

STAT 421 Homework 11

Lecture 27 Problem 1a:

In this problem, we will consider the data from problem 13.1. In problem 13.1, an experiment was performed to investigate the capability of a measurement system. Ten parts were randomly selected, and two randomly selected operators measured each part three times. The tests were made in random order. The data for this experiment are shown below.

■ **TABLE P13.1**
Measurement Systems Data for Problem 13.1

Part No.	Operator 1 Measurements			Operator 2 Measurements		
	1	2	3	1	2	3
1	50	49	50	50	48	51
2	52	52	51	51	51	51
3	53	50	50	54	52	51
4	49	51	50	48	50	51
5	48	49	48	48	49	48
6	52	50	50	52	50	50
7	51	51	51	51	50	50
8	52	50	49	53	48	50
9	50	51	50	51	48	49
10	47	46	49	46	47	48

The goal of this problem is to write code to run a “fixed effects” full model, and identify the numerical values of MS_A , MS_B , MS_{AB} , and MS_E . Note that in the code below, A represents the part factor, B represents the operator factor, and R represents the replication factor. This is done in R below.

```
# Clear environment:
rm(list=ls(all=TRUE))

# Setup important value for the problem:
n <- 3
a <- 10
b <- 2

# Load in the data as a matrix:
y_mat = matrix(c(50, 49, 50, 50, 48, 51, 52, 52, 51, 51, 51, 51, 53, 50, 50, 54,
                 52, 51, 49, 51, 50, 48, 50, 51, 48, 49, 48, 48, 49, 48, 52, 50,
                 50, 52, 50, 50, 51, 51, 51, 51, 50, 50, 52, 50, 49, 53, 48, 50,
                 50, 51, 50, 51, 48, 49, 47, 46, 49, 46, 47, 48), ncol = 6, byrow=T)

# Make the data into a vector (in row order):
```

```

y <- as.vector(t(y_mat))

# Setup the factors for this experiment:
design = gen.factorial(c(3,2,10), 3, varNames=c("R","B","A"), factors="all")
attach(design)

# Confirm that the data in R is the same as the problem:
cbind(R, B, A, y)

```

```

##      R B  A  y
## [1,] 1 1  1 50
## [2,] 2 1  1 49
## [3,] 3 1  1 50
## [4,] 1 2  1 50
## [5,] 2 2  1 48
## [6,] 3 2  1 51
## [7,] 1 1  2 52
## [8,] 2 1  2 52
## [9,] 3 1  2 51
## [10,] 1 2  2 51
## [11,] 2 2  2 51
## [12,] 3 2  2 51
## [13,] 1 1  3 53
## [14,] 2 1  3 50
## [15,] 3 1  3 50
## [16,] 1 2  3 54
## [17,] 2 2  3 52
## [18,] 3 2  3 51
## [19,] 1 1  4 49
## [20,] 2 1  4 51
## [21,] 3 1  4 50
## [22,] 1 2  4 48
## [23,] 2 2  4 50
## [24,] 3 2  4 51
## [25,] 1 1  5 48
## [26,] 2 1  5 49
## [27,] 3 1  5 48
## [28,] 1 2  5 48
## [29,] 2 2  5 49
## [30,] 3 2  5 48
## [31,] 1 1  6 52
## [32,] 2 1  6 50
## [33,] 3 1  6 50
## [34,] 1 2  6 52
## [35,] 2 2  6 50
## [36,] 3 2  6 50
## [37,] 1 1  7 51
## [38,] 2 1  7 51
## [39,] 3 1  7 51
## [40,] 1 2  7 51
## [41,] 2 2  7 50
## [42,] 3 2  7 50
## [43,] 1 1  8 52

```

```
## [44,] 2 1 8 50
## [45,] 3 1 8 49
## [46,] 1 2 8 53
## [47,] 2 2 8 48
## [48,] 3 2 8 50
## [49,] 1 1 9 50
## [50,] 2 1 9 51
## [51,] 3 1 9 50
## [52,] 1 2 9 51
## [53,] 2 2 9 48
## [54,] 3 2 9 49
## [55,] 1 1 10 47
## [56,] 2 1 10 46
## [57,] 3 1 10 49
## [58,] 1 2 10 46
## [59,] 2 2 10 47
## [60,] 3 2 10 48
```

```
# Create the full fixed effects model:
model_1 <- lm(y~A*B)
```

```
# Save the mean square values:
MS_A <- summary.aov(model_1)[[1]][1,3]
MS_B <- summary.aov(model_1)[[1]][2,3]
MS_AB <- summary.aov(model_1)[[1]][3,3]
MS_E <- summary.aov(model_1)[[1]][4,3]
```

```
# Run ANOVA in R to get the results:
summary.aov(model_1)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## A           9  99.02   11.002    7.335 3.22e-06 ***
## B           1   0.42    0.417    0.278   0.601
## A:B         9   5.42    0.602    0.401   0.927
## Residuals   40  60.00    1.500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As seen above, the value of MS_A is 11.0018519, the value of MS_B is 0.4166667, the value of MS_{AB} is 0.6018519, and the value of MS_E is 1.5.

Lecture 27 Problem 1b:

In this problem, using the MS values from part a, we will do random effects modeling. In particular, we will compute the appropriate F-values and p-values for testing $H_0 : \sigma_A^2 = 0$, $H_1 : \sigma_A^2 > 0$, $H_0 : \sigma_B^2 = 0$, $H_1 : \sigma_B^2 > 0$, and $H_0 : \sigma_{AB}^2 = 0$, $H_1 : \sigma_{AB}^2 > 0$. This is done in R below.

```
# Compute the degrees of freedom for each MS term:
df_A <- a-1
df_B <- b-1
df_AB <- (a-1)*(b-1)
df_E <- a*b*(n-1)

# Compute the F-values for the corresponding F tests:
F_A <- MS_A / MS_AB
```

```

F_B <- MS_B / MS_AB
F_AB <- MS_AB / MS_E

# Compute the p-values for the corresponding F tests:
p_val_a <- pf(F_A, df_A, df_AB, lower.tail=F)
p_val_b <- pf(F_B, df_B, df_AB, lower.tail=F)
p_val_ab <- pf(F_AB, df_AB, df_E, lower.tail=F)

```

As shown above, the F-value and p-value for testing $H_0 : \sigma_A^2 = 0$, $H_1 : \sigma_A^2 > 0$ was computed as 18.28 and 9.3810632×10^{-5} , the F-value and p-value for testing $H_0 : \sigma_B^2 = 0$, $H_1 : \sigma_B^2 > 0$ was computed as 0.6923077 and 0.4269057, and the F-value and p-value for testing $H_0 : \sigma_{AB}^2 = 0$, $H_1 : \sigma_{AB}^2 > 0$ was computed as 0.4012346 and 0.9270089.

Lecture 27 Problem 1c:

In this problem, for each of the above tests, we will use $\alpha = 0.05$ and state our conclusions in English.

For testing the competing hypotheses $H_0 : \sigma_A^2 = 0$, $H_1 : \sigma_A^2 > 0$, we obtained a p-value of 9.3810632×10^{-5} . Therefore, at $\alpha = 0.05$, we reject the null hypothesis and thus we have significant evidence that $\sigma_A^2 > 0$. This means that there is a difference in the capability of the measurement system across all parts in the population. In other words, this means that there is evidence of variability across all parts in the population.

For testing the competing hypotheses $H_0 : \sigma_B^2 = 0$, $H_1 : \sigma_B^2 > 0$, we obtained a p-value of 0.4269057. Therefore, at $\alpha = 0.05$, we fail to reject the null hypothesis and thus we do not have significant evidence that $\sigma_B^2 > 0$. This means that we have no evidence that there is a difference in the capability of the measurement system across all operators in the population. In other words, this means that there is no evidence of variability across all operators in the population.

For testing the competing hypotheses $H_0 : \sigma_{AB}^2 = 0$, $H_1 : \sigma_{AB}^2 > 0$, we obtained a p-value of 0.9270089. Therefore, at $\alpha = 0.05$, we fail to reject the null hypothesis and thus we do not have significant evidence that $\sigma_{AB}^2 > 0$. This means that we have no evidence that the effect of all parts on the capability of the measurement system varies across all operators. In other words, there is no evidence of an interaction effect between all of parts and all of the operators.

Lecture 27 Problem 1d:

In this problem, using the formulas from the lecture, we will compute the estimates of the four variance components, and interpret them. This is done below.

```

# Compute the four variance components:
var_A <- (MS_A-MS_AB) / (n*b)
var_B <- (MS_B-MS_AB) / (n*a)
var_AB <- (MS_AB-MS_E) / n
var_E <- MS_E

```

As calculated above, the estimated variance component for the A term is 1.7333333, this component estimates the portion of total variability accounted for by the part factor. Also, the estimated variance component for the B term is -0.0061728, this component estimates the portion of total variability accounted for by the operator factor. Next, the estimated variance component for the AB interaction term is -0.2993827, this component estimates the portion of total variability accounted for by the part-operator interaction. Lastly, the estimated variance component for the error term is 1.5, this component estimates the portion of total variability accounted for by random error. In general, a variance component is an estimate of the contribution that different experimental factors make to the overall variability of the data.

Lecture 27 Problem 1e:

In this problem, we will check if the sum of the four variance component estimates is approximately equal to the sample variance of all of the data, like it is supposed to be. This is done in R below.

```
# Sample variance of the data:
var_y <- var(y)

# Add up all of the variance components:
comp_sum <- var_A + var_B + var_AB + var_E
```

As calculated above, the sum of the estimated variance components is 2.9277778, this is pretty close to the sample variance which was calculated to be 2.7940678.

Lecture 27 Problem 2a:

In this problem, we will use the defining formulas for $\hat{\sigma}_\epsilon^2, \hat{\sigma}_\alpha^2, \dots$ to write

$$\begin{bmatrix} \hat{\sigma}_\epsilon^2 \\ \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\beta^2 \\ \hat{\sigma}_{\alpha\beta}^2 \end{bmatrix} = \begin{bmatrix} \text{matrix of} \\ \text{constants} \\ \text{a, b, n} \end{bmatrix} \begin{bmatrix} MS_E \\ MS_A \\ MS_B \\ MS_{AB} \end{bmatrix}$$

The defining formulas for $\hat{\sigma}_\epsilon^2, \hat{\sigma}_\alpha^2, \dots$ are written below

$$\begin{aligned} \hat{\sigma}_\epsilon^2 &= MS_E \\ \hat{\sigma}_\alpha^2 &= \frac{1}{bn} [MS_A - MS_{AB}] \\ \hat{\sigma}_\beta^2 &= \frac{1}{an} [MS_B - MS_{AB}] \\ \hat{\sigma}_{\alpha\beta}^2 &= \frac{1}{n} [MS_{AB} - MS_E] \end{aligned}$$

We will now use these formulas to form the matrix equation above. This matrix equation with the coefficients matrix filled is shown below

$$\begin{bmatrix} \hat{\sigma}_\epsilon^2 \\ \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\beta^2 \\ \hat{\sigma}_{\alpha\beta}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{bn} & 0 & \frac{-1}{bn} \\ 0 & 0 & \frac{1}{an} & \frac{-1}{an} \\ \frac{-1}{n} & 0 & 0 & \frac{1}{n} \end{bmatrix} \begin{bmatrix} MS_E \\ MS_A \\ MS_B \\ MS_{AB} \end{bmatrix}$$

Lecture 27 Problem 2b:

In this problem, we will use the table of expected values to write

$$\begin{bmatrix} E[MS_E] \\ E[MS_A] \\ E[MS_B] \\ E[MS_{AB}] \end{bmatrix} = \begin{bmatrix} \text{matrix of} \\ \text{constants} \\ \text{a, b, n} \end{bmatrix} \begin{bmatrix} \sigma_\epsilon^2 \\ \sigma_\alpha^2 \\ \sigma_\beta^2 \\ \sigma_{\alpha\beta}^2 \end{bmatrix}$$

The table of expected values is written below

$$\begin{aligned} E[MS_E] &= \sigma_\epsilon^2 \\ E[MS_A] &= \sigma_\epsilon^2 + bn\sigma_\alpha^2 + n\sigma_{\alpha\beta}^2 \\ E[MS_B] &= \sigma_\epsilon^2 + an\sigma_\beta^2 + n\sigma_{\alpha\beta}^2 \\ E[MS_{AB}] &= \sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 \end{aligned}$$

We will now use these formulas to form the matrix equation above. This matrix equation with the coefficients matrix filled is shown below

$$\begin{bmatrix} E[MS_E] \\ E[MS_A] \\ E[MS_B] \\ E[MS_{AB}] \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & bn & 0 & n \\ 1 & 0 & an & n \\ 1 & 0 & 0 & n \end{bmatrix} \begin{bmatrix} \sigma_\epsilon^2 \\ \sigma_\alpha^2 \\ \sigma_\beta^2 \\ \sigma_{\alpha\beta}^2 \end{bmatrix}$$

Lecture 27 Problem 2c:

In this problem, we will put it all together and show that

$$E\left[\begin{bmatrix} \hat{\sigma}_\epsilon^2 \\ \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\beta^2 \\ \hat{\sigma}_{\alpha\beta}^2 \end{bmatrix}\right] = \begin{bmatrix} \sigma_\epsilon^2 \\ \sigma_\alpha^2 \\ \sigma_\beta^2 \\ \sigma_{\alpha\beta}^2 \end{bmatrix}$$

Using the derivation from part a, we can rewrite the left-hand-side of the above matrix equation as

$$E\left[\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{bn} & 0 & \frac{-1}{bn} \\ 0 & 0 & \frac{1}{an} & \frac{-1}{an} \\ \frac{-1}{n} & 0 & 0 & \frac{1}{n} \end{bmatrix} \begin{bmatrix} MS_E \\ MS_A \\ MS_B \\ MS_{AB} \end{bmatrix}\right]$$

Furthermore, if use the fact that for any constant c , $E[cX] = cE[X]$, then we can take the coefficient matrix outside of the expectation operator and simplify the above expression to the following

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{bn} & 0 & \frac{-1}{bn} \\ 0 & 0 & \frac{1}{an} & \frac{-1}{an} \\ \frac{-1}{n} & 0 & 0 & \frac{1}{n} \end{bmatrix} E\left[\begin{bmatrix} MS_E \\ MS_A \\ MS_B \\ MS_{AB} \end{bmatrix}\right]$$

If we now bring the expectation operator inside of the MS vector then we can rewrite the above expression as

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{bn} & 0 & \frac{-1}{bn} \\ 0 & 0 & \frac{1}{an} & \frac{-1}{an} \\ \frac{-1}{n} & 0 & 0 & \frac{1}{n} \end{bmatrix} \begin{bmatrix} E[MS_E] \\ E[MS_A] \\ E[MS_B] \\ E[MS_{AB}] \end{bmatrix}$$

Using our derivation from part b, we can simplify the vector of expectations to the expression below

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{bn} & 0 & \frac{-1}{bn} \\ 0 & 0 & \frac{1}{an} & \frac{-1}{an} \\ \frac{-1}{n} & 0 & 0 & \frac{1}{n} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & bn & 0 & n \\ 1 & 0 & an & n \\ 1 & 0 & 0 & n \end{bmatrix} \begin{bmatrix} \sigma_\epsilon^2 \\ \sigma_\alpha^2 \\ \sigma_\beta^2 \\ \sigma_{\alpha\beta}^2 \end{bmatrix}$$

Doing the matrix multiplication of the two coefficient matrices, we obtain the following expression shown below

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_\epsilon^2 \\ \sigma_\alpha^2 \\ \sigma_\beta^2 \\ \sigma_{\alpha\beta}^2 \end{bmatrix}$$

Since the resulting matrix from the above matrix multiplication is the identity matrix, it follows that our final result is the following vector

$$\begin{bmatrix} \sigma_\epsilon^2 \\ \sigma_\alpha^2 \\ \sigma_\beta^2 \\ \sigma_{\alpha\beta}^2 \end{bmatrix}$$

Hence we have shown that

$$E\left[\begin{bmatrix} \hat{\sigma}_\epsilon^2 \\ \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\beta^2 \\ \hat{\sigma}_{\alpha\beta}^2 \end{bmatrix}\right] = \begin{bmatrix} \sigma_\epsilon^2 \\ \sigma_\alpha^2 \\ \sigma_\beta^2 \\ \sigma_{\alpha\beta}^2 \end{bmatrix}$$

Lecture 27 Problem 3:

In this problem, we will consider the two-factor random-effects model $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$ with $k = 1, \dots, n$ replications. In particular, the goal of this problem is to show that $V[\bar{y}_{i..}] = \frac{1}{bn}E[MS_A] + \frac{1}{abn}E[MS_B] - \frac{1}{abn}E[MS_{AB}]$. To make the derivation easier we will use results already derived in class, namely that $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \implies \bar{y}_{i..} = \mu + \alpha_i + \bar{\beta} + (\bar{\alpha}\bar{\beta})_{i.} + \bar{\epsilon}_{i..}$. Furthermore, if we assume that $\alpha_i \perp \beta_j \perp (\alpha\beta)_{ij} \perp \epsilon_{ijk}$, then there will be no covariance term in our calculation, as it will be zero. Lastly, since the model itself is built on the fact that α_i , β_j , $(\alpha\beta)_{ij}$, and ϵ_{ijk} are all iid random variables, it follows that we can say $V[\alpha_i] = \sigma_\alpha^2$, $V[\beta_j] = \sigma_\beta^2$, $V[(\alpha\beta)_{ij}] = \sigma_{\alpha\beta}^2$, and $V[\epsilon_{ijk}] = \sigma_\epsilon^2$.

The proof also relies on writing the variance components in terms of expected values, the formulas for doing so are shown below.

$$\begin{aligned} \sigma_\alpha^2 &= \frac{1}{bn}(E[MS_A] - E[MS_{AB}]) \\ \sigma_\beta^2 &= \frac{1}{an}(E[MS_B] - E[MS_{AB}]) \\ \sigma_{\alpha\beta}^2 &= \frac{1}{n}(E[MS_{AB}] - E[MS_E]) \\ \sigma_\epsilon^2 &= MS_E \end{aligned}$$

Now that we have went through all of the assumptions and all of the formulas needed for the proof, we can now get to the proof itself. This is shown below.

$$\begin{aligned} V[\bar{y}_{i..}] &= V\left[\mu + \alpha_i + \bar{\beta} + (\bar{\alpha}\bar{\beta})_{i.} + \bar{\epsilon}_{i..}\right] \\ &= V[\mu] + V[\alpha_i] + V[\bar{\beta}] + V[(\bar{\alpha}\bar{\beta})_{i.}] + V[\bar{\epsilon}_{i..}] \\ &= 0 + \sigma_\alpha^2 + V\left[\frac{1}{b} \sum_j \beta_j\right] + V\left[\frac{1}{b} \sum_j (\alpha\beta)_{ij}\right] + V\left[\frac{1}{bn} \sum_j \sum_k \epsilon_{ijk}\right] \\ &= \sigma_\alpha^2 + \frac{1}{b}\sigma_\beta^2 + \frac{1}{b}\sigma_{\alpha\beta}^2 + \frac{1}{bn}\sigma_\epsilon^2 \\ &= \frac{1}{bn}(E[MS_A] - E[MS_{AB}]) + \frac{1}{abn}(E[MS_B] - E[MS_{AB}]) + \frac{1}{bn}(E[MS_{AB}] - E[MS_E]) + \frac{1}{bn}MS_E \\ &= \frac{1}{bn}E[MS_A] - \frac{1}{bn}E[MS_{AB}] + \frac{1}{abn}E[MS_B] - \frac{1}{abn}E[MS_{AB}] + \frac{1}{bn}E[MS_{AB}] - \frac{1}{bn}E[MS_E] + \frac{1}{bn}MS_E \\ &= \frac{1}{bn}E[MS_A] + \frac{1}{abn}E[MS_B] - \frac{1}{abn}E[MS_{AB}] \end{aligned}$$

Hence, as shown above, we have shown that $V[\bar{y}_{i..}] = \frac{1}{bn}E[MS_A] + \frac{1}{abn}E[MS_B] - \frac{1}{abn}E[MS_{AB}]$.

Lecture 28 Problem 1

In this problem, we will find the 95% confidence interval for σ_α^2 in the two-factor random-effects model of example 13.1. As shown in lecture, to compute the approximate confidence interval for σ_α^2 in the two-factor case, we will need to use the following formula

$$\frac{r\hat{\sigma}_\alpha^2}{\chi_{\frac{\alpha}{2}, r}^2} < \sigma_\alpha^2 < \frac{r\hat{\sigma}_\alpha^2}{\chi_{1-\frac{\alpha}{2}, r}^2}$$

where $\chi_{\alpha/2, df}^2$ represents the appropriate critical value under the chi-squared distribution with r degrees of freedom. Also, $\hat{\sigma}_\alpha^2$ and r are defined as follows.

$$\hat{\sigma}_\alpha^2 = \frac{1}{bn}(MS_A - MS_{AB})$$

$$r = \frac{(MS_A - MS_{AB})^2}{\frac{(MS_A)^2}{a-1} + \frac{(MS_{AB})^2}{(a-1)(b-1)}}$$

As computed in lecture 27: $MS_A = 62.391$, $MS_{AB} = 0.712$, $a-1 = 19$, $(a-1)(b-1) = 38$, and $\hat{\sigma}_\alpha^2 = 10.2798$. Putting this all together, we obtain the following value for r

$$r = \frac{(62.391 - 0.712)^2}{\frac{(62.391)^2}{19} + \frac{(0.712)^2}{38}} = 18.568 \approx 19$$

Hence, as computed above, $r = 19$. We will use this to compute our chi-squared critical values, this is done in R below.

```
# Compute the ci-squared critical values for the confidence interval:
upper_chi <- qchisq(1-0.05/2, 19, lower.tail=F)
lower_chi <- qchisq(0.05/2, 19, lower.tail=F)
```

Hence, as computed above, our chi-squared critical values are $\chi_{\alpha/2, r}^2 = 32.8523269$ and $\chi_{1-\alpha/2, r}^2 = 8.9065165$. Putting it altogether, our 95% confidence interval for σ_α^2 is

$$\frac{19 \cdot 10.2798}{32.8523269} < \sigma_\alpha^2 < \frac{19 \cdot 10.2798}{8.9065165}$$

Simplifying this interval to three decimal places we obtain the final 95% confidence interval for σ_α^2 shown below

$$5.94527750179 < \sigma_\alpha^2 < 21.9295838053$$

Lecture 28 Problem 2

In this problem, for the three-factor random-effects model, we will find an approximate confidence interval for σ_α^2 in terms of the various MS values and their corresponding degrees of freedom. The derivation of this interval relies on the following linear combinations of the expectations of mean squares.

$$E[MS_A] + E[MS_{ABC}] = 2\sigma_\epsilon^2 + cn\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha\gamma}^2 + 2n\sigma_{\alpha\beta\gamma}^2 + bcn\sigma_\alpha^2$$

$$E[MS_{AB}] + E[MS_{AC}] = 2\sigma_\epsilon^2 + cn\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha\gamma}^2 + 2n\sigma_{\alpha\beta\gamma}^2$$

The above linear combinations of the expectations of mean squares leads us to the following estimate of σ_α^2 , namely: $\hat{\sigma}_\alpha^2 = [(MS_A + MS_{ABC}) - (MS_{AB} + MS_{AC})]$. Furthermore, in 1946, Satterthwaite showed that $X^2 = r\hat{\sigma}_\alpha^2/\sigma_\alpha^2$ is approximately distributed χ_r^2 . Where X^2 is defined as

$$X^2 = r \frac{(MS_1 + MS_2 + \dots) - (MS'_1 + MS'_2 + \dots)}{E[(MS_1 + MS_2 + \dots) - (MS'_1 + MS'_2 + \dots)]}$$

And r is defined as

$$r = \frac{[(MS_1 + MS_2 + \dots) - (MS'_1 + MS'_2 + \dots)]^2}{\frac{MS_1^2}{df_1} + \frac{MS_2^2}{df_2} + \dots + \frac{MS'^2_1}{df'_1} + \frac{MS'^2_2}{df'_2} + \dots}$$

Thus in our case, the approximate confidence interval for σ_α^2 in terms of the various MS values and their corresponding degrees of freedom can be written as follows

$$\frac{r\hat{\sigma}_\alpha^2}{\chi^2_{\frac{\alpha}{2}, r}} < \sigma_\alpha^2 < \frac{r\hat{\sigma}_\alpha^2}{\chi^2_{1-\frac{\alpha}{2}, r}}$$

Where $\hat{\sigma}_\alpha^2 = [(MS_A + MS_{ABC}) - (MS_{AB} + MS_{AC})]$ and r is written below as

$$r = \frac{[(MS_A + MS_{ABC}) - (MS_{AB} + MS_{AC})]^2}{\frac{MS_A^2}{a-1} + \frac{MS_{ABC}^2}{(a-1)(b-1)(c-1)} + \frac{MS_{AB}^2}{(a-1)(b-1)} + \frac{MS_{AC}^2}{(a-1)(c-1)}}$$

Lecture 28 Optional Problem a:

In this problem, we will consider the data from problem 5.8. In problem 5.8, the factors that influence the breaking strength of a synthetic fiber are being studied. Four production machines and three operators are chosen and a factorial experiment is run using fiber from the same production batch. The data for this experiment are shown below.

Operator	Machine			
	1	2	3	4
1	109	110	108	110
	110	115	109	108
2	110	110	111	114
	112	111	109	112
3	116	112	114	120
	114	115	119	117

The goal of this problem is to treat both factors as fixed and report/interpret the p-values for testing the significance of each effect and their interaction. Note that in the code below, A represents the machine factor, B represents the operator factor, and R represents the replication factor. This is done in R below.

```
# Clear environment:
rm(list=ls(all=TRUE))

# Setup important value for the problem:
n <- 2
a <- 4
b <- 3

# Load in the data as a matrix:
y_mat = matrix(c(109, 110, 108, 110, 110, 115, 109, 108, 110, 110, 111, 114, 112,
                 111, 109, 112, 116, 112, 114, 120, 114, 115, 119, 117),
               ncol = 4, byrow=T)

# Make the data into a vector (in row order):
y <- as.vector(t(y_mat))
```

```
# Setup the factors for this experiment:
design = gen.factorial(c(4, 2, 3), 3, varNames=c("A","R","B"), factors="all")
attach(design)
```

```
# Confirm that the data in R is the same as the problem:
cbind(A, R, B, y)
```

```
##      A R B   y
## [1,] 1 1 1 109
## [2,] 2 1 1 110
## [3,] 3 1 1 108
## [4,] 4 1 1 110
## [5,] 1 2 1 110
## [6,] 2 2 1 115
## [7,] 3 2 1 109
## [8,] 4 2 1 108
## [9,] 1 1 2 110
## [10,] 2 1 2 110
## [11,] 3 1 2 111
## [12,] 4 1 2 114
## [13,] 1 2 2 112
## [14,] 2 2 2 111
## [15,] 3 2 2 109
## [16,] 4 2 2 112
## [17,] 1 1 3 116
## [18,] 2 1 3 112
## [19,] 3 1 3 114
## [20,] 4 1 3 120
## [21,] 1 2 3 114
## [22,] 2 2 3 115
## [23,] 3 2 3 119
## [24,] 4 2 3 117
```

```
# Create the full fixed effects model:
model_2 <- lm(y~A*B)
```

```
# Run ANOVA in R to get the results:
summary.aov(model_2)
```

```
##      Df Sum Sq Mean Sq F value    Pr(>F)
## A      3  12.46    4.15    1.095 0.388753
## B      2 160.33   80.17   21.143 0.000117 ***
## A:B     6  44.67    7.44    1.963 0.150681
## Residuals 12  45.50    3.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As seen above, the *A* effect (machine effect) had a p-value of 0.388753, hence at the 5% level of significance, we fail to reject the null hypothesis and conclude that there is no significant evidence that the mean breaking strength of a synthetic fiber varies across the four machines.

Similarly, the *B* effect (operator effect) had a p-value of 0.000117, hence at the 5% level of significance, we reject the null hypothesis and conclude that there is significant evidence that the mean breaking strength of a synthetic fiber varies across the three operators.

Lastly, the AB interaction effect (machine-operator interaction effect) had a p-value of 0.150681, hence at the 5% level of significance, we fail to reject the null hypothesis and conclude that there is no significant evidence that the effect of the four machines on the mean breaking strength of a synthetic fiber varies across the three operators. In other words, there is no evidence of an interaction between the four machines and the three operators.

Lecture 28 Optional Problem b:

In this problem, we will treat both factors as random and report/interpret the p-values for testing the significance of each effect and their interaction. This is done in R below.

```
# Save the mean square values:
MS_A <- summary.aov(model_2)[[1]][1,3]
MS_B <- summary.aov(model_2)[[1]][2,3]
MS_AB <- summary.aov(model_2)[[1]][3,3]
MS_E <- summary.aov(model_2)[[1]][4,3]

# Compute the degrees of freedom for each MS term:
df_A <- a-1
df_B <- b-1
df_AB <- (a-1)*(b-1)
df_E <- a*b*(n-1)

# Compute the F-values for the corresponding F tests:
F_A <- MS_A / MS_AB
F_B <- MS_B / MS_AB
F_AB <- MS_AB / MS_E

# Compute the p-values for the corresponding F tests:
p_val_a <- pf(F_A, df_A, df_AB, lower.tail=F)
p_val_b <- pf(F_B, df_B, df_AB, lower.tail=F)
p_val_ab <- pf(F_AB, df_AB, df_E, lower.tail=F)
```

For testing the competing hypotheses $H_0 : \sigma_A^2 = 0$, $H_1 : \sigma_A^2 > 0$, we obtained a p-value of 0.6619463. Therefore, at $\alpha = 0.05$, we fail to reject the null hypothesis and thus we do not have significant evidence that $\sigma_A^2 > 0$. This means that we have no significant evidence that the mean breaking strength of a synthetic fiber varies across all machines in the population.

For testing the competing hypotheses $H_0 : \sigma_B^2 = 0$, $H_1 : \sigma_B^2 > 0$, we obtained a p-value of 0.010344. Therefore, at $\alpha = 0.05$, we reject the null hypothesis and thus we have significant evidence that $\sigma_B^2 > 0$. This means that we have significant evidence that the mean breaking strength of a synthetic fiber varies across all operators in the population.

For testing the competing hypotheses $H_0 : \sigma_{AB}^2 = 0$, $H_1 : \sigma_{AB}^2 > 0$, we obtained a p-value of 0.1506807. Therefore, at $\alpha = 0.05$, we fail to reject the null hypothesis and thus we do not have significant evidence that $\sigma_{AB}^2 > 0$. This means that we have no evidence that the effect of all of the machines on the mean breaking strength of a synthetic fiber varies across all operators. In other words, there is no evidence of an interaction between all of machines and all of operators.

Lecture 28 Optional Problem c:

In this problem, we will treat operator as random, but machine as fixed and report/interpret the p-values for testing the significance of each effect and their interaction. This is done in R below.

```
# Compute the F-values for the corresponding F tests:
F_A <- MS_A / MS_AB
F_B <- MS_B / MS_E
F_AB <- MS_AB / MS_E
```

```
# Compute the p-values for the corresponding F tests:
p_val_a <- pf(F_A, df_A, df_AB, lower.tail=F)
p_val_b <- pf(F_B, df_B, df_E, lower.tail=F)
p_val_ab <- pf(F_AB, df_AB, df_E, lower.tail=F)
```

For testing the A effect (machine effect) we computed a p-value of 0.6619463, hence at the 5% level of significance, we fail to reject the null hypothesis and conclude that there is no significant evidence that the mean breaking strength of a synthetic fiber varies across the four machines.

For testing the competing hypotheses $H_0 : \sigma_B^2 = 0$, $H_1 : \sigma_B^2 > 0$, we obtained a p-value of 0.00011667. Therefore, at $\alpha = 0.05$, we reject the null hypothesis and thus we have significant evidence that $\sigma_B^2 > 0$. This means that we have significant evidence that the mean breaking strength of a synthetic fiber varies across all operators in the population.

For testing the competing hypotheses $H_0 : \sigma_{AB}^2 = 0$, $H_1 : \sigma_{AB}^2 > 0$, we obtained a p-value of 0.1506807. Therefore, at $\alpha = 0.05$, we fail to reject the null hypothesis and thus we do not have significant evidence that $\sigma_{AB}^2 > 0$. This means that we have no evidence that the effect of the four machines on the mean breaking strength of a synthetic fiber varies across all operators. In other words, there is no evidence of an interaction between the four machines and all of the operators.