Biostats VHM 801 Course
Winter 2018, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to Midterm Exam, March 2018

The solution is more detailed than required for a 100% score, by including several options in different questions and including answers to all three parts of **e)**. Also the discussion throughout is more verbose than could be reasonably be managed within the time constraints of the exam. The data are from Mitchell et al. (1936), *J. Nutrition* **11**, 257−274.

Question 1

Subquestion a)

The study was an experiment because the rats were given feed of a type decided by the experimenter. The statistical design is two paired samples, where the sets formed by rats from the same litter, of the same sex and of similar weight corresponds to pairs, or blocks. Observations from different sets should (in the absence of information to the contrary) be considered as independent. On the other hand, the values obtained from the two rats in a set should be considered as dependent, where the dependence arises from the similarities among the two rats in the set. When rats have genetic similarity and similar weight, it could be expected that their nutritional responses will be similar.

Subquestion b)

Because of the paired samples, statistical inference to compare the two diets must be based on the differences within pairs. Let us denote by X_i and Y_i the digestibility measured for the rats in set i that were fed raw and roasted peanuts, respectively, for $i = 1, \ldots, 10$. Then we form the differences $D_i = X_i - Y_i$ (raw − roasted), and formulate the statistical model,

$$D_1, \ldots, D_{10} \text{ are a simple random sample (or i.i.d.) from } N(\mu, \sigma).$$

The mean $\mu = \mathrm{E}X_i - \mathrm{E}Y_i$ is the difference in mean digestibility of the two diets. With only 10 observations model checks are difficult, but a simple inspection of the data (or a display of these in a dotplot) shows no reasons for concern with respect the assumed normal distribution. The statistical analysis has the steps:

- *Estimation*: $\hat{\mu} = \bar{D} = 1.240$, $\hat{\sigma} = s = 1.363$,
- *Test* of H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$ (we have no reason to prefer a one-sided H_a):

  $$t = \bar{D}/(s/\sqrt{10}) = 1.240/(1.363/\sqrt{10}) = 2.88, \quad P = 2 \times P(t(9) > 2.88) = 0.018.$$

  The P-value is from the software listing, but we can get $P < 0.02$ from a table for the $t(9)$-distribution for t under H_0 ($t_{.99}(9) = 2.821$),
- *Confidence intervals* for μ: the 95% CI is given in the listing as (0.265, 2.215), and the 90% CI is computed from $t^*(9) = t_{.95}(9) = 1.833$ as (note also $t_{.975}(9) = 2.262$):

  $$\bar{D} \pm t^*(9)\, s/\sqrt{10} = 1.240 \pm 1.833 \cdot 1.363/\sqrt{10} = 1.240 \pm 0.790 = (0.450, 2.030).$$

We conclude that there is moderate significance (certainly evidence at the 0.05 significance level) against H_0, i.e. that the digestibility in raw and roasted peanuts differ; the estimates and confidence intervals show the raw peanuts to have the higher digestibility. A suggested non-technical wording could be: the data provide evidence that the digestibility of raw peanuts is higher than of roasted peanuts; we are 95% confident that the difference is between 0.265% and 2.215%.

Subquestion c)
To compare the difference in digestibility between female and male rats, we split the 10 differences $D_1, \ldots, D_{10}$ analysed above into two samples, one for the female rats and another one for the male rats. Denoting the corresponding parameters and statistics by subscripts $_f$ and $_m$, respectively, the statistical model is:

$$D_1, \ldots, D_4 \text{ are i.i.d. from } N(\mu_f, \sigma_f); \quad D_5, \ldots, D_{10} \text{ are i.i.d. from } N(\mu_m, \sigma_m),$$

and the two samples are independent. The parameters μ_f and μ_m represent the mean differences in digestibility for female and male rats. We have no reason to assume the standard deviations σ_f and σ_m to be equal. The statistical analysis has the steps:

- *Estimation*: $\hat{\mu}_f = \bar{D}_f = 0.975$, $\hat{\sigma}_f = s_f = 1.284$, $\hat{\mu}_m = 1.417$, $\hat{\sigma}_m = 1.504$,
- *Test* of H_0: $\mu_f = \mu_m$ vs. H_a: $\mu_f \neq \mu_m$ (again using a two-sided alternative):

  $$t = (\bar{D}_f - \bar{D}_m)/\sqrt{s_f^2/4 + s_m^2/6} = (0.975 - 1.417)/\sqrt{1.284^2/4 + 1.504^2/6} = -0.50.$$

  The software listing (for the non-pooled variance test) gives $P = 0.634$ from a t-distribution with df = 7. Without access to the software calculation of df, we would only know that df is at least 3 $(4 - 1)$ and at most 8 $(4 - 1 + 6 - 1)$. However, no matter the df, the test is clearly non-significant, $P \gg 0.05$.

We conclude that there is no indication in these data that the difference in digestibility between raw and roasted peanuts (shown under **b)**) would not be the same for female and male rats.

Subquestion d)
Among the 10 observations for raw peanuts $X_1, \ldots, X_{10}$, the first four were for female rats and the last six were for male rats. We then have a two-sample situation analogous to the one in **c)**, and the analysis follows the same steps:

- *Model*: $X_1, \ldots, X_4$ are i.i.d. from $N(\mu_f, \sigma_f)$; $X_5, \ldots, D_{10}$ are i.i.d. from $N(\mu_m, \sigma_m)$,
- *Estimation*: $\hat{\mu}_f = \bar{X}_f = 97.23$, $\hat{\sigma}_f = s_f = 1.544$, $\hat{\mu}_m = 97.50$, $\hat{\sigma}_m = 1.324$,
- *Test* of H_0: $\mu_f = \mu_m$ vs. H_a: $\mu_f \neq \mu_m$: $t = (\bar{X}_f - \bar{X}_m)/\sqrt{s_f^2/4 + s_m^2/6} = -0.29$; as before, this t-value is clearly non-significant; $P \gg 0.05$.

We conclude that there is no indication in these data that the digestibility of raw peanuts differs between female and male rats.

A final note on how this question relates to the previous question, which compared the sexes with respect to the differences in digestibility between raw and roasted peanuts. This question compared the digestibility values themselves, for raw peanuts, between the sexes. It would be very well possible for e.g. the females to have higher digestibility values than the males, and still the difference in digestibility of raw and roasted peanuts to be the same for females and males, or the other way around. In ANOVA terminology (to be discussed later in the course), the former is an interaction between sex and diet, and the latter is the main effect of sex.

Subquestion e.i)

The question suggests a non-parametric analysis. Non-parametric versions of the one-sample t-test for the differences (D_i) in **b)** are the sign test and Wilcoxon's signed rank test. Both of these tests would be for the hypothesis H_0 : median $= 0$ vs. H_a : median $\neq 0$, where the median is for the differences (D_i). Alternatively, it could be phrased in terms of the probability p of higher digestibability for raw than roasted peanuts: $H_0 : p = 0.5$ vs. $H_a : p \neq 0.5$. Without any relevant software listings at hand, the only possibility is to do a sign test, as follows.

Among the 10 sets, 8 differences are positive difference and 2 are negative. The test therefore corresponds to testing $p = 0.5$ based on observing $R = 8$ when $R \sim \text{Bin}(10, p)$. The P-value for the sign test is obtained from the $\text{Bin}(10, 0.5)$ distribution as twice the tail probability: $P = 2 \cdot \text{P}(R \geq 8) = 2 \cdot (0.044 + 0.010 + 0.001) = 0.11$, using a binomial distribution table. We see that this test does not give significance at $P < 0.05$. Whether this is due to the relative weakness of the sign test to the t-test or because the t-test is questionable due to problems with the normality assumptions, is difficult to answer without further analysis. It would clearly be of interest to try also the more powerful Wilcoxon's signed rank test, or other non-parametric tests (beyond the course syllabus).

Subquestion e.ii)

Our sample size calculation will focus on the differences within sets that we used for inference on the primary research question. From the observed ("pilot") data we have our guessed value for $\sigma = 1.363$. In order to achieve a margin of error of at most $m = 0.5$ for a 90% confidence interval (corresponding to $z^* = 1.645$), we will first use the formula that assumes the standard deviation to be known (slide 8L−12).

$$n \geq (z^* \cdot \sigma/m)^2 = (1.645 \cdot 1.363/0.5)^2 = 20.1,$$

leading us to a sample size of 21. However, as we estimate the standard deviation we will need to replace z^* by t^* from a $t(20)$-distribution. Recalculating with $t^* = 1.725$ yields $n \geq 22.1$, so we take $n = 23$. The resulting proposed design therefore consists of 23 sets (pairs) of rats, ideally with an approximately equal distribution of female and male sets, for a total of 46 rats in the experiment.

Subquestion e.iii)

The sets were intended to serve as blocks, so that before the trial treatments the two rats in a set were more similar than two rats from different sets. We should therefore expect some similarity between the two values from a pair, with the digestibility being generally somewhat higher in the rat that was fed raw peanuts. Graphically we could assess the similarity within pairs by plotting the values in a pair against each other (e.g., raw peanut digestibility on the x-axis and roasted peanut digestibility on the y-axis) and look for whether the points seem to show a positive association. An alternative graph would plot the two values from each set with two different symbols against the set number (x-axis). We could then look for whether the points seem to cluster (in their y-values) in the pairs.

The most obvious numerical summary would be the correlation between the two values in each pair; however, correlations have not been covered in the course yet. One idea available to us is to compare the variability (standard deviation) within and between sets. If the two values within a pair were independent, we would have (slide 3L−16),

$$\text{Var}(X_i - Y_i) = \text{Var}(X_i) + \text{Var}(Y_i) \approx 1.337^2 + 1.106^2 = 3.011 = 1.735^2.$$

However, the standard deviation for the differences (D_i) was estimated at 1.363, so somewhat lower. This would indeed correspond to a dependence within each pair, and lower variability within than between pairs.