

MA117 - WORKSHEET 9  
NUMERICAL AND CATEGORICAL VARIABLES  
Week 3, Monday

**Problem 1.** An experiment was conducted to measure and compare the effectiveness of various feed supplements on the growth rate of chickens. Newly hatched chicks were randomly allocated into five groups, and each group was given a different feed supplement. Below are some summary statistics about the weight in grams of chickens from each group. Weight distribution within in each group was roughly normal.

	mean (g)	sd (g)	n
casein	323.58	64.43	12
horsebean	160.20	38.63	10
linseed	218.75	52.24	12
meatmeal	276.91	64.90	11
soybean	246.43	54.13	14

- Calculate and interpret a 97% confidence interval for the difference in mean weights of chickens fed linseed vs chickens fed horsebean.
- Calculate and interpret a p-value for the data under the hypothesis that there is no difference in mean weight of chickens fed linseed and chickens fed horsebean.
- Calculate and interpret a p-value for the data under the hypothesis that there is no difference in mean weight of chickens fed casein and chickens fed soybean.
- Can any of your calculations above be used as evidence that a difference in mean weights is caused by the chicken's diet? Explain why or why not.

**Problem 2.** Kenya has two official languages: English and Swahili. These languages coexist with numerous other languages spoken in smaller, more localized communities. Here is some data set about attitudes towards Swahili among 480 schoolchildren.

<https://sagrawalx.github.io/teaching/data/swahili.csv>

This dataset records the following 4 variables about each child in the study.

- **province** where the child lives (either **NAIROBI** or **PWANI**)
- **sex** of the child (either **female** or **male**)
- **attitude.score** measures how positively the child feels about Swahili (from 40 (most negative) to 200 (most positive)).
- **school** is a code for a specific school the child attends (A through L)

For each of the following, be sure to state your hypotheses and conclusions clearly in the context of the data, and to discuss any concerns you have about the applicability of the relevant test.

- Calculate and interpret 95% confidence interval for the difference in mean **attitude.score** between **NAIROBI** and **PWANI**. Hint: If you called your data frame **df**, try running:

```
t.test(attitude.score ~ province, data = df)
```

Or, if you've installed **tidymodels**, you can also run

```
t_test(df, attitude.score ~ province)
```

- (b) How would you calculate a 98% confidence interval for part (a)?
- (c) Calculate and interpret p-value for the data under the hypothesis that there is no difference in mean `attitude.score` between `male` students and `female` students.
- (d) Calculate and interpret a p-value for the data under the hypothesis that there is no difference in mean `attitude.score` between different `schools`. Hint: To calculate standard deviation of `attitude.score` in each `school`, run

```
summarize(group_by(df, school), sd(attitude.score))
```

Then you can run ANOVA using the following:

```
summary(aov(attitude.score ~ school, data = df))
```