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

Mid-term exam, 25 October 2018

All aids are allowed, except a computer-like device (including tablets and smartphones) and personal assistance. The exam consists of one question with five subquestions (labeled by letters **a)-e)** with weights as indicated) that should all be answered to achieve the maximal 15 points. The subquestions can be answered reasonably independently of each other. The mid-term exam accounts for 15% of the course mark; however, every student may choose to waive the result of the mid-term exam. The duration of the mid-term exam is 1 hour.

Generally, **statistical models and methods should be specified**. If you realize that you need more information than is provided to carry out the analysis, specify what information you need, how you would obtain it using Minitab (or another software), and how you would use it.

Question 1 (*15 points*)

Researchers from Université Laval, Quebéc, conducted a comparative study measuring how good young and elderly people are at keeping their balance (i.e., a stable upright position). It is well-known that the ability to keep one's balance decreases with age, and the novelty of the study was in exploration of the impact of processing additional sensory (auditory) stimuli.

The study involved 17 volunteers, of which nine were classified as elderly (average age 71.1 years) and eight were classified as young (average age 24.6 years). In one trial, each subject stood barefoot on a "force platform" and was asked to maintain a stable upright position while also reacting as quickly as possible to an unpredictable noise by pressing a handheld button. The noise came at random times, and while the subjects concentrated on making their reaction times as fast as possible, the platform automatically measured how much they swayed in both the forward/backward and the side-to-side directions. For each subject, the range of swaying (in *mm*) in both the forward-backward (F/B) and side-to-side (S/S) directions was measured during a 20-second period, and then averaged across two such periods.

Age group	Elderly		Young	
Subject	F/B sway	S/S sway	F/B sway	S/S sway
1	19	14	25	17
2	30	41	21	10
3	20	18	17	16
4	19	11	15	22
5	29	16	14	12
6	25	25	14	14
7	21	18	22	12
8	24	21	17	18
9	50	37		

Descriptive statistics and graphical representations for the two sway variables, both within each age group and overall, have been generated in Minitab and are included in the attachment. The same statistics and displays are also included for a difference in sway variable, computed as: `diff = Fwd/Back - Side/Side`.

a) (*2 points*) Characterize the study type (e.g., experimental or another type) and study design (e.g., one-sample or another design). Discuss briefly any study features of importance for making the two age groups as comparable as possible.

b) (*3 points*) Carry out a brief comparative, descriptive analysis of the side-by-side sway measurements for the two age groups; contrast the distributions using standard descriptors such as the center, spread, shape and extremes.

c) (*3 points*) Compute a 95% confidence interval for the mean side-by-side sway. Does this interval correspond to the expected 95% range for side-by-side sway measurements? If not, compute also an interval to represent this range. Comment briefly on whether the interval(s) should be considered as approximate or exact. (*Note:* If you choose, with a suitable justification, to perform the calculations for each age group separately, it is allowed to limit the calculation to one of the two age groups.)

d) (*2 points*) Use a statistical test to compare the side-by-side sway between the two age groups, as indicated in the rationale for the study (above). Make sure to describe the assumptions you make, the parameters you compare (if any) and the method you use, including any results from the attached Minitab listings, and draw conclusions.

e) (*5 points*) For this subquestion, you should answer **two of the four parts** *i)-iv)* below, each with the same score. It is allowed (but not recommended) to answer more than two parts, in which case your score for subquestion **e)** will be for the best two parts among those answered.

i) (*2.5 points*) In a similar fashion as for **b)-d)**, but concentrating on the most important points, use statistical inference to compare the forward-backward sway between elderly and young people. Also here, make sure to describe the assumptions you make, the parameters you compare (if any) and the method you use. Carry out as much of the analysis as possible from the information provided (see the preamble of the question for the situation that more information is needed to complete the analysis).

ii) (*2.5 points*) The researchers were also interested in comparing the magnitude of sway between the two directions. Along similar lines as in the previous question(s), and while concentrating again on the most important points, use the information provided to carry out a (descriptive and inferential) statistical analysis to compare the sway in the forward-backward and side-by-side directions.

iii) (*2.5 points*) The descriptive statistics and displays for the variables may have indicated the possibility of outlying observations in some distribution tails. For distributions with potential outliers (not necessarily all distributions), carry out a calculation to determine how close the most extreme observations are to be labelled as "suspected outliers". In addition, for the observation you consider as most extreme among all observations, use

the normal distribution to compute the probability of observing at least one observation in a sample of the corresponding size that is as extreme as (or more extreme than) the actual one. Does this probability indicate that the observation should be considered as problematic?

iv) (*2.5 points*) The elderly age group included 6 males and 3 females, whereas all the subjects in the young age group were males. As this might seem as a sizeable difference in the gender distribution between the two groups, there was interest in estimating the probability that it could have happened by chance only. One way of doing this, is to compute the probability that all subjects in the young age group would be males when assuming the overall proportion of males (i.e., both age groups combined) to be valid for the young age group. Compute this probability, and draw conclusions – does the fact that all young subjects were males seem particularly surprising/suspect?

Minitab output:

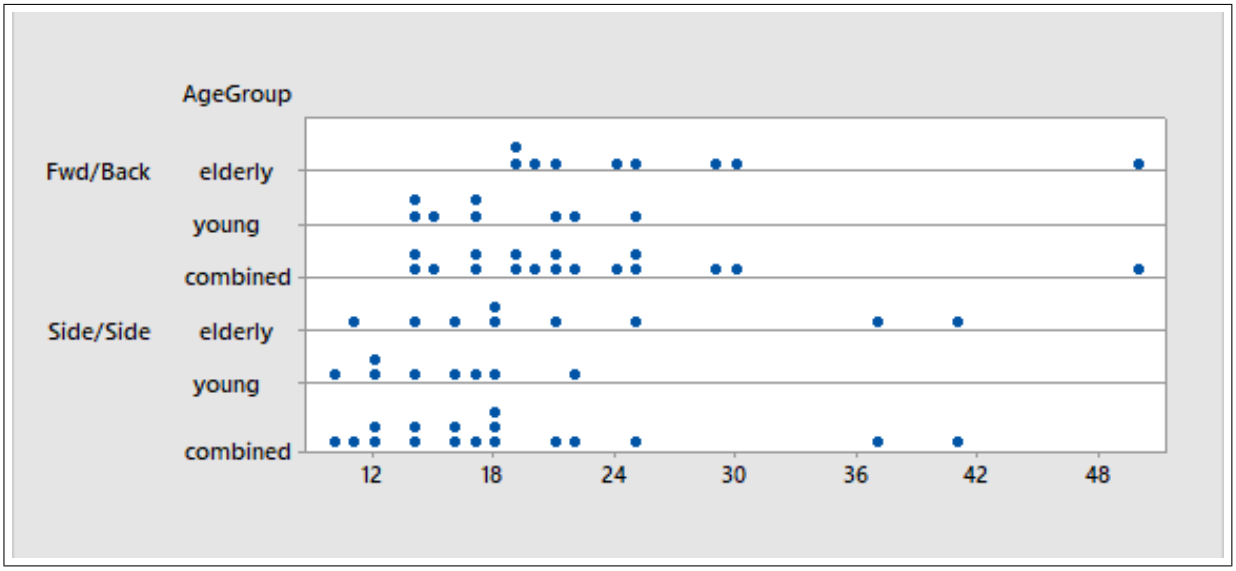
For simplicity of display, the descriptive statistics from Minitab have been transcribed into a formatted table.

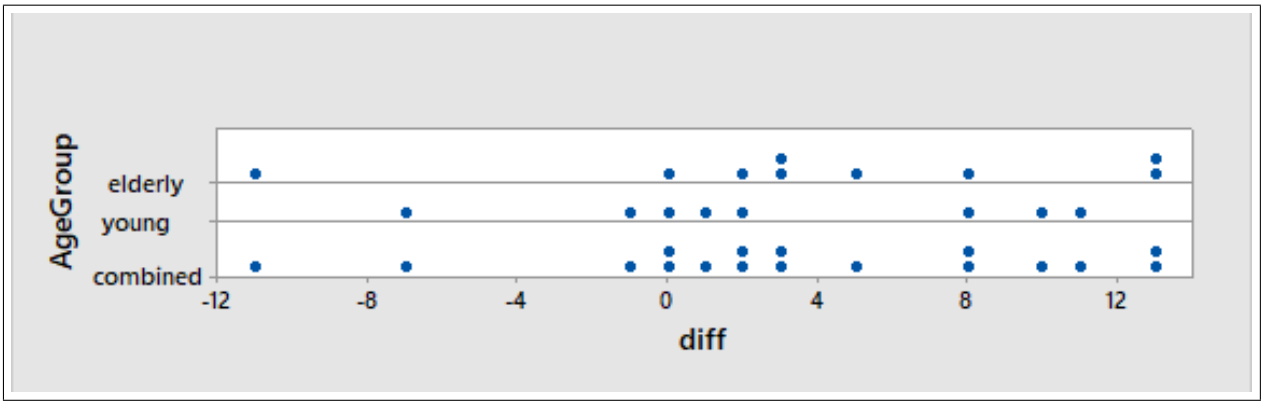
Age group	Elderly ($N=9$)			Young ($N=8$)			All combined ($N=17$)		
Statistic	F/B	S/S	diff	F/B	S/S	diff	F/B	S/S	diff
mean	26.3	22.3	4.0	18.1	15.1	3.0	22.5	18.9	3.5
std.dev.	9.8	10.3	7.3	4.1	3.9	6.2	8.5	8.6	6.6
minimum	19.0	11.0	-11.0	14.0	10.0	-7.0	14.0	10.0	-11.0
Q1	19.5	15.0	1.0	14.3	12.0	-0.8	17.0	13.0	0.0
median	24.0	18.0	3.0	17.0	15.0	1.5	21.0	17.0	3.0
Q3	29.5	31.0	10.5	21.8	17.8	9.5	25.0	21.5	9.0
maximum	50.0	41.0	13.0	25.0	22.0	11.0	50.0	41.0	13.0
skewness	2.09	1.06	-0.80	0.65	0.51	-0.10	2.22	1.65	-0.46
kurtosis	4.89	-0.01	1.50	-0.96	-0.20	-0.81	6.54	2.44	0.14
P-value[a]	0.015	0.11	0.34	0.30	0.82	0.49	0.008	0.005	0.49

[a] Anderson-Darling normality test

```
MTB > Name C5 'diff'
MTB > Let 'diff' = 'Fwd/Back'-'Side/Side'
MTB > Describe 'Fwd/Back' 'Side/Side' 'diff';
SUBC>   By 'AgeGroup';
...
MTB > PPlot 'Fwd/Back' 'Side/Side' 'diff';
SUBC>   Normal;
...
SUBC>   Panel 'AgeGroup'.
```

(continues on the next page)






```
MTB > Dotplot ( 'Fwd/Back' 'Side/Side' ) * 'AgeGroup';
SUBC>   Overlay;
SUBC>     VLast.
MTB > Dotplot ( 'Fwd/Back' 'Side/Side' ) * 'AgeGroup';
SUBC>   Overlay.
MTB > Dotplot ( 'diff' ) * 'AgeGroup';
SUBC>   Overlay.
```

(continues on the next page)

Two-Sample T-Test and CI: Side/Side, AgeGroup

Method

μ_1: mean of Side/Side when AgeGroup = elderly
μ_2: mean of Side/Side when AgeGroup = young
Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

Descriptive Statistics: Side/Side

AgeGroup	N	Mean	StDev	SE Mean
elderly	9	22.3	10.3	3.4
young	8	15.13	3.91	1.4

Estimation for Difference

Difference	Pooled StDev	95% CI for Difference
7.21	7.98	(-1.06, 15.47)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.86	15	0.083

Two-Sample T-Test and CI: Side/Side, AgeGroup

Method

μ_1: mean of Side/Side when AgeGroup = elderly
μ_2: mean of Side/Side when AgeGroup = young
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: Side/Side

AgeGroup	N	Mean	StDev	SE Mean
elderly	9	22.3	10.3	3.4
young	8	15.13	3.91	1.4

Estimation for Difference

Difference	95% CI for Difference
7.21	(-1.03, 15.45)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.95	10	0.080

(continues on the next page)

Mann-Whitney: Side/Side_elderly, Side/Side_young

Method

η_1: median of Side/Side_elderly
η_2: median of Side/Side_young
Difference: $\eta_1 - \eta_2$

Descriptive Statistics

Sample	N	Median
Side/Side_elderly	9	18
Side/Side_young	8	15

Estimation for Difference

Difference	CI for Difference	Achieved Confidence
4	(-1, 15)	95.15%

Test

Null hypothesis	$H_0: \eta_1 - \eta_2 = 0$
Alternative hypothesis	$H_1: \eta_1 - \eta_2 \neq 0$

Method	W-Value	P-Value
Not adjusted for ties	98.00	0.112
Adjusted for ties	98.00	0.111

```
MTB > TwoT 'Side/Side' 'AgeGroup';
SUBC>   Confidence 95.0;
SUBC>   Test 0.0;
SUBC>   Alternative 0;
SUBC>   Pooled.
MTB > TwoT 'Side/Side' 'AgeGroup';
SUBC>   Confidence 95.0;
SUBC>   Test 0.0;
SUBC>   Alternative 0.
MTB > Mann-Whitney 95.0 'Side/Side_elderly' 'Side/Side_young';
SUBC>   Alternative 0.
```