STAT 525 FALL 2018

# Chapter 20
# Two-Factor Studies with One Case per Treatment

Professor Min Zhang

# One Observation Per Cell

- Do not have enough information to estimate **both** the interaction effect and error variance

- With interaction, error degrees of freedom is $ab(n-1) = 0$

- Common to assume there is no interaction (i.e., pooling)

  - SSE$^*$ = SSAB + 0

  - df$_E^*$=df$_{AB}$ + 0

- Can also test for less general type of interaction that requires fewer degrees of freedom

# Tukey's Test for Additivity

- Consider special type of interaction
- Assume following model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta\alpha_i\beta_j + \varepsilon_{ij}$$

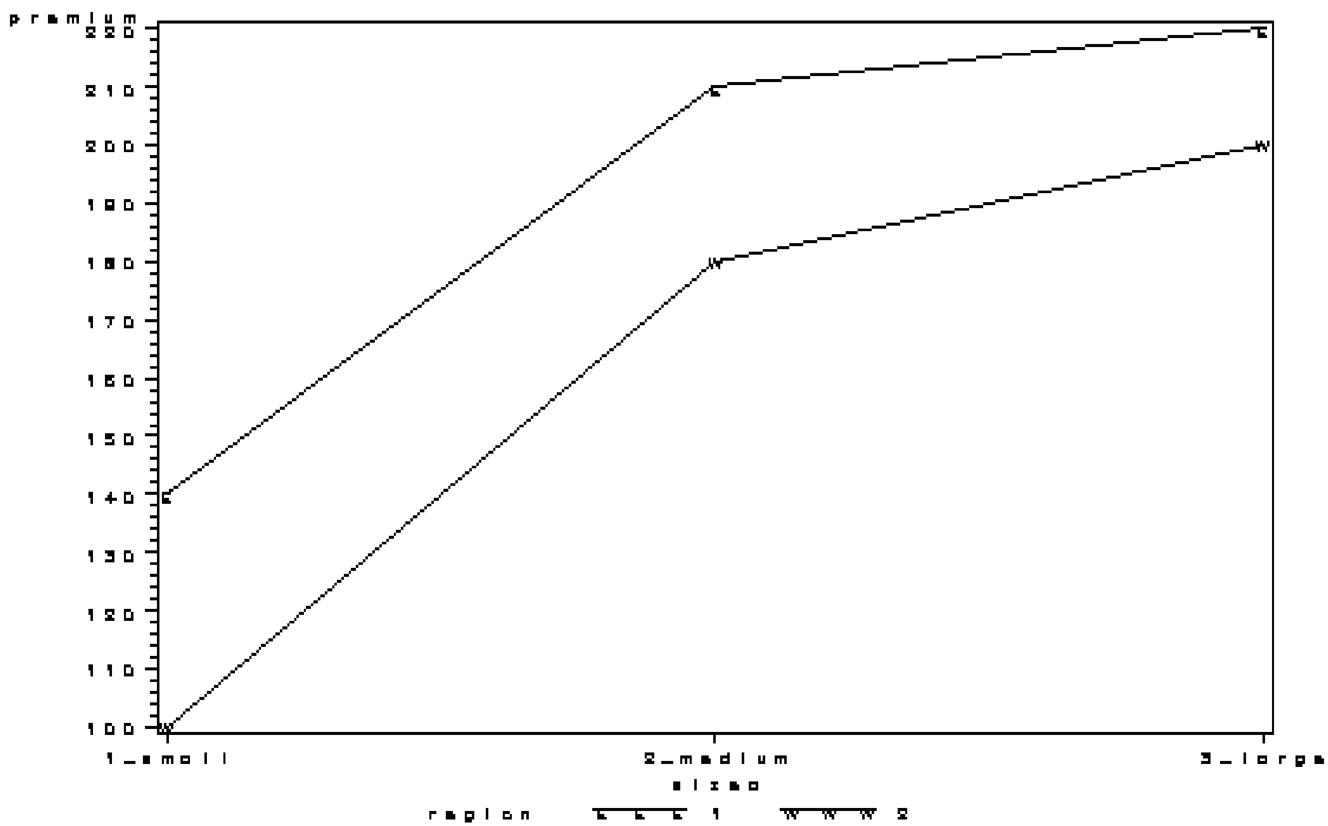
- Uses up only one degree of freedom
- Other variations possible (e.g., $\theta_i\beta_j$)
- Want to test $H_0 : \theta = 0$
- Will use regression after estimating factor effects to test $\theta$

## Example (Page 882)

- $Y$ is the premium for auto insurance
- Factor $A$ is the size of the city
  - $a = 3$: small, medium, large
- Factor $B$ is the region
  - $b = 2$: east, west
- Only one city per cell was observed

```
data a1; infile 'u:\.www\datasets525\CH20TA02.txt';
    input premium size region;
    if size=1 then sizea='1_small ';
    if size=2 then sizea='2_medium';
    if size=3 then sizea='3_large ';

symbol1 v='E' i=join c=black; symbol2 v='W' i=join c=black;
proc gplot data=a1;
    plot premium*sizea=region/frame;
run; quit;
```



```
proc glm data=a1;
    model premium=;
    output out=aall p=muhat;

proc glm data=a1;
    class size;
    model premium=size;
    output out=aA p=muhatA;

proc glm data=a1;
    class region;
    model premium=region;
    output out=aB p=muhatB;

data a2; merge aall aA aB;
    alpha=muhatA-muhat;
    beta=muhatB-muhat;
    atimesb=alpha*beta;

proc print data=a2;
    var size region atimesb;
run; quit;
```

| Obs | size | region | atimesb |
|---|---|---|---|
| 1 | 1 | 1 | -825 |
| 2 | 1 | 2 | 825 |
| 3 | 2 | 1 | 300 |
| 4 | 2 | 2 | -300 |
| 5 | 3 | 1 | 525 |
| 6 | 3 | 2 | -525 |

- These estimates are based on the factor effects model where $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$.
- While not shown, the following were used to compute `atimesb`:
  - $\widehat{\mu} = 175$
  - $\widehat{\mu}_{1.} = 120$
  - $\widehat{\mu}_{2.} = 195$
  - $\widehat{\mu}_{3.} = 210$
  - $\widehat{\mu}_{.1} = 190$
  - $\widehat{\mu}_{.2} = 160$

```
proc glm data=a2;
    class size region;
    model premium=size region atimesb/solution;
run; quit;
                              Sum of
Source              DF       Squares    Mean Square   F Value   Pr > F
Model                4   10737.09677     2684.27419    208.03   0.0519
Error                1      12.90323       12.90323
Corrected Total      5   10750.00000

R-Square     Coeff Var      Root MSE    premium Mean
0.998800      2.052632      3.592106        175.0000

Source              DF     Type I SS    Mean Square   F Value   Pr > F
size                 2   9300.000000    4650.000000    360.37   0.0372
region               1   1350.000000    1350.000000    104.62   0.0620
atimesb              1     87.096774      87.096774      6.75   0.2339

                                      Standard
Parameter           Estimate             Error    t Value    Pr > |t|
Intercept       195.0000000 B   2.93294230      66.49      0.0096
size      1     -90.0000000 B   3.59210604     -25.05      0.0254
size      2     -15.0000000 B   3.59210604      -4.18      0.1496
size      3       0.0000000 B    .                .           .
region    1      30.0000000 B   2.93294230      10.23      0.0620
region    2       0.0000000 B    .                .           .
atimesb          -0.0064516     0.00248323      -2.60      0.2339
```

- The same parameter estimates as the model without the interaction term

# One Quantitative Factor

- Similar to regression with one indicator or categorical variable
- Plot the means vs the quantitative factor for each level of the categorical factor
- Based on this plot,
  - Consider linear/quadratic relationships for the quantitative factor
  - Consider different slopes for the different levels of the categorical factor
  - Can perform lack of fit analysis
- If two quantitative variables, can consider linear and quadratic terms. Interactions modeled as the direct product. Lack of fit test very useful. Again very similar to linear regression models.

## Chapter Review

- Two-Factor Studies with $n_{ij} = 1$
  - No degrees of freedom for interaction
  - Tukey's test for additivity
    * Use only one degree of freedom
    * Can be generalized to use more degrees of freedom
- One or both factors are quantitative
  - A test for interactions effects can be obtained by regression methods
    * Include interactions by taking direct products