

## Übungsserie 1

Nachdem Sie sich im Windows XP eingeloggt haben, klicken Sie links unten auf “Start”, fahren auf “All Programs”, dann auf “Statistics” und schliesslich auf “SYSTAT 10”. Starten Sie Systat mit einem Klick auf das “SYSTAT 10”-Symbol.

Dateien, die wir Ihnen zur Verfügung stellen, finden Sie im Verzeichnis

**Z:\users\s\sfs\datasets\Systat.**

Eigene Daten können Sie im Verzeichnis **T:** ablegen.

1. Die Datei **orangen.syd** enthält Daten über die Beurteilung von 8 Orangensaftvarianten durch 62 KonsumentInnen, wobei die Bewertungsskala von 0 (sehr schlecht) bis 29 (sehr gut) reicht.

Um die Datei zu laden, im Menü **File / Open / Data** das richtige Verzeichnis aussuchen (s.o.) und **orangen.syd** auswählen.

- a) Wie sind die Daten im **Data** Fenster organisiert? Berechnen Sie den Mittelwert jeder Variante.

Im Menü **Statistics / Descriptive Statistics / Basic Statistics** alles ausser **mean** ausschalten und alle Variablen ausser **CONS** (warum?) aktivieren mittels **Add**→.

- b) Erzeugen Sie ein Histogramm der Bewertung der Variante D.

Mit dem **Quickbutton Histogram** erhalten wir das Histogrammenü. Die **X-Variable** bestimmt, über welche Spalte das Histogramm erstellt wird. Wenn wir die **Y-Variable** leer lassen, wird dort automatisch (**density**) ( $\approx$  Häufigkeit) eingesetzt, was wir gerade brauchen. Setzen Sie im Menü **Axes / X-Axis** das Maximum auf 29.5 und das Minimum auf -0.5 und im Menü **Axes / Y-Axis** das Maximum auf 20.

Welche Balkenbreite wurde automatisch gewählt? Ist diese sinnvoll?

Doppelklicken Sie auf die Grafik. Ein separates Grafikenster erscheint. Rechts unten können Sie nun die Anzahl Balken ablesen und auch verändern. (Dies funktioniert nur, wenn im Menü **View** der **Dynamic Explorer** eingeschaltet ist.)

Verändern Sie die Balkenbreite auf folgende Werte: 1, 2, 3, 5 und 6, und vergleichen Sie die entstehenden Histogramme. Welche Balkenbreite ist vorzuziehen? Kurze Begründung!

- c) Erzeugen Sie in einer Grafik Histogramme für alle acht Orangensaftvarianten, einmal mit freier Skalierung der Y-Achse und einmal mit fixer Skalierung für alle Histogramme. Welche der beiden Grafiken eignet sich besser, um die verschiedenen Varianten miteinander zu vergleichen? Kurze Begründung.

Im Histogrammenü alle Sorten als X-Variablen einsetzen, max. und min. der x-Achse wie oben setzen, im Menü **Axes / Y-Axis** das Maximum löschen und in **Options / Histogram Options** Number of bars auf 10 setzen. Für den Plot mit fixer Skalierung nochmals dasselbe, aber im Menü **Axes / Y-Axis** das Maximum auf 24 setzen.

d) Wie sieht das Histogramm der im Durchschnitt verteilten Noten aus?

Erstellen Sie eine neue Variable **MITTEL**, die für jede/n Konsumenten/in die durchschnittlich erteilte Note angibt. Im Menü **Data / Transform / Let 'MITTEL=(A+C+...)/...'** oder **'MITTEL=AVG(A,C,...)'**, dann mit **Quickbutton Histogram** darstellen.

2. Die Datei **wieners.syd** enthält Daten über die Konsistenz von “Wienerlis”.

8 PrüferInnen ( **PRUEFER**) beschreiben die Knackigkeit ( **KNACK**) von 4 verschiedenen Sorten ( **SORTE\$: A, B, C und D**) auf einer Skala von 0 (extrem weich) bis 15 (extrem knackig).

a) Die Tabelle in diesem Beispiel ist anders aufgebaut als im vorherigen (“univariat” statt “multivariat”). Skizzieren Sie die Tabelle, wie sie im multivariaten Format aussehen würde.

b) Berechnen Sie den Mittelwert der verschiedenen Sorten. Welche Sorte scheint am knackigsten zu sein?

Will man jede Sorte separat bearbeiten (in unserem Fall: mitteln), muss man die Daten in Gruppen zusammenfassen. Im Menü **Data** finden wir den Eintrag **By groups ...**. Gefragt ist nun die Gruppierungsvariable: alle Beobachtungen mit demselben Wert der Gruppierungsvariable werden derselben Gruppe zugeordnet.

Die Berechnung des Mittelwerts erfolgt im Menü **Statistics / Descriptive Statistics / Basic Statistics**.

c) Erzeugen Sie je ein Histogramm von den verschiedenen Sorten. Welche Balkenbreite ist als ideal zu erachten?

Histogrammenü mittels **Quickbutton Histogram** aufrufen und in **Axes / X-Axis** Maximum auf 15.5 und Minimum auf -0.5 setzen. Wenn die Gruppierungsvariable immer noch gesetzt ist, braucht bloss noch die X-Variable angegeben zu werden. Wenn wir alle vier Diagramme nebeneinander sehen wollen, dann müssen wir zuerst im **Data / By groups ...** die Gruppierung ausschalten mittels anklicken von **Turn by groups off** und dann im Histogrammenü im Feld **Grouping variable** die Gruppierungsvariable eintragen, sowie wiederum max. und min. der x-Achse setzen.

d) Inwiefern unterscheiden bzw. gleichen sich die verschiedenen PrüferInnen? Tragen Sie zur Beantwortung dieser Frage die Knackigkeit in einem Plot gegen die PrüferInnen auf.

Ein Scatterplot erlaubt uns, die Werte zweier Variablen gegeneinander aufzutragen (**Quickbutton Scatterplot**). In unserem Fall ist die Knackigkeit in y-Richtung gefragt und die PrüferInnen in x-Richtung. Achtung: Gruppierung muss ausgeschaltet sein (s.o.).

Im obigen Plot ersehen wir, welche 4 Werte der oder die PrüferIn verteilt hat, aber wir können nicht sehen, welche Sorte zum jeweiligen Wert gehört. Eine Abhilfe besteht darin, die Punkte mit ihrer Sorte zu beschriften.

Erneut das Scatterplotmenü aufrufen und **Appearance / Symbol and Label...** auswählen. Im untersten Subfenster **Display Case Labels** anklicken und die Variable angeben, mit deren Wert die Punkte beschriftet werden sollen.

Wir können nun zwar die verschiedenen Punkte eindeutig einer Prüferin und einer Sorte zuordnen, aber die Darstellung ist immer noch wenig übersichtlich. Statt die Punkte zu beschriften wollen wir verschiedene Sorten in verschiedenen Farben darstellen.

Im Menü **Appearance / Color and Fill...** erlaubt uns Systat, im Subfenster **Element Color** eine Variable zu selektieren, deren Wert bestimmt, in welcher Farbe der Punkt eines Datensatzes dargestellt wird.

Zu diesem Zweck müssen wir eine neue Variable erstellen, z.B. **FARBE**. Wir wollen

nun den verschiedenen Sorten folgende Farbwerte zuweisen: “A”→1, “B”→2, “C”→4 und “D”→5 (wir lassen den Wert 3 aus, da Systat für 2 und 3 einen sehr ähnlichen Grünwert produziert). Um **FARBE** die richtigen Werte zuzuweisen, verwenden wir das Menü **Data / Transform / If then Let** in der Form ‘If  $SORTE\$=“A”$  Then  $FARBE=1$ ’ und entsprechend für die restlichen Sorten (Achtung: die Anführungszeichen nicht vergessen!). Dann erneut das Scatterplotmenü aufrufen, in **Appearance / Symbol and Label...** das **Display Case Labels** ausschalten und unter **Symbols** die richtige Variable eingeben. Schliesslich in **Appearance / Color and Fill.../Element Color** die Variable **FARBE** auswählen.

Speichern Sie die Daten in Ihrem Verzeichnis als **wieners.syd** ab.

**Bemerkung:**

Serien, Musterlösungen und Organisationsauskünfte sind auch unter “<http://stat.ethz.ch/teaching/lectures>” zu finden.

**Punkteverteilung:**

- Aufgabe 1:**
- a) (1 Punkt) Mittelwerte der Variablen C, F und L.
  - b) (2 Punkte) Automatische und sinnvollste Balkenbreite.
  - c) (1 Punkt) Kurze Begründung.
  - d) (1 Punkt) Kurze Beschreibung.
- Aufgabe 2:**
- a) (1 Punkt) Tabellenskizze.
  - b) (1 Punkt) Mittelwerte der verschiedenen Sorten.
  - c) (1 Punkt) Ideale Balkenbreite.
  - d) (2 Punkte) Interpretation der Grafik.

**Abgabe:** 1 Woche nach der Übungsstunde in der Vorlesung.

**Präsenz:** Montag, 12:00-13:00, LEO C12.1.