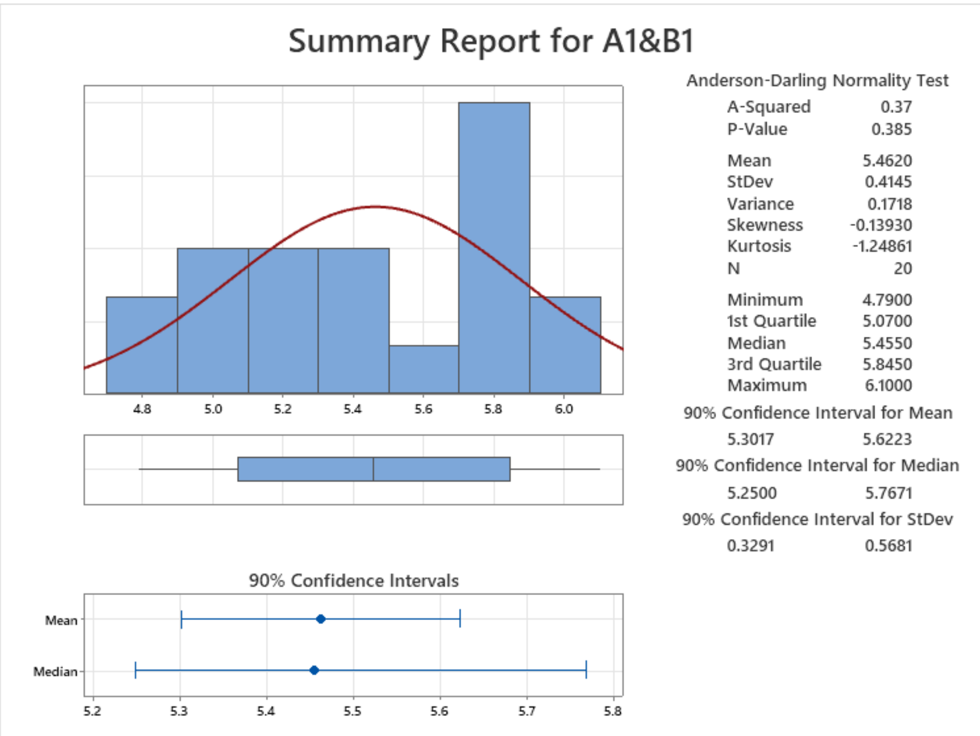
Biostats VHM 801 Course
Fall 2022, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to home assignment II

The solution is more detailed than required for a 100% mark, by covering multiple options for several questions. The solution covers all five questions where only the first three questions would be included for the 10% version of the assignment.

1. Description and inference for samples

The 20 original samples represent the population of samples typically analyzed at the lab. These could be considered as i.i.d. (independent and identically distributed), or a simple random sample, from the distribution in that population, but the 40 values included in the dataset are not because there are two values for each of the original samples. The two values per sample are paired and should not be assumed independent. One option to represent the distribution of the original samples is to select one of the two values we have per sample. From group A, we would most naturally pick A1, in order to avoid any impact of the first sample on A2. For group B, there is no obvious preference for B1 or B2, but one could argue that B1, as the first measurement on each sample, is the one most similar to A1. So combining A1 and B1 yields a sample with 20 values representing the desired distribution. Alternatively we could average the two values A1 and A2 as well as the values B1 and B2. One potential concern is that the two values were not obtained in the same way in groups A and B, and it may also be undesirable that the variability between those 20 means will not represent the variability among the 20 samples: it will be smaller because there is less variation in a mean than in the individual samples. Still, this can be considered as an acceptable approach. For this solution, we will work from the combined A1 and B1 samples. The graphical summary below shows graphics and descriptives for that distribution; the confidence level was set at 90%.



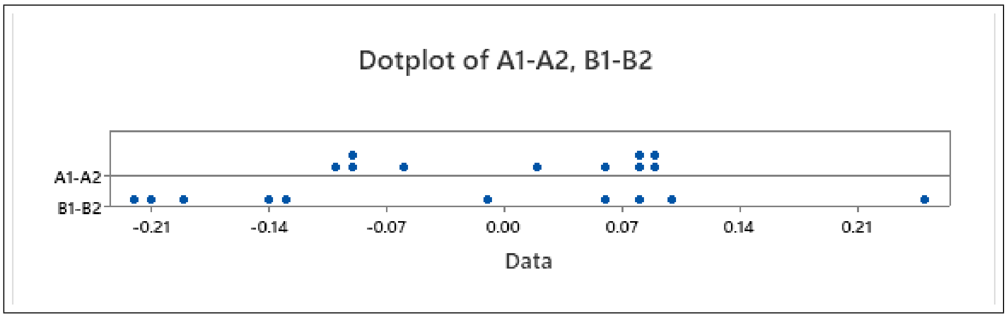
The distribution appears fairly symmetrical, with the mean and median very close, the box also very symmetrical, and a skewness value of -0.14. The histogram gives the impression of a more severe left-skewness, but this is perhaps an artifact caused by having too many bins; a dotplot would represent the values better with n so small. The distribution appears unimodal (again contrasting the too noisy histogram) and centered around 5.45. The boxplot does not indicate any suspected outliers.

For statistical inference we will assume the observations, say $X_1, \ldots, X_{20}$ for the combined values from A1 and B1, to be i.i.d. from $\mathrm{N}(\mu, \sigma)$. There are no concerns with these assumptions: in addition to the description above, it can also be noted that the A-D normality test is completely non-significant at $P = 0.385$. The display shows the 90% confidence interval as $(5.30, 5.62)$; we are 90% confident that the interval includes the population mean. The CI was computed as follows (using $t^*(19) = 1.729$):

$$\hat{\mu} = \bar{X} = 5.462, \ \hat{\sigma} = s_X = 0.4145,$$
$$90\% \text{ CI for } \mu: \ 5.462 \pm 1.729 \cdot 0.4145/\sqrt{20} = 5.46 \pm 0.16 = (5.30, 5.62).$$

2. Differences within group A

We already noted above the measurements A1 and A2 to be paired. If we denote by by $X_1^{A1}, \ldots, X_{10}^{A1}$ and by $X_1^{A2}, \ldots, X_{10}^{A2}$ the measurements for A1 and A2, respectively, our inference will be based on the differences $D_i = X_i^{A1} - X_i^{A2}$, $i = 1, \ldots, 10$. For t-distribution inference, we will need to assume that $D_1, \ldots, D_{10}$ are i.i.d. from $\mathrm{N}(\mu_D, \sigma_D)$. With only 10 observations it is difficult to assess normality assumptions, but a dotplot of these values (for convenience of presentation, the plot below also includes the differences in group B) shows the distribution to not look very nicely normal, perhaps even bimodal with modes on either side of 0, and the A-D normality test gives evidence against the normal distribution at $P = 0.020$. For a start, we will nevertheless continue with analysis based on the normal distribution model.



In a single sample, estimation is by the sample mean and standard deviation: $\hat{\mu}_D = \bar{D} = 0.008$, and $\hat{\sigma}_D = s_D = 0.083$. The hypothesis of interest is no systematic difference between samples A1 and A2 and hence $\mathrm{H}_0: \ \mu_D = 0$, and the default alternative hypothesis (in absence of any reason to do otherwise) is $\mathrm{H}_a: \ \mu_D \neq 0$. Our test statistic is the one-sample t-test with $n = 9$ degrees of freedom,

$$t = \bar{D}/\left(s_D/\sqrt{10}\right) = 0.008/(0.083/\sqrt{10}) = 0.30, \quad P = 2 \cdot P(t(9) \geq 0.30) = 0.77.$$

The test is clearly non-significant and offers no evidence whatsoever of a systematic mean difference between A1 and A2. It remains to take into consideration the problem with model assumptions. Because the test statistic is absolutely non-significant, we may speculate that the violation of assumptions for the t-test will hardly affect the significance. It also seems obvious from the dotplot that the distribution is centered around 0. If one wanted to confirm the result with an additional analysis, a non-parametric test would apply. Both the sign test and Wilcoxon's signed rank test will test hypotheses about the median, but they still correspond to systematic differences between the samples A1 and A2. With four differences below 0 and six above zero, it is no surprise that both tests are completely non-significant. In summary, we can be confident that the data contain no evidence of a systematic difference between A1 and A2 samples.

3. Differences within group B and comparison between groups

For a start, we repeat the analysis for group B, with the same construction of differences (the two subsamples are from the same sample so still paired, even if we may expect the pairing to have less of an impact when the two subsamples are not measured directly after each other), assumptions and analysis. The group B differences look better for the normal distribution assumption (e.g., the dotplot does not show the same apparent bimodal shape, even if the distribution is not clearly centered, and the A-D normality test is non-significant with $P = 0.31$, but then the test is not very powerful in such a small sample). The test statistic is computed as,

$$t = \bar{D}/\left(s_D/\sqrt{10}\right) = -0.041/(0.1602/\sqrt{10}) = -0.81, \quad P = 2 \cdot P(t(9) \geq 0.81) = 0.44.$$

We conclude there is no evidence of a systematic difference between the B1 and B2 samples. With the test outcome being clearly non-significant and the reasonably good (apparent) compliance with the normal distribution assumptions, we can be confident about this conclusion without adding further analysis.

We were also asked for a direct comparison of groups A and B. The fact that there was no significant difference in either group offers an indirect comparison, but because we have *not* proven that the mean difference equals 0 in the two groups, it is still possible that differences could exist between the groups. Groups A and B form two independent samples; there are no links between the samples in the two groups. The obvious analysis is therefore a two-sample comparison between the diffences obtained in groups A and B. The model assumptions are unchanged: two separate normal distributions, and there is no reason to assume the two standard deviations to be equal. The display shows results for the two-sample t-test from Minitab.

Method

μ₁: population mean of A1-A2
μ₂: population mean of B1-B2
Difference: μ₁ - μ₂

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
A1-A2	10	0.0080	0.0831	0.026
B1-B2	10	-0.041	0.160	0.051

Estimation for Difference

Difference	95% CI for Difference
0.0490	(-0.0743, 0.1723)

Test

Null hypothesis H₀: μ₁ - μ₂ = 0
Alternative hypothesis H₁: μ₁ - μ₂ ≠ 0

T-Value	DF	P-Value
0.86	13	0.406

The 95% confidence interval for the difference $\mu_D^A - \mu_D^B$ is shown as $(-0.07, 0.17)$ and thus easily includes 0. The t-test for the null hypothesis of equal means, $\mathrm{H}_0 : \mu_D^A = \mu_D^B$, gives $t = 0.86$ with 13 degrees of freedom and $P = 0.41$, against a two-sided alternative. There is no evidence of a mean difference in the devations betwen the first and second samples between groups A and B. The concern with the normality assumption in group A carries over to this test as well, and we may want to confirm our result with the non-parametric two-sample test (Wilcoxon-Mann-Whitney); it gives $P = 0.34$. We conclude that no systematic mean differences (between the first and second measurements) have been established within any of the two groups, and that also between the groups the mean differences did not show any systematic differences. Thus, when considering the means there is no trace of an effect of the different sampling scenarios in groups A and B.

4. Variability between duplicate measurements

In scenario A, the lab technician is not blinded to the value obtained from the first measurement when conducting the duplicate analysis. In contract, the scenario in group B was specifically designed to not allow the technician to link subsamples from the same sample; hence, the technician was blinded to the value from the first measurement.

The statistic to quantify the variability between measurements is the sample standard deviation, as an estimate of the correponding population value. We already computed the sample standard deviations among the 10 differences between first and second samples in the two groups,

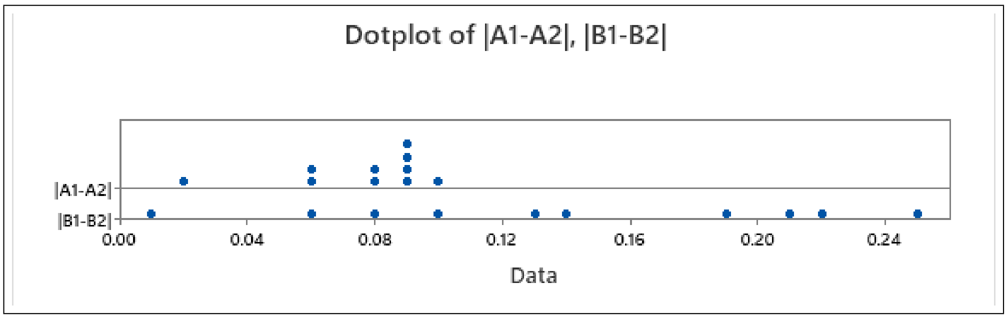
$$s_D^A = 0.083, \quad \text{and} \quad s_D^B = 0.160.$$

It is seen that the spread is almost twice in group B compared to group A. For the spread in group A with the non-blinded duplicate measurements to be smaller does indeed agree with the suspected impact of the first measurement on the second when the technician knows the first value.

Statistical inference for standard deviations (or variances) include Bartlett's test and Levene's test to compare variances from two independent samples. Bartlett's test is the "classical" method, but is nowadays considered less useful because of its strongly reliance on the normality assumption within the two samples (which would be a particular problem for our application). As a robust (to normality assumptions) alternative, Levene's test is more commonly used; it can be found in Minitab's menu for "2 Variances" under "Basic Statistics". Its P-value is just below 0.05, so suggestive, but not strongly conclusive, of a difference between the two standard deviations. For the assignment it was not expected to include any statistical inference to compare the two standard deviations. The confidence intervals for the standard deviations obtained in the Graphical Summary are non-conclusive because overlapping between the two gruops. We will later in the course discuss what conclusions can be drawn from comparing confidence intervals of independent samples (Session 9).

5. Comparison of numerical differences

The smaller variability within the pairs in group A compared to group B should translate into the numerical differences within pairs to be smallest in group A. The reason our first analyses focused on the mean differences within pairs did not pick this up, is that the deviations between the first and second samples can go in either direction, but the direction is not the feature of primary interest when the focus is on how much the two values in a pair differ. Therefore, working with the numerical (or absolute) differences in each pair has a potential to better to pick up if the variability was less in group A than group B. We will effectively be using the same two-sample comparison as in Question 3, only for the numerical differences instead of the (ordinary) differences used there. We start by displaying the numerical differences by dotplots for each group.



The numerical differences clearly have smaller spread in group A, but the dotplot also shows that we again face problems with the normality assumption for group A (the A-D normality test gives $P = 0.019$). Now the main parameters of interest are the means of the two distributions, and we want to test the hypothesis $H_0 : \mu^A_{|D|} = \mu^B_{|D|}$. Because the focus is on establishing a smaller variability in group A, we could justify the one-sided alternative $H_a : \mu^A_D < \mu^B_D$, but the two-sided alternative would also be valid. Results below are given for the two-sided alternative, and therefore P-values should be divided by 2 for the one-sided alternative.

Method

μ₁: population mean of |A1-A2|
μ₂: population mean of |B1-B2|
Difference: μ₁ - μ₂

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
|A1-A2|	10	0.0760	0.0237	0.0075
|B1-B2|	10	0.1390	0.0778	0.025

Estimation for Difference

Difference	95% CI for Difference
-0.0630	(-0.1203, -0.0057)

Test

Null hypothesis H₀: μ₁ - μ₂ = 0
Alternative hypothesis H₁: μ₁ - μ₂ ≠ 0

T-Value	DF	P-Value
-2.45	10	0.034

The mean difference between the two samples of numerical differences equals -0.063, and the two-sample t-test gives $t = -2.45$ and $P = 0.034$ with 10 degrees of freedom. Note that also here we have no interest in assuming the standard deviations to be the same. So there is some evidence in favour of a lower variability in group A, even if we do not know how strongly the problem with the normality assumption in group A affects the results. Our guidelines for 1-sample t-distribution inference required good compliance with the normal distribution when $n < 15$, but for the two-sample t-test one may use the same guidelines with n replaced by the sum of the two sample sizes, $n_1 + n_2$ (Chapter 7 of the IPS textbook). Additionally, the two-sample t-test is very robust when the sample sizes are equal (as in our case). So we should be able to rely on our (weak) significance against equal mean absolute deviations in the two groups. For comparison, the Mann-Whitney non-parametric test gives $P = 0.053$, but the more powerful two-sample randomization test (Minitab's "Resampling" menu found under "Calc") gave $P = 0.027$ (for one run with 10000 resamples). The latter result confirms out 2-sample t-test result because it also tests the hypothesis of equal means.