Problem 1

A famous cereal manufacturer claims that the mean sugar content of his boxed cereal is 0.2 grams per 1 gram of cereals. A random sample of n cereal boxes is selected and the sugar content is recorded. We can assume in general that all measures, $X_1, X_2, \dots X_n$ are independent and normally distributed with expected value μ, which represents the sugar content in the cereals, and standard deviation $\sigma = 0.03$.

A hypothesis test with null hypothesis $H_0 : \mu = 0.2$ against an alternative hypothesis $H_1 : \mu > 0.2$ is conducted, and the null hypothesis is rejected with a significance level of 0.05.

a) Describe how to conduct such a hypothesis test based on the mean $\bar{X}$. What do you conclude if $n = 10$ samples are taken and the mean of the measures is equal to 0.22?

b) Derive the probability to reject the null hypothesis if $n = 10$ cereal boxes are selected and $\mu = 0.22$.

c) Assume that $\mu = 0.22$. Derive how many boxes, n, you have to select such that the probability to reject H_0 becomes bigger than 0.8.

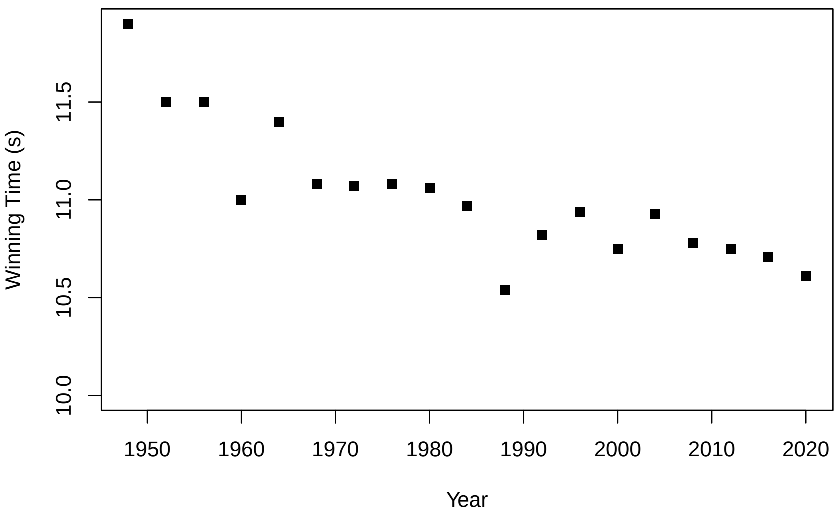
d) A new question arises on whether the colour of the cereal box influences its sales. To test this four different colored boxes are put on sale. The number of boxes of each color sold during the first week were:

Blue = 31, Yellow = 29, Red = 23, Green = 29.

Using a 5% significance level, test the null hypothesis that the number of boxes sold for each of theses four colours is the same.

Problem 2

The following figure shows Olympic 100-metre sprint winner times in seconds for women for the Olympic years from 1948 to 2020:



A linear regression model, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, was fitted to the women's times. Here y_i is the winning time in seconds and x_i denotes the Olympic year. We have a total of 19 observation pairs so that $i = 1, \ldots, 19$. The following parameter estimates were obtained:

$$\hat{\beta}_0 = 36.78, \quad \hat{\beta}_1 = -0.013, \quad \hat{\sigma}^2 = 0.036,$$

Further, $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 9120$ (that is 0.036/9120 is an estimate of the variance of the estimator of β_1).

a) What is the change in expected winning time from one Olympic Games to the next (Olympic Games only occur every 4 years)? What, according to the model, would have been the winning time in year 0?

b) Can we be confident that winner times get faster over time? Propose and perform a hypothesis test to address this question. In doing so, define the null hypothesis and the alternative hypothesis, compute the test statistics and determine the critical region for a significance level of 5%. What do you conclude?

Problem 3

A study investigated whether dogs prefer petting or vocal praise. Researchers randomly placed 14 dogs into two groups of 7 each. In group A the owner would pet the dog, while in group B the owner would provide local praise. The respondence variable is the time, in seconds, that the dog interacted with its owner and is given as:

Group A: X_i	114	203	217	254	256	284	296
Group B: Y_j	4	7	24	25	48	71	294

with $i = 1, \ldots, n_A$ and $j = 1, \ldots, n_B$, where $n_A = n_B = 7$.

We test:

$$H_0 : f_X(x) = f_Y(x) \quad \text{for all } x \in (-\infty, +\infty)$$
$$H_1 : f_X(x) = f_Y(x - c) \quad \text{for all } x \in (-\infty, +\infty)$$

for an arbitrary constant $c \neq 0$.

a) Compute the test statistic and the p-value for Wilcoxons two-sample test (The Wilcoxon Rank Sum Test). What is your conclusion?

Problem 4

Over the past year, it has been of interest to follow how inflation has developed and for which categories it has increased the most. In a survey within the grocery trade focus was set on baked goods, fresh fruit and vegetables, canned fruit and vegetables, meat and the dairy product category. Within each of the 5 categories, 10 products were picked at random and the percentage price increase in a 2-month period for each product was recorded. In the rest of the assignment, you can assume the following model for the observed percentage price increase:

$$Y_{ij} = \mu_j + \epsilon_{ij}, \quad i = 1, 2, \ldots, 10; \quad j = 1, 2, \ldots, 5,$$

with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and independent.

From the observed data, the following values for the total sum of squares and the treatment sum of squares were computed:

$$\text{SSTOT} = \sum_{j=1}^{5} \sum_{i=1}^{10} (y_{ij} - \bar{y}_{..})^2 = 27.5$$

$$\text{SSTR} = \sum_{j=1}^{5} 10(\bar{y}_{.j} - \bar{y}_{..})^2 = 4.6$$

a) Construct the complete analysis of variance table and perform a test of the null hypothesis that the expected price increase is the same within each category. What is the conclusion with a significance level chosen at 0.05?

The average observed increase for each category was: 0.1% for baked goods, 0.5% for fresh fruit and vegetables, 0.6% for canned fruit and vegetables, 1% for meat and 0.3% for dairy products.

b) Construct a contrast to test the hypothesis that the expected price increase for meat is the same as the average price increase for the other 4 categories. Perform the test. What is the conclusion with a significance level chosen at 0.05?

All numbers in the analysis of variance table calculated from the observed data could have been constructed with information about the values for $\bar{y}_{.j}, j = 1, 2, \ldots, 5$ and $\sum_{j=1}^{5} \sum_{i=1}^{10} y_{ij}^2$.

c) Show that SSTOT can be calculated if this information is given. Which value do you get for $\sum_{j=1}^{5} \sum_{i=1}^{10} y_{ij}^2$ using this data?