Assignment 3

STAT 292 Applied Statistics 2A

24/05/2020

1. Skink Temperatures

(a)

In this experiment we're evaluating whether or not the different species of skinks have differences in their preferred temperatures. It's important to keep nuisance factors such as the time of day, light, amount of food recently eaten etc constant. If we don't, we run the risk of introducing variations from sources that are not of interest to us and therefore are part of the error and noise aspect of the analysis. Validity of the experiment can be increased by reducing the possibility that observed effects are due to a cofounding factor. By keeping those extraneous variables consistent between trials, a stronger case could be made of the effect the explanatory variable has on the response variable, addressing the criticism of bias.

(b)

The skinks are evaluated individually and not tested together in the same tank to mitigate the potential of welcoming external factors. These factors may arise from inherent variables that are difficult to control, for example: the possibility that a skink determines the position where they settle based on their peers. Our assumption of independence is then broken.

(c)

$p = 4$

$n = \sum_{i=1}^{p} n_i = 10 \times 4 = 40$

$df_{Treatments} = p - 1 = 4 - 1 = 3$

$df_{Error} = n - p = 40 - 4 = 36$

$df_{Total} = n - 1 = 40 - 1 = 39$

(d)

Model equation: $Y_{ij} = \mu_i + E_{ij}$ for observation j under treament i

Assumptions: Some key assumptions that must be satisfied before performing ANOVA are:

- **Independence of observations between and within groups:** This is impossible to check. Assumptions of independence are validated from the design of the experiment and the process in which the experimental units were selected. It is assumed that simple random sampling was done to ensure independence. The skinks were also tested independently to maintain independence. If independence is not preserved then the results from our ANOVA will not be reliable.
- **Constant variance across groups:** We can check equal variances by plotting the residuals versus the predicted values. If our plot of residuals increases or decreases as our predicted values get larger then we conclude that our assumption of constant variance is violated. What we can see from the residuals plot is that there is a level band and no funneling which does support the assumption of constant variance. Looking at the data, in particular the standard deviations for each group, it appears to be similiar which means similiar variances. Performing a Levene's test for equality of variances can also support this assumption. From the test, we get a p-value of 0.2691 thus with a confidence interval of 0.05 we don't reject the null hypothesis that there is constant variance, supporting our assumption. Boxplots are also a way to examine variances, the key to doing this is to a.) identify outliers and b.) identify huge differences in spread. There is no outliers identified in the boxlots and to compare spreads

we look to the IQR. If the IQR of the biggest box is twice as big as the IQR of the smallest box then there's sufficient evidence that the assumption is violated. Our boxplots are all fairly similiar so there isn't enough evidence to say the variances are not equal.

- **Nearly normal data within groups:** The most direct test of normality is to plot the residuals in a QQ-plot. A QQ plot plots a quantile of the distribution with another quantile and if the points conform to the straight diagonal-line then we can conclude that the data follows that distribution. THe QQ plot follows the line almost perfectly with slight deviations at the tails, from this we can conclude that our data follows a normal distribution. A histogram of residuals can also give us an idea of how normal our data is, our assumption is violated if there's skewness in our residuals. The histogram shows no skewness and is symmetric. Our boxplots show a slight lack of symmetry, with the only group that's symmetric is B but this can be chalked up to small sample sizes within groups ($n_i = 10$). If there are any departures from normality they would be very small and ANOVA is relatively robust to small departures.

Null and alternative hypothesis:

$$\mathcal{H}_0 \ : \ \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$\mathcal{H}_A \ : \ At\ least\ one\ population\ mean\ is\ different\ from\ another$$

ANOVA Table (if relevant), p-value:

One-Way Analysis of Variance

Dependent Variable: Temperature

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	52.0750000	17.3583333	3.06	0.0402
Error	36	203.9000000	5.6638889		
Corrected Total	39	255.9750000			

Statistical conclusions: With a p-value of 0.0402, it is smaller than our confidence interval $\alpha = 0.05$ thus we reject $\mathcal{H}_0$ at the 5% significance level in favour of $\mathcal{H}_A$.

Interpretation: The results of ANOVA tells us that at least one group mean deviates from the population mean. In other words at least one species of skink has a different preferred temperature from the other groups.

2. Nasal Sprays

Model equation: $Y_{ij} = \mu_i + E_{ij}$ for observation j under treatment i

Assumptions: Some key assumptions that must be satisfied before performing ANOVA are:

- **Independence of observations between and within groups:** Without any knowledge of how the experiment was designed, we can't say for certain that the experimental units are independent. We're assuming that simple random sampling was done.
- **Constant variance across groups:** From our residuals plot, there's no funneling and it appears to be levelled so our equal variance assumption holds. However looking at our standard deviations for each group, there seems to be quite a dissonance which suggest the variances aren't equal. Our boxplots also paints a picture of inequal variance, but since the sample sizes in each group is 5 and boxplots highlight 5 points of interest, the boplots are rather useless to us. THe Levene's test for equal variances gives us a p-value of 0.2156 and for a significance level of $\alpha = 0.05$ that means we don't reject the notion that the groups have equal variances.
- **Nearly normal data within groups:** THe QQ plot supports our normality assumption, the points neatly fall into that diagonal-line with very little deviations. THe histogram fits the normal distribution

curve well which also support our assumption. WE could look at our boxplots to try and further support our assumptions but the sample size is too small for the boxplots to have any reliable information.

ANOVA hypothesis:

$$\mathcal{H}_0 \ : \ \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$\mathcal{H}_A \ : \ \textit{At least one population mean is different from another}$$

ANOVA Table (if relevant), p-value:

Dependent Variable: Improvement

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F*
Model	4	1959.440000	489.860000	9.73	0.0002
Error	20	1006.400000	50.320000		
Corrected Total	24	2965.840000			

ANOVA conclusions: With an incredibly low p-value of 0.0002, we reject $\mathcal{H}_0$ at the 5% significance level in favour of $\mathcal{H}_A$.

Kruskal-Wallis hypothesis:

$$\mathcal{H}_0 \ : \ \eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5$$

$$\mathcal{H}_A \ : \ \textit{At least one population median is different from another}$$

Kruskal-Wallis assumptions: A Kruskal-Wallis test is a non-parametric test so it doesn't depend on assumptions of equal variance or normality. The only assumption it requires is independence and a continuous distribution.

Kruskal-Wallis conclusion: The p-value 0.0032 is much lower than our significance level $\alpha = 0.05$ so we reject $\mathcal{H}_0$ concluding that there is a difference between group medians.

Tukey assumptions: Are the same as the ANOVA assumptions, Tukey tests are robust to deviations from the ANOVA assumptions when the design is balanced.

Tukey hypothesis:

$\mathcal{H}_0 : \mu_1 = \mu_2$ versus $\mathcal{H}_A : \mu_1 \neq \mu_2$

$\mathcal{H}_0 : \mu_1 = \mu_3$ versus $\mathcal{H}_A : \mu_1 \neq \mu_3$

$\mathcal{H}_0 : \mu_1 = \mu_4$ versus $\mathcal{H}_A : \mu_1 \neq \mu_4$

$\mathcal{H}_0 : \mu_1 = \mu_5$ versus $\mathcal{H}_A : \mu_1 \neq \mu_5$

$\mathcal{H}_0 : \mu_2 = \mu_3$ versus $\mathcal{H}_A : \mu_2 \neq \mu_3$

$\mathcal{H}_0 : \mu_2 = \mu_4$ versus $\mathcal{H}_A : \mu_2 \neq \mu_4$

$\mathcal{H}_0 : \mu_2 = \mu_5$ versus $\mathcal{H}_A : \mu_2 \neq \mu_5$

$\mathcal{H}_0 : \mu_3 = \mu_4$ versus $\mathcal{H}_A : \mu_3 \neq \mu_4$

$\mathcal{H}_0 : \mu_3 = \mu_5$ versus $\mathcal{H}_A : \mu_3 \neq \mu_5$

$\mathcal{H}_0 : \mu_4 = \mu_5$ versus $\mathcal{H}_A : \mu_4 \neq \mu_5$

Tukey conclusion: Under the 5% significance level, {B, E}, {E, D}, {D, C, A} are the groups of similiar means. THere are significant difference of means between B and D, C, A. Another significant difference of means is found between E and C, A.

Interpretation: Performing an ANOVA tells us there is a significant difference of means at the 5% significance level. THere isn't enough evidence to invalidate our results due to no explicit violations of ANOVA assumptions. In the case where there might be violations on assumptions a Kruskal-Wallis test can

be done since it's not dependent on ANOVA assumptions. Even in this test the group medians appear to be significantly different as well. After establishing that at least one group mean/median is significantly different to the others, we do a post-hoc test to find out how different each group is compared to each other. Performing a Tukey test shows that the means of group B and group E are significantly different to groups C and A, B is also significantly different to D. B and E were assigned the letter A which means that those two have significantly the greatest mean. So recommending these nasal sprays to someone, B and E should be the ones recommended with B possibly being better than E.

3. Forensic dental X-rays

(a)

Only taking samples from the same person for both groups will not be representative of the population. Teeth vary from person to person so only taking it from one person will just yield similiar samples.

(b)

To give a balanced design. A balanced design means that we don't spend nore resources studying one group over the other giving us more power and precision. Any analysis would be unreliable because one group would be underrepresented.

(c)

Model equation: $Y_{ij} = \mu_i + E_{ij}$ for observation j under treatment i

Assumptions:

- **Independence of observations between and within groups:** Taking samples from 16 different people helps us to ensure independence. There is no information about the people being sampled so there's a possibility that a person being sampled is related to another person genetically which might introduce violations of indepence. For the sake of this test we'll assume that the samples truly are independent.
- **Constant variance across groups:** THe boxplots for female and male have very similiar spreads which supports our assumption. Further supported by the standard deviations which are also similiar. THe residuals plot doesn't have any hint towards a lack of equal variance, however there isn't a lot of residuals so the reliability of this plot is questionable. OUr Levene's test is very high with a near 100% probability that our data has equal variance between groups so with a 5% confidence interval $\alpha = 0.05$ we do not reject our hypothesis that our data has equal variance.
- **Nearly normal data within groups:** THe boxplots are symmetric which strongly indicates that our data is normal. The histogram is another way of validating the normality assumption, in our histogram there isn't any noticeable outliers and the data is somewhat symmetric. THe best indication of normality of errors would be the QQ plot and with our QQ plot we can see that the points fit the line perfectly so we can conclude normality.

Null and alternative hypothesis:

$$\mathcal{H}_0 \ : \ \mu_1 = \mu_2$$
$$\mathcal{H}_1 \ : \ \mu_1 \neq \mu_2$$

ANOVA Table (if relevant), p-value:

Dependent Variable: Xray_grad

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F*
Model	1	3.33062500	3.33062500	5.88	0.0294
Error	14	7.92375000	0.56598214		
Corrected Total	15	11.25437500			

Statistical conclusions: With a p-value of 0.0294 at a significance level of 5% ($\alpha = 0.05$), we reject the null hypothesis that the means are equal in favour of our alternate hypothesis where the means are not equal. However we don't reject the null hypothesis under the 1% ($\alpha = 0.01$) significance level due to not having enough evidence to reject the null hypothesis in such a precise test.

Interpretation: Our interpretation is dependent on our choice of significance level. Under the 5% cl: the spectropenetration grade of Males are different to Females. In other words the mean for Males compared to mean for Females are probably from different population means. Under the 1% cl: there isn't sufficient evidence that the means differ.

(d)

Because a Tukey test compares each group means with every other group mean. Since there's only two groups in this experiment, there's no need to use a Tucky test because we're already comparing the groups with every other group.

4. Personality types

(a)

Because the deisgn isn't only taking random samples of the subjects but also a random sample of the possible treatments, the treatments then become some random subset of the set of personality types. Once the treatments are randomly selected then randomly selected subjects are assigned to each treatment.

(b)

Model equation: $Y_{ij} = \mu + A_i + E_{ij}$ for observation j under treatment i, where Y_{ij} is the personality test score j with personality type i, μ is the mean overall, A_i is the random effect of personality type i and E_{ij} are the errors for personality test score j with personality type i.

Assumptions:

The assumptions of a random effects one-way ANOVA is slightly different to a fixed effects one-way ANOVA.

- **Independence and normal distribution of the errors** $E_{ij} \sim iidN(0, \sigma^2)$
- **Independence and normal distribution of the random effects** $A_i \sim iidN(0, \sigma_A^2)$
- **Independence between the random effects and the errors**
- The boxplot shows similiar spread supporting constant variance but the data aren't symmetric which might suggest non-normality. However there aren't any outliers.
- The Levene's test gives us a p-value of 0.6926 so we don't reject $\mathcal{H}_0$ thus the test supports constant variance.
- The residuals plot shows a constant, level band with no funneling or patterns. This also supports our assumption of constant variance.
- The QQ plot displays normality with slight deviations at the tails.
- The histogram shows no noticeable skewness or outliers, also supports our normality assumption.
- The assumption that our random effects and our errors are independent is validated by the random sampling of subjects within a random sample of personality types. The personality type is randomly sampled and within each of those personality types 10 people were randomly sampled.

Random effects analysis:

- **The two components of variance**

$\hat{\sigma}^2 = MSE = 41.852778$

$\hat{\sigma}_A^2 = \frac{MST - MSE}{r} = \frac{93.225000 - 41.852778}{10} = 5.1372222$

- **The total variance**

$\hat{\sigma}^2 + \hat{\sigma}_A^2 = 41.852778 + 5.1372222 = 46.9900002$

- **The percentage of variability in the data due to individual personality type differences and the percentage due to unexplained variance**

$\frac{\hat{\sigma}^2}{\hat{\sigma}_A^2+\hat{\sigma}^2} = \frac{41.852778}{46.9900002} = 0.89067413964$

$\frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2+\hat{\sigma}^2} = \frac{5.1372222}{46.9900002} = 0.10932586035$

Percentage unexplained = 89%

Percentage due to personality type = 11%

Statistical conclusions: 11% of the variation in personality test scores is due to the personality type, whilst the other 89% are due to other unexplained factors.

Interpretation: The conclusion we draw from this type of test is subjective, an 11% variation due to personality type is small for this context. That doesn't tell us there's a big influence of personality types on personality test scores.

5. Phytoremediation

(a)

A 2×4 complete factorial design, first factor is the soil pH (5.5, 7.0) and the second are the four types of plants (lettuce, martin red fescue, alpine pennycress, bladder campion). It's complete because every factor is evaluated by every other factor.

(b)

Using a log is more appropriate because the residuals plot shows funneling. Funneling is an indication that there aren't equal variances between groups. To remediate this we take logs of the residuals. The logs residual plot has shown us that taking the logs fixes our funneling problem.

(c)

Model equation: $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}$ where Y_{ijk} is the response variable, μ is the overall mean, α_i is the effect of factor A, β_j is the effect of factor B, $(\alpha\beta)_{ij}$ is the interaction between factor A and factor B, and E_{ijk} is the error.

Assumptions:

- **Independence of observations between and within groups:**

We don't know how the experiment was designed so we cannot comment for certain on the validity of independence. We have no knowledge of how the plants were collected.

- **Constant variance across groups:**

Constant variance is violated as there is funneling in our residuals plot, one way to remedy this is to take the logs of the residuals. The log(residuals) plot has no funneling and has a constant band which shows us that there's constant variance after taking the logs.

- **Nearly normal data within groups:**

The QQ plot has deviations from the line, it might be normal enough but we need more information. The histogram doesn't look skewed but there's a huge spike in the middle so it's difficult to see relative differences. Looking at the log versions gives a clearer picture, the QQ plot fits the line a lot better and the histogram is a lot more descriptive. There aren't any indication of non-normality.

Interaction test:

$$\mathcal{H}_0 : \textit{There is no interaction } (\textit{all } (\alpha\beta)_{ij} = 0)$$

$$\mathcal{H}_1 : \textit{There is interaction } (\textit{at least one } (\alpha\beta)_{ij} \neq 0)$$

ANOVA Table (if relevant), p-value:

Source	DF	Type I SS	Mean Square	F Value	Pr > F
pH	1	1.46932364	1.46932364	39.68	<.0001
Plant	3	21.65807393	7.21935798	194.97	<.0001
pH*Plant	3	0.90287433	0.30095811	8.13	0.0016

Statistical conclusions: Our p-value is very small so we reject $\mathcal{H}_0$ indicating there is an interaction between factors.

Interpretation: From the test we conclude that there is sufficient evidence to suggest that there is interaction between plant species and soil pH in the effect of zinc uptake. We can further support this via the interation graphs, since the lines are not parralell there's evidence of interaction. The interaction plot also shows plant species growing in acidic soil takes up more zinc, however lettuce has greater zinc uptake in neutral soil.