

Unit 6 – Analysis of Variance  
Practice Problems (2 of 2)

**Due: Wednesday November 15, 2023**

**Last date to submit for credit (-20 points): Wednesday November 29, 2023**

**Before you begin.** Download from the course website  
lbw.xlsx

*(Source: Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression: Third Edition. These data are copyrighted by John Wiley & Sons Inc. and must be acknowledged and used accordingly. Data were collected at Baystate Medical Center, Springfield, Massachusetts during 1986.)*

Low birth weight is an outcome of concern because of its links to infant mortality and birth defects. A woman's behavior during pregnancy (including diet, smoking habits, and receiving prenatal care) can greatly alter the chances of carrying the baby to term and, consequently, of delivering a baby of normal birth weight. The goal of this study was to identify risk factors associated with giving birth to a low birth weight baby (weighing less than 2500 grams). Data were collected on 189 women, 59 of whom had low birth weight babies and 130 of which had normal birth weight babies.

In this homework, we will use three variables to gain practice in performing a two-way analysis of variance: lbw.xlsx has 189 observations on 3 variables.

**Data dictionary/Codebook**

Position	Variable	Label	Type	Codings
1	id	Identification code		Range: 4, 226
2	race	Race	numeric	1 = white 2 = african american 3 = other
3	ftv	Number of visits to physician during 1 <sup>st</sup> trimester	numeric	Range: 0, 6
4	btw	Birthweight (grams)	numeric	Range: 709, 4990

**Outcome Variable**

Y = btw

**Factor I**

racef, coded: 1, 2 or 3

*Note: you will create this from race in exercise #2*

**Factor II**

no\_trimester1, coded: 0, 1

*Note: you will create this from ftv in exercise #2*

#1.

State the analysis of variance model using notation  $\mu$ ,  $\alpha_i$ ,  $\beta_j$ ,  $(\alpha\beta)_{ij}$  and  $\sigma^2$  as appropriate. Define all terms and constraints on the parameters.

#2.

By any means you like, create the following three new variables

(1) **racef** = factor version of **race**

(2) **no\_trimester1** that is a 0/1 indicator of “no visits in the first trimester and defined as follows:

$$\text{no\_trimester1} = \begin{cases} 1 & \text{if } \text{ftv}=0 \\ 0 & \text{for all other values of } \text{ftv} \end{cases}$$

(3) **no\_trimester1f** = factor version of **no\_trimester1**

#3.

By any means you like, produce descriptive statistics of **Y=bwt**, separately for groups defined by **racef** and **no\_trimester1f**.

#4.

Fit the two-way analysis of variance. Show the analysis of variance table.

#5.

This time, perform the two way analysis of variance as a regression. Show.

#6.

By any means you like, perform a partial F-test of the null hypothesis that, controlling for **racef** and **no\_trimester1f**, the extra predictive significance of the interaction of **racef** and **no\_trimester1f** is zero.

#7.

Obtain the predicted means of **bwt** for each group defined by **racef** and **no\_trimester1f**, in two ways: (1) from the analysis of variance; and (2) from the regression. Verify that they are identical.