Biostatistics VHM 801 Course
Fall 2024, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to Mid-term Exam, October 2024

The solution is more detailed than required for a 100% score, by including discussion of multiple options for the answers and generally more detail than can be managed during an exam.

Question 1

The data are from from the article, Batchelor, J. R. & Hackett, M. (1970): HL-A matching in treatment of burned patients with skin allografts, *Lancet* **2**, 581–583. Some patients had multiple grafts of close or poor compatibility, but for simplicity only one graft of each type is included here, generally chosen so as to minimize the difference between the two graft types.

Subquestion a)

The study is an experiment because the grafts were assigned to the patients. It is not stated explicitly, but one would assume the two graft types were randomly allocated to the severely burned areas for each patient. The two survival/rejection times for the grafts applied to the same patient constitute paired samples.

Subquestion b)

The rejection times for poorly matched grafts constitute a single sample (SRS or i.i.d.) of a continuous distribution. The information provided allows us to assess some features of the distribution's shape.

- skewness: on original scale the distribution is somewhat right-skewed (skewness 0.88, mean (23.64) larger than median (19), asymmetric box (Q3−median (10) > median−Q1 (4)); on logarithmic scale the distribution is apparently still right-skewed but to a lesser degree (skewness = 0.23, mean (3.071) larger than median (2.944), asymmetric box (0.423 > 0.236)),
- suspected outliers: none on either scale; for example, on original scale the "box ± 1.5 IQR" rule does not indicate any outliers: $29 + 1.5 \cdot 14 = 50$ but the largest value is 43.

In summary, this analysis shows that the normal distribution will fit worse on original scale (due to the more pronounced right-skewness) than on log-transformed scale, but on neither scale it seems totally off.

Subquestion c)

Following our discussion in **b)**, the calculation is best carried out on log-scale ($\sim Y$); we use the sample mean ($\mu_Y = 3.071$) and standard deviation ($s_Y = 0.447$) to calculate the probability of values exceeding $\ln(30) = 3.401$:

$$\mathrm{P}(Y > 3.401) = \mathrm{P}\left(\frac{Y - 3.071}{0.447} > \frac{3.401 - 3.071}{0.447}\right) = \mathrm{P}(Z > 0.74) = \mathrm{P}(Z < -0.74) \approx 0.23,$$

using PSLS Table B. Alternatively, on original scale the probability would be calculated as $\mathrm{P}(Z > 0.58) \approx 0.28$, with the z-score calculated as $z = (30 - 23.64)/10.88 = 0.5846$.

Subquestion d)

We assume a binomial setting for the event that a skin graft survives for at least 30 days, and hence $X \sim \mathrm{B}(n, p)$, with $n = 11$. The two incomplete observations can be fully utilized because we know the grafts survived for at least 30 days. We observed $X = 5$, and hence $\hat{p} = 5/11 = 0.45$. For a 95% CI, we need to use the "plus four" method (which is valid with $n \geq 10$), and we get $\tilde{p} = 7/15 = 0.467$, and

$$95\%\ \text{CI:}\ \ \tfrac{7}{15} \pm 1.96\sqrt{\tfrac{(7/15)(8/15)}{15}} = 0.467 \pm 0.252 = (0.214, 0.719)\,.$$

Subquestion e)

As noted above, the 11 observations of graft rejection times of closely and poorly compatible grafts are *paired* observations, each being from the same patient. One possible analysis for paired data is for differences (D_i) of close and poor rejection times. The two incomplete observations will however not give the correct differences, thus introducing some bias into the analysis (see below). An alternative method, which utilizes the data fully and takes into account the incomplete data, is by a sign test. This is because even the incomplete rejection times are larger than their corresponding rejection times for the poorly compatible graft, so that the corresponding D_i's are > 0. The statistical assumption is that the differences $D_1, \ldots, D_{11}$ are independent and from the same (continuous) distribution.

Among the 11 differences, 1 is zero, 8 are positive (close > poor) and 2 are negative. Testing the median of the differences equal to zero corresponds to testing $\mathrm{H}_0 : p = 0.5$ in $\mathrm{B}(10, p)$. We choose a one-sided alternative $\mathrm{H}_a : p > 0.5$, because closely compatible grafts would be expected to survive longer. The P-value of our test is $\mathrm{P}(X \geq 8) = \mathrm{P}(X = 8) + \mathrm{P}(X = 9) + \mathrm{P}(X = 10) = 0.044 + 0.010 + 0.001 = 0.055$ (where $X \sim \mathrm{B}(10, 0.5)$), using Table 1 of the Stevens textbook. The P-value is just above the limit of significance at the 5% level. We conclude that there is an indication in the data that closely matched grafts survive longer, but not quite enough for statistical evidence at the 5% level. The result could be labelled as close to significant.

An analysis of differences should be done on logarithmic scale (the distribution of differences appears very right-skewed), and we may include the incomplete observations, noting that their effect is to make the difference between rejection times of closely and poorly matched grafts *smaller*. Excluding the incomplete observations will bias the results even further towards zero (no difference) because both observations are positive, so that is not preferable. We assume $D_i \sim \mathrm{N}(\mu_D, \sigma_D)$ for $i = 1, \ldots, 11$, and test $\mathrm{H}_0 : \mu_D = 0$ against $\mathrm{H}_a : \mu_D > 0$. This gives: $t = 0.397/(.500/\sqrt{11}) = 2.63$, corresponding to $P < 0.02$ (because $t_{.98}(10) = 2.359$ in Table C of PSLS). The analysis therefore gives statistical significance against equal mean rejection times and for a longer mean rejection time in grafts with close compatibility. The incomplete observations may have biased the difference towards smaller and hence most likely the test statistic towards *less* significance. We cannot say for sure because the larger differences will also affect the estimated standard deviation (s_D). Additionally, the validity of the assumed normal distribution may be problematic, but it is difficult to assess based on the small data set and with the two incomplete observations.