School of Mathematics and Statistics

Te Kura Mātai Tatauranga

STAT 292 **Assignment 2: Due Thursday, 6 April 2023 at 11:59 PM**

Note: Your assignment can be typed or handwritten (and scanned). Be sure to submit your assignment as a PDF and follow the instructions specified on the course Nuku page. Where calculations are performed in R, you must include relevant code and output with your answer to receive credit.

Assignments that are submitted late will receive a mark of 0 unless illness, bereavement or other substantial causes occur and have been discussed with the course coordinator.

1. (24 marks)

   Ecotourism researchers investigating attitudes toward forest vegetation and land use collected data on $n = 602$ randomly selected tourists to a national park. Tourists were asked a series of questions about their beliefs and attitudes, and they were also asked questions to collect demographic information. In one question, respondents were asked their level of agreement with the statement "Forests are valuable because they produce timber, jobs, and income for people." Respondents reported their level of agreement according to a seven point Likert scale (1 = "Strongly disagree", 2 = "Disagree", 3 = "Slightly disagree", 4 = "Neutral", 5 = "Slightly agree", 6 = "Agree", 7 = "Strongly agree"). The table below presents counts of the numbers of adults in various age ranges (rows) who specified various levels of agreement with the statement (columns). It is of interest to determine if level of agreement with the statement and age of the respondent are independent.

	1	2	3	4	5	6	7
< 40	34	39	25	12	8	2	2
40-60	52	78	29	16	19	7	5
> 60	54	84	41	20	32	22	21

   a. Suppose that it is of interest to use a chi-square test of independence to test the hypothesis that there is an association between level of agreement with the statement and age of the respondent. Thinking about the conditions of a chi-square test of independence, is such a test appropriate for these data? Justify your answer. (3 marks)
   b. Regardless of your answer to part (a), carry out both Pearson and likelihood ratio chi-square tests of independence, being sure to
      - clearly state the null and alternative hypotheses,
      - present the test statistic and its distribution under the null hypothesis, and
      - report the p-value and your conclusion at the $\alpha = 0.05$ significance level.

      (7 marks)

Now consider the following table which reduces ages to two categories (≤ 60 and > 60) and simply considers whether there is some form of agreement with the statement (values of 5 to 7) or not (values of 1 to 4).

	Disagree or Neutral	Agree
≤ 60	285	43
> 60	199	75

c. Using this contingency table, describe and clearly interpret the association between whether the adult is at least 60 years old and whether there is some form of agreement with the statement using the odds ratio θ (to at least 3dp). (4 marks)

d. Obtain a 95% confidence interval (to at least 3dp) for the odds ratio θ. (6 marks)

e. Test the hypotheses

$$\begin{aligned} \mathcal{H}_0 : \theta &= 1 \\ \mathcal{H}_1 : \theta &> 1 \end{aligned}$$

at the $\alpha = 0.05$ significance level. Be sure to report the type of test you are carrying out, the p-value of the test, and provide a clear interpretation of what the results mean in the context of the data it is applied to. (4 marks)

2. (20 marks)

Children of different ages are timed doing a manipulative task which requires hand-eye coordination. The data are given below.

Age (yrs)	Time (sec.)
3	22, 31, 28, 33
5	18, 18, 12, 25
7	9, 7, 15, 14
9	11, 13, 8, 12

a. Do an ANOVA on these data and a Kruskal-Wallis test, using `R`. Use a 5% significance level for the tests. Check assumptions by including boxplots, a Levene's test and diagnostic graphs.
b. Write up the results as a brief research report, following the Assignment Guidelines given below. Include a statement of the (complete) model equation.

3. (20 marks)

   Insects were reared in a laboratory at various temperatures. Their development times (in days) are given below. Of particular interest to the researcher is whether temperatures 16°C and 20°C have any difference in mean development time.

Temperature (°C)	Development time (days)
12	72, 90, 97, 83, 86, 71
16	75, 93, 80, 70, 63, 75, 69
20	61, 72, 67, 75, 56, 93, 75
24	55, 65, 49, 63, 72, 67

   a. Analyse the experiment (*i.e.*, do the overall test, the ANOVA) using `R`. Be sure to check the ANOVA assumptions.
   b. Test for significant differences between the 16°C and 20°C groups using a Tukey test.
   c. Write up the results as a brief research report, following the Assignment Guidelines below.

4. (20 marks)

   The extent to which X-rays can penetrate tooth enamel has been suggested as a suitable mechanism for differentiating between females and males in forensic medicine. The table below gives *spectropenetration gradients* for one tooth from each of eight females and eight males.

Gender	Spectropenetration gradient
Female	4.7, 5.5, 3.6, 4.2, 5.5, 4.1, 3.5, 4.9
Male	4.8, 5.4, 5.1, 5.3, 5.7, 6.5, 6.4, 4.4

   Note that a high reading reflects a fast drop-off in X-ray penetration, and so less penetration by the X-rays.

   a. Explain why the teeth have been sampled from eight different people of each sex, and not eight teeth from one female and eight from one male.
   b. Given that the researcher could afford to test $n = 16$ subjects, explain the advantages of choosing eight from each group.
   c. Do an ANOVA on these data, using `R`. Use a 5% significance level for any tests that you include. Write up the report using the Assignment Guidelines below.
   d. Explain why there is no point doing a Tukey test with these data.

Assignment Guidelines

The following Assignment Guidelines are relevant for all the assignments in Parts 2 and 3 of the course.

When you do a statistical test of a particular hypothesis, it is assumed you will state the following, **if relevant**:

- Model equation.
- Assumptions about the data, and comments about whether diagnostic graphs support those assumptions.
- Null and alternative hypotheses.
- ANOVA Table (if relevant) and p-value.
- Statistical conclusion(s) (*e.g.*, "We reject $\mathcal{H}_0$ and conclude $\mathcal{H}_1$, that μ_1 and μ_2 differ at the 5% significance level").
- Interpretation of the statistical conclusions back to the original problem, using the original meaning of the response variable and any factors or covariates. For example, if comparing heights of two groups, "Female and male adults have different mean heights, with males being taller on average".