School of Mathematics and Statistics

Te Kura Mātai Tatauranga

STAT 292 **Assignment 3: Solutions and Feedback** **2020**

Note: Sections in *italics* are extra feedback, not part of the required solutions.

1. **Skink Temperatures** [20 marks]

   (a) Other factors are held constant to reduce the amount of extraneous, unexplained variation. Possible differences of means between the species then show up more clearly. *Note: Technically, reducing extraneous variation lowers MSE and hence increases the F statistic, which means there is more power to reject H_0 if it is false.*

   (b) Putting the skinks in the tank together could create dependence in the data. For ANOVA we need to assume independence. *For example, two skinks may prefer the same temperature, but only one can occupy that position in the tank.*

   (c) $n = 40$ and $p = 4$.
   The df in the `Treatments` row (i.e. Species) is $p - 1 = 3$.
   The df in the `Error` row is $n - p = 36$.
   The df in the `Total` row is $n - 1 = 39$.

   (d) **Report**

   Model Equation: $Y_{ij} = \mu_i + E_{ij}$

   where Y_{ij} is the j^{th} skink from species i), μ_i = population mean for group i, and E_{ij} is the error term.

   Assumptions:
   Independence and normality of data within species, plus constant variance – again within species.

   - **Check for Constant Variance.** Boxplots (p.3, Ass3 Questions) for the four species appear to have approximately equal vertical spread, supporting the equal variance assumption. This is supported by Levene's Test (p.3, Ass3 Questions), with $p = 0.2691 > 0.05$, supporting the null hypothesis of equal variances, H_0: $\sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \sigma_D^2$. The top left diagnostic graph (p.4, Ass3 Questions), Residual vs Predicted Value, also shows a level band across the page.
   - **Check for Normality.** Boxplots for the four species show approximate symmetry, supporting the normality assumption. The Q-Q plot (middle left diagnostic graph, Residual vs Quantile) shows a straight line, also indicating normality.
   - **Check for Independence.** Although separate trials helps with the independence, assumption we can't fully check. We'd need to know **all** the details of how the experiment was run. *e.g. How were skinks sampled for the trials?*

   Hypotheses:
   H_0: $\mu_A = \mu_B = \mu_C = \mu_D$ vs
   H_A: at least one population mean is different from at least one other.

ANOVA Table:

Source	df	Sum of Squares	Mean Square	F value	p-value
Species	3	52.075	17.358	3.06	0.0402
Error	36	203.900	5.664		
Total	39	255.975			

Note: this can be copied from the given SAS output in the case of a one-way design. For factorial designs, SAS presents results in two tables, which need amalgamating.

Statistical Conclusion:
At the 5% significance level we reject H_0, since $p = 0.0402 < 0.05$. At least one population mean preferred temperature is different from at least one other.

Interpretation:
We have statistically significant evidence of differences in average preferred temperature among the four different skink species.
Note: more detail could be obtained using a Tukey test.

2. **Nasal Sprays** [20 marks]

The ANOVA Model:
The (complete) model is that for one-way ANOVA,

$$Y_{ij} = \mu_i + E_{ij}$$

where
Y_{ij} is the j^{th} data point in the i^{th} Type (treatment, group),
μ_i = population mean for group i, and
E_{ij} is the error term.

Assumptions:
We assume the errors come independently from a $N(0, \sigma^2)$ distribution (or, equivalently, the data come independently from normal distributions with constant variance).

Checking assumptions:

Note: The boxplots may be used to check normality and constant variance. Levene's test checks for constant variance. The first diagnostic graph is used to check constant variance, and the second, centre left, checks for normality.

Normality? The boxplots (p.5, Ass3 Questions) show approximate symmetry, suggesting normality.

Note: However, Types A, C and D don't have the median halfway between the lower and upper quartiles, so possibly there isn't normality. It's very hard to tell with only five observations per group.

The Q-Q plot (middle left in the diagnostic graphs, p.7, Ass3 Questions) shows a straight line, suggesting normality of the residuals.

Note: Why the possible contradiction? The boxplots are looking for normality within each group, while the Q-Q plot pools the residuals. Apparent non-normality within a group can get subsumed by overall normality when considering all the groups.

Constant Variance?

The boxplots (p.5, Ass3 Questions) don't show very equal vertical spread, e.g. the spread for Type A is quite a bit less than the spread for Type B.

Note: The Types with low mean have smaller spread, and those with high mean have more spread. This suggests that a log transformation might be useful.

Levene's Test (p.6, Ass3 Questions) has a p-value of 0.2156, indicating at a 5% level we would not reject the null hypothesis of equal variances.

Note: Why the contradiction with the boxplots? Levene's test isn't infallible – for example a small data set can fail to reject H_0 simply because there isn't enough evidence. But then the boxplots aren't very good here either, as a boxplot of only 5 observations can't reliably show patterns in the data.

The Residual versus Predicted Value graph (first diagnostic plot, p.7, Ass3 Questions) shows a fairly level band, suggesting constant variance.

Independence?

We cannot confirm independence as we do not know how the people were selected or how the experiment was run. Assume independence.

ANOVA Hypotheses:

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (or equivalently, "All population means are equal")

H_A: At least one population mean is different from at least one other.

ANOVA Table:

The SAS ANOVA Table output is on page 5 of the Ass3 Questions.

A more informative ANOVA Table would change "Model" to "Type", as follows.

Source	df	Sum of Squares	Mean Square	F value	p-value
Type	4	1959.44	489.86	9.73	0.0002
Error	20	1006.40	50.32		
Total	24	2965.84			

Statistical Conclusion:

Since $p = 0.0002 < 0.05$, we reject H_0 at the 5% level. There is evidence that at least one population mean is different.

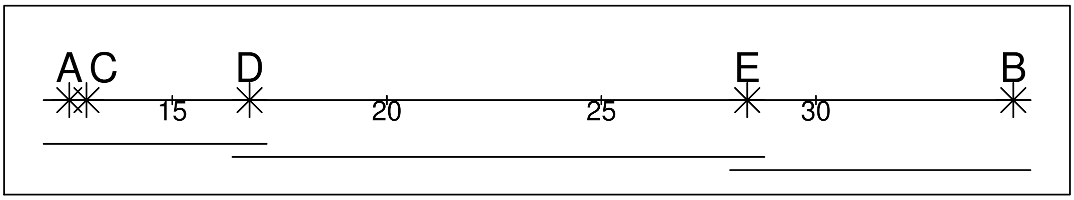
Tukey Test:

Page 6 of the Ass3 Questions shows the results of a Tukey Test, which compares the mean of every Type with every other Type (all possible pairwise tests, e.g. H_0: $\mu_1 = \mu_2$ versus H_A: $\mu_1 \neq \mu_2$). At an experimentwise significance level of 5%:
pairs of Types which differ are {(B,D), (B,C), (B,A), (E,C), (E,A)}, and
pairs of Types which are not significantly different are {(B,E), (E,D), (D,C), (D,A), (C,A)}.

This is shown on the underlining diagram (next page).

Interpretation:

The five different types of nasal spray show some difference in mean improvement in airflow for people with nasal congestion.



Recommendation:

Type B has the highest mean, but as Type E does not differ from it significantly (shown by the Tukey Test), the recommendation is to use either Type B or Type E.

Kruskal-Wallis Test:

If we decided that the assumptions of the ANOVA were not satisfied, we could use the Kruskal-Wallis Test, with results given on page 7 of the Ass3 Questions.

Since $p = 0.0032 < 0.05$, the null hypothesis is rejected, and we accept the alternative that at least one Type has a different distribution from at least one other Type.

The interpretation is that when measuring improvement in airflow, at least one type of nasal spray has a different distribution from at least one other type.

Notes:

The Kruskal-Wallis is the better test to use if there are serious doubts about the validity of the ANOVA. It only assumes independent data from continuous distributions.

It is a test of whether the data from different Types have the same distribution (H_0), versus an alternative that at least one Type has a different distribution.

If we can assume similar shaped distributions, it becomes a test of equal medians,
H_0*:* $\eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5$ *(or, equivalently, "All population medians are equal") vs*
H_A*: At least one population median is different from at least one other.*

3. **Forensic dental X-rays** [20 marks]

   (a) Data must be independent. Any two teeth from the same person are likely to have positively correlated X-ray spectropenetration gradient.

   (b) With n fixed at 16, the advantage of balance (equal numbers, 8, in each group) is that the power of the hypothesis test to detect a difference of means is maximised. Also, the precision of a confidence interval for $\mu_1 - \mu_2$ is maximised, i.e. the confidence interval would be as narrow as possible.

   (c) **Report**

   This is a one-way analysis of variance, with response variable Y = "X-ray spectropenetration gradient" of teeth, and factor "Gender" at two levels, `Female` and `Male`.

   The ANOVA Model:
   The (complete) model is that for one-way ANOVA, $Y_{ij} = \mu_i + E_{ij}$ where
   Y_{ij} is the j^{th} data point in the i^{th} Gender (treatment, group),
   μ_i = population mean for group i, and
   E_{ij} is the error term.

Assumptions:

We assume data come independently from groups, with group i having a normal distribution with mean μ_i and variance σ^2, where the variance is the same over all groups (i.e. σ^2 rather than σ_i^2).

Checking assumptions:

Normality? The boxplots (p.9, Ass3 Questions) show approximate symmetry, suggesting normality. The Q-Q plot (middle left in the diagnostic graphs, p.10, Ass3 Questions) shows a straight line, suggesting normality of the residuals.

Constant Variance?

The boxplots show approximately equal vertical spread, supporting the assumption of constant variance. Levene's Test (p.10, Ass3 Questions) has a p-value of 0.9305, indicating at a 5% level we would not reject the null hypothesis of equal variances. The Residual versus Predicted Value graph (first diagnostic plot, p.10, Ass3 Questions) shows the residuals from the two groups have approximately the same spread, again confirming constant variance.

Independence?

We cannot confirm independence as we do not know how the subjects were selected or how the experiment was run. Assume independence.

Hypotheses:

H_0: $\mu_1 = \mu_2$ (there is no difference of population means)
H_A: At least one is different (OR, since there are only two groups, $\mu_1 \neq \mu_2$).

ANOVA Table of X-ray spectropenetration gradient of teeth:

Source	df	Sum of Squares	Mean Square	F value	p-value
Gender	1	3.3306	3.3306	5.88	0.0294
Error	14	7.9238	0.5660		
Total	15	11.2544			

Statistical Conclusion:

Since the p-value is 0.0294, which is less than 0.05, we have enough evidence to reject H_0 at the 5% significance level, i.e. a significant difference of means has been found.

Interpretation:

We have found a significant difference of mean X-ray spectropenetration gradient between teeth from females and males. Since the sample mean is greater for males than females (5.4250 vs. 4.5125) it seems there is less penetration by X-rays in teeth from males. This suggests the technique could be useful in forensic analysis.

(d) The Tukey test is used to control the experiment-wise error rate when conducting *post hoc* tests in exploratory research. Since there are only two Gender groups here, only one comparison can be made in this case (male vs. female). So there is no need to make any adjustment for multiple comparisons.

4. **Personality Types** [20 marks]

(a) This is a random effects design because the four personality types investigated are a random sample from a large number of personality types.
Note: The investigator assumes there is some effect from different personality types, and wants to find how important it is, by estimating the amount of variability from that source.

(b) **Report**

The model: $Y_{ij} = \mu + A_i + E_{ij}$
where Y, A and E are random variables and μ is a fixed parameter, the overall mean.
Y_{ij} is the j^{th} person in the i^{th} personality type,
A_i is the random effect of personality type i and E_{ij} is the error term.

Assumptions, graphs and comments:

- The assumptions are that the errors E_{ij} come independently from a $N(0, \sigma^2)$ distribution, the A_i values are independently distributed as $N(0, \sigma_A^2)$, and the A_i are independent of the E_{ij} values.
- The assumptions of normality and constant variance appear to be satisfied since:
  - the boxplots show some symmetry, confirming a possible normal distribution within each group,
  - the quantile plot of residuals is reasonably close to a straight line, supporting the assumption of a normal distribution of errors.
  - the boxplots show approximately equal vertical spread, supporting the constant variance assumption,
  - Levene's test has $p = 0.6926$, indicating that the assumption of equal variance is satisfied,
  - the plot of Residual versus Predicted Value shows a level band across the graph, supporting the constant variance assumption.
- Independence: We cannot fully check for independence of the data. Random sampling of subjects from within each personality type helps preserve independence, but we must assume that the experiment was run in such a way as to retain independence.

ANOVA Table: (as far as the mean square column)

Source	df	Sum of Squares	Mean Square
Personality Type	3	279.675	93.225
Error	36	1506.700	41.853
Total	39	1786.375	

Components of variance:
Using P = personality type,

$$\hat{\sigma}^2 = MSE = 41.853, \quad \hat{\sigma}_P^2 = \frac{MSP - MSE}{r} = \frac{93.225 - 41.853}{10} = 5.1372.$$

Total variance = 41.853 + 5.1372 = 46.9902.

Percentage due to personality type: $\frac{5.1372}{46.9902} \times 100 = 10.9\%$.

Percentage of variance unexplained = 89.1%.

Interpretation:

The personality type is not very important in determining the score on the test, as it accounts for only 10.9% of the total variation in the score.

Note: Whether it is seen as important depends on the context. It is possible that for this situation in psychology, a value of 10.9% is seen as useful and important. If someone stated this they still got the marks. A judgment call was needed here.

5. **Phytoremediation** [20 marks]

(a) **Design and model equation:** This is a 4×2 (or 2×4) factorial design.

Note: One factor is Species (plant name) at 4 levels (Lettuce, Martin red fescue, Alpine pennycress and Bladder campion), the other is pH at two levels (5.5 and 7).

The model equation for the k^{th} response (zinc uptake) with the first factor at level i and the second at level j is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}.$$

There are also constraints on the parameters:
$\sum \alpha_i = 0$, $\sum \beta_j = 0$, for each i, $\sum_j (\alpha\beta)_{ij} = 0$, for each j, $\sum_i (\alpha\beta)_{ij} = 0$.

(b) **Analysis of the data:**

Using the raw data (p.15, Ass3 Questions), the plot of residuals versus predicted values shows clear funnelling, indicating variance increasing with mean. This suggests a log transformation of the response variable, Zinc, would be appropriate to stabilise variances. The quantile plot of the residuals shows slight deviations from a straight line, indicating possible non-normality.

Using log(Zinc) (p.16, Ass3 Questions), the plot of residuals versus predicted values shows a much more level band. The assumption of constant variance is supported for the log(Zinc) data. The quantile plot of the residuals is nearer to a straight line than for the raw data, indicating better conformance with the assumption of normality.

For these reasons, it is better to work with the logarithm of zinc uptake.

Note that pH was chosen as the first factor, and Plant as the second. This was done so that the interaction graph, discussed below, would have the quantitative factor, pH, on the x axis.

(c) **Report**

The model was given above, in part (a).

Assumptions, diagnostic graphs and comments:

We assume the errors are independent and normally distributed, with constant variance. Alternatively, $E_{ijk} \sim N(0, \sigma^2)$, independently.

The logarithm of zinc uptake was used as the dependent variable in the further SAS output (pp.17 to 19, Ass3 Questions). For comments on the diagnostic graphs (p.16, Ass3 Questions) see part (b).

ANOVA Table, log(Zinc) data:

Source	df	Sum of Squares	Mean Square	F value	p-value
pH	1	1.469	1.469	39.68	<0.0001
Plant	3	21.658	7.219	194.97	<0.0001
pH × Plant	3	0.903	0.301	8.13	0.0016
Error	16	0.592	0.037		
Total	23	24.623			

Note: (i) The table above used natural log. If $\log_{10}$ is used, the SS and MS columns are different but the F statistics and p-values are the same. Any log transform can be used to stabilize variances. (ii) SAS splits the ANOVA Table into two parts, which should be amalgamated as above for presentation of the results.

Hypotheses:

We start with the interaction test.

H_0: There is no interaction (i.e. all $(\alpha\beta)_{ij} = 0$)

versus H_A: There is interaction (at least one $(\alpha\beta)_{ij} \neq 0$).

Statistical conclusion:

The null hypothesis is rejected at either the 5% or the 1% significance level; p-value $= 0.0016 < 0.01 < 0.05$ (see p.17, Ass3 Questions).

Significant interaction has been detected.

Since there is significant interaction, we do not proceed to any main effects tests. They would be meaningless.

Interpretation and comments on interaction graph:

Different species have different patterns of change in log(Zinc uptake) as pH is changed.

The interaction graph is on page 18 of Ass3 Questions.

The four species show significantly different patterns of zinc uptake. Lettuce has low zinc uptake in **both** acid and neutral soil. The other three species all show higher zinc uptake, with **better uptake in acid soil than in neutral soil**.

There is a choice of which factor to choose for the x-axis and which factor forms the trace lines. Since pH is numerical, it is best on the x axis. For comparison, the alternative interaction graph is on page 19, Ass3 Questions.

Extra comments on Question 5:

- *The best species for high uptake overall appears to be Alpine pennycress, although we didn't specifically test this. We can't do a Tukey test (multiple comparisons) over the four species, as there is interaction present, and the Tukey test would average over the other factor – a meaningless procedure in the presence of interaction. However, we could restrict the data set to acid soil only and then do a Tukey test on the log(data). This would indicate if Alpine pennycress differs from the other three in acid soil.*
- *Another way of explaining the interaction is to say that we can't talk about the effect of one factor on the response variable without specifying the level of the other factor. In this example, we can't talk about the effect of soil pH on zinc uptake unless we say which plant species we are considering. There are different patterns. With lettuce, a more acid soil doesn't increase uptake, but with the other three species it does.*