Homework 2

See general homework tips and submit your files via the course website.

In the first exercise, we will investigate the relationship between peanut allergy and early consumption of peanuts. For exercise 1 use the **peanut** data set defined in **HW2Data.sas**. The **peanut** data contains peanut allergy data for 247 at-risk children. Out of these 247 children, 117 of them were in the intervention group (early consumption of peanuts during infancy), and 130 children were in the control group (complete avoidance of peanuts during infancy). The data showed whether these children were allergic to peanuts or not when they turned 5 years old. The data came from the LEAP study paper[1] from the New England Journal of Medicine.

The second and third exercises will investigate the relationship between different fish species and their weight size. We will use the **fish** data set defined in **HW2Data.sas**. **Fish** data set in the **HW2Data.sas** file is based on the fish data from SAS help library.

The variables are:

- **species**: bream, perch, roach
- **weight**: fish's weight
- **weight_cat:** the weight category of the fish ('1-Light', '2-Medium', '3-Heavy')

Exercise 1
a) Construct a contingency table for **treat** and **allergy**. What's the % of allergic children in the control group, and the % of allergic children in the intervention group?
b) Perform and comment on appropriate tests of association, and interpret the results. Can we use the chi-square tests in this case, why or why not? Is it appropriate to use the Mantel-Haenszel test? Is there a statistically significant association? What's the magnitude of the association?
c) Test (using risk differences) if the control group has a significantly higher rate of having peanut allergy than the intervention group. What's the risk difference and 95% CI?

Exercise 2
a) Analyze the relationship between **species** and **weight_cat** for the **fish** data as given in **HW2Data.sas**, and comment on appropriate tests of associations. Can we use the chi-square tests in this case, why or why not? Is it appropriate to use the Mantel-Haenszel test? Is there a statistically significant association? What's the magnitude of the association?
b) Focus only on bream and perch, and light versus heavy categories. Test whether bream has larger proportion in the heavy weight category than perch and state your conclusion.

Exercise 3

For the fish data set defined in **HW2Data.sas**, consider a one-way ANOVA model to identify weight differences between different fish species. We will skip the normality test for now, so you can proceed without testing normality. Since there's a large variability in weight, we will transform the values of weight using the log function.

a) Create a new variable in the fish data, **lweight**=log(**weight**). Perform a one-way ANOVA for l**weight** with **species** as the categorical predictor. State the null hypothesis, alternative hypothesis, and conclusion. What are the assumptions with one-way ANOVA?

b) Test any assumptions of the model that should be tested (aside from normality, which you do not need to test), comment on the significance of the model, and the variation described by the model.

c) Comment on any significantly different log(weight) means as determined by the best test for comparing all pairwise differences, and explain the difference in log(weight) between pairs of fish species (eg, differences of means, and 95% CIs).

---

[1] "Randomized Trial of Peanut Consumption in Infants at Risk for Peanut Allergy", The New England Journal of Medicine, Du Toit, G., et. al., Vol. 372 No.9, February 26, 2015.