6 Combining continuous and categorical variables

6.2 Combining continuous and categorical variables

To combine continuous and categorical variables, the word equation and continuous variables are specified as before (using the covariate subcommand or the covariates sub-dialog box). Minitab will assume that all variables not specified as continuous are categorical. This is illustrated below with the *leprosy* dataset.

MINITAB COMMANDS FOR BOX 6.1 **A treatment for leprosy**

Commands

```
glm BACAFTER = BACBEF + TREATMT;
covariate BACBEF;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model
`BACAFTER` → Response
`BACBEF + TREATMT` → Model
Covariates…
`BACBEF` → Covariates
Results…
⊙ In addition, coefficients for all terms

MINITAB OUTPUT FOR BOX 6.1 **A treatment for leprosy**

General Linear Model: `BACAFTER` versus `TREATMT`

```
Factor             Type Levels Values
TREATMT            fixed     3 1 2 3

Analysis of Variance for BACAFTER, using Adjusted SS for Tests

Source            DF       Seq SS      Adj SS       Adj MS          F          P
BACBEF             1       587.48      515.01       515.01      35.22      0.000
TREATMT            2        83.35       83.35        41.67       2.85      0.076
Error             26       380.15      380.15        14.62
Total             29      1050.98
```

```
Term                      Coef         SE Coef            T            P
Constant                -0.013           1.806        -0.01        0.994
BACBEF                  0.8831          0.1488         5.93        0.000
TREATMT
1                       -1.590           1.012        -1.57        0.128
2                       -0.726           1.002        -0.72        0.475

Unusual Observations for BACAFTER

Obs             BACAFTER         Fit      SE Fit      Residual        St Resid
13               16.0530      6.6802      1.3434        9.3728           2.62R

R denotes an observation with a large standardized residual
```

With the *fats* dataset, combining continuous and categorical variables can reveal hitherto hidden relationships, as discussed in the main text.

MINITAB COMMANDS FOR BOX 6.2 **Both weight and sex as explanatory variables for fat content**

Commands

```
glm FAT = WEIGHT + SEX;
covariate WEIGHT;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model
`FAT` → Response
`WEIGHT + SEX` → Model
Covariates...
`FAT` → Covariates
Results...
⊙ In addition, coefficients for all terms

MINITAB OUTPUT FOR BOX 6.2 **Both weight and sex as explanatory variables for fat content**

General Linear Model: `FAT` versus `SEX`

```
Factor       Type Levels Values
SEX          fixed     2 1 2

Analysis of Variance for FAT, using Adjusted SS for Tests

Source          DF        Seq SS       Adj SS       Adj MS          F          P
WEIGHT           1         1.328       87.105       87.105      34.00      0.000
SEX              1       176.098      176.098      176.098      68.73      0.000
Error           16        40.995       40.995        2.562
Total           18       218.421
```

```
Term                      Coef        SE Coef           T             P
Constant                13.010          2.679        4.86         0.000
WEIGHT                 0.21715        0.03724        5.83         0.000
SEX
1                       3.9519         0.4767        8.29         0.000

Unusual Observations for FAT

Obs                  FAT           Fit       SE Fit       Residual   St Resid
 13            35.0000       31.9454       0.5769         3.0546      2.05R
 16            26.0000       29.7740       0.5551        -3.7740     -2.51R

R denotes an observation with a large standardized residual
```

6.3 Orthogonality in the context of continuous and categorical variables

This example illustrates orthogonality between a continuous and categorical explanatory variable using the *bacterial growth* dataset.

MINITAB COMMANDS FOR BOX 6.3 Bacterial growth at two levels of lactose

Commands

```
glm BACTERIA = DAY + LACTOSE;
covariate DAY;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model
BACTERIA → Response
DAY + LACTOSE → Model
Covariates…
DAY → Covariates
Results…
⊙ In addition, coefficients for all terms

MINITAB OUTPUT FOR BOX 6.3 **Bacterial growth at two levels of lactose**

General Linear Model: BACTERIA versus LACTOSE

```
Factor         Type Levels Values
LACTOSE        fixed     2 1 2

Analysis of Variance for BACTERIA, using Adjusted SS for Tests

Source              DF      Seq SS      Adj SS      Adj MS          F          P
DAY                  1      297.97      297.97      297.97     130.56      0.000
LACTOSE              1      397.07      397.07      397.07     173.99      0.000
Error               17       38.80       38.80        2.28
Total               19      733.85

Term                       Coef        SE Coef           T           P
Constant                 3.0939         0.7922        3.91       0.001
DAY                      2.7293         0.2389       11.43       0.000
LACTOSE
1                       -4.4557         0.3378      -13.19       0.000

Unusual Observations for BACTERIA

Obs             BACTERIA           Fit      SE Fit      Residual      St Resid
 15              24.1410       21.1964      0.6756        2.9446         2.18R

R denotes an observation with a large standardized residual
```

Orthogonality may then be demonstrated more directly, by using DAY as the response variable.

MINITAB COMMANDS FOR BOX 6.4 **Illustrating orthogonality between continuous and categorical variables**

Commands

```
glm DAY = LACTOSE;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model
DAY → Response
LACTOSE → Model
Results…
⊙ In addition, coefficients for all terms

MINITAB OUTPUT FOR BOX 6.4 Illustrating orthogonality between continuous and categorical variables

General Linear Model: DAY versus LACTOSE

```
Factor         Type Levels Values
LACTOSE        fixed     2 1 2

Analysis of Variance for DAY, using Adjusted SS for Tests

Source            DF       Seq SS       Adj SS       Adj MS         F         P
LACTOSE            1        0.000        0.000        0.000      0.00     1.000
Error             18       40.000       40.000        2.222
Total             19       40.000

Term                      Coef          SE Coef             T             P
Constant                3.0000           0.3333          9.00         0.000
LACTOSE
1                      -0.0000           0.3333         -0.00         1.000
```

6.4 Treating variables as continuous or categorical

Some variables may be legitimately treated as continuous or categorical. DAY in the *bacterial growth* dataset is an example of this.

MINITAB COMMANDS FOR BOX 6.5 Analysing bacterial growth with DAY as a categorical

Commands

```
glm BACTERIA = DAY + LACTOSE;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model
BACTERIA → Response
DAY + LACTOSE → Model
Results…
⊙ In addition, coefficients for all terms

MINITAB OUTPUT FOR BOX 6.5 Analysing bacterial growth with DAY as a categorical variable

General Linear Model: BACTERIA versus DAY, LACTOSE

```
Factor            Type Levels Values
DAY               fixed     5 1 2 3 4 5
LACTOSE           fixed     2 1 2
```

```
Analysis of Variance for BACTERIA, using Adjusted SS for Tests

Source          DF        Seq SS       Adj SS       Adj MS           F         P
DAY              4        298.51       298.51        74.63       27.30     0.000
LACTOSE          1        397.07       397.07       397.07      145.28     0.000
Error           14         38.26        38.26         2.73
Total           19        733.85

Term                      Coef       SE Coef            T            P
Constant               11.2820        0.3697        30.52        0.000
DAY
1                      -5.2729        0.7393        -7.13        0.000
2                      -2.8055        0.7393        -3.79        0.002
3                      -0.2044        0.7393        -0.28        0.786
4                       2.6236        0.7393         3.55        0.003
LACTOSE
1                      -4.4557        0.3697       -12.05        0.000
```

6.7 Exercises

Conservation and its influence on biomass

These data are stored in the *conservation* dataset.

MINITAB COMMANDS FOR BOX 6.7 **Conservation and biomass analysis**

Commands

```
glm BIOMASS = CONS + ALT + SOIL;
covariate ALT;
brief 3.
```

Menu route

Stat > ANOVA > General Linear Model
`BIOMASS` → Response
`CONS + ALT + SOIL` → Model
Covariates…
`ALT` → Covariates
Results…
⊙ In addition, coefficients for all terms

MINITAB OUTPUT FOR BOX 6.7 **Conservation and biomass analysis**

General Linear Model: `BIOMASS` versus `CONS`, `SOIL`

```
Factor       Type Levels Values
CONS         fixed     2 1 2
SOIL         fixed     3 1 2 3
```

```
Analysis of Variance for BIOMASS, using Adjusted SS for Tests

Source          DF       Seq SS       Adj SS       Adj MS          F          P
CONS             1       0.7176       0.0249       0.0249       2.80      0.101
ALT              1       5.8793       4.4273       4.4273     498.10      0.000
SOIL             2       0.3953       0.3953       0.1977      22.24      0.000
Error           45       0.4000       0.4000       0.0089
Total           49       7.3922

Term                    Coef          SE Coef             T             P
Constant             2.21156          0.02486         88.97         0.000
CONS
1                   -0.02443          0.01460         -1.67         0.101
ALT                -0.002907         0.000130        -22.32         0.000
SOIL
1                    0.10574          0.02057          5.14         0.000
2                    0.01952          0.01889          1.03         0.307

Unusual Observations for BIOMASS

Obs         BIOMASS           Fit         SE Fit        Residual        St Resid
 7          0.98000       1.16311       0.03508        -0.18311          -2.09R
18          1.53000       1.34099       0.02896         0.18901           2.11R
23          2.14000       1.94560       0.03599         0.19440           2.23R
50          1.18000       0.98815       0.04317         0.19185           2.29R

R denotes an observation with a large standardized residual
```

Determinants of the Grade Point Average

See Minitab output for this exercise in the answers to exercises in Chapter 14.