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

Solution to Midterm Exam, October 2018

The solution is more detailed than required for a 100% score, by including answers to all four 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 Teasdale et al. (1993), On the cognitive penetrability of posture control, *Experimental Aging Research* **19**, 1–13.

Question 1

Subquestion a)

The study involved an experimental setup for the sway measurements, but because no treatments were imposed on the subjects for the data discussed here (age is not an assigned treatment), it is probably better described as observational (possibly a cross-sectional study because it involved only a single time point). The study design is two independent samples (young and elderly subjects), and the sway measurements in the two directions could be considered as paired data (for each subject). As for important study features, it can be said that no randomization seems to be involved in the study execution; however, the researchers would have needed to ensure that the two age groups were similar with regard to important (potential) confounders, such as gender, body weight and any health conditions that could affect balance (preferably subjects with such conditions should not be included). It is good that the sample sizes in the two groups are almost the same (9 and 8 individuals).

Subquestion b)

From the dotplots and descriptive statistics we may summarize the side-by-side sway distributions for elderly and young as follows:

- *center*: the mean is a bit higher and the median somewhat higher for the elderly than young (7.2 mm and 3.0 mm, respectively),
- *spread*: the spread is much larger among elderly than young people – both the standard deviation and the interquartile range are $2-3$ times higher for the elderly,
- *shape*: both distributions are somewhat right-skewed, and the skewness is more pronounced for the elderly (1.06) than the young (0.51),
- *extremes*: none of the distributions have single extreme values (there are two high values among the elderly, but in such a small sample it is difficult to say whether they correspond to outliers or just a long right tail),
- *normality*: the normality test is non-significant for both groups, with $P = 0.11$ not too far off significance for the elderly, probably reflecting the moderate right-skewness, but overall the tests give no formal evidence against a normal distribution.

Subquestion c)

Calculations should be done for the two groups separately, by the presumed difference between them and the differences in the distributions found in the descriptive analysis. We describe the assumptions and calculations for the elderly group, and summarize the results for both groups in the table below. Let $X_1, \ldots, X_9$ denote the side-by-side sway among the $n = 9$ elderly subjects, and assume these to be independent and normally distributed as $\mathrm{N}(\mu, \sigma)$. The unknown model parameters μ and σ are

estimated by their sample values: $\hat{\mu} = \bar{X} = 22.3$ and $\hat{\sigma} = s = 10.3$. For a 95% confidence interval, the relevant t-distribution has $n - 1 = 9 - 1 = 8$ df, and $t^* = 2.306$, so we get:

$$95\%\text{ CI for } \mu\text{: } \bar{X} \pm t^* s/\sqrt{n} = 22.3 \pm 2.306 \cdot 10.3/\sqrt{9} = 22.3 \pm 7.9 = (14.4, 30.2).$$

We are 95% confident that the true mean sway in a population of elderly people will be included in this interval. The interval does however not represent a 95% range of sways in the population. Based on the normal distribution and the 68-95-99.7 rule, the 95% range would be estimated as $\mu \pm 2\sigma$,

$$95\%\text{ range for values: } \bar{X} \pm 2s = 22.3 \pm 2 \cdot 10.3 = 22.3 \pm 20.6 = (1.7, 42.9).$$

The estimated range goes close to zero (perhaps unreasonably so), reflecting that the distribution is right-skewed and that the approximation therefore is probably quite poor (and definitely not exact). For the 95% CI, the right-skewness of the distribution indicates the interval to be approximate, but probably reasonably good (due the approximate normality of $\bar{X}$). In the young group, the normality assumption seems reasonable despite some minor right-skewness, so the CI should be exact, whereas the range depends on the distribution being truly normal (which it probably is not).

Statistic/feature	Elderly group	Young group
95% CI	$22.3 \pm 2.306 \cdot 10.3/\sqrt{9}$ $22.3 \pm 7.9 = (14.4, 30.2)$	$15.1 \pm 2.365 \cdot 3.9/\sqrt{8}$ $15.1 \pm 3.3 = (11.8, 18.4)$
assessment	approximate (good)	exact
95% range	$22.3 \pm 2 \cdot 10.3$ $22.3 \pm 20.6 = (1.7, 42.9)$	$15.1 \pm 2 \cdot 3.9$ $15.1 \pm 7.8 = (7.3, 22.9)$
assessment	approximate (poor)	approximate (good)

Subquestion d)

The age groups form two independent samples. For the statistical inference, we have a choice between continuing with the normal distribution models already used in each of the samples, or switching to a nonparametric analysis using the Wilcoxon-Mann-Whitney (WMW) test (included in the Minitab listing). With both distributions somewhat, but not strongly, right-skewed both choices can be considered as reasonable. In favour of the normality assumption, we could say that the approximation appears to be reasonably good, and that t-distribution inference gives inference about a relevant parameter, namely the mean difference. In favour of the nonparametric analysis, we could say that its inference does not suffer from any uncertainty from non-normal distributions. The two distributions seem to differ clearly in their spread (as earlier noted), and this has implications for both analyses. For the t-test, we should use the version that does not assume equal variances. For the WMW test, the added "Δ-assumption" is not tenable (because the different spread will give the distributions different shapes). This in turn means that the confidence intervals for the difference between medians are not valid, so the method will not provide any confidence intervals. We summarize the two analyses in the table below, where index 1 $\sim$ elderly group, and index 2 $\sim$ young group.

Statistic/feature	2-sample t-test	Wilcoxon-Mann-Whitney test
assumption (per sample)	normal distribution (μ, σ)	no distributional assumption
parameter of interest	mean difference $\mu_1 - \mu_2$	no parameters
null hypothesis H_0	equal means: $\mu_1 = \mu_2$	equal distributions: $\mathcal{P}_1 = \mathcal{P}_2$
alternative hypothesis H_a	higher means: $\mu_1 > \mu_2$	$\mathcal{P}_1$ systematically larger than $\mathcal{P}_2$
test statistic	$t = 1.95$ (Minitab)	rank sum = 98 (Minitab)
P-value	$0.080/2 = 0.040$	$0.111/2 = 0.056$
conclusion	reject H_0, weak diff. demonstrated	cannot reject H_0, no evidence of diff.

Note that in both cases a one-sided alternative hypothesis is chosen, because it is "well-known" (as stated in the question) that elderly people have less balance than young people. The P-values against the one-sided alternative were both computed by halving the P-values in the Minitab listing for the two-sided alternative; this was valid because the estimate was in the direction of the one-sided alternative of interest.

Subquestion e), part i)

The setup and design for the forward-backward sway are entirely analogous to that for the side-by-side sway analyzed up till now. A comparative descriptive analysis shows that also for this variable the distribution is centred at higher values for the elderly than young. It however also shows a rather extreme value for subject 9 in the elderly group, and a normality assumption for this group seems no longer acceptable (e.g., the normality test is clearly significant). Therefore medians seem preferable to means for comparing groups, with values of 24.0 and 17.0 for the elderly and young groups, respectively. For statistical inference without a normality assumption, we will need a rank-based approach, i.e. the WMW test which is not included among the Minitab listings. The WMW analysis will proceed along similar lines as indicated under **d)**, e.g. once more with a one-sided alternative hypothesis. It is difficult to assess whether the "Δ-assumption" of equal shapes can be justified, among other things because we don't know how much the single extreme value affects the descriptive statistics. For that reason it might be of interest to recompute descriptive statistics without the extreme value; if that comparison shows the distributions for the two age groups to have reasonably similar shape, the "Δ-assumption" could be justified. In that case, the WMW procedure would provide inference about the difference between medians.

Subquestion e), part ii)

The two directions of sway measured for the same subject are paired observations, and therefore a comparison between the two directions should be based on the differences, e.g. the `diff` variable constructed. The descriptive statistics show the distributions of the differences to be quite similar in the two age groups, and it may therefore be acceptable to explore this question in the combined data for both age groups. One could perform a comparison between the age groups as an initial step and only proceed to the combined data if that comparison showed no reason to suspect an age effect. Although this could be done with a normal distribution analysis from the information provided, such an additional step is hardly feasible within the time constraints of the exam. (Indeed there is no evidence of an age effect; results not shown.) The statistical design is now a single sample, say from a distribution with mean μ_D and standard deviation σ_D, as well as its median$_D$. A normal distribution analysis seems well justified from the descriptive statistics (and a clearly non-significant P-value of the normality test, $P = 0.49$), but the data also allow a nonparametric analysis with the sign test (whereas the Wilcoxon signed rank test would require additional Minitab prints). We show again the results in tabular form. Alternative hypotheses are now chosen as two-sided, in absence of any

information to indicate a particular direction.

Statistic/feature	1-sample t-test	sign test
assumption	normal distribution (μ, σ)	no distributional assumption
parameter of interest	mean μ_D	median_D
95% confidence interval	$3.5 \pm 2.120 \cdot 6.6/\sqrt{17}$	not computable by hand
null hypothesis H_0	zero mean: $\mu_D = 0$	zero median: $\text{median}_D = 0$
alternative hyp. H_a	non-zero mean: $\mu_D \neq 0$	non-zero median: $\text{median}_D \neq 0$
test statistic	$t = 3.5/(6.6/\sqrt{17}) = 2.19$	$X = 12$ out of $n = 15$
P-value	$2 \cdot P(t(16) > 2.19) < 0.05$	$2 \cdot P(X \geq 12) = 2(0.014 + 0.003) = 0.034$
conclusion	reject H_0, weak diff. demonstrated	reject H_0, weak diff. demonstrated

Subquestion e), part iii)

A visual assessment of the dotplots suggest that only two observations are of interest as potential outliers: the large `Fwd/Back` sway (50) for subject 9 in the elderly group, and the low (negative) difference (-11) for subject 2 in the elderly group. Both of these potential outliers should be assessed within the elderly group. For the `Fwd/Back` sway, we get IQR $= 10$ and an upper cut-off of $29.5 + 1.5 \cdot 10 = 44.5$, which is exceeded by the observed value of 50. For the sway `diff`, we get IQR $= 9.5$ and a lower cut-off of $1.0 - 1.5 \cdot 9.5 = -13.3$, which is not exceeded by the observed value. We therefore continue our analysis with the large value for forward-backward sway only.

Using the estimated $\bar{Y} = 26.3$ and $s = 9.8$, its z-score equals $z = (50 - 26.3)/9.8 = 2.42$, with a corresponding tail probability of 0.0078. Therefore the probability of one value being as extreme (in either direction) as 50 is estimated at $2 \times 0.0078 = 0.0156$. The estimated probability of at least 1 such extreme value in a sample of 9 is computed from a binomial distribution $B(9, 0.0156)$ as $P = 1 - (1 - 0.0156)^9 = 0.13$. From that we conclude that this particular extreme value is still within what we could expect by chance alone, and there is no evidence of a concern about extreme values in these data. Note that these calculations are entirely similar to those in the midterm exam of 2015.

Subquestion e), part iv)

The question asks us to assess $X_{\text{obs}} = 8$ in the binomial distribution $X \sim B(8, p)$, where $p = 14/17 = 0.8235$ is overall proportion of males across the two samples. Therefore,

$$P(X = 8) = p^8 = 0.8235^8 = 0.23,$$

and the event of all people in the young age group being males does therefore not seem particularly surprising. Thus we have no reason to suspect a real difference in the gender distribution between the age groups. The calculation here is in fact closely related to (but less precise than) the calculations for Fisher's exact test, which also points towards non-significance for these data (not shown).