STAT 525

Chapter 20
Two-Factor Studies with One Case per Treatment

Dr. Qifan Song

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
  - $\text{df}^*_E = \text{df}_{AB} + 0$
- 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 $H_0 : \theta = 0$

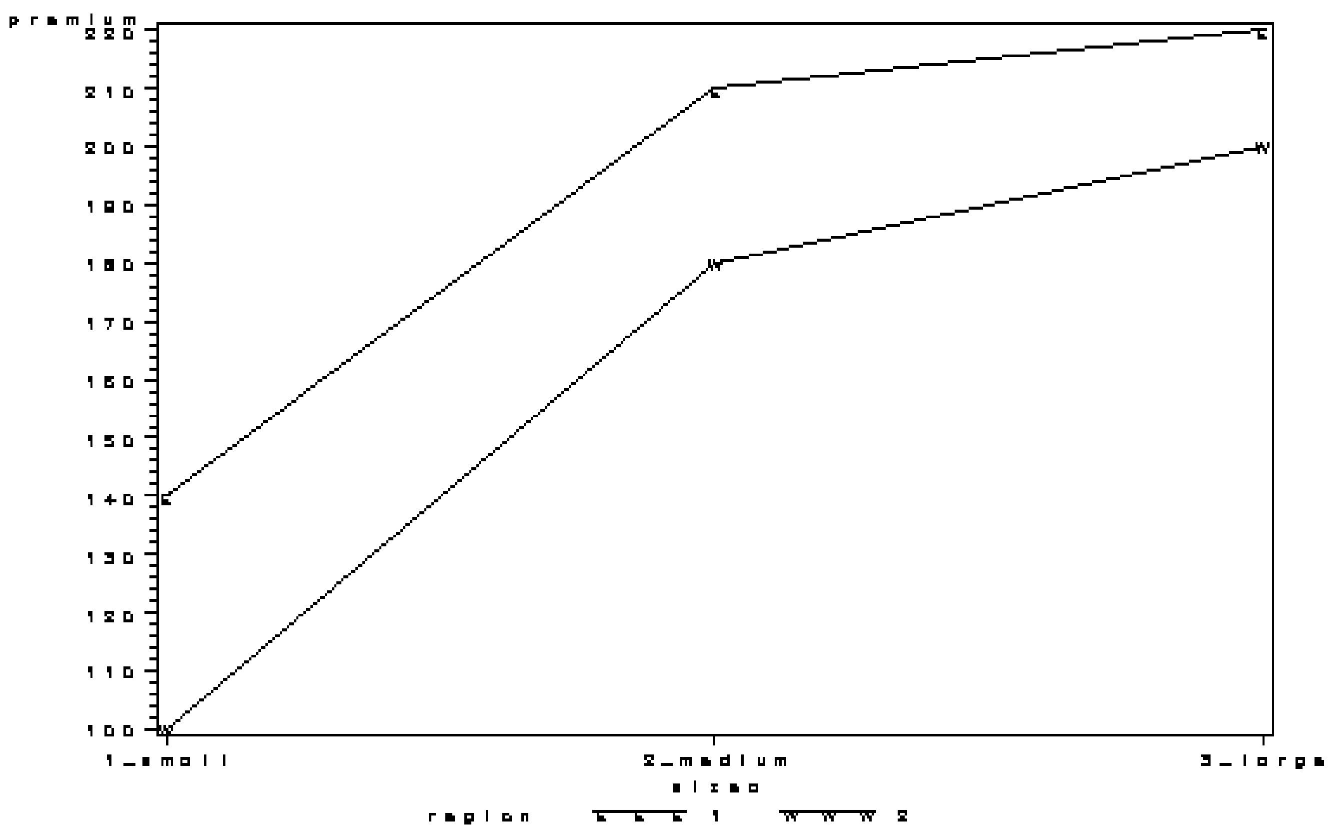
- Will use regression after estimating factor effects to test θ, i.e., $Y_{ij} = \mu + \alpha_i + \beta_j + \theta\widehat{\alpha}_i\widehat{\beta}_j + \varepsilon_{ij}$.
- It is equivalent to model $Y_{ij} = \mu + \alpha'_i + \beta'_j + \theta(\widehat{Y}_{ij})^2 + \varepsilon_{ij}$.

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;
    class size region;
    model premium=size region ;
output out=diag P=yhat;
run;

proc glm data=diag;
    class size region;
    model premium=size region yhat*yhat;
run;
```

```
                              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
yhat*yhat          1       87.096774      87.096774      6.75   0.2339
```

We fail to reject the additive hypothesis.

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