

The goal of this project is to assess how correctly you can apply the skills in statistical reasoning, data analysis, and interpretation you learned in class to solve the following problems.

It helps you increase your understanding of real-world problems.

This project requires you to use statistical software R to analyze the data and make a statistical conclusion. To answer each question,

1. Write null and alternative hypotheses when necessary.
2. Write what statistical procedures, for example, two-sample, t-test, one-sample t-test, Chi-square test, one-way ANOVA, regression analysis, etc., you choose for a particular question and why?
3. Produce appropriate graphs to support your claim,
4. Calculate summary statistics
5. Use more rigorous statistical methods to measure the evidence with p-values, and to construct a confidence interval of the parameter of interest if needed.
6. Interpret and decide for each question after calculating all necessary steps.

Write your answer underneath each question, put your graphs, summary statistics, and results, interpretation, and R code. **Document your results with R Markdown and submit your results as html file.**

Your grading will be given based on clarity, organization, the correctness of your answers, correct interpretation, and correct R code. **You won't receive No points if you don't produce your results with R Markdown.**

For this project, use the data set called "NutrionStudy", which you can find on canvas to answer the questions presented below. The NutrionStudy dataset gives nutrition levels in people's blood as well as their eating habits. Use a 5% level of significance.

A dataset with 315 observations on the following 17 variables.

| | |
|---------------|--|
| ID | ID number for each subject in this sample |
| Age | Subject's age (in years) |
| Smoke | Smoker? coded as No or Yes |
| Quetelet | Weight/(Height ²) |
| Vitamin | Vitamin use: coded as 1=Regular, 2=Occasionally, or 3=No |
| Calories | Number of calories consumed per day |
| Fat | Grams of fat consumed per day |
| Fiber | Grams of fiber consumed per day |
| Alcohol | Number of alcoholic drinks consumed per week |
| Cholesterol | Cholesterol consumed (mg per day) |
| BetaDiet | Dietary beta-carotene consumed (mcg per day) |
| RetinolDiet | Dietary retinol consumed (mcg per day) |
| BetaPlasma | Plasma beta-carotene (ng/ml) |
| RetinolPlasma | Plasma retinol (ng/ml) |
| Sex | Coded as Female or Male |
| VitaminUse | Coded as No Occasional Regular |
| PriorSmoke | Smoking status: coded as 1=Never, 2=Former, or 3=Current |

1. Find a 99% confidence interval for the average number of grams of fiber per day that people eat. Interpret the results.
2.
 - a) Construct 95% confidence interval for the effect of gender on smoking.
 - b) Do the sample data provide evidence that the effect of gender on smoking is statistically significant? In other words, Is there evidence of a difference in the percentage of current smokers by gender?
3. Does smoking decrease beta-carotene levels in the blood (betaPlasma)? Justify your answer with confidence interval and a p-value.
4.
 - a) Construct 90% confidence interval for the effect of sex on beta-carotene levels in the blood.
 - b) Does sex have any significant effect on beta-carotene levels in the blood?
5. Are the amount of alcohol consumption and the age of the people significantly related?
6. Which variables listed in the dataset are significantly related to plasma beta carotene?
7. Is there sufficient evidence that the proportions of smoking status (prior-smoke)- Never, former and current are equal?
8. Consider the following variables: fat grams consumed in a day, cholesterol consumed in mg per day, and age in years. Which of the above factors significantly affects the variable calories consumed in a day at a 5% level of significance?
9.
 - a) Estimate the effect of smoking on cholesterol levels and provide a 95% confidence interval for it.
 - b) Is the effect of smoking on cholesterol levels significant? Make your decision based on confidence interval and p-value.
10. Do the sample data provide evidence that vitaminuse affects retinol (micronutrient in the blood) levels? Justify your answer with a p-value.
11. Who (Man or Woman) is more likely to take a vitamin? Is there a significant association between vitaminuse and the sex of the participants?