Homework 2

Ricardo Andres Calvo Mendez

February 16, 2026

Question 1

Use the etch rate data on lecture slide to answer this question.

Power (W)	**Obs 1**	**Obs 2**	**Obs 3**	**Obs 4**	**Obs 5**	**Total**	**Mean**
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

a. Fit the three types of models on the lecture notes. Write down the three models.

a.1 Dummy coding: $y_{ij} = \beta_0 + \beta_1 X_{180} + \beta_2 X_{200} + \beta_3 X_{220}$

Using RF 160 W as base, the values of β are:

$$\beta_0 = \overline{y}_{160}$$
$$\beta_0 = 551.2$$

$$\beta_1 = \overline{y}_{180} - \overline{y}_{160}$$
$$\beta_1 = 587.4 - 551.2$$
$$\beta_1 = 36.2$$

$$\beta_2 = \overline{y}_{200} - \overline{y}_{160}$$
$$\beta_2 = 625.4 - 551.2$$
$$\beta_2 = 74.2$$

$$\beta_3 = \overline{y}_{220} - \overline{y}_{160}$$
$$\beta_3 = 707.0 - 551.2$$
$$\beta_3 = 155.8$$

a.2 Mean model: $y_{ij} = \mu_i + \epsilon_{ij}$

This can be directly inferred from the table:

$$\mu_0 = \overline{y}_{160}$$
$$\mu_0 = 551.2$$

$$\mu_1 = \overline{y}_{180}$$
$$\mu_1 = 587.4$$

$$\mu_2 = \overline{y}_{200}$$
$$\mu_2 = 625.4$$

$$\mu_3 = \overline{y}_{220}$$
$$\mu_3 = 707.0$$

a.3 Effects model: $y_{ij} = \mu + \tau_i + \epsilon_{ij}, \Sigma(\tau_i) = 0$

Computing the overall mean:

$$\overline{y} = \frac{\overline{y}_{160} + \overline{y}_{180} + \overline{y}_{200} + \overline{y}_{220}}{4}$$
$$\overline{y} = \frac{551.2 + 587.4 + 625.4 + 707.0}{4}$$
$$\overline{y} = 617.75$$

Now computing each effect:

$$\tau_0 = \overline{y}_{160} - \overline{y}$$
$$\tau_0 = 551.2 - 617.75$$
$$\tau_0 = -66.55$$

$$\tau_1 = \overline{y}_{180} - \overline{y}$$
$$\tau_1 = 587.4 - 617.75$$
$$\tau_1 = -30.35$$

$$\tau_2 = \overline{y}_{200} - \overline{y}$$
$$\tau_2 = 625.4 - 617.75$$
$$\tau_2 = 7.65$$

$$\tau_3 = \overline{y}_{220} - \overline{y}$$
$$\tau_3 = 707.0 - 617.75$$
$$\tau_3 = 89.25$$

b. What is your prediction of etch rate when RF Power=200 based on each of the model you obtained in a.? Are the predictions the same from the three models?

b.1 Dummy coding: $y_{ij} = \beta_0 + \beta_1 X_{180} + \beta_2 X_{200} + \beta_3 X_{220}$

$$\hat{\mu}_2 = \beta_0 + \beta_2$$
$$\hat{\mu}_2 = 551.2 + 74.2$$
$$\hat{\mu}_2 = 625.4$$

b.2 Mean model: $y_{ij} = \mu_i + \epsilon_{ij}$

For this model, the value is directly the mean of the RF 200 W:

$$\hat{\mu}_2 = \mu_2$$
$$\hat{\mu}_2 = 625.4$$

b.3 Effects model: $y_{ij} = \mu + \tau_i + \epsilon_{ij}, \Sigma(\tau_i) = 0$

$$\hat{\mu}_2 = \mu + \tau_2$$
$$\hat{\mu}_2 = 617.75 + 7.65$$
$$\hat{\mu}_2 = 625.4$$

The prediction is the same for the three models.

c. Are the etch rates significantly different with RF Power=160 and 180?

To respond this, we do a hypothesis test:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

We compute the t-statistic:

$$t_0 = \frac{\overline{y}_1 - \overline{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

First compute S_1^2,S_2^2 and S_p^2

$$S_1^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{y}_{160})^2}{n-1}$$
$$S_1^2 = \frac{\sum_{i=1}^{5}(x_i - \overline{y}_{160})^2}{5-1}$$
$$S_1^2 = \frac{1602.8}{4}$$
$$S_1^2 = 400.7$$

$$S_1^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{y}_{180})^2}{n-1}$$
$$S_1^2 = \frac{\sum_{i=1}^{5}(x_i - \overline{y}_{180})^2}{5-1}$$
$$S_1^2 = \frac{1121.2}{4}$$
$$S_1^2 = 280.3$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$
$$S_p^2 = \frac{(5-1)(400.7) + (5-1)(280.3)}{5+5-2}$$
$$S_p^2 = \frac{(4)(400.7) + (4)(280.3)}{8}$$
$$S_p^2 = \frac{1602.8 + 1121.2}{8}$$
$$S_p^2 = 340.5$$

Now with this information, compute t_0

$$
\begin{aligned}
t_0 &= \frac{\overline{y}_1 - \overline{y}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\\
t_0 &= \frac{551.2 - 587.4}{\sqrt{340.5}\sqrt{\frac{1}{5}+\frac{1}{5}}}\\
t_0 &= \frac{551.2 - 587.4}{\sqrt{340.5}\sqrt{\frac{1}{5}+\frac{1}{5}}}\\
t_0 &= -3.1018
\end{aligned}
$$

Let $\alpha = 0.05$

$$
\begin{aligned}
p_{value}(|t_0|, n_1 + n_2 - 2) &< \alpha\\
p_{value}(3.1018, 8) &< 0.05\\
0.014625 &< 0.05
\end{aligned}
$$

We conclude that we have enough information to conclude that the rates 160 and 180 are significantly different at $\alpha = 0.05$

d. Which RF Power provides the largest etch rate?

We can observe this directly from the data:
$\overline{y}_{220} = 707.0$ is the largest etch rate at RF Power 220 W

Question 2

A computer ANOVA output is shown below. Fill in the blanks

Source	DF	SS	MS	F	P
Factor	4	987.71	246.93	33.0949	1.18×10^{-9}
Error	25	186.53	7.4612		
Total	29	1174.24			

To compute the missing data, we compute the following:

$$
\begin{aligned}
F_{df} &= T_{df} - E_{df}\\
F_{df} &= 29 - 25\\
F_{df} &= 4
\end{aligned}
$$

$$SS_{factor} = SS_{total} - SS_{error}$$
$$SS_{factor} = 1174.24 - 186.53$$
$$SS_{factor} = 987.71$$

$$MS_{error} = \frac{SS_{error}}{E_{df}}$$
$$MS_{error} = \frac{186.53}{25}$$
$$MS_{error} = 7.4612$$

$$F = \frac{MS_{factor}}{MS_{factor}}$$
$$F = \frac{246.93}{7.4612}$$
$$F = 33.0949$$

For P value we use R code:

```
pf(f_stat, df1 = df_factor, df2 = df_error, lower.tail = FALSE)
```

$$P = 1.18 \times 10^{-9}$$

Question 3

The tensile strength of Portland cement is being studied. Four different mixing techniques can be used economically. A completely randomized experiment was conducted and the following data were collected:

Mixing Technique	**Tensile Strength (lb/in^2)**			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

a. Test the hypothesis that mixing techniques affect the strength of the cement. Use $\alpha = 0.05$

Define the hypothesis test:

$$
\begin{aligned}
H_0 &: \mu_1 = \mu_2 = \mu_3 = \mu_4 \\
H_1 &: \mu_i \neq \mu_j \text{for at least one pair } (i, j)
\end{aligned}
$$

Let:

$$
\begin{aligned}
N &= 16 \\
a &= 4 \\
\overline{y}_{1.} &= 2971 \\
\overline{y}_{2.} &= 3156.25 \\
\overline{y}_{3.} &= 2933.75 \\
\overline{y}_{4.} &= 2666.25 \\
\overline{y}_{..} &= 2931.812
\end{aligned}
$$

We compute the Sums of the Squares (SS)

$$
\begin{aligned}
SS_{Treatments} &= \sum_{i=1}^{a} n_i(\overline{y}_{i.} - \overline{y}_{..})^2 \\
SS_{Treatments} &= 489740.2
\end{aligned}
$$

$$
\begin{aligned}
SS_{Total} &= \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{..})^2 \\
SS_{Total} &= 643648.4
\end{aligned}
$$

$$
\begin{aligned}
SS_E &= SS_{Total} - SS_{Treatments} \\
SS_E &= 153908.2
\end{aligned}
$$

Now compute the F Statistic:

$$
\begin{aligned}
F_0 &= \frac{SS_{\text{Treatments}}/(a-1)}{SS_E/(N-a)} \\
F_0 &= 12.72811
\end{aligned}
$$

For P value we use R code: `pf(f_stat, df1 = df_treat, df2 = df_error, lower.tail = FALSE)`

$$P_{value} = 0.000489$$

Now we observe that:

$$P_{value} < \alpha$$
$$0.000489 < 0.05$$

We can conclude that we have strong evidence to reject the null hypothesis H_0. Mixing techniques significantly affect tensile strength

b. Use the derivation of ANOVA table on lecture notes, reproduce the ANOVA table by hand (or with your own code): compute the group means, grand mean, sum of squares, df, MS, and F, and verify they match your software output (R/Python/Matlab).

This is the ANOVA table made by my manual computations, some of the data was previously computed in a.:

Source	DF	SS	MS	F
Factor	3	489740.2	163246.73	12.72811
Error	12	153908.2	12825.69	
Total	15	643648.4		

$$df_{Factor} = a - 1$$
$$df_{Factor} = 3$$

$$df_{Error} = N - a$$
$$df_{Error} = 12$$

$$df_{Total} = N - 1$$
$$df_{Total} = 15$$

$$MS_{Factor} = SS_{\text{Factor}}/(a-1)$$
$$MS_{Factor} = 163246.73$$

$$MS_{Error} = SS_{\text{Error}}/(N-a)$$
$$MS_{Error} = 12825.69$$

Now this is the R code used to compute ANOVA table

R code

```
technique <- factor(rep(1:4, each = 4))
strength <- c(
  3129, 3000, 2865, 2890,
  3200, 3300, 2975, 3150,
  2800, 2900, 2985, 3050,
  2600, 2700, 2600, 2765
)
data <- data.frame(Technique = technique, Strength =
↪  strength)
fit <- lm(Strength ~ Technique, data = data)
anova(fit)
```

The results are summarized below:

R output

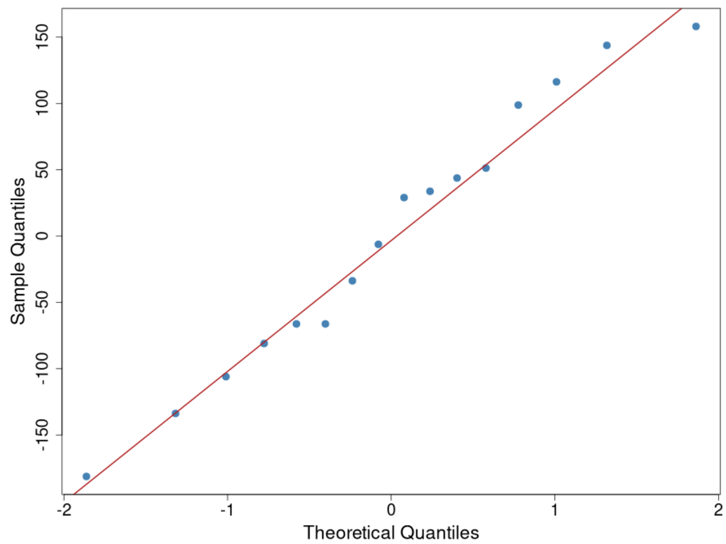
```
Analysis of Variance Table

Response: Strength
          Df Sum Sq Mean Sq F value    Pr(>F)
Technique  3 489740  163247  12.728 0.0004887 ***
Residuals 12 153908   12826
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We can see that the values are the same, it only changes some digits because of precision

c. Construct a normal probability plot of the residuals. What conclusion would you draw about the validity of the normality assumption?

This is the plot created using R `resid()` function

Figure 1: Residual Plot for Checking Normality

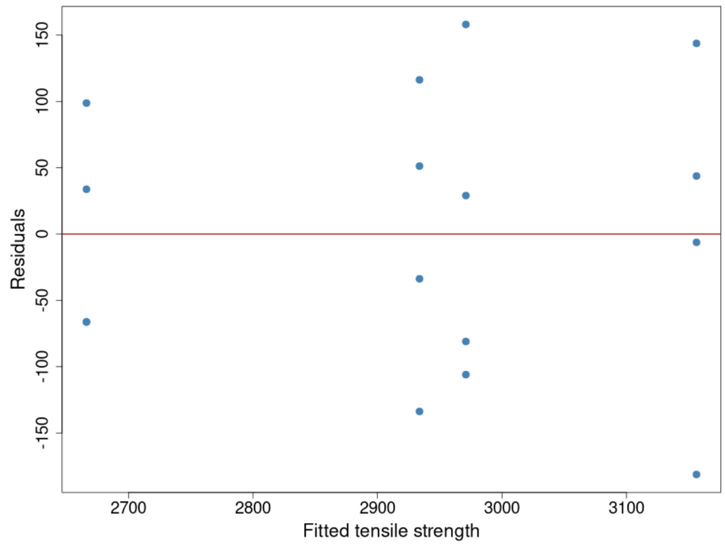


The normal probability plot of the residuals shows that the points closely follow the normal reference line, indicating that the residuals follow the normality pattern expected under the model.

d. Plot the residuals versus the predicted tensile strength. Comment on the plot.

This is the plot created using R `resid()` and `fitted()` functions

Figure 2: Plot of Residuals versus fitted values



The residuals-versus-fitted plot shows points scattered around zero, with no clear trend or funnel shape. This suggests that the constant-variance assumption is reasonable and that the one-way ANOVA model form is adequate for these data.

Question 4

A product developer is investigating the tensile strength of a new synthetic fiber that will be used to make cloth for men's shirts. Strength is usually affected by the percentage of cotton used in the blend of materials for the fiber. The engineer conducts a completely randomized experiment with five levels of cotton content and replicates the experiment five times. The data are shown in the following table.

Cotton Weight Percent	Observations				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	19	19	18	18
30	19	25	22	19	23
35	7	10	11	15	11

a. Is there evidence to support the claim that cotton content affects the mean tensile strength? Use $\alpha = 0.05$

To respond this, let define the following hypothesis test::

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$
$$H_1 : \mu_i \neq \mu_j \text{for at least one pair } (i, j)$$

Let:

$$N = 25$$
$$a = 5$$
$$\overline{y}_{1.} = 9.8$$
$$\overline{y}_{2.} = 15.4$$
$$\overline{y}_{3.} = 17.6$$
$$\overline{y}_{4.} = 21.6$$
$$\overline{y}_{5.} = 10.8$$
$$\overline{y}_{..} = 15.04$$

We compute the Sums of the Squares (SS)

$$SS_{Factor} = \sum_{i=1}^{a} n_i(\overline{y}_{i.} - \overline{y}_{..})^2$$
$$SS_{Factor} = 475.76$$

$$SS_{Total} = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \overline{y}_{..})^2$$
$$SS_{Total} = 636.96$$

$$SS_E = SS_{Total} - SS_{Factor}$$
$$SS_E = 161.2$$

Now compute the F Statistic:

$$F_0 = \frac{SS_{\text{Factor}}/(a-1)}{SS_E/(N-a)}$$
$$F_0 = 14.75682$$

For P value we use R code: `pf(f_stat, df1 = df_treat, df2 = df_error, lower.tail = FALSE)`

$$P_{value} = 9.12 \times 10^{-6}$$

Now we observe that:

$$P_{value} < \alpha$$
$$9.12 \times 10^{-6} < 0.05$$

We can reject the null hypothesis. Cotton content significantly affects mean tensile strength

b. Analyze the residuals from this experiment and comment on model adequacy.

For this, we construct a normal probability plot of the residuals and a plot the residuals versus the predicted tensile strength

Figure 3: Residual Plot for Checking Normality

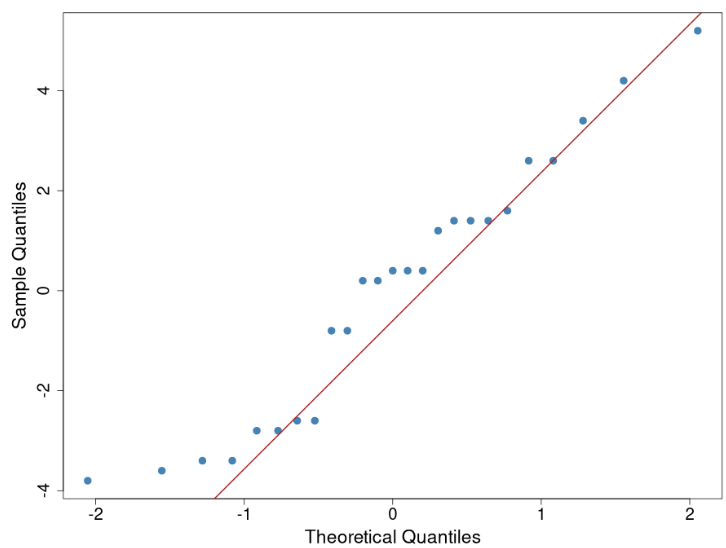
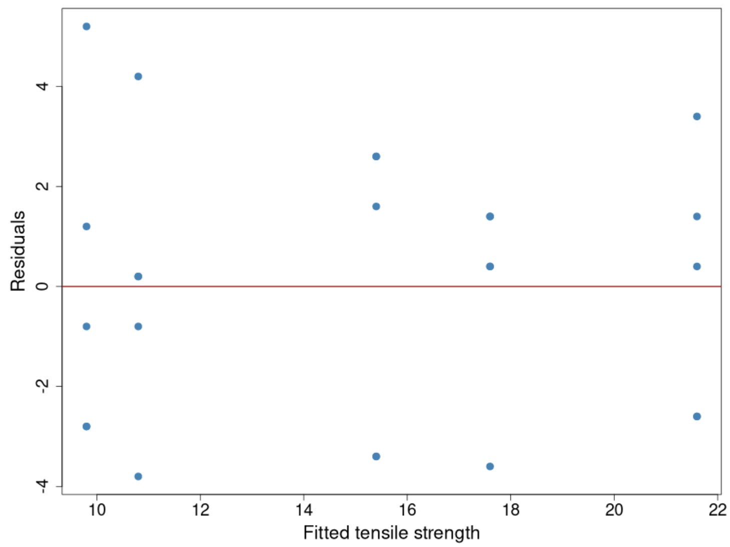


Figure 4: Plot of Residuals versus fitted values



As shown in fig. 3, the residual points lie close to the red reference line, which supports the normality assumption. In fig. 4, the residuals are randomly scattered around zero with no clear pattern, suggesting constant variance and indicating that the model is adequate.