\alpha_1 \\ \vdots \\ \mu+\alpha_6 \end{pmatrix} \qquad V(Y_i) = \begin{pmatrix} \sigma^2+\sigma_\eta^2 & & \\ & \ddots & \\ & & \end{pmatrix} \quad \sigma_\eta^2 = \Sigma$$

(b) The model in part (a) could be written in the form

$$\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{e},$$

4

where X is an appropriate model matrix, $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)^T$, and Σ is a block diagonal matrix with each 6×6 block corresponding to the covariance matrix you reported in part (a) for a set of six measurements taken on a single rabbit. Assuming that the variance components σ_η^2 and σ^2 are known for the model in part (a). What is the best linear unbiased estimator for an estimable function of $\boldsymbol{\beta}$, say $\mathbf{c}'\boldsymbol{\beta}$?

$$\hat{\beta}_{GLS} = (X^T \tilde{\Sigma}^{-1} X)^{-1} (X^T \tilde{\Sigma}^{-1} \underset{\sim}{Y}).$$

where

$\tilde{\Sigma}$ is a block-diagonal matrix with Block is Σ.

$$\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{Y}_1 \\ \vdots \\ \underset{\sim}{Y}_{12} \end{pmatrix}.$$

(c) Under the model in (a), construct a column of *source of variation* and *degree of freedom* of ANOVA table and specify the sources of error that should be used used to test the fixed does effects.

repeated.

8

	Source	df
	Trt	$6-1=5$
(rabbit)	Subject	$12-1=11$
(	Error	55
	Total	$12\times 6-1=\boxed{71}$

(d) ~~Using the sample mean of y_{i1} and y_{i6},~~ Provide an unbiased estimator of $\alpha_6-\alpha_1$, which is the mean difference in blood pressure between does level 0.1 and 3.0. Also give a formula for the confidence interval of $\alpha_6-\alpha_1$.

8 Specify the degree of freedom you used to define your confidence interval.

$$(\bar{y}_{\cdot 1}-\bar{y}_{\cdot 6}) \pm t_{(55), \alpha/2}\sqrt{MSE \cdot 1/6}$$

$$\begin{array}{r} 71 \\ -16 \\ \hline 55 \end{array}$$

$$Var(\bar{y}_{\cdot 1}-\bar{y}_{\cdot 6}) = Var(\cancel{\mu}+\cancel{\eta_{\cdot}}+\cancel{\alpha_1}+\bar{\varepsilon}_{\cdot 1} - \cancel{\mu}-\cancel{\eta_{\cdot}}-\cancel{\alpha_6}-\bar{\varepsilon}_{\cdot 6})$$

$$= \frac{1}{12}\sigma^2+\frac{1}{12}\sigma^2$$

$$= \frac{1}{6}\sigma^2.$$

Your Score: ______/100