Appendix

Solutions to Exercises

Solutions for Chapter 2 Exercises

Exercise 2.1: Box plot of survived and died, *t*-test and CI of difference

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **babies**.

2. **Graph➢Box Plot➢Box Plot**.
   Task Roles: birthweight as the vertical and **status** as the **horizontal** variable; then **Run**.

3. **Analyze➢Anova➢t Test➢Two sample**.
   Task Roles: birthweight is an **Analysis** variable and **status** is the **Group by** variable.

4. **Run**.

Exercise 2.2: Begin with a box plot, then a *t*-test on all data. Repeat the *t*-test after removing the one clear outlier. Perform a nonparametric test.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata** and open **choles**.

2. **Graph➢Box Plot➢Box Plot**.
   Task Roles: **cholesterol** is the vertical, and **type** is the horizontal variable.

3. **Run**.

4. **Analyze➢Anova ➢t Test➢Two sample**.
   Task Roles: **cholesterol** is an **Analysis variable** and **type** is the **Group by** variable.

5. **Run**.

6. **Data➢Filter and Query➢Filter Data➢Add a new filter➢Cholesterol➢less than 420. Run**.

7. Repeat *t*-test on the resulting data set and then perform a Wilcoxon test.

8. **Analyze➢Anova ➢Nonparametric One-Way ANOVA**.
   Task Roles: **Cholesterol** is **Dependent**, and **type** is **independent**.

9. **Analysis: Select only Wilcoxon.**

The results confirm those from the *t*-test.

Exercise 2.3: Calculate a variable for the change, do a box plot of the change, then a paired *t*-test.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata** and open **diet**.

2. **Data➢Filter and Query➢Computed Columns➢New➢Build Expression➢ final - baseline➢OK**.

3. Rename **Calculation1** to **Change** and **Run**.

4. **Describe➢Distribution Analysis**.
   Task Roles: **Change** is the **Analysis variable.**
   Plots. Select **Box plot**.

5. **Run**.

6. The output from **Distribution analysis** of the **change** variable gives the equivalent of paired *t*-test. Compare the Student's t result under **Tests for Location: Mu0=0** with the output from the paired *t*-test below.

7. **Analyze➢ANOVA➢t Test➢Paired**.
   Task Roles: **baseline** and **final** are the two paired variables.

There is little evidence that the diet affects triglyceride levels.

Solutions for Chapter 3 Exercises

Exercise 3.1: This is a similar question to that posed of the horse racing data, but here the data are in the form of counts.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **crash**.

2. **Describe➢One-Way Frequencies**.
   Task Roles: **quarter** is the **Analysis variabl**e and **N** is the **Frequency count**.
   Statistics: **Chi-square goodness of fit**, select **Asymptotic test** and **Exact p-values**.

3. **Run**.

There is no evidence to support quarterly variation.

Exercise 3.2: Apply a chi-squared test.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **fear**.

2. **Describe➢Table Analysis**.
   Task Roles: **Fear** and **sex** are **Table variables**, **num** is the **Frequency count**.
   Tables: drag **fear** to the row and **sex** to the column position.
   Table Statistics➢Association➢Tests of Association: select **Chi-square tests**.

3. **Run**.

Greatest fear is clearly not independent of sex.

Exercise 3.3: As the numbers with suicidal feelings are small look at the results of the Fisher's exact test and compare them with those from chi-squared test.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **suicidal**.
2. **Describe➢Table Analysis**.
   Task Roles: **type** and **suicidal** are **Table variables**, **num** is the **Frequency count**.
   Tables: Drag **type** to the column and **suicidal** to the row position.
   Table Statistics➢Association➢Tests of Association: select **Chi-square tests** and **Fisher's exact test for r x c tables**.
3. **Run**.

Neither result suggests an association.

Exercise 3.4: The process of matching pairs will mean that the members of a pair are not independent. Apply McNemar's test.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata** and open **cancer**.
2. **Describe➢Table Analysis**.
   Task Roles: **caseuse** and controluse are **Table variables**, and **num** is the **Frequency count**.
   Tables: Drag **caseuse** to the column and **controluse** to the row position.
   Table Statistics➢Agreement, select **Measures.**
3. **Run**.

The highly significant result provides strong evidence for an association.

Solutions for Chapter 4 Exercises

Exercise 4.1: Scatterplots of mortality against latitude and longitude and correlation coefficients

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **mortality**.
2. **Graph➢Scatter Plot.**
   Scatter Plot: **2 D scatter plot.**
   Task Roles: **Mortality** is **vertical** and **latitude** is **horizontal**.
3. **Run**.
4. Open **Scatter Plot** task (double-click, right-click **Open**).
   Task Roles: **longitude** is **horizontal**.
5. **Run**.
6. **No** to **Would you like to replace the results from the previous run?**
7. **Analyze➢Multivariate➢Correlations.**
   Task Roles: **mortality**, **latitude**, and **longitude** are **Analysis variables**.
8. **Run**.

Exercise 4.2: Scatterplot of the expired ventilation and oxygen uptake data with fitted *quadratic* curve

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **anaerob**.
2. **Graph➢Scatter Plot.**
   Scatter Plot: **2 D scatter plot**
   Task Roles: **airout** is **vertical** and **o2in** is **horizontal**.
   Appearance➢Interpolations➢Interpolation Method: select **Regression,** select **Quadratic** as **Type**.
3. **Run**.

Exercise 4.3: To plot both indexes over time, a multiple line plot is needed.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **index**.
2. **Graph➢Line Plot**
   Line Plot: **Multiple vertical line plots using overlay**
   Task Roles: **year** is **horizontal, Food** and **House** are **vertical**.
   Appearance➢Plots could be used to distinguish the two series.
3. **Run**.

4. Rerun specifying a custom chart size under **Chart Area**.

Solutions for Chapter 5 Exercises

Exercise 5.1: Begin with summary statistics for each cell, then apply anova, try a log transform, and repeat the anova.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata** and open **rats**.
2. **Describe➢Summary Tables**.
   Task Roles: **hours** is an **Analysis variable**, and **Poison** and **treatment** are **Classification variables**.
   Summary Tables: drag **hours** to the column position, **treatment** to the top edge of the **hours** box, and **Poison** to the row position. Drag **Mean** and **StdDev** below **hours**.
3. **Run**.
4. **Analyze➢ANOVA➢Linear Models.**
   Task Roles: **hours** is the **Dependent variable,** and **Poison**, and **treatment** are **Classification variables.**
   Model: Select **Poison** and **treatment**; click **Factorial**.
5. **Run**.
6. **Data➢Filter and Query➢Computed Columns➢New➢Build Expression**, type **Log(hours)** or use the **Functions** and **Data boxes,** click **OK**, and rename **Calculation1** to **loghours**.
7. **Run**.
8. Repeat the **Linear Models** task above.

Exercise 5.2: Construct box plots. Are there any outliers?

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **knee**.
2. **Graph➢Box Plot**.
   Task Roles: **Distance** is **vertical** and **angle** is **horizontal**.
3. **Run**.

Apply Anova. Try out two multiple comparison methods.

1. **Analyze➢ANOVA➢One-Way ANOVA**.
   Task Roles: **distance** is the **Dependent variable,** and **angle** is the **Independent variable**.
   Means➢Comparison: Select **Scheffe's** and **Tukey's** tests.

2. **Run**.

Exercise 5.3: Initial Anova with interaction diagrams

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **hypertension**.

2. **Analyze➢ANOVA➢Linear Models.**
   Task Roles: **bp** is the **Dependent variable,** and **drug, diet,** and **biofeed** are **Classification variables**.
   Model: Select all three: **drug**, **diet**, and **biofeed;** then click **Factorial**.
   Plots➢Means: Dependent means for **two-way effects**.

3. **Run**.

Try log transform.

1. **Data➢Filter and Query➢Computed Columns➢New➢Build Expression**, type **Log(bp)** or use the **Functions** and **Data boxes**, click **OK**, and rename **Calculation1** to **logbp**.

2. **Run**.

3. Repeat the **Linear Models** task above.

Exercise 5.4: This is an unbalanced design so try entering the variables in different orders using Type I SS.

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **genotypes**.

2. **Analyze➢ANOVA➢Linear Models.**
   Task Roles: **weight** is the **Dependent variable,** and **litter** and **mother** are **Classification variables**.
   Model: select **litter,** click **Main**, then select **mother**, and click **Main**.

3. **Run**.

4. Repeat above but select and enter **mother**, and then **litter**.

Solutions for Chapter 6 Exercises

Exercise 6.1: Fit price and then fit price and temperature comments on change in model SS, etc. Import the data, as described, if necessary.

1. **Analyze➤Regression➤Linear.**
   Task Roles: **consumption** is the **Dependent variable**, and **price** is the only **explanatory variable**.
2. **Run**.
3. Repeat with both **price** and **temperature** as explanatory variables.

Exercise 6.2: Remove 1 and 15. Compare results with those in text. Import the data, as described, if necessary.

1. **Data➤Filter and Query➤Filter Data**, drag **rainfall** to the filter pane, select **Less than** as the **Operator**, and type **11** for the **Value**. Click **OK** and **Run**.
2. Repeat the analysis described in the chapter on the resulting data set.

Exercise 6.3: Fit two models and two diagrams.

1. **File➤Open➤Data➤Local Computer**, browse to **c:\saseg\sasdata**, and open **fat**.
2. **Graph➤Line Plot➤Multiple line plots by group column.**
   Task Roles: **bodyfat** is **vertical**, **age** is **horizontal**, **sex** is **group**.
   Appearance➤Interpolations: Select **Scatter** for both **M** and **F**.
3. **Run**.

As **sex** is a character variable, the **Linear Models** task will be more convenient than the **Linear regression** task.

1. **Analyze➤ANOVA➤Linear Models.**
   Task Roles: **bodyfat** is the **Dependent variable**, **age** is a **Quantitative variable**, and **sex** is a **Classification variable**.
   Model: Select **age** and **sex**, and click **Main**.
   Predictions: Select **Original sample** as the **Data to predict**.
2. **Run**.

Select the resulting predictions data set and produce another scatterplot as before with the predicted values of bodyfat.

1. **Graph➢Line Plot➢Multiple line plots by group column.**
   Task Roles: **predicted_bodyfat** is **vertical**, **age** is **horizontal, sex** is **group**.
   Appearance➢Interpolations: select **Scatter** for both **M** and **F**.

2. **Run**.

3. Reopen the **Linear models** task (double-click, or right-click **Open**).
   Model: Select **age** and **sex**, and click **Cross**.

4. **Run**. Do not replace the results of previous run.

Repeat the graph with the predicted values from this model.

Exercise 6.4: Import the data as described in Chapter 1.

Begin with scatterplots of the response variable against explanatory variables.

1. **Graph➢Scatter Plot➢2 D Scatter Plot.**
   Task Roles: **so2** is **vertical**, and **temperature** is **horizontal**.

2. **Run**.

3. Repeat with remaining variables as horizontal variables.

There is one clear outlier both in terms of number of factories and population size. Remove this before proceeding.

1. **Data➢Filter and Query➢Filter Data**. Drag **population** to the filter pane, select **Less than** as the **Operator**, and type **3000** for the **Value**. Click **OK** and **Run**.

Using the resulting data set, examine the correlation matrix of explanatory variables.

1. **Analyze➢Multivariate➢Correlations.**
   Task Roles: all except **so2** are analysis variables.

2. **Run**.

Population and **factories** are highly correlated. Use only **factories** in the regression.

Ensure that the data set with the outlier removed is the active data set (click on its icon).

1. **Analyze➤Regression➤Linear.**
   Task Roles: **so2** is the **Dependent variable**, and the remaining variables are explanatory variables (omitting **population**).

2. **Run**.

Solutions for Chapter 7 Exercises

Exercise 7.1: Fit logistic model

1. **File➤Open➤Data➤Local Computer**, browse to **c:\saseg\sasdata**, and open **plasma**.

2. **Analyze➤Regression➤Logistic**.
   Task Roles: **esr** is the **Dependent variable** (change the Sort order to Descending), and **gamma** and **fibrinogen** are **quantitative variables**.
   Model➤Effects: enter the main effects of both **fibrinogen** and **gamma**.
   Model➤Selection: choose **Backward elimination**.
   Predictions: select **Original sample** as the **Data to predict**.

3. **Run**.

Using the resulting data set, graph the predicted probabilities and the original data.

1. **Graph➤Line Plot➤Multiple vertical column line plots using overlay.**
   Task Roles: **fibrinogen** is **horizontal** and **esr** and **IP_1** are **vertical**.
   Appearance➤Interpolations: **esr** is **Scatter**, **IP_1** is **Line**. Check **Sort the data by the Horizontal column before plotting.**

2. **Run**.

Exercise 7.2: Logistic model with an interaction

1. **File⮚Open⮚Data⮚Local Computer**, browse to **c:\saseg\sasdata**, and open **leukaemia2**.

2. **Analyze⮚Regression⮚Logistic.**
   Task Roles: **survived** is the **Dependent variable** (change the **Sort order** to **Descending**), **whitecells** is a **Quantitative variable**, and **AG** is a **classification variable**.
   Model⮚Effects: Enter the main effects of both plus their interaction.
   Model⮚Selection: Choose **Backward elimination**.

3. **Run**.

Exercise 7.3: Logistic model for Low Infant Birthweight data

1. **File⮚Open⮚Data⮚Local Computer**, browse to **c:\saseg\sasdata**, and open **lowbwgt**.

2. **Analyze⮚Regression⮚Logistic.**
   Task Roles: **low** is the **Dependent variable** (change the **Sort order** to **Descending**), **age**, **lwt**, and **ftv** are **Quantitative variables**, and **race** is a **Classification variable** (choose **Reference** as the **Coding style**).
   Model⮚Effects: enter the main effects of all predictors.

3. **Run**.

Solutions for Chapter 8 Exercises

Exercise 8.1: Survival functions for breast cancer data

1. **File⮚Open⮚Data⮚Local Computer**, browse to **c:\saseg\sasdata**, and open **lowbwgt**.

2. **Analyze⮚Survival Analysis⮚Life Tables.**
   Task Roles: **days** is the **Survival time**, **censor** is the **censoring variable** with a value of 1 indicating censoring, and **Metastasized** is a **Strata variable**.
   Methods: **Product Limit** is the default.
   Plots: select **Show survival function plot**, **Show censored values**, and **Overlay strata on a single plot**.

3. **Run**.

Exercise 8.2: Cox regression of prostate cancer data

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **prostate**.
2. **Analyze➢Survival Analysis➢Proportional Hazards.**
   Task Roles: **time** is the **Survival time**, **status** is the **censoring variable** with a value of 0 indicating censoring; **treatment**, **age**, **haem**, and **gleason** are **Explanatory variables**.
3. **Run**.

Exercise 8.3: Cox regression of heroin treatment times

1. **File➢Open➢Data➢Local Computer**, browse to **c:\saseg\sasdata**, and open **heroin**.
2. **Analyze➢Survival Analysis➢Proportional Hazards.**
   Task Roles: **time** is the **Survival time**, **status** is the **censoring variable** with a value of 0 indicating censoring, **prison** and **dose** are **Explanatory variables**, and **clinic** is a **Strata variable.**
   Methods: select **Compute confidence limits for hazard ratio.**
   Plots: select **Survival function**.
3. **Run**.