Experimental and Statistical Methods in Biological Sciences – First Exam

Instructions: There are five assignments that should *all* be completed. Maximum score for each assignment is 6. To get the top grade, you need to score 30 points; to pass the exam you need at least 16 points. Please read all the assignments before beginning. Plan your answers carefully. Good luck!

1. ***Define the following concepts***
    a. Regression coefficient (β)
    b. *p*-value
    c. Standard deviation
    d. Standard error of mean
    e. Chi-squared (χ^2) test
    f. Reliability

2. ***T-tests. Answer briefly the following questions related to t-tests***
    a. What are the three different t-test types?
    b. How do the t-tests use the standard error of mean for testing the difference between the means?
    c. Describe the t-distribution
    d. In general, are large or small t-values statistically significant? Why?
    e. What happens to t-value in the one-sample t-test if everything else remains constant and
        i. difference between population and sample mean increases
        ii. standard deviation decreases
        iii. sample size increases?
    f. Give an example of a study whose data could be analyzed with a t-test

3. ***Short essay: Describe the analysis of variance and discuss its applications in biomedical research***

4. **Applied linear regression analysis.** In this study the researchers addressed whether children's (n = 10) working memory capacity (WMC) is associated with regulation of aggressive acts. First, WMC was measured using a standard operation span task. After this, participants took a short test on the computer. They were shown pictures of various social encounters that involved potentially aggression-triggering elements. The participants were asked to choose how they would solve each situation. The researchers computed the number of aggressive solutions, and this score constitutes the variable *Number of aggressive acts*. The data were analyzed with linear regression analysis. Assess the data and visualizations below and answer the following questions:

a. Estimate (roughly) the correlation coefficient between working memory capacity and the number of aggressive acts.
b. Define (roughly) the regression coefficient (β_1) for the variable 'number of aggressive acts' and justify your conclusion.
c. Define the constant term (β_0) for the regression line and justify your conclusion
d. Compute (roughly) the sum of squared residuals
e. Report the results of the regression analysis you conducted above
f. Interpret the results

Table 2 Data matrix for the study

Subject ID	Working memory capacity	Aggressive acts
1	1	5
2	5	1
3	4	2
4	2	3
5	3	4
6	3	3
7	2	5
8	1	4
9	5	2
10	4	3

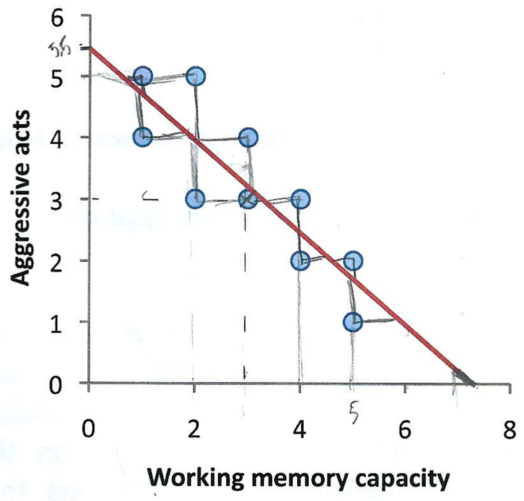

Figure 1. Scatterplot and regression line for variables *Working memory capacity* and *Aggressive acts*.

5. **Interpretation of R outputs.** This dataset comes from a study by Willerman et al. (1991) where the relationship between intelligence and brain size was examined. Intelligence was measured with four subtests of the Wechsler Adult Intelligence Scale-Revised (1981). The data includes the total IQ and two subscales (visual IQ and performance IQ) built using two subtests each (Vocabulary and Similarities, and Block Design and Picture Completion, respectively). Brain size was measured using Magnetic Resonance Imaging (MRI) and defined as number of voxels. The following variables are included:

    1. IQ: total IQ score
    2. VIQ: visual IQ score (based on visual subtests)
    3. PIQ: performance IQ score (based on performance subtests)
    4. Gender: Male, Female
    5. BMI: body mass index (calculated as the relationship between weight and height)
    6. BMI_group: BMI values categorized into three classes (lean – overweight - obese)
    7. MRI_voxels: number of voxels in the brain (the higher the number, the bigger the brain is)

The researchers were particularly interested in assessing whether gender and body mass index would be associated with brain volume. Study the descriptive statistics and outputs below, and answer the following questions:

a. Based on the descriptive statistics, what can you tell about the distribution of IQ, BMI and brain volume?
b. Are there gender differences in brain size? Report the results of the analysis as you would report it in a scientific article.
c. Are there differences in brain size between different BMI groups? Report the results of the analysis as you would report it in a scientific article.

Summaries of the data set

```
> summary(data)
   Gender         IQ              VIQ              PIQ             MRI_voxels           BMI          BMI_group
 Female:20   Min.   : 77.00   Min.   : 71.0   Min.   : 72.00   Min.   : 790619   Min.   :18.69   1:13
 Male  :20   1st Qu.: 89.75   1st Qu.: 90.0   1st Qu.: 88.25   1st Qu.: 855919   1st Qu.:20.69   2:14
             Median :116.50   Median :113.0   Median :115.00   Median : 905399   Median :22.30   3:13
             Mean   :113.45   Mean   :112.3   Mean   :111.03   Mean   : 908755   Mean   :22.60
             3rd Qu.:135.50   3rd Qu.:129.8   3rd Qu.:128.00   3rd Qu.: 950078   3rd Qu.:23.62
             Max.   :144.00   Max.   :150.0   Max.   :150.00   Max.   :1079549   Max.   :29.88
                                                                                 NA's   :2

> describe(data)
           var  n      mean       sd    median  trimmed      mad       min        max     range  ske
Gender*      1 40      1.50     0.51      1.5      1.50     0.74      1.00       2.00      1.00  0.0
IQ           2 40    113.45    24.08    116.5    114.09    33.36     77.00     144.00     67.00 -0.1
VIQ          3 40    112.35    23.62    113.0    112.03    28.91     71.00     150.00     79.00  0.0
PIQ          4 40    111.03    22.47    115.0    110.91    26.69     72.00     150.00     78.00 -0.0
MRI_voxels   5 40 908755.00 72282.05 905399.0 905557.22 71568.07 790619.00 1079549.00 288930.00  0.3
BMI          6 38     22.60     2.57     22.3     22.42     2.34     18.69      29.88     11.19  0.6
BMI_group*   7 40      2.00     0.82      2.0      2.00     1.48      1.00       3.00      2.00  0.0
```

Output 1

```
> t.test(MRI_voxels ~ Gender)

        Welch Two Sample t-test

data:  MRI_voxels by Gender
t = -5.2156, df = 38, p-value = 6.758e-06    0,000068
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -127987.86  -56413.74
sample estimates:
mean in group Female   mean in group Male
            862654.6             954855.4
```

Output 2

```
> tapply(MRI_voxels, BMI_group, mean)
       1        2        3
884526.8 916373.8 924778.4

> anova1 <- aov(MRI_voxels ~ BMI_group)
> summary(anova1)
            Df    Sum Sq   Mean Sq F value Pr(>F)
BMI_group    2 1.178e+10 5.891e+09   1.135  0.332
Residuals   37 1.920e+11 5.189e+09
```