

Question 1

A process engineer is testing the yield of a product manufactured on randomly selected three machines. Each machine can be operated at a large number of power settings and only two power settings are randomly selected for this study. Furthermore, a machine has only three stations on which the product is formed. An experiment is conducted in which each machine is tested at both power settings, and three observations on yield are taken from each station. The runs are made in random order. Part of the ANOVA table computed for the above experiment incorrectly treating all main effects as fixed and all combinations of factors as crossed is given below.

| Source | SS |
|---|---------|
| Machine | 21.436 |
| Power | 845.698 |
| Station | 16.980 |
| Machine \times Power | 0.383 |
| Machine \times Station | 16.603 |
| Power \times Station | 16.303 |
| Machine \times Power \times Station | 12.905 |
| Error | 61.760 |
| Total | 992.068 |

The model of the nested & crossed Design is

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_{k(i)} + (\tau\beta)_{ij} + (\beta\gamma)_{jk(i)} + \varepsilon_{l(ijk)}$$

for $i = 1, 2, 3$ is the number of Machine being compared; $j = 1, 2$ is the number of Power being compared; $k = 1, 2, 3$ is the number of stations; $l = 1, 2, 3$ is the number of replication;

μ is the overall true mean response;

τ_i is the random effect of i^{th} level of Machine;

β_j is the random effect of j^{th} level of Power;

$\gamma_{k(i)}$ is the main fixed effect of k^{th} stations nested in i^{th} level of Machine;

$(\tau\beta)_{ij}$ is the interaction effect of i^{th} level of Machine and j^{th} level of Power;

$(\beta\gamma)_{jk(i)}$ is the interaction effect of j^{th} level of Power and k^{th} stations nested in i^{th} level of Machine.

y_{ijkl} is response value for the l^{th} replication for k^{th} stations nested in i^{th} level of Machine when j^{th} level of Power is applied;

$\varepsilon_{l(ijk)}$ is random error for the l^{th} replication for k^{th} stations nested in i^{th} level of Machine when j^{th} level of Power is applied.

Assumptions:

| | | |
|---|---|--|
| $\varepsilon_{l(ijk)} \sim iidN(0, \sigma^2)$ | $\tau_i \sim iidN(0, \sigma_\tau^2)$ | $\beta_j \sim iidN(0, \sigma_\beta^2)$ |
| $\sum_{k=1}^3 \gamma_{k(i)} = 0$ | $(\tau\beta)_{ij} \sim iidN(0, \sigma_{\tau\beta}^2)$ | $(\beta\gamma)_{jk(i)} \sim iidN(0, \frac{3-1}{3} \sigma_{\beta\gamma}^2)$ |

$\varepsilon_{l(ijk)}, \tau_i, \beta_j, (\tau\beta)_{ij}$, and $(\beta\gamma)_{jk(i)}$ are independent. For restricted model, $\sum_{k=1}^3 (\beta\gamma)_{jk(i)} = 0$

Produce the corrected ANOVA table with Source, df, SS, MS, EMS, F value along with numerator and denominator df, and p value for each test.

| source | i(r) | j(r) | k(f) | l(r) | df | SS | EMS | F |
|-------------------------|------|------|------|------|-------------|----------------------|--|--------------------------------|
| A | 1 | b | c | n | a-1 | SS_A | $\sigma^2 + cn\sigma_{\tau\beta}^2 + bc n\sigma_{\tau_i}^2$ | $\frac{MS_A}{MS_{AB}}$ |
| τ_i | | 2 | 3 | 3 | 3-1 | 21.436 | $\sigma^2 + 9\sigma_{\tau\beta}^2 + 18\sigma_{\tau_i}^2$ | $df_{2,2}$ |
| (r) | | | | | | | | |
| B | a | 1 | c | n | b-1 | SS_B | $\sigma^2 + cn\sigma_{\beta\gamma}^2 + ac n\sigma_{\tau\beta}^2$ | $\frac{MS_B}{MS_{AB}}$ |
| β_j | 3 | | 3 | 3 | 2-1 | 845.698 | $\sigma^2 + 9\sigma_{\beta\gamma}^2 + 27\sigma_{\tau\beta}^2$ | $df_{1,2}$ |
| (r) | | | | | | | | |
| C(A) | 1 | b | 0 | n | a(c-1) | $SS_C + SS_{AC}$ | $\sigma^2 + n\sigma_{\beta\gamma}^2 + \frac{bn \sum_{k=1}^c \gamma_{k(i)}^2}{c-1}$ | $\frac{MS_{C(A)}}{MS_{BC(A)}}$ |
| $\gamma_{k(i)}$ | | 2 | | 3 | 3(3-1) | 16.980+16.603 | $\sigma^2 + 3\sigma_{\beta\gamma}^2 + 3 \sum_{k=1}^3 \gamma_{k(i)}^2$ | $df_{6,6}$ |
| (f) | | | | | | | | |
| AB | 1 | 1 | c | n | (a-1)(b-1) | SS_{AB} | $\sigma^2 + cn\sigma_{\tau\beta}^2$ | $\frac{MS_{AB}}{MS_E}$ |
| $(\tau\beta)_{ij}$ | | | 3 | 3 | (3-1)(2-1) | 0.383 | $\sigma^2 + 9\sigma_{\tau\beta}^2$ | $df_{2,36}$ |
| (r) | | | | | | | | |
| BC(A) | 1 | 1 | 0 | n | a(b-1)(c-1) | $SS_{BC} + SS_{ABC}$ | $\sigma^2 + n\sigma_{\beta\gamma}^2$ | $\frac{MS_{BC(A)}}{MS_E}$ |
| $(\beta\gamma)_{jk(i)}$ | | | | 3 | 3(2-1)(3-1) | 16.303+12.905 | $\sigma^2 + 3\sigma_{\beta\gamma}^2$ | $df_{6,36}$ |
| (r) | | | | | | | | |
| Error | 1 | 1 | 1 | 1 | abc(n-1) | SS_E | σ^2 | |
| (r) | | | | | 3×2×3(3-1) | 61.760 | | |
| Total | | | | | abcn-1 | SS_T | | |
| | | | | | 3×2×3×3-1 | 992.068 | | |

Compute the p value assuming you don't have access to a computer. If an exact F test is not available, construct a Satterthwaite (approximate) F statistic and its df values. Provide the numerical values for df, SS, MS, F in the table. The numerical values:

| source | df | SS | MS | F | p |
|--------|----|---------|--------------------|---------------------------------------|--------------|
| A | 2 | 21.436 | 21.436/2=10.718 | $\frac{10.718}{0.1915} = 55.96867$ | 0.01755351 |
| B | 1 | 845.698 | 845.698 | $\frac{845.698}{0.1915} = 4416.178$ | 0.0002263633 |
| C(A) | 6 | 33.583 | 33.583/6=5.597167 | $\frac{5.597167}{4.868} = 1.149788$ | 0.4348901 |
| AB | 2 | 0.383 | 0.383/2=0.1915 | $\frac{0.1915}{1.715556} = 0.1116256$ | 0.8946874 |
| BC(A) | 6 | 29.208 | 29.208/6=4.868 | $\frac{4.868}{1.715556} = 2.837564$ | 0.02292305 |
| Error | 36 | 61.760 | 61.760/36=1.715556 | | |
| Total | 53 | 992.068 | | | |

Question 2

Consider the linear model for a two-stage nested design with B nested in A as given below. Using only the given information here,

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}; \text{ for } i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, n;$$

Assumptions: $\varepsilon_{(ij)k} \sim iidN(0, \sigma^2)$; $\tau_i \sim iidN(0, \sigma_\tau^2)$; $\sum_{j=1}^b \beta_{j(i)} = 0$; $\varepsilon_{(ij)k}$, τ_i are independent.

$$\bar{y}_{i..} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n y_{ijk}; \bar{y}_{.j.} = \frac{1}{an} \sum_{i=1}^a \sum_{k=1}^n y_{ijk}; \bar{y}_{ij.} = \frac{1}{n} \sum_{k=1}^n y_{ijk}; \bar{y}_{...} = \frac{1}{abn} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$$

(a). Derive the least squares estimator of $\beta_{j(i)}$. Provide the appropriate constraints about estimators used when solving normal equations. (Use only sum to zero constraints)

This is a two-stage nested design with fixed factor B nested in random factor A. $\varepsilon_{(ij)k} = y_{ijk} - \mu - \tau_i - \beta_{j(i)}$

$$SSE = \sum_i^a \sum_j^b \sum_k^n (y_{ijk} - \mu - \tau_i - \beta_{j(i)})^2$$

Derive For $i = 1, \dots, a, j = 1, \dots, b,$

$$\frac{\partial SSE}{\partial \tau_i} = 2 \sum_j^b \sum_k^n (y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_{j(i)}) (-1) = 0$$

$$\text{For } \sum_j^b \hat{\beta}_{j(i)} = 0, \bar{y}_{i..} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n y_{ijk},$$

$$\sum_j^b \sum_k^n y_{ijk} = bn\hat{\mu} + bn\hat{\tau}_i + n \sum_j^b \hat{\beta}_{j(i)} = bn\bar{y}_{i..} \implies \hat{\mu} + \hat{\tau}_i = \bar{y}_{i..}$$

$$\frac{\partial SSE}{\partial \beta_{j(i)}} = 2 \sum_k^n (y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_{j(i)}) (-1) = 0$$

$$\text{For } \bar{y}_{ij.} = \frac{1}{n} \sum_{k=1}^n y_{ijk},$$

$$\sum_k^n y_{ijk} = n\hat{\mu} + n\hat{\tau}_i + n\hat{\beta}_{j(i)} = n\bar{y}_{ij.} \implies \hat{\mu} + \hat{\tau}_i + \hat{\beta}_{j(i)} = \bar{y}_{ij.}$$

$$\implies \hat{\beta}_{j(i)} = \bar{y}_{ij.} - \bar{y}_{i..}$$

(b). Derive $E(MS_{B(A)})$ for this model.

$$\text{For } \sum_j^b \hat{\beta}_{j(i)} = 0,$$

$$\bar{y}_{i..} = \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n y_{ijk} = \mu + \tau_i + \frac{1}{b} \sum_{j=1}^b \beta_{j(i)} + \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{ijk} = \mu + \tau_i + \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{ijk}$$

$$\bar{y}_{ij.} = \frac{1}{n} \sum_{k=1}^n y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \frac{1}{n} \sum_{k=1}^n \varepsilon_{ijk}$$

$$\bar{y}_{ij.} - \bar{y}_{i..} = \beta_{j(i)} + \frac{1}{n} \sum_{k=1}^n \varepsilon_{ijk} - \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{ijk}$$

For B has fixed effect, $E[\beta_{j(i)}] = \beta_{j(i)}, V[\beta_{j(i)}] = 0, Cov[\beta_{j(i)}, \frac{1}{n} \sum_{k=1}^n \varepsilon_{(ij)k} - \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{ijk}] = 0$

For $\varepsilon_{(ij)k} \sim iidN(0, \sigma^2), E[\varepsilon_{(ij)k}] = 0, V[\varepsilon_{(ij)k}] = \sigma^2$

$$E[\bar{y}_{ij.} - \bar{y}_{i..}] = E[\beta_{j(i)} + \frac{1}{n} \sum_{k=1}^n \varepsilon_{(ij)k} - \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{(ij)k}] = \beta_{j(i)}$$

$$\begin{aligned}
\text{Var}[\bar{y}_{ij.} - \bar{y}_{i..}] &= \text{Var}[\beta_{j(i)} + \frac{1}{n} \sum_{k=1}^n \varepsilon_{(ij)k} - \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{(ij)k}] \\
&= \text{Var}[\beta_{j(i)}] + \text{Var}[\frac{1}{n} \sum_{k=1}^n \varepsilon_{ijk} - \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{ijk}] + 2\text{Cov}[\beta_{j(i)}, \frac{1}{n} \sum_{k=1}^n \varepsilon_{ijk} - \frac{1}{bn} \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{ijk}] \\
&= 0 + \frac{1}{n^2} \sum_{k=1}^n \text{Var}[\varepsilon_{ijk}] + \frac{1}{b^2 n^2} \sum_{j=1}^b \sum_{k=1}^n \text{Var}[\varepsilon_{ijk}] - \frac{2}{bn^2} \text{Cov}[\sum_{k=1}^n \varepsilon_{ijk}, \sum_{j=1}^b \sum_{k=1}^n \varepsilon_{ijk}] + 0 \\
&= \frac{1}{n} \sigma^2 + \frac{\sigma^2}{bn} - \frac{2\sigma^2}{bn} = \frac{(b-1)\sigma^2}{bn}
\end{aligned}$$

$$\text{For } MS_{B(A)} = SS_{B(A)} / df_{B(A)} = \frac{n}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b \hat{\beta}_{j(i)}^2 = \frac{n}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..})^2$$

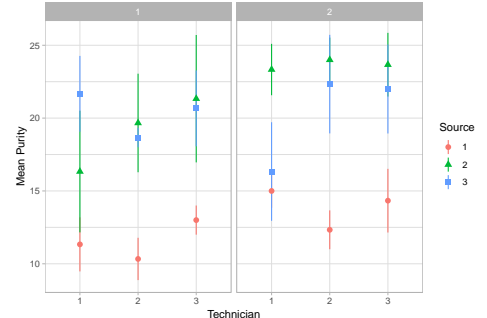
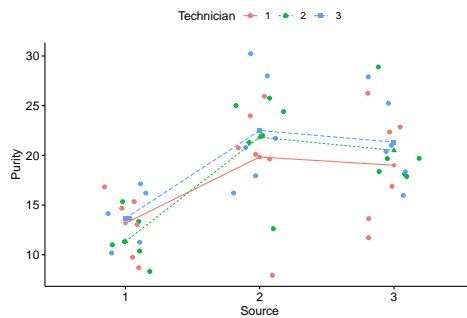
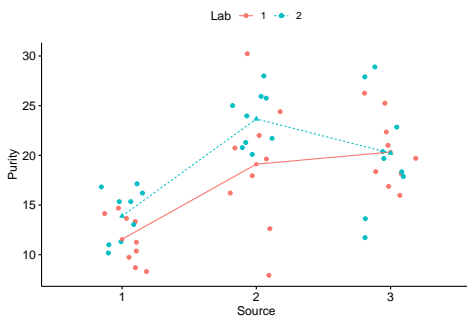
$$\begin{aligned}
E(MS_{B(A)}) &= \frac{n}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b E[(\bar{y}_{ij.} - \bar{y}_{i..})^2] = \frac{n}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b \{ \text{Var}[\bar{y}_{ij.} - \bar{y}_{i..}] + E[(\bar{y}_{ij.} - \bar{y}_{i..})^2] \} \\
&= \frac{n}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b \left\{ \beta_{j(i)}^2 + \frac{(b-1)\sigma^2}{bn} \right\} = \sigma^2 + \frac{n}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b \beta_{j(i)}^2
\end{aligned}$$

Question 3

Two laboratories were used to determine the purity of a chemical compound synthesized from 3 sources. Within each of these laboratories, 3 technicians were used to carry out the analysis.

Analyze the data. If you have to decide between unrestricted and restricted models, then make a decision and provide reasons.

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    54 obs. of  4 variables:
## $ Lab      : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ Technician: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 2 ...
## $ Source    : Factor w/ 3 levels "1","2","3": 1 1 1 2 2 2 3 3 3 1 ...
## $ Purity    : num  10 9 15 8 20 21 26 22 17 8 ...
```



The above plots show that: The average purity analysed in Lab 1 is lower than that from Lab 2. There is not much difference in the average purity among technicians. The average purity of a chemical compound synthesized from source 2 and 3 are higher than that from source 1. There is not much difference in the average purity between sources 2 and 3.

| Lab | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|-----|-----|----|--------|------|-----|----------|----------|----|---------|
| 1 | 8 | 13 | 17 | 21.0 | 30 | 17.00000 | 5.824352 | 27 | 0 |
| 2 | 10 | 15 | 20 | 23.5 | 29 | 19.25926 | 5.633639 | 27 | 0 |

| Technician | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|------------|-----|-------|--------|-------|-----|----------|----------|----|---------|
| 1 | 8 | 13.25 | 17 | 21.75 | 26 | 17.33333 | 5.718906 | 18 | 0 |
| 2 | 8 | 13.00 | 18 | 21.75 | 29 | 17.88889 | 6.057459 | 18 | 0 |
| 3 | 10 | 16.00 | 18 | 21.75 | 30 | 19.16667 | 5.762454 | 18 | 0 |

| Source | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|--------|-----|-------|--------|-------|-----|----------|----------|----|---------|
| 1 | 8 | 10.25 | 13.0 | 15.00 | 17 | 12.72222 | 2.803476 | 18 | 0 |
| 2 | 8 | 20.00 | 21.5 | 24.75 | 30 | 21.38889 | 5.326282 | 18 | 0 |
| 3 | 12 | 18.00 | 20.0 | 22.75 | 29 | 20.27778 | 4.599304 | 18 | 0 |

There is not much difference in the average purity of sources 3 between lab 1 and 2.

Not all the lines are parallel in the interaction plot. Therefore, in the model, there is the interaction effect of source level and technicians nested in the lab. The Tables show the same thing with the numerical summaries for each factor level and their combinations.

| Lab.Source | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|------------|-----|----|--------|----|-----|----------|----------|---|---------|
| 1.1 | 8 | 10 | 11 | 14 | 15 | 11.55556 | 2.505549 | 9 | 0 |
| 2.1 | 10 | 11 | 15 | 16 | 17 | 13.88889 | 2.713137 | 9 | 0 |
| 1.2 | 8 | 16 | 20 | 22 | 30 | 19.11111 | 6.392270 | 9 | 0 |
| 2.2 | 20 | 21 | 24 | 26 | 28 | 23.66667 | 2.783882 | 9 | 0 |
| 1.3 | 16 | 18 | 20 | 22 | 26 | 20.33333 | 3.500000 | 9 | 0 |
| 2.3 | 12 | 18 | 20 | 23 | 29 | 20.22222 | 5.717906 | 9 | 0 |

| Technician.Source | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|-------------------|-----|-------|--------|-------|-----|----------|----------|---|---------|
| 1.1 | 9 | 10.75 | 14.0 | 15.00 | 17 | 13.16667 | 3.125167 | 6 | 0 |
| 2.1 | 8 | 10.25 | 11.0 | 12.50 | 15 | 11.33333 | 2.422120 | 6 | 0 |
| 3.1 | 10 | 11.75 | 14.0 | 15.50 | 17 | 13.66667 | 2.732520 | 6 | 0 |
| 1.2 | 8 | 20.00 | 20.5 | 23.25 | 26 | 19.83333 | 6.274286 | 6 | 0 |
| 2.2 | 13 | 21.25 | 23.0 | 24.75 | 26 | 21.83333 | 4.708149 | 6 | 0 |
| 3.2 | 16 | 18.75 | 21.5 | 26.50 | 30 | 22.50000 | 5.504544 | 6 | 0 |
| 1.3 | 12 | 14.75 | 19.5 | 22.75 | 26 | 19.00000 | 5.513619 | 6 | 0 |
| 2.3 | 18 | 18.00 | 19.0 | 20.00 | 29 | 20.50000 | 4.277850 | 6 | 0 |
| 3.3 | 16 | 18.50 | 20.5 | 24.00 | 28 | 21.33333 | 4.457204 | 6 | 0 |

This is a neseted and crossed design. Three fixed sources apply on all random selected technicians nested in fixed labs.

$$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{j(i)k} + \varepsilon_{(ijk)l}$$

for $i = 1, 2; j = 1, 2, 3; k = 1, 2, 3; l = 1, 2, 3;$

μ is the overall true mean response;

τ_i is the fixed main effect of i^{th} level of labs;

$\beta_{j(i)}$ is the random effect of j^{th} level of technicians nested in i^{th} level of labs;

γ_k is the main fixed effect of k^{th} level of sources;

$(\tau\gamma)_{ik}$ is the interaction effect of i^{th} level of labs and k^{th} level of sources;

$(\beta\gamma)_{j(i)k}$ is the interaction random effect of k^{th} level of sources and j^{th} level of technicians nested in i^{th} level of labs.

y_{ijkl} is response value for the l^{th} replication for j^{th} level of technicians nested in i^{th} level of labs when k^{th} level of sources is applied;

$\varepsilon_{(ijk)l}$ is random error for the l^{th} replication for j^{th} level of technicians nested in i^{th} level of labs when k^{th} level of sources is applied.

Assumptions: Usually, the technicians in a lab are skillful. From the above plots, the technicians' performances are stable. The covariance between two observations from the same level of the random factor can be either positive or negative. Thus we assume this is an **restricted model**. $\varepsilon_{(ijk)l}$, $\beta_{j(i)}$, and $(\beta\gamma)_{j(i)k}$ are independent.

| | | | |
|---|--------------------------------------|--|--|
| $\varepsilon_{(ijk)l} \sim iidN(0, \sigma^2)$ | $\sum_{i=1}^2 \tau_i = 0$ | $\sum_{k=1}^3 \gamma_k = 0$ | $\beta_{j(i)} \sim iidN(0, \sigma_\beta^2)$ |
| $\sum_{i=1}^2 (\tau\gamma)_{ik} = 0$ | $\sum_{k=1}^3 (\tau\gamma)_{ik} = 0$ | $\sum_{i=1}^2 (\beta\gamma)_{j(i)k} = 0$ | $(\beta\gamma)_{j(i)k} \sim iidN(0, \frac{2-1}{2} \sigma_{\beta\gamma}^2)$ |

Table 6: Analysis of Variance Table

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|------------------------------------|----|--------|---------|---------|-----------|
| Lab_f | 1 | 68.91 | 68.91 | 6.955 | 0.05775 |
| Source_f | 2 | 800.6 | 400.3 | 30.62 | 0.0001783 |
| Lab_f:Source_f | 2 | 49.04 | 24.52 | 1.875 | 0.2148 |
| Lab_f:Technician_r | 4 | 39.63 | 9.907 | 0.5 | 0.7358 |
| Lab_f:Source_f:Technician_r | 8 | 104.6 | 13.07 | 0.6598 | 0.7226 |
| Residual | 36 | 713.3 | 19.81 | NA | NA |

The ANOVA table shows that only sources have significant effects on the average purity of a chemical compound synthesized at 0.05 significance level (p-value=0.0001783).

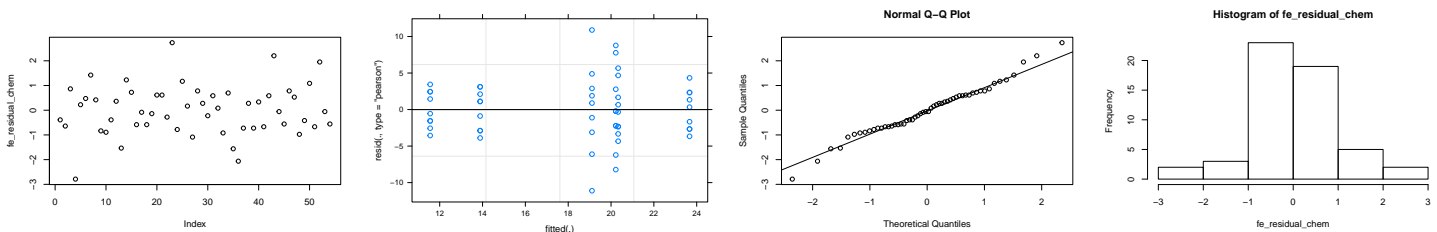
| Groups Name | | Variance | Std.Dev. | |
|-----------------------|-------------|----------|----------|------------------------------|
| Source:Lab:Technician | (Intercept) | 0.00 | 0.000 | $\sigma_{\beta\gamma}^2 = 0$ |
| Lab:Technician | (Intercept) | 0.00 | 0.000 | $\sigma_\beta^2 = 0$ |
| Residual | | 17.87 | 4.227 | $\sigma^2 = 17.87$ |

| | 2.5 % | 97.5 % | |
|---------------|-------|--------|--|
| .sig01 | 0 | 1.539 | $CI_{\sigma_{\beta\gamma}^2} : (0, 1.539^2)$ |
| .sig02 | 0 | 1.603 | $CI_{\sigma_\beta^2} : (0, 1.603^2)$ |
| .sigma | 3.337 | 4.873 | $CI_{\sigma^2} : (3.337^2, 4.873^2)$ |

The results of variance components and condidence intervals show that none of the effects related with technician has significant variance on average value of purity at 0.05 significance level. The variance of interaction effect between sources and technicians nested in labs is zero with confidence intervals $(0, 1.539^2)$ at 0.05 significance level. The variance of technicians nested in labs is zero with confidence

intervals $(0, 1.603^2)$ at 0.05 significance level.

Conclusion: As the main effects of sources shown in the above tables, the average purity is different with sources. The average purity from source 1 is lowest (12.72222). The the average purity from source 2 and 3 are 21.38889 and 20.27778 respectively. The selections of labs and technicians don't change this result.



In the plots of residuals versus predicted value of purity, there is no significant pattern on this plot besides that more values occur when the predicted value is higher. Therefore, the fitted model is good enough to describe the relationship between the mean value of purity and the labs, technician, and sources.

The residuals in this plot are almost symmetrically distributed about zero and hence zero mean assumption is not violated. Further, the vertical deviation of the residuals from zero is about same for each predicted value and hence the constant variance assumption is not violated.

The points are along the straight line in the normal qq plot shown at bottom left and the histogram of residuals shown at the top right is about normal. These plots show no violation of normal distribution assumption of residuals.

Report your code and/or output at the end of the analysis.

```
favstats(Purity ~ Source, data = fe_table_chem)
favstats(Purity ~ Lab, data = fe_table_chem)
favstats(Purity ~ Lab + Source, data = fe_table_chem)
favstats(Purity ~ Technician, data = fe_table_chem)
favstats(Purity ~ Technician + Source, data = fe_table_chem)
fe_table_chem$Lab_f <- as.fixed(fe_table_chem$Lab)
fe_table_chem$Technician_r <- as.random(fe_table_chem$Technician)
fe_table_chem$Source_f <- as.fixed(fe_table_chem$Source)
fe_model_chem1 <- aov(formula = Purity ~ Lab_f * Source_f + Technician_r %in%
  Lab_f + Source_f * Technician_r %in% Lab_f, data = fe_table_chem)
gad(fe_model_chem1)

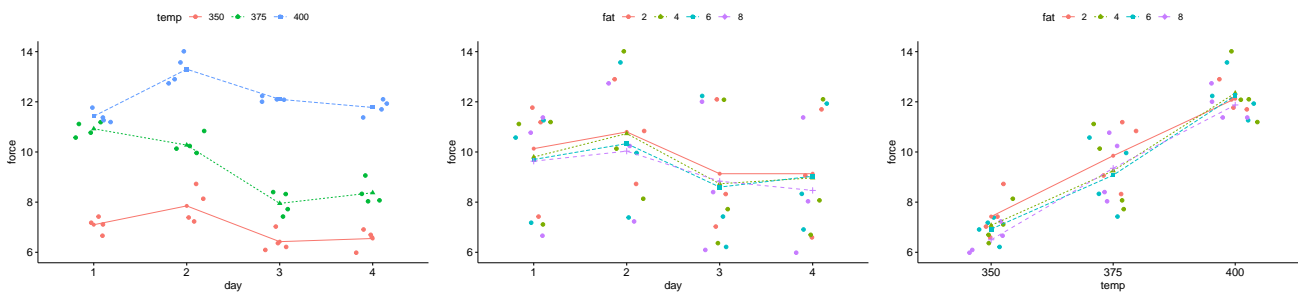
fe_model_chem2 <- lmer(formula = Purity ~ Lab * Source + (1 | Lab:Technician) +
  (1 | Source:Lab:Technician), data = fe_table_chem, REML = TRUE)
summary(fe_model_chem2)
confint(fe_model_chem2)

fe_residual_chem <- rstudent(fe_model_chem2)
plot(fe_residual_chem)
plot(fe_model_chem2)
qqnorm(fe_residual_chem)
qqline(fe_residual_chem)
hist(fe_residual_chem)
```

Question 4

A baker wanted to determine the effect that the amount of fat in a recipe of cookie dough would have on the texture of the cookie. The baker also wanted to determine whether the temperature (°F) at which the cookies were baked would have an influence on the texture of the surface. The texture of the cookie is measured by determining the amount of force (g) required to penetrate the cookie surface. On a given day, the baker made a batch of cookie dough for each of the 4 recipes and baked one cookie from each batch in the oven at one time. The baker continued this process each of 4 days so that a single cookie is baked at 3 different temperatures in one day. The data are given in Cookie Excel file. Analyze the data and answer the baker's questions. If you have to decide between unrestricted and restricted models, then make a decision and provide reasons.

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   48 obs. of  4 variables:
## $ day : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...
## $ temp : Factor w/ 3 levels "350","375","400": 1 1 1 1 2 2 2 2 3 3 ...
## $ fat : Factor w/ 4 levels "2","4","6","8": 1 2 3 4 1 2 3 4 1 2 ...
## $ force: num  7.4 7.1 7.2 6.7 11.2 11.1 10.6 10.8 11.8 11.2 ...
```



The above plots show that: There is not much difference in the average force of the cookie from different days or from different fat. The tables show the same thing with the numerical summaries for each factor level and their combinations. The average force of the cookie are higher when the temperature is higher.

| day | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|-----|-----|------|--------|--------|------|-----------|----------|----|---------|
| 1 | 6.7 | 7.35 | 10.95 | 11.225 | 11.8 | 9.816667 | 2.033284 | 12 | 0 |
| 2 | 7.2 | 8.55 | 10.15 | 12.750 | 14.0 | 10.475000 | 2.384467 | 12 | 0 |
| 3 | 6.1 | 6.85 | 8.00 | 12.025 | 12.2 | 8.825000 | 2.526271 | 12 | 0 |
| 4 | 6.0 | 6.85 | 8.20 | 11.475 | 12.1 | 8.900000 | 2.290693 | 12 | 0 |

| fat | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|-----|-----|-------|--------|--------|------|----------|----------|----|---------|
| 2 | 6.6 | 8.075 | 9.95 | 11.725 | 12.9 | 9.800000 | 2.200413 | 12 | 0 |
| 4 | 6.4 | 7.550 | 9.10 | 11.425 | 14.0 | 9.558333 | 2.520988 | 12 | 0 |
| 6 | 6.2 | 7.350 | 9.15 | 11.450 | 13.6 | 9.416667 | 2.480408 | 12 | 0 |
| 8 | 6.0 | 7.075 | 9.30 | 11.400 | 12.7 | 9.241667 | 2.439992 | 12 | 0 |

| temp | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|------|------|--------|--------|--------|------|----------|-----------|----|---------|
| 350 | 6.0 | 6.550 | 6.95 | 7.250 | 8.7 | 6.98125 | 0.7148135 | 16 | 0 |
| 375 | 7.4 | 8.250 | 9.55 | 10.650 | 11.2 | 9.38125 | 1.3422463 | 16 | 0 |
| 400 | 11.2 | 11.625 | 12.05 | 12.325 | 14.0 | 12.15000 | 0.7983316 | 16 | 0 |

Not all the lines are parallel in the interaction plot. Therefore, in the model, there is the interaction effect of day and temperature.

| day.fat | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|---------|-----|------|--------|-------|------|-----------|----------|---|---------|
| 1.2 | 7.4 | 9.30 | 11.2 | 11.50 | 11.8 | 10.133333 | 2.386071 | 3 | 0 |
| 2.2 | 8.7 | 9.75 | 10.8 | 11.85 | 12.9 | 10.800000 | 2.100000 | 3 | 0 |
| 3.2 | 7.0 | 7.65 | 8.3 | 10.20 | 12.1 | 9.133333 | 2.650157 | 3 | 0 |
| 4.2 | 6.6 | 7.85 | 9.1 | 10.40 | 11.7 | 9.133333 | 2.550163 | 3 | 0 |
| 1.4 | 7.1 | 9.10 | 11.1 | 11.15 | 11.2 | 9.800000 | 2.338803 | 3 | 0 |
| 2.4 | 8.1 | 9.10 | 10.1 | 12.05 | 14.0 | 10.733333 | 3.000555 | 3 | 0 |
| 3.4 | 6.4 | 7.05 | 7.7 | 9.90 | 12.1 | 8.733333 | 2.987195 | 3 | 0 |
| 4.4 | 6.7 | 7.40 | 8.1 | 10.10 | 12.1 | 8.966667 | 2.802380 | 3 | 0 |
| 1.6 | 7.2 | 8.90 | 10.6 | 10.95 | 11.3 | 9.700000 | 2.193171 | 3 | 0 |
| 2.6 | 7.4 | 8.70 | 10.0 | 11.80 | 13.6 | 10.333333 | 3.113412 | 3 | 0 |
| 3.6 | 6.2 | 6.80 | 7.4 | 9.80 | 12.2 | 8.600000 | 3.174902 | 3 | 0 |
| 4.6 | 6.9 | 7.60 | 8.3 | 10.10 | 11.9 | 9.033333 | 2.579406 | 3 | 0 |
| 1.8 | 6.7 | 8.75 | 10.8 | 11.10 | 11.4 | 9.633333 | 2.557994 | 3 | 0 |
| 2.8 | 7.2 | 8.70 | 10.2 | 11.45 | 12.7 | 10.033333 | 2.753785 | 3 | 0 |
| 3.8 | 6.1 | 7.25 | 8.4 | 10.20 | 12.0 | 8.833333 | 2.973774 | 3 | 0 |
| 4.8 | 6.0 | 7.00 | 8.0 | 9.70 | 11.4 | 8.466667 | 2.730079 | 3 | 0 |

| temp.fat | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|----------|------|--------|--------|--------|------|--------|-----------|---|---------|
| 350.2 | 6.6 | 6.900 | 7.20 | 7.725 | 8.7 | 7.425 | 0.9105859 | 4 | 0 |
| 375.2 | 8.3 | 8.900 | 9.95 | 10.900 | 11.2 | 9.850 | 1.3771952 | 4 | 0 |
| 400.2 | 11.7 | 11.775 | 11.95 | 12.300 | 12.9 | 12.125 | 0.5439056 | 4 | 0 |
| 350.4 | 6.4 | 6.625 | 6.90 | 7.350 | 8.1 | 7.075 | 0.7410578 | 4 | 0 |
| 375.4 | 7.7 | 8.000 | 9.10 | 10.350 | 11.1 | 9.250 | 1.6196707 | 4 | 0 |
| 400.4 | 11.2 | 11.875 | 12.10 | 12.575 | 14.0 | 12.350 | 1.1789826 | 4 | 0 |
| 350.6 | 6.2 | 6.725 | 7.05 | 7.250 | 7.4 | 6.925 | 0.5251984 | 4 | 0 |
| 375.6 | 7.4 | 8.075 | 9.15 | 10.150 | 10.6 | 9.075 | 1.4818344 | 4 | 0 |
| 400.6 | 11.3 | 11.750 | 12.05 | 12.550 | 13.6 | 12.250 | 0.9746794 | 4 | 0 |
| 350.8 | 6.0 | 6.075 | 6.40 | 6.825 | 7.2 | 6.500 | 0.5597619 | 4 | 0 |
| 375.8 | 8.0 | 8.300 | 9.30 | 10.350 | 10.8 | 9.350 | 1.3601471 | 4 | 0 |
| 400.8 | 11.4 | 11.400 | 11.70 | 12.175 | 12.7 | 11.875 | 0.6184658 | 4 | 0 |

| temp.day | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|----------|------|--------|--------|--------|------|--------|-----------|---|---------|
| 350.1 | 6.7 | 7.000 | 7.15 | 7.250 | 7.4 | 7.100 | 0.2943920 | 4 | 0 |
| 375.1 | 10.6 | 10.750 | 10.95 | 11.125 | 11.2 | 10.925 | 0.2753785 | 4 | 0 |
| 400.1 | 11.2 | 11.275 | 11.35 | 11.500 | 11.8 | 11.425 | 0.2629956 | 4 | 0 |
| 350.2 | 7.2 | 7.350 | 7.75 | 8.250 | 8.7 | 7.850 | 0.6855655 | 4 | 0 |
| 375.2 | 10.0 | 10.075 | 10.15 | 10.350 | 10.8 | 10.275 | 0.3593976 | 4 | 0 |
| 400.2 | 12.7 | 12.850 | 13.25 | 13.700 | 14.0 | 13.300 | 0.6055301 | 4 | 0 |
| 350.3 | 6.1 | 6.175 | 6.30 | 6.550 | 7.0 | 6.425 | 0.4031129 | 4 | 0 |
| 375.3 | 7.4 | 7.625 | 8.00 | 8.325 | 8.4 | 7.950 | 0.4795832 | 4 | 0 |
| 400.3 | 12.0 | 12.075 | 12.10 | 12.125 | 12.2 | 12.100 | 0.0816497 | 4 | 0 |
| 350.4 | 6.0 | 6.450 | 6.65 | 6.750 | 6.9 | 6.550 | 0.3872983 | 4 | 0 |
| 375.4 | 8.0 | 8.075 | 8.20 | 8.500 | 9.1 | 8.375 | 0.4991660 | 4 | 0 |
| 400.4 | 11.4 | 11.625 | 11.80 | 11.950 | 12.1 | 11.775 | 0.2986079 | 4 | 0 |

This is a simple Split-Plot design model (fat is whole-plot factor and temperature is split-plot factor)

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijk}$$

for $i = 1, 2, 3, 4; j = 1, 2, 3; k = 1, 2, 3, 4$

μ is the overall true mean response;

τ_i is the effect of i^{th} replication of days;

β_j is the main effect of j^{th} level of temperature (effect of split-plot factor);

$(\tau\beta)_{ij}$ is the interaction effect of i^{th} replication and j^{th} level of temperature;

γ_k is the main effect of k^{th} level of fat (effect of whole-plot factor);

$(\tau\gamma)_{ik}$ is the interaction effect of i^{th} replicatin and k^{th} level of fat(whole-plot error);

$(\beta\gamma)_{jk}$ is the interaction effect of j^{th} level of temperature and k^{th} level of fat;

$(\tau\beta\gamma)_{ijk}$ is the interaction effect of i^{th} replicatin, j^{th} level of temperature and k^{th} level of fat (sub-plot error);

y_{ijk} is response value for the i^{th} replication when j^{th} level of temperature and k^{th} level of fat are applied;

ε_{ijk} is random error for the i^{th} replication when j^{th} level of temperature and k^{th} level of fat are applied.

Assumptions: For an experienced baker, he/she will try to let the recipe and temperature are accurate in each day. the covariance between two observations from the same level of the random factor can be either positive or negative. Thus, we assume this is a **restricted model**.

| | | |
|--|--|--|
| $\varepsilon_{ijk} \sim iidN(0, \sigma^2)$ | $\tau_i \sim iidN(0, \sigma_\tau^2)$ | |
| $\sum_{j=1}^3 \beta_j = 0$ | $\sum_{j=1}^3 (\tau\beta)_{ij} = 0$ | $(\tau\beta)_{ij} \sim iidN(0, \frac{3-1}{3} \sigma_{\tau\beta}^2)$ |
| $\sum_{k=1}^4 \gamma_k = 0$ | $\sum_{k=1}^4 (\tau\gamma)_{ik} = 0$ | $(\tau\gamma)_{ik} \sim iidN(0, \frac{4-1}{4} \sigma_{\tau\gamma}^2)$ |
| $\sum_{j=1}^3 (\beta\gamma)_{jk} = 0$ | $\sum_{k=1}^4 (\beta\gamma)_{jk} = 0$ | |
| $\sum_{j=1}^3 (\tau\beta\gamma)_{ijk} = 0$ | $\sum_{k=1}^4 (\tau\beta\gamma)_{ijk} = 0$ | $(\tau\beta\gamma)_{ijk} \sim iidN(0, \frac{(3-1)(4-1)}{3 \times 4} \sigma_{\tau\beta\gamma}^2)$ |

ε_{ijk} , τ_i , $(\tau\beta)_{ij}$, $(\tau\gamma)_{ik}$, $(\beta\gamma)_{jk}$, and $(\tau\beta\gamma)_{ijk}$ are independent.

Since this is a simple replicated factorial design, I use $(\tau\beta\gamma)_{ijk}$ to compute SSE and df.

Table 10: Analysis of Variance Table

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---------------------|----|--------|---------|---------|-----------|
| day_r | 3 | 22.4 | 7.466 | 74.5 | 2.415e-10 |
| fat_f | 3 | 2.004 | 0.6681 | 7.012 | 0.009915 |
| temp_f | 2 | 214.1 | 107 | 41.18 | 0.0003131 |
| day_r:fat_f | 9 | 0.8575 | 0.09528 | 0.9508 | 0.5082 |
| day_r:temp_f | 6 | 15.6 | 2.599 | 25.94 | 6.251e-08 |
| fat_f:temp_f | 6 | 1.59 | 0.2649 | 2.644 | 0.05113 |
| Residual | 18 | 1.804 | 0.1002 | NA | NA |

The results show the interaction effect of day and fat is not significant at 0.05 significance level (P-value=0.5082).

The interaction effect of day and temperature is significant at 0.05 significance level (P-value= 6.251×10^{-08}).

The interaction effect of temperature and fat is not significant at 0.05 significance level (P-value=0.05113).

| Groups | Name | Variance | Std.Dev. | |
|----------|-------------|-----------|-----------|---------------------------------------|
| day:fat | (Intercept) | 1.697e-15 | 4.120e-08 | $\hat{\sigma}_{\tau\gamma}^2 = 0$ |
| day:temp | (Intercept) | 6.252e-01 | 7.907e-01 | $\hat{\sigma}_{\tau\beta}^2 = 0.6252$ |
| day | (Intercept) | 4.055e-01 | 6.368e-01 | $\hat{\sigma}_{\tau}^2 = 0.4055$ |
| Residual | | 9.856e-02 | 3.140e-01 | $\hat{\sigma}^2 = 0.9856$ |

| | 2.5 % | 97.5 % | |
|---------------|--------|--------|---|
| .sig01 | 0 | 0.1906 | $CI_{\sigma_{\tau\gamma}^2} : (0, 0.1906^2)$ |
| .sig02 | 0.4371 | 1.242 | $CI_{\sigma_{\tau\beta}^2} : (0.4371^2, 1.242^2)$ |
| .sig03 | 0 | 1.626 | $CI_{\sigma_{\tau}^2} : (0, 1.626^2)$ |
| .sigma | 0.2116 | 0.3492 | $CI_{\sigma^2} : (0.2116^2, 0.3492^2)$ |

The results of variance components show the variance of interaction term of days and fat is negligible and hence dropping interaction term of days and fat.

Table 13: Analysis of Variance Table

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---------------------|----|--------|---------|---------|-----------|
| day_r | 3 | 22.4 | 7.466 | 75.75 | 2.881e-13 |
| fat_f | 3 | 2.004 | 0.6681 | 6.778 | 0.00149 |
| temp_f | 2 | 214.1 | 107 | 41.18 | 0.0003131 |
| day_r:temp_f | 6 | 15.6 | 2.599 | 26.37 | 4.298e-10 |
| fat_f:temp_f | 6 | 1.59 | 0.2649 | 2.688 | 0.03544 |
| Residual | 27 | 2.661 | 0.09856 | NA | NA |

The ANOVA table of new model shows that there is a significant interaction effect from the day and temperature, on average amount of force (g) (p-value= 4.298×10^{-10}). Similarly, there is a significant interaction effect from the fat and temperature, on average amount of force (g) (p-value=0.03544).

This means that the effects of day v.s.temperature and fat v.s.temperature on the force are not independent. Hence, the simple effects must be tested. The variances and confidence intervals don't change.

| Groups | Name | Variance | Std.Dev. | |
|----------|-------------|----------|----------|---------------------------------------|
| day:temp | (Intercept) | 0.62520 | 0.7907 | $\hat{\sigma}_{\tau\beta}^2 = 0.6252$ |
| day | (Intercept) | 0.40554 | 0.6368 | $\hat{\sigma}_{\tau}^2 = 0.40554$ |
| Residual | | 0.09856 | 0.3140 | $\hat{\sigma}^2 = 0.9856$ |

| | 2.5 % | 97.5 % | |
|---------------|--------|--------|---|
| .sig02 | 0.4371 | 1.242 | $CI_{\sigma_{\tau\beta}^2} : (0.4371^2, 1.242^2)$ |
| .sig03 | 0 | 1.626 | $CI_{\sigma_{\tau}^2} : (0, 1.626^2)$ |
| .sigma | 0.2116 | 0.3492 | $CI_{\sigma^2} : (0.2116^2, 0.3492^2)$ |

Since there are several simple effects comparison tests, the Tukey's adjustment was used to compute p value to reduce the experimentwise error rate. The Tables below show the summary of all those simple effect comparison tests.

For all levels of fat, the mean forces are significantly different between 350°F and 400°F (p value=0.0011369, 0.0005290, 0.0004966, 0.0004663 respectively).

When the fat was 4, 6 and 8, the mean forces are significantly different between 375°F and 400°F (p value=0.0145686, 0.0127251, 0.0432588 respectively).

When the fat was 8, the mean forces are significantly different between 350°F and 375°F (p value=0.0231342).

When temperature was 350°F, the mean forces are significantly different between the fat of 2 and 8 (p value=0.0063518).

When temperature was 375°F, the mean forces are significantly different between the fat of 2 and 6 (p value=0.0326101).

For all the rest of temperature, the mean forces are not significantly different between any value of fat.

When the day 1, the mean forces are not significantly different between 375°F and 400°F (p value=0.3674143).

For all the rest of days, the mean forces are significantly different between any value of temperature.

For all levels of temperature, the mean forces are significantly different between the day 2 v.s. day 3, and day 2 v.s. day 4 (p value<0.0002).

When temperature was 350°F and 400°F, the mean forces are significantly different between the day 1 and day 2 (p value=0.0421959, 0.0000001).

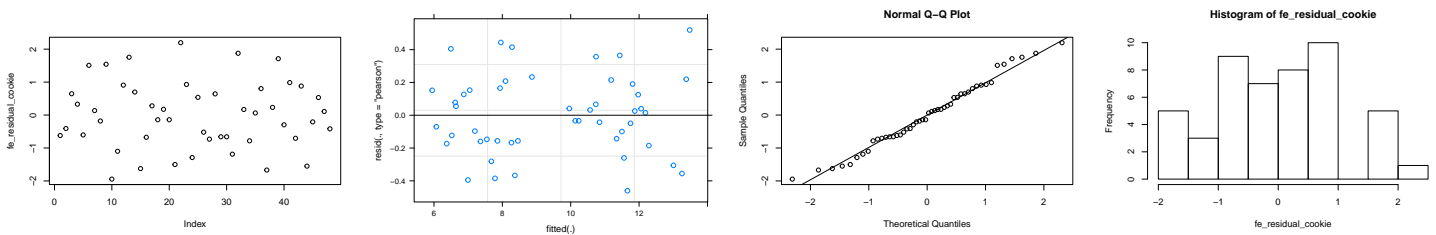
When temperature was 375°F, the mean forces are significantly different between the day 1 v.s. day 3, and day 1 v.s. day 4 (p value=0.0000000).

| fat | temp | contrast | estimate | SE | df | t.ratio | p.value |
|-----|------|-----------|---------------|------------------|------------------|-------------------|------------------|
| 2 | . | 350 - 375 | -2.425 | 0.6015677 | 7.421373 | -4.0311341 | 0.0527024 |
| 2 | . | 350 - 400 | -4.700 | 0.6015677 | 7.421373 | -7.8129198 | 0.0011369 |
| 2 | . | 375 - 400 | -2.275 | 0.6015677 | 7.421373 | -3.7817856 | 0.0711075 |
| 4 | . | 350 - 375 | -2.175 | 0.6015677 | 7.421373 | -3.6155533 | 0.0869670 |
| 4 | . | 350 - 400 | -5.275 | 0.6015677 | 7.421373 | -8.7687557 | 0.0005290 |
| 4 | . | 375 - 400 | -3.100 | 0.6015677 | 7.421373 | -5.1532024 | 0.0145686 |
| 6 | . | 350 - 375 | -2.150 | 0.6015677 | 7.421373 | -3.5739952 | 0.0914689 |
| 6 | . | 350 - 400 | -5.325 | 0.6015677 | 7.421373 | -8.8518719 | 0.0004966 |
| 6 | . | 375 - 400 | -3.175 | 0.6015677 | 7.421373 | -5.2778766 | 0.0127251 |
| 8 | . | 350 - 375 | -2.850 | 0.6015677 | 7.421373 | -4.7376216 | 0.0231342 |
| 8 | . | 350 - 400 | -5.375 | 0.6015677 | 7.421373 | -8.9349880 | 0.0004663 |
| 8 | . | 375 - 400 | -2.525 | 0.6015677 | 7.421373 | -4.1973665 | 0.0432588 |
| . | 350 | 2 - 4 | 0.350 | 0.2219964 | 27.000000 | 1.5766020 | 0.7734870 |
| . | 350 | 2 - 6 | 0.500 | 0.2219964 | 27.000000 | 2.2522886 | 0.3674143 |
| . | 350 | 2 - 8 | 0.925 | 0.2219964 | 27.000000 | 4.1667340 | 0.0063518 |
| . | 350 | 4 - 6 | 0.150 | 0.2219964 | 27.000000 | 0.6756866 | 0.9976187 |
| . | 350 | 4 - 8 | 0.575 | 0.2219964 | 27.000000 | 2.5901319 | 0.2117431 |
| . | 350 | 6 - 8 | 0.425 | 0.2219964 | 27.000000 | 1.9144453 | 0.5690221 |
| . | 375 | 2 - 4 | 0.600 | 0.2219964 | 27.000000 | 2.7027464 | 0.1724320 |
| . | 375 | 2 - 6 | 0.775 | 0.2219964 | 27.000000 | 3.4910474 | 0.0326101 |
| . | 375 | 2 - 8 | 0.500 | 0.2219964 | 27.000000 | 2.2522886 | 0.3674143 |
| . | 375 | 4 - 6 | 0.175 | 0.2219964 | 27.000000 | 0.7883010 | 0.9936958 |
| . | 375 | 4 - 8 | -0.100 | 0.2219964 | 27.000000 | -0.4504577 | 0.9998421 |
| . | 375 | 6 - 8 | -0.275 | 0.2219964 | 27.000000 | -1.2387587 | 0.9209247 |
| . | 400 | 2 - 4 | -0.225 | 0.2219964 | 27.000000 | -1.0135299 | 0.9723239 |
| . | 400 | 2 - 6 | -0.125 | 0.2219964 | 27.000000 | -0.5630722 | 0.9992813 |
| . | 400 | 2 - 8 | 0.250 | 0.2219964 | 27.000000 | 1.1261443 | 0.9511428 |
| . | 400 | 4 - 6 | 0.100 | 0.2219964 | 27.000000 | 0.4504577 | 0.9998421 |
| . | 400 | 4 - 8 | 0.475 | 0.2219964 | 27.000000 | 2.1396742 | 0.4309173 |
| . | 400 | 6 - 8 | 0.375 | 0.2219964 | 27.000000 | 1.6892165 | 0.7088627 |

| day_r | temp_f | contrast | estimate | SE | df | t.ratio | p.value |
|-------|--------|-----------|---------------|------------------|-----------|--------------------|------------------|
| 1 | . | 350 - 375 | -3.825 | 0.2219964 | 27 | -17.2300081 | 0.0000000 |
| 1 | . | 350 - 400 | -4.325 | 0.2219964 | 27 | -19.4822968 | 0.0000000 |
| 1 | . | 375 - 400 | -0.500 | 0.2219964 | 27 | -2.2522886 | 0.3674143 |
| 2 | . | 350 - 375 | -2.425 | 0.2219964 | 27 | -10.9235999 | 0.0000000 |
| 2 | . | 350 - 400 | -5.450 | 0.2219964 | 27 | -24.5499462 | 0.0000000 |
| 2 | . | 375 - 400 | -3.025 | 0.2219964 | 27 | -13.6263463 | 0.0000000 |
| 3 | . | 350 - 375 | -1.525 | 0.2219964 | 27 | -6.8694804 | 0.0000060 |
| 3 | . | 350 - 400 | -5.675 | 0.2219964 | 27 | -25.5634761 | 0.0000000 |
| 3 | . | 375 - 400 | -4.150 | 0.2219964 | 27 | -18.6939957 | 0.0000000 |
| 4 | . | 350 - 375 | -1.825 | 0.2219964 | 27 | -8.2208535 | 0.0000002 |
| 4 | . | 350 - 400 | -5.225 | 0.2219964 | 27 | -23.5364163 | 0.0000000 |
| 4 | . | 375 - 400 | -3.400 | 0.2219964 | 27 | -15.3155628 | 0.0000000 |
| . | 350 | 1 - 2 | -0.750 | 0.2219964 | 27 | -3.3784330 | 0.0421959 |
| . | 350 | 1 - 3 | 0.675 | 0.2219964 | 27 | 3.0405897 | 0.0881989 |
| . | 350 | 1 - 4 | 0.550 | 0.2219964 | 27 | 2.4775175 | 0.2573498 |
| . | 350 | 2 - 3 | 1.425 | 0.2219964 | 27 | 6.4190226 | 0.0000188 |
| . | 350 | 2 - 4 | 1.300 | 0.2219964 | 27 | 5.8559505 | 0.0000805 |
| . | 350 | 3 - 4 | -0.125 | 0.2219964 | 27 | -0.5630722 | 0.9992813 |
| . | 375 | 1 - 2 | 0.650 | 0.2219964 | 27 | 2.9279752 | 0.1112061 |
| . | 375 | 1 - 3 | 2.975 | 0.2219964 | 27 | 13.4011174 | 0.0000000 |
| . | 375 | 1 - 4 | 2.550 | 0.2219964 | 27 | 11.4866721 | 0.0000000 |
| . | 375 | 2 - 3 | 2.325 | 0.2219964 | 27 | 10.4731422 | 0.0000000 |
| . | 375 | 2 - 4 | 1.900 | 0.2219964 | 27 | 8.5586968 | 0.0000001 |
| . | 375 | 3 - 4 | -0.425 | 0.2219964 | 27 | -1.9144453 | 0.5690221 |
| . | 400 | 1 - 2 | -1.875 | 0.2219964 | 27 | -8.4460824 | 0.0000001 |
| . | 400 | 1 - 3 | -0.675 | 0.2219964 | 27 | -3.0405897 | 0.0881989 |
| . | 400 | 1 - 4 | -0.350 | 0.2219964 | 27 | -1.5766021 | 0.7734870 |
| . | 400 | 2 - 3 | 1.200 | 0.2219964 | 27 | 5.4054927 | 0.0002601 |
| . | 400 | 2 - 4 | 1.525 | 0.2219964 | 27 | 6.8694804 | 0.0000060 |
| . | 400 | 3 - 4 | 0.325 | 0.2219964 | 27 | 1.4639876 | 0.8314608 |

Conclusion: Choosing a higher temperature for a given amount of fat can get a harder texture of the surface. In most of the cases, changing fat for a given temperature doesn't change the texture.

For a given temperature, almost all of the amount of fat cannot change the texture but different days don't give consistent results. If the baker wants to examine the texture for a specific temperature, he/she should check what other factors in different days may affect the results and redo the experiment.



In the plots of residuals versus predicted value of purity, there is no significant pattern on this plot. Therefore, the fitted model is good enough to describe the relationship between the mean value of force and the days, temperature, and fats.

The residuals in this plot are almost symmetrically distributed about zero and hence zero mean assumption is not violated. Further, the vertical deviation of the residuals from zero is about same for each predicted value and hence the constant variance assumption is not violated.

The points are along the straight line in the normal qq plot shown at bottom left and the histogram of residuals shown at the top right is about normal. These plots show no violation of normal distribution assumption of residuals.

```
favstats(force ~ day, data = fe_table_cookie)
favstats(force ~ fat, data = fe_table_cookie)
favstats(force ~ temp, data = fe_table_cookie)
favstats(force ~ day + fat, data = fe_table_cookie)
favstats(force ~ temp + fat, data = fe_table_cookie)
favstats(force ~ temp + day, data = fe_table_cookie)

fe_table_cookie$day_r <- as.random(fe_table_cookie$day)
fe_table_cookie$temp_f <- as.fixed(fe_table_cookie$temp)
fe_table_cookie$fat_f <- as.fixed(fe_table_cookie$fat)
fe_model_cookie1 <- aov(formula = force ~ day_r + fat_f + fat_f %in% day_r +
  temp_f %in% day_r + temp_f + temp_f:fat_f, data = fe_table_cookie)
gad(fe_model_cookie1)

fe_model_cookie2 <- lmer(formula = force ~ (1 | day) + fat + (1 | day:fat) +
  temp + (1 | day:temp) + temp:fat, data = fe_table_cookie, REML = TRUE)
summary(fe_model_cookie2)
confint(fe_model_cookie2)

fe_model_cookie3 <- aov(formula = force ~ day_r + fat_f + temp_f + temp_f %in%
  day_r + temp_f:fat_f, data = fe_table_cookie)
gad(fe_model_cookie3)
fe_model_cookie4 <- lmer(formula = force ~ (1 | day) + fat + temp + (1 | day:temp) +
  temp * fat, data = fe_table_cookie, REML = TRUE)
summary(fe_model_cookie4)
confint(fe_model_cookie4)

temp_fat <- pairs(lsmeans(fe_model_cookie4, ~temp | fat))
```

```

fat_temp <- pairs(lsmeans(fe_model_cookie4, ~fat | temp))
test(rbind(temp_fat, fat_temp), adjust = "tukey")
temp_day <- pairs(lsmeans(fe_model_cookie3, ~temp_f | day_r))
day_temp <- pairs(lsmeans(fe_model_cookie3, ~day_r | temp_f))
test(rbind(temp_day, day_temp), adjust = "tukey")

fe_residual_cookie <- rstudent(fe_model_cookie4)
plot(fe_residual_cookie)
plot(fe_model_cookie4)
qqnorm(fe_residual_cookie)
qqline(fe_residual_cookie)
hist(fe_residual_cookie)

```