Biostats VHM 801 Course
Winter 2026, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to Mid-term Exam, March 2026

The solution is more detailed than required for a 100% score, both by including both options for **e)** and by providing more detail than could reasonably be expected within the short time. Some specific questions related to interpretation were waived in the marking. The data are from Altman (1990), *Practical Statistics for Medical Research*, and their analysis is discussed extensively in the book.

Question 1

Sub-question a)

The description of the study does not indicate any treatments to have been applied, and if outcomes are merely observed for this group of patients, the study is observational. A comparison with another patient group is mentioned in **e)**, but these groups are clearly not a result of assignment of treatments or conditions.

The values of T_4 and T_8 cells from each patient are paired samples, both obtained from the same blood sample (and patient).

Sub-question b)

The descriptive statistics on original scale are of principal interest for describing the distributions. The dotplots additionally give an impression of distribution shapes. The description of the distributions should include the following points:

- Both distributions (T_4 and T_8 cell counts) are right-skewed, as seen by the positive and moderately large values of skewness, the means being larger than the medians, the medians being closer to Q1 than Q3, and also visually in the dotplots. The log-transformed values are much more symmetrically distributed.
- The highest value for T_4 might look like an outlier on original scale, but the log-transformation makes it no longer extreme relative to the distribution. Both this value and the largest value for T_8 are flagged as suspected outliers by the boxplot rule, but neither should be considered as real outliers.
- On original scale, the probability plots on original scale are curved, and the P-values for the A-D normality tests are close to 0.05 (but significant for T_8 only). On log-scale, the probability plots show no obvious systematic deviations from the line, and the normality tests have much larger and clearly non-significant P-values. These results show that a normal distribution is a reasonable approximation only on log-scale, for both T_4 and T_8 counts.

Sub-question c)

For analysis on original scale, denote by $X_1, \ldots, X_{20}$ the T_4 cell counts for the 20 patients. We assume this to be a simple random sample (or the $X_1, \ldots, X_{20}$ to be i.i.d.) from a normal distribution $\mathrm{N}(\mu, \sigma)$. For a 95% confidence interval we use $t^* = 2.093$ from a t-distribution with 19 df,

$$(T_4)\ 95\%\ \mathrm{CI}:\ \bar{X} \pm t^* s/\sqrt{n} = 823.2 \pm 2.093 \cdot 566.385/\sqrt{20} = 823.2 \pm 265.1 = (558.1, 1088.3) \approx (558, 1088).$$

We are 95% confident that the population mean T_4 count is within the limits. Because the T_4 distribution is not normal, as discussed above, the CI is approximate. According to the guidelines on use of t-distribution inference for non-normal data (slide 6L−9), the interval is acceptable (no strong skewness or outliers), but it is certainly not (statistically) exact.

Analysis on log-scale has the same assumptions and uses the same calculation.

$$(\ln T_4)\ \ 95\%\ \text{CI}:\ \ 6.4869 \pm 2.093 \cdot 0.7083/\sqrt{20} = 6.4869 \pm 0.3315 = (6.1554, 6.8184) \approx (6.155, 6.818).$$

The assumed normal distribution seems very reasonable from our descriptive analysis, so the CI can be considered as (statistically) exact. In order to make it an interval for T_4 cell counts, we can back-transform both the estimate and CI endpoints: $\exp(6.4869) = 656.5$ and $(\exp(6.1554), \exp(6.8184)) = (471.3, 914.5)$. These values are for the median (not mean) on original scale; hence we are 95% confident of the population median being included in this interval.

Note. A confidence interval is (statistically) exact if it has the correct coverage (e.g. 95%). Confidence intervals based on the t-distribution are exact when the model assumptions are met.

Sub-question d)

The T_4 and T_8 samples are paired, and analysis should use the pairwise differences. The question is whether to compute differences on original or log-scale. The Minitab listings do not include descriptive statistics or graphs for the differences. One may suspect that the differences of the counts will be affected by the non-normality of the T_4 and T_8 values, but it should be suggested to explore the distribution of the differences.

The differences constitute a single sample, so the statistical model and methods are the same as for **c)**. The last Minitab listing shows the results. The analyses of the differences on both scales give some (weak) evidence against the null hypothesis ($\text{H}_0 : \mu_D = 0$) against the two-sided alternative ($\text{H}_a : \mu_D \neq 0$)), but the P-values of 0.080 and 0.064 are not quite strong enough for formal significance at the 5% significance level. Because the P-values are still relatively low and close to 0.05, we should, despite the formally non-significant results, phrase our conclusion more carefully than just saying that there is no significance and insufficient evidence.

The CIs for differences show that the plausible values for μ_D are strongly on the side of T_4 being larger than T_8. The original scale CI could be interpreted as 95% confidence that the mean T_4 count is at most 28 units lower than the (population) mean T_8 count and up to 446 units larger. The log-scale CI could, upon backtransformation and using the shown formula, be interpreted as 95% confidence that the (population) median T_4 count is not less than 98.4% ($= \exp(-0.016) \cdot 100$) of the median T_8 count and up to 167.5% ($= \exp(0.516) \cdot 100$) of it.

Sub-question e)

Part i) For each of the 20 patients, we observe whether the T_4 cell count is higher than the T_8 count. This corresponds to a binomial setting, and we let $Y \sim B(20, p)$ denote the number of patients with $T_4 > T_8$. The binomial parameter p is the probability of a Hodgkin's patient to have their $T_4 > T_8$. From the data listing, we observe $Y = 12$, and hence $\hat{p} = 12/20 = 0.60$. For the confidence interval, we cannot use the classical (normal approximation) method because the observed count of events is less than 15. However, the "plus four" interval applies (because $n \geq 10$). We calculate $\tilde{p} = 14/24 = 0.583$ and $\text{SE}(\tilde{p}) = \sqrt{0.583(1-0.583)/24} = 0.101$, and finally

$$95\%\ \text{CI for}\ p\ :\ \tilde{p} \pm z^* \cdot \text{SE}(\tilde{p}) = 0.583 \pm 1.96 \cdot 0.101 = 0.583 \pm 0.198 = (0.385, 0.781).$$

We are 95% confident that the (population) probability lies within this interval. The observed probability of 1 (or 100%) in the non-Hodgkin's group is far beyond this interval, and even without

carrying out a statistical test one would expect to have evidence of a difference between the groups. The statistical method would be a 2-sample comparison for proportions and was not expected for the exam.

Part ii) The non-Hodgkin's patients can hardly be linked to the Hodgkin's patients in any way (and nothing in the description indicates that they would be), so this must be two independent samples. The natural statistical model is therefore two independent normally distributed samples. The distribution of T_4 values in the Hodgkin's group was not normal on original scale, but much better approximated by a normal distribution on log-scale. It can be suggested to investigate whether the situation is the same for the non-Hodgkin's group. If yes, analysis on log-scale seems most natural. The parameter of interest would therefore be the difference between the log-scale means in the two groups, and the analysis would include a two-sample t-test and a confidence interval based on the t-distribution. The alternative hypothesis for the t-test should be two-sided, in absence of any interest in a particular direction. We have no way of knowing whether the standard deviations could be the same in the two groups, so it seems most natural to not make that assumption. The confidence interval should be backtransformed to original scale, in a similar way as discussed in **d)**.

Alternatively, one could consider the non-parametric test Wilcoxon-Mann-Whitney test, and if the two distributions on original scale look reasonably similar, the parameter of interest would be the difference between the medians in the two groups. A confidence interval could then be obtained, in addition to the P-value for the WMW test. Without reasonably similar looking distributions on original scale, the interpretation of the test in terms of medians is more speculative, and one should focus only on the interpretation of the P-value as evidence against the two distributions being equal. Also here, regardless of whether hypotheses are phrased in terms of medians or the distributions themselves, the alternative hypothesis should be two-sided.