

Unit 7 – Analysis of Variance
Practice Problems - Week #12

SOLUTIONS

Consider the little data set of $n=12$ observations on page 62 of Lecture Notes 7. *Introduction to Analysis of Variance*.

Light (LIGHT)	Water Temp (TEMP)	Fish Growth (GROWTH)
1=low	1=cold	4.55
1=low	1=cold	4.24
1=low	2=lukewarm	4.89
1=low	2=lukewarm	4.88
1=low	3=warm	5.01
1=low	3=warm	5.11
2=high	1=cold	5.55
2=high	1=cold	4.08
2=high	2=lukewarm	6.09
2=high	2=lukewarm	5.01
2=high	3=warm	7.01
2=high	3=warm	6.92

SAS program

```
*
*
* PubHlth 640 - 2009
* Week 12 Practice Problems
* Analysis of Variance: 2 Way Fixed Effects
*
* code: week12_2009
* _____;
options nocenter nodate;

/* Dictionary of Discrete Value Codes */
proc format;
  value lightf 1="1=low"
              2="2=high";
  value tempf 1="1=cold"
             2="2=lukewarm"
             3="3=warm";
run;
quit;
*
* _____;
* (1) Create a data set.
* _____;
data temp;
  input light temp growth;
  *
  * ___ note use of logical operators to create 0/1 variables___;
  i_high=(light=2);
  i_luke=(temp=2);
  i_warm=(temp=3);
  hi_luke=(i_high)*(i_luke);
  hi_warm=(i_high)*(i_warm);

  label i_high="I_HIGH: High Light"
        i_luke="I_LUKE: Lukewarm Water"
        i_warm="I_WARM: Warm Water"
        hi_luke="HI_LUKE: High Light/Lukewarm"
        hi_warm="HI_WARM: High Light/Warm";
  format light LIGHTF. temp TEMPF.;
cards;
1 1 4.55
1 1 4.24
1 2 4.89
1 2 4.88
1 3 5.01
1 3 5.11
2 1 5.55
2 1 4.08
2 2 6.09
2 2 5.01
2 3 7.01
2 3 6.92
;
```

```

run;
quit;
*
*
* (2) Two Way Fixed Effects Analysis of Variance
* as an ANALYSIS OF VARIANCE
*
* (4) Obtain means by conditions of light and temp
* MEANS statement gives observed means
* LSMEANS statement gives predicted means from model
* _____;
proc glm data=temp;
  class light temp;
  model growth=light temp light*temp;
  means light*temp;
  lsmeans light*temp;
  title "Question 2";
run;
quit;

*
*
* (3) Two Way Fixed Effects Analysis of Variance
* as a REGRESSION
*
* (4) Obtain means by conditions of light and temp
* OUPUT statement creates new data set
* PREDICTED = gives name to predicted values
* _____;
proc glm data=temp;
  model growth=i_high i_luke i_warm hi_luke hi_warm;
  output out=temp2
         predicted=yhat;
title "Question 3";
run;
quit;
proc sort data=temp2;
  by light temp;
run;
quit;
proc means data=temp2 mean;
  var yhat;
  by light temp;
  title "Question 3";
run;
quit;

*
*
* (5) Compare predicted means w actual means
* _____;
proc means data=temp2 mean;
  var growth yhat;
  by light temp;
  title "Question 5";
run;
quit;

```

```

SAS Output
*
*
* (2) Two Way Fixed Effects Analysis of Variance
* (4) Obtain means by conditions of light and temp
*
The GLM Procedure
Dependent Variable: growth

Source                DF          Sum of
                    Squares      Mean Square    F Value    Pr > F
Model                  5      8.23196667      1.64639333      5.74    0.0276
Error                  6      1.72080000      0.28680000
Corrected Total       11      9.95276667

--- some output omitted ---

The GLM Procedure
Least Squares Means

light    temp          growth
                    LSMEAN
1=low    1=cold      4.39500000
1=low    2=lukewarm  4.88500000
1=low    3=warm      5.06000000
2=high   1=cold      4.81500000
2=high   2=lukewarm  5.55000000
2=high   3=warm      6.96500000

*
*
* (3) Two Way Fixed Effects Analysis of Variance
*      as a REGRESSION
* (4) Obtain means by conditions of light and temp
*
The GLM Procedure
Dependent Variable: growth

Source                DF          Sum of
                    Squares      Mean Square    F Value    Pr > F
Model                  5      8.23196667      1.64639333      5.74    0.0276
Error                  6      1.72080000      0.28680000
    
```

```

--- some output omitted ---

```

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	4.395000000	0.37868192	11.61	<.0001
i_high	0.420000000	0.53553711	0.78	0.4627
i_luke	0.490000000	0.53553711	0.91	0.3955
i_warm	0.665000000	0.53553711	1.24	0.2607
hi_luke	0.245000000	0.75736385	0.32	0.7573
hi_warm	1.485000000	0.75736385	1.96	0.0976

```

--- some output omitted ---
light=1=low temp=1=cold
Analysis Variable : yhat
  Mean
-----
 4.3950000
-----

light=1=low temp=2=lukewarm
Analysis Variable : yhat
  Mean
-----
 4.8850000
-----

light=1=low temp=3=warm
Analysis Variable : yhat
  Mean
-----
 5.0600000
-----

light=2=high temp=1=cold
Analysis Variable : yhat
  Mean
-----
 4.8150000
-----

light=2=high temp=2=lukewarm
Analysis Variable : yhat
  Mean
-----
 5.5500000
-----

light=2=high temp=3=warm
Analysis Variable : yhat
  Mean
-----
 6.9650000
-----

```

```

*
*
* (5) Compare predicted means w actual means
*

```

```

light=1=low temp=1=cold
Variable      Mean
-----
growth        4.3950000
yhat          4.3950000

```

```

light=1=low temp=2=lukewarm
Variable      Mean
-----
growth        4.8850000
yhat          4.8850000

```

```

light=1=low temp=3=warm
Variable      Mean
-----
growth        5.0600000
yhat          5.0600000

```

```

light=2=high temp=1=cold
Variable      Mean
-----
growth        4.8150000
yhat          4.8150000

```

```

light=2=high temp=2=lukewarm
Variable      Mean
-----
growth        5.5500000
yhat          5.5500000

```

```

light=2=high temp=3=warm
Variable      Mean
-----
growth        6.9650000
yhat          6.9650000

```

```

Stata v 10

. * Use DATA > DATA EDITOR to create data set.
. save "/Users/carolbigelow/Desktop/week12_2009.dta"
file /Users/carolbigelow/Desktop/week12_2009.dta saved

. set more off

. * Question 1
. * Create indicators and interactions
. label define lightf 1 "1=low" 2 "2=high"
. label define tempf 1 "1=cold" 2 "2=lukewarm" 3 "3=warm"
. label values light lightf
. label values temp tempf
. generate i_high=(light==2)
. generate i_luke=(temp==2)
. generate i_warm=(temp==3)
. generate hi_luke=i_high*i_luke
. generate hi_warm=i_high*i_warm

. ** Question 2
. ** Obtain two way anova
. ** ouput means
. anova growth light temp light*temp

              Number of obs =      12      R-squared      = 0.8271
              Root MSE      = .535537      Adj R-squared = 0.6830

              Source | Partial SS   df      MS          F      Prob > F
-----+-----
              Model | 8.23196772   5   1.64639354   5.74   0.0276
              |
              light | 2.98003383   1   2.98003383  10.39   0.0181
              temp | 3.9843175    2   1.99215875   6.95   0.0274
              light*temp | 1.26761639   2   .633808197   2.21   0.1909
              |
              Residual | 1.72080044   6   .286800074
              -----+-----
              Total | 9.95276816  11   .904797106

. * Use command PREDICT to obtain the predicted values.
. predict yhat
(option xb assumed; fitted values)
. table light temp,contents(mean yhat)

-----+-----
              |               temp
              |      1=cold  2=lukewarm  3=warm
-----+-----
              +-----+-----+-----
1=low |      4.395      4.885      5.06
2=high |      4.815      5.55      6.965
              +-----+-----+-----
    
```

```

. ** Question 3
. * Perform a regression analysis
. regress growth i_high i_luke i_warm hi_luke hi_warm

-----+-----
Source |           SS       df       MS                Number of obs =      12
-----+-----
Model |  8.23196772       5   1.64639354            F( 5, 6) =      5.74
Residual |  1.72080044       6   .286800074            Prob > F      =  0.0276
-----+-----
Total |  9.95276816      11   .904797106            R-squared     =  0.8271
                                           Adj R-squared =  0.6830
                                           Root MSE    =  .53554

-----+-----
growth |           Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
i_high |   .4200001   .5355372     0.78   0.463   - .8904122   1.730412
i_luke |    .49       .5355372     0.91   0.395   - .8204123   1.800412
i_warm |   .6650002   .5355372     1.24   0.261   - .6454121   1.975412
hi_luke |   .2450001   .7573639     0.32   0.757   -1.608203   2.098203
hi_warm |    1.485     .7573639     1.96   0.098   - .3682029   3.338203
 _cons |    4.395     .378682     11.61   0.000    3.468399   5.321601
-----+-----

. predict yhat_reg
(option xb assumed; fitted values)

. table light temp,contents(mean yhat_reg)

-----+-----
light |           temp
-----+-----
      |   1=cold  2=lukewarm  3=warm
-----+-----
1=low |    4.395    4.885    5.06
2=high |    4.815    5.55    6.965
-----+-----

. ** Question 5
. ** Compare with actual means
. table light temp,contents(mean growth)

-----+-----
light |           temp
-----+-----
      |   1=cold  2=lukewarm  3=warm
-----+-----
1=low |    4.395    4.885    5.06
2=high |    4.815    5.55    6.965
-----+-----

. * save data for later!!
. save "/Users/carolbigelow/Desktop/week12_2009.dta", replace
file /Users/carolbigelow/Desktop/week12_2009.dta saved

. log close
    
```