Stat 514 Test 3

By – Aditya Gaitonde

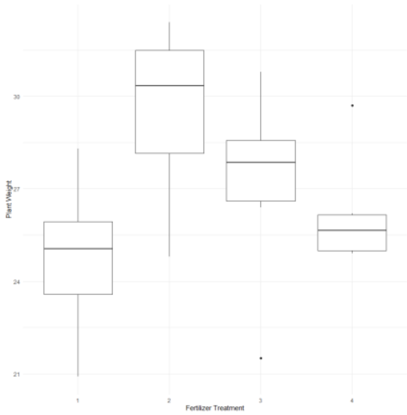
T1)
```
data <- data.frame(
  Block = c(rep(1, 4), rep(2, 4), rep(3, 4), rep(1, 4), rep(2, 4), rep(3, 4)),
  Ferti = factor(rep(1:4, 6)),
  Perc = c(13, 7, 10, 8, 12, 5, 9, 9, 7, 6, 6, 7, 9, 7, 6, 7, 8, 9, 5, 9, 10, 13, 15, 10),
  Weight = c(23.1, 30.1, 26.4, 26.2, 20.9, 31.8, 27.2, 25.3, 28.3, 32.4, 28.6, 29.7,
25.0, 30.6, 28.5, 26.0, 25.1, 27.5, 30.8, 24.9, 26.2, 24.8, 21.5, 24.9),
  Greenhouse = factor(c(rep(1, 12), rep(2, 12))))
model <- lmer(Weight ~ Ferti + Perc + Greenhouse + (1|Block), data =
data)
summary(model)
ggplot(data, aes(x = factor(Ferti), y = Weight)) +
  geom_boxplot() +
  labs(x = "Fertilizer Treatment", y = "Plant Weight") +
  theme_minimal()
ggplot(data, aes(x = Perc, y = Weight)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, linetype = "dashed", color =
"blue") +
  labs(x = "Blight Percentage", y = "Plant Weight") +
  theme_minimal()
ggplot(data, aes(x = factor(Greenhouse), y = Weight)) +
  geom_boxplot() +
  labs(x = "Greenhouse", y = "Plant Weight") +
  theme_minimal()
```

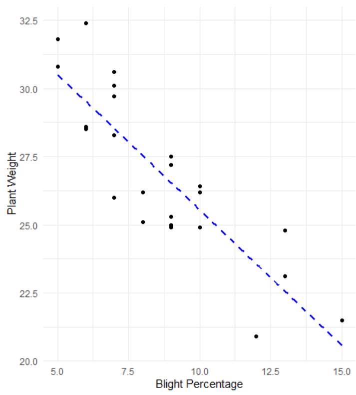
Output –

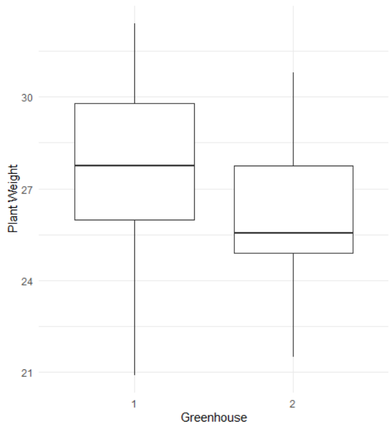
```
Linear mixed model fit by REML ['lmerMod']
Formula: Weight ~ Ferti + Perc + Greenhouse + (1 | Block)
   Data: data
REML criterion at convergence: 66.7
Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.78312 -0.36336 -0.06154  0.51216  1.59616
Random effects:
 Groups   Name        Variance Std.Dev.
 Block    (Intercept) 0.3038   0.5511
 Residual             0.9553   0.9774
Number of obs: 24, groups:  Block, 3
Fixed effects:
            Estimate Std. Error t value
(Intercept) 33.96251    0.95960  35.392
Ferti2       2.94728    0.58837   5.009
Ferti3       1.18708    0.57512   2.064
Ferti4       0.03546    0.57796   0.061
Perc        -0.90969    0.08329 -10.922
Greenhouse2 -0.50106    0.40388  -1.241
Correlation of Fixed Effects:
            (Intr) Ferti2 Ferti3 Ferti4 Perc
Ferti2      -0.514
Ferti3      -0.447  0.525
Ferti4      -0.465  0.529  0.521
```

Perc   -0.821 0.283 0.193 0.216

Greenhouse2 -0.078 -0.044 -0.030 -0.033 -0.155







From the plots and the results of the analysis it can be concluded that the fertilizer treatment and blight percentage had a significant impact on the weight of Mahogany plants. Fertilizer treatments 2 and 3 showed positive effects on plant weight, while blight percentage had a negative effect. The greenhouse effect did not show a significant influence, and block variability was accounted for as a random effect.

T2) a) Layout –

Week 1:

Chamber I (2°C): a - Plastic, b - Cardboard, c - Waxed Cardboard, d - Open

Chamber II (6°C): a - Waxed Cardboard, b - Open, c - Plastic, d - Cardboard

Chamber III (10°C): a - Open, b - Plastic, c - Cardboard, d - Waxed Cardboard

Week 2:

Chamber I (2°C): a - Cardboard, b - Waxed Cardboard, c - Open, d - Plastic

Chamber II (6°C): a - Open, b - Plastic, c - Cardboard, d - Waxed Cardboard

Chamber III (10°C): a - Waxed Cardboard, b - Open, c - Plastic, d - Cardboard

Week 3:

Chamber I (2°C): a - Waxed Cardboard, b - Open, c - Plastic, d - Cardboard

Chamber II (6°C): a - Cardboard, b - Waxed Cardboard, c - Open, d - Plastic

Chamber III (10°C): a - Open, b - Plastic, c - Cardboard, d - Waxed Cardboard

Week 4:

Chamber I (2°C): a - Open, b - Plastic, c - Cardboard, d - Waxed Cardboard

Chamber II (6°C): a - Waxed Cardboard, b - Open, c - Plastic, d - Cardboard

Chamber III (10°C): a - Cardboard, b - Waxed Cardboard, c - Open, d - Plastic b)We can use a split plot design with temperature as a whole-plot factor and container type as the subplot factor for analysis.

The model will be as such –

$$Y_{ijk} = \mu + T_i + B_j + C_k + (TC)_{ik} + \varepsilon_w + \varepsilon_s$$

Y_{ijkl} is the response, μ is the overall mean, T_i is the Temperature Effect and is the whole-plot factor, B_j is the Effect of Block, each chamber represents a different block in the design, C_k is the Container Effect is the subplot factor, $(TC)_{ik}$ is the Interaction effect between temperature and container, ε_w is the Whole plot error term, ε_s is the Subplot error term

ANOVA Table –

Source of Variation	df	SS	MS	F
Temperature	2	SS_T	MS_T	$F_T = \frac{MS_T}{MS_{\varepsilon_w}}$
Block	2	SS_B	MS_B	$F_B = \frac{MS_B}{MS_{\varepsilon_w}}$
Container	3	SS_C	MS_C	$F_C = \frac{MS_C}{MS_{\varepsilon_w}}$
Temperature x Container	6	SS_{TXC}	MS_{TXC}	$F_{TxC} = \frac{MS_{TxC}}{MS_{\varepsilon_w}}$
Whole plot Error	6	SS_{ε_w}	MS_{ε_w}	F_{ε_w}
Subplot Error	24	SS_{ε_S}	MS_{ε_S}	-
Total	47	SS_T	-	-

T3) a) To ensure a fair and balanced study, we will use a design where all 36 treatment combinations will be performed by each worker exactly once during the 18-day research period. Given 6 men, we have decided that each man will be allocated to work for 3 non-consecutive days. On each of those days, the two working men will be assigned to complete 6 tasks each.

One way to arrange it would be –

Man 1: (1 pail, Distance 3, Flat, Load 1), (1 pail, Distance 1, 4% incline, Load 2), (1 pail, Distance 2, Flat, Load 3), (2 pails, Distance 2, 4% incline, Load 1), (2 pails, Distance 1, Flat, Load 2), (2 pails, Distance 3, 4% incline, Load 3)

Man 2: (1 pail, Distance 1, Flat, Load 1), (1 pail, Distance 3, 4% incline, Load 2), (1 pail, Distance 2, 4% incline, Load 1), (2 pails, Distance 3, Flat, Load 1), (2 pails, Distance 2, 4% incline, Load 2), (2 pails, Distance 1, Flat, Load 3)

Repeat this process for each day, making sure all treatment combinations are covered once.

b) To ensure randomization in our study

- Randomly ordering the tasks for each worker on each day:
- Randomize the pairing of men on each workday
- Randomly selecting treatment combinations
- Randomly assigning the men to workdays

c) Model used –

$$Y_{ijklm} = \mu + P_i + D_j + I_k + L_l + (PD)_{ij} + (PI)_{ik} + (PL)_{il} + (DI)_{jk} + (DL)_{jl} + (IL)_{kl} + R_m + \varepsilon_{ijklm}$$

Y_{ijklm} is the oxygen uptake, μ is the overall mean, P_i, D_j, I_k, L_l, are the main effects of Pail, distance, incline and load respectively, $(PD)_{ij}, (PI)_{ik}, (PL)_{il}, (DI)_{jk}, (DL)_{jl}, (IL)_{kl}$ are the two factor interactions, R_m is the random effect of 'm' man and ε_{ijklm} is the random error term.

ANOVA Table –

Source	df	Mean Sq
P	1	$\sigma^2 + 36\sigma^2(R) + 18\sigma^2(P) + 6\sigma^2(PD) + 6\sigma^2(PI) + 6\sigma^2(PL)$
D	2	$\sigma^2 + 18\sigma^2(R) + 6\sigma^2(PD)$
I	1	$\sigma^2 + 36\sigma^2(R) + 18\sigma^2(PI) + 6\sigma^2(PI)$
L	2	$\sigma^2 + 18\sigma^2(R) + 6\sigma^2(PL)$
PxD	2	$\sigma^2 + 6\sigma^2(R) + 3\sigma^2(PD)$
PxI	1	$\sigma^2 + 18\sigma^2(R) + 9\sigma^2(PI)$
PxL	2	$\sigma^2 + 6\sigma^2(R) + 3\sigma^2(PL)$
DxI	2	$\sigma^2 + 6\sigma^2(R) + 3\sigma^2(DC)$
DxL	4	$\sigma^2 + 3\sigma^2(R) + 1.5\sigma^2(DL)$
IxL	2	$\sigma^2 + 6\sigma^2(R) + 3\sigma^2(IL)$
R	5	$\sigma^2 + 2\sigma^2(R)$
Error	180	σ^2

T4) Objective of the study – The primary aim is to investigate the influence of three factors (food diet, chemical treatment, and analytical technique) on the glycogen content in the livers of rats.

Type of experimental units – The rats serve as the experimental units, and their livers are divided into four segments for examination.

Constraints –

• 12 rats are allocated to three food diets (T1, T2, T3), with four rats per diet.

• Each rat's liver is divided into four segments.

• Each segment is treated with one of two chemicals (P1 and P2), with two segments per chemical.

• Two distinct analytical methods (A and B) are employed on the two segments that have been treated with the same chemical.

Restriction on randomization –

- Limited rat numbers: With only 12 rats in the study, randomization is constrained by the fixed number of rats per diet.
- Liver segment constraint: Each rat's liver is divided into four segments, which restricts the random assignment of chemicals and analytical methods.
- Allocation limitation: Due to two chemicals and two methods, each segment is prepared with one chemical and one method, limiting the randomization of exposure.

Despite these restrictions, the statistical model used incorporates random effects for rats, chemicals, and methods to account for these limitations and ensure valid results.

Model –

The selected model is a linear mixed-effects model:

model <- lmer(glycogen ~ food * prep * method + (1 | food:rat), data = data)

Output –

	Estimate	Std. Error	t value
(Intercept)	138.000	3.739	36.912
foodT2	17.750	5.287	3.357
foodT3	-6.250	5.287	-1.182
prepP2	-6.000	4.683	-1.281
methodB	-3.000	4.683	-0.641
foodT2:prepP2	1.500	6.623	0.226
foodT3:prepP2	6.000	6.623	0.906
foodT2:methodB	1.250	6.623	0.189
foodT3:methodB	0.250	6.623	0.038
prepP2:methodB	3.500	6.623	0.528
foodT2:prepP2:methodB	-2.000	9.366	-0.214
foodT3:prepP2:methodB	-1.250	9.366	-0.133

Findings of the analysis:

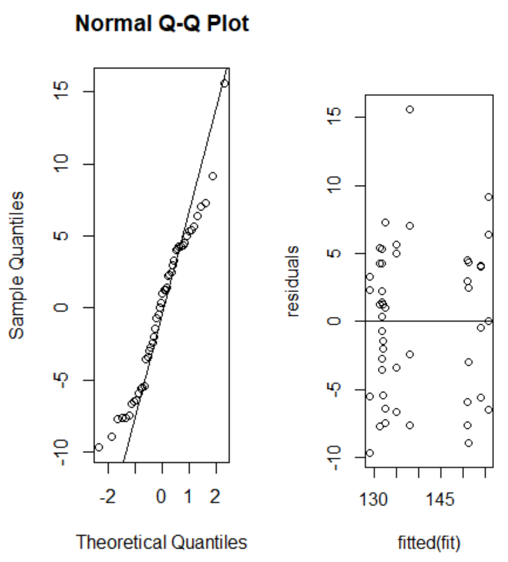
The analysis results indicate that there is a significant increase in glycogen levels for rats on food diet T2 in comparison to those on diet T1. Nonetheless, the differences in glycogen levels between the two chemical preparations and the analytical methods are not statistically significant. Furthermore, the interaction effects between the food diets, chemical preparations, and analytical methods do not exhibit any significant impact on glycogen levels.

Verification of model assumptions:

Code –

```
residuals <- residuals(model1)
qqnorm(residuals)
qqline(residuals)
plot(fitted(fit), residuals)
abline(h=0)
```

Plots –



The plots show no significant deviations from normality, indicating that standard assumptions are met.

Improved design recommendations –

- Increase rat sample size.
- Add a control group.
- Incorporate more food diets.
- Randomize treatments and methods.
- Study various durations.
- Complete factorial design

Design change for four chemicals and two methods –

- Stratified randomization:
    - Divide rats into three strata (T1, T2, T3).
    - Assign rats to chemicals (C1, C2, C3, C4) within strata.
    - Apply methods (A, B) on segments from different chemicals.

- Nested design:
    - Assign rats to chemicals (C1, C2, C3, C4).
    - Assign liver segments to methods (A, B) within chemicals.