Vrije Universiteit Amsterdam	Statistical Data Analysis, Exam II
Faculty of Sciences	29 May 2019

SOLUTION

Question 1 [2+2+2=6 points]

a. Correct. In this case, midranks are used instead of the usual ranks.

b. Incorrect. For example, if the sample size is small and the data is non-normally distributed, the t-test is in general not valid, but the sign test is.

c. Correct. The permutation test then usually controls the type I error probability exactly, whereas the bootstrap test controls it only approximately.

Question 2 [3+2=5 points]

a. i. C_1 (Wilcoxon tests): **preferred** because data are apparently symmetrically distributed (and we assume continuity); hence, the Wilcoxon test is applicable. Also, it is a very powerful test, so it is usually preferred.
ii. C_2 (sign tests): **not preferred** because the Wilcoxon test is usually more powerful.

iii. C_3 (t-tests): **not preferred** because the data is apparently non-normal (yet symmetric): the tails are a bit too heavy as apparent from the histogram and the QQ-plot. Also, the outliers could cause problems.

b. C_1: one should then use mid-ranks instead of the usual ranks and one could use a suitable normal approximation to approximately find the quantiles of the distribution of the test statistic.

C_3: one could apply a bootstrap-version of the t-test and find the corresponding confidence interval. This would be much more reliable than the result from the usual t-test.

Question 3 [1.5+1.5+2+1+1=7 points]

a. We have one sample of size $n = 65$. The underlying distribution is multinomial (9-nomial) with parameters 65, cell probabilities $p_{ij} \geq 0, i, j = 1, ..., 3$, where

$$\sum_{i=1}^{3}\sum_{j=1}^{3} p_{ij} = 1.$$

For that model the hypothesis of the independence of the row and column variables is tested (that is whether body temperature and heart rate are independent). The hypotheses can be stated as

$$H_0 : p_{ij} = p_{i\cdot}p_{\cdot j}, \qquad \text{for all } i = 1, 2, 3,\ j = 1, 2, 3$$
$$H_1 : p_{ij} \neq p_{i\cdot}p_{\cdot j}, \qquad \text{for some } i = 1, 2, 3\ j = 1, 2, 3$$

where

$$p_{i\cdot} = \sum_{j=1}^{3} p_{ij}, \quad i = 1, 2, 3, \qquad p_{\cdot j} = \sum_{i=1}^{3} p_{ij}, \quad j = 1, 2, 3.$$

Or, equivalently expressed in words:

H_0 : "Heart rate and body temperature are independent."
H_1 : "Heart rate and body temperature are dependent."

b. The rule of thumb: at least 80% of the entries of the table of the expected values under H_0 have to be greater than 5, and all entries of that table have to be greater than 1.

   In our case, the smallest expected number under the null hypothesis corresponds to heart rate $\in [71, 78]$ and body temperature > 37.1, which is equal to $20 \cdot 19/65 \approx 5.85 \geq 5$. So, the rule of thumb is satisfied.

c. Under the null hypothesis, the test statistic X^2 approximately follows the $\chi^2_{2\cdot 2} = \chi^2_4$ distribution, for which the critical value equals 9.49 (for $\gamma = 0.95$). Therefore, the critical region is $[9.49, \infty)$. Since the observed value is not in the critical region ($4.79 < 9.49$), we do not reject the null hypothesis of independence.

d. With Kendall's test you would in this case reject the null hypothesis of independence because $0.033 < 0.05$. Hence, there is enough evidence that heart rate and body temperature are dependent.

e. The χ^2 test from c. did not reject the null hypothesis of independence, whereas Kendall's test did.

   This could be because Kendall's test is more powerful for many data distributions because it uses ranks instead of only "signs". (Also, the categorization of the observations was perhaps not optimal.)

Question 4 [2.5+2.5+2+2=9 points]

a. The t-scores for the different covariates are given as $\hat{\beta}_j/se(\hat{\beta}_j)$. The resulting t-scores are thus -3.83 (Intercept), 6.64 (thorax), -1 (sleep). Hence, for a two-sided t-test, sleep would have the least significant result. The critical value is found as the 97.5% quantile of the t-distribution with $n - p - 1 = 25 - 2 - 1 = 22$ degrees of freedom, hence, 2.074. Because $|-1| < 2.074$, we see that sleep does not have a significance influence on the full linear model. Hence, we remove sleep from the full model in the first step of the step down procedure.

b. The first plot in Figure 3, the normal QQ-plot of the residuals, can be used to investigate the normality of the residuals. Since the residuals are the estimated errors, one can (approximately) check the normality assumption of the errors.

   The second plot, the scatter plot of the residuals against the response variable values, can be used to check the constant error variance assumption.

   The line in the QQ-plot is quite straight, hence the normality assumption is plausible for these data.

   The constant error variance assumption is also acceptable because the spread of the points for small values on the horizontal axis is approximately equals to the spread for large values.

c. The last plot in Figure 3, i.e. the added variable plot for the variable sleep regressed with respect to intercept and thorax, shows what the additional inclusion of sleep would add to the linear model (only) without sleep.

   In particular, the residuals for sleep are compared with the residuals for longevity.

   From this plot, one can see that the residuals for both models seem quite independent of each other; there is no obvious pattern.

   Apparently, an inclusion of sleep into the model would not result in a better model, which is why we would keep out final model from b.

d. A leverage point is a point which has a hat-value greater than $2 * (p+1)/n = 2 * 3/25 = 0.24$.

   An influence point is a point with a Cook's distance of close to or greater than 1.

   In our case, all the hat-values are at most $0.21 \leq 0.24$ and Cook's distances are very far from 1 (i.e. below 1), hence there are no leverage or influence points.