

Stat 642 Summer 2015

Solutions for Homework 5

1. (10points) In the Cell Means model: $Y_{ijk} = \mu_{ij} + e_{ijk}$

a. H_o : No interaction $\Rightarrow H_o : \mu_{ij} - \mu_{kj} = \mu_{ih} - \mu_{kh}$ for all $(i, k, j, h) \Rightarrow$

For all $j = 1, \dots, b : \mu_{2j} - \mu_{1j} = C_1; \mu_{3j} - \mu_{2j} = C_2; \dots; \mu_{aj} - \mu_{a-1j} = C_{a-1} \Rightarrow$

Under H_o only $b + (a-1)$ parameters are needed to express μ_{ij} , namely,

$(\mu_{11}, \mu_{12}, \dots, \mu_{1b}, C_1, C_2, \dots, C_{a-1})$, that is, given these parameters we have

$$\mu_{ij} = \mu_{1j} + (\mu_{2j} - \mu_{1j}) + (\mu_{3j} - \mu_{2j}) + \dots + (\mu_{ij} - \mu_{i-1j}) = \mu_{1j} + C_1 + C_2 + \dots + C_{i-1}$$

Thus, $df_{A*B} = (\# \text{ parameters under } H_1) - (\# \text{ parameters under } H_o) = ab - (b + a - 1) = (a-1)(b-1)$

b. H_o : No main effect Factor A $\Rightarrow H_o : \bar{\mu}_{1.} = \bar{\mu}_{2.} = \dots = \bar{\mu}_{a.} = C \Rightarrow$

$$\bar{\mu}_{i.} = C \Rightarrow \frac{1}{b} \sum_{j=1}^b \mu_{ij} = C \Rightarrow \mu_{ib} = bC - \sum_{j=1}^{b-1} \mu_{ij} \Rightarrow$$

Under H_o only $1 + a(b-1)$ parameters are needed to express μ_{ij} , namely,

C plus, for each i , need $b-1$ parameters $\mu_{i1}, \mu_{i2}, \dots, \mu_{ib-1}$ to express μ_{ij} under H_o

Thus, $df_A = (\# \text{ parameters under } H_1) - (\# \text{ parameters under } H_o) = ab - (1 + a(b-1)) = a - 1$

2. (25 points) This is a CRD experiment involving three treatments with fixed levels, Soil Type, and 10 replications/treatment. The EU's are the fields and there is subsampling with the MU's being the locations within the fields.

a. $Y_{ijk} = \mu + \tau_i + e_{ij} + d_{ijk}$, $i = 1, \dots, 15$, $j = 1, 2$, $k = 1, \dots, n_{ij}$, where μ is overall mean, τ_i is the fixed effect due to the Soil Type, e_{ij} is the random effect due to the selected fields, d_{ijk} is the random effect due to variation in Soil subsamples within the same Soil Type-Field and all other sources.

$\tau_3 = 0$, $e_{ij} \sim iid N(0, \sigma_e^2)$, $d_{ijk} \sim iid N(0, \sigma_d^2)$, and e_{ij} and d_{ijk} are independent

b. $t = 3$, $n_i = r = 10$, $m_{ij} = 1$ or 2 :

Source	df	SS	MS	EMS	F	Pr > F
SoilType	2	36.67107	18.33554	$\sigma_d^2 + 1.6342\sigma_e^2 + 14.66\theta_\tau$	29.92	< .0001
Field(SoilType)	27	15.67045	0.58039	$\sigma_d^2 + 1.448\sigma_e^2$	1.769	0.1318
Location(Field)	14	4.59362	0.32812	σ_d^2		
Total	43	56.93514				

Variance for Fields= σ_e^2 , Variance for Locations= σ_d^2 :

$$A = \frac{25}{15} + \frac{22}{14} + \frac{25}{15} = 4.904762, \quad B = 16(1)^2 + 14(2)^2 = 72, \quad D = (15)^2 + (14)^2 + (15)^2 = 646 \Rightarrow$$

$$C_1 = \frac{1}{3-1} (4.904762 - \frac{72}{44}) = 1.6342, \quad C_2 = \frac{1}{3-1} (44 - \frac{646}{44}) = 14.66, \quad C_3 = \frac{1}{30-3} (44 - 4.904762) = 1.448$$

c. The variance components are computed as follows:

$$E(MS_{SUB}) = \sigma_d^2 \Rightarrow \hat{\sigma}_d^2 = MS_{SUB} = 0.32812$$

$$E(MSE) = \sigma_d^2 + c_3\sigma_e^2 \Rightarrow \hat{\sigma}_e^2 = \frac{MSE - MS_{SUB}}{c_3} = \frac{.58039 - .32812}{1.448} = .17422$$

Using the more accurate REML (Restricted Maximum Likelihood) estimators from the SAS function PROC MIXED, we obtain $\hat{\sigma}_d^2 = 0.3244$, and $\hat{\sigma}_e^2 = 0.1780$

Because of the unequal number of subsamples, REML produces estimates which are a little different from the AOV-MOM estimates.

From the model, the variance in the individual porosity readings, $\sigma_y^2 = \sigma_e^2 + \sigma_d^2$. Therefore,

Proportion of variation due to Fields is $\frac{\sigma_e^2}{\sigma_e^2 + \sigma_d^2} \approx \frac{.3244}{.3244 + .1780} = 64.6\%$

Proportion of variation due to Locations in the field is $\frac{\sigma_d^2}{\sigma_e^2 + \sigma_d^2} \approx \frac{.1780}{.3244 + .1780} = 35.4\%$

- d. Using the results from SAS, PROC MIXED, we have the following groupings, where Soil Types within a group are not significantly different:

$$G_1 = \{CLAY, LOAM\}, \quad G_2 = \{SANDY\}$$

Note that if you use PROC GLM to do the grouping, the Standard Errors of the estimates of the Treatment means $\widehat{SE}(\hat{\mu}_i)$ are smaller than the values from PROC MIXED:

PROC GLM - $\widehat{SE}(\hat{\mu}_i) = .15687, .16202, .15687$; which are incorrect estimates because they ignore the value of σ_e^2

PROC MIXED - $\widehat{SE}(\hat{\mu}_i) = .2014, .2054, .2014$; which are larger because they include $\hat{\sigma}_e^2$ in the estimation

This results in GLM having different groupings: $G_1 = \{CLAY\}$, $G_2 = \{LOAM\}$, $G_3 = \{SANDY\}$

- e. For testing $H_0 : \sigma_e^2 = 0$ vs $H_1 : \sigma_e^2 > 0$, test statistic is $F = \frac{MSE}{MSE_{SUB}} = \frac{.58039}{.32812} = 1.7688$. Because $F = 1.7688 < 2.326 = F_{0.05, 27, 14}$ and $p\text{-value} = 1 - pf(1.7688, 27, 14) = 0.1318 > 0.05$, we fail to reject H_0 and conclude that there is not significant evidence of a difference in porosity in fields of the same soil type.

3. (24 points) This is a 4×3 CRD with two crossed factors, Type of Crop and Nitrogen, having fixed levels. There are four reps/treatment with EU=MU=Growth Chamber.

- a. Cell Mean Model: $Y_{ijk} = \mu_{ij} + e_{ijk}$; $i = 1, 2, 3$; $j = 1, 2, 3, 4$; $k = 1, 2, 3, 4$, where Y_{ijk} is acetylene reduction from the k th growth chamber receiving i th Nitrogen level with the j th Crop, μ_{ij} is the mean response of i th Nitrogen level, j th Crop, and $e_{ijk} \sim iid N(0, \sigma_e^2)$.

Effects Model: $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$, where τ_i is the fixed effect of i th Nitrogen level, β_j is the fixed effect of j th Crop, $(\tau\beta)_{ij}$ are interaction effects between Nitrogen and Crop, with $\tau_3 = 0$, $\beta_4 = 0$, $(\tau\beta)_{3j} = (\tau\beta)_{i4} = 0$ for $i = 1, 2, 3$; $j = 1, 2, 3, 4$.

- b. Based on the B-F-L test which has p-value = 0.0389 and the plot of the residuals, there is an indication that the assumption of equality of variance is invalid.

The Shapiro-Wilk test has p-value < .0001, the Stem-Leaf plot appears heavy tailed with numerous outliers, and the normal probability plot has the residuals both above and below a straight line; therefore, the normal condition appears not to be valid.

There is not an index of time or space relative to the measurements or experimental units so a valid measure of correlation in the residuals is not available.

- c. Using the Box-Cox procedure, a transformation $X = \log(Y)$ was suggested. An evaluation of the AOV conditions using $\log(Y)$ yielded:

- Based on the B-F-L test which has p-value = 0.1940 and the plot of the residuals, the evidence indicates that the assumption of equality of variance is valid.

The Shapiro-Wilk test has p-value = .9289, the Stem-Leaf plot appears symmetric with no outliers, and the normal probability plot has the residuals very close to a straight line; therefore, the normal condition appears to be valid for the transformed data.

The ANOVA table from SAS is given below:

Dependent Variable: X = log(ACE-CONC)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	41.23719964	3.74883633	18.41	<.0001
Error	36	7.32907897	0.20358553		
Corrected Total	47	48.56627862			

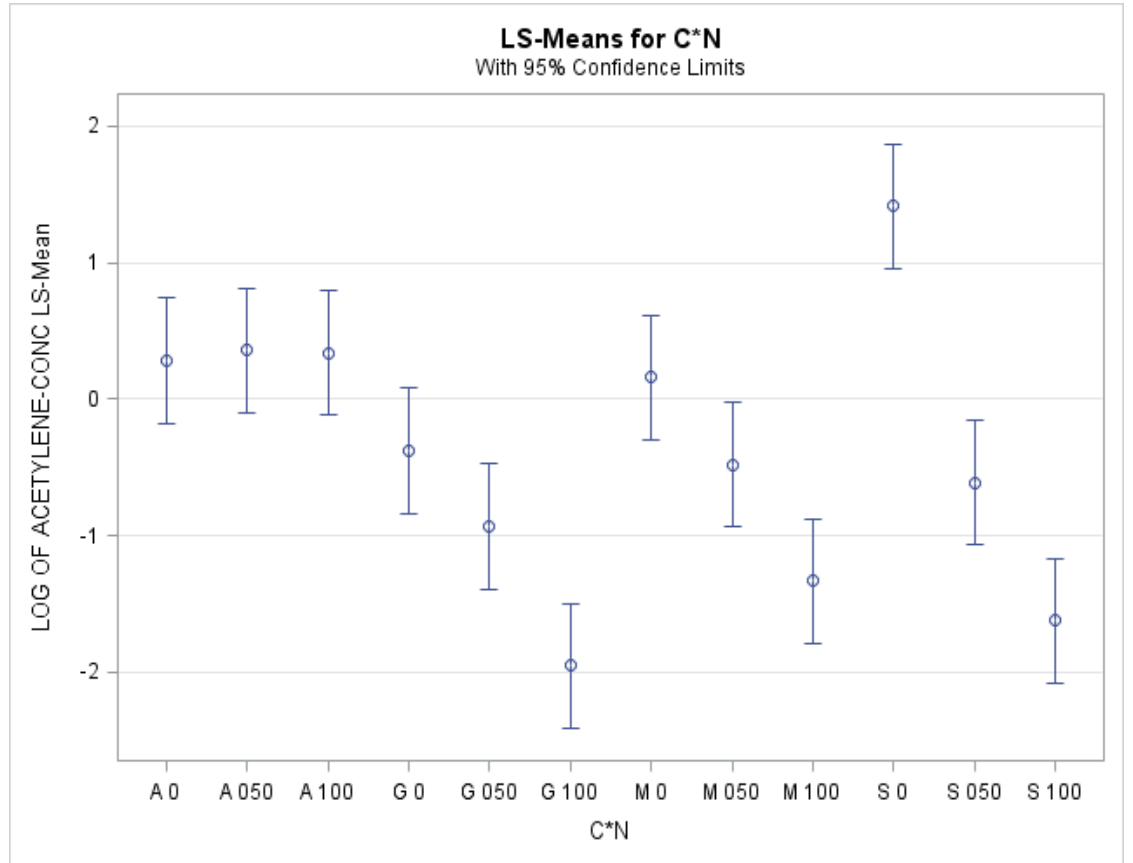
Source	DF	Type III SS	Mean Square	F Value	Pr > F
C	3	12.46379492	4.15459831	20.41	<.0001
N	2	18.34306101	9.17153051	45.05	<.0001
C*N	6	10.43034372	1.73839062	8.54	<.0001

- (d) Test for Interaction: $H_0 : \mu_{ij} - \mu_{ij'} = \mu_{i'j} - \mu_{i'j'}$ for all (i, i', j, j') vs $H_1 : \mu_{ij} - \mu_{ij'} \neq \mu_{i'j} - \mu_{i'j'}$ for some choice (i, i', j, j') :

Test statistic is $F = \frac{MS_{C \times N}}{MSE} = \frac{1.73839}{0.20358553} = 8.54$ with $F = 8.54 > 2.36 = F_{0.05, 6, 36}$ and

$p - value = 1 - pf(8.54, 6, 36) = 8.494528e - 06 < 0.05$, we reject H_0 . Thus, we conclude that there is significant evidence of an interaction between Nitrogen level and Crop. The ANOVA table shows both main effect have very small p-values and hence there is strong evidence of a main effect of both Nitrogen and Crop however, neither of these results have meaningful interpretation.

- (e) The profile plot is given here and the graph confirms the conclusion from the test of hypotheses. There is very little change in the mean level of log-acetylene with increasing levels of nitrogen for Alfalfa but there is a strong decrease in log-acetylene with increasing nitrogen for the other three crops.



- (f) Because of the significant interaction, the Nitrogen levels will be grouped separately for each of the 4 Types of Crops using the Tukey adjusted p-values:

Crop = Alfalfa: $G1 = \{0, 50, 100\}$

Crop = Soybean: $G1 = \{0\}$ $G2 = \{50, 100\}$

Crop = Guar: $G1 = \{0, 50\}$ $G2 = \{50, 100\}$

Crop = Mungbean: $G1 = \{0, 50\}$ $G2 = \{50, 100\}$

- The groupings are a bit surprising after viewing the profile plot. The 0 and 50 level of Nitrogen were not found to be different for Mungbean whereas in the profile plot there appears to be a sizable difference in the two means. This seeming contradiction reflects the simultaneous testing of all pairs which results in fewer pairs being declared different.

4. (20points)

- Cell Mean Model: $Y_{ijk} = \mu_{ij} + e_{ijk}$; $i = 1, 2, 3$; $j = 1, 2, 3, 4$; $k = 1, 2, 3, 4$, where Y_{ijk} is acetylene reduction from the kth receiving ith Nitrogen level with the jth Crop, μ_{ij} is the mean response of ith Nitrogen level, jth Crop, and $e_{ijk} \sim iid N(0, \sigma_e^2)$.
- Effects Model: $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$, where τ_i is the fixed effect of ith Nitrogen level, β_j is the fixed effect of jth Crop, $(\tau\beta)_{ij}$ are interaction effects between Nitrogen and Crop, with $\tau_3 = 0$, $\beta_4 = 0$, $(\tau\beta)_{3j} = (\tau\beta)_{i4} = 0$ for $i = 1, 2, 3$; $j = 1, 2, 3, 4$.
- The matrix form for the cell means model is

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

where

$$\begin{aligned} \mathbf{Y} &= (Y_{111}, Y_{112}, Y_{113}, Y_{114}, Y_{121}, Y_{122}, Y_{123}, Y_{124}, Y_{131}, Y_{132}, Y_{133}, Y_{134}, Y_{141}, Y_{142}, Y_{143}, Y_{144}, \\ &Y_{211}, Y_{212}, Y_{213}, Y_{214}, Y_{221}, Y_{222}, Y_{223}, Y_{224}, Y_{231}, Y_{232}, Y_{233}, Y_{234}, Y_{241}, Y_{242}, Y_{243}, Y_{244}, \\ &Y_{311}, Y_{312}, Y_{313}, Y_{314}, Y_{321}, Y_{322}, Y_{323}, Y_{324}, Y_{331}, Y_{332}, Y_{333}, Y_{334}, Y_{341}, Y_{342}, Y_{343}, Y_{344},)^T, \\ \mathbf{e} &= (e_{111}, e_{112}, e_{113}, e_{114}, e_{121}, e_{122}, e_{123}, e_{124}, e_{131}, e_{132}, e_{133}, e_{134}, e_{141}, e_{142}, e_{143}, e_{144}, \\ &e_{211}, e_{212}, e_{213}, e_{214}, e_{221}, e_{222}, e_{223}, e_{224}, e_{231}, e_{232}, e_{233}, e_{234}, e_{241}, e_{242}, e_{243}, e_{244}, \\ &e_{311}, e_{312}, e_{313}, e_{314}, e_{321}, e_{322}, e_{323}, e_{324}, e_{331}, e_{332}, e_{333}, e_{334}, e_{341}, e_{342}, e_{343}, e_{344},)^T, \end{aligned}$$

$$X = \begin{pmatrix} 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 \\ 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 \end{pmatrix}_{48 \times 12} \quad \boldsymbol{\beta} = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{14} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{24} \\ \mu_{31} \\ \mu_{32} \\ \mu_{33} \\ \mu_{34} \end{pmatrix}_{12 \times 1}$$

d. The matrix form for the effects model with constraints is

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

where \mathbf{Y} and \mathbf{e} are the same as those in c,

$$X = \begin{pmatrix} 1_4 & 1_4 & 0_4 & 1_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 1_4 & 1_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 1_4 & 1_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 \\ 1_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 1_4 & 0_4 & 1_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 \\ 1_4 & 0_4 & 1_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 \\ 1_4 & 0_4 & 1_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 \\ 1_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 1_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 1_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \\ 1_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 & 0_4 \end{pmatrix}_{48 \times 12}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ (\tau\gamma)_{11} \\ (\tau\gamma)_{12} \\ (\tau\gamma)_{13} \\ (\tau\gamma)_{21} \\ (\tau\gamma)_{22} \\ (\tau\gamma)_{23} \end{pmatrix}_{12 \times 1}$$

- 5.(21 points) 1. Prove that there is an $F_1 * F_2 * F_3$ interaction, that is, show that the $F_2 * F_3$ interaction effect at $F_1 = 1$ is different from the $F_2 * F_3$ interaction effect at $F_1 = 2$.

Verify that $[(\mu_{111} - \mu_{112}) - (\mu_{121} - \mu_{122})] \neq [(\mu_{211} - \mu_{212}) - (\mu_{221} - \mu_{222})]$

$$[(\mu_{111} - \mu_{112}) - (\mu_{121} - \mu_{122})] = [(8 - 4) - (2 - 4)] = 6 \text{ and}$$

$$[(\mu_{211} - \mu_{212}) - (\mu_{221} - \mu_{222})] = [(4 - 6) - (6 - 2)] = -6 \neq 6 \Rightarrow \text{There is an } F_1 * F_2 * F_3 \text{ interaction.}$$

2. Prove that there is not an $F_2 * F_3$ interaction (ignore the fact that there is an $F_1 * F_2 * F_3$ interaction)..

Verify that $(\bar{\mu}_{.11} - \bar{\mu}_{.12}) = (\bar{\mu}_{.21} - \bar{\mu}_{.22})$

$$(\bar{\mu}_{.11} - \bar{\mu}_{.12}) = ((8 + 4)/2 - (4 + 6)/2) = 1 \text{ and } (\bar{\mu}_{.21} - \bar{\mu}_{.22}) = ((2 + 6)/2 - (4 + 2)/2) = 1 \Rightarrow \text{There is not an } F_2 * F_3 \text{ interaction.}$$

3. Prove that there is a main effect due to F_3 (ignore the fact that there is an $F_1 * F_2 * F_3$ interaction).

Verify that $\bar{\mu}_{..1} \neq \bar{\mu}_{..2}$

$$\bar{\mu}_{..1} = (8 + 2 + 4 + 6)/4 = 5 \text{ and } \bar{\mu}_{..2} = (4 + 4 + 6 + 2)/4 = 4 \neq 5 \Rightarrow \text{There is a } F_3 \text{ main effect.}$$

-

-