Vrije Universiteit Amsterdam	Statistical Data Analysis, Exam II
Faculty of Science	1 June 2022

EXEMPLARY SOLUTIONS

Question 1 [9 points]

a. ARE(Wilcoxon,sign)=2 for a certain distribution means that the sign test would need the double sample size to reach the same (asymptotic) power of the Wilcoxon test.

b. The two-sample Wilcoxon test is based on the ranks of the observations of both samples within the pooled sample. Under the null hypothesis, every rank allocation is equally likely, thus the distribution of the test statistic doesn't depend on the data distribution.

c. $D = \max_{x\in\mathbb{R}} |\hat{F}_n(x) - \hat{G}_m(x)|$,
where $\hat{F}_n$ is the empirical distribution function of the first sample and
$\hat{G}_m$ is the empirical distribution function of the second sample.

d. Spearman's rank correlation test is based on ranks,
that's why the test statistic is much more robust against outliers.

e. Permutation tests have the advantage that they are exact in the special case of two exchangeable samples. (This is not so for bootstrap tests.)

Question 2 [8 points]

a. $(N_{11}, \dots, N_{33}) \sim Mult(92, p_{11}, \dots, p_{33})$ with
$H_0 : p_{jk} = p_{\cdot k} \cdot p_{j\cdot}$ for all $j, k = 1, 2, 3$ vs.
$H_a : p_{jk} \neq p_{\cdot k} \cdot p_{j\cdot}$ for at least one combination of $j, k = 1, 2, 3$.

(In words: H_0 : independence of favorite game and favorite snacks, H_a : dependence of favorite game and favorite snacks.)

b. Rule of thumb: at least 80% of the expected entries of the table under the null hypothesis should be at least 5; all of them have to be at least 1.

To check the rule of thumb, we start checking the smallest entry: $E_{13} = \frac{26\cdot25}{92} \approx 7.1 > 5$.
Thus, all entries are greater than 5 (and in particular greater than 1).
We conclude that the rule of thumb is satisfied.

c. For the test, we will use the 95%-quantile of the chi-squared distribution with $(3-1)^2 = 4$ degrees of freedom, i.e. $\chi^2_{4;0.95} = 9.49$.

The test score $X^2 = 11.438$ is greater than that critical vaule, hence we reject the null hypothesis of independence.

As a conclusion, game and snack preferences are not independent.

d. In the bootstrap for contingency tables, many contingency tables are randomly generated.
They are generated in a way such that the marginals are kept fixed, but the entries are randomly re-arranged as if the null hypothesis of independence was true.

Question 3 [10 points]

a. Step up: in the first step we would add the explanatory variables thorax to the linear model since that would result in the highest multiple R^2 value.

   Let us test the significance of this variable: $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$.

   The t-test score is $t = 125/18.94 \approx 6.6$.

   This is greater than the 97.5%-quantile of the t-distribution with $25 - 2 = 23$ degrees of freedom, thus the variable significantly improves the model, so it is added to the model.

b. That's the overall F-test.

c. The scatterplot of the residuals against thorax (in the model) does not show a clear pattern (nor a change of spread for small/larger thorax values).

   Hence, based on this, we don't see any reason for including also higher order terms of thorax in the model.

d. The scatterplot of the residuals against sleep (not in the model) does not show a clear pattern (nor a change of spread for small/larger sleep values).

   Hence, based on this, we don't see any reason for adding sleep to the model.

e. In the mean shift outlier model, we add an explanatory variable for a single data point, say i.

   That variable, say u, is 0 except for the i-th for which it equals 1.

   We then test based on a t-test whether the parameter corresponding to u is different from 0.