

Курсовую практикум

# **SAS Student Series: SAS® Enterprise Guide, SAS BASE, SAS Macro & SQL**

Практические  
материалы курса

*SAS Student Series: SAS® Enterprise Guide, SAS BASE, SAS Macro & SQL Course Notes* was developed by Stacey Syphus. Additional contributions were made by Richard Bell and Davetta Dunlap. Editing and production support was provided by the Curriculum Development and Support Department.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

**SAS Student Series: SAS® Enterprise Guide, SAS BASE, SAS Macro & SQL Course Notes**

Copyright © 2015 SAS Institute Inc. Cary, NC, USA. All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

Book code STE0020, course code EGRU1/PGRU2/PGRU3/MSRU4, prepared date 03Mar2020.

# Содержание

<b>День 1. SAS(R) Enterprise Guide(R): Essentials .....</b>	<b>7</b>
Введение в SAS Enterprise Guide	7
<b>Обзор курса</b>	<b>15</b>
Рабочая область SAS Enterprise Guide	20
Упражнения	26
<b>Сведения о структуре данных SAS</b>	<b>27</b>
<b>Доступ к данным SAS и DBMS</b>	<b>37</b>
Создание библиотеки SAS	46
Упражнения	48
<b>Импорт файлов с данными</b>	<b>49</b>
Использование задачи «Импорт данных»	52
Использование задачи «Импорт данных» (продолжение)	55
Упражнения	58
<b>Задачи и мастера настроек</b>	<b>59</b>
<b>Создание частотного отчёта</b>	<b>69</b>
Создание одномерного частотного отчёта	74
Упражнения	76
<b>Создание вывода в форматах HTML, PDF и RTF</b>	<b>77</b>
Изменение формата результатов	82
<b>Создание распечатки (Списочный отчёт)</b>	<b>85</b>
Использование задачи «Списочный отчет» (List Data) для создания отчета	87
<b>Фильтрация данных в задаче</b>	<b>91</b>
Фильтрация данных в задаче	93
Упражнения	97
<b>Создание графика</b>	<b>99</b>
Создание столбчатой диаграммы	103
Упражнения	108

<b>Фильтрация и сортировка данных</b>	<b>109</b>
Выбор столбцов и фильтрация строк	114
Упражнения	125
<b>Создание новых столбцов при помощи выражений</b>	<b>127</b>
Создание столбцов с помощью выражений	133
Упражнения	135
<b>Группировка и агрегация данных в запросе</b>	<b>137</b>
Сведение и фильтрация по группам	141
Упражнения	143
<b>Объединение таблиц</b>	<b>145</b>
Объединение таблиц	152
Упражнения	155
<b>Создание отчёта Сводная статистика (Summary Statistics)</b>	<b>157</b>
Создание отчёта Сводная статистика	160
Упражнения	163
<b>Создание Сводного отчёта (Summary Report)</b>	<b>165</b>
Создание сводного табличного отчета	167
Упражнения	172
<b>День 2. SAS(R) Base: Essentials</b>	<b>173</b>
Выполнение программы SAS в SAS Studio	173
Выполнение программы SAS в SAS Enterprise Guide	174
Выполнение программы SAS в оконном интерфейсе SAS	175
Упражнения. Создание данных курса.	176
Диагностика и исправление синтаксических ошибок	206
Упражнения. Исправление непарных кавычек	207
Упражнения. Изучение компоненты данных набора данных SAS	218
Просмотр библиотек SAS при помощи программы	219
Упражнения. Создание и просмотр содержимого библиотеки	221
Упражнения. Создание набора данных SAS	229

Упражнения. Выбор наблюдения по двум условиям	230
Упражнения. Создание новых переменных	231
Упражнения. Использование условной обработки	232
Упражнения. Подсчет уровней переменной	234
Упражнения. Создание сводного отчета	236
<b>Приложение (День 2) .....</b>	<b>237</b>
Задача 1.	237
Задача 2*.	237
Задача 3.	238
Задача 4.	238
Задача 5.	239
Задача 6.	239
Задача 7.	239
Задача 8*.	239
<b>День 3. SAS(R) Base: Data Preparation.....</b>	<b>241</b>
Изучение ошибок данных	241
Упражнения. Чтение текстового файла с разделителем	242
Использование списочного ввода: важность модификатора "двоеточие"	243
Упражнения	245
Упражнения. Чтение текстового файла с пропущенными значениями	247
Упражнения. Конкатенация наборов данных.	248
Упражнения. Слияние наборов данных.	250
Упражнения. Использование параметра IN=	251
Упражнения. Выполнение вычислений в циклах DO	252
Упражнения. Вычисления в циклах DO с условием	253
Упражнения. Использование массивов	254
Упражнения. Массивы для табличного поиска	255

<b>День 4. SAS(R) Macro Language and SQL: Essentials .....</b>	<b>257</b>
Exercises. Writing Text to the SAS Log	257
Exercises. Automatic Macro Variables	258
Exercises. Defining and Using Macro Variables	259
Exercises. Consecutive Macro Variable References	260
Exercises. Macro Functions	261
Exercises	262
Macros with Positional Parameters	263
Macros with Keyword Parameters	264
Macros with Mixed Parameter Lists	265
Exercises. Defining and Using Macro Parameters	267
Exercises. Quering a Table	268
Exercises. Eliminating Duplicates	269

# День 1. SAS(R) Enterprise Guide(R): Essentials

## Введение в SAS Enterprise Guide

**Цели**

- Познакомиться с возможностями и основными функциями SAS Enterprise Guide.
- Изучить как работает Enterprise Guide в различных конфигурациях.

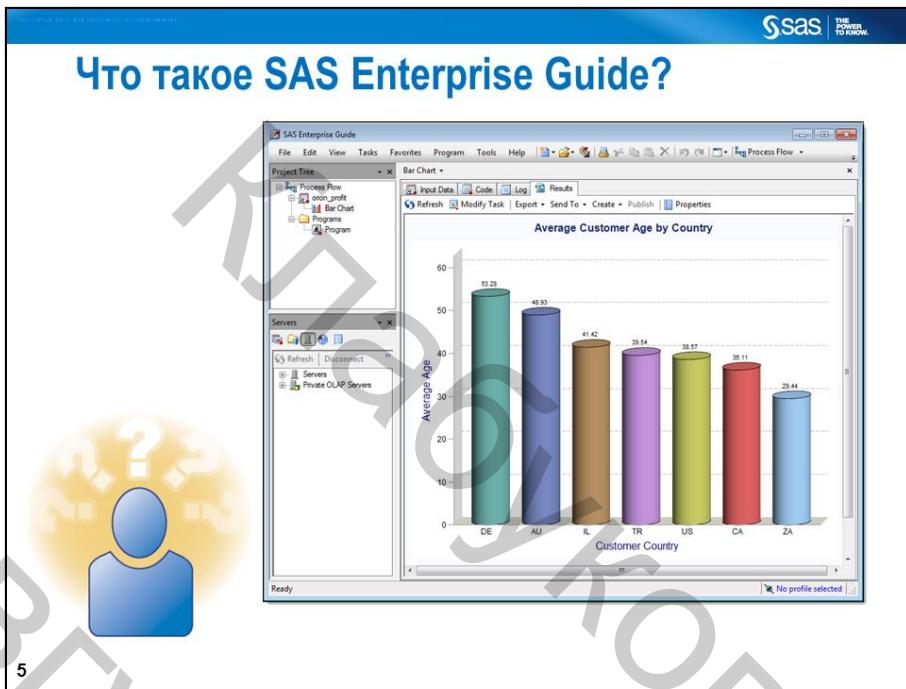
3

**Что такое SAS?**

SAS – это набор решений и технологий для решения различных бизнес-задач.

<b>Statistical Analysis</b>		<b>High-Performance Analytics</b>				<b>Business Solutions</b>		<b>Time Series Analysis</b>		<b>Business Intelligence</b>				<b>Scoring Acceleration</b>		<b>Information Management</b>					
Analysis of Variance	R Integration	Process Capability Analysis	Statistical Process Control	Categorical Data Analysis	Social Network Analysis	Reliability Analysis	Scheduling	Linear Programming	Predictive Modeling	Sample Size Computations	ARIMA Models	Mixed-Integer Programming	Nonparametric Analysis	X11 & X12 Models	D-Optimal	Survival Analysis	Ensemble Models	Decision Trees	High-Performance Forecasting	Information Theory	Multivariate Analysis
Design of Experiments	Cluster Analysis	Vector Autoregressive Models	Network Flow Models	Network Flow Models	Vector Autoregressive Models	Network Flow Models	Linear Programming	Predictive Modeling	Sample Size Computations	ARIMA Models	Mixed-Integer Programming	Nonparametric Analysis	X11 & X12 Models	D-Optimal	Survival Analysis	Ensemble Models	Decision Trees	High-Performance Forecasting	Information Theory	Multivariate Analysis	
Nonlinear Programming																					

4



5

Курс создан для версии SAS Enterprise Guide 7.1. Чтобы определить версию Enterprise Guide, которую используете вы, выберите в меню Справка ⇒ О приложении SAS Enterprise Guide (Help ⇒ About SAS Enterprise Guide).



6

...

**Что такое SAS Enterprise Guide?**

Графический интерфейс  
для SAS

прозрачный доступ к  
данным

...  
7

**Что такое SAS Enterprise Guide?**

Графический интерфейс  
для SAS

прозрачный доступ к  
данным

окна для интерактивной  
настройки задач

...  
8

**Что такое SAS Enterprise Guide?**

- Графический интерфейс для SAS
- прозрачный доступ к данным
- окна для интерактивной настройки задач
- возможности экспорта

9 ...

**Что такое SAS Enterprise Guide?**

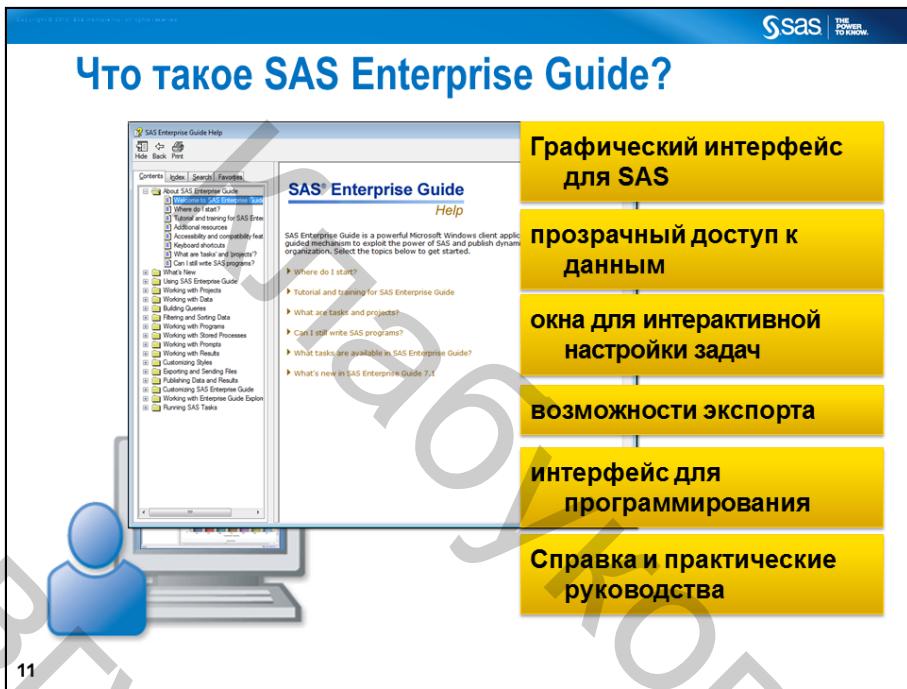
- Графический интерфейс для SAS
- прозрачный доступ к данным
- окна для интерактивной настройки задач
- возможности экспорта
- интерфейс для программирования

10 ...

Программисты SAS могут использовать редактор SAS Enterprise Guide для создания кода или изменения существующих программ SAS. Редактор программ предлагает ряд функций для сокращения времени программирования, включая возможность автодополнения множества инструкций SAS, процедур и параметров. Этот редактор также содержит встроенную справку по синтаксису и функциям, а также обеспечивает автоматическое форматирование кода.



Сведения о написании и выполнении программ на языке SAS в SAS Enterprise Guide см. в приложении А.



**Что такое SAS Enterprise Guide?**

- Графический интерфейс для SAS
- прозрачный доступ к данным
- окна для интерактивной настройки задач
- возможности экспорта
- интерфейс для программирования
- Справка и практические руководства

11

## 1.01 Вопрос с несколькими вариантами ответа

Опишите свой опыт работы с SAS Enterprise Guide.

- До прохождения этого курса мне ничего не было известно об этом приложении.
- Мне известно, что другие люди используют его.
- У меня есть небольшой опыт работы.
- У меня большой опыт работы.

12

Обсуждение

Зачем вы изучаете SAS Enterprise Guide?

13

За кулисами

Enterprise Guide создаёт код на языке SAS либо с помощью визуальных интерфейсов, либо он вводится пользователем. Затем код отправляется на выполнение в SAS.

14

...

Чтобы настроить результаты и получить доступ к функциям SAS, отсутствующим в графическом интерфейсе, можно изменить программный код.

Код можно сохранить и затем выполнить в пакетном режиме.

**За кулисами**

После выполнения кода SAS возвращает в Enterprise Guide результаты.

The diagram illustrates the workflow between a user's computer and a SAS server. A blue user icon is shown at a computer monitor displaying a bar chart. A blue arrow points from the monitor to a document icon, which then points to a server rack labeled 'Sas'. Another blue arrow points from the server back to the document icon, indicating the return of results. A third blue arrow points from the document icon back to the user's computer monitor.

15

**Конфигурация**

Enterprise Guide может быть настроен для работы с SAS либо на том же самом компьютере, либо на удалённом сервере.

The diagram compares two configuration options for Enterprise Guide. On the left, under 'Рабочая станция' (Workstation), a blue user icon is shown at a computer monitor with a 'Sas' logo. On the right, under 'Клиент/Сервер' (Client/Server), a blue user icon is shown at a computer monitor with a 'Sas' logo, and a separate server rack with a 'Sas' logo is shown to its right, connected by a line.

16

**Конфигурация**

Enterprise Guide также может являться частью платформы SAS для бизнес-аналитики. Enterprise Guide содержит дополнительные средства для интеграции с платформой SAS.

The diagram shows the SAS Business Analytics architecture across three tiers:

- Client Tier:** SAS Information Map Studio, SAS Add-in for Microsoft Office, SAS Private JRE, SAS/CONNECT Server, Platform Suite for SAS, batch servers, OLAP Server, Stored Process Server, SAS/SHARE Server.
- Web Tier:** SAS Management Console, SAS Data Integration Studio, Web Browser, Java Server Container, SAS Web Report Studio, SAS Information Delivery Portal, SAS Stored Process Web Application, J2SE SDK, WebDAV Server, SAS Services Application.
- Data Tier:** Metadata Server, Metadata Repositories, RDBMS Tables, SAS Data Sets, OLAP Cubes, SPD Files, ERP data structures, Other Data Sources.

A central circle labeled "Metadata Server" is connected to "Metadata Repositories". The entire diagram is set against a background showing a user interface with a SAS logo and a server icon.

17

Функции, которые используются при работе с платформой SAS Business Analytics, отмечены значком .

## 1.02 Вопрос с несколькими вариантами ответа

Опишите конфигурацию Enterprise Guide, которая используется на вашем рабочем месте.

- SAS установлен на моём компьютере.
- SAS установлен на сервере.
- Наша компания использует платформу SAS для бизнес-аналитики.
- Другое
- Не знаю

18

## Обзор курса

**Цели**

- Познакомиться со сценариями для демонстраций и упражнений.
- Изучить шаги для построения проекта в Enterprise Guide.
- Перечислить темы, которые изучаются в этом курсе.
- Узнать о различиях между уровнями упражнений в этом курсе.
- Ознакомиться с дополнительными ресурсами в сети интернет.

20

**Orion Star Sports & Outdoors**

Orion Star Sports & Outdoors – это вымышленная компания, продающая товары для туризма и спорта по всему миру. Для демонстраций и упражнений вам нужно использовать данные о клиентах, товарах, заказах и сотрудниках компании Orion.

21

**Проект Orion Star**

В этом курсе вы будете использовать Enterprise Guide чтобы получить данные из различных источников, создать запросы на их основе для фильтрации и объединения таблиц, создать по этим данным сводные отчёты и графики.

22

**Интерфейс Enterprise Guide: Проект**

Проект – это один файл, содержащий

- источники данных
- программы SAS и журналы их выполнения
- задачи и запросы
- результаты
- комментарии и информацию для документирования.

Пользователь может контролировать содержимое проекта, упорядочивать его составляющие и обновлять его.

e101d01.egp

23

**Создание проекта в Enterprise Guide**

Чтобы начать работу в Enterprise Guide, вам нужно

1. создать новый проект
2. добавить данные в проект
3. запустить задачи для обработки данных.

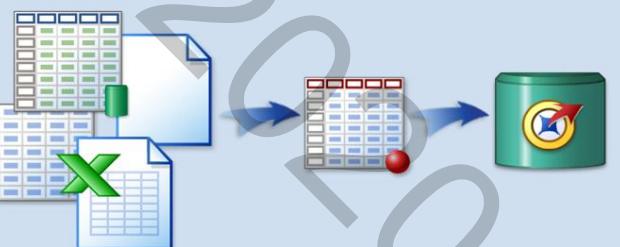


Дополнительно вы можете

4. настроить, обновить и передать результаты
5. настроить автоматизацию в процессе.

24

**Об этом курсе**



**Глава 2**  
**Работа с данными в проекте**

25

**Об этом курсе**

**Глава 3**  
Начало работы с задачами

26

**Orion Star Customer List**

Customer ID	Customer Name	Age Group
29	Candy Kinney	61+ years
41	Veronica Lindsey	31-45 years
53	Karen Pickens	31-45 years
111	Xanina Duker	31-45 years
171	Elaine Gandy	31-45 years
183	Duncan Robertshaw	61+ years
205	Erica Ladd	31-45 years
293	Garnet Trentham	46-60 years

**Average Customer Age by Country**

**Customer Country-CA**

**Number of Products per Category**

Product Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Cloud	100	21.05
Golf	56	11.64
Indoor Sports	15	3.12
Outdoor Sports	97	20.20
Racket Sports	14	2.91
Running - Jogging	25	5.20
Shoes	34	7.00
Swim Sports	7	1.46
Team Sports	23	4.78
Winter Sports	17	3.53

**Об этом курсе**

**Глава 4**  
Создание простых запросов

27

**Об этом курсе**

Summary of Customer Age by Country							
Analysis Variable : Customer_Age Customer_Age							
Customer Country	N	Obs	Mean	Std Dev	Minimum	Maximum	N
AU	8	54.0	14.5	34.0	75.0	8	
CA	15	49.4	22.0	19.0	75.0	15	
DE	10	48.6	19.6	21.0	75.0	10	
IL	5	41.8	10.3	25.0	50.0	5	
TR	7	41.0	12.6	29.0	65.0	7	
Total	28	42.3	14.9	23.1	75.0	28	
	4	36.6	19.2	21.0	60.0	4	

Profits by Customer Age Group					Total
	15-30 years	31-45 years	46-60 years	61+ years	Total
Children	\$419	\$238	\$195	\$25	\$867
Children Sports	\$179	\$122	\$84	\$21	\$386
Clothes & Shoes	\$171	\$107	\$94	\$19	\$387
Clothes	\$87	\$56	\$44	\$10	\$197
Outdoors	\$1,081	\$1,000	\$2,111	\$1,046	\$8,038
Product Line Subtotal	\$1,081	\$1,000	\$2,111	\$1,046	\$8,038
Sports	\$1,385	\$6,264	\$1,117	\$1,175	\$115,478
Amateur Sports Articles	\$1,385	\$6,264	\$1,117	\$1,175	\$115,478
Golf	\$1,784	\$3,150	\$1,037	\$762	\$9,773
Indoor Sports	\$1,051	\$1,259	\$1,259	\$1,259	\$5,038
Racket Sports	\$851	\$627	\$1,016	\$347	\$3,231
Running, Jogging	\$201	\$1,216	\$332	\$106	\$1,751
Swim Sports	\$400	\$191	\$212	\$779	\$1,002
Tennis Sports	\$1,051	\$1,188	\$1,188	\$1,188	\$5,511
Product Line Subtotal	\$10,854	\$11,723	\$4,491	\$2,749	\$28,669
Total	\$10,854	\$11,723	\$4,491	\$2,749	\$10,869 \$14,168

**Глава 5**  
**Создание сводных отчётов**

28



## Рабочая область SAS Enterprise Guide

Запустите приложение SAS Enterprise Guide и приступите к созданию проекта.

1. Откройте SAS Enterprise Guide.

При входе в SAS Enterprise Guide автоматически открывается диалоговое окно «Welcome to SAS Enterprise Guide» («Вас приветствует SAS Enterprise Guide»), где можно начать создавать новый или открыть существующий проект. Все проекты, находящиеся в разделе «Open a project» (открытие проекта), имеют расширение EGP.

2. Выберите **Создать проект (New Project)**.

Если диалоговое окно не открылось при запуске SAS Enterprise Guide, выберите **File (файл) ⇒ Open (открыть) ⇒ Project (проект)**, чтобы открыть его. Чтобы создать новый проект, последовательно выберите **File ⇒ New (создать) ⇒ Project**.

3. По умолчанию в SAS Enterprise Guide отображаются три основных окна. Эти окна называются:

- Project Tree (Дерево Проекта)
- Process Flow (Схема Процесса)
- Server List (Список Серверов)

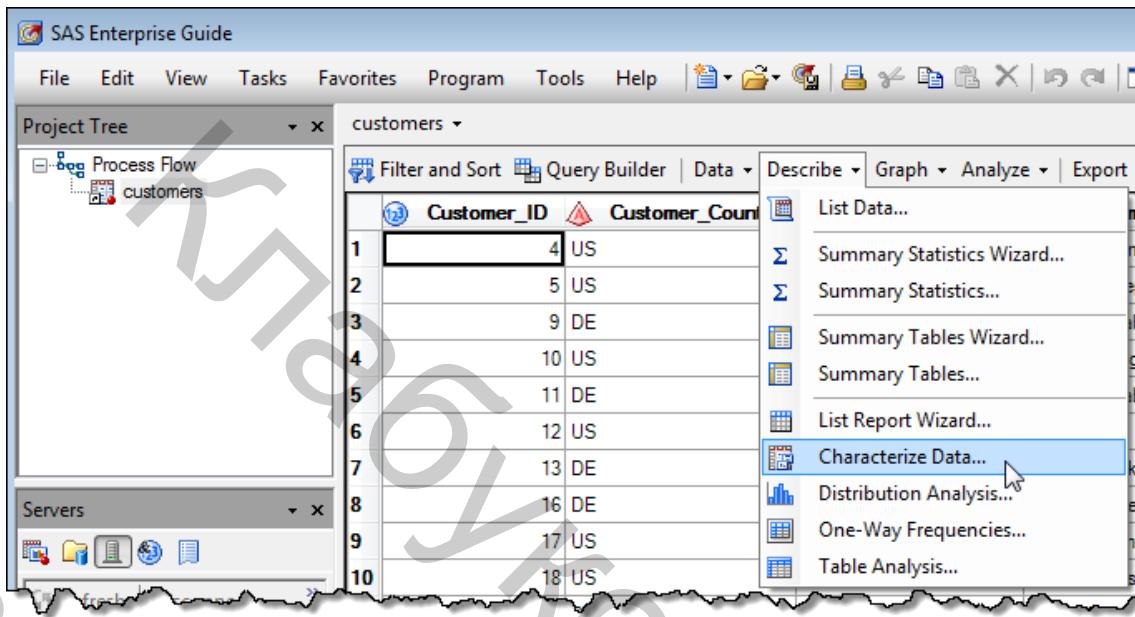
4. По умолчанию окна Дерево Проекта (Project Tree) и Список Серверов (Server List) закреплены в левой части. Чтобы закрепить окно в левой части или автоматически скрыть его, щелкните кнопку . Режим Скрывать автоматически (Auto-hide) скрывает окно, но отображает его имя на границе окна SAS Enterprise Guide. Чтобы отобразить окно, наведите на его имя указатель мыши.

Области Дерева проекта («Project Tree»), Схемы процесса («Process Flow») и другие окна можно открыть из меню Вид («View»). Чтобы восстановить исходное положение окон, в строке меню выберите Сервис ⇒ Параметры ⇒ Восстановить разметку окна (**Tools ⇒ Options ⇒ Restore Window Layout**). Затем нажмите **OK**.

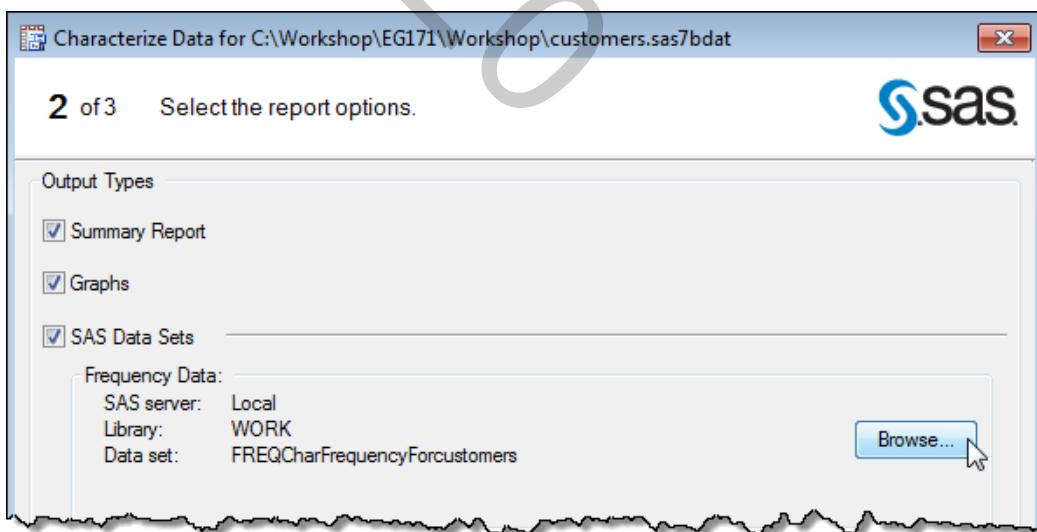
5. Добавьте в проект существующий набор данных SAS, выбрав **File ⇒ Open ⇒ Data (данные)**.
6. Чтобы перейти к данным, сохраненным на ПК, выберите **My Computer (мой компьютер)**. Перейдите в расположение, указанное преподавателем, и выберите **customers ⇒ Open**.

В дереве проекта и в окне «Process Flow» добавляется ярлык для набора данных **customers**. По умолчанию в рабочей области открывается средство просмотра («сетка данных»), в котором появятся несколько строк из источника данных.

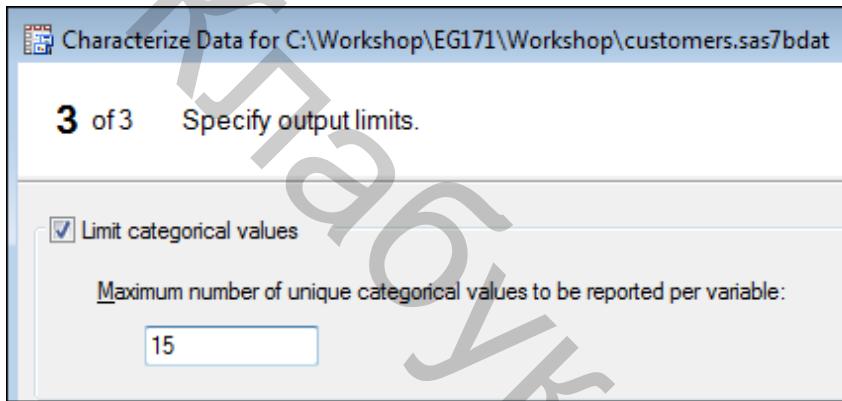
7. Данные, добавленные в проект, можно использовать для анализа и создания отчетов. Оставив набор данных **customers** открытым, выберите **Describe (описательная статистика) ⇒ Characterize Data (характеризовать данные)**.



8. Задача «Характеризовать данные» (Characterize Data) доступна в режиме Мастера настроек, который содержит три шага для настройки задачи и изменения её параметров. На шаге 1 вы проверяете, какой источник данных выбран для анализа. Нажмите кнопку **Next (далее)**.
9. Чтобы получить дополнительные сведения о каждой задаче, из окна с настройками задачи можно открыть справку по SAS Enterprise Guide. Чтобы получить дополнительные сведения о задаче «Характеризовать данные» (Characterize Data), нажмите кнопку **Help (справка)** и просмотрите информацию о задаче и параметрах отчетов. Закройте окно справки, нажав кнопку .
10. На шаге 2 можно настроить параметры отчета, в том числе создать сводный отчет, набор выходных данных или построить графики. Чтобы изменить имя набора данных, включающего частоты для столбцов символьного типа, в области «Frequency Data» (данные частот) нажмите кнопку **Browse (Обзор)**.



11. В поле **File name (имя файла)** введите **CustomerCounts** и нажмите кнопки **Save (сохранить)** ⇒ **Next (далее)**.
12. На шаге 3 ограничьте количество уникальных категориальных значений для каждой переменной, указав **15**.



13. Чтобы выполнить задачу и просмотреть результаты, нажмите кнопку **Finish (готово)**.

Чтобы отследить состояние задачи во время ее выполнения, в левом нижнем углу в приложении SAS Enterprise Guide нажмите кнопку **Details (подробнее)**. Откроется окно «Task Status» (состояние задачи) с отображением выполняющейся задачи, ее состояния и сервера.

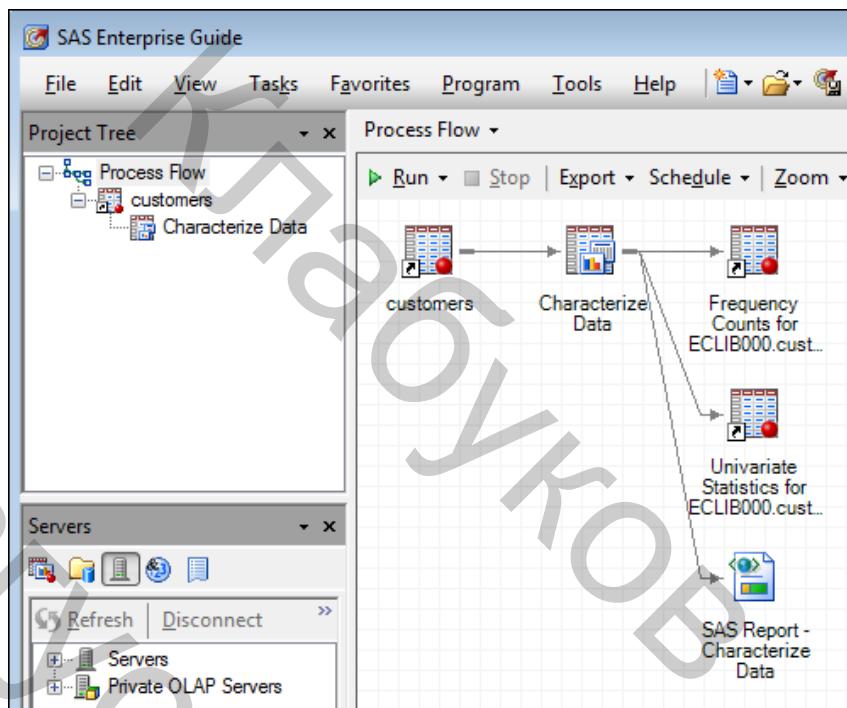
Результаты содержат подсчет частот для символьных столбцов, сводные статистики для числовых столбцов и графики, описывающие данные в них.

Фрагмент вывода

Characterize Data					
Input Data	Code	Log	Output Data (2)	Results	
<b>Summary of Categorical Variables for ECLIB000.customers</b>					
<b>Limited to the 15 Most Frequent Distinct Values per Variable</b>					
Variable	Label	Value	Frequency Count	Percent of Total Frequency	
Customer_Age_Group	Customer Age Group	31-45 years	27	35.0649	
		15-30 years	22	28.5714	
		46-60 years	14	18.1818	
		61+ years	14	18.1818	
Variable	Label	Value	Frequency Count	Percent of Total Frequency	
Customer_Country	Customer Country	US	28	36.3636	
			15	48%	

14. Чтобы просмотреть содержимое текущего проекта, щелкните **Process Flow** на панели инструментов. Дважды щелкните **Process Flow** в дереве проекта или нажмите клавишу F4.

Задача «Характеризовать данные» (Characterize Data) связана с набором данных **customers** и добавлена в дерево проекта и схему процесса.



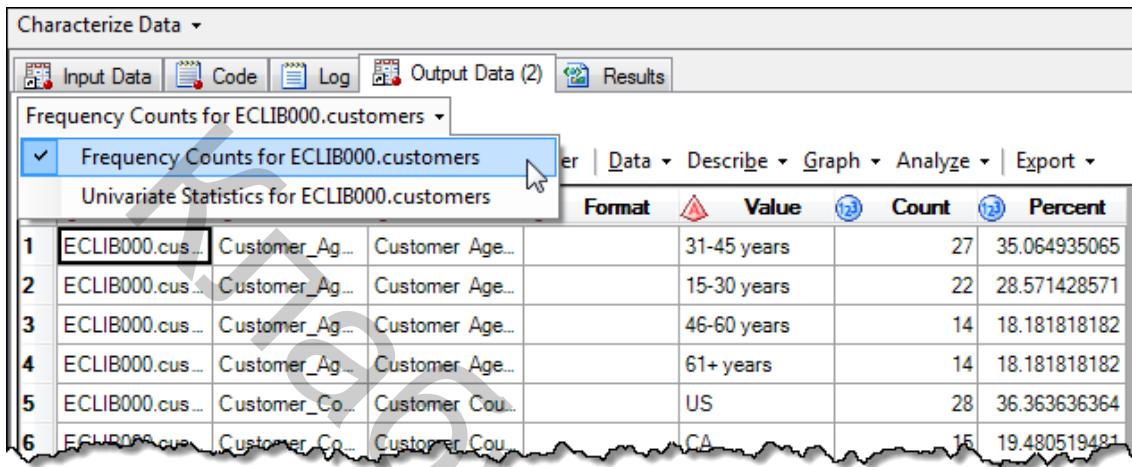
15. Вернитесь к результатам задачи, дважды щелкнув значок (**Characterize Data**) в Дереве проекта или в Схеме процесса . Чтобы просмотреть код задачи, откройте вкладку **Code** (**код**).

Код SAS можно сохранить и изменить для более детальной настройки результатов задачи или использования тех настроек SAS, для которых в среде SAS Enterprise Guide не предусмотрен соответствующий интерфейс.

16. В окне «Log» (журнал) отображаются сообщения из SAS для каждой выполняемой задачи. Чтобы просмотреть журнал, откройте вкладку **Журнал** (**Log**).

17. В результате выполнения этой задачи формируются два набора данных, но на вкладке открывается только набор, созданный первым. Перейдите на вкладку **Output Data (2)**, чтобы просмотреть набор данных, содержащий рассчитанные значения частот.

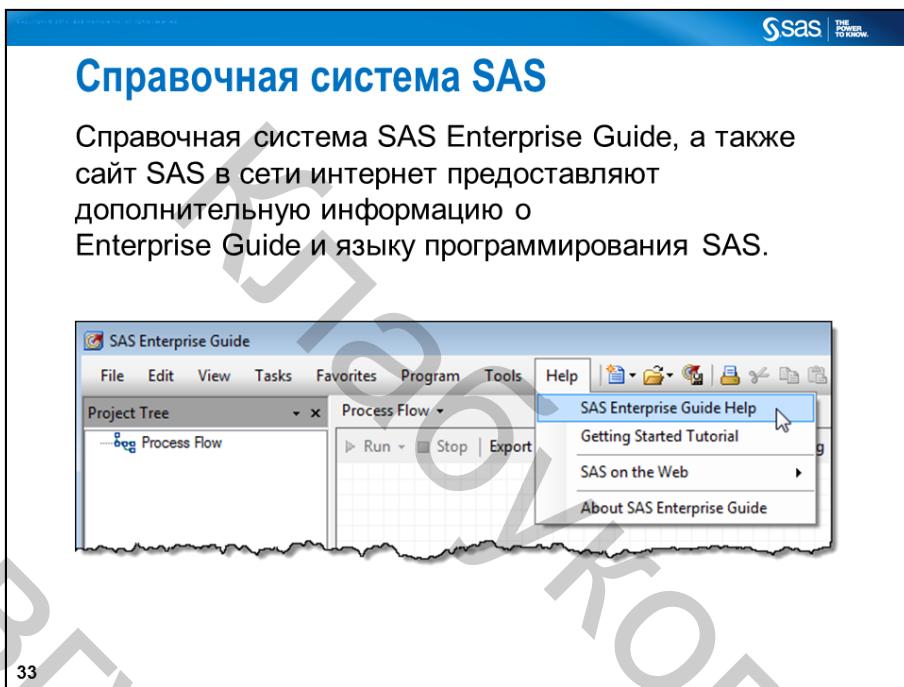
Другой набор данных, содержащий сводную статистику, доступен либо из раскрывающегося списка на вкладке «Output Data», либо из схемы процесса.



18. Выходные данные можно экспортить в различные форматы, включая Microsoft Excel. Чтобы автоматически открыть данные в Excel, в окне «Process Flow» выберите Отправить (Send To) ⇒ Microsoft Excel. Закройте Excel и не сохраняйте изменения.
19. Чтобы сохранить задачи, код и результаты, имеющиеся на данный момент, необходимо сохранить проект. В меню выберите Файл⇒Сохранить проект как (File ⇒ Save Project As) и щелкните значок My Computer. Перейдите в требуемое место расположения файла и в качестве имени файла введите Chapter1. Нажмите кнопку Save.

Проекты можно сохранить на клиентском компьютере или на удаленном сервере, где установлен SAS.

20. (Необязательно) Для упорядочения элементов проекта в нём можно создать несколько схем процесса, при этом каждую из них можно назвать своим именем. В дереве проекта нажмите правой кнопкой мыши на схему Process Flow и выберите Переименовать (Rename). Введите Demos.
21. Для упражнений можно создать другую схему процесса. В меню выберите Файл⇒Создать⇒Схема процесса (File ⇒ New ⇒ Process Flow). Можно также переименовать схему процесса нажав кнопку Свойства (Properties) и в поле Ярлык (Label) ввести Exercises.
  - Новые источники данных добавляются в активную схему процесса. Перед добавлением источников данных убедитесь, что вы выбрали требуемую схему процесса в Дереве проекта
  - В дереве проекта вы можете копировать элементы проекта из одной схемы процесса в другую.
22. На панели инструментов нажмите кнопку (Сохранить проект/Save Project).



## Дополнительные ресурсы

- Сообщество пользователей SAS Enterprise Guide:  
<https://communities.sas.com/>
- Блог по SAS Enterprise Guide:  
<http://blogs.sas.com/sasdumy/>
- Документация по системе SAS:  
<http://support.sas.com/documentation>

The SAS Dummy  
A SAS® blog for the rest of us

SAS BLOG HOME - THE SAS DUMMY - NEW SAS PROGRAMMING FEATURES IN SAS ENTERPRISE GUIDE 7.1

Chris Hemedinger | OCTOBER 12, 2014

4364 25 3 1 3 1

New SAS programming features in SAS Enterprise Guide 7.1

SAS Enterprise Guide 7.1 began shipping last week. Of the many new features, some are more subtle. My favorite new features are those for SAS programmers, including several customers ask for specifically. I'll describe them briefly here; the SAS Enterprise Guide online details.

**SAS Enterprise Guide**

Welcome to SAS Enterprise Guide

SAS Enterprise Guide is a desktop application to help you program, report generation, and data analysis. Use this community to share your experiences, ask questions, and ideas with other SAS Enterprise Guide users. For questions requiring

Recent Discussions in the SAS Enterprise Guide Community

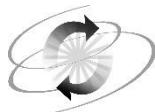
Re: Creating a new Character string with Query Builder? 1 hour ago

Re: Reorder Bargraph made with sgplot 2 hours ago

Re: Options to Prefilter dataset in SAS 2 hours ago

Re: length issue 2 hours ago

34



## Упражнения

---

### 1. Создание проекта и настройка задачи «Характеризовать данные»

- a. Добавьте в проект **Chapter1** набор данных **employee\_organization**.



Если вы в предыдущей демонстрации создали схему процесса Exercises, перед добавлением данных в проект убедитесь, что в дереве проекта выбрана именно она.



Если проект **Chapter1** не был сохранен, создайте новый проект, выбрав **New Project (создать проект)** в окне «Welcome to SAS Enterprise Guide» или последовательно выбрав **File (файл) ⇒ New (создать) ⇒ Project (проект)** в меню SAS Enterprise Guide.

- 1) В строке меню выберите **File (файл) ⇒ Open (открыть) ⇒ Data (данные)**.
- 2) В левой части окна «Open Data» (открытие данных) выберите **My Computer (мой компьютер)**.
- 3) Перейдите к папке, в которой хранятся данные курса.
- 4) Выберите набор данных **employee\_organization** и нажмите кнопку **Open (открыть)**.

- b. С помощью задачи «Характеризовать данные» (Characterize Data) изучите распределения значений всех столбцов в наборе данных **employee\_organization**.

- 1) В меню сетки данных выберите **Describe (описательная статистика) ⇒ Characterize Data (характеризовать данные)**.
- 2) Ограничьте количество уникальных категориальных значений, выводимых для каждой переменной, до 15.
- 3) Выполните задачу и изучите результаты.

Фрагмент результатов

Summary of Categorical Variables for ECLIB000.employee_organization				
Limited to the 15 Most Frequent Distinct Values per Variable				
Variable	Label	Value	Frequency Count	Percent of Total Frequency
Department		Sales	201	47.4057
		Administration	34	8.0189
		Stock & Shipping	26	6.1321
		IS	25	5.8962
		Marketing	20	4.7170
		Group HR Management	18	4.2453

- c. Сохраните проект с именем **Chapter1** в расположении, указанном преподавателем.

## Сведения о структуре данных SAS

Книгуakov Maxim

3

**Цели**

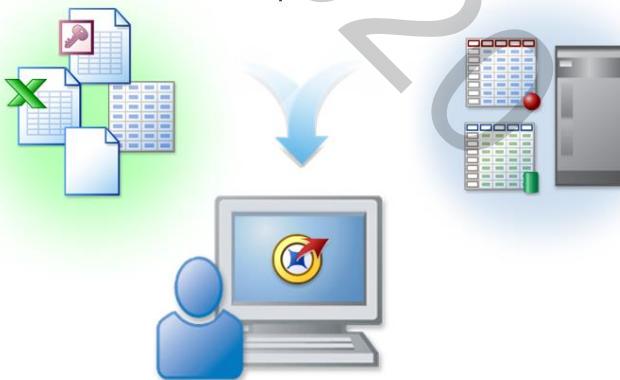
- Понять, что такое набор данных SAS.
- Узнать, как хранятся данные в наборе данных SAS.

Книгуakov Maxim

4

**Бизнес-сценарий**

Компания Orion хранит и поддерживает данные в различных форматах и использует разные источники. Аналитику необходимо получить доступ к этим данным из SAS Enterprise Guide.



**Общие форматы данных**

SAS Enterprise Guide может считывать и использовать данные в разных форматах и из разных источников

The diagram illustrates the connectivity of SAS Enterprise Guide. At the center is a computer monitor displaying a SAS logo. Arrows point from the monitor to two main data storage units: one labeled 'DB2' and another labeled 'ORACLE'. From these units, further arrows point to specific data formats: a grid icon, a green 'X' icon, and a blue 'a' icon. On the left side, there are icons for Microsoft Excel (blue 'X' icon) and Microsoft Access (green 'a' icon). A small number '5' is located at the bottom left corner of the slide.



SAS Enterprise Guide может читать информационные карты SAS. *Информационная карта* представляет собой абстракцию на уровне метаданных, которая создана поверх источников данных в хранилище. Метаданные — это сведения о структуре и содержимом данных. В информационной карте отсутствуют физические данные. Информационные карты дают бизнес-пользователям более удобную возможность для составления запросов к данным и получения необходимых им результатов.

**2.01 Вопрос с несколькими вариантами ответа**

С какими типами данных вы работаете?

- Microsoft Excel
- Microsoft Access
- Таблицы DBMS (например, Oracle или DB2)
- Наборы данных SAS
- Текстовые файлы (файлы с разделителями)
- Другие типы

The slide contains a question and a list of six options. A small number '6' is located at the bottom left corner of the slide.

**Набор данных SAS**

Набор данных SAS (или таблица) – это «плоская» таблица, состоящая из строк из столбцов.

	Customer_Name	Customer_Age	Order_Date	Profit
1	James Klisurich	38	11JAN2010	\$9.05
2	Sandrina Stephano	28	15JAN2010	\$137.95
3	Dianne Patchin	28	20JAN2010	\$19.75
4	Wendell Summersby	43	28JAN2010	\$19.00
5	Duncan Robertshawe	63	27FEB2010	\$37.20
6	Najma Hicks	21	02MAR2010	\$118.65
7	Tulio Devereaux	58	03MAR2010	\$19.25
8			03MAY2010	\$42.10

Строки (наблюдения)

Столбцы (переменные)

**Имена столбцов**

Имена столбцов должны содержать не более 32 символов. Также рекомендуется, чтобы они

- начинались с буквы латинского алфавита или знака подчёркивания
- содержали только буквы латинского алфавита, знаки подчёркивания и цифры.

	Customer_Name	Customer_Age	Order_Date	Profit
1	James Klisurich	38	11JAN2010	\$9.05
2	Sandrina Stephano	28	15JAN2010	\$137.95
3	Dianne Patchin	28	20JAN2010	\$19.75
4	Wendell Summersby	43	28JAN2010	\$19.00
5	Duncan Robertshawe	63	27FEB2010	\$37.20
6	Najma Hicks	21	02MAR2010	\$118.65
7	Tulio Devereaux	58	03MAR2010	\$19.25
8			03MAY2010	\$42.10

В Enterprise Guide можно использовать пробелы и специальные символы в именах столбцов, поскольку он автоматически включает настройку сеанса VALIDVARNAME=ANY, когда подключается к SAS. Однако, не все окружения SAS используют эту настройку. Если вы будете использовать код, который создаётся в Enterprise Guide, в других приложениях SAS, вам рекомендуется следовать стандартным правилам для наименования столбцов. Стандартные правила предполагают, что имя столбца начинается с буквы латинского алфавита или знака подчёркивания, и состоит из цифр, знака подчёркивания и букв латинского алфавита.



В Enterprise Guide 6.1 или более поздних версиях есть настройка правил для наименования столбцов. Чтобы форсировать использование стандартных правил SAS, выберите Сервис ⇒ Параметры ⇒ Основные данные (Tools ⇒ Options ⇒ Data General) и на панели Naming Options измените значение настройки Valid variable names на Basic variable names (V7).

**2.02 Вопрос с несколькими вариантами ответа**

Какие имена столбцов соответствуют стандартным правилам SAS?

- a. data5mon
- b. 5monthsdata
- c. data#5
- d. five months data
- e. five\_months\_data
- f. FiveMonthsData

9

**Ярлыки столбцов**

Для столбца может быть задан ярлык – дополнительная информация, использующаяся при создании отчётов. Ярлыки могут иметь до 256 символов и содержать любые символы.

Customer Name	Customer Age	Date Order was Placed	Total Profit
James Klisurich	38	11JAN2010	\$9.05
Sandrina Stephano	28	15JAN2010	\$137.95
Dianne Patchin	28	20JAN2010	\$19.75
Wendell Summersby	43	28JAN2010	\$19.00
Duncan Robertshawe	63	27FEB2010	\$37.20
Najma Hicks	21	02MAR2010	\$118.65
Tulio Devereaux	58	03MAR2010	\$19.25
Tulio Devereaux	58	03MAR2010	\$40.10

11

По умолчанию в заголовке сетки данных и в задачах отображаются имена столбцов. Чтобы вместо имён выводились ярлыки, можно включить в меню Сервис ⇒ Параметры ⇒ Основные данные

(Tools ⇒ Options ⇒ Data General) настройку Использовать ярлыки для имён столбцов (Use labels for column names).

**Тип столбца**

Тип столбца может быть или символьным (текстовым),  
или числовым.

	Customer_Name	Customer_Age	Order_Date	Profit
1	James Kisurich	38	18273	9.05
2	Sandrina Stephano	28	18277	137.95
3	Dianne Patchin	28	18282	19.75
4	Wendell Summersby	43	18290	19.00
5	Duncan Robertshawe	63	18320	37.20
6	Najma Hicks	21	18323	118.65
7	Tulio Devereaux	58	18324	19.25
8	Тим Денисих	58	18324	0.10

12

Текстовые значения -- длиной от 1 до 32,767 символов.

Числовые значения могут содержать  
 • числа  
 • валюту  
 • дату (дни с 01 января 1960 г.)  
 • время(секунды с полуночи).

Наряду с именем и типом, обязательным атрибутом столбца является его *длина* (*length*). Длина — это число байтов, используемых для хранения значения каждой переменной в наборе данных SAS. Символьные переменные имеют длину от 1 до 32 767 байт. Один байт соответствует одному символу. Числовые значения хранятся в виде чисел с плавающей запятой и по умолчанию занимают 8 байт. Этого достаточно для хранения целочисленных значений с 16 или 17 значащими цифрами.

**Форматы**

Формат используется для вывода данных в определённом виде. Форматы не влияют на хранение данных.

	Customer_Name	Customer_Age	Order_Date	Profit
1	James Klurich	38	11JAN2010	\$9.05
2	Sandrina Stephano	28	15JAN2010	\$137.95
3	Dianne Patchin	28	20JAN2010	\$19.75
4	Wendell Summersby	43	28JAN2010	\$19.00
5	Duncan Robertshawe	63	27FEB2010	\$37.20
6	Najma Hicks	21	02MAR2010	\$118.65
7	Tulio Devereaux	58	03MAR2010	\$19.25
8	Tull Devereaux	59	03MAR2010	\$19.25

Формат: DATE  
Ширина: 9  
Исходное значение: 18324

Формат: DOLLAR  
Ширина: 10  
Число дробных разрядов: 2  
Исх. значение: 19.25

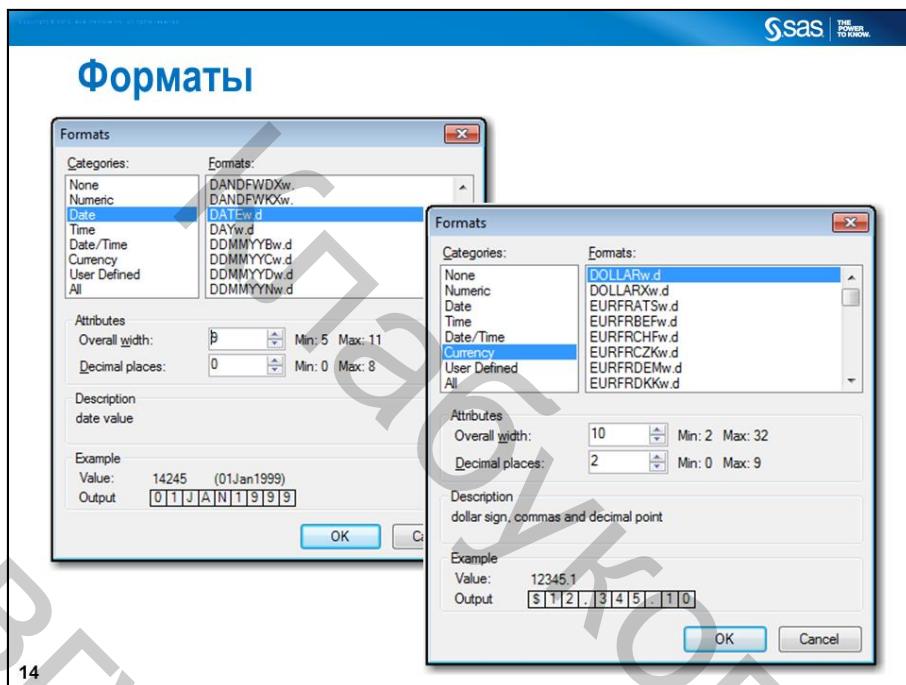
13

*Формат (format)* — это инструкция, применяемая к столбцу для его отображения. Формат сообщает SAS Enterprise Guide о том, как должны отображаться значения данных на экране или в отчёте. Форматы используются для управления отображением значений данных или для группировки значений данных для анализа.

*Входной формат (Informat)* — это инструкция, которую SAS Enterprise Guide применяет для считывания значений данных в переменную. Например, следующее значение содержит символ доллара и запятые:

\$1,000,000

Чтобы удалить символ доллара (\$) и запятые (,) перед сохранением числового значения 1000000 в переменную, это значение нужно прочитать, используя informat «DOLLAR10.»..



14

The screenshot shows the SAS Data Editor interface. A table named 'CUSTOMERORDERS' is displayed with columns: Customer\_Name, Customer\_Age, Order\_Date, and Profit. A red arrow points to the 'Properties' button in the toolbar above the table. A 'Properties for CUSTOMERORDERS' dialog box is open, showing the 'Columns' tab with details for each column: Customer\_Name (Character, 40, Format: \$12.1, Informat: DOLLAR10.2, Label: Customer Name), Customer\_Age (Numeric, 8, Format: 12.1, Informat: 12.1, Label: Customer Age), Order\_Date (Date, 8, Format: DATE9., Informat: DATE9., Label: Date Order was Placed), and Profit (Currency, 8, Format: \$12.1, Informat: DOLLAR10.2, Label: Total Profit).

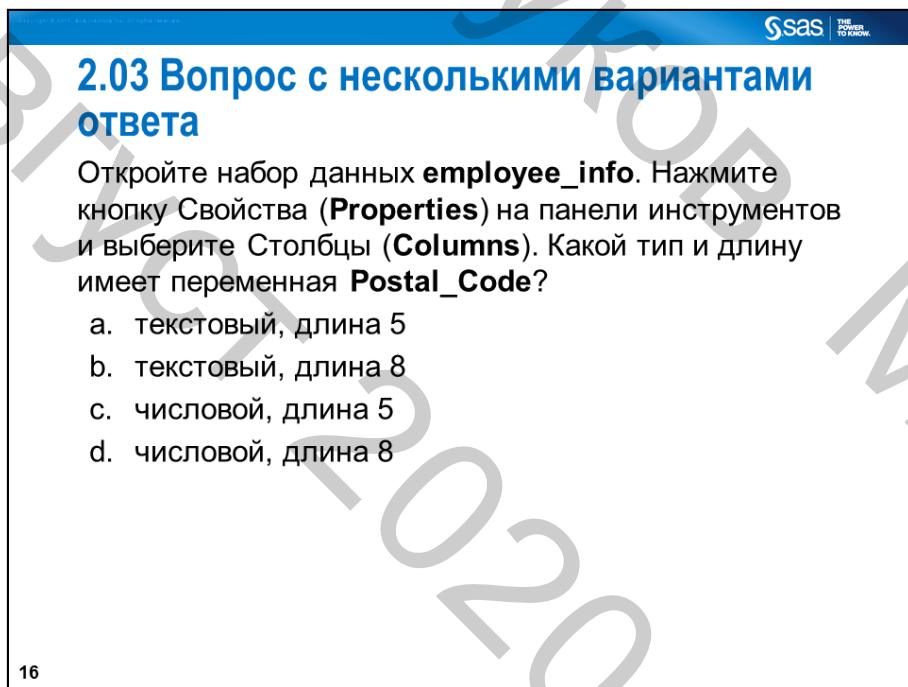
15

Атрибуты набора данных SAS можно посмотреть в свойствах. Там содержатся:

- имя набора данных;
- место расположения хранилища;
- дата последнего изменения;
- все атрибуты столбцов (имя, тип, длина);
- число строк и столбцов в наборе данных;
- сервер, на котором находятся данные.

Чтобы просмотреть свойства источника данных, нажмите кнопку  на панели инструментов сетки данных или щелкните правой кнопкой мыши в дереве проекта или в окне схемы процесса интересующий вас элемент и в контекстном меню выберите пункт **Properties (свойства)**.

В документации SAS свойства набора данных называются его *Дескриптором (descriptor portion)*.



The screenshot shows a question from a SAS exam. The title is "2.03 Вопрос с несколькими вариантами ответа". The text asks to open the "employee\_info" dataset and check the properties of the "Postal\_Code" column. The correct answer is listed as option "a".

Откройте набор данных **employee\_info**. Нажмите кнопку Свойства (**Properties**) на панели инструментов и выберите Столбцы (**Columns**). Какой тип и длину имеет переменная **Postal\_Code**?

a. текстовый, длина 5  
b. текстовый, длина 8  
c. числовой, длина 5  
d. числовой, длина 8

16

**Пропущенные значения**

Если в ячейке отсутствуют данные, её значение называется пропущенным (missing).

- Пропущенное текстовое значение отображается как пробел.
- Пропущенное числовое значение отображается как точка.

	Customer_Name	Customer_Age	Order_Date	Profit
1	James Klisurich	.	11JAN2010	\$9.05
2	Sandrina Stephano	28	15JAN2010	\$137.95
3		28	20JAN2010	\$19.75
4	Wendell Summersby	43	28JAN2010	\$19.00
5	Duncan Robertshawe	63	27FEB2010	\$37.20
6	Najma Hicks	21	02MAR2010	\$118.65
7	Tulio Devereaux	58	03MAR2010	\$19.25
8		59	07MAR2010	0.00

18

Во многих задачах Enterprise Guide имеются настройки, позволяющие указать, как нужно обрабатывать пропущенные значения в отчёте или при анализе данных.

Клабуков Максим  
Август 2020

## Доступ к данным SAS и DBMS

Кабуков Максим

20

**Цели**

- Познакомиться с разными способами организации доступа к данным.
- Узнать, что такое библиотеки SAS.
- Научиться добавлять данные в проект из библиотеки.
- Познакомиться с интерактивным исследованием данных.

sas THE POWER TO KNOW.

Кабуков Максим

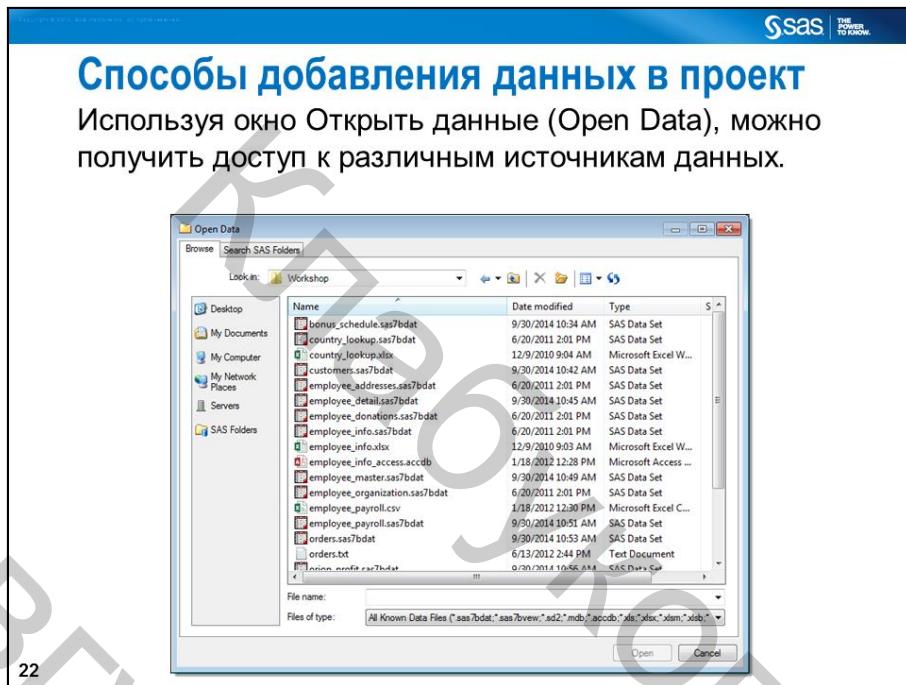
21

**Бизнес-сценарий**

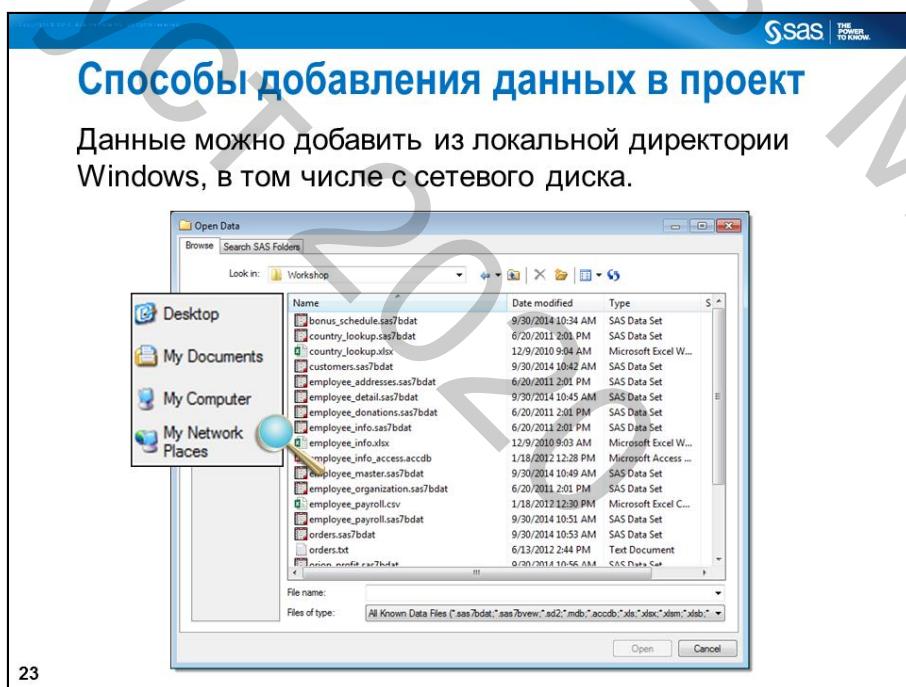
Компания Orion Star хочет использовать SAS Enterprise Guide для доступа к таблицам SAS и DBMS, содержащим информацию о клиентах и заказах.



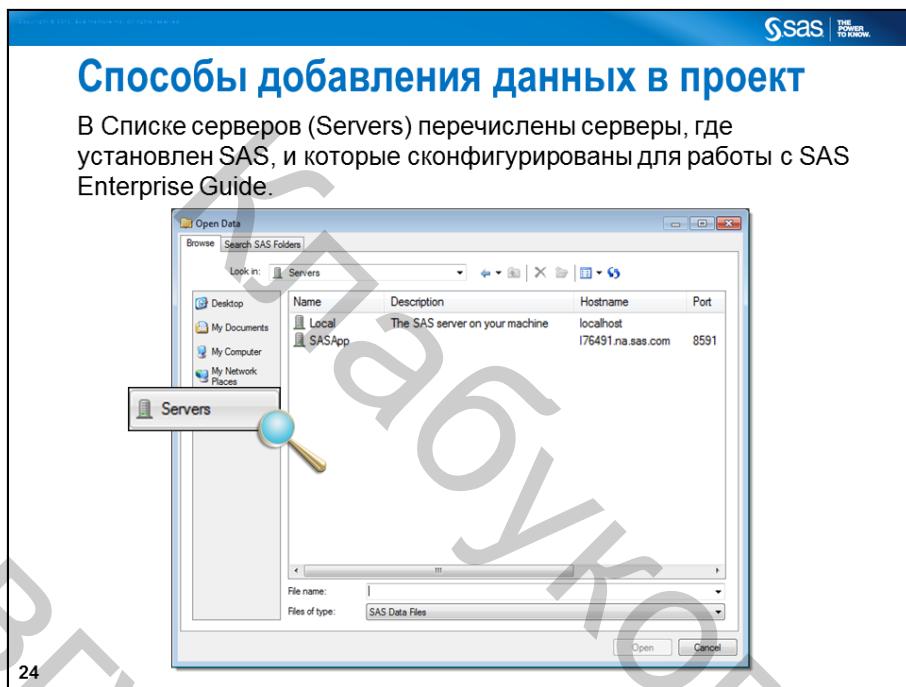
DBMS — это сокращение для *системы управления базой данных* (*database management system*).



22

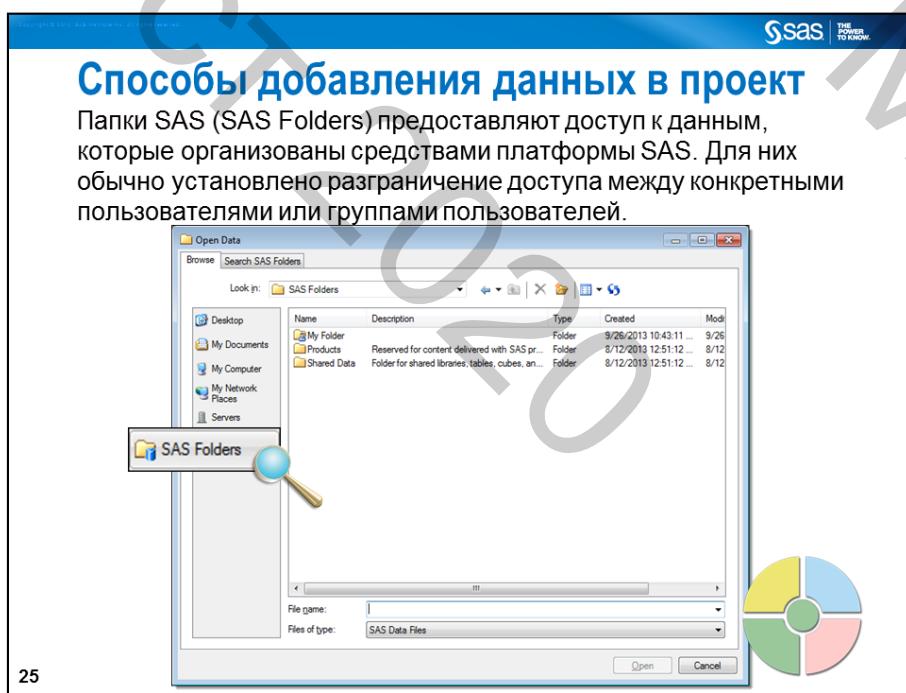


23



24

Администратор SAS Enterprise Guide должен определить удаленные серверы.



25

**Доступ к данным через Серверы или Папки SAS**

Панель ресурсов предоставляет ещё один способ для просмотра и доступа к файлам и источникам данных.

26

При наличии необходимых прав можно выполнять различные задачи по управлению серверами и файлами, например, удалять, копировать или переименовывать файлы, выбирая элементы (серверы, папки, библиотеки, наборы данных, файлы) правой кнопкой мыши.

## 2.04 Тест

- Выберите Файл ⇒ Создать ⇒ Проект (**File ⇒ New ⇒ Project**).
- Выберите Вид ⇒ Серверы (**View ⇒ Servers**) или нажмите на значок Серверы (**Servers**) в панели ресурсов.

- Раскройте список Серверы ⇒ Локальный ⇒ Файлы (**Servers ⇒ Local ⇒ Files**).
- Перейдите в расположение данных курса.

Что происходит, когда вы щёлкаете два раза указателем мыши на **customers.sas7bdat**?

27

**Библиотеки SAS**

Ещё один способ организации доступа к данным SAS или DBMS – через библиотеку SAS. Библиотека SAS – это коллекция файлов, которые создаются, распознаются и используются в SAS.

29

Как правило, библиотеки SAS создаются и обслуживаются администраторами SAS или пользователями других приложений SAS.

**Библиотеки SAS**

Чтобы указать тип данных в библиотеке, и то, как их нужно считывать, используется специальный механизм (*engine*).

30

**Библиотека SAS**

Библиотеку **orion** можно создать с помощью механизма Base SAS. Через неё можно обращаться ко всем наборам данных SAS в директории (например, s:\workshop).

31

В операционных системах Windows и UNIX библиотека SAS является директорией. В z/OS библиотека SAS представляет собой файл операционной системы. Доступ к библиотекам SAS осуществляется по их именам, например **Sasuser** или **orion\_db2**.

**Библиотека Microsoft Access**

Библиотеку **Orionacs** можно создать с помощью механизма SAS/ACCESS Interface to PC Files. Через него можно прочитать все таблицы в базе данных Microsoft Access.

32

Имена библиотек должны начинаться с буквы латинского алфавита или знака подчеркивания, состоять только из букв латинского алфавита, знаков подчеркивания или чисел, а их длина не должна превышать 8 символов.

**Библиотека DBMS**

Библиотеку **Oriondb** можно создать при помощи механизма SAS/ACCESS Interface to Teradata. С его помощью из базы данных Teradata можно считывать все доступные таблицы.

33

### 2.05 Тест

- Выберите Вид ⇒ Серверы (**View** ⇒ **Servers**) или щёлкните на значок Серверы (**Servers**) на панели ресурсов под деревом проекта.
- Раскройте Серверы ⇒ Локальный ⇒ Библиотеки (**Servers** ⇒ **Local** ⇒ **Libraries**).

Какие библиотеки перечислены в списке?

34

**Доступные библиотеки**

Некоторые библиотеки создаются в SAS автоматически.



**Work** – временная библиотека



**Sasuser** – постоянная библиотека для определённого пользователя

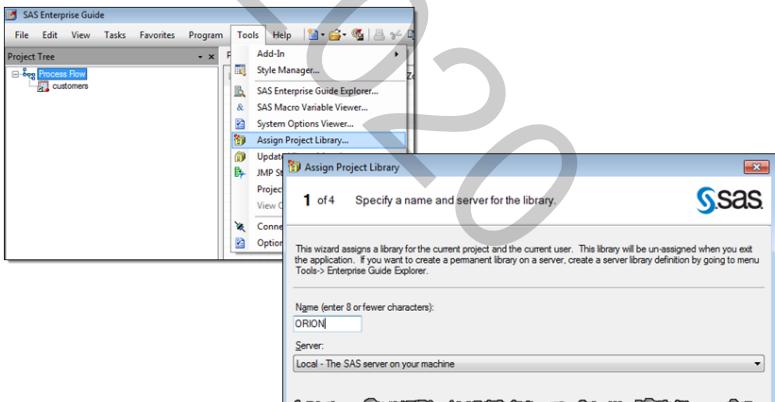


**Sashelp, Maps** – библиотеки, содержащие таблицы с примерами данных или служебные таблицы

36

**Пользовательские библиотеки**

Пользовательские библиотеки SAS обычно настраиваются вашим администратором. Однако, при помощи мастера создания библиотек в Enterprise Guide можно назначать библиотеки для источников данных SAS или DBMS, к которым у вас есть хотя бы доступ на чтение.



37

Библиотеки, созданные с помощью мастера назначения библиотек, остаются назначенными в течение всего сеанса работы с SAS или до отмены их назначения вручную. Чтобы назначить библиотеку проекта в новом сеансе работы с SAS Enterprise Guide, задачу «Назначить библиотеку проекта» потребуется выполнить повторно.

**Добавление данных в проект**

После того, как библиотека создана, щёлкните на источнике данных два раза или перенесите его из списка серверов в дерево проекта или на диаграмму процесса. Эти действия добавят ярлык для источника данных в проект.

38

Источник данных, добавляемый в проект, автоматически открывается в сетке данных. Чтобы изменить это поведение, выберите в меню Сервис Параметры Основные данные (**Tools** ⇒ **Options** ⇒ **Data General**) и снимите флажок Автоматически открывать данные при добавлении к проекту (**Automatically open data when added to the project**).



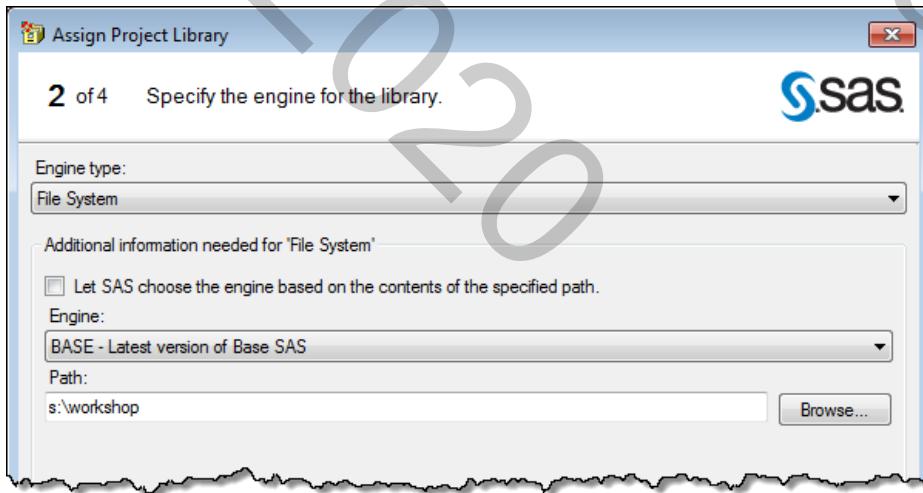
## Создание библиотеки SAS

В демонстрации показано, как создать библиотеку и добавить из неё набор данных в проект.

1. Выберите в меню Файл ⇒ Создать ⇒ Проект (File ⇒ New ⇒ Project).
2. Чтобы создать библиотеку, выберите в меню Сервис ⇒ Назначить библиотеку проекта (Tools ⇒ Assign Project Library).

 Как правило, настройка библиотек, доступных на сервере SAS, является обязанностью администратора SAS. Однако, если требуется создать библиотеку SAS только для определённого проекта или для таблиц DBMS, доступных для чтения, можно воспользоваться задачей «Назначить библиотеку проекта».

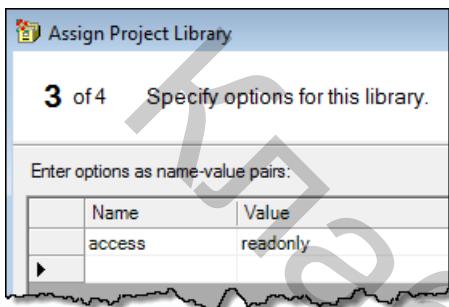
3. На шаге 1 Мастера настроек **Name (имя)** введите **ORION**. Выберите сервер, на котором находятся данные курса, а затем нажмите кнопку **Next (Вперед)**.
  4. На шаге 2 определяется механизм доступа к данным (Engine) и указываются сведения, необходимые для подключения к данным. Заданные по умолчанию настройки типа механизма (Engine type), Файловая система (File system), и названия механизма (Engine), Base SAS, позволяют читать наборы данных SAS. Укажите путь к данным курса и нажмите кнопку **Next (Вперед)**.
-  Чтобы подключиться к DBMS, в поле **Engine Type (тип механизма)** выберите значение **Database System (система баз данных)**, и в раскрывающемся списке **Engine (механизм)** выберите соответствующий механизм SAS/ACCESS.



5. На шаге 3 введите **access** в поле Имя (**Name**) и **readonly** в поле Значение (**Value**). Это гарантирует, что в случае предоставления прав на запись и изменение данных на уровне операционной системы или базы данных, SAS не позволит вносить изменения в источник данных в библиотеке. Нажмите кнопку Вперёд (**Next**).



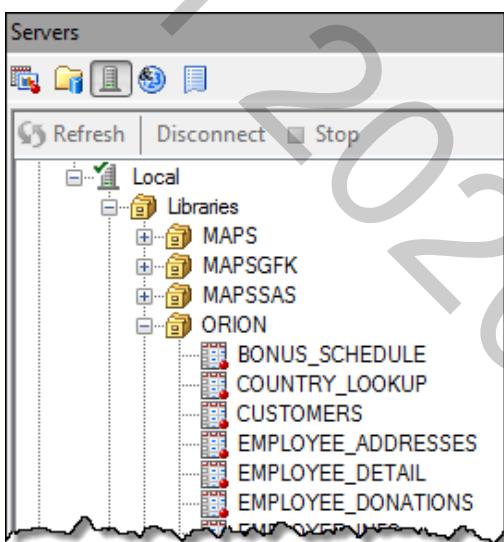
Имена и значения параметров, указываемые на третьем шаге, соответствуют ключевым словам и значениям, используемым в операторе LIBNAME.



6. На шаге 4 проверьте параметры и нажмите кнопку **Test Library** (проверить библиотеку). Должно появиться сообщение «OK». Чтобы завершить создание библиотеки для проекта, нажмите кнопку **Finish** (готово).
7. Для доступа к источникам данных в библиотеке **orion** последовательно выберите **View (вид) ⇒ Server List (серверы)** или нажмите кнопку в области ресурсов. Раскройте узел **Servers (серверы) ⇒ <имя сервера> ⇒ Libraries (библиотеки) ⇒ ORION**. Дважды щелкните **ORDERS**, чтобы добавить ярлык в проект.



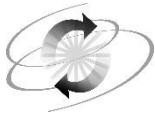
Чтобы просмотреть текущее содержимое библиотеки **orion**, возможно, потребуется обновить список библиотек данного сервера. Для этого нужно выделить имя сервера и нажать кнопку **Refresh**.



8. Сохраните проект как **Chapter2**.

## Упражнения

---



### Уровень 1

#### 1. Назначение библиотеки проекта

Создайте библиотеку **orion** для доступа к наборам данных SAS, содержащим данные о клиентах и сотрудниках компании Orion.

- С помощью задачи «Назначить библиотеку проекта» (Assign Project Library) создайте библиотеку **orion** и укажите путь к расположению данных курса.
- На шаге 3 задачи включите параметр **access=readonly**. Затем проверьте назначение библиотеки (кнопка Test) и запустите задачу.
- Убедитесь, что библиотека **orion** присутствует в списке серверов. Если библиотека или наборы данных отсутствуют, выберите сервер и нажмите кнопку Refresh.
- Сохраните проект **Chapter2**.

### Уровень 2

#### 2. Использование списка серверов для операций с файлами

Используйте список серверов для копирования, вставки и удаления наборов данных.

- С помощью списка серверов можно создать копию набора данных. Чтобы создать временную копию, щелкните правой кнопкой мыши таблицу **travel\_expenses** в библиотеке **orion** и в контекстном меню выберите команду **Copy (копировать)**. Затем щелкните правой кнопкой мыши библиотеку **Work** и выберите команду **Paste (вставить)**.
- В библиотеке **Work** щелкните два раза таблицу **travel\_expenses**, чтобы добавить ее в проект.
- Щелкните правой кнопкой мыши таблицу **travel\_expenses** в дереве проекта или в окне «Process Flow» и выберите команду **Delete (удалить)**.  
Удалилась ли таблица из библиотеки?
- В списке серверов найдите библиотеку **Work**, щелкните в ней правой кнопкой мыши таблицу **travel\_expenses** и выберите команду **Delete (удалить)**.  
Удалось ли удалить таблицу?
- В библиотеке **orion** щелкните правой кнопкой мыши таблицу **travel\_expenses** в списке серверов и выберите команду **Delete**.  
Удалось ли удалить таблицу?

## Импорт файлов с данными

Книгуakov Maxim

Слайд 44 из 100. Всё внимание на Стартовую страницу.

**Цели**

- Научиться импортировать файлы Microsoft Excel и создавать из них наборы данных SAS.
- Импортировать текстовый файл с полями фиксированной ширины в набор данных SAS.

44

Книгуakov Maxim

Слайд 45 из 100. Всё внимание на Стартовую страницу.

**Бизнес-сценарий**

У компании Orion есть файл **products** в формате Microsoft Excel и текстовый файл **orders**. Чтобы использовать данные из них в SAS Enterprise Guide, сначала их нужно импортировать в наборы данных.



45

**2.06 Тест**

Выберите в меню Файл ⇒ Открыть ⇒ Данные (**File** ⇒ **Open** ⇒ **Data**) и перейдите в расположение данных курса. Что произойдёт, если вы откроете файл формата Excel с названием **products**?

46

**Задача Импорт данных (Import Data)**

Эту задачу можно использовать для импорта текстовых файлов или файлов других форматов, в том числе Microsoft Excel и Microsoft Access, для создания из них наборов данных SAS.

48

Мастер импорта данных (Import Data) может считывать следующие форматы данных:

- Microsoft Access 1.x, 2.0, Access 95, Access 97, Access 2000, Access 2002 (XP), Access 2003, Access 2007

В 64-битной версии Enterprise Guide нельзя открыть данные Microsoft Access. Если вам нужна эта функция, воспользуйтесь 32-битной версией Enterprise Guide.

- Microsoft Excel 4-7, Excel 97, Excel 2000, Excel 2002 (XP), Excel 2003, Excel 2007, Excel 2010, Excel 2013
- HTML
- Текстовый формат (с разделителями полей или столбцами фиксированной ширины)
- Все версии Stata (\*.DTA) в Microsoft Windows
- Все версии SPSS (\*.SAV) в Microsoft Windows
- JMP



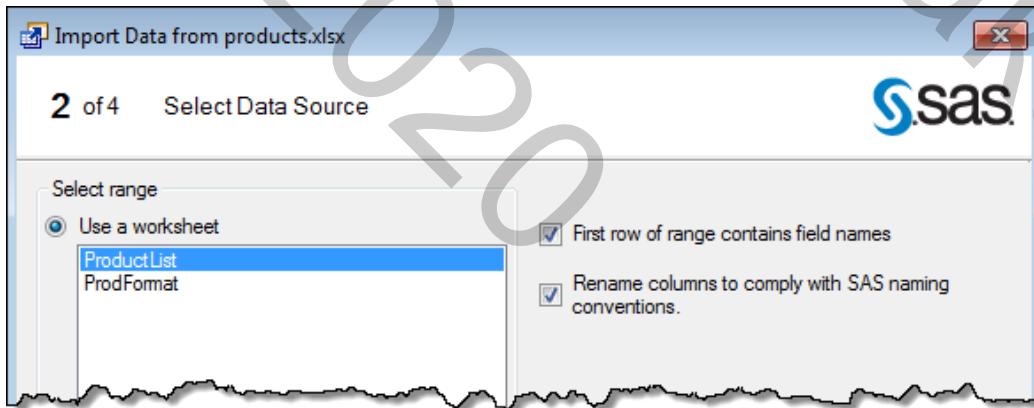
### Обсуждение

- Зачем нужно обязательно использовать задачу Импорт данных перед считыванием текстовых файлов или файлов в формате Excel?
- Какую информацию нужно указать, чтобы данные были импортированы корректно?

## Использование задачи «Импорт данных»

### Часть 1: Импорт электронной таблицы Microsoft Excel с названием Products

- Используйте проект **Chapter2**. Для импорта файла Excel в меню выберите Файл  $\Rightarrow$  Открыть  $\Rightarrow$  Данные (**File  $\Rightarrow$  Open  $\Rightarrow$  Data**). Перейдите к расположению данных курса и выберите файл Excel **products**. Нажмите кнопку Открыть (**Open**).
  - Автоматически откроется Мастер импорта данных. Убедитесь, что исходным файлом с данными является страница Excel **products**, а выходным набором данных SAS — **products**. Нажмите кнопку Вперёд (**Next**).
-  Можно изменить место хранения итогового набора данных SAS и выбрать другую библиотеку. Для определения библиотеки по умолчанию SAS Enterprise Guide выполняет поиск первой доступной для записи библиотеки в следующем порядке: **egtask, gridwork, work, sasuser**. Чтобы изменить библиотеку по умолчанию, последовательно выберите Сервис  $\Rightarrow$  Параметры  $\Rightarrow$  Библиотека выходных данных (**Tools  $\Rightarrow$  Options  $\Rightarrow$  Output Library**).
- На шаге 2 убедитесь, что выделена страница **ProductList** и установлены флажки **Первая строка диапазона содержит имена полей** (**First row of range contains field names**) и **Переименовать столбцы в соответствии с требованиями SAS** (**Rename columns to comply with SAS naming conventions**). Нажмите кнопку Вперёд (**Next**).
-  Параметр **переименовать столбцы в соответствии с требованиями SAS** (**Rename columns to comply with SAS naming conventions**) сокращает имена до 32 символов и заменяет пробелы и специальные знаки знаком подчеркивания.



- На шаге 3 измените имя в первой строке на **Product\_ID**, а ярлык — на **Product ID**.
- Также измените тип **Supplier\_ID** на **String** (строка). Будет выведен запрос, в котором нужно указать способ выбора длины символьной переменной. Убедитесь, что выбран параметр **Проверять все значения в столбце** (**Scan all values in the column**) и нажмите кнопки **OK**  $\Rightarrow$  Далее (**Next**).



Длина символьного столбца соответствует максимальному числу символов, которое может быть сохранено в одном значении данных.

- На шаге 4 примите значения по умолчанию для дополнительных параметров. (Ничего не выбрано.) Нажмите кнопку **Finish**, чтобы завершить импорт.



Для получения сведений о дополнительных параметрах нажмите кнопку **Help (справка)**.

- Проверьте результаты на вкладке «Output Data» (выходные данные). Чтобы просмотреть созданный код шага DATA, откройте вкладку **Code (код)**.

Фрагмент результатов (7 строк из 481)

The screenshot shows the SAS Studio interface with the 'Import Data' window open. The 'Output Data' tab is selected, displaying a preview of 7 rows from the 'ProductList' dataset. The columns are Product\_ID, Product\_Line, Product\_Category, Product\_Group, and Product\_Name. Below this, the 'Code' tab is selected, showing the generated SAS code:

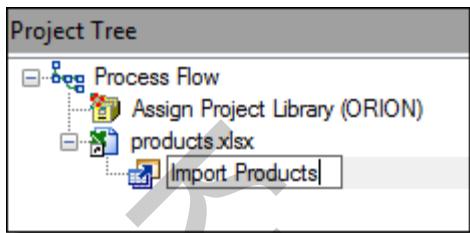
```

/*
  This DATA step reads the data values from a temporary text file
  created by the Import Data wizard. The values within the temporary
  text file were extracted from the Excel source file.
*/

DATA WORK.PRODUCTS_0000;
LENGTH
  Product_ID      8
  Product_Line    $ 15
  Product_Category $ 24
  Product_Group   $ 24
  Product_Name    $ 45
  Supplier_Country $ 2
  Supplier_Name   $ 28
  Supplier_ID     $ 5 ;

```

- Чтобы получить доступ к задаче и новому набору данных SAS, дважды щелкните значок **Import Data (импорт данных)** в проекте. Задачу «Импорт данных» можно переименовать, чтобы выделить в его названии имя создаваемого набора данных SAS. Щелкните правой кнопкой мыши значок **Import Data** в дереве проекта или в окне «Process Flow» и в контекстном меню выберите команду **переименовать (Rename)**. Введите **Import Products**.



9. Сохраните проект **Chapter2**.

АВГУСТ 2020 Год  
Лабуков Максим

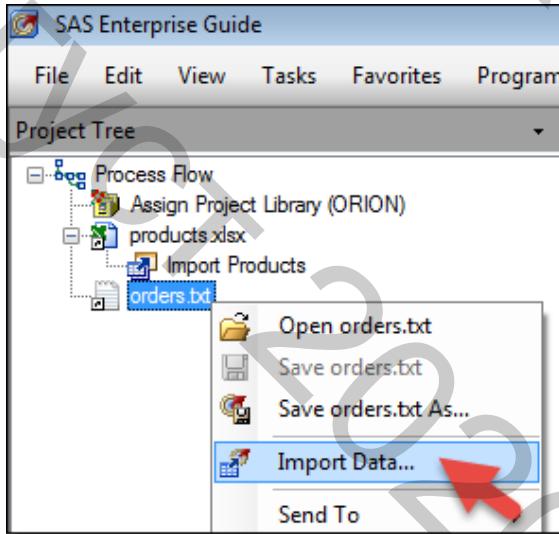


## Использование задачи «Импорт данных» (продолжение)

### Часть 2. Импорт текстового файла «Orders» с полями фиксированной ширины

1. В проекте **Chapter2** последовательно выберите Файл  $\Rightarrow$  Открыть  $\Rightarrow$  Данные (File  $\Rightarrow$  Open  $\Rightarrow$  Data). Перейдите в расположение данных курса и откройте файл **orders.txt**. Файл добавляется в проект, но поскольку он не является структурированным набором данных, его нельзя использовать для создания отчетов или запросов.
2. Чтобы создать пользовательский набор данных SAS из текстового файла с помощью мастера импорта данных, щелкните правой кнопкой мыши файл **orders.txt** в дереве проекта или окне схемы процесса и в контекстном меню выберите команду Импорт данных (**Import Data**).  

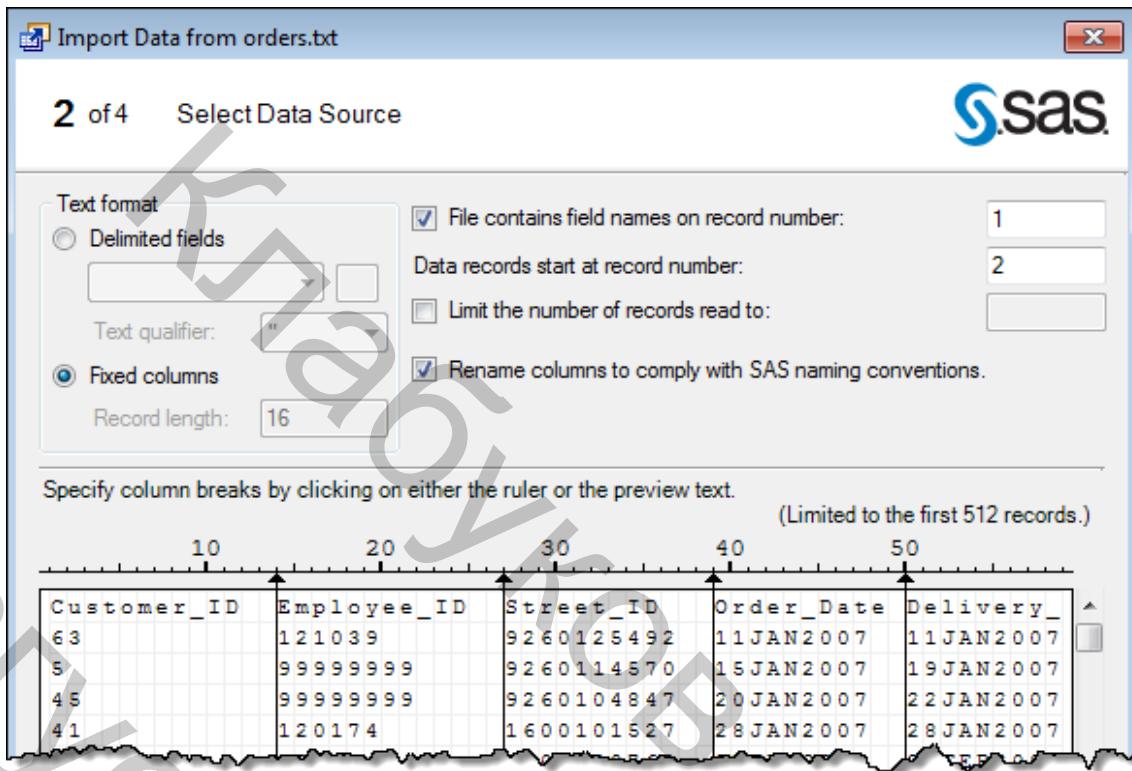
Ее можно также выбрать с помощью меню Файл  $\Rightarrow$  Импорт данных (File  $\Rightarrow$  Import Data).



3. На шаге 1 Мастера импорта данных убедитесь, что для выходного набора данных задано имя **orders**. Не изменяйте библиотеку по умолчанию и нажмите кнопку Вперед (Next).
4. На шаге 2 выберите режим **Фиксированные столбцы (Fixed Columns)**. Установите флажок **Файл содержит имена полей в записи (File contains field names on record number)** и убедитесь, что задано значение **1**. Щелкните мышью в начале каждого поля, чтобы определить границы столбцов. Нажмите кнопку Вперед (Next).



Перед первым полем не требуется указывать границу.



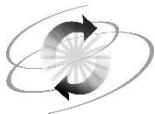
5. На шаге 3 измените свойства переменных, как описано ниже.
  - a. Исключите переменную **Employee\_ID**, сняв флажок Вкл (Inc).
  - b. Отформатируйте **Discount** как процентное значение, дважды щелкнув в поле **формат вывода (Output Format)** и нажав кнопку **...**. В категории числовых форматов выберите **PERCENTw.d**. В качестве общего числа знаков выберите **5** и нажмите кнопку **OK**.
  - c. Отформатируйте **Profit** как значение валюты, выбрав в раскрывающемся списке в столбце **Type (тип)** пункт **Currency (валюта)**. Нажмите кнопку **Вперед (Next)**.
6. На шаге 4 не вносите никаких изменений. Нажмите кнопку Готово (**Finish**), чтобы создать импортированный набор данных SAS.  
Задача «Импорт данных» и набор данных **orders** добавляются в проект.

Фрагмент результатов (6 строк из 617)

The screenshot shows the SAS Enterprise Guide interface. On the left, the Project Tree displays a process flow named 'Process Flow' containing tasks like 'Assign Project Library (ORION)', 'products.xlsx', 'Import Products', and 'orders.txt'. Under 'orders.txt', there is a task named 'Import Data (orders.txt)'. On the right, a data grid titled 'Import Data (orders.txt)' shows six rows of data with columns: Customer\_ID, Street\_ID, and Order\_Date. The data is as follows:

	Customer_ID	Street_ID	Order_Date
1	63	9260125492	11JAN2007
2	5	9260114570	15JAN2007
3	45	9260104847	20JAN2007
4	41	1600101527	28JAN2007
5	183	1600100760	27FEB2007
6	79	9260101874	02MAR2007
7	23	9260126679	03MAR2007
8	23	9260100070	03MAR2007

7. Сохраните проект **Chapter2**.



## Упражнения

### 3. Создание набора данных SAS из таблицы Excel

Добавьте в проект таблицу **Addresses** из книги Excel **employee\_info**.

- В проекте **Chapter2** с помощью Мастера импорта данных откройте книгу Excel **employee\_info**. Создайте набор данных SAS с именем **employee\_addresses**.
  - Набор данных можно сохранить в библиотеке SAS по умолчанию, например **Sasuser**, **Work** или **EGTASK**.
- Считайте данные из таблицы **Addresses** и укажите, что первая строка содержит имена переменных.
- В качестве значения типа переменной **Postal\_Code** укажите **String (строка)**.
  - Проверьте все значения в столбце, чтобы определить подходящую длину для этой переменной.
- Измените ярлык для каждой переменной, заменив знаки подчеркивания пробелами.
- Завершите работу мастера импорта данных и просмотрите набор данных.

Имена переменных отображаются в сетке данных, а ярлыки — в отчетах.

Фрагмент результатов (7 строк из 424)

Import Data (employee_info.xlsx[Addresses])									
	Employee_ID	Employee_Name	Street_ID	Street_Number	Street_Name	City	State	Postal_Code	Country
1	121044	Abbott, Ray	9260116912	2267	Edwards Mill Rd	Miami-Dade	FL	33135	US
2	120145	Aisbit, Sandy	1600101803	30	Bingera Street	Melbourne		2001	AU
3	120761	Akinfolarin, Tameaka	9260121030	5	Donnybrook Rd	Philadelphia	PA	19145	US
4	120656	Amos, Salley	9260123736	3524	Calico Ct	San Diego	CA	92116	US
5	121107	Anger, Rose	9260120989	744	Chapwith Rd	Philadelphia	PA	19142	US
6	121038	Anstey, David	9260116991	939	Hilltop Needmore Rd	Miami-Dade	FL	33157	US
7	120273	Antonini, Doris	9260116925	681	Ferguson Rd	Miami-Dade	FL	33141	US

- Переименуйте задачу «Import Data» (импорт данных) в **Import Employee Addresses**.
- Сохраните проект как **Chapter2**.

Program Log Output Data					
	Employee_ID	Employee_Gender	Salary	Birth_Date	Employee_Hire_Date
1	120101	M	\$163,040.00	18-08-1981	01JUL2008
2	120102	M	\$108,255.00	11-08-1974	01JUN1994
3	120103	M	\$87,975.00	22-01-1954	01JAN1979
4	120104	F	\$46,230.00	11-05-1959	01JAN1986
5	120105	F	\$27,110.00	21-12-1979	01MAY2004
6	120106	M	\$26,960.00	23-12-1949	01JAN1979
7	120107	F	\$30,475.00	21-01-1954	01FEB1979
8	120108	F	\$27,660.00	23-02-1989	01AUG2011

- Сохраните проект **Chapter2**.

## Задачи и мастера настроек

Книги уков

**Цели**

- Понять, как устроена задача в SAS Enterprise Guide.
- Узнать, как можно создать задачу и настроить её.
- Познакомиться с функциями и назначением общих элементов интерфейса для настройки задач.

3

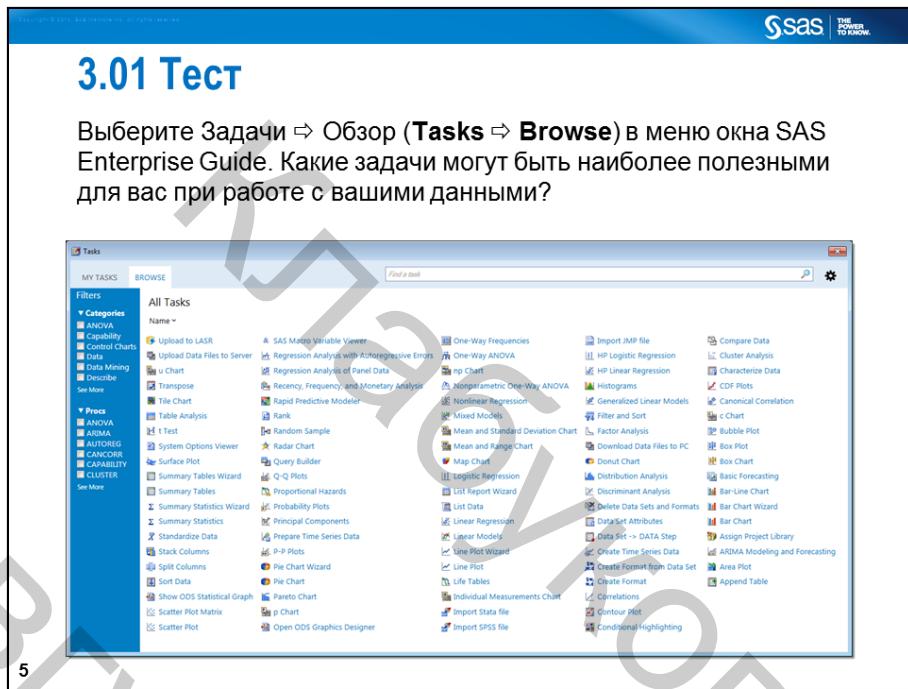
АВГУСТ 2020 Максим

**Что такое задача?**

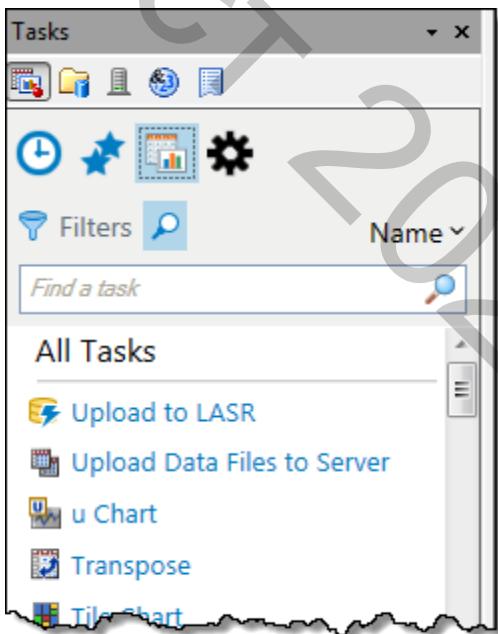
Задача (*task*) – это определённый вид анализа данных, отчёта, или трансформации данных в проекте.

4

Задачи SAS Enterprise Guide создают код SAS и результаты его работы. Задачи включают в себя процедуры SAS — от простого вывода данных в отчёт и до выполнения более сложных аналитических процедур.



Задачи можно запустить из окна Задачи (Tasks) в панели ресурсов. Это окно можно открыть из меню Вид ⇒ Задачи (View ⇒ Tasks).



**Избранные задачи (Favorites)**

Задержите курсор мыши над любой задачей и нажмите звезду. Задача будет добавлена в список избранных.

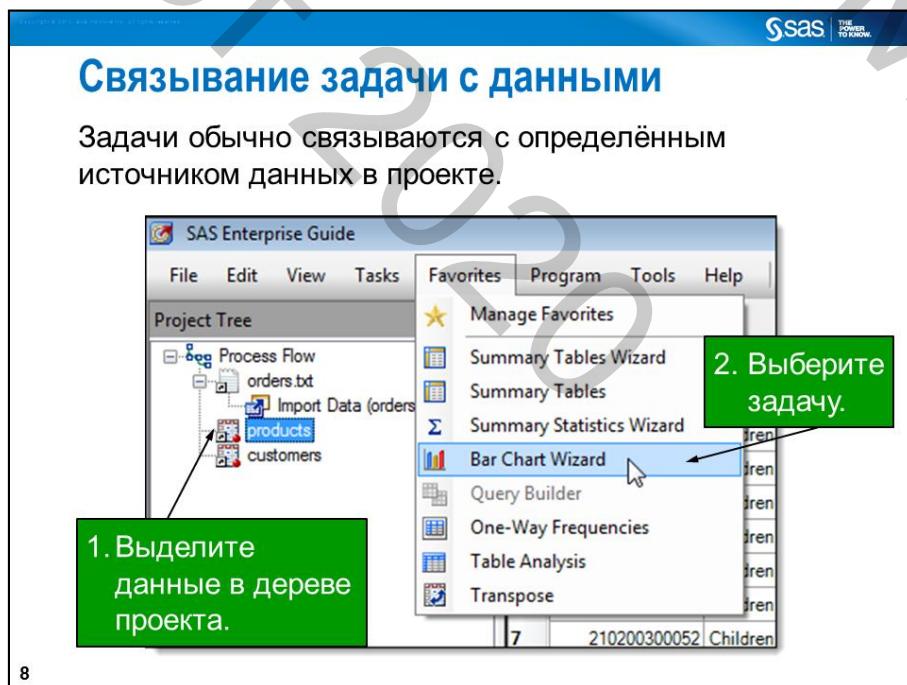
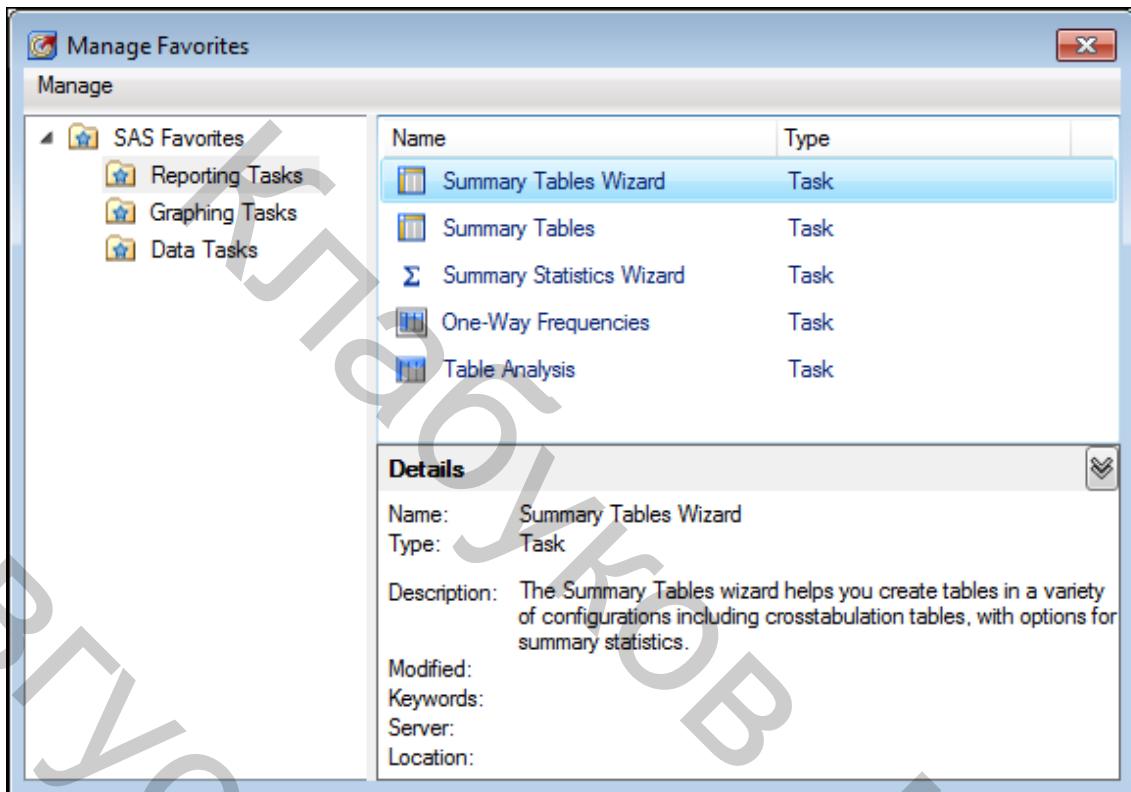
6

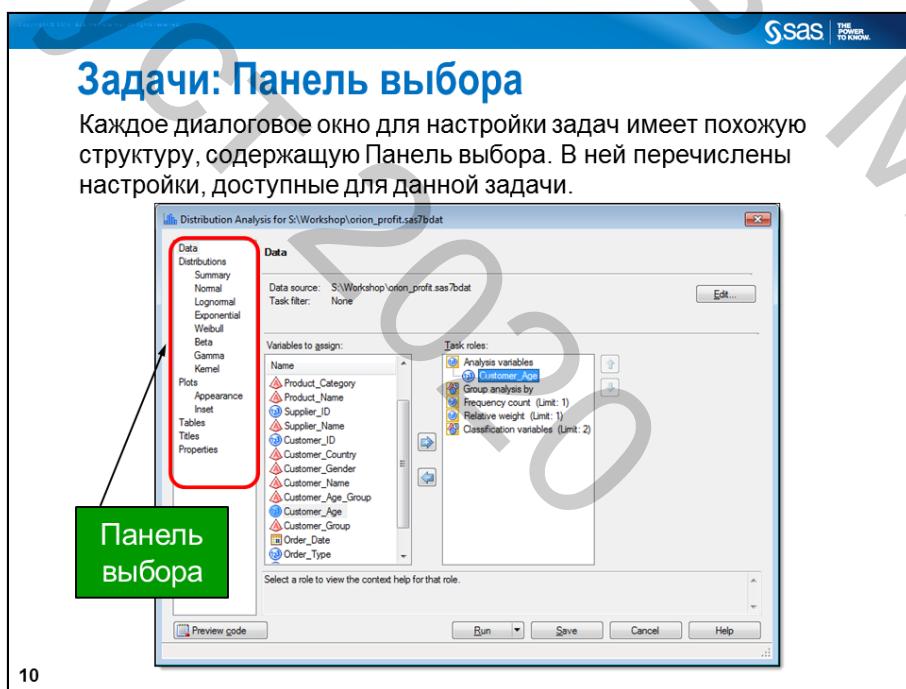
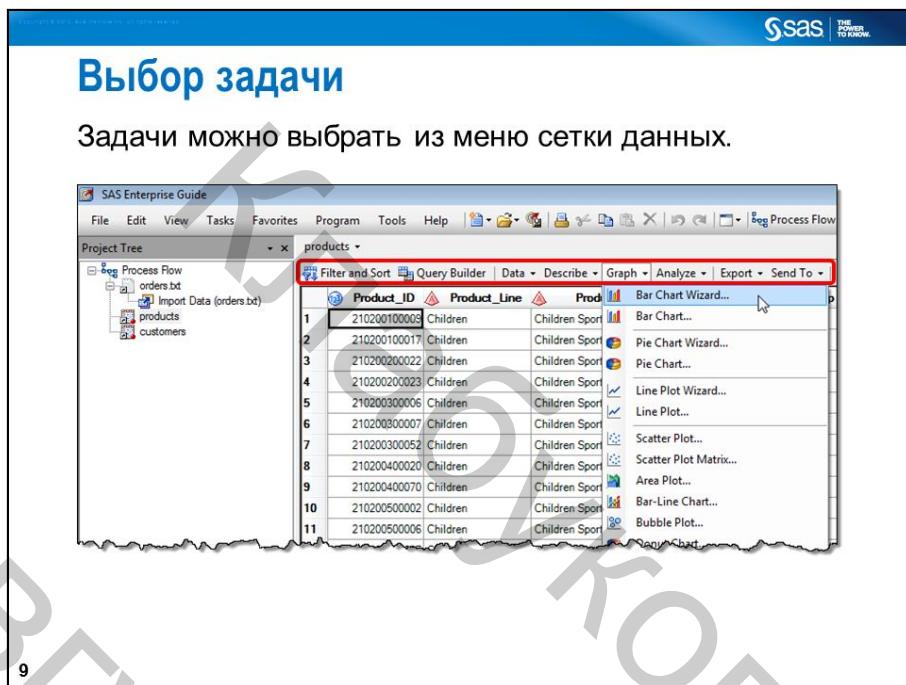
**Использование списка Избранных задач**

Список избранных задач можно открыть из меню Избранное или в секции Задачи (Tasks) Панели ресурсов.

7

Вы можете настроить список избранных задач в меню Избранное ⇒ Упорядочить избранное (Favorites ⇒ Manage Favorites). Для группировки задач можно создать для них отдельные папки.





**Задачи: Данные**

Для настройки того, как будут использоваться столбцы при создании отчёта, выберите **Данные (Data)**.

Distribution Analysis for S:\Workshop\orion\_profit.sas7bdat

Data

Data source: S:\Workshop\orion\_profit.sas7bdat  
Task filter: None

Variables to assign:

- Name
- Product\_Category
- Product\_Name
- Supplier\_ID
- Supplier\_Name
- Customer\_ID
- Customer\_Country
- Customer\_Gender
- Customer\_Name
- Customer\_Age\_Group
- Customer\_Age
- Customer\_Group
- Order\_Date
- Order\_Type

Task roles:

- Analysis variables (Customer\_Age)
- Group analysis by
- Frequency count (Limit: 1)
- Relative weight (Limit: 1)
- Classification variables (Limit: 2)

Specifies the numeric variables to be analyzed. An analysis variable is required and you can specify up to two additional variables for this role.

Preview code Run Save Cancel Help

11

### 3.02 Тест

- Выберите Файл ⇒ Создать ⇒ Проект (**File ⇒ New ⇒ Project**).  
Перейдите в расположение данных курса.
- Выберите **Orion\_Profit** ⇒ Открыть (**Open**).
- В сетке данных нажмите Описательная статистика ⇒ Анализ распределения (**Describe ⇒ Distribution Analysis**).
- Перенесите **Customer\_Age** в роль Переменные анализа (**Analysis variables**).
- Щёлкните правую кнопку мыши на элементе **Customer\_Age** и нажмите Свойства (**Properties**).

Какие атрибуты столбца вы можете поменять?

12

**Задачи: Заголовки и сноски**

Выберите Заголовки (Titles) для добавления заголовка и сносок в отчёте.

Distribution Analysis for S:\Workshop\orion\_profit.sas7bdat

Titles

Section: Analysis Titles

Text for section: Analysis Titles

Use default text

Distribution Analysis for Customer Age

14

Ведите текст заголовка или сноски

**Задачи: Свойства**

Здесь можно переименовать задачу, поменять её атрибуты и настройки вывода.

Distribution Analysis for S:\Workshop\orion\_profit.sas7bdat

Properties

Label: Distribution Analysis

Server: Local

Created: 10/2/2014 10:04:58 AM

Last modified by:

Last execution time:

Limits on data:

Prompts:

Result formats (Style):

Graph format:

Properties

Edit...

Distribution Analysis task

General

Label: Distribution Analysis

Last execution time: Unknown

Server: Local

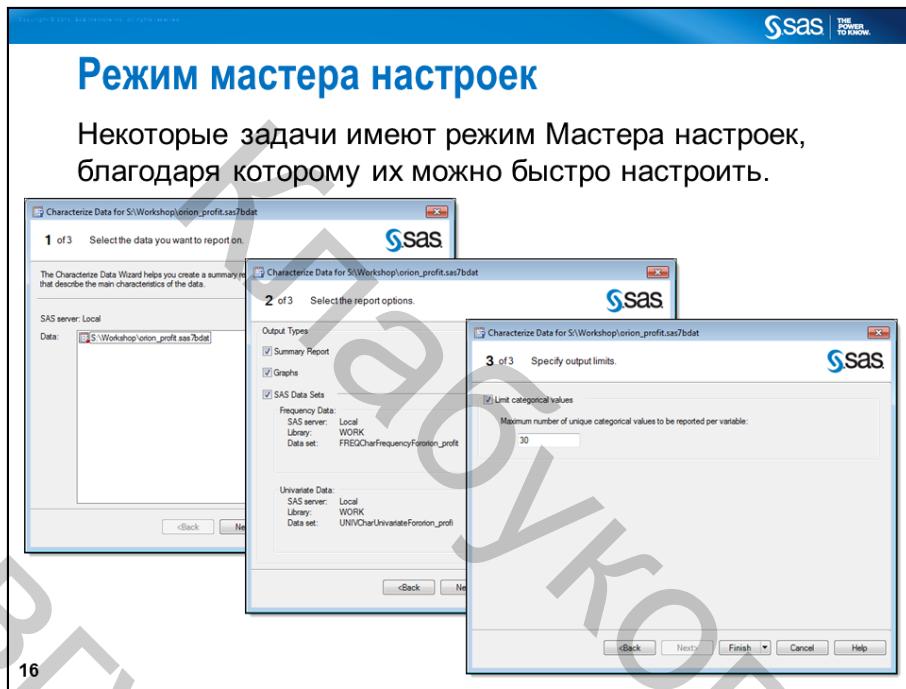
Input data: orion\_profit

Data limits

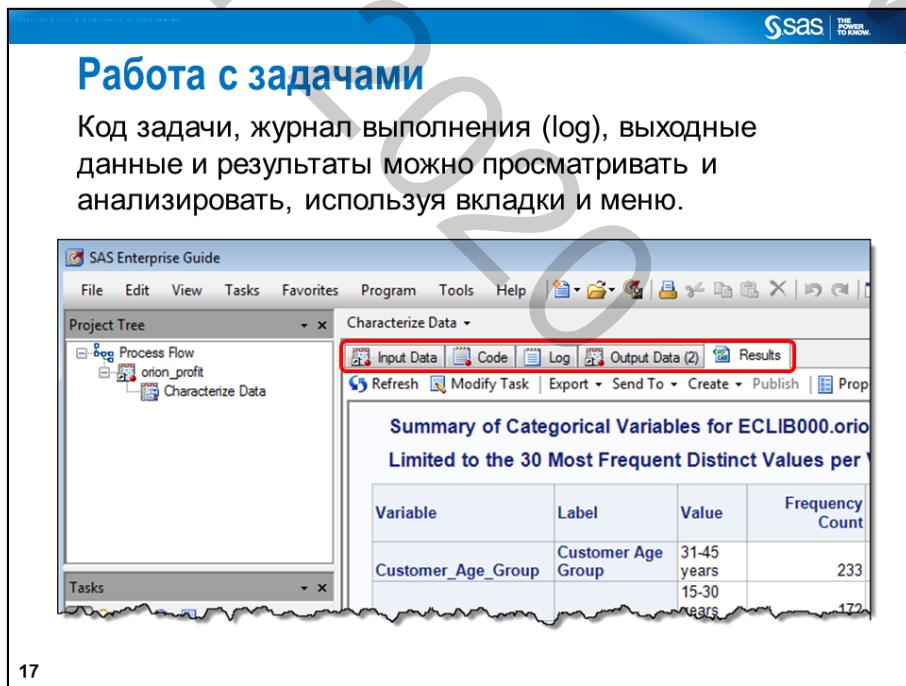
Start at row: 1

Limit the number of rows used: 10000

15



Задачи Сводная статистика (Summary Statistics), Сводные таблицы (Summary Tables), Столбчатая диаграмма (Bar Chart), Секторная диаграмма (Pie Chart) и Линейный график (Line Plot) имеют режим Мастера настроек. Если задача имеет два режима настроек (традиционный и Мастер настроек), то они указываются отдельно в меню выбора задач.



## Работа с задачами

Задачи можно обновлять или изменять из дерева проекта, диаграммы процесса или вкладки результатов.

The screenshot shows the SAS Enterprise Guide interface. In the top right corner, the SAS logo is visible. The main window title is 'SAS Enterprise Guide'. Below it, the menu bar includes File, Edit, View, Tasks, Favorites, Program, Tools, and Help. A 'Project Tree' pane on the left shows a 'Process Flow' node with a sub-node 'onion\_profit' and a 'Characterize Data' task. To the right of the tree is a toolbar with icons for Input Data, Code, Refresh, and Modify. A context menu is open over the 'Characterize Data' task, with options 'Open', 'Run Characterize Data', and 'Modify Characterize Data'. The 'Run Characterize Data' option is highlighted with a red box. Below the tree, a 'Select Input Data' button is shown. On the far right, a results pane titled 'Characterize Data' displays a summary of categorical variables for ECLIB01, limited to the 30 most frequent distinct values. The results pane has tabs for Input Data, Code, Log, Output Data (2), Results, Refresh, and Modify Task. The 'Results' tab is selected. The results table shows columns for Variable, Label, Value, and Freq. The first few rows are: Variable, Label, Value, Freq. The number 18 is located at the bottom left of the screenshot area.

18

Клабуков Максим  
Август 2020

## Создание частотного отчёта

21

**Цели**

- Познакомиться с задачей Одномерный анализ частот (One-Way Frequencies).
- Создать частотный отчёт и столбчатую диаграмму.

22

**Бизнес-сценарий**

Руководству компании Ogiон нужен отчёт, где подсчитано число предлагаемых продуктов для каждой категории.

Product_ID	Product_Line	Product_Category
433	240500200122	Sports
434	240500200130	Sports
435	240600100010	Sports
436	240600100016	Sports
437	240600100017	Sports
438	240600100080	Sports
439	240600100102	Sports
440	240600100181	Sports
441	240600100185	Sports
442	240700100001	Sports
443	240700100034	Sports

Product Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Indoor Sports	15	3.12
Outdoors	87	18.09
Racket Sports	14	2.91
Running - Jogging	25	5.20
Shoes	34	7.07
Swim Sports	7	1.46
Team Sports	23	4.78
Winter Sports	17	3.53

**Задача Одномерный анализ частот**

Задачу Одномерный анализ частот (One-way frequencies) можно использовать для подсчёта частот, процентных долей, создания графиков и расчёта статистических тестов для анализа распределения данных.

**Customers in Each Country**

Customer_Country	Customer Country		Cumulative Frequency	Cumulative Percent
	Frequency	Percent		
AU	8	10.39	8	10.39
CA	15	19.48	23	29.87
DE	10	12.99	33	42.86
IL	5	6.49	38	49.35
TR	7	9.09	45	58.44
US	28	36.36	73	94.81
ZA	4	5.19	77	100.00

**Distribution of Customer\_Country**

23

Отчёт с одномерным анализом частот показывает распределение значений переменной.

Чтобы запустить задачу Одномерный анализ частот (One-Way Frequencies), выберите в меню Задачи⇒Описательная статистика⇒Одномерный анализ частот (**Tasks** ⇒ **Describe** ⇒ **One-Way Frequencies**). Также можно выбрать Описательная статистика⇒Одномерный анализ частот (**Describe** ⇒ **One-Way Frequencies**) в меню сетки данных.

**3.03 Вопрос с выбором нескольких вариантов ответа**

Выделите любой источник данных в проекте, выберите в меню Задачи ⇒ Описательная статистика ⇒ Одномерный анализ частот (**Tasks** ⇒ **Describe** ⇒ **One-Way Frequencies**).

Какую роль требуется обязательно задать?

- Переменные анализа (Analysis variables)
- Подсчёт частот (Frequency count)
- Группирующие переменные (Group analysis by)

24

**Одномерный анализ частот: Роли**

Задача Одномерный анализ частот создаёт частотную таблицу для каждой переменной, которая добавлена в роль Переменные анализа (Analysis variables).

Product_Category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Assorted Sports Articles	68	14.14	68	14.14
Children Sports	27	5.61	95	19.75
Clothes	108	22.45	203	42.20
Golf	56	11.64	259	53.85
Indoor Sports	15	3.12	274	56.96
Outdoors	87	18.09	361	75.05
Racket Sports	14	2.91	375	77.96
Running - Jogging	25	5.20	400	83.16
Shoes	34	7.07	434	90.23
Swim Sports	7	1.46	441	91.68
Team Sports	23	4.78	464	96.47
Winter Sports	17	3.53	481	100.00

26

Для создания двумерных таблиц (таблиц сопряженности, таблиц кросс табуляции) нужно использовать задачу Табличный анализ (Table Analysis).

**Одномерный анализ частот: Роли**

Роль Группирующие переменные (Group analysis) можно использовать для разбивки отчёта по уникальным значениям переменной (или комбинации переменных).

Product_Category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Clothes	3	50.00	3	50.00
Outdoors	3	50.00	6	100.00

Product_Category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Running - Jogging	14	100.00	14	100.00

27

**Настройка для Теста**

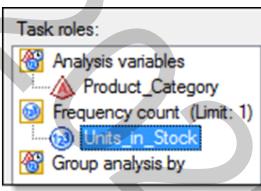
Снова откройте задачу Одномерный анализ частот (One-Way Frequencies) выбрав Задачи  
 ⇨ Описательная статистика ⇨ Одномерный анализ частот (**Tasks** ⇨ **Describe** ⇨ **One-Way Frequencies**). В окне Роли задачи (Tasks Roles) нажмите Справка (**Help**) и прочитайте описание роли Подсчёт частот (Frequency count).

28

**3.04 Вопрос с выбором нескольких вариантов ответа**

Каким будет значение частоты для **Children Sports**, если в роль Подсчет частот (Frequency count) добавлена переменная **Units\_in\_Stock**?

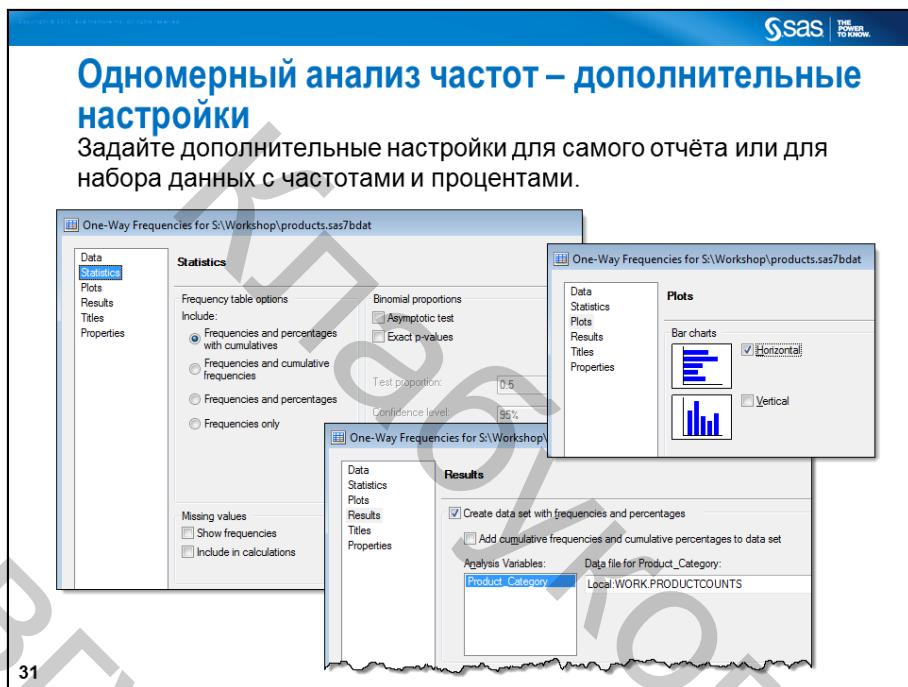
a. 3  
 b. 25  
 c. 50  
 d. 70



Product_Name	Product_Category	Units_in_Stock
Children's Mitten	Children Sports	25
Sunfit Swimming Trunks	Children Sports	15
Osprey Nylon Shorts	Children Sports	10
Liga Football Boot	Shoes	5
Pytossage Bathing Sandal	Shoes	15

Product_Category	Frequency
Children Sports	??
Shoes	??

29



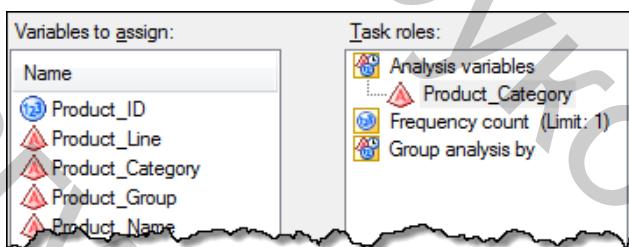
31



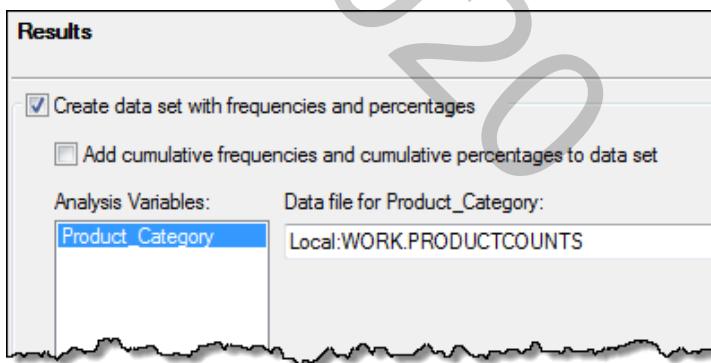
## Создание одномерного частотного отчёта

Руководителям компании Orion требуется сводный отчет по количеству продуктов, предлагаемых в каждой категории. Создайте этот отчет с помощью задачи «Одномерный анализ частот» (One-Way Frequencies).

1. Создайте проект и добавьте в него набор данных SAS **products** из данных курса.
2. Чтобы открыть задачу «Одномерный анализ частот» (One-Way Frequencies), в строке меню выберите Задачи  $\Rightarrow$  Описательная статистика  $\Rightarrow$  Одномерный анализ частот (**Tasks**  $\Rightarrow$  **Describe**  $\Rightarrow$  **One-Way Frequencies**).
3. Перетащите переменную **Product\_Category** из области Переменные для назначения (Variables to assign) в роль Переменные анализа (Analysis variables) в области «Task roles» (Роли задачи).



4. В области выбора настроек выберите Статистика (Statistics). Чтобы включить только вывод частот и процентов, выберите Частоты и процентные доли (Frequencies and percentages) в области параметров таблицы частоты.
5. В области выбора настроек щелкните Результаты (Results). Чтобы создать набор данных с подсчитанными частотами и процентами, установите флажок Создать набор данных с частотами и процентными долями (Create data set with frequencies and percentages). Нажмите кнопку Обзор (Browse) и введите ProductCounts в поле Имя файла (File name). Нажмите кнопку Сохранить (Save).



6. В области выбора настроек выберите Заголовки (Titles). Чтобы изменить заголовок отчёта, снимите флажок использовать текст по умолчанию (Use default text). В поле Текст (Text) удалите исходный заголовок по умолчанию One-Way Frequencies Results и введите Number of Products per Category.



Если вы хотите поменять нижний колонтитул для этого отчёта, в области выбора раздела выберите Сноска (Footnotes) вместо Анализ (Analysis).

7. Чтобы задать для значка задачи «Одномерный анализ частот» (One-Way Frequencies) более подходящую подпись, в области выбора щелкните **Свойства (Properties)** и выберите команду **изменить (Edit)**. В поле **Ярлык (Label)** введите **Products per Category** и нажмите кнопку **OK**.
8. Нажмите кнопку Пуск (**Run**), чтобы создать отчет и проверить результаты. Просмотрите отчет на вкладке Результаты (Results). Перейдите на вкладку **выходные данные (Output Data)**, чтобы просмотреть набор данных.

Фрагмент результатов

Number of Products per Category		
The FREQ Procedure		
Product_Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Indoor Sports	15	3.12
	8	0.00

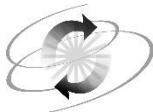
9. Чтобы удалить строчку **FREQ Procedure**, которая автоматически добавляется в отчёт, выберите Сервис ⇔ Параметры (**Tools** ⇔ **Options**). В области выбора настроек выберите **Общие для задач (Tasks General)**. Снимите флажок **Включать заголовки процедур SAS в результаты (Include SAS procedure titles in results)**. Нажмите кнопку **OK**, чтобы закрыть окно.

 В той же области вы можете изменить или удалить используемый по умолчанию текст нижнего колонтитула.

10. Чтобы повторно выполнить задачу, нажмите кнопку Обновить (**Refresh**) на вкладке Результаты (Results).

Number of Products per Category		
Product_Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Indoor Sports	15	3.12
Outdoors	87	18.09
Racket Sports	14	2.91
Running - Jogging	25	5.20
Shoes	34	7.07
Swim Sports	7	1.46
Team Sports	23	4.78
Winter Sports	17	3.53

11. Сохраните проект как **Chapter3**.



## Упражнения

---

### 1. Использование задачи «Одномерный анализ частот» для подсчета частот и процентов

- a. Добавьте в проект **Chapter3** набор данных **employee\_organization**.
  - Если проект **Chapter3** отсутствует, создайте его.
- b. С помощью задачи «Одномерный анализ частот» (One-Way Frequencies) создайте отчет, анализирующий столбец **Department**.
- c. Включите в отчет горизонтальную столбчатую диаграмму.
- d. Добавьте название **Employee Counts by Department** и удалите нижний колонтитул.
- e. Переименуйте задачу в **Dept Freq Report**.
- f. Сформируйте отчет и проверьте результаты.
- g. Измените настройки созданной вами задачи так, чтобы в таблице отображались только частоты и проценты. Повторно выполните отчет и замените результаты.
- h. Сохраните проект как **Chapter3**.

Фрагмент результатов (график не показан)

Employee Counts by Department		
Department	Frequency	Percent
Accounts	17	4.01
Accounts Management	9	2.12
Administration	34	8.02
Concession Management	11	2.59
Engineering	9	2.12
Executives	4	0.94
Group Financials	3	0.71
Group HR Management	18	4.25
		5.90



Предполагается, что при создании окончательных результатов параметр «Включать заголовки процедур SAS в результаты» («Include SAS procedure titles in results») был отключен. Этот параметр можно изменить в меню Сервис Параметры Общие для задач (**Tools** ⇒ **Options** ⇒ **Tasks General**).

## Создание вывода в форматах HTML, PDF и RTF

**Цели**

- Узнать, какие форматы вывода можно настроить.
- Сравнить различные форматы вывода.

37

**Бизнес-сценарий**

Руководству компании Orion требуется разместить частотный отчет по категориям продуктов во внутренней сети в формате HTML.

Number of Products per Category		
Product_Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Indoor Sports	15	3.12
Outdoors	87	18.09
Racket Sports	14	2.91
Running - Jogging	25	5.20
Shoes	34	7.07
Swim Sports	7	1.46
Team Sports	23	4.78
Winter Sports	17	3.53

38

**Форматы вывода задач**

По умолчанию SAS Enterprise Guide создаёт отчёты в формате SAS. Вы можете также выбрать форматы вывода HTML, PDF, RTF или текстовый.

Number of Products per Category

Product_Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Teddy Sports	15	3.24

Number of Products per Category

Product_Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Teddy Sports	15	3.24

Number of Products per Category

Product_Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Teddy Sports	15	3.24

Number of Products per Category

Product_Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Teddy Sports	15	3.24

39

**Формат отчёта SAS**

Формат отчёта SAS позволяет объединить несколько таких отчётов в один. В свою очередь, этот объединённый отчёт можно просматривать в приложениях SAS или экспортirовать в форматы HTML, XML или PDF.

ORION STAR SPORTS & OUTDOORS

Profit by Product Line

Customers in Each Country

Customer_Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AU	15	19.49	23	29.57
CA	10	12.99	33	42.86
DE	5	6.49	38	48.25
IL	1	1.29	39	49.44
TR	1	1.29	40	50.00

Summary of Profit by Country

Customer_Country	Analysis Variable : Profit	Mean	Sum	Min	Max
AU	71.56	9231.34	44.30	126.94	6926.48
CA	126.94	6926.48	64.50	138.33	65.00
DE	43.39	824.40	31.25	104.44	7115.50
IL	104.44	7115.50	67.05	126.94	65.00
TR	20	140	10.00	20	140

40

 THE POWER TO KNOW.

### 3.05 Вопрос с выбором нескольких вариантов ответа

Какие форматы отчетов будут нужны в вашей работе?

- a. Формат отчёта SAS
- b. HTML
- c. PDF
- d. RTF
- e. Текстовый формат

41

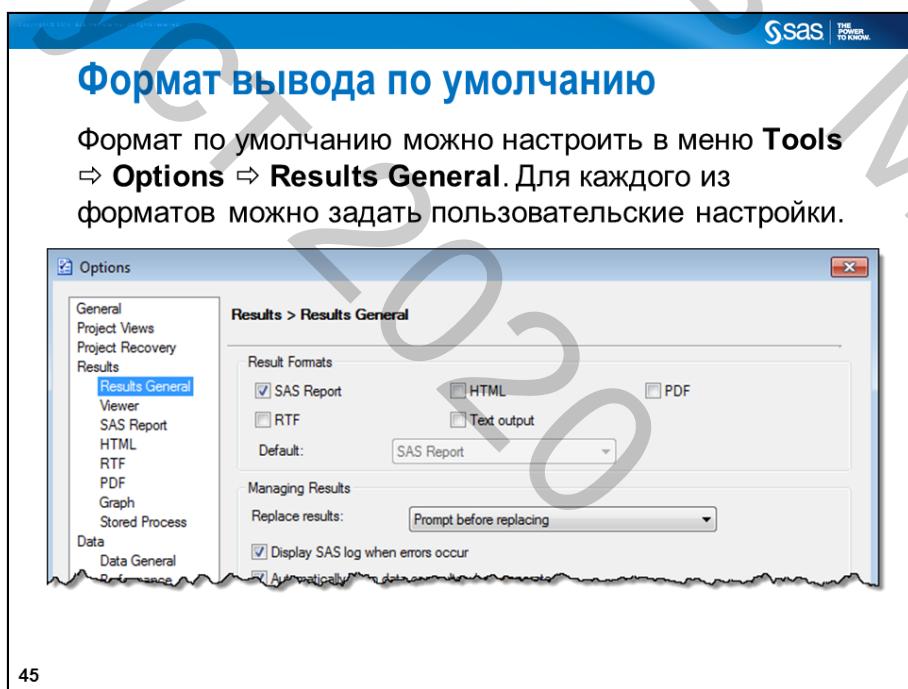
 THE POWER TO KNOW.

### 3.06 Тест

Какими преимуществами обладает каждый формат отчёта?

42

	Отчёт SAS	HTML	RTF	PDF	Текстовый
Программа для просмотра без SAS Enterprise Guide	Другие приложения SAS	Браузер (Internet Explorer)	Любой текстовый процессор (Microsoft Word)	Adobe Reader	Любой редактор текста (Блокнот)
Можно ли объединять результаты в один документ?	Да ( <b>File</b> ⇒ <b>New</b> ⇒ <b>Report</b> )	Да ( <b>Tools</b> ⇒ <b>HTML Document</b> )	Нет (только код)	Нет (только код)	Нет (только код)
Можно ли редактировать результаты в программе для просмотра?	Да	Нет	Да	Нет	Да
Результаты отформатированы для печати?	Да	Нет	Да	Да	Да



Скачано 2015 годом Максимом

## Переопределение формата вывода

В настройках каждой отдельной задачи можно переопределить формат вывода и создать отчет в других форматах.

The screenshot shows two overlapping windows. The top window is titled 'Properties' for a task named 'Products per Category'. It lists details like Label, Server, Created, Last modified by, Last execution time, and Limits on data. The 'Properties' tab is selected. The bottom window is titled 'Properties for Products per Category' and has a 'Results' tab selected. This window contains options for customizing result formats, styles, and behavior. Under the 'Results' tab, the 'Customize result formats, styles, and behavior' radio button is selected. Underneath, several checkboxes are checked: 'SAS Report (HtmlBlue)', 'HTML', 'PDF', 'RTF', and 'Text'. To the right of these checkboxes are dropdown menus for 'Format' (set to 'HtmlBlue'), 'Style' (set to 'EGDefault'), 'Printer' (set to 'Printer'), and 'Rtf' (set to 'Rtf'). A checkbox 'Automatically open data or results when generated' is also checked. At the bottom of this window are 'Reset Options' and 'OK' buttons.

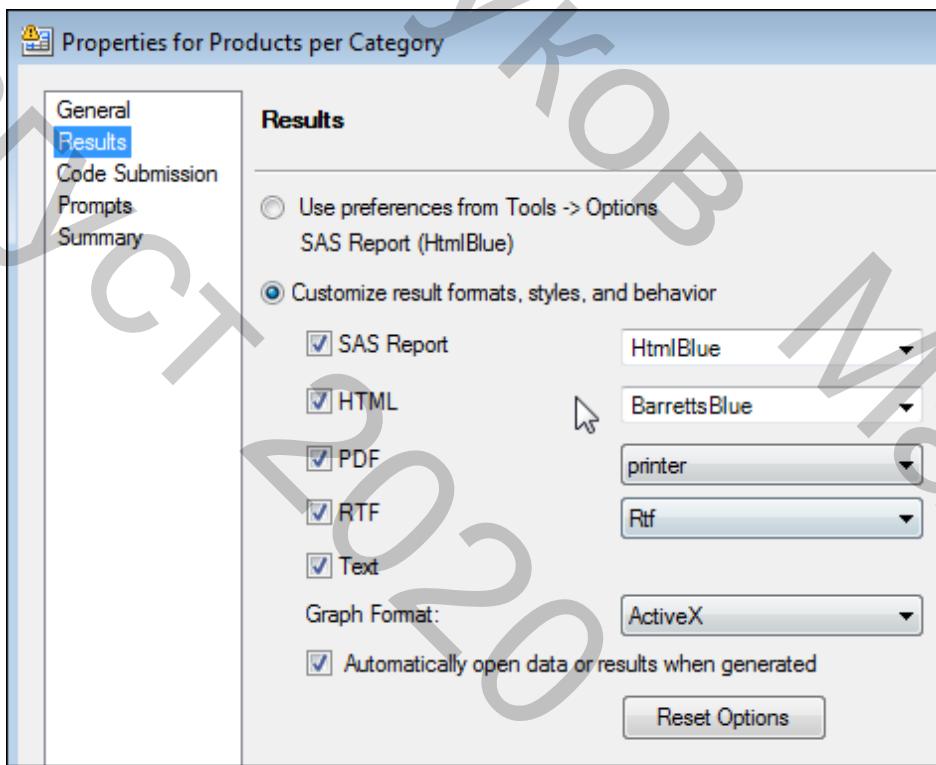
46



## Изменение формата результатов

В этой демонстрации показано, как изменить свойства задачи, чтобы вывести результаты в форматах HTML, PDF, RTF или в текстовом формате.

1. В проекте **Chapter3** щелкните правой кнопкой мыши задачу **Products per Category** в дереве проекта или в окне схемы процесса и просмотрите вкладку Результаты (Results). Нажмите кнопку Изменить задачу (Modify Task).
2. В области выбора настроек выберите Свойства (Properties) и нажмите кнопку Изменить (Edit).
3. В области выбора настроек выберите Результаты (Results). Установите флажок Настроить форматы, стили и функции результатов (Customize result formats, styles, and behavior). Выберите **HTML, PDF, RTF** и **Text** (текст), чтобы просмотреть результаты во всех возможных форматах вывода. В меню рядом с **HTML** выберите **BarrettsBlue** и нажмите кнопку **OK**.



Нажмите кнопку **Run**, чтобы сформировать отчет во всех форматах вывода. При запросе на замену результатов нажмите кнопку **Да (Yes)**. Чтобы просмотреть отчеты, щелкните вкладки с результатами.

Фрагмент полученных результатов (Обратите внимание на вкладки «SAS Report», «HTML», «PDF», «RTF» и «Listing».)

Product_Category	Frequency	Percent
Assorted Sports Articles	68	14.14
Children Sports	27	5.61
Clothes	108	22.45
Golf	56	11.64
Indoor Sports	15	3.12

4. Сохраните проект **Chapter3**.

Клабуков Максим  
Август 2020

## Создание распечатки (Списочный отчёт)

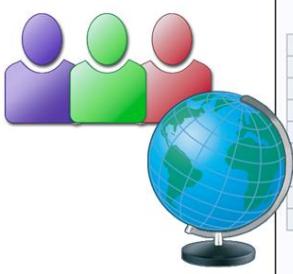
**Цели**

- Узнать про задачу Списочный отчёт (List Data).
- Научиться настраивать параметры отчёта.
- Научиться разбивать отчёт на группы.

50

**Бизнес-сценарий**

Отделу маркетинга в компании Orion Star нужно создать список клиентов, включающий их код, имя и возрастную категорию. Отчёт должен быть сгруппирован по странам.



Orion Star Customer List		
Customer Country=AU		
Customer ID	Name	Age Group
29	Candy Kinsey	61+ years
41	Wendell Summersby	31-45 years
53	Dericka Pockran	46-60 years
111	Karolina Dokter	31-45 years
171	Robert Bowerman	31-45 years
183	Duncan Robertsshawe	61+ years
195	Cosi Rimmington	61+ years
215	Ramesh Trentholme	46-60 years

Customer Country=CA		
Customer ID	Name	Age Group
11171	Bill Cuddy	15-30 years
17023	Susan Krasowski	46-60 years

51

**Задача Списочный отчёт**

Эта задача выводит строки набора данных, позволяя объединять их в группы, а также вычислять суммы по столбцам.

onion\_profit ▾

Filter and Sort ▾ Query Builder | Data ▾ Describe ▾ Graph ▾ Analyze

	Order_ID	Product_Category	Quantity	Profit
1	1230058123	Clothes	1	\$9.05
2	1230080101	Outdoors	1	\$137.95
3	1230106883	Swim Sports	1	\$19.75
4	1230147441	Swim Sports	2	\$19.00
5	1230315085	Golf	3	\$37.20
6	1230333319	Assorted Sports Articles	1	\$118.65
7	1230338566	Winter Sports	1	\$19.25
8	1230338566	Winter Sports	2	\$40.10
9	1230371142	Running - Jogging	2	\$71.00
10	1230371142	Running - Jogging	1	\$34.50

Product Category=Indoor Sports			
Order ID	Quantity Ordered	Profit	
1231976710	4	\$47.50	
1231976710	2	\$77.90	
1233092596	1	\$7.20	
1233315988	1	\$825.45	
1233315988	1	\$100.20	
1233378724	3	\$850.65	
	1	\$30.55	
	1	\$31.10	
1240512244	2	\$14.70	
1241235281	1	\$36.35	
1242185055	1	\$11.00	
1243797399	2	\$563.70	
1243797399	1	\$175.15	
Product_Category	27	\$3,412.65	

Product Category=Outdoors			
Order ID	Quantity Ordered	Profit	
12300		\$137.95	
12304		\$66.60	
12304		\$64.00	
12304		\$34.30	
12304		\$34.15	
123064227	1	\$67.50	

52



## Использование задачи «Списочный отчет» (List Data) для создания отчета

С помощью задачи «Списочный отчет» (List Data) создайте список клиентов, включающий их идентификаторы, имена и возрастные группы. Для каждой страны необходимо создать отдельный отчет.

- Добавьте в проект **Chapter3** набор данных **customers**.

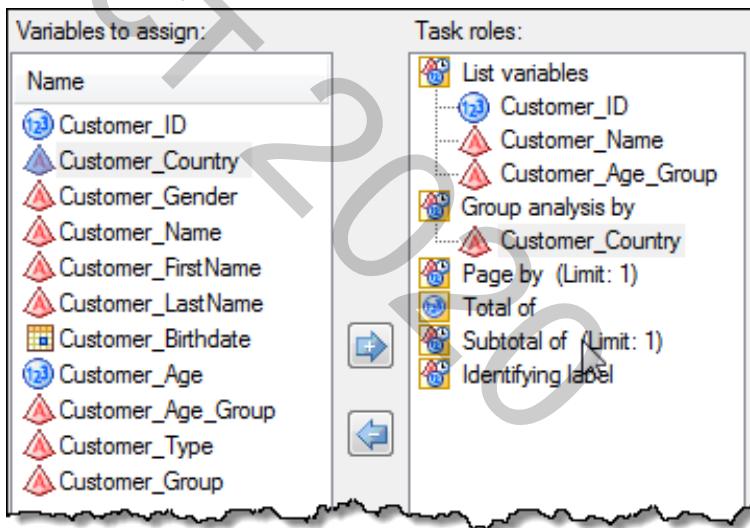


Если проект **Chapter3** отсутствует, создайте его.

- Чтобы открыть задачу «Списочный отчет» (List Data), в сетке данных выберите Описательная статистика  $\Rightarrow$  Списочный отчет (**Describe**  $\Rightarrow$  **List Data**). Откроется диалоговое окно с настройками задачи, в котором можно выбрать столбцы и определить для них роли в отчете.
- Перетащите переменные **Customer\_ID**, **Customer\_Name** и **Customer\_Age\_Group** из области «Переменные для назначения» (Variables to assign) в роль Перечислить переменные (**List variables**). Столбцы, помещенные в эту роль, отображаются в отчете в том порядке, в котором они указаны. Перетащите переменную **Customer\_Country** в роль Группирующие переменные (**Group analysis by**). Для каждой страны создается отдельный отчет.



Для назначения переменных ролям также можно использовать кнопку .



- Чтобы задать название для отчета, в области выбора выберите **Заголовки** (Titles). Снимите флажок **Использовать текст по умолчанию** (Use default text). В поле **Text** удалите заголовок по умолчанию и введите **Orion Star Customer List**.
- Нажмите кнопку Пуск (Run), чтобы сформировать отчет.

Фрагмент результатов

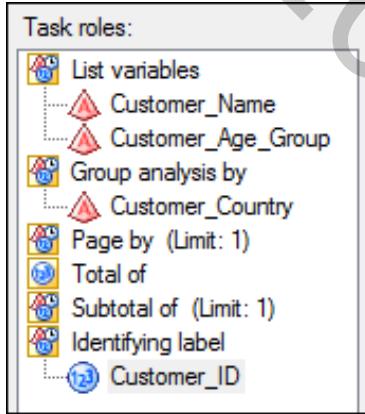
Orion Star Customer List			
Customer Country=AU			
Row number	Customer ID	Customer Name	Customer Age Group
1	29	Candy Kinsey	61+ years
2	41	Wendell Summersby	31-45 years
3	53	Dericka Pockran	46-60 years
4	111	Karolina Dokter	31-45 years
5	171	Robert Bowerman	31-45 years
6	183	Duncan Robertshawe	61+ years
7	195	Cosi Rimmington	61+ years
8	215	Ramesh Trentholme	46-60 years

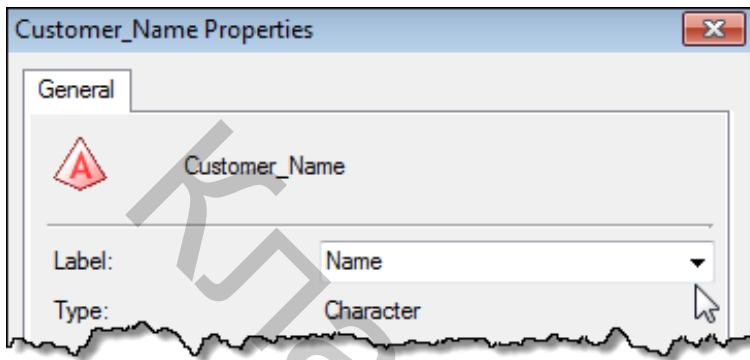
Customer Country=CA			
Row number	Customer ID	Customer Name	Customer Age Group
9	11171	Bill Cuddy	15-30 years
10	17023	Susan Krasowski	46-60 years

6. Чтобы изменить ярлыки столбцов и удалить номера строк в задаче, на вкладке «Results» выберите команду **Modify Task**.
7. Роль **Идентифицирующий ярлык (Identifying label)** может использоваться для замены столбца с номером строки столбцом из данных. Перетащите переменную **Customer\_ID** из роли **Перечислить переменные (List variables)** в роль Идентифицирующий ярлык (**Identifying label**).

 Чтобы удалить нумерацию строк, не используя роль Идентифицирующий ярлык (**Identifying label**), в области выбора настроек выберите Параметры (**Options**) и снимите флажок **Печатать номер строки (Print the row number)**.



8. Чтобы указать ярлык, который будет использоваться для столбца **Customer\_Name** в этой задаче, щелкните правой кнопкой мыши имя столбца в области Роли задачи (Task roles). В меню выберите Свойства (**Properties**).
9. В появившемся диалоговом окне в поле **Ярлык (Label)** введите **Name** и нажмите кнопку **OK**.



10. Повторите шаг 8, чтобы изменить ярлык **Customer\_Age\_Group** на **Age Group**.

В диалоговом окне задачи по-прежнему отображаются имена переменных по умолчанию. Чтобы просмотреть ярлыки, щелкните правой кнопкой мыши переменную и выберите **Show Labels (показать ярлыки)**.

11. По умолчанию значок задачи в проекте подписан как **List Data**. Задайте для задачи другое название, чтобы ее можно было легко идентифицировать в проекте. Выберите Свойства ⇒ Изменить (**Properties** ⇒ **Edit**) и в поле Ярлык (**Label**) введите **Customer List by Country**. Нажмите кнопку **OK**.
12. Нажмите кнопку Пуск (**Run**), чтобы сформировать отчет. При запросе на замену результатов нажмите кнопку Да (**Yes**).

Фрагмент результатов

<b>Orion Star Customer List</b>		
Customer Country=AU		
<b>Customer ID</b>	<b>Name</b>	<b>Age Group</b>
29	Candy Kinsey	61+ years
41	Wendell Summersby	31-45 years
53	Dericka Pockran	46-60 years
111	Karolina Dokter	31-45 years
171	Robert Bowerman	31-45 years
183	Duncan Robertshawe	61+ years
195	Cosi Rimmington	61+ years
215	Ramesh Trentholme	46-60 years

Customer Country=CA		
<b>Customer ID</b>	<b>Name</b>	<b>Age Group</b>
11171	Bill Cuddy	15-30 years
17023	Susan Krasowski	46-60 years
26		

13. В меню нажмите кнопку , чтобы сохранить проект **Chapter3**.

Клабуков Максим  
Август 2020

## Фильтрация данных в задаче

**Цели**

- Научиться фильтровать данные прямо в задаче.
- Научиться создавать фильтры, используя разные операторы.

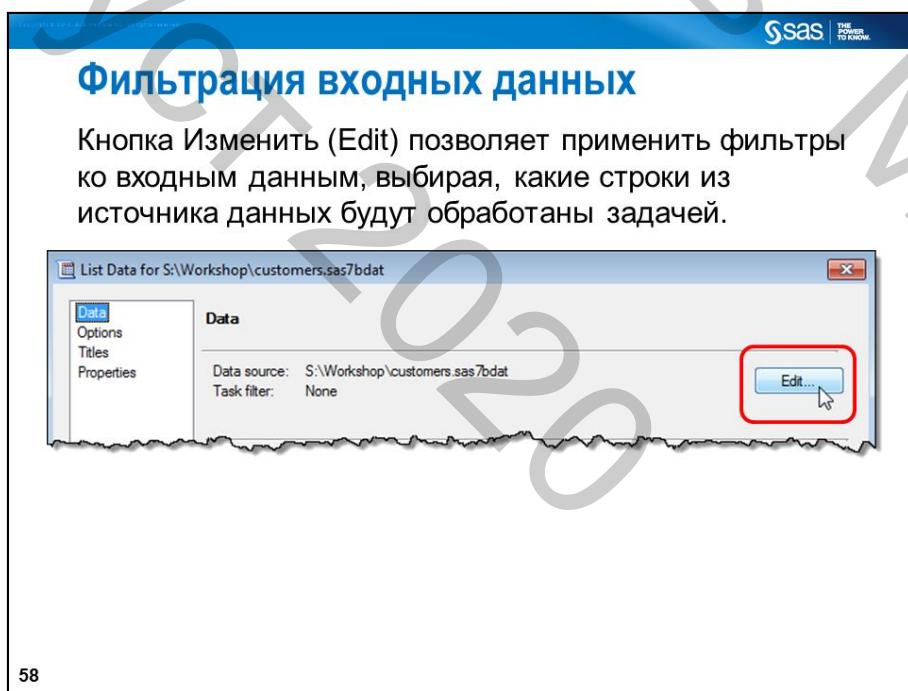
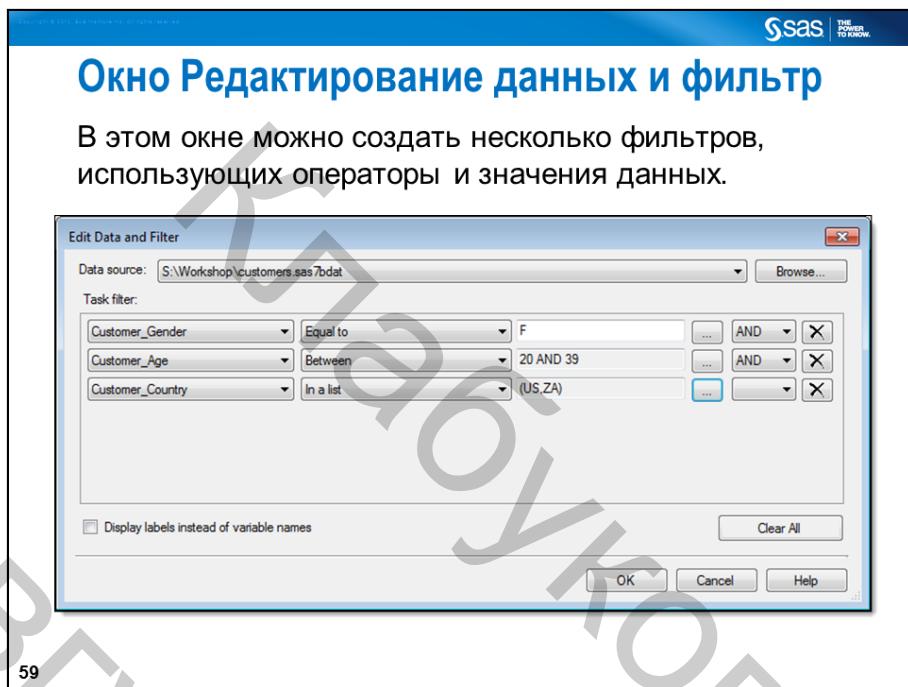
56

**Бизнес-сценарий**

Отделу маркетинга компании Orion Star нужно провести рекламную кампанию среди активных членов клуба Orion, рожденных в 1970 году или позднее.

Orion Star Customer List		
High Activity Customers Born 1970 or Later		
Customer ID	Customer Name	Customer Type Name
10	Karen Ballinger	Orion Club members high activity
11	Elke Wallstab	Orion Club members high activity
39	Alphone Greenwald	Orion Club Gold members high activity
49	Annmarie Leveille	Orion Club Gold members high activity
171	Robert Bowerman	Orion Club members high activity
908	Avni Umran	Orion Club Gold members high activity
14703	Eyal Bloch	Orion Club members high activity
19873	Avinoam Tuvia	Orion Club Gold members high activity
46966	Lauren Krasowski	Orion Club members high activity

57



Для создания фильтров можно использовать только определённые операторы, доступные в окне Редактирование данных и фильтр (Edit Data and Filter). Более сложные фильтры, использующие функции или составные операторы сравнения, можно создать в задаче «Фильтрация и сортировка» или с помощью конструктора запросов.

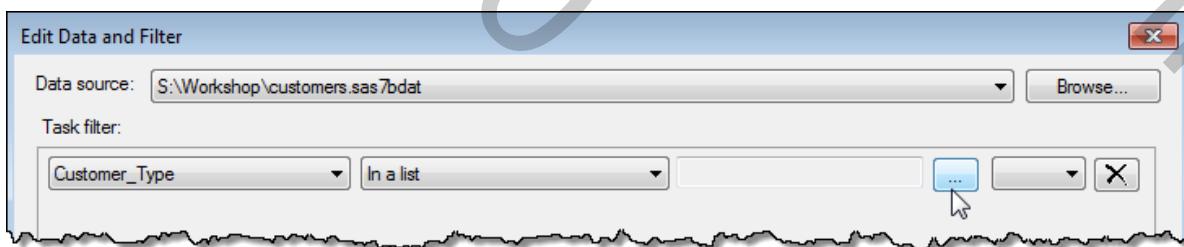
Доступны следующие операторы:

- Equal to/Not equal to (равно/не равно)
- In a list/Not in a list (в списке/отсутствует в списке)
- Less than/Less than or equal to (меньше/меньше или равно)
- Greater than/Greater than or equal to (больше/больше или равно)
- Between/Not between (в указанном интервале, включая границы интервала/ вне указанного интервала)
- Contains/Does not contain (содержит/не содержит)
- Is missing/Is not missing (пропущенное значение/ не пропущенное значение)

## Фильтрация данных в задаче

Отделу маркетинга в компании Orion Star требуется провести специальную рекламную акцию для членов клуба Orion с высоким уровнем активности, которые родились в 1970 году или позднее. Измените задачу «Списочный отчет» (List Data), добавив в нее соответствующий фильтр.

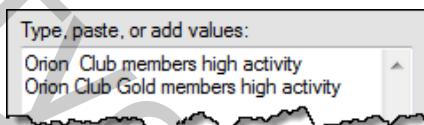
1. В проекте **Chapter3** дважды щелкните набор данных **customers** в дереве проекта или в окне схемы процесса «Process Flow».
2. В сетке данных выберите Описательная статистика  $\Rightarrow$  Списочный отчет (Describe  $\Rightarrow$  List Data).
3. Чтобы включить только клиентов с высоким уровнем активности, родившихся в 1970 году или позднее, сначала нажмите кнопку Изменить (Edit).
4. В первом раскрывающемся списке выберите **Customer\_Type**, а во втором списке — В списке (In a list).



5. Рядом с третьим полем нажмите кнопку ..., чтобы выбрать конкретные значения. Нажмите кнопку **Добавить значения (Add Values)**, чтобы просмотреть различные значения в столбце **Customer\_Type**. Чтобы увидеть все значения, возможно, потребуется растянуть окно. Удерживая нажатой клавишу CTRL, выделите второе и четвертое значения, указывающие на высокую активность. Нажмите кнопку **OK**.

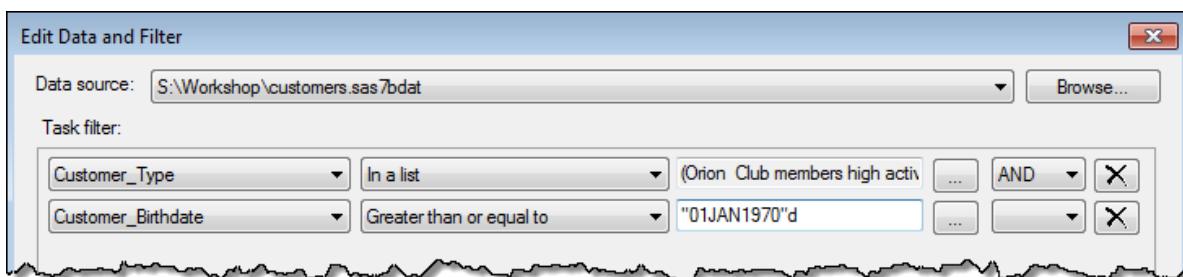
Value	Formatted Value
Internet/Catalog Customers	Internet/Catalog Customers
Orion Club members high activity	Orion Club members high activity
Orion Club members medium activity	Orion Club members medium activity
Orion Club Gold members high activity	Orion Club Gold members high activ...
Orion Club Gold members low activity	Orion Club Gold members low activity
Orion Club Gold members medium ac...	Orion Club Gold members medium a...
Orion Club members low activity	Orion Club members low activity

6. Убедитесь, что выбраны следующие значения, и нажмите кнопку **OK**:

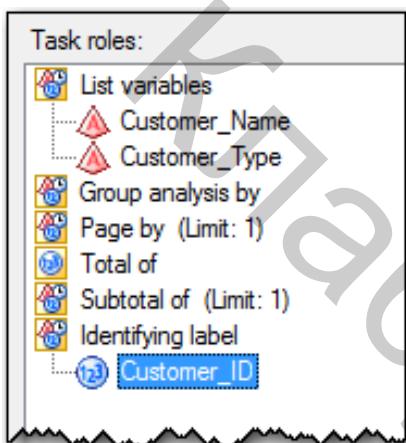


7. Чтобы добавить другой фильтр, в самом правом раскрывающемся списке (рядом со значениями данных) выберите **AND**. Чтобы строка выводилась в отчет, требуется выполнение обоих условий фильтра, объединённых оператором AND.
8. В первом списке выберите **Customer\_BirthDate**, а во втором — Больше чем или равно (**Greater than or equal to**). Нажмите кнопку ..., чтобы просмотреть значения данных. Обратите внимание, что значения являются датами в формате SAS, то есть число дней между 1 января 1960 г и указанной датой.
9. Чтобы выполнить фильтрацию по конкретной дате, можно использовать константу даты SAS. Нажмите кнопку **Cancel (отмена)**, чтобы вернуться в третье поле, и введите "**01JAN1970"d**. Перед фильтрацией данных система SAS преобразует дату в кавычках в формат даты SAS. Нажмите кнопку **OK**.

В выражениях необходимо указывать константу даты SAS исключительно в формате "ddMONyyyy"d (например, "18APR1987"d).



10. В области настроек Данные (Data) добавьте переменные **Customer\_Name** и **Customer\_Type** в роль Перечислить переменные (List variables), а переменную **Customer\_ID** — в роль Идентифицирующий ярлык (Identifying label).

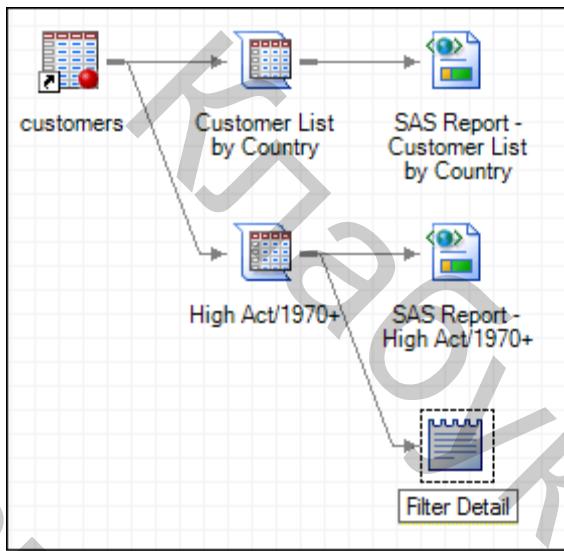


11. Измените заголовок отчёта так, чтобы в нём указывалось условие фильтрации. В области выбора выберите Заголовки (Titles). Удалите заголовок по умолчанию и в первой строке введите **Orion Star Customers List**, а во второй — **High Activity Customers Born 1970 or Later**.
12. Переименуйте задачу. Последовательно выберите Свойства ⇒ Изменить (Properties ⇒ Edit). В поле Ярлык (Label) введите **High Act/1970+**. Нажмите кнопку **OK**.
13. Нажмите кнопку Пуск (Run) и просмотрите окончательный отчет.

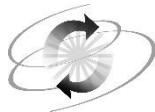
Orion Star Customer List		
High Activity Customers Born 1970 or Later		
Customer ID	Customer Name	Customer Type Name
10	Karen Ballinger	Orion Club members high activity
11	Elke Wallstab	Orion Club members high activity
39	Alphone Greenwald	Orion Club Gold members high activity
49	Annmarie Leveille	Orion Club Gold members high activity
171	Robert Bowerman	Orion Club members high activity
908	Avni Umran	Orion Club Gold members high activity
14703	Eyal Bloch	Orion Club members high activity
19873	Avinoam Tuvia	Orion Club Gold members high activity
46966	Lauren Krasowski	Orion Club members high activity

14. Чтобы добавить в проект комментарий о применении фильтра в этой задаче, можно также использовать примечание (Note). Выделите задачу **High Act/1970+**, щелкните и выберите **Примечание (Note)**.
15. Примечание может содержать любой текст, поясняющий или документирующий элементы в проекте. Введите в появившемся редакторе текст: «**В этот отчет включены клиенты из групп с высоким уровнем активности, родившиеся в 1970 г. или позднее**».

16. Чтобы переименовать значок примечания в проекте, нажмите кнопку Свойства (**Properties**) и в поле Ярлык (**Label**) введите **Filter Detail**. Нажмите кнопку **OK**.



17. Сохраните проект **Chapter3**.



## Упражнения

2. Использование задачи «Списочный отчет» (List Data) для вывода списка сотрудников из Сан-Диего

- a. При необходимости добавьте в проект Chapter3 таблицу **employee\_addresses**.

Если проект **Chapter3** отсутствует, создайте его.

- b. С помощью задачи «Списочный отчет» (List Data) создайте отчет и выведите столбцы **Employee\_Name**, **Street\_Number**, **Street\_Name**, **City** и **Postal\_Code**. Для каждой строки вместо её номера укажите значение столбца **Employee\_ID**.
- c. Измените свойства столбцов так, чтобы в отчете отображались следующие подписи столбцов: **Employee ID**, **Name**, **Street Number**, **Street Name**, **City** и **Postal Code**.
- d. Создайте отчет.
- e. Измените задачу «Списочный отчет» (List Data), включив в отчет только сотрудников из Сан-Диего (San Diego).
- f. Удалите из отчета столбец **City**.
- g. Измените заголовок отчёта на **Employee List for San Diego** и удалите нижний колонтитул.
- h. Переименуйте задачу в **San Diego List**.
- i. Выполните задачу и замените результаты.
- j. Сохраните проект как **Chapter3**.

Фрагмент результатов

<b>Employee List for San Diego</b>				
<b>Employee ID</b>	<b>Name</b>	<b>Street Number</b>	<b>Street Name</b>	<b>Postal Code</b>
120656	Amos, Salley	3524	Calico Ct	92116
120759	Apr, Nishan	105	Brack Penny Rd	92071
121017	Arizmendi, Gilbert	379	Englehardt Dr	91950
121062	Armant, Debra	10398	Crown Forest Ct	92025
121049	Bataineh, Perrior	892	Birchland Dr	92126
120998	Benedicto, Tondelayo	2902	Gable Ridge Ln	92118
120800	Benyami, Fred	1498	Falconwood Dr	92069
121025	Blankley, L	41	Dubose St	92110

Клабуков Максим  
Август 2020

## Создание графика

**Цели**

- Узнать типы графиков, которые поддерживаются в SAS Enterprise Guide.
- Понять разницу между форматами результатов, которые поддерживаются в SAS Enterprise Guide.
- Создать столбчатую диаграмму.

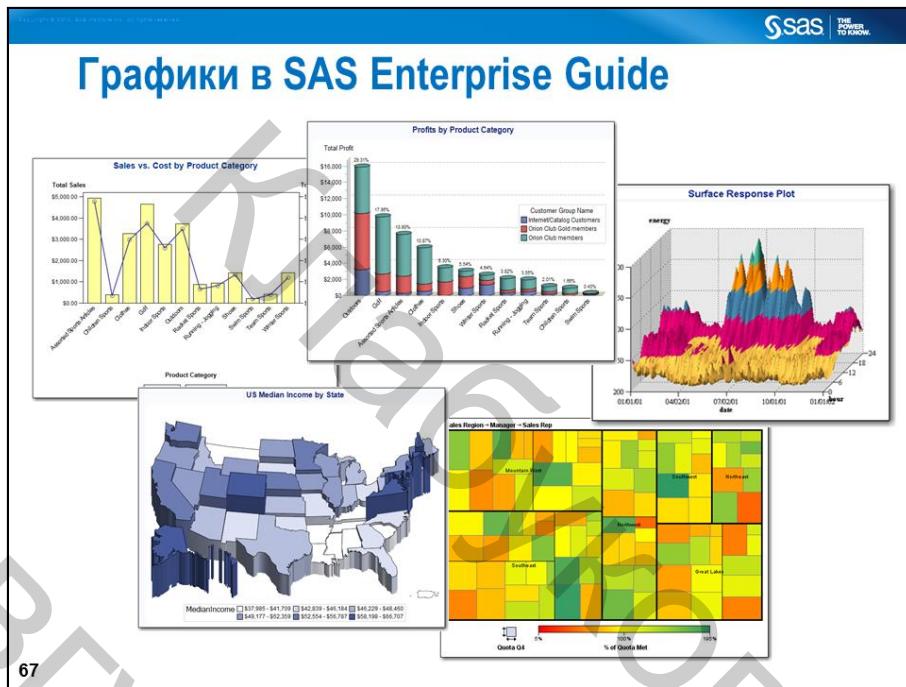
65

**Бизнес-сценарий**

Руководству нужно сравнение среднего возраста клиентов в разных странах в графическом виде.

Country	Average Customer Age
AU	54.0
DE	49.0
US	42.3
CA	41.8
IL	41.8
TR	41.0
ZA	38.8

66



67

Для получения дополнительных сведений о форматах вывода графика, выберите в меню Справка ⇒ Справка приложения SAS Enterprise Guide (Help ⇒ SAS Enterprise Guide Help). Выполните поиск по запросу **setting results and graph options**.

### 3.08 Вопрос с несколькими вариантами ответа

Выберите Задачи ⇒ График (Tasks ⇒ Graph) и просмотрите доступные типы графиков. Какие из них были бы полезны для изображения ваших данных?

- столбчатая диаграмма (bar chart)
- секторная диаграмма (pie chart)
- линейный график (line plot)
- диаграмма рассеяния (scatter plot)
- график поверхности (surface plot)
- карта (map)
- другие

68

**Мастер настроек столбчатой диаграммы**

Этот мастер настроек (Bar Chart Wizard) позволяет выполнять следующие настройки:

1. выбирать и фильтровать данные
2. задавать роли для переменных
3. настраивать вид графика
4. задавать сноски и заголовки отчёта

69

**Столбчатая диаграмма: Расширенный режим**

Для более детальных настроек графика вы можете открыть Расширенный режим с дополнительными настройками (Advanced View). Этот режим содержит настройки, которых нет в режиме Мастера настроек. Вы не сможете вернуться в режим Мастера настроек после того, как откроете задачу в Расширенном режиме.

70

**Задача столбчатая диаграмма: Расширенный режим**

Расширенный режим задачи Столбчатая диаграмма содержит дополнительные настройки, такие как форма столбцов, форматирование значений данных, настройки опорных линий.

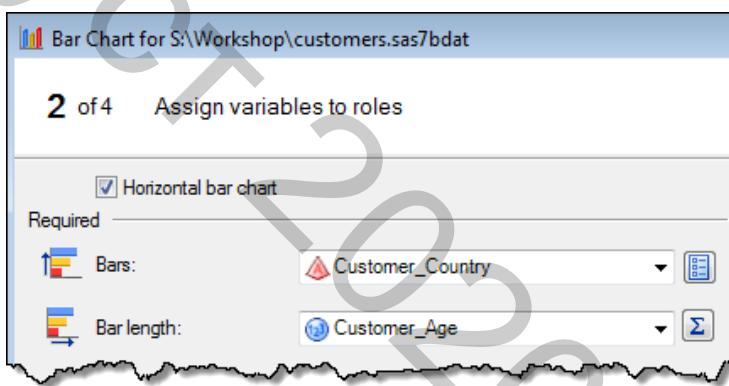
The screenshot shows the SAS Data Editor interface for a 'Bar Chart' task. The left sidebar lists categories like Bar Chart, Data, Appearance, Bars, Layout, Axes, Chart Area, Advanced, Titles, and Properties. The main pane is titled 'Data' and shows a data source path 'S:\Workshop\orion\_profit.sas7bdat'. A 'Task filter' section is set to 'None'. Below these are two panes: 'Columns to assign:' containing a list of variables (Order\_ID, Product\_ID, Product\_Line, Product\_Category, Product\_Name, Supplier\_ID, Supplier\_Name, Customer\_ID, Customer\_Country, Customer\_Gender, Customer\_Name, Customer\_Age\_Group, Customer\_Age, Customer\_Group, Order\_Date) with icons indicating their role; and 'Task roles:' which includes 'Column to chart (Limit: 1)' (selected), 'Sum of (Limit: 1)' (selected), 'Group charts by' (selected), and 'Product\_Line' (selected). The number '71' is visible at the bottom left of the window.

## Создание столбчатой диаграммы

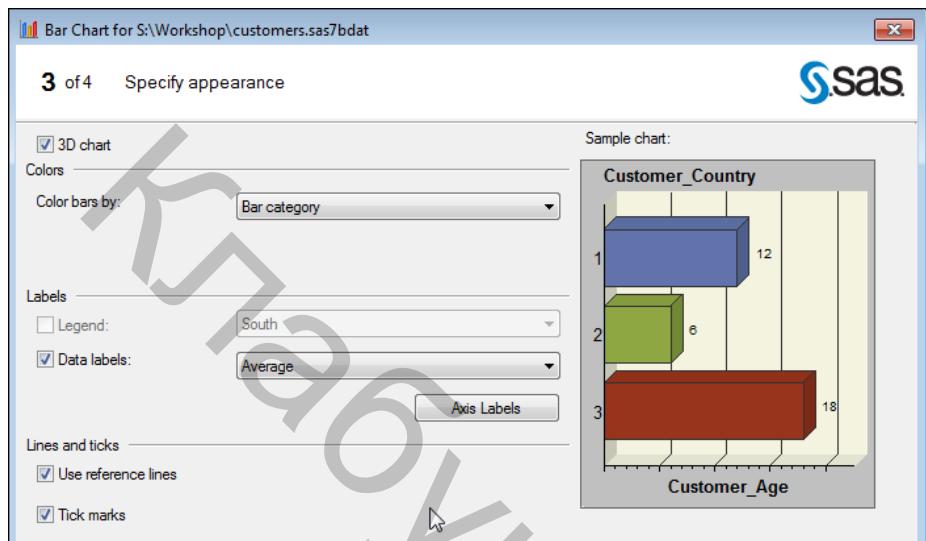


Чтобы лучше представить демографический профиль клиентов Orion Star, руководству компании требуются данные об их среднем возрасте в разбивке по странам. Создайте графический отчёт с помощью Мастера создания столбчатых диаграмм и затем дополнительно настройте задачу.

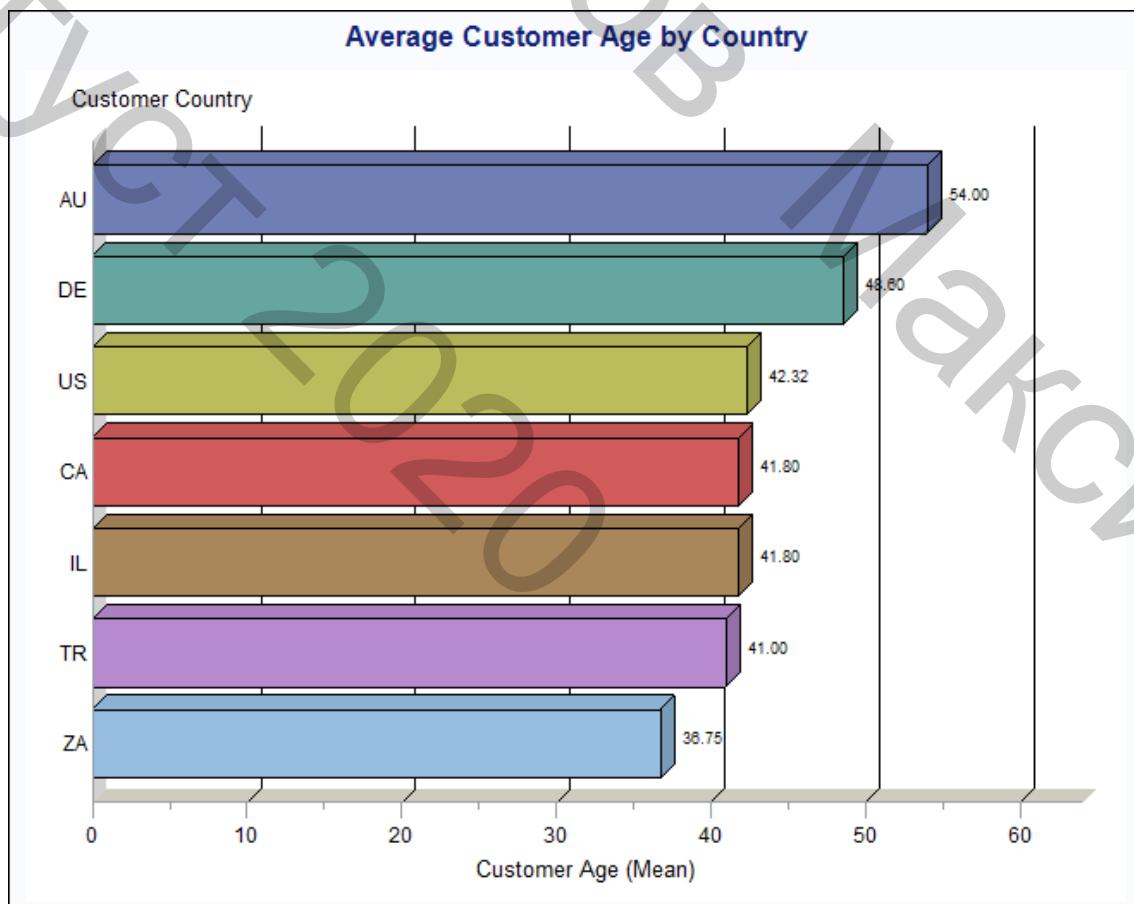
1. В проекте **Chapter3** выберите набор данных **customers** в дереве проекта или в окне схемы процесса. Выберите **Tasks (задачи) ⇒ Graph (график) ⇒ Bar Chart Wizard (мастер столбчатых диаграмм)**.
2. На шаге 1 убедитесь, что работа ведется с набором данных **customers**. Нажмите кнопку **Вперед (Next)**.
3. На шаге 2 установите флажок **горизонтальная столбчатая диаграмма (Horizontal bar chart)**. Чтобы создать отдельный столбик для каждой страны, в раскрывающемся списке «Bars» (столбики) выберите **Customer\_Country**.
4. Чтобы упорядочить столбики по убыванию значения высоты или длины, нажмите кнопку . В появившемся окне выберите **убывающая высота столбика (Descending bar height)** ⇒ **OK**.
5. Чтобы задать длину столбиков на основании значения среднего возраста, в раскрывающемся списке Высота столбика (Bar length) выберите **Customer\_Age**.



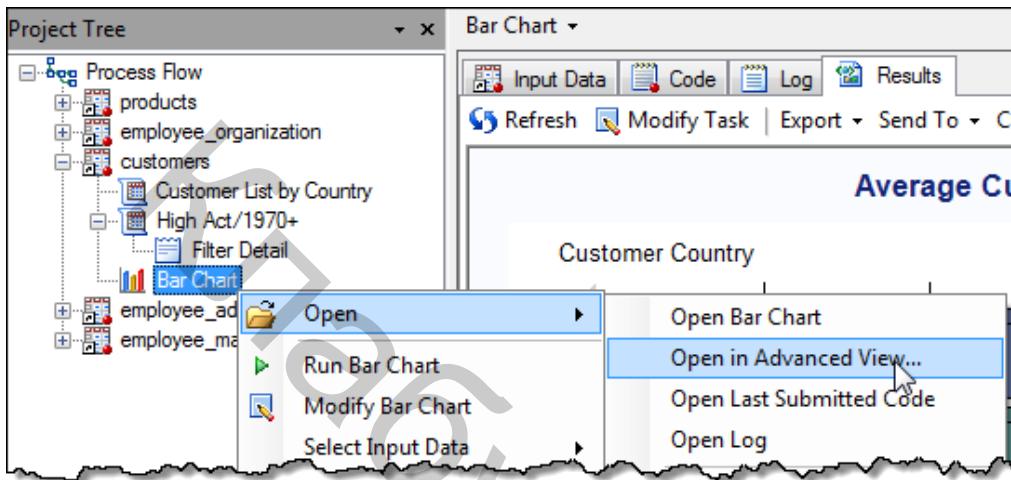
6. Нажмите кнопку , выберите **Average (среднее)** и нажмите кнопку **OK**. Нажмите кнопку **Вперед (Next)**.
7. На шаге 3 внесите следующие изменения.
  - a. Установите флажок **трехмерная диаграмма (3D chart)**.
  - b. В раскрывающемся списке **Раскрасить столбики по (Color bars by)** выберите **Категория столбика (Bar category)**.
  - c. Установите флажок **Ярлыки данных (Data labels)** и в раскрывающемся списке выберите **Среднее (Average)**.
  - d. Установите флажок **Use reference lines (использовать опорные линии)**.
  - e. Нажмите кнопку **Вперед (Next)**.



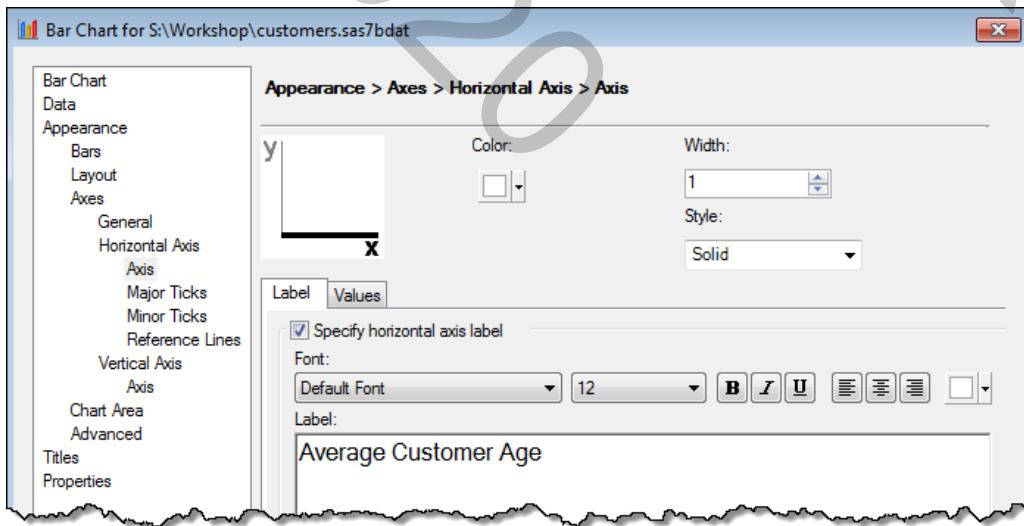
8. На шаге 4 измените заголовок отчёта на **Average Customer Age by Country** и нажмите кнопку **Finish (готово)**.



9. Чтобы внести дополнительные изменения, которые невозможно выполнить в Мастере создания столбчатых диаграмм, откройте задачу в расширенном представлении. Щелкните правой кнопкой мыши значок **Bar Chart** в дереве проекта или в схеме процесса. Выберите **Open (открыть) ⇒ Open in Advanced View (открыть в режиме расширенных настроек)**.

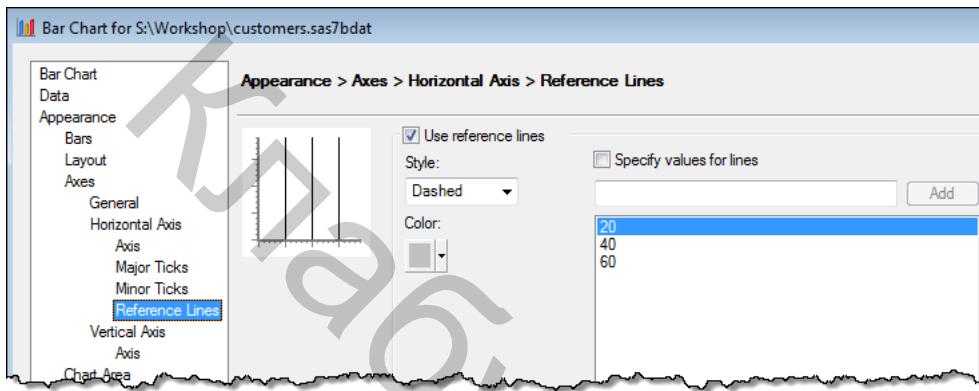


10. В расширенном представлении можно изменить отображение значений, применив формат к переменным. Перейдите в настройки Данных (Data). Чтобы отформатировать переменную **Customer\_Age** так, чтобы подписи данных округлялись до одного десятичного разряда, щелкните переменную **Customer\_Age** правой кнопкой мыши и выберите Свойства (Properties).
11. Нажмите кнопку Изменить (Change), чтобы выбрать формат. В окне Форматы (Formats) в области Категории (Categories) выберите Числовую (Numeric) категорию форматов. В области Форматы (Formats) выберите **w.d.** В качестве значения общего числа знаков выберите **4**, а для числа десятичных разрядов — **1**. Нажмите кнопки **OK** ⇒ **OK**, чтобы вернуться в диалоговое окно задачи.
12. В области выбора настроек выберите Разметка (Layout). Измените форму столбика, выберите Цилиндр (Cylinder) в поле Форма (Shape).
13. В области выбора настроек выберите Горизонтальная ось (Horizontal Axis). В поле Ярлык (Label) введите **Average Customer Age** и задайте размер шрифта **12**.

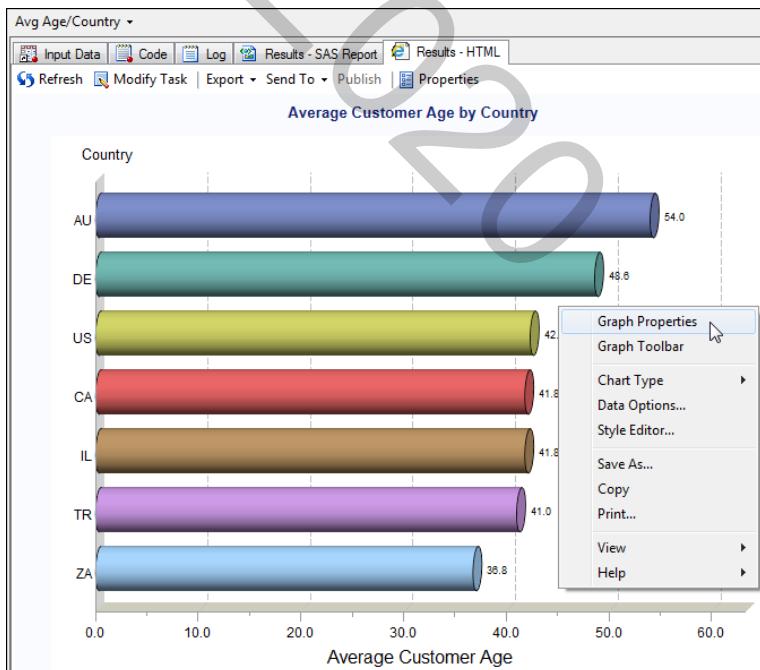


14. В области выбора настроек выберите Reference Lines (опорные линии). Установите флажок **Specify values for lines** (указать значения для делений), введите **20**, а затем нажмите кнопку

**Add (добавить).** Повторите то же самое для ввода значений **40** и **60**. В качестве стиля выберите **Dashed (пунктирный)** и задайте цвет светло-серый.



15. В области выбора настроек выберите **вертикальная ось (Vertical Axis)**. В поле Ярлык (Label) введите **Country**.
16. В области выбора настроек выберите Свойства (Properties) и щелкните Изменить (Edit). В области Общие (General) измените Ярлык (Label) на **Avg Age/Country**.
17. Формат графика по умолчанию — ActiveX. Чтобы график был доступен для работы после выполнения задачи, он должен иметь формат HTML. В области Результаты (Results) установите флажок Настроить форматы, стили и функции результатов (Customize result formats, styles, and behavior) и выберите HTML. Нажмите кнопку OK.
18. Нажмите кнопку Пуск (Run), чтобы создать график. Перейдите на вкладку Results – HTML . Наведите указатель мыши на график и нажмите правую кнопку мыши. В контекстном меню выберите свойства диаграммы (Graph Properties).



19. На вкладке График (Graph) выберите стиль **Curve**. Будет применена новая цветовая схема.

20. Откройте вкладку **столбчатая диаграмма (Bar)** и выберите **шрифт (Font)**. Увеличьте размер шрифта для подписи данных до **10**. Последовательно нажмите кнопки **OK** ⇒ **OK**, чтобы вернуться к графику и применить изменения.
21. Можно также изменить тип диаграммы. Для этого щелкните график правой кнопкой мыши и выберите **Тип диаграммы (Chart Type)** ⇒ **вертикальная столбчатая (Vertical Bar)**.
22. Повернуть, выбрать данные для отображения или изменить размер графика можно с помощью панели инструментов. Щелкните график правой кнопкой мыши и выберите **панель графика (Graph Toolbar)**. Щелкните  и перетащите указатель мыши через три верхних столбика. График будет перестроен, и в него будет включена только выделенная часть. Щелкните , чтобы сбросить график.
23. Можно сохранить график как JPG-файл (все изменения, которые были интерактивно внесены в график, будут сохранены в виде статического изображения) или экспортовать его как HTML-файл (интерактивность сохраняется, но внесенные в интерактивном режиме изменения теряются). Чтобы экспортовать результат в формате HTML, выберите **Export (экспорт)** ⇒ **Export HTML - Avg Age/Country**. Перейдите в нужное расположение и нажмите кнопку **Save (сохранить)**.



Графики можно копировать и вставлять в документы Microsoft Word, Excel и PowerPoint. Щелкните правой кнопкой мыши график в приложении SAS Enterprise Guide и выберите команду **Copy (копировать)**. Затем щелкните правой кнопкой мыши в документе Word, Excel или PowerPoint и выберите команду **Paste (вставить)**.

Чтобы сохранить интерактивные возможности ActiveX в этих приложениях, примите во внимание следующие факторы.

- Microsoft Excel. График автоматически сохраняет функциональные возможности ActiveX.
- Microsoft Word. После вставки графика в документ необходимо выйти из режима конструктора. Для этого в Office 2007 выберите **Developer (разработчик)** ⇒ **Design Mode (режим конструктора)**.
- Microsoft PowerPoint. Чтобы получить возможность работать с графиком ActiveX в режиме просмотра в PowerPoint, сначала необходимо изменить свойство графика и лишь затем скопировать и вставить график. Щелкните правой кнопкой мыши график в приложении SAS Enterprise Guide и выберите **Graph Properties (свойства графика)**. На вкладке «Graph» нажмите кнопку **Advanced (дополнительно)**. В окне «Advanced Settings» (дополнительные параметры) установите флажок **Display interactive graphs (отображать интерактивные графики)**. После изменения свойств графика скопируйте и вставьте его в презентацию PowerPoint.

24. Сохраните проект **Chapter3**.

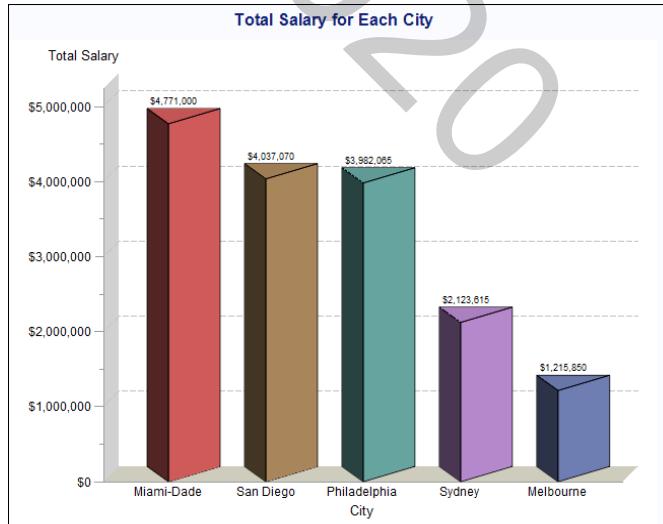


## Упражнения

### 3. Создание трехмерной вертикальной столбчатой диаграммы

- Добавьте в проект **Chapter3** набор данных SAS **employee\_detail**.
 

Если проект **Chapter3** отсутствует, создайте его.
- В сетке данных откройте таблицу **employee\_detail** и запустите мастер создания столбчатых диаграмм. На шаге 2 мастера настройте вертикальную столбчатую диаграмму, которая отображает сумму столбца **Salary** для каждого города (столбец **City**). Расположите столбики на диаграмме по убыванию высоты.
- На шаге 3 мастера создания столбчатых диаграмм измените следующие атрибуты.
  - Создайте трехмерную диаграмму.
  - Окрасьте столбики в разные цвета.
  - Выполните подписи в верхней части каждого столбика с суммой столбца **Salary**.
  - Подпишите вертикальную ось — **Total Salary**.
  - Включите отображение опорных линий.
- На шаге 4 мастера задайте заголовок для отчёта **Total Salary for Each City**.
- Создайте график.
- Щелкните правой кнопкой мыши задачу **Bar Chart** в дереве проекта или в окне схемы процесса. Выберите Открыть ⇔ Открыть в режиме расширенных настроек (**Open ⇔ Open in Advanced View**). Внесите следующие изменения.
  - Измените формат, применяемый к столбцу **Salary**, чтобы значения округлялись до ближайшего целого числа.
  - В качестве фигуры столбика выберите призму.
  - Для опорных делений выберите светло-серый цвет и отображение пунктиром.
- Измените свойства созданной вами задачи, чтобы отчет создавался не только в используемом по умолчанию формате «Отчет SAS», но и в формате HTML.
- Выполните задачу и замените результаты. Сохраните проект как **Chapter3**.



## Фильтрация и сортировка данных

**Цели**

- Исследовать возможности задач Фильтр и сортировка (Filter and Sort), Конструктор запросов (Query Builder) и Исследование данных (Data Explorer).
- Сравнить функциональность этих задач.
- Научиться применять фильтр в запросе.
- Узнать, как исключить столбцы из запроса.
- Научиться задавать сортировку строк.

3

**Изменение данных**

Часто требуется изменить ваши данные перед тем, как использовать их в задачах. Например, если нужно исследовать только часть строк или столбцов, предварительно вычислить новый столбец, объединить данные из двух или более таблиц.

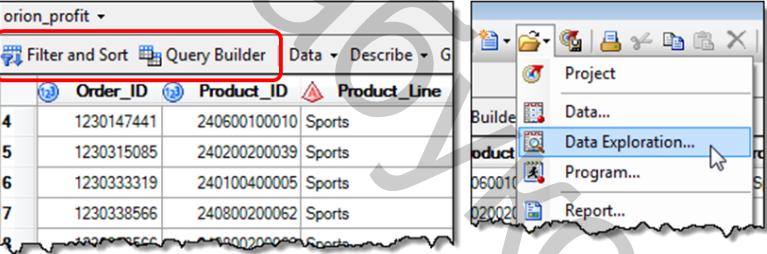


4

**Построение запросов к данным**

Для создания новых источников данных в Enterprise Guide есть три задачи, способные изменять данные и строить запросы.

- Конструктор запросов (Query Builder)
- Фильтр и сортировка (Filter and Sort)
- Исследование данных (Data exploration)



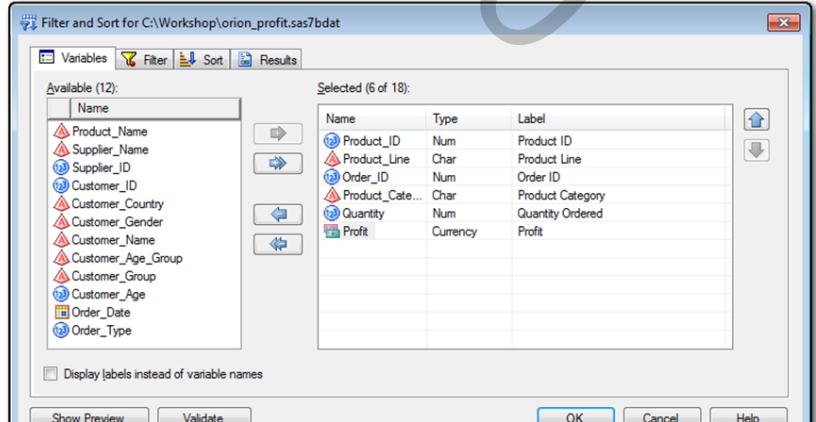
5

Запрос — это набор инструкций, позволяющий выполнять действия с конкретным набором данных. Для создания этих инструкций можно использовать задачу «Фильтр и сортировка» (Filter and Sort) и конструктор запросов (Query Builder). Код, создаваемый этими задачами, основан на Языке структурированных запросов (SQL).

- ✍ В Enterprise Guide для создания запросов также можно использовать Исследование данных (Data Explorer).

**Задача Фильтр и сортировка**

Эта задача позволяет создавать новые таблицы, выбирая строки, столбцы, задавая сортировку.



8

**4.02 Тест**

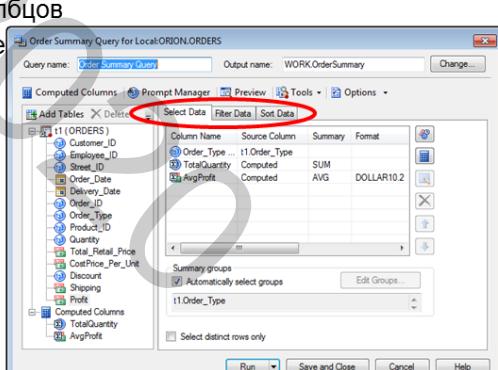
Закройте задачу Фильтр и сортировка, вернитесь в сетку данных. Выберите Конструктор запросов (**Query Builder**). Какие настройки запроса, доступные в этой задаче, отсутствуют в задаче Фильтр и сортировка (Filter and Sort)?

9

## Конструктор запросов (Query Builder)

Конструктор запросов также содержит вкладки для выбора столбцов, фильтрации и сортировки строк, а также имеет дополнительные функции:

- изменение свойств столбцов
- группировка и сведение данных
- применение форматов
- выбор отличающихся строк
- объединение таблиц



The screenshot shows the 'Order Summary Query for Local:ORION.ORDERS' dialog box. The 'Select Data' tab is highlighted with a red circle. The main area displays a table with columns: Order\_Type, Total\_Quantity, and AvgProfit. The 'Source Column' column shows '1| Order\_Type', 'Computed', and 'Computed' respectively. The 'Summary' column shows 'SUM', 'AVG', and 'DOLLAR10.2'. The 'Format' column shows icons for date, time, and currency. On the left, there's a tree view of tables: ORION.ORDERS, ORION.EMPLOYEES, and ORION.CUSTOMERS. At the bottom, there are buttons for Run, Save and Close, Cancel, and Help.

11

С помощью конструктора запросов можно также создавать представления и отчеты SAS. Для получения дополнительных сведений, выберите в меню Справка ⇨ Справка приложения SAS Enterprise guide (**Help ⇨ SAS Enterprise Guide Help**). Выполните поиск по запросу **query output**. Выберите второй пункт в списке результатов поиска, **Setting Options for the Current Query**.

**Бизнес-сценарий**

Компании Orion Star нужно исследовать продажи через интернет с 2013 года. Чтобы подготовить данные для анализа, нужно создать новую таблицу из таблицы **orders**.

12

**Бизнес-сценарий**

Подготовка нового набора данных требует только применения фильтра и сортировки, поэтому можно использовать любую из задач.

Order_ID	Customer_ID
1	63
2	5
3	45
4	41

13

На вкладке «Filter» (фильтрация) создается оператор WHERE языка запросов SQL, который просматривает все строки данных и возвращает только те из них, которые соответствуют условиям фильтра.

При создании нескольких условий для фильтра можно задать логические отношения между ними — AND (И) или OR (ИЛИ).

Отношение AND возвращает строки данных, только если оба выражения, связанные им, верны. OR возвращает строки данных, если верно хотя бы одно из связанных выражений. Если при использовании AND или OR верные выражения отсутствуют, строки данных не возвращаются.

-  При вводе данных в третье поле помните, что символьные значения зависят от регистра. Например, CA ≠ Ca.

После нажатия кнопки  в третьем поле, в список возвращается максимум 250 уникальных значений из столбца, указанного в первом поле. Чтобы получить следующие 250 уникальных значений, выберите Ещё значения (More Values). Каждый раз при нажатии на Ещё значения (More Values) в список значений добавляется следующие 250 уникальных значений для переменной. Количество получаемых каждый раз значений определяется настройкой, которая находится в меню Tools (сервис) ⇒ Options (параметры).

В число операторов сравнения, доступных в диалоговом окне Edit Filter Condition (изменение условия фильтра), входят стандартные операторы сравнения, а также следующие:

IN ( <i>список значений</i> )	соответствует элементу в списке. Пример: <b>category in ("BREAD","MEAT")</b>
BETWEEN	роверяет вхождение в диапазон (включая границы). Пример. <b>income between 60000 and 80000</b>
IS MISSING	выполняет проверку на наличие пропущенных значений.
CONTAINS	роверяет символьное значение на содержание подстроки. Пример. <b>country contains "US"</b> соответствует значениям "USA", "RUSSIA"

С помощью конструктора расширенного фильтра можно создать выражение с использованием констант, столбцов, математических операторов и функций. Эти элементы можно добавить с помощью конструктора или же вручную ввести в поле в верхней части редактора.



## Выбор столбцов и фильтрация строк

Компании Orion Star требуется проанализировать интернет-продажи с начала 2013 года. Чтобы подготовить данные для различных аналитических задач, компании необходимо создать новую выборку из таблицы **orders**. Для создания необходимого запроса можно использовать задачу «Фильтр и сортировка» или конструктор запросов.

1. Создайте проект и добавьте в него набор данных SAS **orders**.
2. Чтобы создать набор данных, включающий заказы, размещенные в Интернете 1 января 2013 г. или позднее, Нажмите **Фильтр и сортировка (Filter and Sort)** в сетке данных.

Можно также щелкнуть правой кнопкой мыши **orders** в дереве проекта или окне схемы процесса «Process Flow» и выбрать **Фильтр и сортировка (Filter and Sort)**.

3. На вкладке **Переменные (Variables)** нажмите кнопку , чтобы добавить все столбцы в запрос. Для всех интернет-заказов в столбце **Employee\_ID** указано значение **99999999**, поэтому его можно удалить. Выделите столбец **Employee\_ID** в области **Выбрано (Selected)** и нажмите кнопку .

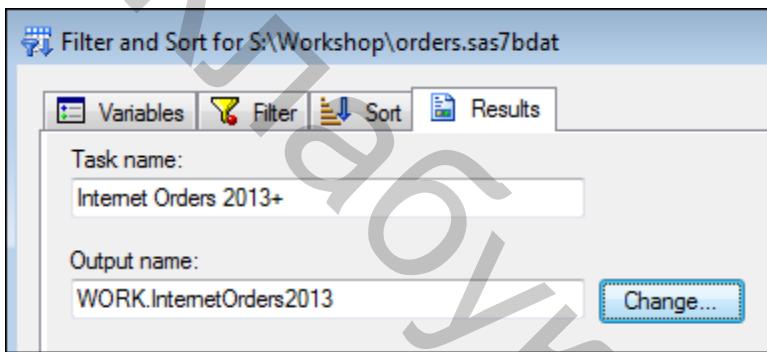
Name	Type	Label
Order_ID	Num	Order ID
Customer_ID	Num	Customer ID
Order_Date	Date	Order Date
Delivery_Date	Date	Delivery Date

4. Перейдите на вкладку **Фильтр (Filter)**. Определите два фильтра, показанные на экране ниже.

Order_Type	Equal to	3	AND
Order_Date	Greater than or equal to	"01JAN2013" d	X

- В столбце **Order\_Type** значение 1 означает розничные продажи, 2 — продажи по каталогу, а 3 — интернет-продажи.
- Даты SAS сохраняются в виде количества дней с 1 января 1960 г. Чтобы создать фильтр на основе конкретной даты, можно выбрать ее в списке, нажав кнопку , или ввести константу даты SAS. Для этого следует использовать формат "**DDMONYYYY**"**d**. Например, 1 января 2013 г. вводится как "**01JAN2013**"**d**.

5. Перейдите на вкладку **Сортировка (Sort)**. В качестве переменной сортировки выберите **Order\_Date**, а для порядка сортировки установите значение **По убыванию (Descending)**.
6. Перейдите на вкладку **Результаты (Results)**. В поле **Имя задачи (Task name)** введите **Internet Orders 2013+**. Нажмите кнопку **изменить (Change)** и введите **internetorders2013** в поле **Имя файла (File name)**. Нажмите кнопку **Сохранить (Save)**.



7. Нажмите кнопку **OK** и проверьте результаты. Полученный источник данных представляет собой таблицу SAS, которую можно экспортить или использовать в качестве входных данных для других задач.

Фрагмент результатов (8 строк из 55, 5 столбцов из 12)

	Order_ID	Customer_ID	Order_Date	Delivery_Date	Order_Type	
1	1244086685	14104	03DEC2014	06DEC2014	3	
2	1244086685	14104	03DEC2014	06DEC2014	3	
3	1243644877	70079	16OCT2014	19OCT2014	3	
4	1243485097	11	21SEP2014	02OCT2014	3	
5	1243462945	24	25SEP2014	28SEP2014	3	
6	1243279343	27	05SEP2014	10SEP2014	3	
7	1243165497	70201	24AUG2014	29AUG2014	3	
8	1243165497	70201	24AUG2014	29AUG2014	3	

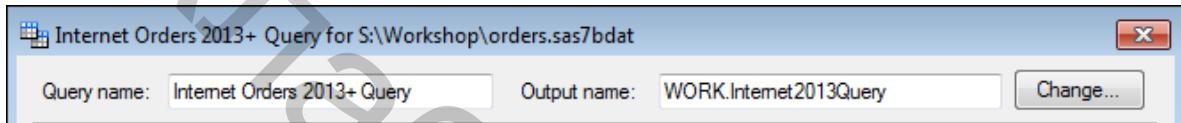
8. Для получения аналогичных результатов можно использовать конструктор запросов (Query builder). Сравним обе задачи, для этого опять откроите двойным щелчком просмотр таблицы **orders**. В сетке данных выберите **Конструктор запросов (Query Builder)**.

Можно также щелкнуть правой кнопкой мыши таблицу **orders** в проекте и выбрать Конструктор запросов (Query Builder).

9. Конструктор запросов позволяет подписать значок запроса и указать имя и расположение созданной таблицы SAS. В поле **Имя запроса (Query name)** введите **Internet Orders 2013+ Query**. Нажмите кнопку **Изменить (Change)** рядом с полем **Выходное имя (Output name)**. В поле **Имя файла (File name)** введите **internet2013query** и нажмите кнопку **Save**.



Таблица сохраняется в стандартной библиотеке SAS. Можно также выбрать любую другую библиотеку.



10. Чтобы выбрать все столбцы, выделите мышью элемент **t1 (orders)** в области выбора и перетащите его на вкладку Выбор данных (Select Data). Все столбцы будут добавлены в запрос. Удалите столбец **Employee\_ID**, выделив его на вкладке Выбор данных (Select Data) и нажав кнопку .

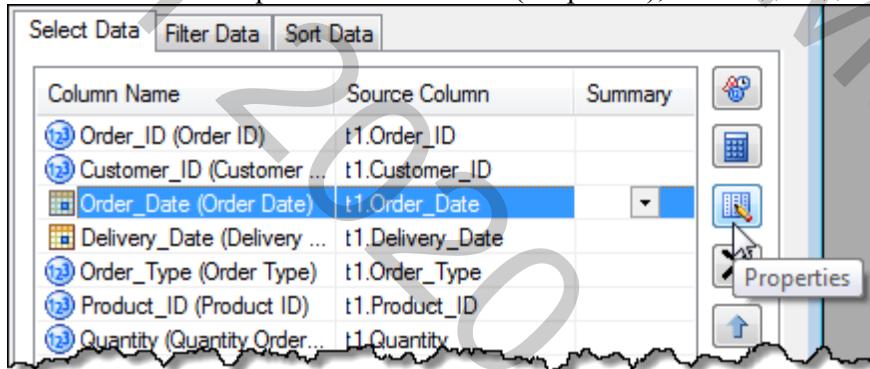


Столбцы можно добавлять по отдельности с помощью двойного щелчка или перетаскивания. Чтобы выбрать несколько столбцов, удерживайте нажатой клавишу CTRL. Для выбора диапазона столбцов используйте клавишу SHIFT.

11. Свойства столбца можно просмотреть или изменить на вкладке Выбор данных (Select Data). Выделите **Order\_Date** и нажмите кнопку  , чтобы открыть окно Свойства (Properties).



Чтобы открыть окно Свойства (Properties), можно дважды щелкнуть столбец.



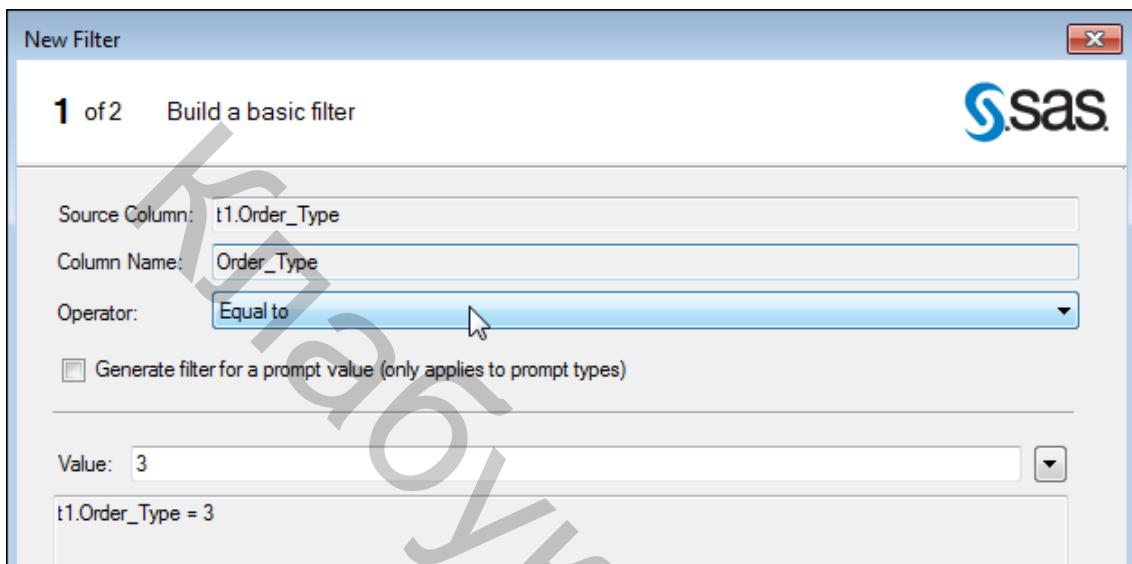
12. Чтобы применить формат к этому столбцу, нажмите кнопку **Изменить (Change)**. В окне Форматы (Formats) в области Категории (Categories) выберите Дата (Date), а в области Форматы (Formats) — MMDDYYw.d. Измените общее число знаков на 10. Нажмите **OK** ⇒ **OK**, чтобы вернуться в конструктор запросов.

13. Чтобы включить только интернет-заказы, размещенные не ранее 1 января 2013 г., откройте вкладку **Фильтрация данных (Filter Data)**. Перетащите **Order\_Type** на эту вкладку, чтобы запустить мастер создания фильтра.

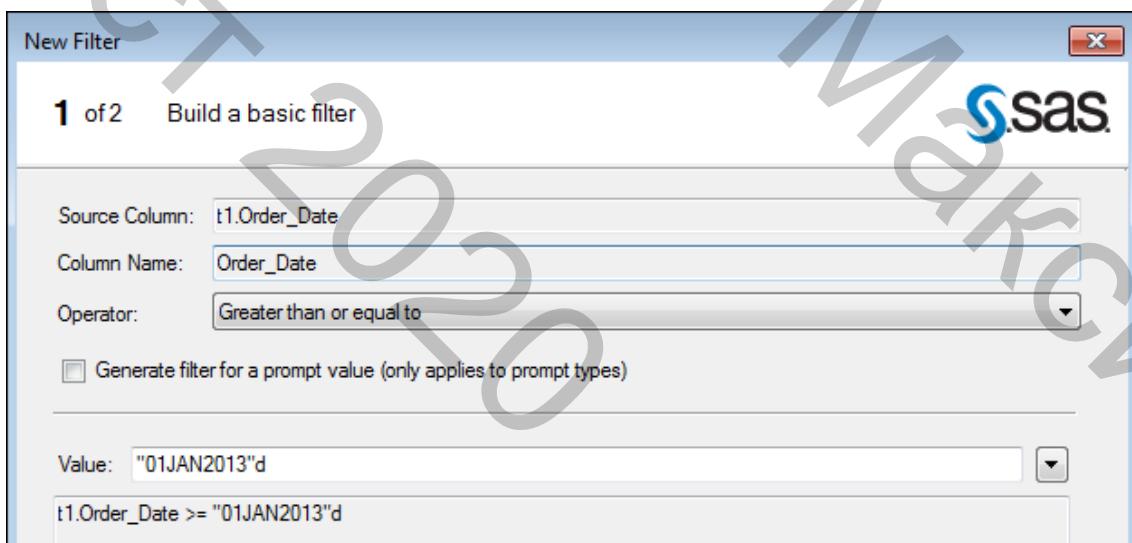


Можно также нажать кнопку  и выбрать **Создать фильтр (Basic Filter)**.

14. На шаге 1 выберите оператор **Равно (Equal to)** и введите **3** в поле **Значение (Value)**.

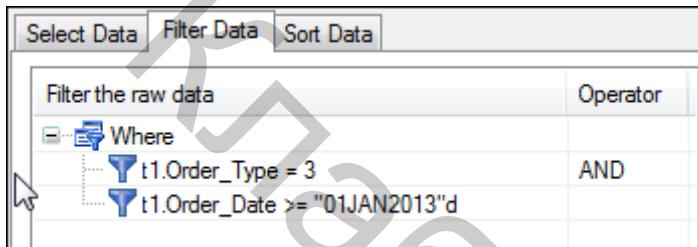


15. Нажмите кнопку Вперёд (Next), чтобы проверить настройки фильтра, а затем нажмите кнопку Готово (Finish).
16. Перетащите Order\_Date на вкладку Фильтрация данных (Filter Data). На шаге 1 выберите оператор Больше или равно (Greater than or equal to) и введите "01JAN2013"d в поле Значение (Value).

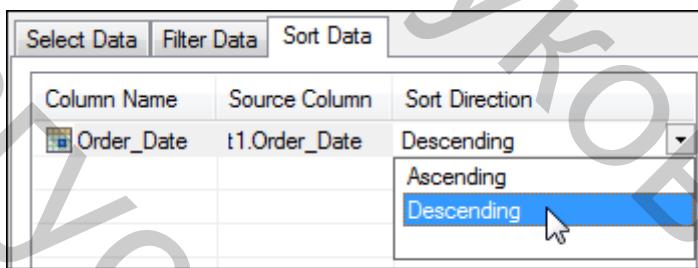


17. Нажмите кнопку Вперед (**Next**), чтобы проверить фильтр, а затем нажмите кнопку Готово (**Finish**).

Два фильтра автоматически связываются с помощью оператора AND.



18. Перейдите на вкладку Сортировка данных (**Sort Data**). Перетащите **Order\_Date** на эту вкладку и в качестве направления сортировки выберите По убыванию (**Descending**).



19. Чтобы выполнить запрос и просмотреть итоговую таблицу SAS, нажмите кнопку **Выполнить** (**Run**).

Фрагмент результатов (8 строк из 55, 6 столбцов из 12)

	Order_ID	Customer_ID	Order_Date	Delivery_Date	Order_Type	Pr
1	1244086685	14104	12/03/2014	06DEC2014	3	2
2	1244086685	14104	12/03/2014	06DEC2014	3	2
3	1243644877	70079	10/16/2014	19OCT2014	3	2
4	1243485097	11	09/28/2014	02OCT2014	3	2
5	1243462945	24	09/25/2014	28SEP2014	3	2
6	1243279343	27	09/05/2014	10SEP2014	3	24
7	1243165497	70201	08/24/2014	29AUG2014	3	2
8	1243165497	70201	08/24/2014	29AUG2014	3	24

20. Сохраните проект как **Chapter4**.

**Исследование данных**

Исследование данных – это альтернативный способ для интерактивной фильтрации и сортировки данных, а также построения запроса в проекте.

15

**Панель исследования данных**

На этой панели находятся следующие настройки:

- выбор столбцов
- фильтрация строк
- сортировка строк
- просмотр свойств данных
- исследование статистик

16

**Исследование данных**

Все обновления настроек запроса ставятся в очередь и не применяются пока вы не нажмёте Применить обновления (**Apply Updates**). При этом создаётся новое представление данных.

Viewing: ORDERS

Pending Updates

- Custom Select
- Custom Filter
- Order\_Date is sorted oldest to youngest

Apply Updates

	Quantity	Total_Retail_Price	Profit
2	2	\$114.20	\$55.90
3	2	\$80.40	\$40.00
4	1	\$14.60	\$7.85
5	2	\$78.20	\$38.90
6	1	\$44.10	\$22.05
7	2	\$73.40	\$45.50
8	2	\$215.80	\$109.10
9	3	\$658.50	\$326.80
10	2	\$201.20	\$97.40
11	2	\$180.40	\$103.80
12	2	\$16.60	\$9.30
13	3	\$405.50	\$235.35
14	2	\$70.20	\$35.00
15	4	\$700.00	\$343.80
16	4	\$186.00	\$88.00
17	2	\$277.80	\$141.30
18	2	\$120.40	\$65.60
19	5	\$238.50	\$124.28
20	2	\$403.00	\$230.80
21	2	\$91.68	\$52.60
22	2	\$23.40	\$12.10

Viewing: ORDERS (View 1)

Pending Updates

Selected Columns

Properties

17

### 4.03 Тест

- Выберите Файл  $\Rightarrow$  Открыть  $\Rightarrow$  Исследование данных (File  $\Rightarrow$  Open  $\Rightarrow$  Data Exploration).
- Кликните два раза набор данных, расположенный на сервере Local  $\Rightarrow$  Libraries  $\Rightarrow$  SASHELP  $\Rightarrow$  Cars, чтобы открыть его.
- Нажмите Изменить (Edit) в области Выбранные столбцы (Selected Columns). Добавьте столбцы Make, Model, Type, MSRP, MPG\_City, MPG\_Highway в область Выбранные столбцы (Selected columns).
- Нажмите Добавить к обновлениям, ожидающим применения (Add to Pending Updates).

Selected Columns

Available columns:

Selected columns:

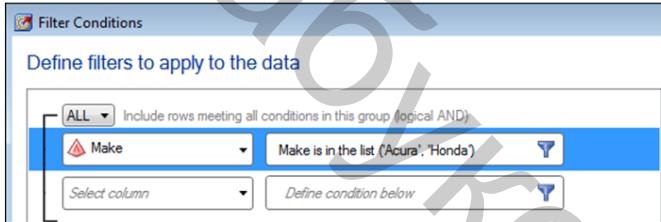
Add to Pending Updates

18

продолжается...

**4.03 Тест**

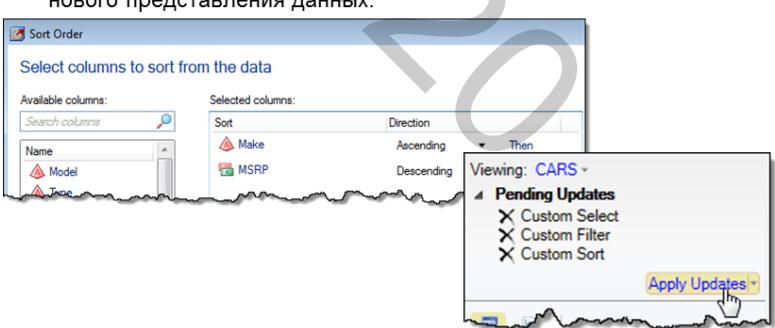
5. Нажмите Изменить (**Edit**) в области Настройки фильтра (**Filter Conditions**).
6. Выберите **Make** в выпадающем списке.
7. Нажмите на область Задать условие (**Define condition below**) и затем Щёлкните чтобы загрузить значения (**Click to load values**). Выберите значения **Acura** и **Honda**.
8. Нажмите Добавить к обновлениям, ожидающим применения (**Add to Pending Updates**).



19 продолжается...

**4.03 Тест**

9. Нажмите Изменить (**Edit**) в области Порядок сортировки (**Sort Order**).
10. Добавьте **Make** и **MSRP** в панель Выбранные столбцы (**Selected columns**). Выберите порядок сортировки по убыванию (**Descending**) для **MSRP**.
11. Нажмите Добавить к обновлениям, ожидающим применения (**Add to Pending Updates**).
12. Нажмите Применить обновления (**Apply Updates**) для создания нового представления данных.



20

**Добавление исследования в диаграмму процесса**

В диаграмму процесса можно добавить запрос с инструкциями по фильтрации и сортировке данных.

22

**Управление исследованиями**

Для изменения и доступа к исследованиям данных используйте вкладку История исследования данных (Data Exploration History) в панели ресурсов.

Name	Location
CARS	SASHHELP (Local)
ORDERS	EC10003 (Local)

23

**Настройка для вопроса**

1. Откройте Историю исследования данных (Data Exploration History) в панели ресурсов, двойным щелчком откройте исследование CARS.
2. Нажмите ссылку «Щёлкните для выполнения всех статистик» (**Click to run all statistics**).
3. Нажмите на созданные графики и эскизы отчётов. Просмотрите результаты.



24

#### 4.04 Вопрос с несколькими вариантами ответа

Какие два значения переменная **Type** принимает наиболее часто?

- a. Hybrid
- b. SUV
- c. Sedan
- d. Sports
- e. Truck
- f. Wagon

25

**4.05 Вопрос с несколькими вариантами ответа**

Какое среднее значение у переменной MPG\_Highway?

- 31.7
- 25.3
- 28.9
- 30.5

27

**Инструменты для построения запросов**

	Фильтр и сортировка	Конструктор запросов	Исследование данных
Сортировка данных	✓	✓	✓
Выбор строк, столбцов	✓	✓	✓
Создание новой таблицы	✓	✓	✓
Создание нового столбца	✗	✓	✗
Объединение таблиц	✗	✓	✗
Группировка, агрегация	✗	✓	✗
Атрибуты столбцов	✗	✓	✗
Удаление дубликатов	✗	✓	✓
Быстрое исследование агрегатов (Quick stats)	✗	✗	✓

29



## Упражнения

---

### 1. Создание таблицы в задаче «Фильтр и сортировка» с простым фильтром

С помощью задачи «Фильтр и сортировка» создайте таблицу, в которой сотрудники компании в Сан-Диего выводятся в отсортированном порядке по почтовому индексу.

- Создайте новый проект или воспользуйтесь существующим проектом **Chapter4**. Добавьте в проект таблицу **employee\_addresses**.
- С помощью задачи «Фильтр и сортировка» создайте таблицу для руководителя офиса в Сан-Диего. В эту таблицу должны входить столбцы **Employee\_ID**, **Employee\_Name**, **Street\_Number**, **Street\_Name** и **Postal\_Code**.
- Создайте фильтр, необходимый для включения в выходную таблицу только сотрудников из Сан-Диего (San Diego).
- Отсортируйте таблицу по возрастанию почтового индекса (столбец **Postal\_Code**).
- Присвойте задаче и таблице имя **SanDiegoEmployees**.
- Запустите задачу по созданию таблицы.

Фрагмент результатов (8 строк из 112)

		<b>Employee_ID</b>		<b>Employee_Name</b>		<b>Street_Number</b>		<b>Street_Name</b>		<b>Postal_Code</b>
1		121103		Farnsworth, Brian		8458		Crispwood Ct		91901
2		120278		Jongleux, Binit		3053		Blackpine Ct		91910
3		121127		Mangini, Keyna		9		Garner Rd		91910
4		120260		Fletcher, Christine		2416		Cove Point Dr		91910
5		120765		Kokoszka, Nikeisha		2416		Cove Point Dr		91910
6		121027		Rudder, Allan		941		Dixon Dr		91911
7		121059		Carhde, Jacquin		5637		Greenfield Dr		91911
8		120777		Sacher, Kary		2547		Allenby Dr		91913
9		121000		Helpin, Melvin		1000		Alley Ln		91902

- Сохраните проект как **Chapter4**.

Клабуков Максим  
Август 2020

## Создание новых столбцов при помощи выражений

**Цели**

- Узнать, как создать новый столбец в запросе при помощи выражения.

34

**Бизнес-сценарий**

Компании Orion Star нужно исследовать способы доставки, вычислив, сколько дней проходит с даты заказа до даты доставки. Также необходимо вычислить полную сумму счёта, сложив стоимость заказа и стоимость доставки.

Orders				
Order _Date	Delivery _Date	Total_Retail _Price	Shipping	

способы доставки

полная сумма счёта

35

**Вычисляемые столбцы (Computed Columns)**

В конструкторе запросов можно добавить в запрос новые агрегированные столбцы, перекодированные столбцы, столбцы, основанные на выражении.

Query name: Query Builder   Output name: WORK.QUERY\_FOR\_ORION\_PROFIT   Change...

Computed Columns | Prompt Manager | Preview | Tools | Options

Add Tables | Delete | Join Tables

Select Data | Filter Data | Sort Data

Column Name	Source Column	Summary
Order_ID (Order ID)	t1.Order_ID	
Product_ID (Product ID)	t1.Product_ID	
Product_Line (Product Line)	t1.Product_Line	
Product_Category (Product Category)	t1.Product_Category	
Product_Name (Product Name)	t1.Product_Name	

Чтобы начать создание нового столбца, нажмите на кнопку

36

**Мастер создания вычисляемого столбца**

В процессе создания нового столбца мастер запрашивает его атрибуты: имя, ярлык, формат.

New Computed Column  
1 of 4 Select a type

Summarized column  
Recorded column  
**Advanced expression**  
From an existing computed column

New Computed Column  
2 of 4 Build an advanced expression

Enter an expression:  
t1.Profit / t1.Quantity

New Computed Column  
3 of 4 Modify additional options

Column Name: Profitper  
Label: Profit per item  
Summary: NONE  
Expression: t1.Profit / t1.Quantity  
Format: DOLLAR10.2

37

**Функции SAS**

Функция SAS – это подпрограмма, которая возвращает значение, определяемое указанными аргументами.

Общий синтаксис функции SAS:

```
function-name(argument1,argument2,...)
```

Пример:

```
sum(Salary,Bonus)
```

38

**Использование функций SAS**

Фрагмент вывода PROC PRINT

Employee_ID	Qtr1	Qtr2	Qtr3	Qtr4	Total
120267	15	15	15	15	60
120269	20	20	20	20	80
120271	20	20	20	20	80
120275	15	15	15	15	60
120660	25	25	25	25	100
120669	15	15	15	15	60
120671	20	20	20	20	80

39

**SAS** THE POWER TO KNOW.

## Списки переменных SAS

Вместо раздельного ввода переменных можно использовать список переменных SAS.

```
data contrib;
  set orion.employee_donations;
  Total=sum(of Qtr1-Qtr4);
  if Total ge 50;
run;
```

При использовании списка переменных SAS в функции SAS укажите ключевое слово OF перед именем первой переменной в списке.

40 p205d01

**SAS** THE POWER TO KNOW.

## Списки переменных SAS

Список переменных SAS — это сокращенный способ указания группы имен переменных. SAS допускает использование следующих списков переменных:

- пронумерованный диапазон;
- диапазон имен;
- префикс имени;
- специальные списки имен SAS.

41

**Использование функций SAS**

Функции SAS могут делать следующие операции:

- выполнять арифметические действия
- вычислять статистики (сумму, среднее, стандартное отклонение)
- выполнять действие с датами в формате SAS
- выполнять операции с текстовыми значениями
- а также много других задач

 Статистические функции игнорируют пропущенные значения.

42

**4.06 Вопрос с несколькими вариантами ответа**

Каким будет результат выражения при данных значениях **Var1**, **Var2** и **Var3**?

- a. . (пропущенное значение)
- b. 3
- c. 9
- d. 12

Var1	Var2	Var3
9	.	3

**Var1+Var2+Var3**

43

 THE SAS INSTITUTE INC.  
4.07 Вопрос с несколькими вариантами ответа

Каким будет результат выражения при данных значениях **Var1**, **Var2** и **Var3**?

a. . (пропущенное значение)  
b. 3  
c. 9  
d. 12

Var1	Var2	Var3
9	.	3

**sum(Var1,Var2,Var3)**

45

## Создание столбцов с помощью выражений



Компании Orion Star требуется проанализировать способы доставки заказов. Для этого нужно определить количество дней, прошедших с момента оформления каждого заказа до даты его доставки. Компании также необходимо вычислить общую сумму счета, выставленного клиенту, которая складывается из итоговой розничной цены и расходов на доставку. Для этого создайте с помощью конструктора запросов таблицу **shipping** с новыми вычисленными столбцами.

1. В проекте **Chapter4** двойным щелчком откройте таблицу **orders**. На панели инструментов сетки данных выберите Конструктор запросов (**Query Builder**).
 

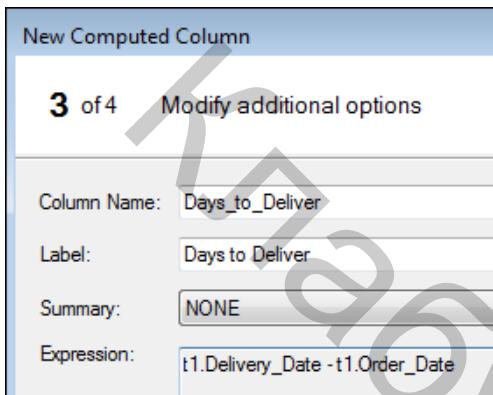
Если проект **Chapter4** не был сохранен, создайте новый проект и добавьте в него таблицу **orders**.
2. В поле Имя запроса (**Query name**) введите **Shipping Detail Query**. Нажмите кнопку Изменить (**Change**) и введите **shipping** в поле Имя файла (**File name**). Нажмите кнопку Сохранить (**Save**).



3. Убедитесь, что вкладка Выбор данных (**Select Data**) активна. Добавьте следующие столбцы из таблицы **orders**: **Order\_ID**, **Order\_Date**, **Delivery\_Date**, **Product\_ID**, **Total\_Retail\_Price**, **Shipping** и **Profit**.
4. Чтобы добавить в запрос новый столбец, щелкните кнопку на вкладке Выбор данных (**Select Data**) или выберите Вычисленные столбцы (**Computed Columns**)  $\Rightarrow$  Создать (**New**).
5. На шаге 1 выберите **Расширенное выражение** (**Advanced expression**)  $\Rightarrow$  Вперёд (**Next**).
6. На шаге 2 разверните узел **Выбранные столбцы** (**Selected Columns**). Дважды щелкните **Delivery\_Date**, чтобы добавить столбец в выражение. Выберите или введите знак минус «-», а затем дважды щелкните **Order\_Date**, чтобы завершить выражение. Нажмите кнопку Вперёд (**Next**).



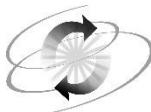
7. На шаге 3 в поле Имя столбца (**Column Name**) введите **Days\_to\_Deliver**. Введите **Days to Deliver** (**число дней до доставки**) в поле Ярлык (**Label**). Нажмите кнопку Вперед (**Next**).



8. На шаге 4 проверьте сводку с настройками нового столбца и нажмите кнопку Готово (**Finish**).
9. Повторите процесс, чтобы создать столбец **Invoice\_Amt**. Сначала нажмите кнопку .
10. На шаге 1 выберите Расширенное выражение  $\Rightarrow$  Вперед (**Advanced expression  $\Rightarrow$  Next**).
11. На шаге 2 прокрутите список вниз. В папке Функции (Functions) найдите функцию SUM. Дважды щелкните **SUM Function** (**функция SUM**), чтобы добавить ее в выражение. Синтаксис для функции SUM указан справа от списка функций.
12. Перейдите в верхнюю часть папки Функции (Functions) и сверните ее. Разверните узел Выбранные столбцы (**Selected Columns**). Дважды щелкните **Total\_Retail\_Price**, а затем введите или выберите запятую. Дважды щелкните **Shipping**, чтобы выражение имело следующий вид:
- 
13. Нажмите кнопку Вперед (**Next**). На шаге 3 в поле Имя столбца (**Column Name**) введите **Invoice\_Amt**. В поле Ярлык (**Label**) введите **Invoice Amount**. Чтобы значения отображались в виде денежных единиц, нажмите кнопку Изменить (**Change**) рядом с полем Формат (**Format**). В окне Формат (**Format**) в области Категории (**Categories**) выберите Валюта (**Currency**), а в области Форматы (**Formats**) выберите **DOLLARw.d**. Измените общее число знаков на **8**, а число десятичных разрядов — на **2**. Нажмите кнопки **OK  $\Rightarrow$  Вперед (Next)**.
14. Просмотрите сводные данные о настройках и нажмите кнопку Готово (**Finish**). Новые столбцы будут добавлены в область выбора и на вкладку Выбор данных (Select Data).
15. Чтобы выделить заказы с самым большим сроком доставки, отсортируйте таблицу в убывающем порядке по столбцу **Days\_to\_Deliver**. Откройте вкладку Сортировка данных (**Sort Data**) и перетащите на нее столбец **Days\_to\_Deliver**. В качестве направления сортировки выберите По убыванию (**Descending**).
16. Выполните запрос и изучите результаты. Сохраните проект **Chapter4**.

Фрагмент результатов (8 строк из 617)

	Order_ID	Order_Date	Delivery_Date	Product_ID	Total_Retail_Price	Shipping	Profit	Days_to_Deliver	Invoice_Amt
1	1236483576	22JUL2012	02AUG2012	240200200071	\$74.80	\$10.95	\$34.80	11	\$85.75
2	1237165927	27SEP2012	08OCT2012	240500100057	\$76.80	\$10.95	\$42.60	11	\$87.75
3	1236965430	08SEP2012	18SEP2012	240200100050	\$27.00	\$6.95	\$13.70	10	\$33.95
4	1243227745	31AUG2014	10SEP2014	230100500004	\$6.40	\$6.95	\$3.70	10	\$13.35
5	1241298131	29JAN2014	08FEB2014	240100400058	\$37.40	\$8.95	\$7.75	10	\$46.35
6	1231305521	27AUG2010	04SEP2010	220200100035	\$125.20	\$15.95	\$62.40	8	\$141.15
7	1232648239	07APR2011	15APR2011	210201000126	\$6.50	\$6.95	\$4.20	8	\$13.45
8	1233998114	27OCT2011	03NOV2011	220100400022	\$98.90	\$12.95	\$51.20	7	\$111.85



## Упражнения

### Уровень 1

#### 2. Использование конструктора запросов для создания таблицы с вычисленным столбцом

С помощью конструктора запросов создайте столбец **New\_Salary**, значение которого складывается из текущей заработной платы и двухпроцентной надбавки.

- Добавьте в проект **Chapter4** таблицу SAS **employee\_payroll**.



Если проект **Chapter4** отсутствует, создайте его.

- С помощью конструктора запросов создайте запрос **New Salary Query**. Присвойте выходной таблице имя **New\_Salary**.
- Включите в таблицу столбцы **Employee\_ID**, **Employee\_Gender**, **Salary**, **Birth\_Date** и **Employee\_Hire\_Date**. Измените имя столбца **Salary** на **Old\_Salary** в его свойствах.
- Включите только действующих сотрудников, у которых отсутствует значение в столбце **Employee\_Term\_Date**.

Создайте столбец с именем **New\_Salary**, значение которого складывается из текущей заработной платы и двухпроцентной надбавки. Отформатируйте его так, чтобы в нем указывался символ доллара и два десятичных разряда.

Подсказка: Умножьте значение в столбце **Old\_Salary** на 1.02.

- Выполните запрос и проверьте результаты.

Фрагмент результатов (8 строк из 308)

	Employee_ID	Employee_Gender	Old_Salary	Birth_Date	Employee_Hire_Date	New_Salary
1	120101	M	163040	18AUG81	01JUL08	\$166,300.80
2	120102	M	108255	11AUG74	01JUN94	\$110,420.10
3	120103	M	87975	22JAN54	01JAN79	\$89,734.50
4	120104	F	46230	11MAY59	01JAN86	\$47,154.60
5	120105	F	27110	21DEC79	01MAY04	\$27,652.20
6	120106	M	26960	23DEC49	01JAN79	\$27,499.20
7	120107	F	30475	21JAN54	01FEB79	\$31,084.50
8	120108	F	27660	23FEB89	01AUG11	\$28,213.20
9			26495	15DEC91		\$7,944.00

f. Сохраните проект **Chapter4**.

## Группировка и агрегация данных в запросе

**Цели**

- Научиться добавлять группирующую переменную в запрос.
- Научиться выбирать переменную для анализа и задавать функцию для подсчета агрегатов.
- Узнать, как фильтровать агрегированные данные в том же запросе.

52

**Бизнес-сценарий**

Компании Orion Star нужно спланировать рекламную кампанию для наиболее прибыльных товаров. Необходимо создать список товаров, для которых суммарное значение прибыли превышает \$500.

**Orders** (фрагмент)

Product_ID	Profit
230100500026	\$551.80
230100500026	\$137.95
230100500056	\$4.50
230100500056	\$9.00
230100500056	\$4.50

**Query\_for\_Orders** (фрагмент)

Product_ID	Total_Profit
230100500026	✓ \$689.75
230100500056	✗ \$18.00

53

**Группировка данных**

Product_Category	Quantity	Profit
Golf	2	\$84.50
Golf	1	\$1.50
Indoor Sports	2	\$600.20
Indoor Sports	4	\$32.20
Indoor Sports	1	\$36.35
Indoor Sports	2	\$14.70
Indoor Sports	3	\$825.45
Indoor Sports	1	\$7.20
Indoor Sports	1	\$77.90
Indoor Sports	1	\$47.50
Indoor Sports	1	\$8.80
Indoor Sports	2	\$563.70
Indoor Sports	1	\$31.10
Indoor Sports	1	\$30.55
Indoor Sports	1	\$175.15
Indoor Sports	3	\$850.65
Indoor Sports	2	\$100.20
Indoor Sports	1	\$11.00
Outdoors	2	\$206.90
Outdoors	2	\$142.40

54

Для группировки и агрегации данных можно использовать Конструктор запросов.

Product_Category	SUM_of_Quantity	AVG_of_Profit
Assorted Sports Articles	121	\$95.54
Children Sports	49	\$30.93
Clothes	206	\$48.65
Golf	143	\$131.39
Indoor Sports	27	\$213.29
Outdoors	240	\$107.99
Racket Sports	37	\$90.04
Running - Jogging	64	\$64.19
Shoes	69	\$78.94
Swim Sports	11	\$28.85
Team Sports	75	\$31.08
Winter Sports	36	\$147.72

**Группировка данных**

На вкладке Выбор данных (Select Data) задаются параметры группировки и агрегации данных.

Выберите функцию для агрегируемых столбцов.

Столбцы, для которых не назначены агрегаты, автоматически определяют группы.

55

Мастер создания вычисленных столбцов также можно использовать для создания столбцов с агрегированными данными.

Порядок столбцов на вкладке Выбор данных (Select Data) определяет исходную иерархию для группировки. Однако ее можно изменить в области Сводные группы (Summary Groups). Для ручной настройки группирующих переменных снимите флажок **Выбирать группы автоматически (Automatically select groups)** и нажмите **Изменить группы (Edit Groups)**.

Имеются следующие функции для агрегирования данных:

AVG, MEAN (Среднее)	STD (стандартное отклонение)
COUNT, FREQ, N (число непропущенных значений)	STDERR (стандартная ошибка среднего)
CSS (скорректированная сумма квадратов)	SUM (Сумма)
CV (коэффициент вариации)	SUMWGT (сумма значений переменной WEIGHT)
MAX или MIN (максимум, минимум)	T ( $t$ -значение для проверки гипотезы о том, что математическое ожидание равно нулю)
NMISS (число пропущенных значений)	USS (нескорректированная сумма квадратов)
PRT (двустороннее $p$ -значение для $t$ -статистики Стьюдента)	VAR (дисперсия)
RANGE (диапазон значений)	



Некоторые функции агрегирования можно использовать с модификатором **DISTINCT**. Например, функция **AVG** вычисляет среднее значение всех наблюдений в каждой группе. Однако функция **AVG DISTINCT** сначала исключает из расчётов повторяющиеся значения (дубликаты), поэтому она выдаёт среднее значение только для уникальных значений в каждой группе.

**Группировка по значениям столбца**

Результат запроса будет сгруппирован так, что каждой уникальной комбинации группирующих столбцов будет отвечать одна строка в выводе и посчитанные для этих групп агрегаты.

Product_Category	SUM_of_Quantity	AVG_of_Profit
Assorted Sports Articles	121	\$95.54
Children Sports	49	\$30.93
Clothes	206	\$48.65
Golf	143	\$131.39
Indoor Sports	27	\$213.29
Outdoors	240	\$107.99
Racket Sports	137	\$90.04
Running - Jogging	64	\$64.19
Shoes	69	\$78.94
Swim Sports	11	\$28.85
Team Sports	75	\$31.08
Winter Sports	36	\$147.72

56

### 4.08 Тест

1. Откройте Конструктор запросов (**Query Builder**) для любого источника данных в текущем проекте.
2. Откройте вкладку Фильтрация данных (**Filter Data**) и изучите её содержимое.
3. Вернитесь на вкладку Выбор данных (**Select Data**) и добавьте на неё два любых столбца.
4. Для одного из столбцов задайте агрегат **Count** в поле **Summary**.
5. Повторно откройте вкладку Фильтрация данных (**Filter Data**).

Как изменилась вкладка Фильтрация данных (**Filter Data**) после включения агрегации в запросе?

57

**Фильтрация данных**

Вкладка Фильтрация данных (Filter Data) позволяет создавать фильтры для исходных данных и для агрегатов.

59

Сначала запрос извлекает из таблицы все строки, удовлетворяющие условиям, которые указаны в области «Фильтровать необработанные данные» (Filter the raw data). Затем полученные данные группируются по назначенным группирующими переменным, и для каждой уникальной группы переменных (определенными на вкладке Выбор данных(Select Data)) вычисляются агрегированные значения. После этого к агрегированным значениям применяются фильтры, заданные в области «Фильтровать суммарные данные» (Filter summarized data).

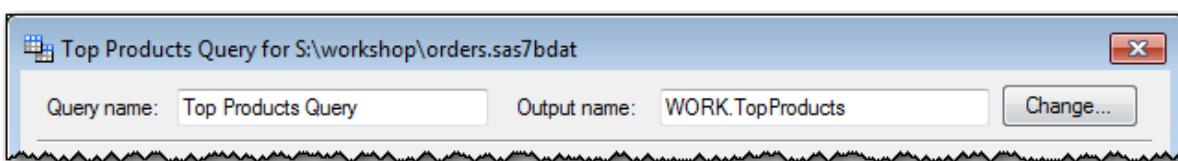
В запросе SQL фильтрация по агрегатам задаётся в операторе HAVING.



## Сведение и фильтрация по группам

Компания Orion Star хочет организовать рекламную акцию для продвижения наиболее популярных продуктов. Компании требуется список всех продуктов, суммарная прибыль по которым превышает 500 долларов США. Воспользуйтесь конструктором запросов, чтобы вычислить суммарную прибыль, сгруппировать и отфильтровать набор данных **orders**.

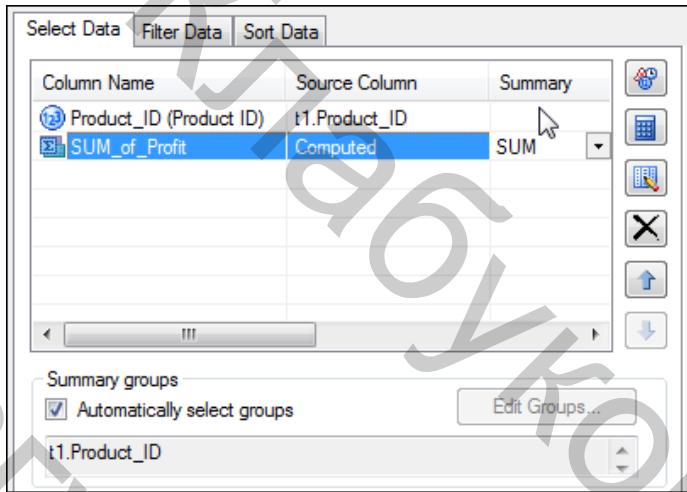
1. В проекте **Chapter4** щелкните правой кнопкой мыши таблицу **orders** в дереве проекта или окне схемы процесса «Process Flow» и выберите Конструктор запросов (Query Builder).
2. Присвойте запросу имя **Top Products Query**, а выходной таблице — **topproducts**.



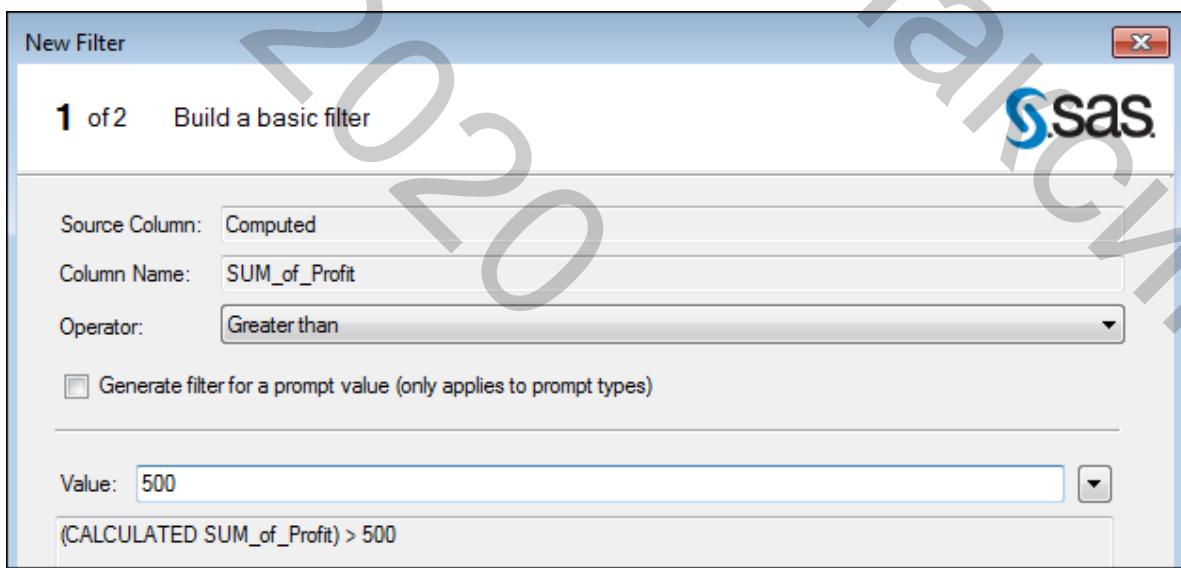
3. Дважды щелкните столбцы **Product\_ID** и **Profit** в области выбора, чтобы добавить их на вкладку Выбор данных (Select Data).

4. На вкладке Выбор данных (Select Data) щелкните столбец **Summary** (Сводная информация) для столбца **Profit**. В раскрывающемся списке выберите статистику **SUM**.

 При выборе сводной статистики устанавливается флажок Выбирать группы автоматически (Automatically select groups).



5. Перейдите на вкладку Фильтрация данных (Filter Data). Перетащите **SUM\_of\_Profit** из области выбора в область Фильтровать суммарные данные (Filter the summarized data).
6. В мастере создания фильтра убедитесь, что фильтр основан на столбце **SUM\_of\_Profit**. Выберите **Больше чем (Greater than)** в поле **Оператор (Operator)**. Введите **500** в поле **Значение (Value)** и нажмите кнопку Готово (Finish).



7. Перейдите на вкладку Сортировка данных (Sort Data). Перетащите **SUM\_of\_Profit** на эту вкладку и выберите порядок сортировки По убыванию (Descending).

8. Нажмите кнопку Выполнить (**Run**) и проверьте результаты. Сохраните проект **Chapter4**.

Фрагмент результатов (8 строк из 19)

	Product_ID	SUM_of_Profit
1	230100700009	\$1,660.65
2	230100700008	\$1,578.50
3	230100700011	\$1,322.00
4	240200200013	\$1,139.40
5	240200100173	\$946.40
6	240400200097	\$877.80
7	240300300090	\$850.65
8	240300300070	\$825.45

## Упражнения

### Уровень 1

#### 3. Использование конструктора запросов для группировки и сведения данных

С помощью конструктора запросов создайте выходную таблицу со средней заработной платой по городу.

- a. Добавьте в проект **Chapter4** таблицу **employee\_master**.

 Также можно создать новый проект.

- b. С помощью конструктора запросов создайте запрос с именем **Average Salary per City Query** и выходную таблицу с именем **AvgSal\_City**.

- c. Включите в запрос только столбцы **City** и **Salary**.

- d. Выберите статистику **AVG** (среднее значение) для **Salary**.

 После выбора сводной статистики в левой нижней части вкладки Выбор данных (Select Data) появится раздел Сводные группы (Summary groups). Также будет установлен флагок Выбирать группы автоматически (**Automatically select groups**).

- e. Отсортируйте таблицу в убывающем порядке по столбцу **AVG\_of\_Salary**.

- f. Выполните запрос и просмотрите результаты. Сохраните проект **Chapter4**.

## Результаты

	City	Avg_of_Salary
1	Miami-Dade	\$46,933.59
2	Philadelphia	\$45,292.87
3	San Diego	\$36,696.65
4	Sydney	\$33,451.81
5	Melbourne	\$31,394.62

## Объединение таблиц

**Цели**

- Научиться объединять таблицы по общим столбцам.
- Узнать, как оставить в выводе только совпадающие строки.

65

**Бизнес-сценарий**

После определения товаров, для которых прибыль превышает \$500, аналитикам требуется узнать о них больше подробной информации, в том числе название, категорию, поставщика и страну. Эти столбцы содержатся в трёх различных таблицах.

	Product_ID	SUM_of_Profit	Product_Category	Product_Name	Supplier_Name	Country_Name
1	230100700009	\$1,660.65	Outdoors	Family Holiday 6	Petterson AB	Sweden
2	230100700008	\$1,578.50	Outdoors	Family Holiday 4	Petterson AB	Sweden
3	230100700011	\$1,322.00	Outdoors	Hurricane 4	Petterson AB	Sweden
4	240200200013	\$1,139.40	Golf	Master Gof Rain Suit	Mike Schaeffer Inc	United States
5	240200100173	\$946.40	Golf	Proplay Executive Bi-Metal Graphite	Van Dammeren International	Netherlands
6	240400200097	\$877.80	Racket Sport	Smasher Tg 70 Tennis String Roll	British Sports Ltd	United Kingdom
7	240300300090	\$850.65	Indoor Sports	Top R&D Long Jacket	Top Sports Inc	United States
8	240300300070	\$825.45	Indoor Sports	Top Men's R&D Ultimate Jacket	Top Sports Inc	United States
9	2409100026	\$999.95	Men's Casual Dress	Top Men's Casual Dress	Top Sports Inc	United States

topproducts      products      country\_lookup

66

**Бизнес-сценарий**

Чтобы включить в результаты все необходимые столбцы, нужно объединить таблицы **topproducts**, **products** и данные в файле Excel **country\_lookup**.

67

Объединение таблиц может потребоваться для решения следующих задач:

- создание отчета со столбцами из нескольких таблиц;
- вычисление нового столбца на основе столбцов, которые находятся в разных таблицах;
- подстановка сведений из таблицы-справочника в базовую таблицу;
- определение значений в столбце, которые присутствуют или отсутствуют в других таблицах.

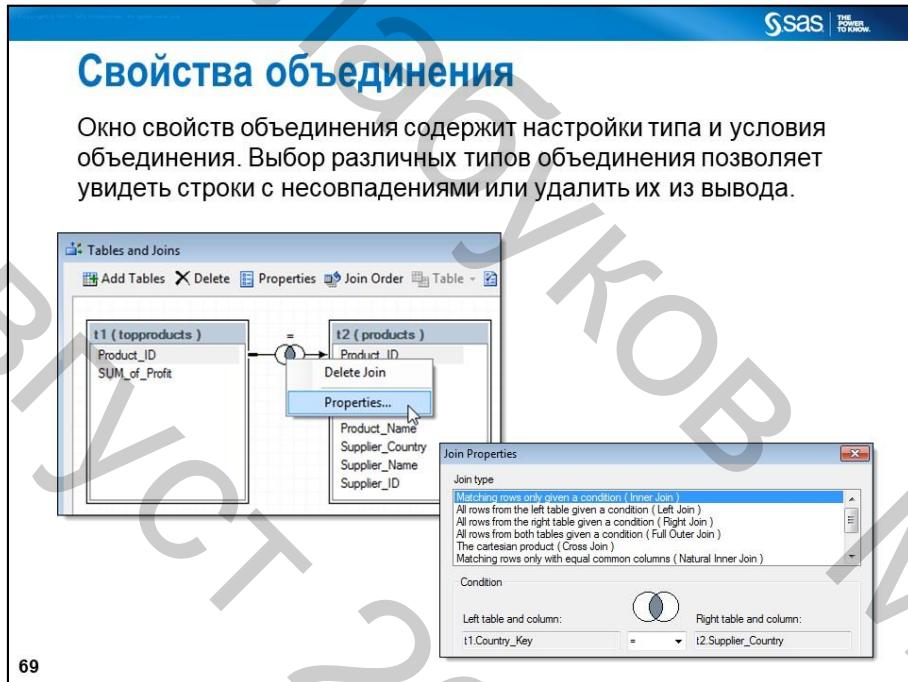
**Окно объединения таблиц**

Объединение таблиц позволяет извлекать и обрабатывать данные из нескольких источников одновременно.

68

Выберите **Добавить таблицы (Add Tables)**, чтобы включить дополнительные таблицы в запрос. Конструктор запросов попытается объединить таблицы по столбцам с одинаковым именем и типом. Если такие столбцы отсутствуют, выводится предупреждение о необходимости объединения столбцов вручную.

 Чтобы исключить объединение таблиц по совпадающим столбцам, выберите Сервис ⇨ Параметры ⇨ Запрос (Tools ⇨ Options ⇨ Query) и снимите флагок «автоматически пытаться объединять таблицы в запросе» (Automatically attempt to join tables in query).



69

Для получения дополнительных сведений о типах объединений нажмите Справка (Help) в окне «Tables and Joins» (таблицы и объединения), а затем выберите **Understanding the types of joins**.

**Типы объединения**

По умолчанию Конструктор запросов выполняет внутреннее объединение (*inner join*), при котором в результатах запроса остаются только совпадающие строки.

**Employee\_Payroll**

Employee_ID	Salary
120101	163040
120102	108255
120103	87975
120104	92500

**Employee\_Organization**

Employee_ID	Department
120101	Sales Management
120102	Sales Management
120103	Engineering
120105	Administration

70

**Сокращение кода**

Псевдоним таблицы-это временное альтернативное имя таблицы. С помощью псевдонимов мы можем упростить запрос.

```
SELECT alias-1.object-item<, ...alias-2.object-item>
  FROM table-name <AS> alias-1,
       table-name <AS> alias-2
 WHERE join-condition(s)
   <other clauses>;
```

Ключевое слово AS необязательно в синтаксисе псевдонима таблицы.

71

**Альтернативный синтаксис Join.**

Этот альтернативный синтаксис используется для указания типа и условий соединения с помощью ключевого слова ON.

```
select c.ID, Name, Action, Amount
  from customers as c
    inner join
      transactions as t
    on c.ID=t.ID;
```

```
SELECT object-item <, ...object-item>
  FROM table-name <<AS> alias>
    INNER JOIN
      table-name <<AS> alias>
    ON join-condition(s)
    WHERE sql-expression
      <other clauses>;
```

72

## Типы объединения

Полное внешнее объединение (*full outer join*) включает все строки из обеих таблиц.

Employee_Payroll		Employee_Organization	
Employee_ID	Salary	Employee_ID	Department
120101	163040	120101	Sales Management
120102	108255	120102	Sales Management
120103	87975	120103	Engineering
120104	92500	120105	Administration

Employee_ID	Salary	Department
120101	163040	Sales Management
120102	108255	Sales Management
120103	87975	Engineering
120104	92500	
120105		Administration

73

**Типы объединения**

Левостороннее объединение (*left join*) включает все строки из таблицы **Employee\_Payroll**.

Employee_Payroll		Employee_Organization	
Employee_ID	Salary	Employee_ID	Department
120101	163040	120101	Sales Management
120102	108255	120102	Sales Management
120103	87975	120103	Engineering
120104	92500	120105	Administration

Employee_ID	Salary	Department
120101	163040	Sales Management
120102	108255	Sales Management
120103	87975	Engineering
120104	92500	

74

## Outer Joins

Синтаксис Outer join аналогичен альтернативному синтаксису inner join .

```
proc sql;
  title 'All Customers';
  select *
    from customers as c
    left join
      transactions as t
    on c.ID=t.ID;
  quit;
```

Ключевое слово ON определяет условия соединения.

**SELECT object-item <, ...object-item>**  
**FROM table-name <> AS alias>**  
**LEFT|RIGHT|FULL JOIN**  
**table-name <> AS alias>**  
**ON join-condition(s)**  
**<other clauses>;**

75

**4.09 Вопрос с несколькими вариантами ответа**

Какие значения **Employee\_ID** будут включены в результаты правостороннего объединения (right join)?

- 120101
- 120102
- 120103
- 120104
- 120105

**Employee\_Payroll**

Employee_ID	Salary
120101	163040
120102	108255
120103	87975
120104	92500

**Employee\_Organization**

Employee_ID	Department
120101	Sales Management
120102	Sales Management
120103	Engineering
120105	Administration

76

**Результаты объединения**

При объединении таблиц в Конструкторе запросов можно задать фильтр или порядок сортировки результатов для любых столбцов из входных таблиц. Так же можно вычислить новые столбцы, задать группировку и включить подсчёт агрегатов.

The screenshot shows the SAS Query Builder interface with three tables joined:

- t1 (topproducts)**: Contains columns Product\_ID, SUM\_of\_Profit.
- t2 (products)**: Contains columns Product\_ID, Product\_Line, Product\_Name, Product\_Group, Supplier\_Name, Supplier\_Country, Supplier\_ID.
- t3 (COUNTRY\_LOOKUP)**: Contains columns Country\_Key, Country\_Name.

The join condition is `t2.Product_Line = 'Clothes & Shoes'`. A filter is applied to `t2` where `Product_Line = 'Clothes & Shoes'`. The resulting query output shows a single column `SUM_of_Profit` with a sort direction of `Descending`.

78



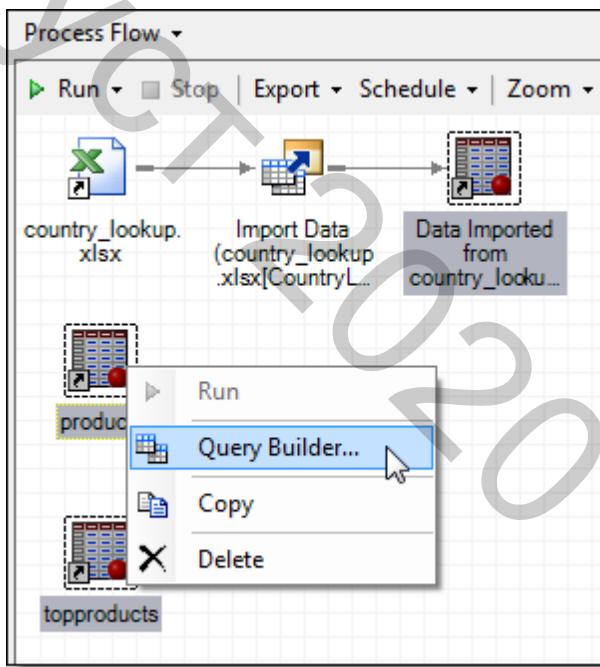
## Объединение таблиц

В предыдущем запросе были определены продукты, суммарная прибыль по которым превышает 500 долларов США. Аналитикам требуется более подробные сведения об этих популярных продуктах, включая название, категорию, поставщика и страну. Чтобы добавить необходимые столбцы, с помощью конструктора запросов объедините таблицу SAS **topproducts** с таблицей SAS **products** и таблицей в формате Excel **country\_lookup**.

1. При необходимости добавьте в проект **Chapter4** таблицы SAS **topproducts** и **products**.
2. Добавьте в проект файл Excel **country\_lookup**. Присвойте импортированной таблице имя **country\_lookup** и установите флажок **First row of range contains field names** (первая строка диапазона содержит имена полей). Не изменяйте атрибуты полей.
3. Открыв схему процесса, удерживайте нажатой клавишу **Ctrl key** и выделите таблицы **TopProducts**, **Products** и **Data Imported from country\_lookup**. Нажмите правой кнопкой мыши на любую из этих таблиц и выберите Конструктор запросов (**Query Builder**).



Если вы используете более старую версию Enterprise Guide, запустите Конструктор запросов для любой из трёх таблиц. Затем нажмите кнопку Добавить таблицы (**Add Tables**) и добавьте два других набора данных.



4. Появится сообщение о том, что SAS Enterprise Guide не удалось найти совпадающие столбцы, по которым нужно выполнить объединение всех трех таблиц. Нажмите кнопку **OK**.
5. Автоматически откроется окно Таблицы и соединения (Tables and Joins). Чтобы объединить таблицы **products** и **country\_lookup**, сначала выберите **Supplier\_Country** в таблице **products**. Выберите переменную **Supplier\_Country** еще раз и перетащите ее, чтобы связать с переменной **Country\_Key** в таблице **country\_lookup**.



Можно также щелкнуть **Supplier\_Country** правой кнопкой мыши и выбрать **Join with (соединить с) ⇒ <имя таблицы> ⇒ Country\_Key**.

- Автоматически откроется окно «Свойства соединения» (Join Properties). Убедитесь, что выбран тип объединения **внутреннее соединение (Matching rows only given a condition (Inner Join))** и условие **t2.Supplier\_Country = t3.Country\_Key**.



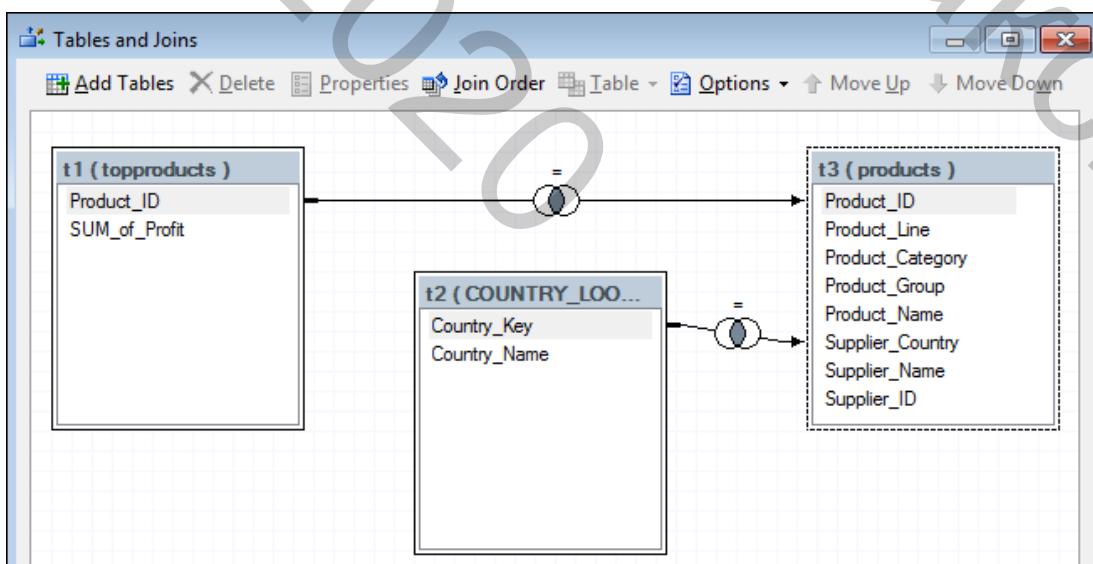
Псевдонимы таблиц (**t1**, **t2** или **t3**) могут отличаться в зависимости от порядка добавления таблиц в запрос.



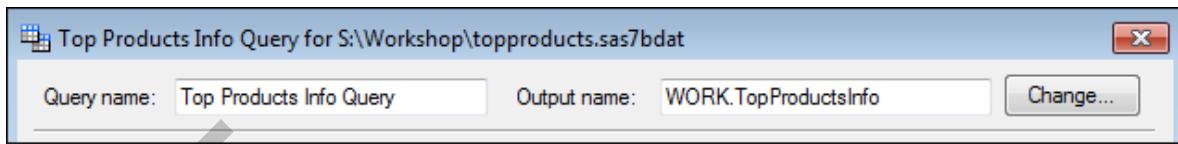
- Нажмите **OK** и проверьте, что три таблицы связаны так, как показано ниже. Нажмите кнопку **Закрыть (Close)** для возврата на вкладку Выбор данных (Select Data).



Расположение таблиц в окне Таблицы и соединения (Tables and Joins) может отличаться.



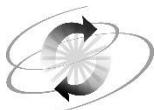
- Задайте имя запроса **Top Products Info Query** и имя выходной таблицы **TopProductsInfo**.



9. Двойным щелчком добавьте следующие столбцы на вкладку Выбор данных (Select Data): **Product\_ID, SUM\_of\_Profit, Product\_Category, Product\_Name, Supplier\_Name** и **Country\_Name**.
10. Перейдите на вкладку Сортировка данных (Sort Data) и задайте сортировку результатов по убыванию столбца **SUM\_of\_Profit**.
11. Нажмите кнопку Выполнить (Run) и запустите запрос. Проверьте результаты и сохраните проект **Chapter4**.

Фрагмент результатов (8 строк из 19)

	Product_ID	SUM_of_Profit	Product_Category	Product_Name	Supplier_Name	Country_Name
1	230100700009	\$1,660.65	Outdoors	Family Holiday 6	Petterson AB	Sweden
2	230100700008	\$1,578.50	Outdoors	Family Holiday 4	Petterson AB	Sweden
3	230100700011	\$1,322.00	Outdoors	Hurricane 4	Petterson AB	Sweden
4	240200200013	\$1,139.40	Golf	Master Golf Rain Suit	Mike Schaeffer Inc	United States
5	240200100173	\$946.40	Golf	Proplay Executive Bi-Metal Graphite	Van Dammeren International	Netherlands
6	240400200097	\$877.80	Racket Sports	Smasher Tg 70 Tennis String Roll	British Sports Ltd	United Kingdom
7	240300300090	\$850.65	Indoor Sports	Top R&D Long Jacket	Top Sports Inc	United States
8	240300300070	\$825.45	Indoor Sports	Top Men's R&D Ultimate Jacket	Top Sports Inc	United States
9	240100000026	\$800.00	Men's Fleece D	Top Men's Fleece D	Two	United States



## Упражнения

### Уровень 1

#### 4. Объединение данных из двух таблиц

С помощью конструктора запросов создайте таблицу, данные которой являются результатом объединения двух таблиц. Создайте новый столбец.

- Добавьте таблицы **employee\_payroll** и **employee\_addresses** в проект **Chapter4**, если это необходимо.
- С помощью конструктора запросов объедините таблицы **employee\_payroll** и **employee\_addresses**, чтобы создать таблицу **employee\_payroll\_location**. Присвойте запросу имя **Payroll Location Join Query**.
- Откройте окно Таблицы и соединения (Tables and Joins). Убедитесь, что две таблицы объединены по столбцу **Employee\_ID**.
- Добавьте на вкладку Выбор данных (Select Data) следующие столбцы: **Employee\_ID**, **Employee\_Name**, **Employee\_Gender**, **Birth\_Date**, **Salary**, **Street\_Number**, **Street\_Name**, **City**, **State** и **Country**.
- Отформатируйте столбец **Salary** так, чтобы в нем указывался символ доллара, точка и два десятичных разряда. Отформатируйте столбец **Birth\_Date**, чтобы в нём выводилась дата в соответствии с форматом 01JAN2004.
- Создайте столбец с именем **Bonus**, значение которого составляет 1,5% от значения столбца **Salary**. Отформатируйте значения так, чтобы они включали символ доллара, точку и два десятичных разряда.
- Включите в выходную таблицу только действующих сотрудников, у которых отсутствует значение в столбце **Employee\_Term\_Date**.
- Выполните запрос и просмотрите результаты. Сохраните проект **Chapter4**.

Фрагмент результатов (8 строк из 308)

	Employee_ID	Employee_Name	Employee_Gender	Birth_Date	Salary	Street_Number	Street_Name	City	State	Country	Bonus
1	121044	Abbott, Ray	M	11DEC1959	\$25,660.00	2267	Edwards Mill Rd	Miami-Dade	FL	US	\$384.90
2	120145	Aisbitt, Sandy	M	22JAN1969	\$26,060.00	30	Bingera Street	Melbourne	AU		\$390.90
3	120761	Akinfolarin, Tameaka	F	28DEC1991	\$30,960.00	5	Donnybrook Rd	Philadelphia	PA	US	\$464.40
4	120656	Amos, Salley	F	28JAN1979	\$42,570.00	3524	Calico Ct	San Diego	CA	US	\$638.55
5	121107	Anger, Rose	F	24APR1991	\$31,380.00	744	Chapwith Rd	Philadelphia	PA	US	\$470.70
6	120273	Antonini, Doris	F	07JUN1991	\$28,455.00	681	Ferguson Rd	Miami-Dade	FL	US	\$426.83
7	120759	Apr, Nishan	M	04NOV1969	\$36,230.00	105	Brack Penny Rd	San Diego	CA	US	\$543.45
8	120798	Ardskin, Elizabeth	F	23JUN1964	\$80,755.00	701	Glenridge Dr	Miami-Dade	FL	US	\$1,211.33

### Уровень 2

#### 5. Объединение данных из трех таблиц

- Выполните упражнение уровня 1.
- Добавьте в проект таблицу **country\_lookup**.

- c. Измените запрос **Payroll Location Join Query**. Вручную объедините таблицы **country\_lookup** и **employee\_addresses** по общему столбцу.



Общий столбец может иметь разные имена в двух таблицах.

- d. На вкладке Выбор данных (Select Data) удалите столбец **Country** и замените его столбцом из таблицы **country\_lookup**, содержащим полное название страны.
- e. Выполните запрос повторно и просмотрите результаты. Сохраните проект как **Chapter4**.

Фрагмент результатов (8 строк из 304)

	Employee_ID	Employee_Name	Employee_Gender	Birth_Date	Salary	Street_Number	Street_Name	City	State	Country_Name	Bonus
1	121044	Abbott, Ray	M	11DEC1959	\$25,660.00	2267	Edwards Mill Rd	Miami-Dade	FL	United States	\$384.90
2	120145	Aisbett, Sandy	M	22JAN1969	\$26,060.00	30	Bingera Street	Melbourne		Australia	\$390.90
3	120761	Akinfolarin, Tameaka	F	28DEC1991	\$30,960.00	5	Donnybrook Rd	Philadelphia	PA	United States	\$464.40
4	120656	Amos, Salley	F	28JAN1979	\$42,570.00	3524	Calico Ct	San Diego	CA	United States	\$638.55
5	121107	Anger, Rose	F	24APR1991	\$31,380.00	744	Chapwith Rd	Philadelphia	PA	United States	\$470.70
6	120273	Antonini, Doris	F	07JUN1991	\$28,455.00	681	Ferguson Rd	Miami-Dade	FL	United States	\$426.83
7	120759	Apr, Nishan	M	04NOV1969	\$36,230.00	105	Brack Penny Rd	San Diego	CA	United States	\$543.45
8	120798	Ardskin, Elizabeth	F	23JUN1964	\$80,755.00	701	Glenridge Dr	Miami-Dade	FL	United States	\$1,211.33

## Создание отчёта Сводная статистика (Summary Statistics)

Создание отчёта Сводная статистика

3

**Цели**

- Научиться создавать табличный отчёт «Сводная статистика».
- Научиться экспортить подсчитанные статистики в Excel.

Sas THE POWER TO KNOW.

Создание отчёта Сводная статистика

4

**Бизнес-сценарий**

Руководству требуется отчёт в формате HTML и файл в формате Excel, содержащий анализ прибыли по категориям и линейкам товаров.

Orion\_Profit

Sas THE POWER TO KNOW.

**Задача Сводная статистика (Summary Statistics)**

Эту задачу можно использовать для вычисления сводных статистик в разбивке по группам. Вы можете выводить результаты в отчёты, графики и наборы данных.

**Summary of Customer Age by Country**

Customer Country	N	Obs	Mean	Std Dev	Minimum	Maximum	N
AU	8	54.0	14.9	34.0	75.0	8	
CA	15	41.8	19.4	22.0	75.0	15	
DE	10	48.6	19.6	21.0	75.0	10	
IL	5	41.8	10.3	25.0	50.0	5	
TR	7	41.0	12.6	29.0	65.0	7	
US	28	42.3	14.9	23.0	75.0	28	
ZA	4	36.8	19.2	21.0	60.0	4	

Customer_Country	_WAY	_TYPE	_FREQ	Customer_Age_Mean	Customer_Age_StdDev	Customer_Age_Min	Customer_Age_Max	Customer_Age_N
1 AU	1	1	8	54	14.9	34	75	8
2 CA	1	1	15	41.8	19.4	22	75	15
3 DE	1	1	10	48.6	19.6	21	75	10
4 IL	1	1	5	41.8	10.3	25	50	5
5 TR	1	1	7	41.0	12.6	29	65	7
6 US	1	1	28	42.3	14.9	23	75	28
7 ZA	1	1	4	36.8	19.2	21	60	4

5

**Задача Сводная статистика**

Задача может быть настроена в режиме Мастера настроек и в стандартном режиме.

The screenshot shows the SAS interface with two windows open:

- Summary Statistics Wizard...**: A dialog box where "Summary Statistics..." is selected from a list of options.
- Assign variables to roles**: A sub-dialog where variables like Customer\_ID, Customer\_Age, and Customer\_Country are assigned to various task roles such as Analysis variables, Class level, Frequency count, and Group analysis by.

При настройке задачи в расширенном режиме в нём доступны настройки, отсутствующие в мастере настроек:

- сортировка по классифицирующей переменной;
- применение формата к классифицирующей переменной;
- включение пропущенных значений в вычисления;
- указание сочетаний классифицирующих переменных.

 При переходе из мастера в расширенный режим вернуться в мастер настроек невозможно.

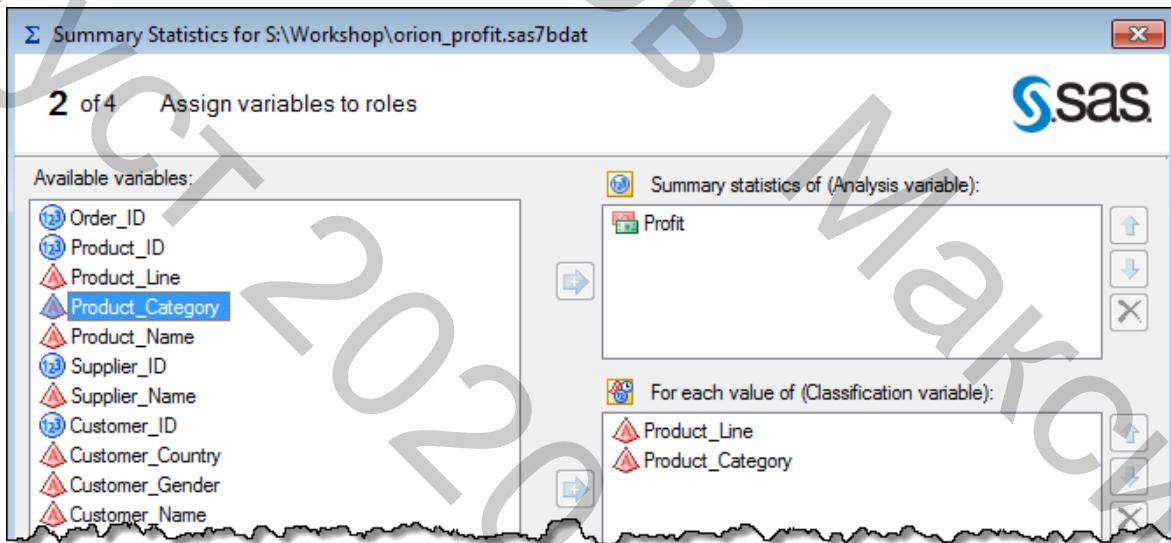
Клабуков Максим  
Август 2020



## Создание отчёта Сводная статистика

Руководству компании требуется отчет со сводкой по прибыли для каждой линейки и категории продуктов. HTML-отчет вместе с файлом Excel должны храниться на общем сетевом диске.

1. Создайте проект и добавьте в него таблицу **orion\_profit**.
2. В сетке данных выберите **Описательная статистика**  $\Rightarrow$  **Мастер сводной статистики** (**Describe**  $\Rightarrow$  **Summary Statistics Wizard**).
3. На шаге 1 проверьте данные и при необходимости примените фильтр. Примите значения параметров, заданные по умолчанию, и нажмите кнопку Вперед (**Next**).
4. На шаге 2 настраиваются роли для переменных в задаче. Перетащите переменную **Profit** в область Сводная статистика (Summary statistics of). В область Для каждого значения (For each value of) сначала перетащите переменную **Product\_Line**, а затем -- переменную **Product\_Category**. Нажмите кнопку Вперед (**Next**).



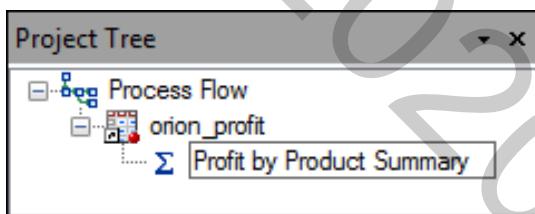
5. На шаге 3 выберите статистики, которые требуется включить в отчет. Нажмите кнопку **Изменить (Edit)**.
  - a. На вкладке Основные (Basic) снимите флагшки **Стандартное отклонение (Standard deviation)**, **Минимум (Minimum)** и **Максимум (Maximum)**.
  - b. Нажмите кнопку **Сумма (Sum)**.
  - c. Измените число десятичных разрядов на **2**.
  - d. Откройте вкладку **Процентиль (Percentile)** и выберите **Медиана (Median)**  $\Rightarrow$  **OK**.
6. Установите флагок **Сохранить статистику в набор данных (Save statistics to data set)** и нажмите кнопку **Обзор (Browse)**.

- a. В поле **Имя файла (File name)** введите **profit\_summary** и нажмите кнопку Сохранить (Save). Нажмите кнопку Вперед (Next).
7. На шаге 4 удалите название анализа, используемое по умолчанию, и введите **Summary of Profit by Product Line/Category**. Нажмите кнопку Готово (Finish).
8. Обратите внимание на то, что в результаты задачи входит несколько вкладок, включая вкладки Результаты (Results) и Выходные данные (Output Data).

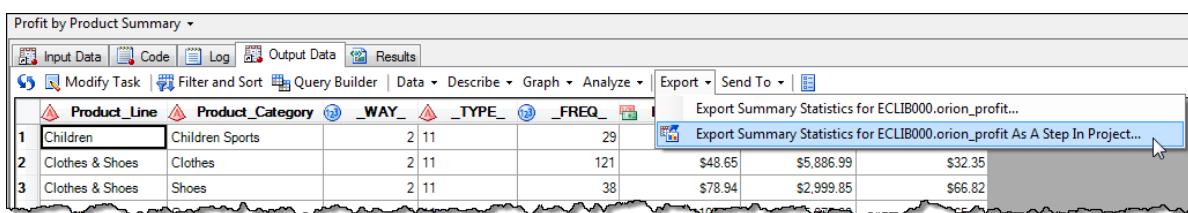
Фрагмент результатов

<b>Summary of Profit by Product Line/Category</b>				
<b>Analysis Variable : Profit</b>				
<b>Product Line</b>	<b>Product Category</b>	<b>Mean</b>	<b>Sum</b>	<b>Median</b>
Children	Children Sports	30.93	896.90	24.10
Clothes & Shoes	Clothes	48.65	5886.99	32.35
	Shoes	78.94	2999.85	66.82
Outdoors	Outdoors	107.99	15875.00	65.05
Sports	Assorted Sports Articles	95.54	7547.35	66.90
	Golf	131.39	9722.83	85.95
	Indoor Sports	213.29	3412.65	41.93
	Racket Sports	90.04	2070.85	30.40
	Running	64.19	1925.55	44.05

9. Задайте более понятное имя для задачи.
- a. Щелкните правой кнопкой мыши значок созданной вами задачи **Сводная статистика (Summary Statistics)** в дереве проекта или окне схемы процесса «Process Flow» и выберите Переименовать Rename.
- b. Введите **Profit by Product Summary**.



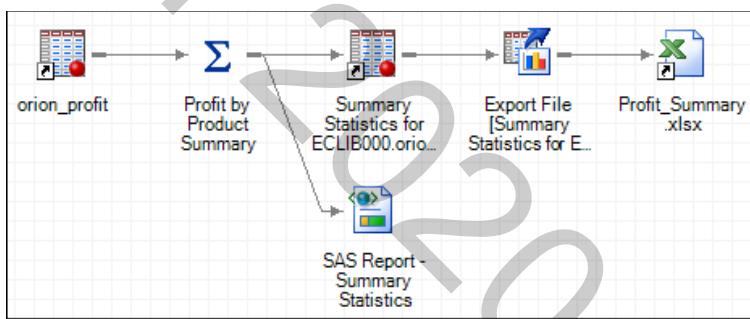
10. Чтобы автоматически экспорттировать выходные данные в Excel, сначала откройте вкладку **Output Data**. Выберите Экспорт ⇔ Экспортировать как шаг в проекте (Export ⇔ Export Summary Statistics for ECLIB000.ORION\_PROFIT As A Step In Project).



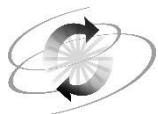
11. На шаге 1 мастера экспорта убедитесь, что выбрана таблица **Summary Statistics for <библиотека>.ORION\_PROFIT**. Нажмите кнопку Вперед (Next).
12. На шаге 2 выберите **Microsoft Excel Workbooks (\*.xlsx)** в качестве типа выходного файла. Нажмите кнопку Вперед (Next).
13. На шаге 3 установите флажок **Использовать ярлыки для имен столбцов (Use labels for column names)**. Нажмите кнопку Вперед (Next).
14. На шаге 4 измените имя выходного файла.
  - a. Нажмите кнопку Обзор (Browse).
  - b. Перейдите к расположению, указанному преподавателем, и введите **Profit\_Summary** в поле Имя файла (File name).
  - c. Нажмите кнопку Сохранить (Save).

 Если флажок **Перезаписать существующий выходной файл (Overwrite existing output)** снят и файл с таким именем уже существует, создается новое имя файла путем добавления текущей даты и времени к существующему имени.
15. Нажмите кнопку Вперед (Next), чтобы проверить параметры экспорта, а затем нажмите кнопку Готово (Finish). Файл **Profit\_Summary.xlsx** будет экспортирован, а в проект будет добавлена задача «Экспорт файла» (Export File).
 

 При повторном выполнении проекта запускаются задачи «Сводная статистика» (Summary Statistics) и «Экспорт файла» для обновления всех результатов.



16. Сохраните проект как **Chapter5**.



## Упражнения

---

### Уровень 1

#### 1. Создание сводного отчета по заработной плате по отделу

С помощью мастера сводной статистики создайте отчет, анализирующий заработную плату всех сотрудников по переменной **Department**.

- a. Создайте новый проект или воспользуйтесь проектом **Chapter5**. Добавьте в него таблицу **employee\_master**.
- b. С помощью мастера сводной статистики создайте отчет для столбца **Salary** по переменной **Department**.
- c. Включите значения средней, минимальной, максимальной заработной платы и медианы для каждой группы и округлите их до ближайшего целого числа.
- d. Создайте набор данных SAS с именем **salary\_summary**, включающий вычисленную статистику.
- e. Введите название **Summary Statistics for Salary by Department** и удалите нижний колонтитул.
- f. Выполните задачу и просмотрите отчет. Переименуйте задачу **Salary by Dept Summary**.
- g. Создайте отдельный узел в проекте, чтобы экспортить результаты в файл Excel с именем **SalaryStats.xlsx**.
- h. Сохраните проект как **Chapter5**.

Фрагмент результатов

Summary Statistics for Salary by Department				
Analysis Variable : Salary				
Department	Mean	Minimum	Maximum	Median
Accounts	41331	26345	58200	40015
Accounts Management	44041	32965	53400	45100
Administration	29919	25020	48600	27290
Concession Management	34320	26840	48435	32560
Engineering	30698	26140	48075	28510
Executives	288333	207885	433800	255823
Group Financials	37510	37510	37510	37510

Клабуков Максим  
Август 2020

## Создание Сводного отчёта (Summary Report)

**Цели**

- Узнать возможности задачи Сводный отчёт.
- Научиться настраивать разметку табличного отчёта и вычисляемые статистики.
- Научиться задавать заголовки для столбцов и строк.
- Изучить, как задать ярлык для пропущенных значений и как добавить заголовки в отчёт.
- Создать сводный табличный отчёт.

12

**Бизнес-сценарий**

Для лучшего понимания поведения покупателей, отделу маркетинга компании Orion требуется отчёт, где собраны данные о прибыли по каждой возрастной группе и категории продуктов.

13

Sas THE WISE CHOICE TO KNOW.

## Сводные таблицы

Мастер сводных таблиц (Summary Tables Wizard) или одноимённую задачу можно использовать для настройки и создания табличного отчёта.

		Profit by Customer Age Group				Total
		15-30 years	31-45 years	46-60 years	61-75 years	
Children	Children Sports	\$419	\$338	\$115	\$25	\$897
	Product Line Subtotal	<b>\$419</b>	<b>\$338</b>	<b>\$115</b>	<b>\$25</b>	<b>\$897</b>
Clothes & Shoes	Clothes	\$1,002	\$2,463	\$1,344	\$1,078	\$5,887
	Shoes	\$578	\$567	\$964	\$891	\$3,000
	Product Line Subtotal	<b>\$1,581</b>	<b>\$3,030</b>	<b>\$2,308</b>	<b>\$1,969</b>	<b>\$8,887</b>
Outdoors	Outdoors	\$3,385	\$6,204	\$1,111	\$5,175	\$15,875
	Product Line Subtotal	<b>\$3,385</b>	<b>\$6,204</b>	<b>\$1,111</b>	<b>\$5,175</b>	<b>\$15,875</b>
Sports	Assorted Sports Articles	\$3,638	\$2,153	\$748	\$1,008	\$7,547
	Golf	\$4,784	\$3,150	\$1,027	\$762	\$9,723
	Indoor Sports	\$1,697	\$750	**	\$966	\$3,413
	Racket Sports	\$81	\$627	\$1,016	\$347	\$2,071
	Running - Jogging	\$309	\$1,216	\$232	\$168	\$1,926
	Swim Sports	\$20	\$49	\$162	**	\$231
	Team Sports	\$406	\$191	\$212	\$279	\$1,088
	Winter Sports	\$319	\$587	\$1,416	\$190	\$2,511
	Product Line Subtotal	<b>\$11,264</b>	<b>\$8,723</b>	<b>\$4,813</b>	<b>\$3,720</b>	<b>\$28,509</b>
	<b>Total</b>	<b>\$16,638</b>	<b>\$18,294</b>	<b>\$8,347</b>	<b>\$10,889</b>	<b>\$54,168</b>

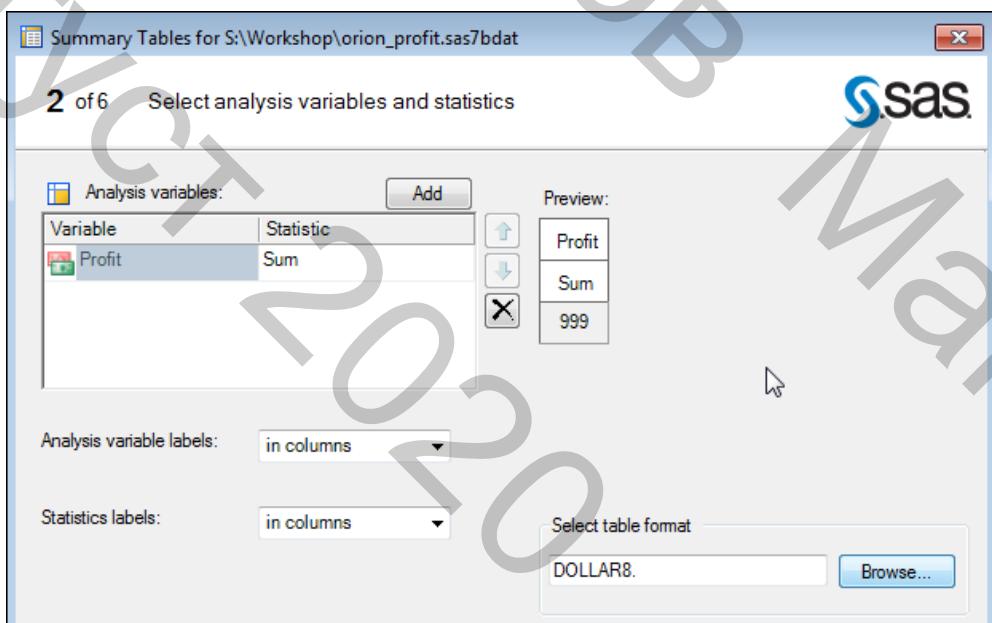
14



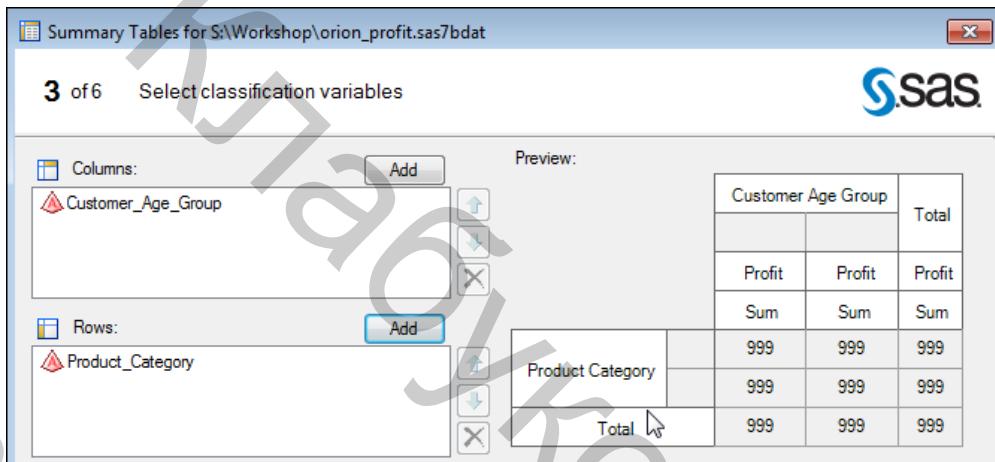
## Создание сводного табличного отчета

Используя таблицу **orion\_profit**, создайте сводный табличный отчет для отображения суммы прибыли по каждой категории продуктов и возрастной группе клиентов.

1. В проекте **Chapter5** дважды щелкните таблицу **orion\_profit**, чтобы открыть ее в сетке данных.
2. На панели инструментов выберите Описательная статистика  $\Rightarrow$  **Мастер сводных таблиц** (**Describe**  $\Rightarrow$  **Summary Tables Wizard**).
3. На шаге 1 убедитесь, что активным источником данных является таблица **orion\_profit**. Нажмите кнопку Вперед (Next).
4. На шаге 2 назначаются переменные анализа и статистика. Чтобы вычислить сумму прибыли, выберите **Добавить (Add)**  $\Rightarrow$  **Profit**.
5. Чтобы применить формат для отображения значений в таблице в виде денежных единиц, нажмите кнопку Обзор (Browse). В области Категории (Categories) выберите Валюта (Currency), а в области Форматы (Formats) — **DOLLARw.d**. Для общей ширины выберите **8**, а для числа десятичных разрядов — **0**. Нажмите кнопки **OK**  $\Rightarrow$  Вперед (Next).



6. На шаге 3 нужно назначить классифицирующие переменные для определения строк и столбцов в таблице. Выберите **Add (Добавить)**  $\Rightarrow$  **Customer\_Age\_Group** в области Столбцы (Columns) и **Add (Добавить)**  $\Rightarrow$  **Product\_Category** в области Строки (Rows). Нажмите кнопку Вперед (Next).

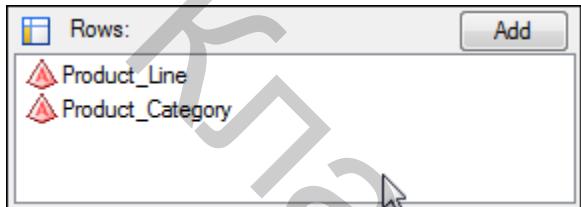


7. На шаге 4 убедитесь, что для строк и столбцов включены итоговые значения. Нажмите кнопку Вперед (Next).
8. Дважды нажмите кнопку Вперед (Next), чтобы выполнить два заключительных шага, а затем нажмите кнопку Готово (Finish), чтобы просмотреть промежуточные результаты.

Product Category	Customer Age Group					Total
	15-30 years	31-45 years	46-60 years	61-75 years		
	Profit	Profit	Profit	Profit		
	Sum	Sum	Sum	Sum		
Assorted Sports Articles	\$3,638	\$2,153	\$748	\$1,008	\$7,547	
Children Sports	\$419	\$338	\$115	\$25	\$897	
Clothes	\$1,002	\$2,463	\$1,344	\$1,078	\$5,887	
Golf	\$4,784	\$3,150	\$1,027	\$762	\$9,723	
Indoor Sports	\$1,697	\$750	-	\$966	\$3,413	
Outdoors	\$3,385	\$6,204	\$1,111	\$5,175	\$15,875	

9. Для улучшения вида отчета в мастере сводных таблиц можно внести дополнительные изменения. На вкладке Результаты (Results) нажмите кнопку Изменить задачу (Modify Task), чтобы повторно открыть мастер.
10. Чтобы удалить заголовки «Profit» и «Sum», нажмите кнопку Вперед (Next) для перехода к шагу 2. Измените значение в полях Ярлыки анализируемых переменных (Analysis variable labels) и Ярлыки статистики (Statistics labels) на Скрыто (hidden).

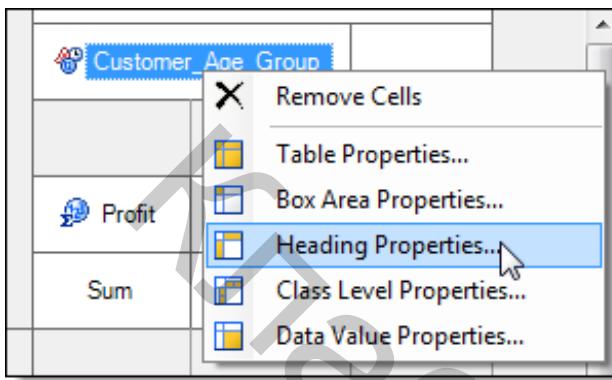
11. Нажмите кнопку Вперед (Next) для перехода к шагу 3. Чтобы добавить ещё одну переменную для разбивки отчёта, выберите Add  $\Rightarrow$  Product\_Line в поле «Rows». Выделите Product\_Line и нажмите кнопку  , чтобы переместить строку вверх списка.



12. Чтобы удалить ярлыки для строк Product\_Category, Product\_Line, а также дополнительный пробел в первой строке таблицы, нажмите кнопку Дополнительные параметры (More Options). Установите флагок Скрыть заголовки строк (Hide the row headers) и нажмите кнопки OK  $\Rightarrow$  Вперед (Next).
13. Чтобы добавить промежуточные суммы для каждого значения Product\_Line, укажите в поле Rows (Строки) значение Итоги по каждому уровню (Totals at each level).
14. Нажмите кнопку Готово (Finish). При запросе на замену результатов нажмите кнопку Да (Yes).

		Customer Age Group				Total
		15-30 years	31-45 years	46-60 years	61-75 years	
Children	Children Sports	\$419	\$338	\$115	\$25	\$897
	Total	\$419	\$338	\$115	\$25	\$897
	Clothes	\$1,002	\$2,463	\$1,344	\$1,078	\$5,887
Clothes & Shoes	Shoes	\$578	\$567	\$964	\$891	\$3,000
	Total	\$1,581	\$3,030	\$2,308	\$1,969	\$8,887
	Outdoors	\$3,385	\$6,204	\$1,111	\$5,175	\$15,875
Outdoors	Total	\$3,385	\$6,204	\$1,111	\$5,175	\$15,875
	Assorted Sports Articles	\$3,638	\$2,153	\$748	\$1,008	\$7,547
	Golf	\$4,784	\$3,150	\$1,027	\$762	\$9,723

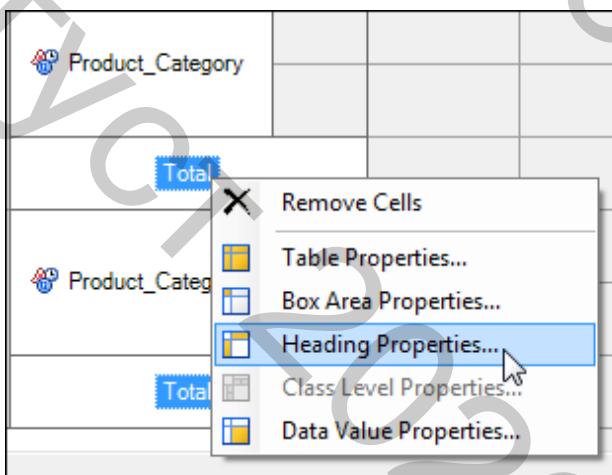
15. Чтобы еще больше изменить внешний вид таблицы, необходимо воспользоваться расширенным представлением задачи «Сводные таблицы» (Summary Tables). Правой кнопкой мыши щелкните значок созданной вами задачи Summary Tables в дереве проекта или окне схемы процесса «Process Flow» и в контекстном меню выберите Открыть  $\Rightarrow$  Открыть в режиме расширенных настроек (Open  $\Rightarrow$  Open in Advanced View).
16. В области выбора выберите Сводные таблицы (Summary Tables). Чтобы изменить заголовок столбца Customer\_Age\_Group, щелкните правой кнопкой мыши Customer\_Age\_Group в области Предварительный просмотр (Preview) и выберите Свойства заголовка (Heading Properties).



17. В диалоговом окне Свойства заголовка (Heading Properties) в поле Ярлык (Label) введите **Profit by Customer Age Group**. Нажмите кнопку OK.

Полное удаление ярлыка приводит к удалению поля из отчета.

18. Измените заголовок столбца для промежуточных итогов **Product\_Line**, щелкнув правой кнопкой мыши на элемент Итог (Total) и выбрав Свойства заголовка (Heading Properties).

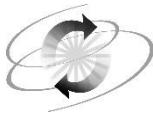


19. В поле Ярлык (**Label**) введите **Product Line Subtotal**. Нажмите кнопку **OK**.
20. Чтобы выделить промежуточные итоги другим цветом фона и шрифтом, правой кнопкой мыши щелкните элемент Итог (**Total**) и выберите **Свойства значений данных (Data Value Properties)**.
21. В появившемся окне откройте вкладку Шрифт (**Font**). Выберите шрифт **Bold Italic** и желтый цвет фона. Нажмите кнопку **OK**.
22. Чтобы указать ярлыки для пропущенных значений, щелкните правой кнопкой мыши в любом месте таблицы и выберите **Свойства таблицы (Table Properties)**.
23. В появившемся окне откройте вкладку Общие (**General**). В поле **Ярлык для отсутствующих значений (Label for missing values)** удалите ярлык, используемый по умолчанию, и введите **\*\***. Нажмите кнопку **OK**.
24. Нажмите кнопку Пуск (**Run**), чтобы создать окончательный отчет. При запросе на замену результатов нажмите кнопку Да (**Yes**).

		<b>Profit by Customer Age Group</b>				<b>Total</b>
		<b>15-30 years</b>	<b>31-45 years</b>	<b>46-60 years</b>	<b>61-75 years</b>	
<b>Children</b>	<b>Children Sports</b>	\$419	\$338	\$115	\$25	\$897
	<b>Product Line Subtotal</b>	<b>\$419</b>	<b>\$338</b>	<b>\$115</b>	<b>\$25</b>	<b>\$897</b>
<b>Clothes &amp; Shoes</b>	<b>Clothes</b>	\$1,002	\$2,463	\$1,344	\$1,078	\$5,887
	<b>Shoes</b>	\$578	\$567	\$964	\$891	\$3,000
<b>Outdoors</b>	<b>Product Line Subtotal</b>	<b>\$1,581</b>	<b>\$3,030</b>	<b>\$2,308</b>	<b>\$1,969</b>	<b>\$8,887</b>
	<b>Outdoors</b>	\$3,385	\$6,204	\$1,111	\$5,175	\$15,875
<b>Sports</b>	<b>Product Line Subtotal</b>	<b>\$3,385</b>	<b>\$6,204</b>	<b>\$1,111</b>	<b>\$5,175</b>	<b>\$15,875</b>
	<b>Assorted Sports Articles</b>	\$3,638	\$2,153	\$748	\$1,008	\$7,547
	<b>Golf</b>	\$4,784	\$3,150	\$1,027	\$762	\$9,723
	<b>Total Sports</b>	\$1,697	\$750	**	\$966	\$2413

25. Сохраните проект **Chapter5**.

## Упражнения



### Уровень 1

#### 4. Создание сводного табличного отчета с несколькими статистиками

С помощью задачи «Сводные таблицы» создайте отчет, вычисляющей минимальную и максимальную заработную плату для каждой страны и отдела.

- В проекте **Chapter5** откройте таблицу **employee\_master** и Мастер настроек сводной таблицы.
- Два раза добавьте переменную **Salary** в список Анализируемые переменные (Analysis variables). Выберите Минимум (**Minimum**) в первой строке и **Максимум (максимум)** — во второй.
- Скройте ярлыки анализируемых переменных и примените к результатам в таблице формат **DOLLARw.d** со значением **10** для общей ширины и **0** для количества десятичных разрядов.
- Добавьте **Country** в роль Столбцы (Columns) и **Department** в область Строки (Rows). Скройте заголовки строк.
- Исключите из отчёта итоговую строку и столбец.
- Создайте отчет и проверьте результаты.

	Country			
	AU		US	
	Minimum	Maximum	Minimum	Maximum
Accounts	-	-	\$26,545	\$58,200
Accounts Management	-	-	\$32,965	\$53,400
Administration	\$26,495	\$46,230	\$25,020	\$48,600
Concession Management	-	-	\$26,840	\$48,435
Engineering	\$27,645	\$31,670	\$26,140	\$48,075
Executives	-	-	\$207,885	\$433,800
Group Financials	-	-	\$37,510	\$37,510

- Измените настройки созданной вами задачи «Сводные таблицы», открыв ее в расширенном представлении. Измените свойства заголовков для столбцов **Minimum** и **Maximum** так, чтобы в них отображались значения **Low** и **High**.
- Измените свойства заголовка столбцов (задаваемых переменной **Country**) так, чтобы там отображался следующий текст: **Salary Ranges by Country**.
- Измените свойства таблицы так, чтобы пропущенные значения выводились в виде дефиса.
- Переименуйте задачу в **Salary Ranges by Country**.
- Создайте отчет и сохраните проект **Chapter5**.

	Salary Ranges by Country			
	AU		US	
	Low	High	Low	High
Accounts	-	-	\$26,545	\$58,200
Accounts Management	-	-	\$32,965	\$53,400
Administration	\$26,495	\$46,230	\$25,020	\$48,600
Concession Management	-	-	\$26,840	\$48,435
Engineering	\$27,645	\$31,670	\$26,140	\$48,075
Executives	-	-	\$207,885	\$433,800
Group Financials	-	-	\$37,510	\$37,510
Group HR Management	-	-	\$27,365	\$63,915

## День 2. SAS(R) Base: Essentials



### Выполнение программы SAS в SAS Studio



В демонстрации используется SAS Studio 3.3.

#### Выполнение программы `cre8data`

1. Запустите SAS Studio. Главное окно SAS Studio состоит из панели навигации слева и рабочей области справа.
2. В области Folders (Папки) панели навигации перейдите в указанное место файловой системы, где находится программа `cre8data`.
3. Двойным щелчком откройте программу `cre8data`. Вместо двойного щелчка вы можете мышью перетащить эту программу в рабочую область.
4. На вкладке CODE (Код) программы `cre8data` найдите оператор %LET.
5. Если файлы данных нужно создать в директории, отличной от `s:\workshop`, измените значение, присваиваемое макропеременной `path`, указав в ней требуемую директорию. Если файлы данных для курса создаются в директории `s:\workshop`, никаких изменений не нужно.



В программе `cre8data` для указания путей используется знак "прямого слэша" (/), чтобы обеспечить переносимость между различными операционными средами. В UNIX и Linux для этого требуется прямой слэш. Windows принимает прямой слэш и может преобразовывать его в "обратный слэш" (\).

6. Для выполнения программы нажмите  (Run, Выполнить) на вкладке CODE (Код) или клавишу F3.
7. На вкладке RESULTS (Результаты) убедитесь, что вывод содержит список файлов данных.

#### Выполнение программы `p102d01`

1. В области Folders панели навигации перейдите в место расположения программы `p102d01`.
2. Двойным щелчком откройте программу `p102d01`. Вместо двойного щелчка вы можете мышью перетащить эту программу в рабочую область.
3. Для выполнения программы нажмите  (Run, Выполнить) на вкладке CODE (Код) или клавишу F3.
4. На вкладке RESULTS (Результаты) просмотрите вывод PROC PRINT и PROC MEANS.
5. Откройте вкладку LOG (Журнал) и проверьте, содержит ли она предупреждения (warning) или ошибки (error).



## Выполнение программы SAS в SAS Enterprise Guide



В демонстрации используется SAS Enterprise Guide 7.1.

### Выполнение программы `cre8data`

1. Откройте SAS Enterprise Guide. Закройте диалоговое окно с приветствием, нажав на кнопку **X** в правом верхнем углу окна. По умолчанию SAS Enterprise Guide содержит окна Project Tree и Servers. Они расположены слева. Окно Process Flow (Схема процесса) находится в правой части интерфейса.
2. Выберите **File**  $\Rightarrow$  **Open**  $\Rightarrow$  **Program** (**Файл**  $\Rightarrow$  **Открыть**  $\Rightarrow$  **Программа**) или щёлкните  (**Open, Открыть**)  $\Rightarrow$  **Program** (**Программа**).
3. В окне Open Program (Открыть программу) перейдите в место расположения программы **cre8data** и нажмите кнопку **Open** (**Открыть**).
4. На вкладке Program (Программа) программы **cre8data** найдите оператор %LET.
5. Если файлы данных нужно создать в директории, отличной от **s:\workshop**, измените значение, присваиваемое макропеременной **path**, указав в ней требуемую директорию. Если файлы данных для курса создаются в директории **s:\workshop**, никаких изменений не нужно.



В программе **cre8data** для указания путей используется знак "прямого слэша" (/), чтобы обеспечить переносимость между различными операционными средами. В UNIX и Linux для этого требуется прямой слэш. Windows принимает прямой слэш и может преобразовывать его в "обратный слэш" (\).

6. Для выполнения программы нажмите **Run** (**Выполнить**) на вкладке Program (Программа) или клавишу F3.
7. На вкладке Results (Результаты) проверьте, что вывод содержит список файлов данных.

### Выполнение программы `p102d01`

1. Выберите **File**  $\Rightarrow$  **Open**  $\Rightarrow$  **Program** (**Файл**  $\Rightarrow$  **Открыть**  $\Rightarrow$  **Программа**) или щёлкните  (**Open, Открыть**)  $\Rightarrow$  **Program** (**Программа**).
2. В окне Open Program (Открыть программу) перейдите в место расположения программы **p102d01** и нажмите кнопку **Open** (**Открыть**).
3. Чтобы выполнить программу, нажмите **Run** (**Выполнить**) на вкладке Program (Программа) или клавишу F3.
4. На вкладке Results (Результаты) просмотрите вывод PROC PRINT и PROC MEANS.
5. Откройте вкладку **Log** (**Журнал**) и проверьте, содержит ли она предупреждения (warning) или ошибки (error).

## **Выполнение программы SAS в оконном интерфейсе SAS**



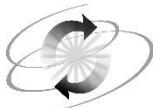
В демонстрации используется SAS 9.4M2.

### **Выполнение программы `cre8data`**

3. Запустите оконный интерфейс SAS. Если вы видите диалог, содержащий описание изменений или советы по работе с SAS, нажмите **Close**. По умолчанию оконный интерфейс SAS состоит из левой панели, содержащей окна Results и Explorer, и правой панели с окнами Output, Log и Editor.
  4. Нажмите  (**Open**) при активном окне Editor или выберите в меню **File** ⇒ **Open Program**.
  5. В окне Open перейдите в место расположения программы **cre8data** и нажмите кнопку **Open**.
  6. В окне Editor программы **cre8data** найдите оператор %LET.
  7. Если файлы данных нужно создать в директории, отличной от **s:\workshop**, измените значение, присваиваемое макропеременной **path**, указав в ней требуемую директорию. Если файлы данных для курса создаются в директории **s:\workshop**, никаких изменений не нужно.
-  В программе **cre8data** для указания путей используется знак "прямого слэша" (/), чтобы обеспечить переносимость между различными операционными средами. В UNIX и Linux для этого требуется прямой слэш. Windows принимает прямой слэш и может преобразовывать его в "обратный слэш" (\).
8. Запустите программу, нажмите  (**Submit**) или клавишу F3.
  9. В окне Results Viewer проверьте, что вывод содержит список файлов данных.
  10. Для очистки всего содержимого окна журнала (Log), откройте это окно и нажмите  (**New**) или откройте его и выберите в меню **Edit** ⇒ **Clear All**.

### **Выполнение программы `p102d01`**

1. Нажмите  (**Open**) при активном окне Editor или выберите в меню **File** ⇒ **Open Program**.
2. В окне Open перейдите в место расположения программы **p102d01** и нажмите **Open**.
3. Запустите программу, нажмите  (**Submit**) или клавишу F3.
4. В окне Results Viewer просмотрите вывод PROC PRINT и PROC MEANS.
5. Перейдите в окно Log и проверьте, содержит ли оно предупреждения (warning) или ошибки (error).



## Упражнения. Создание данных курса.



Для выполнения упражнений в этой главе просто следуйте инструкциям.

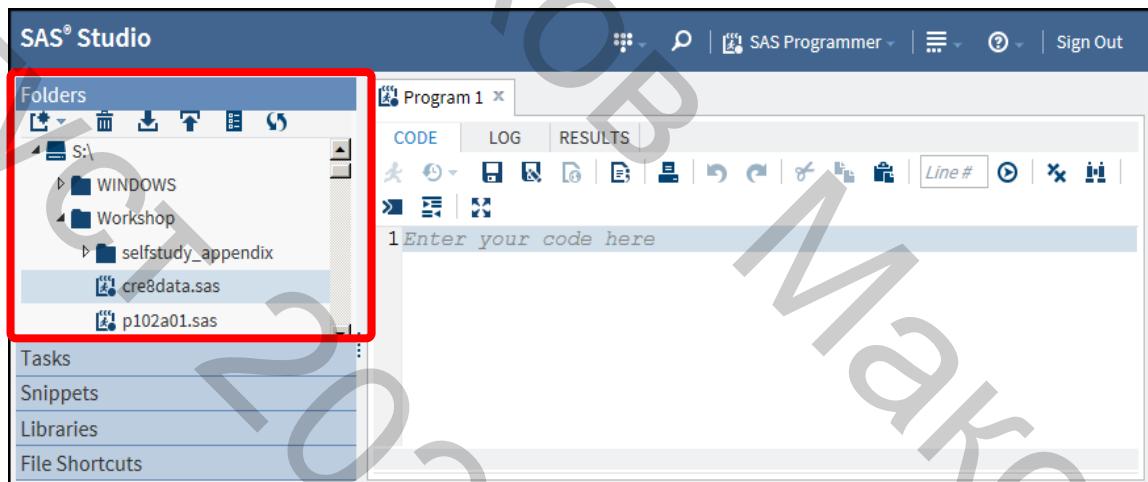
### SAS Studio



**Вы должны** выполнить это упражнение для создания файлов с данными курса. Если вы не создадите файлы с данными, то большинство программ в этом курсе не будет работать.

#### Создание данных курса в SAS Studio

- Запустите SAS Studio. В основном окне SAS Studio находятся панель навигации (слева) и рабочая область (справа).
- В области Folders (Папки) на панели навигации перейдите в указанное место файловой системы, где находится программа **cre8data**.



- Двойным щелчком откройте программу **cre8data**. Вместо двойного щелчка вы можете мышью перетащить эту программу в рабочую область.
- На вкладке CODE (Код) программы **cre8data** найдите оператор %LET, показанный ниже.

```

15
16 %let path=s:/workshop;
17
18

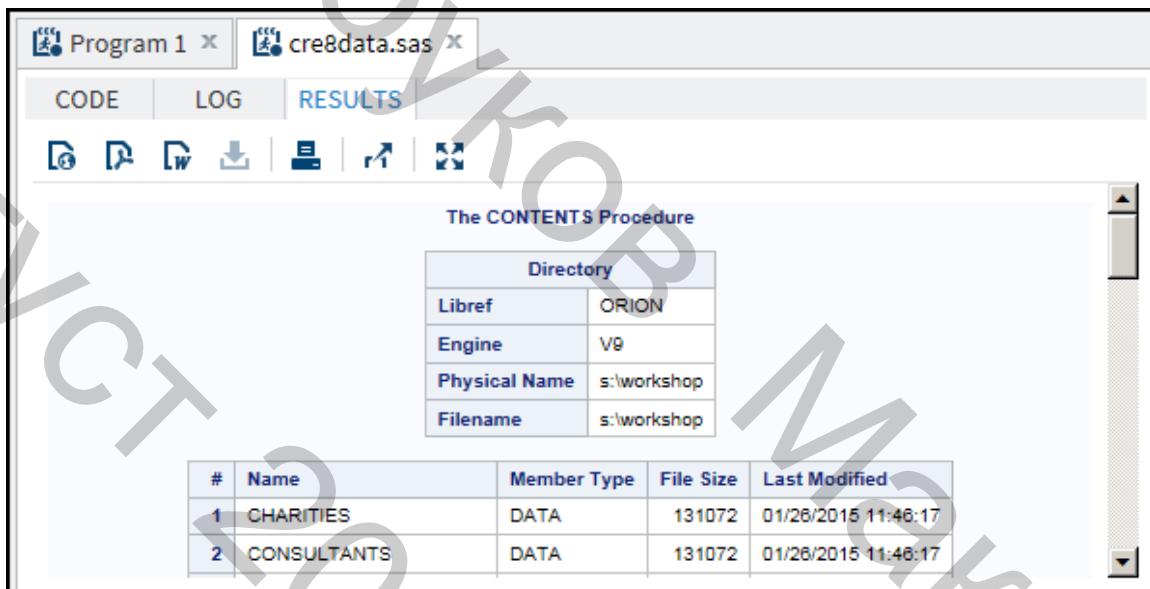
```

Line 20, Column 21

- e. Если файлы данных нужно создать в директории, отличной от **s:\workshop**, измените значение, присваиваемое макропеременной **path**, указав в ней требуемую директорию. Если файлы данных для курса создаются в директории **s:\workshop**, никаких изменений не нужно.

 В программе **cre8data** для указания путей используется знак "прямого слэша" (/), чтобы обеспечить переносимость между различными операционными средами. В UNIX и Linux для этого требуется прямой слэш. Windows принимает прямой слэш и может преобразовывать его в "обратный слэш" (\).

- f. На вкладке CODE (Код) нажмите  (Run, Выполнить) или нажмите F3.  
g. На вкладке RESULTS (Результаты) проверьте, что вывод содержит список файлов данных.



#	Name	Member Type	File Size	Last Modified
1	CHARITIES	DATA	131072	01/26/2015 11:46:17
2	CONSULTANTS	DATA	131072	01/26/2015 11:46:17

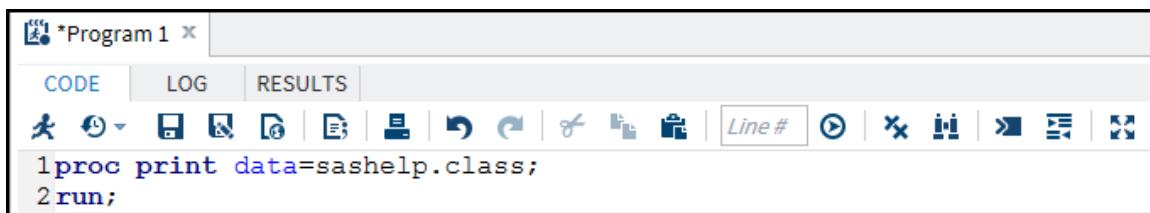
- h. Закройте вкладку с программой **cre8data**, нажав на значок X справа от названия.



## Знакомство с SAS Studio

### Основные вкладки интерфейса

- a. На вкладке CODE (Код) вкладки Program 1 введите шаг PROC PRINT, показанный ниже.



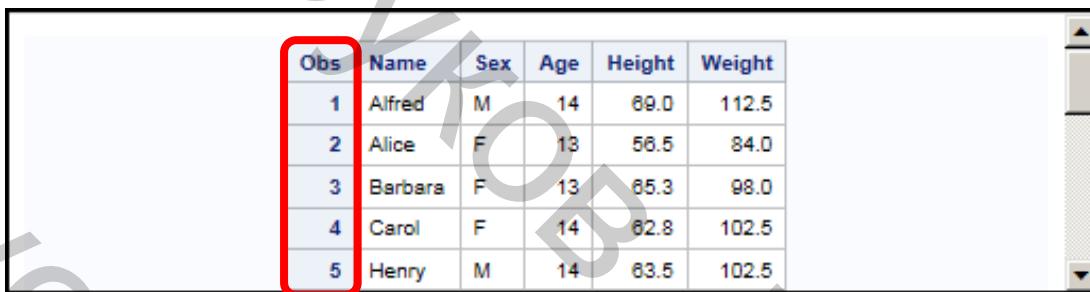
```
1proc print data=sashelp.class;
2run;
```

b. Выполните программу, нажав  (Run, Выполнить) на вкладке CODE (Код) или клавишу F3. Если при запуске программы не возникает ошибок, автоматически открывается вкладка RESULTS (Результаты), в которой находится вывод PROC PRINT.

- 1) Откройте вкладку **LOG (Журнал)** и проверьте наличие в ней двух примечаний, показанных ниже. Если вы видите какие-либо предупреждения (warning) или ошибки (error), вернитесь на вкладку CODE (Код) для их исправления, а затем повторно запустите программу.

**NOTE: There were 19 observations read from the data set SASHELP.CLASS.**  
**NOTE: PROCEDURE PRINT used (Total process time):**

- 2) Откройте вкладку **RESULTS (Результаты)**. Обратите внимание, что вывод PROC PRINT содержит столбец Obs.



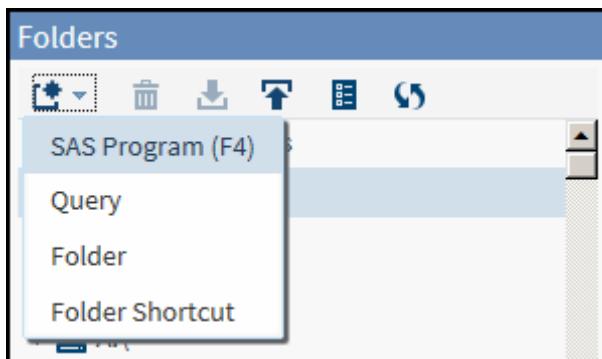
Obs	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69.0	112.5
2	Alice	F	13	56.5	84.0
3	Barbara	F	13	65.3	98.0
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5

- c. Откройте вкладку **CODE** и добавьте параметр NOOBS для того, чтобы убрать столбец Obs.

```
1 proc print data=sashelp.class noobs;
2 run;
```

 Параметр NOOBS нужно указать в операторе PROC PRINT перед точкой с запятой.

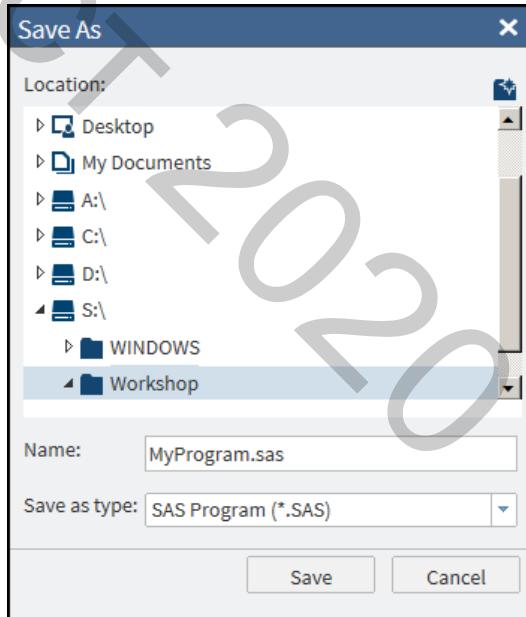
- 1) Для выполнения программы нажмите  (Run, Выполнить) на вкладке CODE (Код) или клавишу F3.
  - 2) Изучите содержимое вкладок RESULTS и LOG после перезапуска программы.
- d. Создайте другую программу. В области Folders (Папки) панели навигации нажмите  (New, Новая) ⇒ SAS Program (F4) (Программа SAS (F4)) или нажмите клавишу F4.



- 1) На вкладке CODE, находящейся внутри вкладки Program 2, введите шаги PROC PRINT и PROC MEANS, показанные ниже.

```
1 proc print data=sashelp.class;
2 run;
3 proc means data=sashelp.class;
4 run;
```

- 2) Для выполнения программы нажмите (Run, Выполнить) на вкладке CODE (Код) или клавишу F3. Если программа выполнится успешно, то автоматически открывается вкладка RESULTS (Результаты), содержащая вывод PROC PRINT и PROC MEANS.
- 3) Откройте вкладку LOG. Проверьте примечания в журнале. Если вы видите какие-либо предупреждения (warning) или ошибки (error), вернитесь на вкладку CODE (Код) для исправления ошибок, и затем повторно запустите программу.
- 4) Сохраните программу, для этого перейдите на вкладку CODE и нажмите (Save As, Сохранить как).
- 5) В окне Save As (Сохранить как) выберите место расположения файла, например, s:\workshop, назовите файл MyProgram и нажмите кнопку Save.

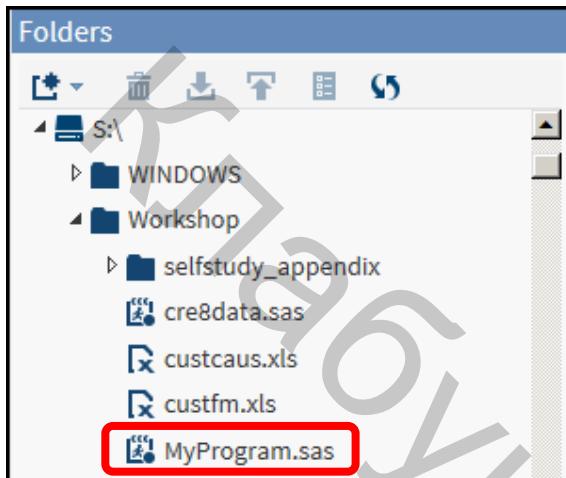


- 6) Обратите внимание, что вкладка Program 2 теперь называется так, как вы назвали вашу программу. Закройте вкладку с программой MyProgram, нажав на значок X рядом с названием.



Звёздочка (\*), выводимая перед названием файла, означает, что программа не была сохранена.

- e. Откройте программу, которую вы сохранили и закрыли. В области Folders (Папки) панели навигации перейдите к месту расположения файла **MyProgram.sas**.



- 1) Дважды щёлкните **MyProgram.sas**, чтобы открыть эту программу. Вместо двойного щелчка вы можете мышью перетащить эту программу в рабочую область.
- 2) Выполните фрагмент программы. На вкладке CODE (Код) выделите шаг PROC MEANS (две строчки кода) и нажмите (Run, Выполнить) или клавишу F3. При этом выполняется только выделенный фрагмент кода.

```

*Program 1 * MyProgram.sas
CODE LOG RESULTS
1 proc print data=sashelp.class;
2 run;
3 proc means data=sashelp.class;
4 run;

```

- 3) Просмотрите вкладки RESULTS и LOG. Обратите внимание, что информация в этих вкладках относится только к шагу PROC MEANS.

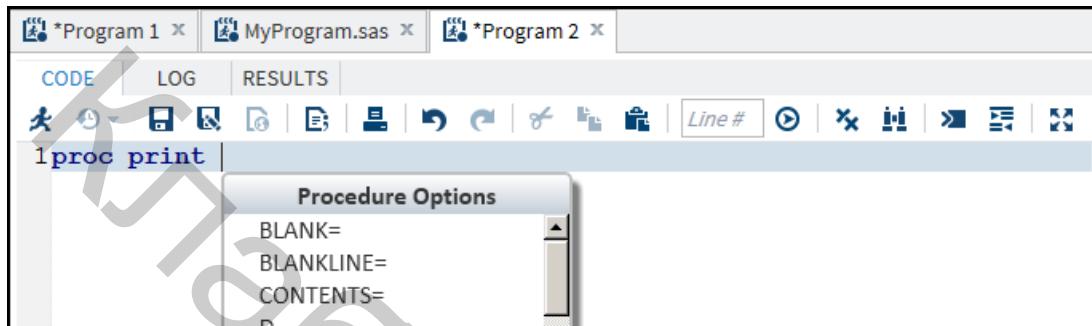
#### Использование функций вкладки CODE (Код)

- a. На панели инструментов вкладки CODE (Код) выберите (Maximize View, Развернуть представление), чтобы увеличить рабочую область и скрыть панель навигации.

После того, как вы развернёте окно до максимального размера, щёлкните (Exit Maximize View, Выйти из развернутого представления), чтобы снова отобразить панель навигации.

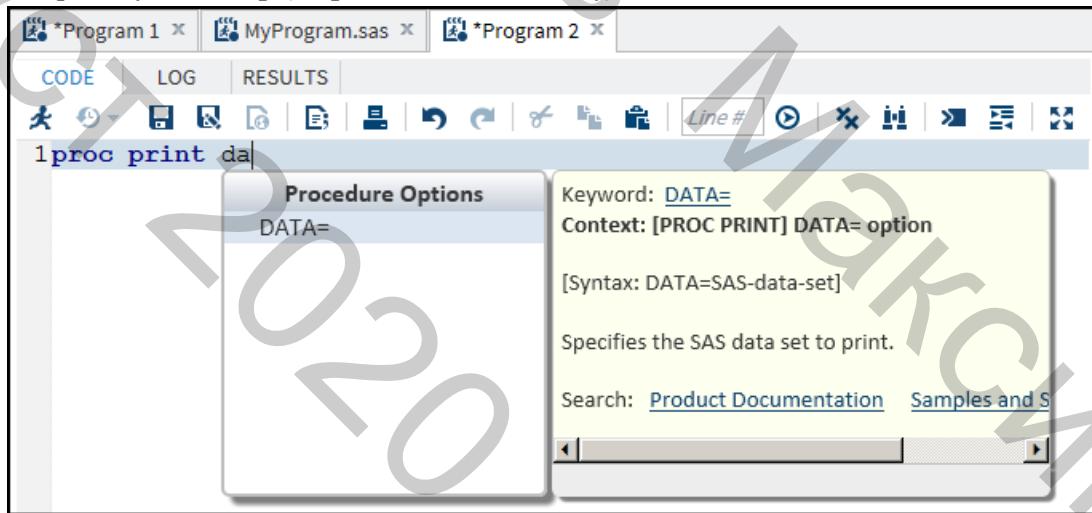
- b. Для создания новой программы, в области Folders (Папки) панели навигации щёлкните (New, Новая) ⇔ SAS Program (F4) (Программа SAS (F4)) или нажмите F4.
- 1) На вкладке CODE (Код) программы Program 2 начните вводить шаг PROC PRINT. Обратите внимание на функцию автозаполнения, которая появляется в процессе ввода текста программы. Функция автозаполнения позволяет вставить в программу возможные ключевые слова, выбрав их во всплывающем окне. Например, после ввода

**proc print** появляется всплывающее окно, содержащее параметры оператора PROC PRINT.



Чтобы перейти к нужному ключевому слову в этом окне, прокрутите их список клавишами "вверх" и "вниз", Page Up или Page Down с клавиатуры или прокрутите его мышью, используя полосу прокрутки. Для добавления выбранного ключевого слова в программу дважды щёлкните его мышью, или нажмите клавишу Enter.

- 2) Начните вводить параметр DATA=. Справка по синтаксису появляется, как только вы уменьшаете список релевантных ключевых слов. Справка по синтаксису также появляется, если щёлкнуть правой кнопкой мыши на ключевое слово в программе, и выбрать Syntax Help (Справка по синтаксису).

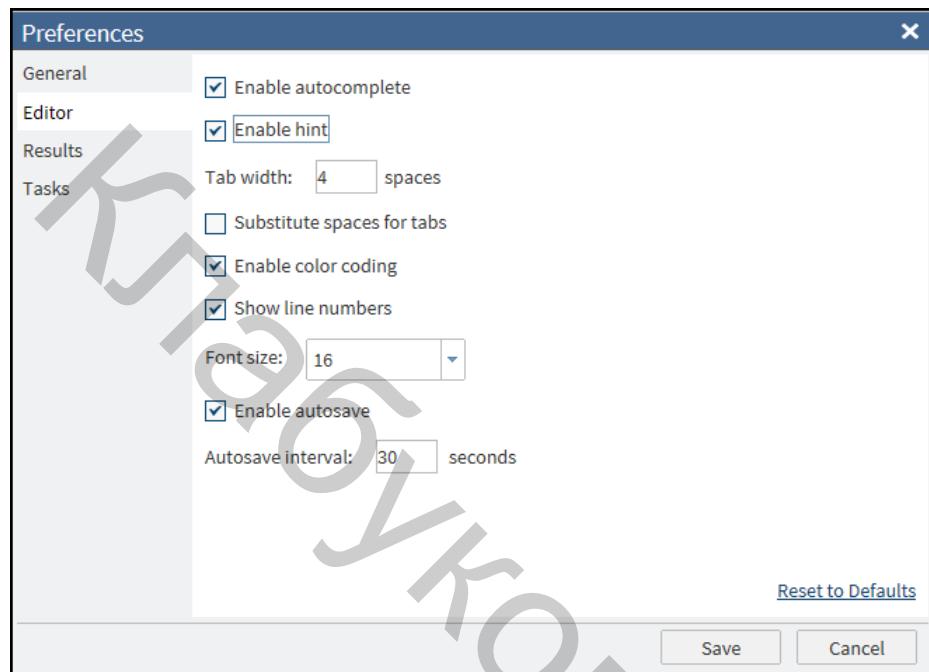


- 3) Закончите шаг PROC PRINT. Обратите внимание на автозаполнение и справку по синтаксису, появляющиеся по мере ввода текста программы.

```
proc print data=sashelp.class;
run;
```

Вы также можете обращаться к справке по синтаксису просто расположив указатель мыши над ключевым словом в вашей программе, если включите эту функцию в настройках. Включите её, нажав (More Application Options, Настройки) на главной панели инструментов.

- 4) Выберите пункт Редактор и изучите список настроек редактора.
- 5) Включите флагок Enable hint (Включить подсказку) и нажмите Save (Сохранить).



- 6) Расположите указатель мыши над словом **print** в вашей программе и просмотрите справку по синтаксису.

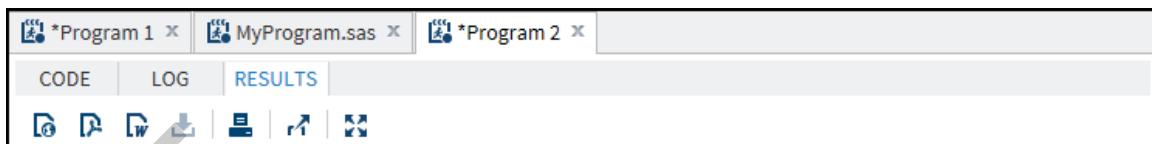
#### Использование функций вкладки RESULTS (Результаты)

- На вкладке CODE программы Program 2 нажмите кнопку (Run, Выполнить) или нажмите F3 для выполнения шага PROC PRINT.  
Если программа успешно выполняется, автоматически открывается вкладка RESULTS (Результаты), в которой находится вывод PROC PRINT.
- По умолчанию результаты на вкладке RESULTS (Результаты) выводятся в формате HTML5. На панели инструментов RESULTS нажмите (Open In A New Browser Tab, Открыть в новой вкладке), чтобы открыть вывод в формате HTML5 в другой вкладке браузера.
- Просмотрите вывод в новой вкладке и закройте её.

Obs	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69.0	112.5
2	Alice	F	13	66.5	84.0
3	Barbara	F	13	65.3	98.0
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5
6	Laura	F	12	65.0	92.5

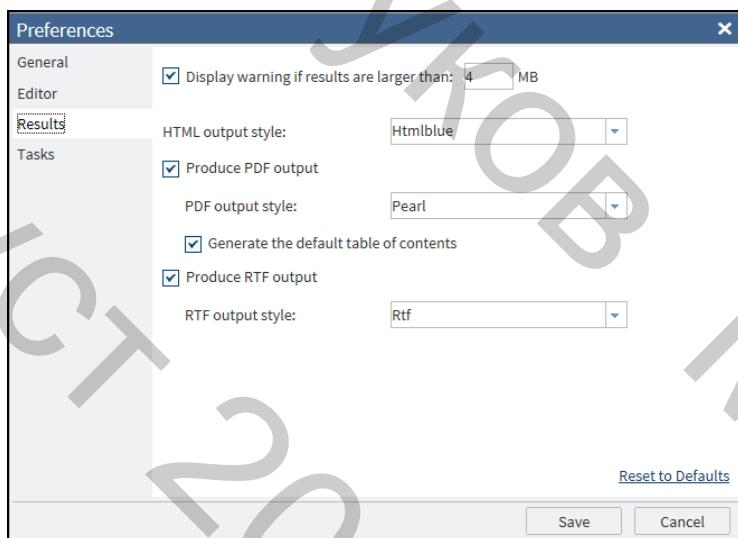


В дополнение к формату HTML5, по умолчанию вывод также создаётся в форматах PDF и RTF. Форматы PDF и RTF не отображаются на вкладке RESULTS (Результаты). Вы можете скачать вывод в форматах HTML5, PDF и RTF, нажав соответствующую кнопку.



- d. Нажмите (Download Results As A PDF File, Загрузить результаты как файл PDF) или (Download Results As An RTF File, Загрузить результаты как файл RTF) для загрузки вывода в форматах PDF или RTF. Следуйте инструкциям, чтобы открыть файл. После просмотра файла закройте его.
- e. Вы можете изменить настройки вывода результатов в окне Preferences (Настройки). Откройте настройки, нажав (More Application Options, Настройки).

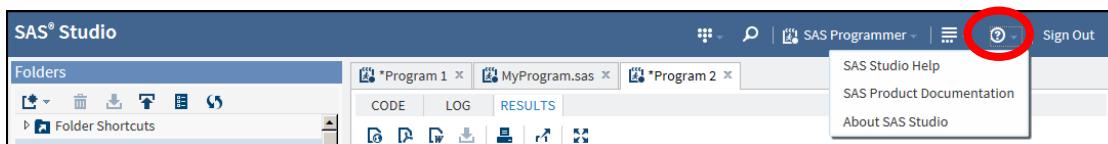
1) Затем выберите пункт Results (Результаты) и изучите список настроек результатов.



2) Нажмите Cancel (Отмена), чтобы закрыть окно настроек.

## Доступ к справке и документации

- a. Вы можете открыть справку (SAS Help) и документацию прямо из главной панели инструментов.



- b. Нажмите ⇒ SAS Studio Help (Справка SAS Studio) для перехода на страницу документации по SAS Studio. Эта веб-страница содержит справку по интерфейсу SAS Studio. Закройте окно после просмотра этой страницы.

The screenshot shows the SAS Customer Support website. At the top, there's a navigation bar with links for 'support.sas.com', 'Knowledge Base' (which is highlighted in orange), 'Support', 'Training & Books', 'Happenings', 'Store', and 'Support Communities'. On the far right of the header are 'Log in', 'Create profile', a search bar with 'Search support.sas.com' and a 'Search' button, and 'Advanced Search'.

The main content area has a title 'SAS CUSTOMER SUPPORT / SAS Products'. To the left, there's a sidebar titled 'KNOWLEDGE BASE' with sections like 'Products & Solutions' (listing SAS 9.4, SAS 9.3, SAS 9.2, Maintenance & Product Upgrades), 'System Requirements', 'Install Center', 'Third-Party Software Reference', 'Documentation', 'Papers', 'Samples & SAS Notes', and 'Focus Areas'. The main content area starts with a section for 'SAS® Studio' with tabs for 'Overview', 'Documentation' (which is selected), 'Training', and 'Samples and Tips'. Below this are links to 'Online Documentation' and 'Technical Papers'. A 'Free Online Documentation' section lists several resources, including the 'Most recent release for SAS Studio' (with links to 'Accessibility Features of SAS Studio 3.3 HTML', 'Getting Started with Programming in SAS Studio 3.3 PDF | HTML', 'SAS Studio 3.3: Administrator's Guide PDF | HTML', 'SAS Studio 3.3: Developer's Guide to Writing Custom Tasks PDF | HTML', 'SAS Studio 3.3: System Requirements PDF', and 'SAS Studio 3.3: User's Guide PDF | HTML'). There are also sections for 'Videos' (linking to 'SAS Studio: Video Library') and 'Programming References' (linking to 'SAS 9.4 Programmer's Bookshelf').

On the right side, there's a 'Stay Informed' box with links to 'SAS Notes for SAS 9 RSS', 'SAS Samples RSS', 'SAS Bookshelf Blog', 'SAS Training and Book Report', and 'SAS Training Blog'.

Внешний адрес на страницу SAS Studio:  
**<http://support.sas.com/software/products/sasstudio/>**

- c. Нажмите ⇒ **SAS Product Documentation (Документация по продуктам SAS)** для перехода на веб-страницу с описанием решений и продуктов SAS. Эта веб-страница содержит материалы о продуктах SAS.

The screenshot shows the SAS Knowledge Base homepage. At the top, there's a navigation bar with links for 'Log in', 'Create profile', 'support.sas.com', 'Knowledge Base' (which is highlighted in orange), 'Support', 'Training & Books', 'Happenings', 'Store', and 'Support Communities'. Below the navigation is a search bar with a 'Search support.sas.com' button and an 'Advanced Search' link. A large banner at the top says 'KNOWLEDGE BASE / SAS PRODUCTS & SOLUTIONS'. On the left, there's a sidebar titled 'KNOWLEDGE BASE' with a list of categories: 'Products & Solutions' (with sub-links for SAS 9.4, 9.3, 9.2, and Maintenance & Product Upgrades), 'System Requirements', 'Install Center', 'Third-Party Software Reference', 'Documentation' (which is highlighted in blue), 'Papers', 'Samples & SAS Notes', and 'Focus Areas'. The main content area has a title 'SAS® Products and Solutions' and a sub-section 'SAS® 9.4'. It describes SAS 9.4 as the second maintenance release, mentioning Analytical Products 13.2, availability starting in August 2014, and a more highly available and resilient platform. There's a link to 'Read more about SAS 9.4 and the Analytical Products'. Below this is a 'SAS Product Listing: Index A-Z' section with a grid of letters from A to W. Under letter A, there's a list of products: SAS/ACCESS, SAS Activity-Based Management, SAS Add-In for Microsoft Office, SAS Anti-Money Laundering, SAS/AF, SAS Analytics Pro, and SAS AppDev Studio. Under letter B, there's a list of products: Base SAS, SAS BI Dashboard, and SAS BI Server. At the bottom right of the main content area is a box titled 'Interested in SAS 9.4?' with instructions to request it via a form or contact the SAS account team.

Внешний адрес на страницу о продуктах SAS: <http://support.sas.com/software/>

- d. На указанной выше странице выберите **Documentation** на панели навигации KNOWLEDGE BASE.

This screenshot shows the 'KNOWLEDGE BASE' sidebar from the previous page. The 'Documentation' link is highlighted with a red rectangle. The other links in the sidebar are: 'Products & Solutions' (with sub-links for SAS 9.4, 9.3, 9.2, and Maintenance & Product Upgrades), 'System Requirements', 'Install Center', 'Third-Party Software Reference', 'Papers', 'Samples & SAS Notes', and 'Focus Areas'.

- e. Откроется веб-страница с документацией о продуктах SAS. Эта веб-страница будет полезна для поиска документации по синтаксису, который используется в различных продуктах SAS.



Внешний адрес страницы документации по продуктам SAS:  
<http://support.sas.com/documentation/>

- f. На этой странице нажмите ссылку **Programmer's Bookshelf**. Затем выберите ссылку в соответствии с используемой версией SAS.

The screenshot shows the SAS Support website with the URL [support.sas.com/documentation/](http://support.sas.com/documentation/). The main navigation bar includes links for Knowledge Base, Support, Training & Books, Happenings, Store, and Support Communities. Below the navigation bar, a sub-menu for 'KNOWLEDGE BASE / DOCUMENTATION' is displayed, with 'SAS Product Documentation' selected. Under 'Starting Points', there are links for 'Product Index A-Z', 'What's New in SAS', 'Programmer's Bookshelf' (which is highlighted with a red box), and 'Documentation by Title'.



Страница Programmer's Bookshelf содержит избранную документацию по продуктам SAS (таким, как Base SAS).

- g. Просмотрите страницу Programmer's Bookshelf. После этого закройте окно.

## SAS Enterprise Guide



Вы **должны** выполнить это упражнение для создания файлов с данными курса. Если вы не создадите файлы с данными, то большинство программ в этом курсе не будет работать.

### Создание данных для курса в SAS Enterprise Guide

- a. Запустите SAS Enterprise Guide. Закройте диалоговое окно с приветствием (Welcome to SAS Enterprise Guide), нажав на значок X в правом углу этого окна. По умолчанию SAS Enterprise Guide содержит окна Project Tree и Servers. Они расположены слева. Окно Process Flow (Схема процесса) находится в правой части интерфейса.
- b. Выберите **File** ⇒ **Open** ⇒ **Program** (**Файл** ⇒ **Открыть** ⇒ **Программа**) или щёлкните (**Open**, **Открыть**) ⇒ **Program** (**Программа**).
- c. В окне Open Program (Открыть программу) перейдите в место расположения программы **cre8data** и нажмите кнопку **Open** (**Открыть**).

File Name
selfstudy_appendix
cre8data.sas
p102a01.sas
p103e02.sas
p103e03.sas
p103e07.sas
p103e08.sas
p104d08.sas
p104d09.sas
p104d10.sas
p104d11.sas
p104d01.sas

- d. На вкладке Program (Программа) программы **cre8data** найдите оператор %LET, показанный ниже.

```
%let path=s:\workshop;
```

- e. Если файлы данных нужно создать в директории, отличной от **s:\workshop**, измените значение, присваиваемое макропеременной **path**, указав в ней требуемую директорию. Если файлы данных для курса создаются в директории **s:\workshop**, никаких изменений не нужно.

В программе **cre8data** для указания путей используется знак "прямого слэша" (/), чтобы обеспечить переносимость между различными операционными средами. В UNIX и Linux для этого требуется прямой слэш. Windows принимает прямой слэш и может преобразовывать его в "обратный слэш" (\).

- 1) Чтобы выполнить программу, нажмите (**Run, Выполнить**) или клавишу F3.
- 2) На вкладке Results (Результаты) проверьте, что вывод содержит список файлов данных.

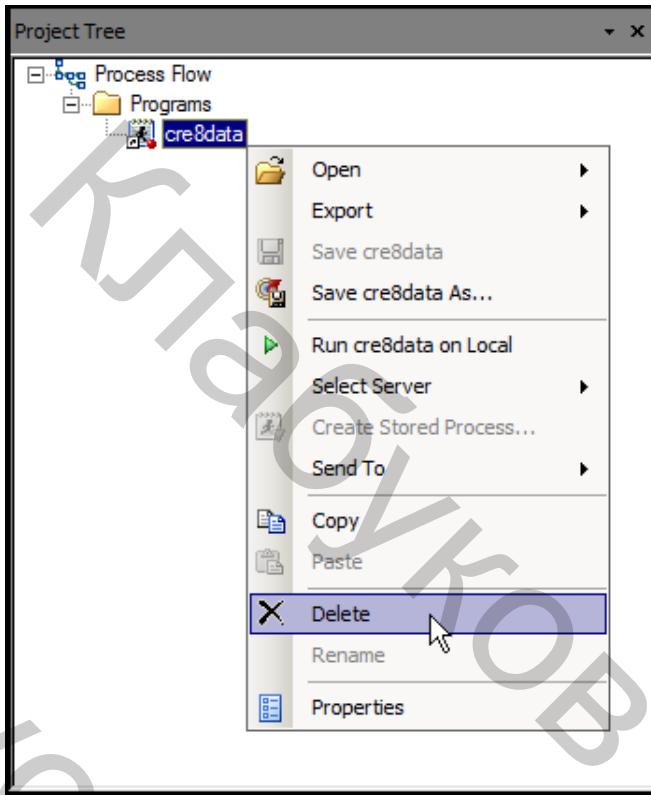
The CONTENTS Procedure

Directory	
Libref	ORION
Engine	V9
Physical Name	s:\workshop
Filename	s:\workshop

#	Name	Member Type	File Size	Last Modified
1	CHARITIES	DATA	131072	01/26/2015 11:33:25
2	CONSULTANTS	DATA	131072	01/26/2015 11:33:25

- 3) Удалите **cre8data** из проекта. Нажмите правой кнопкой мыши **cre8data** в окне Дерево проекта (Project Tree), и выберите **Delete (Удалить)**. Нажмите **Yes (Да)**, чтобы удалить все элементы, указанные в списке.



## Знакомство с SAS Enterprise Guide

### Основные вкладки интерфейса

- Создайте новую программу, выбрав в меню **File**  $\Rightarrow$  **New**  $\Rightarrow$  **Program** (Файл  $\Rightarrow$  Создать  $\Rightarrow$  Программа) или нажав на кнопку (New, Создать)  $\Rightarrow$  **Program** (Программа).

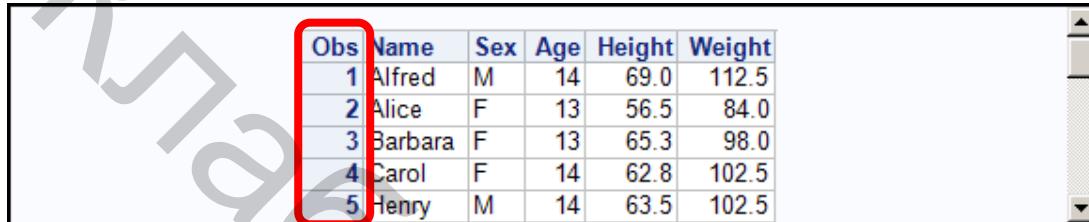
1) На вкладке Program (Программа) введите шаг PROC PRINT, показанный ниже.

```
proc print data=sashelp.class;
run;
```

- Выполните программу, нажав на вкладке Program (Программа) кнопку **Run** (Запустить) или нажмите F3. Если программа успешно выполняется, автоматически открывается вкладка Results (Результаты), в которой находится вывод PROC PRINT.
- Откройте вкладку **Log (Журнал)** и проверьте наличие в ней двух примечаний, показанных ниже. Если вы видите какие-либо предупреждения (warning) или ошибки (error), вернитесь на вкладку Program (Программа), чтобы исправить их, а затем повторно запустите программу.

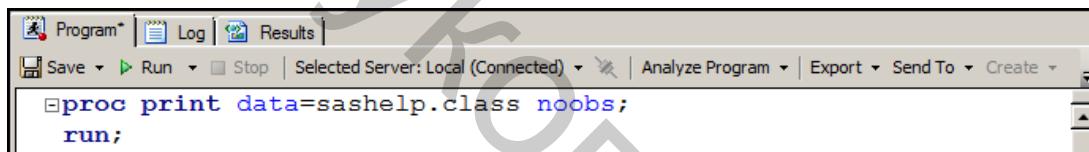
NOTE: There were 19 observations read from the data set SASHELP.CLASS.  
 NOTE: PROCEDURE PRINT used (Total process time):

- 4) Откройте вкладку **Results (Результаты)**. Обратите внимание, что вывод PROC PRINT содержит столбец **Obs**.



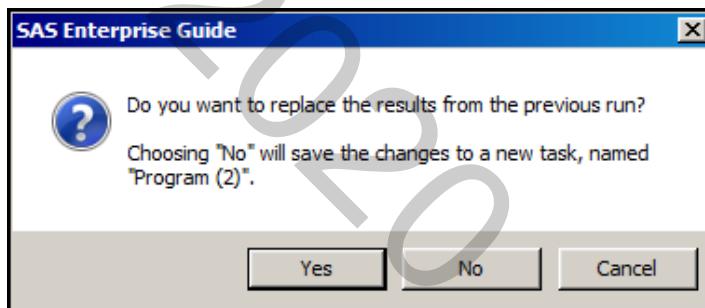
Obs	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69.0	112.5
2	Alice	F	13	56.5	84.0
3	Barbara	F	13	65.3	98.0
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5

- 5) Откройте вкладку **Program (Программа)** и добавьте параметр NOOBS, чтобы убрать этот столбец из отчёта.



 Параметр NOOBS нужно указать в операторе PROC PRINT перед точкой с запятой.

- 6) Выполните программу, нажав на вкладке **Program (Программа)** кнопку **Run (Выполнить)** или клавишу F3 на клавиатуре.
- 7) Нажмите **Yes (Да)** для замены результатов предыдущего запуска.



- 8) Просмотрите содержимое вкладок **Results (Результаты)** и **Log (Журнал)** после перезапуска программы.
- b. Создайте другую программу. Выберите в меню **File**  $\Rightarrow$  **New**  $\Rightarrow$  **Program (Файл**  $\Rightarrow$  **Создать**  $\Rightarrow$  **Программа**) или нажмите кнопку  (**New, Создать**)  $\Rightarrow$  **Program (Программа)**.
- 1) Введите текст новой программы, содержащей шаги PROC PRINT и PROC MEANS, как показано ниже.

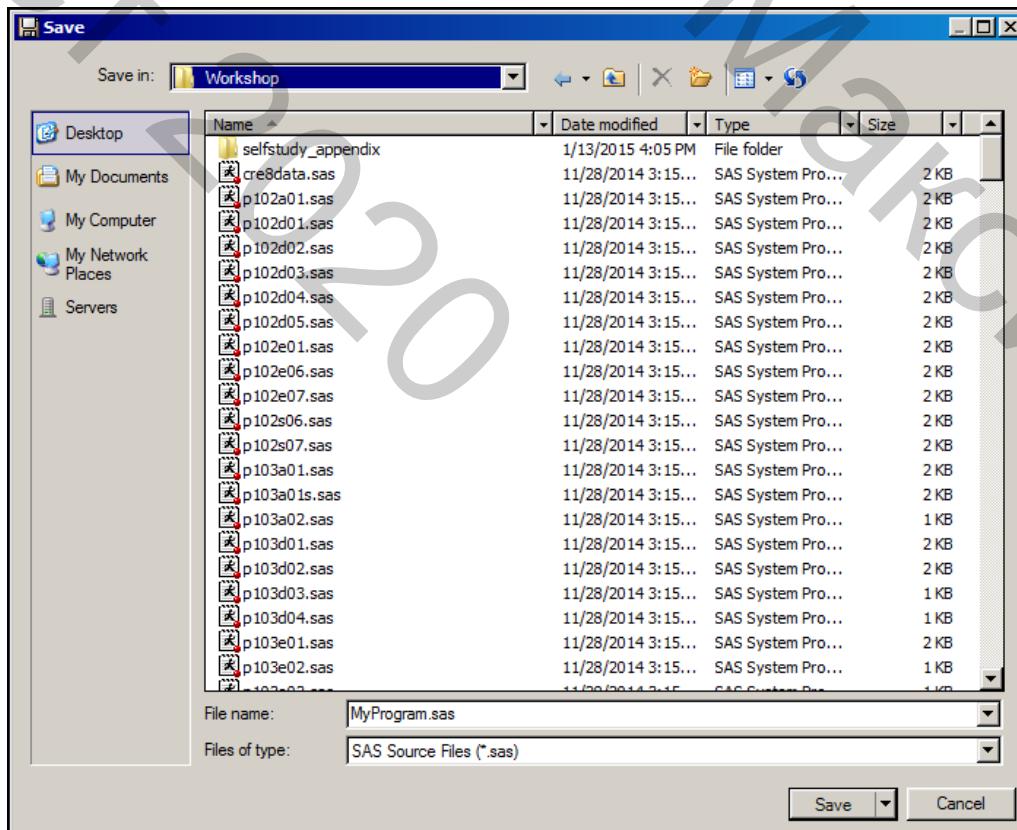
```

Program*
Save Run Stop Selected Server: Local (Connected) Analyze Program Export Send To Create
proc print data=sashelp.class;
run;
proc means data=sashelp.class;
run;

```

Звёздочка (\*) после слова **Program** на вкладке Program (Программа) означает, что программа не была сохранена.

- 2) Выполните программу, нажав на вкладке Program (Программа) кнопку **Run** (**Выполнить**) или клавишу F3 на клавиатуре. Если программа выполняется успешно, автоматически открывается вкладка Results (Результаты), содержащая вывод PROC PRINT и PROC MEANS.
- 3) Откройте вкладку **Log**. Проверьте примечания в журнале. Если вы видите какие-либо предупреждения (warning) или ошибки (error), вернитесь на вкладку Program (Программа), чтобы исправить их, а затем повторно запустите программу.
- 4) Откройте вкладку **Program** (Программа). Сохраните программу, выберите **File ⇨ Save Program (2) As** (**Файл ⇨ Сохранить Программа (2) как**) или нажмите на кнопку **Save As** (**Сохранить как**).
- 5) В окне Save (Сохранить) выберите место расположения файла, например, **s:\workshop**, назовите файл **MyProgram** и нажмите кнопку **Save**.



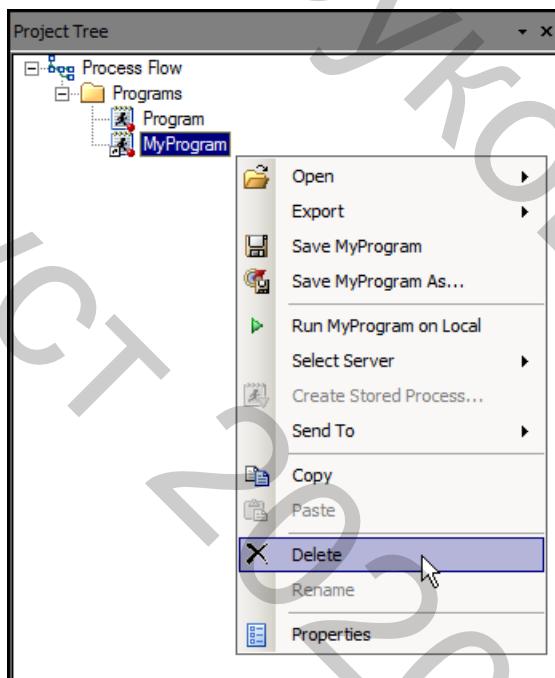
- 6) Обратите внимание, что рабочая область содержит название программы над вкладками Program, Log и Results (Программа, Журнал и Результаты).

```

proc print data=sashelp.class;
run;
proc means data=sashelp.class;
run;

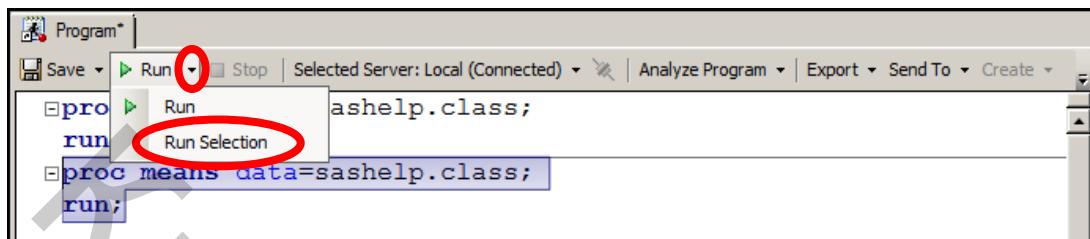
```

- 7) Удалите **MyProgram** из проекта. Щёлкните правой кнопкой мыши **MyProgram** в Дереве проекта (Project Tree) и выберите **Delete** (Удалить). Нажмите **Yes** (Да), чтобы удалить все элементы, указанные в списке.



Этими действиями вы удаляете программу из проекта. Однако, программа по прежнему существует как файл на жёстком диске.

- c. Откройте программу, которую вы сохранили и удалили из проекта. Выберите **File**  $\Rightarrow$  **Open**  $\Rightarrow$  **Program** (**Файл**  $\Rightarrow$  **Открыть**  $\Rightarrow$  **Программа**) или щёлкните (**Open**, **Открыть**)  $\Rightarrow$  **Program** (**Программа**).
- 1) В окне Open Program (Открыть программу) перейдите в место расположения файла **MyProgram**. Нажмите кнопку **Open** (**Открыть**).
  - 2) Выполните фрагмент программы. На вкладке **Program** (Программа) выделите шаг PROC MEANS (две строчки кода). Чтобы выполнить выделенный код, нажмите стрелку рядом с кнопкой **Run** (выполнить) и выберите **Run Selection** (**Выполнить выделение**) или клавишу F3.



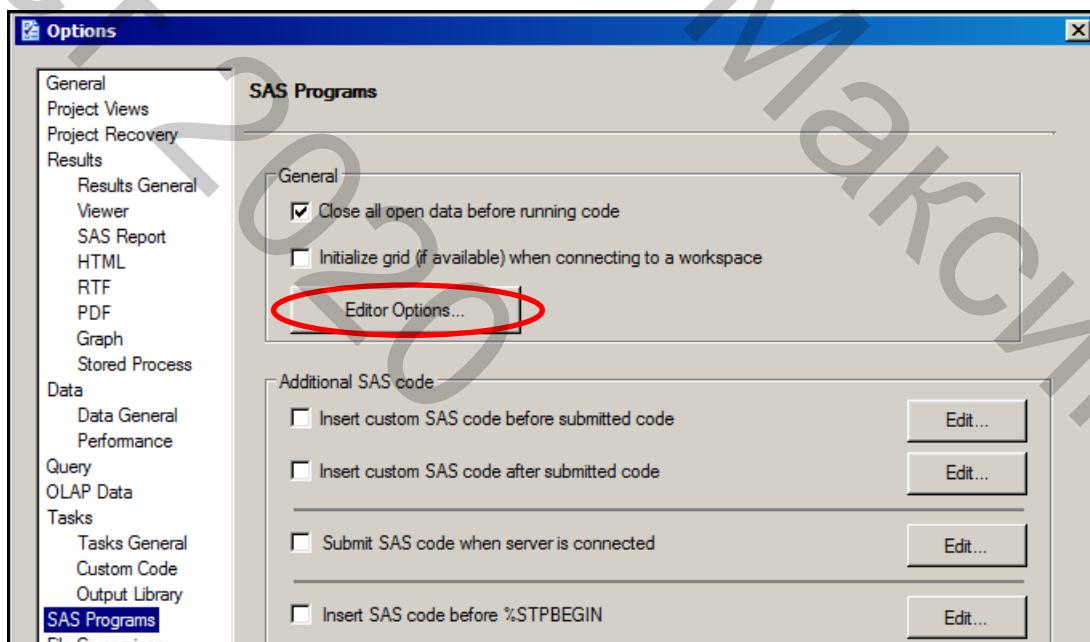
- 3) Просмотрите содержимое вкладок Results и Log (Результаты и Журнал). Обратите внимание, что информация в этих вкладках относится только к шагу PROC MEANS.

### Использование функций вкладки Program (Программа)

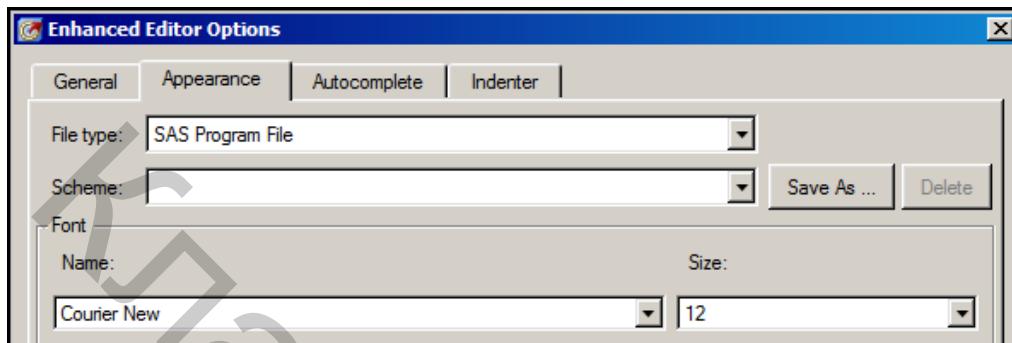
- Выберите View  $\Rightarrow$  Maximize Workspace (Вид  $\Rightarrow$  Развернуть рабочую область), чтобы развернуть окно с программой и свернуть окна Project Tree и Servers (Дерева проекта и Серверы).

Чтобы вновь отобразить эти окна, ещё раз нажмите View  $\Rightarrow$  Maximize Workspace (Вид  $\Rightarrow$  Развернуть рабочую область).

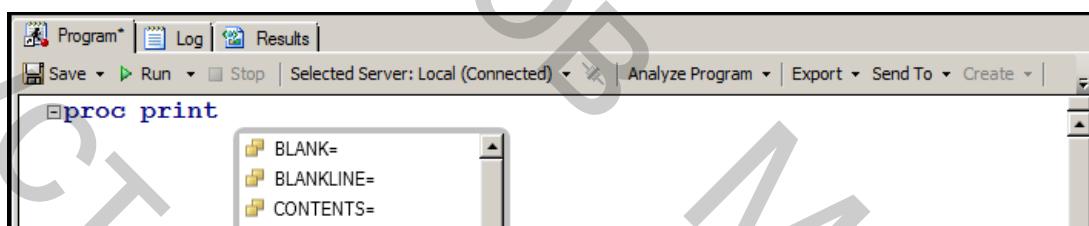
- Откройте вкладку Program (Программа) программы MyProgram.
- Внешний вид и функционал вкладки Program (Программа) можно настраивать. Для перехода к настройкам, выберите Tools  $\Rightarrow$  Options  $\Rightarrow$  SAS Programs (Сервис  $\Rightarrow$  Параметры  $\Rightarrow$  Программа SAS). Затем нажмите Editor Options (Параметры редактора).



- В окне Enhanced Editor Options (Параметры редактора) на четырёх вкладках расположены различные настройки. На вкладке Appearance (Отображение) можно увеличить размер шрифта и нажать OK. Нажмите OK, чтобы закрыть окно с настройками параметров. Обратите внимание на изменившийся размер шрифта у текста программы на вкладке Program (Программа).

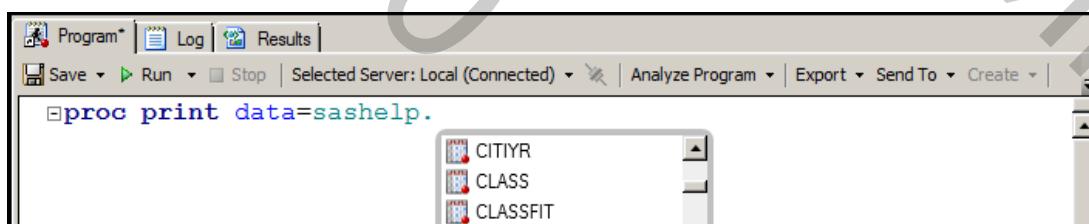


- c. Создайте новую программу. Выберите в меню **File** **⇒** **New** **⇒** **Program** (**Файл** **⇒** **Создать** **⇒** **Программа**) или нажмите кнопку (**New, Создать**) **⇒** **Program** (**Программа**).
- 1) Во вкладке **Program** (**Программа**) начните вводить шаг **PROC PRINT**. Обратите внимание на функцию автозаполнения, которая появляется в процессе ввода текста программы. Функция автозаполнения позволяет вставлять в программу возможные ключевые слова, выбирая их во всплывающем окне. Например, после ввода **proc print** появляется всплывающее окно с параметрами оператора **PROC PRINT**.



Для перехода к нужному ключевому слову в этом окне, прокрутите их список клавишами "вверх" и "вниз", Page Up или Page Down с клавиатуры или прокрутите его мышью, используя полосу прокрутки. Чтобы добавить ключевое слово в программу, дважды щёлкните ключевое слово или нажмите клавишу Ввод (Enter) или пробел.

- 2) Помимо вывода допустимых ключевых слов, автозаполнение может подсказывать названия библиотек, наборов данных и переменных.



- 3) Закончите шаг **PROC PRINT**. Обратите внимание на функцию автозаполнения, которая появляется в процессе ввода текста программы.

```
proc print data=sashelp.class;
run;
```

- d. Вы можете воспользоваться справкой, наведя указатель мыши на ключевое слово в программе. Расположите указатель мыши над словом **print** в программе и просмотрите справку по синтаксису.

Keyword: [PRINT](#)

**Context: [PROCEDURE DEFINITION] PROC PRINT**

Syntax: PROC PRINT <option(s)>;  
 BY <DESCENDING> variable-1 <...<DESCENDING> variable-n><NOTSORTED>;  
 PAGEBY BY-variable;  
 SUMBY BY-variable;

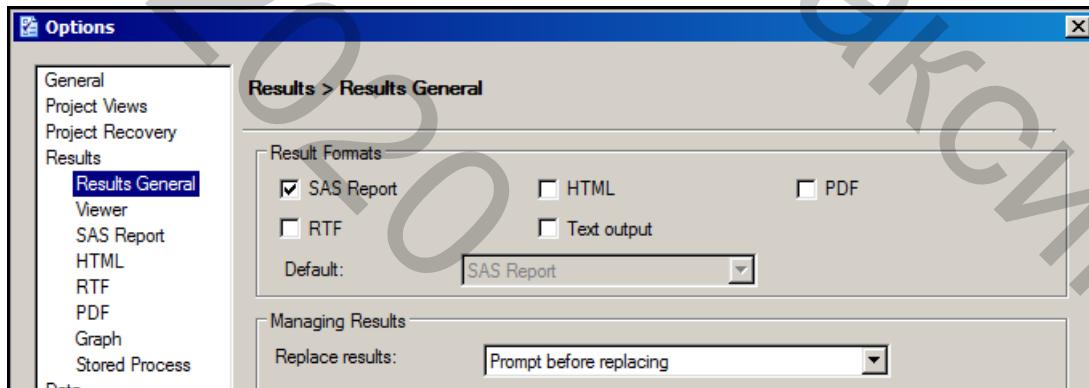
ID variable(s) <option>;  
 SUM variable(s) <option>;  
 VAR variable(s) <option>;

The PRINT procedure prints the observations in a SAS data set, using all or some of the variables. You can create a variety of reports ranging from a simple listing to a highly customized report that groups the data and calculates totals and subtotals for numeric variables.

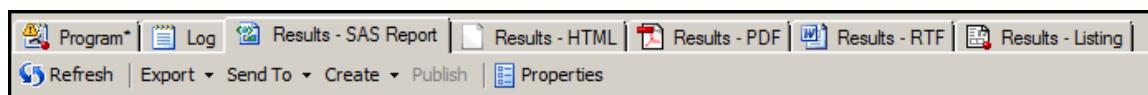
Search: [The Product Documentation](#), [Samples & SAS Notes](#), [Papers](#)

## Настройка результатов

- Выполните программу, нажав на вкладке Program (Программа) кнопку **Run (Выполнить)** или клавишу F3 на клавиатуре. Если программа успешно выполняется, автоматически открывается вкладка Results (Результаты), в которой находится вывод PROC PRINT.
- По умолчанию результаты, выводимые на вкладку Results (Результаты), имеют формат Отчёта SAS (SAS Report). Формат SAS Report - это файл XML, который можно просматривать в различных приложениях SAS.
- Помимо вывода в формате SAS Report, можно создавать отчёты и в других форматах. Выберите **Tools** ⇒ **Options** ⇒ **Results General** (Сервис ⇒ Параметры ⇒ Общие результаты).



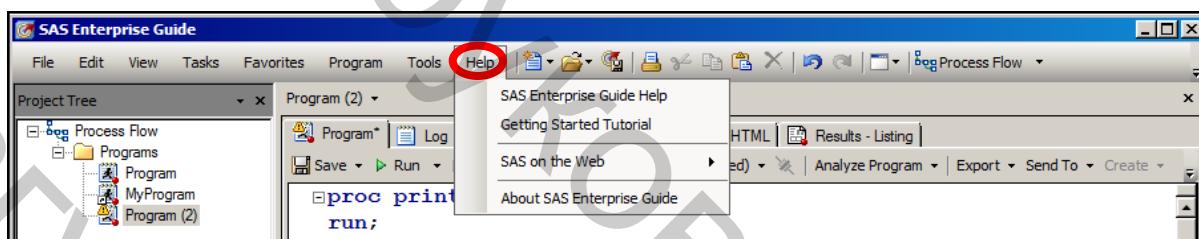
- Флажок SAS Report (Отчёт SAS) уже включен. Включите флажки для форматов **HTML**, **PDF**, **RTF** и **Text output** (Текстовый вывод). Нажмите **OK**.
- Выполните шаг PROC PRINT. Нажмите **Yes (Да)** для замены результатов предыдущего запуска. Появятся пять вкладок с результатами программы.



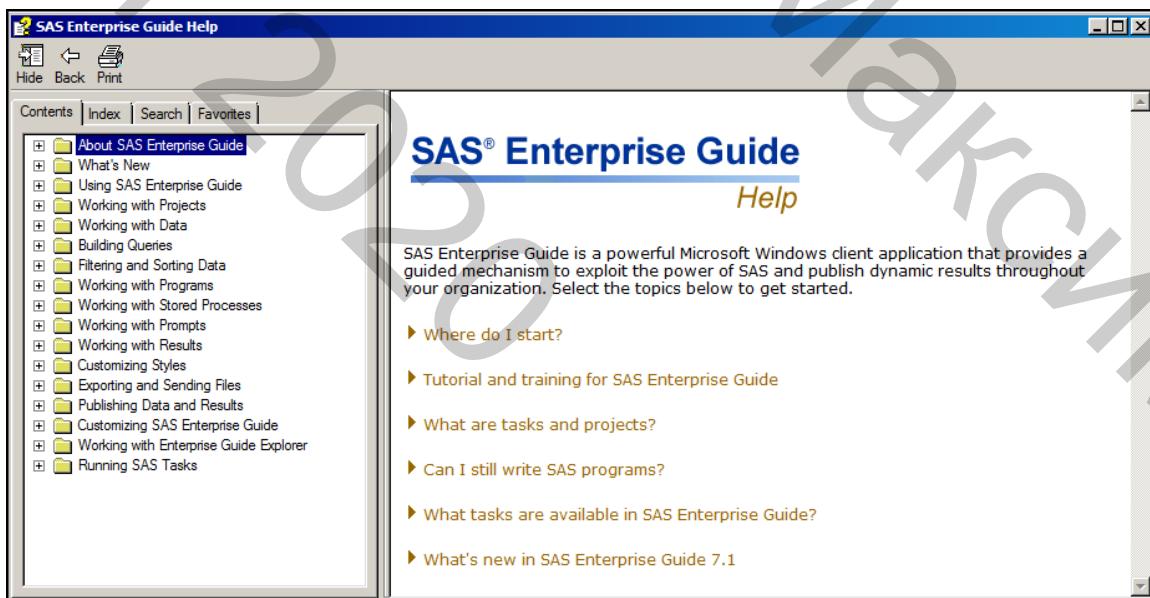
- c. Откройте каждую вкладку и просмотрите результаты. Когда вы откроете вкладку PDF, отчёт в формате PDF открывается в программе Adobe (например, Adobe Reader). Когда вы открываете вкладку RTF, отчёт в формате RTF открывается в текстовом редакторе (например, Microsoft Word). После нажатия на вкладки PDF и RTF они исчезают, потому что эти файлы открываются в сторонних программах. Вкладка Listing содержит текстовый вывод.
- d. Выберите **Tools**  $\Rightarrow$  **Options**  $\Rightarrow$  **Results General** (Сервис  $\Rightarrow$  Параметры  $\Rightarrow$  Общие результаты). Снимите флажки для форматов **HTML**, **PDF** и **RTF**. Не изменяйте флажки SAS Report (Отчёт SAS) и Text output (Текстовый вывод). Нажмите **OK**.

### Доступ к справке и документации

Перейти к справке и документации по SAS можно из меню Help (Справка).



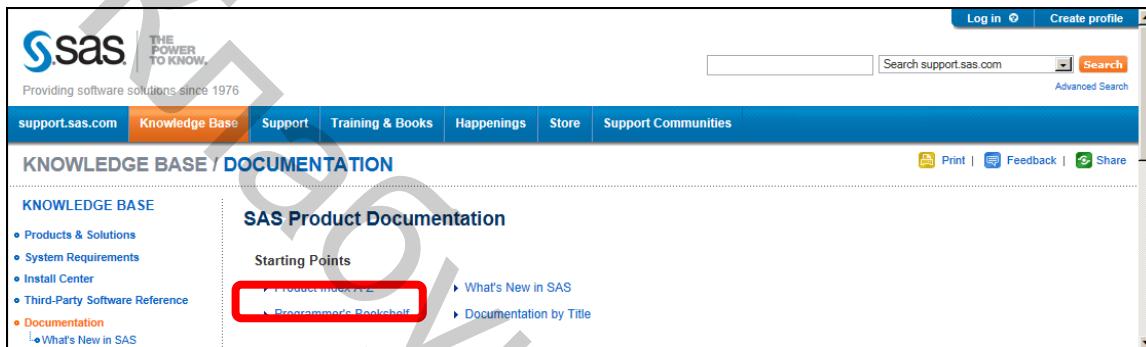
- a. Выберите **Help**  $\Rightarrow$  **SAS Enterprise Guide Help** (Справка  $\Rightarrow$  Справка приложения SAS Enterprise Guide) и откройте окно со справкой по SAS Enterprise Guide. Это окно содержит справку по интерфейсу SAS Enterprise Guide. Просмотрите окно справки и закройте его.



- b. Выберите **Help**  $\Rightarrow$  **SAS on the Web**  $\Rightarrow$  **SAS Product Documentation** (Справка  $\Rightarrow$  SAS на веб-страницах  $\Rightarrow$  Документация по продуктам SAS) для перехода на страницу документации по продуктам SAS. Эта веб-страница будет полезна для поиска документации по синтаксису, который используется в различных продуктах SAS.

 Внешний адрес страницы документации по продуктам SAS:  
<http://support.sas.com/documentation/>

- На этой странице нажмите ссылку **Programmer's Bookshelf**. Затем выберите ссылку в соответствии с используемой версией SAS.



 Страница Programmer's Bookshelf содержит избранную документацию по продуктам SAS (таким, как Base SAS).

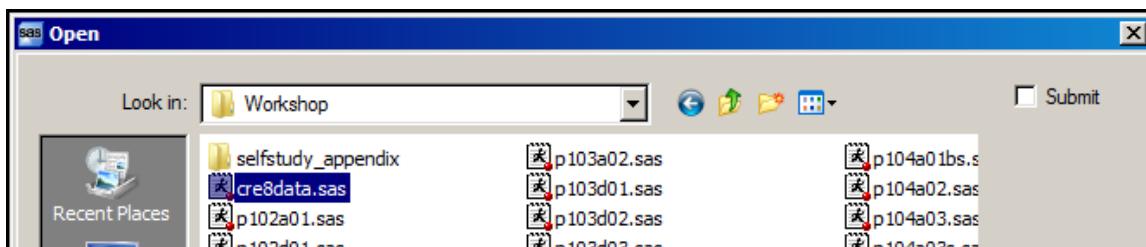
- Просмотрите страницу Programmer's Bookshelf. После этого закройте окно.

## Оконный интерфейс SAS (SAS Windowing Environment)

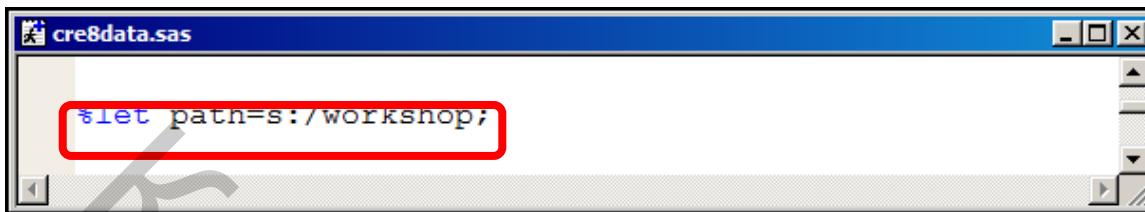
 Вы **должны** выполнить это упражнение для создания файлов с данными курса. Если вы не создадите файлы с данными, то большинство программ в этом курсе не будет работать.

### Создание данных для курса в оконном интерфейсе SAS

- Запустите оконный интерфейс SAS. Если вы видите диалог, содержащий описание изменений или советы по работе с SAS, нажмите **Close**. По умолчанию оконный интерфейс SAS состоит из левой панели, содержащей окна Results и Explorer, и правой панели с окнами Output, Log и Editor.
- Нажмите  (**Open**) при активном окне Editor или выберите в меню **File** ⇒ **Open Program**.
- В окне Open Program (Открыть программу) перейдите в место расположения программы **cre8data** и нажмите кнопку **Open** (Открыть).



- В окне Editor программы **cre8data** найдите оператор %LET.



- e. Если файлы данных нужно создать в директории, отличной от **s:\workshop**, измените значение, присваиваемое макропеременной **path**, указав в ней требуемую директорию. Если файлы данных для курса создаются в директории **s:\workshop**, никаких изменений не нужно.

В программе **cre8data** для указания путей используется знак "прямого слэша" (/), чтобы обеспечить переносимость между различными операционными средами. В UNIX и Linux для этого требуется прямой слэш. Windows принимает прямой слэш и может преобразовывать его в "обратный слэш" (\).

- f. Запустите программу, нажмите (**Submit**) или клавишу F3.  
g. В окне Results Viewer проверьте, что вывод содержит список файлов данных.

Directory	
Libref	ORION
Engine	V9
Physical Name	s:\workshop
Filename	s:\workshop

#	Name	Member Type	File Size	Last Modified
1	CHARITIES	DATA	131072	01/26/2015 11:42:34
2	CONSULTANTS	DATA	131072	01/26/2015 11:42:34

- h. Для очистки всего содержимого окна журнала (Log), откройте это окно и нажмите (**New**) или откройте его и выберите в меню **Edit** ⇒ **Clear All**.  
i. Закройте окно редактора (Editor) программы **cre8data** нажав значок X в правом верхнем углу окна.



## Знакомство с оконным интерфейсом SAS

### Использование основных окон интерфейса

- a. Выберите **View** ⇒ **Enhanced Editor** чтобы открыть новое окно редактора.  
1) В окне редактора введите шаг PROC PRINT, показанный ниже.

```
Editor - Untitled2 *
proc print data=sashelp.class;
run;
```

Для редактирования программы используется Enhanced Editor (расширенный редактор).

- 2) Запустите программу, нажмите (Submit) или клавишу F3. Если программа успешно выполняется, автоматически открывается окно Results Viewer, в котором находится вывод PROC PRINT.
- 3) Перейдите в окно журнала (Log) и проверьте наличие в нём двух примечаний, показанных ниже. Если вы видите какие-либо предупреждения (warning) или ошибки (error), вернитесь в окно редактора, чтобы исправить их, а затем повторно запустите программу.

**NOTE: There were 19 observations read from the data set SASHHELP.CLASS.**  
**NOTE: PROCEDURE PRINT used (Total process time):**

- 4) Снова перейдите в окно Results Viewer. Обратите внимание, что вывод PROC PRINT содержит столбец Obs.

The SAS System

Obs	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69.0	112.5
2	Alice	F	13	56.5	84.0
3	Barbara	F	13	65.3	98.0
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5

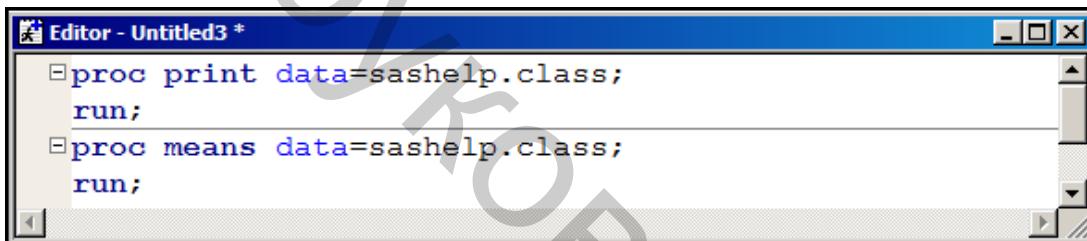
- 5) Откройте окно редактора и добавьте параметр NOOBS для того, чтобы убрать столбец Obs.

```
Editor - Untitled2 *
proc print data=sashelp.class noobs;
run;
```

Параметр NOOBS нужно указать в операторе PROC PRINT перед точкой с запятой.

- 6) Запустите программу, нажмите (Submit) или клавишу F3.

- 7) Изучите содержимое окон Results Viewer и Log. Обратите внимание, что информация в этих окнах накапливается. Самая свежая информация добавляется снизу.
  - 8) Закройте окно Log, нажав на значок X в правом углу окна. Выберите View  $\Rightarrow$  Log, чтобы повторно открыть журнал. Окно Log всё ещё содержит информацию о выполнении обоих фрагментов кода.
  - 9) Очистите окно журнала, нажав  (New) при активном окне Log. Также для этого можно выбрать в меню Edit  $\Rightarrow$  Clear All .
- b. Создайте другую программу. Нажмите  (New) из открытого окна редактора или выберите в меню View  $\Rightarrow$  Enhanced Editor.
- 1) В новом окне редактора введите шаги PROC PRINT и PROC MEANS, показанные ниже.



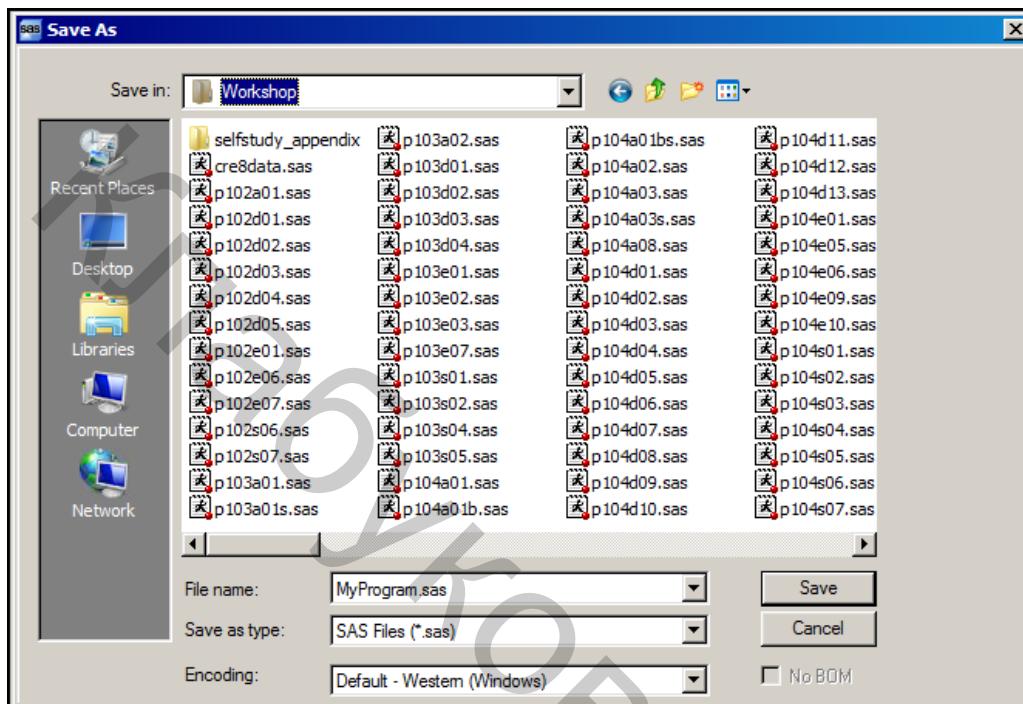
```

Editor - Untitled3 *
proc print data=sashelp.class;
run;
proc means data=sashelp.class;
run;

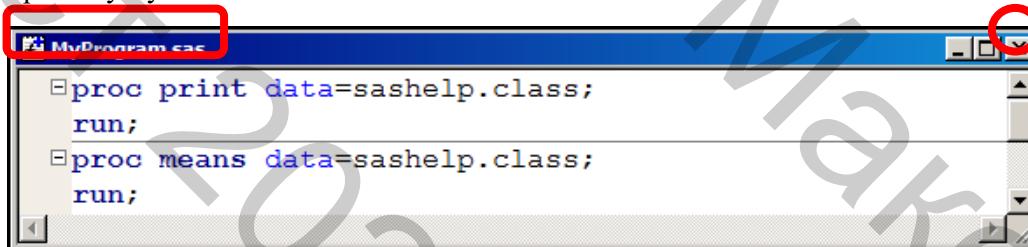
```

 Звёздочка (\*), выводимая после названия файла, означает, что программа не была сохранена.

- 2) Запустите программу, нажмите  (Submit) или клавишу F3. Если программа успешно выполняется, то автоматически будет открыта вкладка Results Viewer, содержащая вывод PROC PRINT и PROC MEANS. Не забывайте, что окно Results Viewer накапливает информацию. Самая свежая информация добавляется снизу.
- 3) Перейдите в окно Log. Проверьте примечания в журнале. Если вы видите какие-либо предупреждения (warning) или ошибки (error), вернитесь в окно редактора, чтобы исправить их, а затем повторно запустите программу.
- 4) Вернитесь в окно Editor – Untitled3. Нажмите  (Save) или выберите в меню File  $\Rightarrow$  Save As для сохранения программы.
- 5) В окне Save As выберите место расположения файла, например, s:\workshop, назовите файл MyProgram и нажмите кнопку Save.

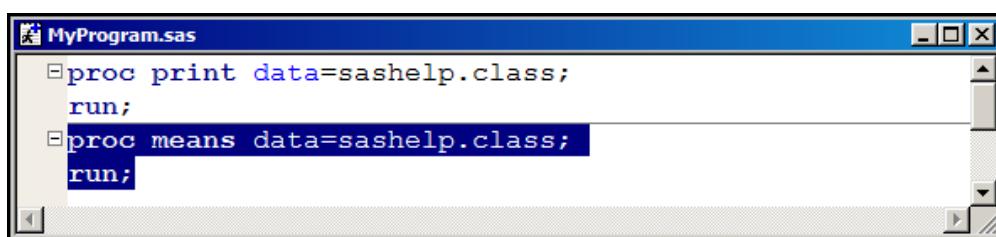


- 6) Обратите внимание, что заголовок окна редактора теперь содержит название программы. Закройте окно редактора программы **MyProgram**, нажав на значок X в правом углу окна.



- c. Откройте программу, которую вы сохранили и закрыли. Чтобы открыть программу, нажмите (Open) при активном окне редактора или выберите в меню File ⇨ Open Program.

- 1) В окне Open перейдите к месту расположения файла **MyProgram** и нажмите кнопку **Open**.
- 2) Выполните фрагмент программы. В окне редактора выделите шаг PROC MEANS (две строчки кода). Для выполнения выделенного фрагмента кода нажмите (Submit) или клавишу F3.



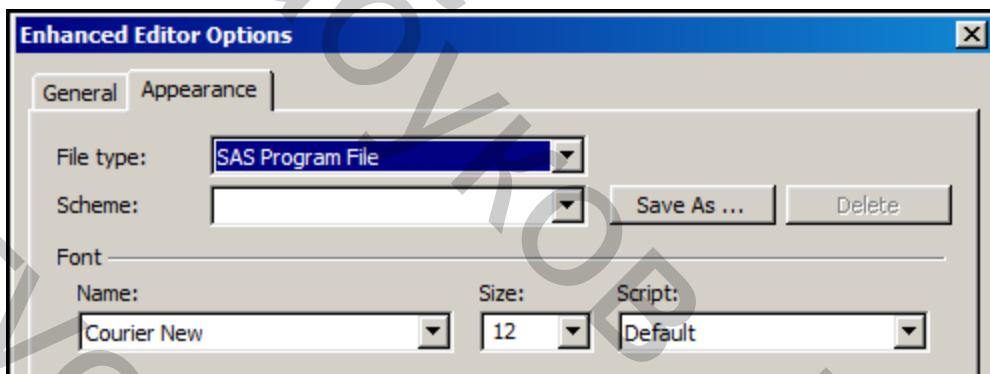
- 3) Просмотрите содержимое окон Results Viewer и Log. Обратите внимание, что при выполнении фрагмента программы выполнился только шаг PROC MEANS.

### Настройка редактора

- a. Вернитесь в окно редактора программы **MyProgram**. Если вы хотите, можете нажать кнопку **Развернуть** в правом верхнем углу окна, чтобы развернуть его.



- b. Внешний вид и функционал редактора можно настраивать. Для настройки редактора выберите в меню **Tools** ⇒ **Options** ⇒ **Enhanced Editor**.



- 1) В окне Enhanced Editor Options на двух вкладках расположены параметры редактора. На вкладке Appearance можно увеличить размер шрифта и нажать **OK**. Обратите внимание на изменение размера шрифта у текста программы в редакторе.
- 2) Список функциональных клавиш оконного интерфейса можно отредактировать, выбрав в меню **Tools** ⇒ **Options** ⇒ **Keys** и открыв окно Keys. Обратите внимание на следующие основные клавиши:
  - F1 - вызов справки.
  - F3 - команда end.
  - F6 - переход к журналу.
  - F8 - команды zoom off и submit.

Команда End - это псевдоним для команды Submit, когда она вызывается в окне редактора.

- 3) Функциональные клавиши можно настраивать. Например, F12 можно настроить для очистки журнала и выполнения программы. Затем эту клавишу можно будет использовать из окна редактора. Введите следующую команду в определении (столбец Definition) клавиши F12:

```
clear log; submit
```

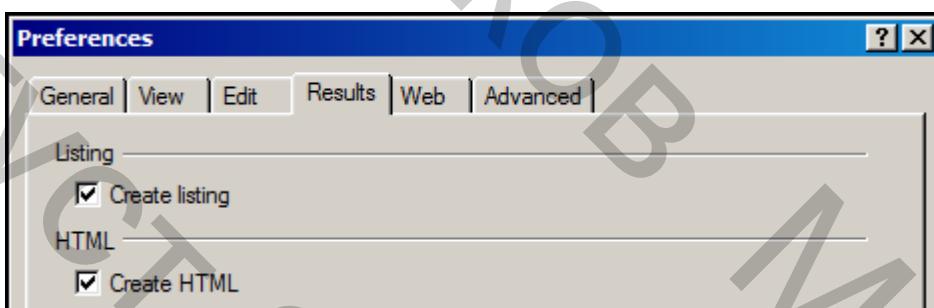
- 4) Закройте окно Keys, нажав на значок X в правом углу окна или нажав клавишу F3.
- c. Вернитесь в окно редактора программы **MyProgram**. Нажмите F12. Убедитесь, что окно журнала содержит только сообщения, относящиеся к последнему запуску шагов PROC PRINT и PROC MEANS.

## Настройка результатов

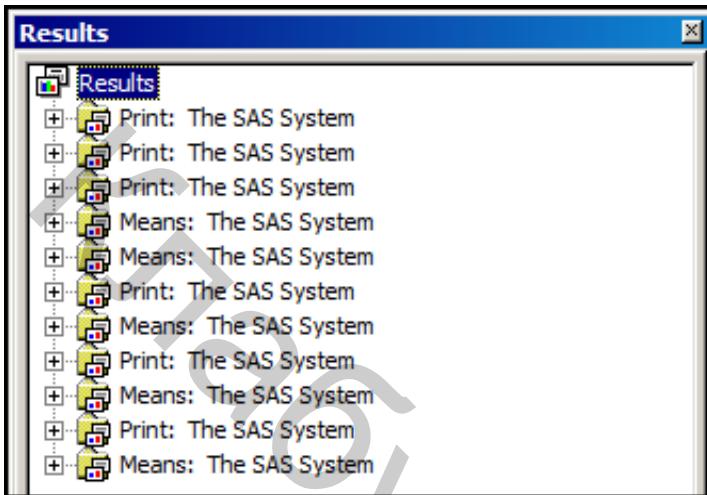
- a. По умолчанию вывод в окне Results Viewer имеет вид накапливающихся результатов в формате HTML4. Чтобы очистить предыдущий вывод, можно добавить в программу операторы ODS, которые закрывают существующий файл вывода и открывают новый. Добавьте два следующих оператора ODS в начало программы **MyProgram**:

```
ods html close; ods html;
```

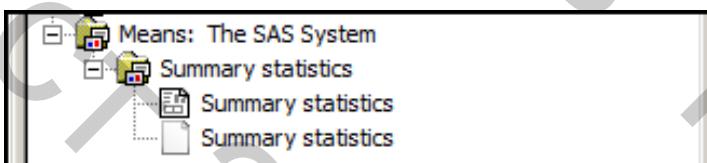
- b. Для выполнения изменённой программы нажмите  (Submit) или клавишу F3. Проверьте, что окно Results Viewer содержит только вывод PROC PRINT и PROC MEANS.
- c. Помимо вывода в формате HTML4 в окно Results Viewer, можно создать текстовый вывод в окне Output.
- 1) Выберите **Tools**  $\Rightarrow$  **Options**  $\Rightarrow$  **Preferences**, затем откройте вкладку **Results**. Флажок **Create HTML** уже установлен.
  - 2) Включите флажок **Create listing** для создания текстового вывода. Нажмите **OK**.



- d. Выполните программу.
- 1) Проверьте, что окно Results Viewer содержит вывод PROC PRINT и PROC MEANS. В окне Results Viewer можно отключить накопление результатов, если перед запуском программ выполнять два указанных выше оператора ODS.
  - 2) Кроме того, проверьте, что в окне Output также виден вывод PROC PRINT и PROC MEANS в текстовом формате.  
Окно Output также накапливает результаты. Самая свежая информация добавляется снизу.
  - 3) Окно Results используется для перемещения по результатам. Окно содержит закладки, которые можно разворачивать и сворачивать.



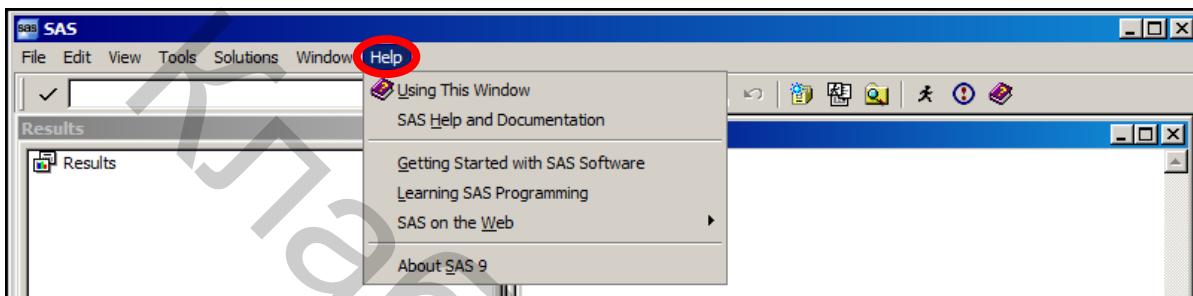
- e. Выберите в окне Results последний узел Means.
- 1) Щёлкните значок плюса перед Means, чтобы развернуть этот узел.
  - 2) Щёлкните значок плюса перед Summary statistics, чтобы развернуть этот узел. Каждая закладка содержит значок, указывающий тип файла, например HTML или LISTING (текстовый вывод).



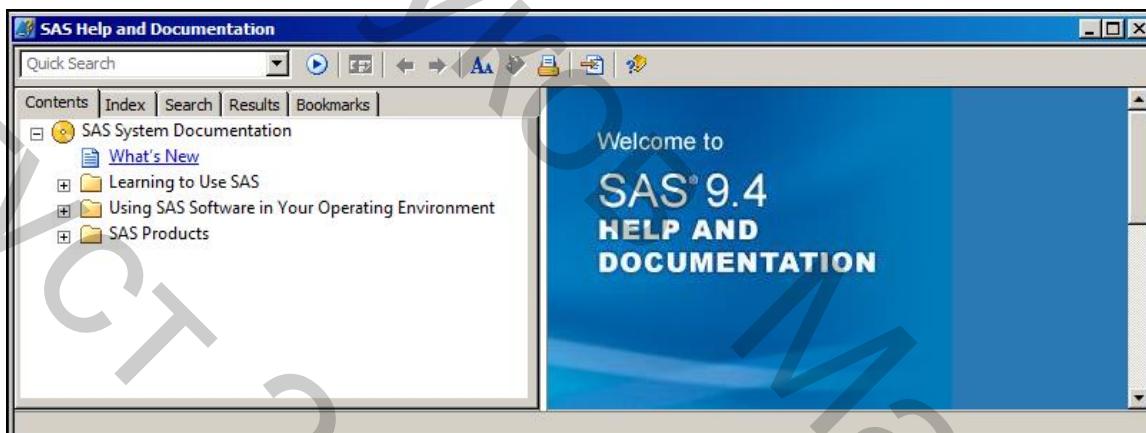
- 3) Двойным щелчком по одной из закладок откройте в соответствующем окне любой отчёт.
- f. Можно изменить настройку клавиши F12 для очистки окон Log, Output и Results.
- 1) Для изменения настройки клавиши F12, выберите в меню Tools ⇨ Options ⇨ Keys и откройте окно Keys.
  - 2) Введите следующую команду в столбец Definition для F12:
- ```
clear log; clear output; odsresults; clear;
```
- 3) Закройте окно Keys, нажав на значок X в правом углу окна или нажав клавишу F3.
  - 4) Нажмите F12. Убедитесь, что окна Log, Output и Results очищены.
- g. Для сохранения программы вернитесь в окно редактора программы MyProgram и нажмите (Save) или выберите File ⇨ Save.
- h. Закройте окно редактора программы MyProgram, нажав на значок X в правом углу окна.

## Доступ к справке и документации

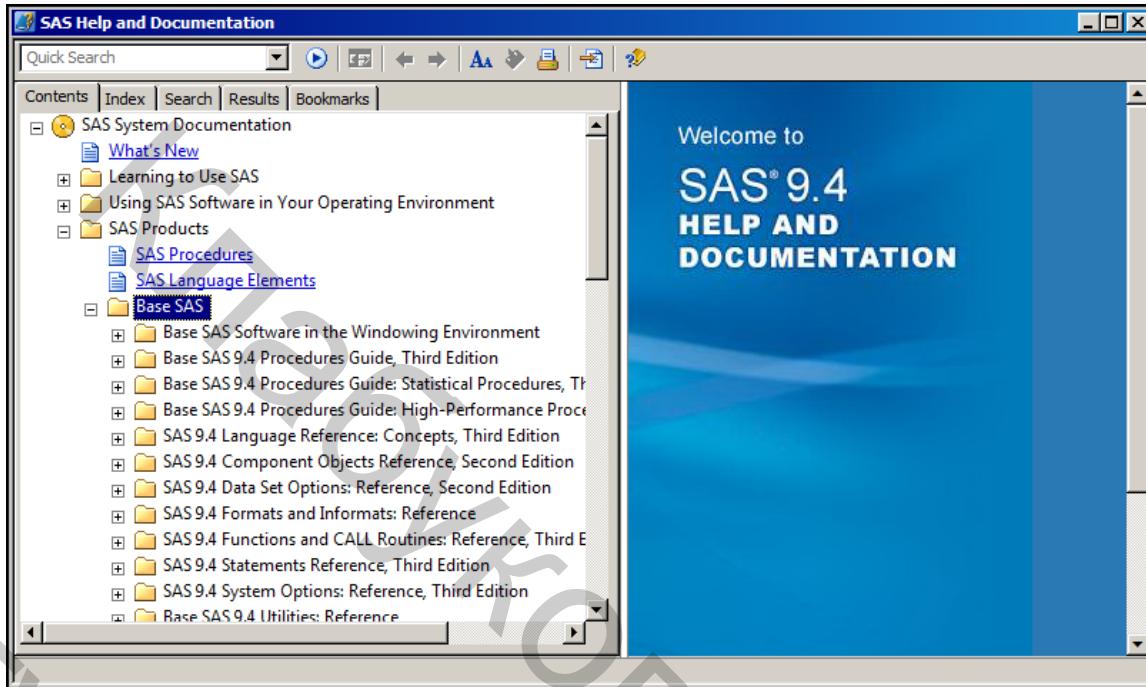
Перейти к справке и документации по SAS можно из меню **Help**.



- Нажмите **Help** ⇒ **SAS Help and Documentation** для перехода в окно со справкой и документацией по SAS.



- Чтобы просмотреть справку по синтаксису для конкретного продукта, раскройте **SAS Products** на вкладке **Contents**. Выберите продукт, например, Base SAS, и перейдите к документации по нему. Закройте окно, когда закончите просмотр.



- c. Кроме того, вы можете получить доступ к документации по продуктам SAS через веб-страницу <http://support.sas.com/documentation/>.
- d. На этой странице перейдите по ссылке **Programmer's Bookshelf**. Затем выберите ссылку в соответствии с используемой версией SAS.

 Страница **Programmer's Bookshelf** содержит избранную документацию по продуктам SAS (таким, как Base SAS).

- e. Просмотрите страницу **Programmer's Bookshelf**. Закройте окно после просмотра.

## Диагностика и исправление синтаксических ошибок



### p102d04

1. Откройте и запустите программу **p102d04**.
2. Перейдите в журнал и проверьте в нём примечания, предупреждения и ошибки.

Журнал показывает, какие действия выполнил SAS:

- предположил, что ключевое слово DATA было написано с ошибкой, и выполнил шаг DATA
- интерпретировал слово RUN как параметр оператора PROC PRINT (из-за пропущенной точки с запятой), поэтому шаг PROC PRINT не был выполнен
- не распознал слово AVERAGE как допустимый параметр оператора PROC MEANS, поэтому шаг PROC MEANS также не был выполнен

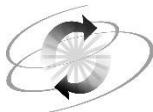
3. Вернитесь к программе и отредактируйте её.
  - a. Исправьте написание оператора DATA.
  - b. Поставьте точку с запятой в конце оператора PROC PRINT.
  - c. Измените слово AVERAGE на MEAN в операторе PROC MEANS.

```
data work.newsalesemps;
length First_Name $ 12
      Last_Name $ 18 Job_Title $ 25;
infile "&path\newemps.csv" dlm=',';
input First_Name $ Last_Name $ 
      Job_Title $ Salary;
run;

proc print data=work.newsalesemps;
run;

proc means data=work.newsalesemps mean min;
  var Salary;
run;
```

4. Выполните программу. Проверьте, что она выполняется без предупреждений или ошибок и создаёт результаты.
5. Сохраните исправленную программу.



## Упражнения. Исправление непарных кавычек

### SAS Studio

#### Исправление непарных кавычек в SAS Studio

- В SAS Studio воспользуйтесь панелью навигации и перейдите в область Folders (Папки), найдите в ней программу **p102e10**.
- Двойным щелчком откройте программу **p102e10**. Обратите внимание на отсутствие закрывающей кавычки в параметре DLM= в операторе INFILE.
- Для выполнения программы нажмите (Run, Выполнить) на вкладке CODE (Код) или клавишу F3.
- На вкладке LOG (Журнал) обратите внимание на отсутствие сообщений после каждого шага. Такое поведение часто указывает на наличие несбалансированных (непарных) кавычек.

#### Фрагмент журнала SAS

```

71   data work.newsalesemps;
72     length First_Name $ 12
73       Last_Name $ 18 Job_Title $ 25;
74     infile "&path\newemps.csv" dlm=',';
75     input First_Name $ Last_Name $
76       Job_Title $ Salary;
77   run;
78
79   proc print data=work.newsalesemps;
80   run;
81
82   proc means data=work.newsalesemps;
83     var Salary;
84   run;

```

- Вернитесь в программу. Чтобы исправить программу, добавьте закрывающую кавычку в параметр DLM= оператора INFILE.

```
infile "&path\newemps.csv" dlm=', '
```

- Для выполнения программы нажмите (Run, Выполнить) на вкладке CODE (Код) или клавишу F3.
- На вкладке RESULTS (Результаты) просмотрите вывод PROC PRINT и PROC MEANS.
- На вкладке LOG (Журнал) проверьте отсутствие ошибок или предупреждений.
- Для сохранения программы, на вкладке CODE (Код) нажмите (Save Program, Сохранить программу).

#### Диагностика ошибок в SAS Studio

- a. Создайте новую программу, для этого в области Folders (Папки) панели навигации нажмите (New, Новая) ⇒ SAS Program (F4) (Программа SAS (F4)) или нажмите клавишу F4.

- b. На вкладке CODE (Код) введите следующую программу, отформатированную как показано ниже:

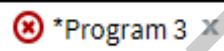
```
proc print data= sashelp.class; run; proc means  
data=sashelp.class; run;
```

- c. Для более удобного чтения вашей программы нажмите (Format Code, Формат кода) на панели инструментов вкладки CODE (Код). Эта кнопка автоматически форматирует ваш код, добавляя в него переносы строк и выравнивая его в соответствии с уровнями вложенности.
- d. Добавьте в эту программу две ошибки. Удалите букву **o** из первого шага PROC и удалите точку с запятой после последнего названия набора данных **sashelp.class**.

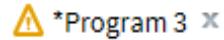
```
proc print data=sashelp.class;  
run;  
proc means data=sashelp.class  
run;
```

- e. Чтобы выполнить программу, нажмите (Run, Выполнить) на вкладке CODE (Код) или клавишу F3. На вкладке Program (Программа) находится значок (перед названием программы), показывающий наличие ошибок или предупреждений в программе.

Ошибки либо ошибки и предупреждения:



Предупреждения без ошибок:



- f. Обратите внимание, что вкладка RESULTS (Результаты) содержит вывод PROC PRINT, но не содержит вывод PROC MEANS.
- g. Прокрутите журнал на вкладке LOG (Журнал). Примечания (Note) выделяются голубым цветом, предупреждения (warning) - зелёным, а ошибки (error) - красным.
- h. В начале окна журнала разверните секции со списками ошибок (Errors), примечаний (Notes) и предупреждений (Warnings).

Errors, Warnings, Notes

Errors (2)

ERROR 22-322: Syntax error, expecting one of the following: :, (, ALPHA, CHARTYPE, CLASSDATA, CLM,  
ERROR 202-322: The option or parameter is not recognized and will be ignored.

Warnings (1)

WARNING 14-169: Assuming the symbol PROC was misspelled as prc.

- i. Нажмите на сообщение в любой из этих секций. Сообщение выделяется в журнале, который выводится снизу.

The screenshot shows the SAS Studio interface. The top navigation bar includes 'File', 'Edit', 'View', 'Project', 'Help', and a 'SAS' logo. Below the bar, the 'Errors, Warnings, Notes' panel is open, displaying two errors and one warning:

- Errors (2)**
  - ERROR 22-322: Syntax error, expecting one of the following: ;, (, ALPHA, CHARTYPE, CLASSDATA, CLM,
  - ERROR 202-322: The option or parameter is not recognized and will be ignored.
- Warnings (1)**
  - WARNING 14-169: Assuming the symbol PROC was misspelled as prc.

The code editor below shows the following SAS code:

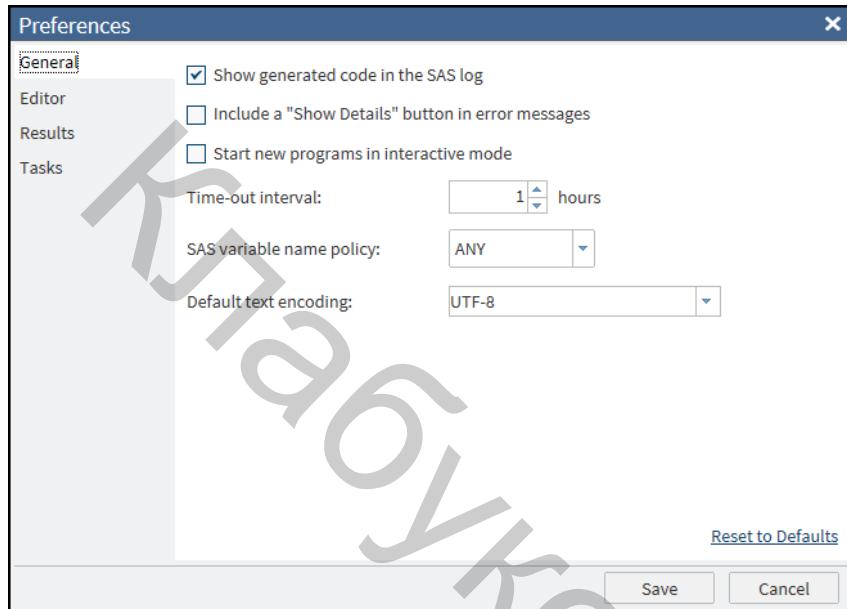
```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
57
58      prc print data=sashelp.class;
14
59      run;

NOTE: There were 19 observations read from the data set SASHELP.CLASS.

```

- j. Вернитесь на вкладку CODE (Код) и исправьте две ошибки. Выполните программу и проверьте, что она не выводит никаких ошибок или предупреждений в журнал.
- k. Обратите внимание на оператор OPTIONS в журнале, указанный в начале и в конце программы. SAS Studio автоматически включает эти инструкции до и после вашей программы. Кроме того, SAS Studio включает дополнительный код до и после вашей программы. По умолчанию этот код не отображается в журнале.
- l. Для просмотра полного кода программы, исполняемого SAS и содержащего, в том числе, код, добавленный автоматически, нажмите (More Application Options, Дополнительные параметры приложения) ⇨ Preferences ⇨ General (Настройки ⇨ Общие).
- m. Включите флагок Show generated code in the SAS log (Отобразить сгенерированный код в журнале SAS) и нажмите Save (Сохранить).



- n. Выполните программу и изучите дополнительный код в журнале.
- o. Вернитесь в окно дополнительных параметров и снимите флажок **Show generated code in the SAS log** (Отобразить сгенерированный код в журнале SAS). Нажмите **Save** (Сохранить).

## SAS Enterprise Guide

### Исправление непарных кавычек в SAS Enterprise Guide

- a. В меню SAS Enterprise Guide выберите **File**  $\Rightarrow$  **Open**  $\Rightarrow$  **Program** (Файл  $\Rightarrow$  Открыть  $\Rightarrow$  Программа) или нажмите (Open, Открыть)  $\Rightarrow$  **Program** (Программа).
- b. В окне Open Program (Открыть программу) перейдите в место расположения программы **p102e12** и нажмите **Open** (Открыть). Обратите внимание, что закрывающая кавычка в параметре DLM= оператора INFILE пропущена.
- c. Чтобы запустить программу, нажмите на вкладке Program (Программа) кнопку **Run** (Выполнить), или нажмите клавишу F3.
- d. На вкладке Log (Журнал) обратите внимание на отсутствие сообщений после каждого шага. Такое поведение часто указывает на наличие несбалансированных (непарных) кавычек.

### Фрагмент журнала SAS

```

71      data work.newsalesemps;
72          length First_Name $ 12
73              Last_Name $ 18 Job_Title $ 25;
74          infile "&path\\newemps.csv" dlm=',';
75          input First_Name $ Last_Name $
76              Job_Title $ Salary;
77      run;
78
79      proc print data=work.newsalesemps;
80      run;

```

```

81
82      proc means data=work.newsalesemps;
83          var Salary;
84      run;

```

- e. Вернитесь в программу. Чтобы исправить программу, добавьте закрывающую кавычку в параметр DLM= оператора INFILE.

```
infile "&path\newemps.csv" dlm=', ';
```

- f. Выполните программу, нажав Run (Выполнить) на вкладке Program (Программа) или клавишу F3. Нажмите Yes (Да), чтобы заменить результаты.
- g. На вкладке Results (Результаты) просмотрите вывод PROC PRINT и PROC MEANS.
- h. На вкладке Log (Журнал) проверьте отсутствие ошибок или предупреждений.
- i. Сохраните программу, для этого нажмите Save (Сохранить) на вкладке Program (Программа).

#### Диагностика ошибок в SAS Enterprise Guide

- a. Для создания новой программы выберите в меню File ⇒ New ⇒ Program (Файл ⇒ Создать ⇒ Программа) или нажмите (New, Создать) ⇒ Program (Программа).
- b. На вкладке Program (Программа) введите следующую программу, отформатированную как показано ниже.

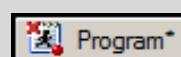
```
proc print data=sashelp.class;
run; proc means data=sashelp.class;    run;
```

- c. Нажмите Edit ⇒ Format Code (Изменить ⇒ Форматировать код), чтобы вашу программу было удобнее читать. Эта кнопка автоматически форматирует ваш код, добавляя в него переносы строк и выравнивая его в соответствии с уровнями вложенности.
- d. Добавьте в эту программу две ошибки. Удалите букву **o** из первого шага PROC и удалите точку с запятой после последнего названия набора данных sashelp.class.

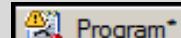
```
proc print data=sashelp.class;
run;
proc means data=sashelp.class
run;
```

- e. Запустите программу, для этого на вкладке Program (Программа) нажмите кнопку Run (Выполнить) или клавишу F3. На вкладке Program (Программа) находится значок (перед названием программы), показывающий наличие ошибок или предупреждений в программе.

Ошибки либо ошибки и предупреждения:

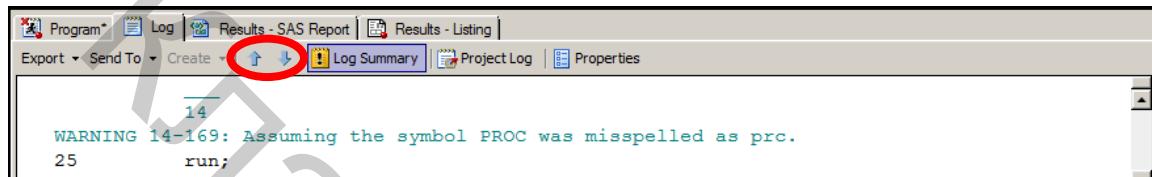


Предупреждения без ошибок:



- f. Обратите внимание, что вкладка Results (Результаты) содержит вывод PROC PRINT, но не содержит вывод PROC MEANS.

- g. Откройте вкладку LOG (Журнал) и прокрутите журнал. Примечания (Note) выделяются зелёным цветом, предупреждения (warning) - синим, а ошибки (error) - красным. Используйте кнопки со стрелками вверх и вниз на вкладке Log (Журнал), чтобы найти предыдущее или последующее предупреждение или ошибку.



- h. Перейдите в окно Log Summary (Сводная информация), расположенное под вкладкой Log (Журнал). Сводная информация содержит все ошибки, предупреждения и примечания, полученные при запуске программы, а также связанные с ними номера строк и код, вызвавший их. По умолчанию выводятся все три типа сообщений. Можно отфильтровать сообщения, включив или отключив соответствующий тип сообщения при помощи кнопок.

| Log Summary |                                                                                          |      |                               |
|-------------|------------------------------------------------------------------------------------------|------|-------------------------------|
|             | Description                                                                              | Line | Affected Code                 |
| ✖           | Errors (2)                                                                               |      |                               |
| ⚠           | Warnings (1)                                                                             |      |                               |
| ⓘ           | Notes (6)                                                                                |      |                               |
|             |                                                                                          |      |                               |
| ⓘ           | NOTE: Writing TAGSETS.SASREPORT13(EGSR) Body file: EGSR                                  | 24   | options(rolap="on")           |
| ⚠           | WARNING 14-169: Assuming the symbol PROC was misspelled as prc.                          | 32   | prc print data=sashelp.class; |
| ⓘ           | NOTE: There were 19 observations read from the data set SASHELP.CLASS.                   | 35   | run;                          |
| ⓘ           | NOTE: The PROCEDURE PRINT printed page 1.                                                | 36   | run;                          |
| ⓘ           | NOTE: PROCEDURE PRINT used (Total process time):                                         | 37   | run;                          |
| ✖           | ERROR 22-322: Syntax error, expecting one of the following: :, (, ALPHA, CHARTYPE, CL... | 48   | run;                          |
| ✖           | ERROR 202-322: The option or parameter is not recognized and will be ignored.            | 54   | run;                          |
| ⓘ           | NOTE: The SAS System stopped processing this step because of errors.                     | 66   | /*";*/quit;                   |
| ⓘ           | NOTE: PROCEDURE MEANS used (Total process time):                                         | 67   | /*";*/quit;                   |

- i. На вкладке Log (Журнал) в окне Log Summary (Сводная информация) нажмите описание сообщения и изучите само сообщение.

A screenshot of the SAS Log window and Log Summary table. The Log window shows the same log entries as before, with the warning message highlighted in blue:

```

24      prc print data=sashelp.class;

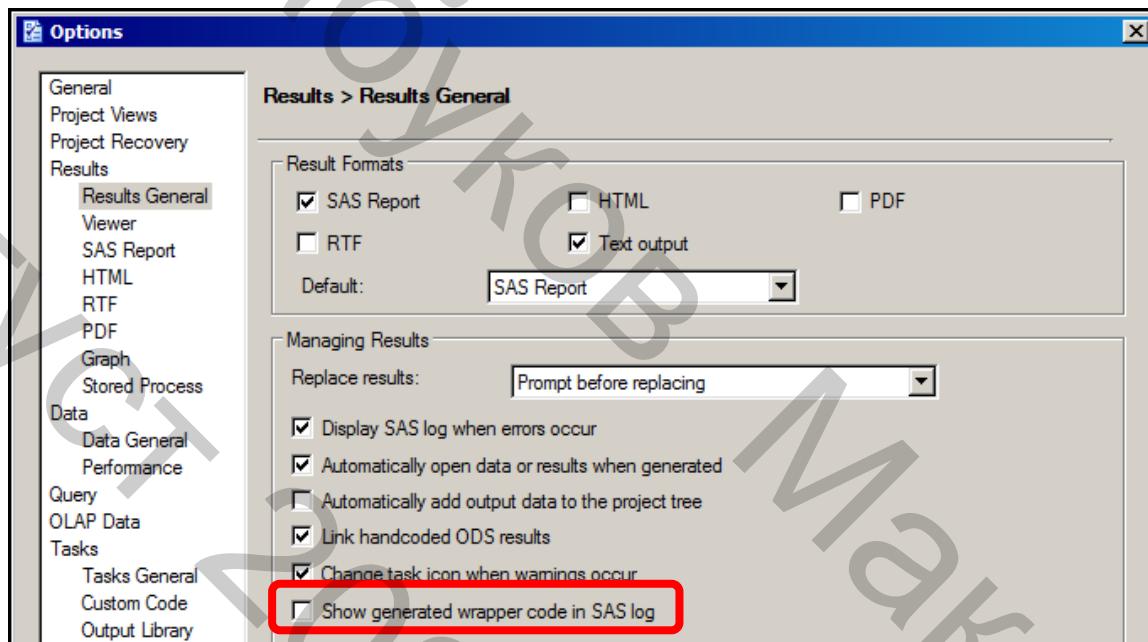
14
WARNING 14-169: Assuming the symbol PROC was misspelled as prc.
25      run;

```

The Log Summary table below shows the detailed information for the warning message:

| Log Summary |                                                                        |      |                               |
|-------------|------------------------------------------------------------------------|------|-------------------------------|
|             | Description                                                            | Line | Affected Code                 |
| ✖           | Errors (2)                                                             |      |                               |
| ⚠           | Warnings (1)                                                           |      |                               |
| ⓘ           | Notes (6)                                                              |      |                               |
| ⓘ           | NOTE: Writing TAGSETS.SASREPORT13(EGSR) Body file: EGSR                | 24   | options(rolap="on")           |
| ⚠           | WARNING 14-169: Assuming the symbol PROC was misspelled as prc.        | 32   | prc print data=sashelp.class; |
| ⓘ           | NOTE: There were 19 observations read from the data set SASHELP.CLASS. | 35   | run;                          |
| ⓘ           | NOTE: The PROCEDURE PRINT printed page 1.                              | 36   | run;                          |
| ⓘ           | NOTE: PROCEDURE PRINT used (Total process time):                       | 37   | run;                          |

- j. Откройте вкладку **Program** (Программа) и исправьте две ошибки. Выполните программу и нажмите **Yes** (Да) для замены результатов предыдущего запуска. Проверьте, что в журнале отсутствуют ошибки или предупреждения.
- k. Обратите внимание, что в журнале содержится дополнительный код, который SAS Enterprise Guide включает до и после вашей программы. Этот сгенерированный код включается в журнал по умолчанию.
- l. Чтобы скрыть его, выберите в меню **Tools**  $\Rightarrow$  **Options**  $\Rightarrow$  **Results General** (Сервис  $\Rightarrow$  Параметры  $\Rightarrow$  Общие результаты). Снимите флажок **Show generated wrapper code in SAS log** (Показывать созданный код программы-оболочки в журнале SAS) и нажмите **OK**.



- m. Выполните программу и изучите журнал. Обратите внимание, что большая часть автоматически сгенерированного кода теперь не выводится в журнал.

## Оконный интерфейс SAS (SAS Windowing Environment)

### Исправление кавычек в оконном интерфейсе SAS

- В оконном интерфейсе SAS нажмите (Open) при активном окне редактора или выберите **File**  $\Rightarrow$  **Open Program**.
- В окне Open перейдите в место расположения программы **p102e14** и нажмите **Open**. Обратите внимание, что в параметре DLM= оператора INFILE пропущена закрывающая кавычка.
- Запустите программу, нажмите (Submit) или клавишу F3.
- В окне Log обратите внимание на отсутствие сообщений после каждого шага. Такое поведение часто указывает на наличие несбалансированных (непарных) кавычек.

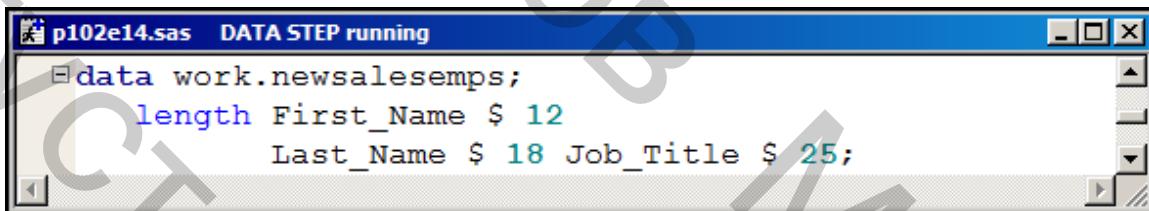
Фрагмент журнала SAS

```

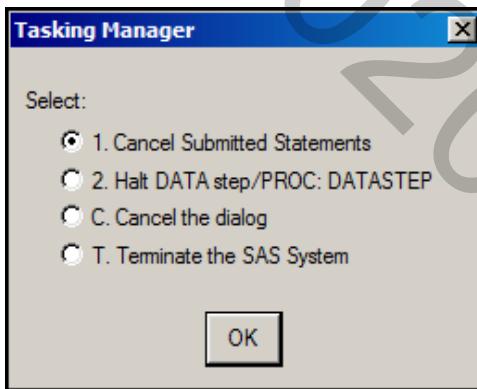
71      data work.newsalesemps;
72          length First_Name $ 12
73              Last_Name $ 18 Job_Title $ 25;
74      infile "&path\\newemps.csv" dlm=',';
75      input First_Name $ Last_Name $
76          Job_Title $ Salary;
77      run;
78
79      proc print data=work.newsalesemps;
80      run;
81
82      proc means data=work.newsalesemps;
83          var Salary;
84      run;

```

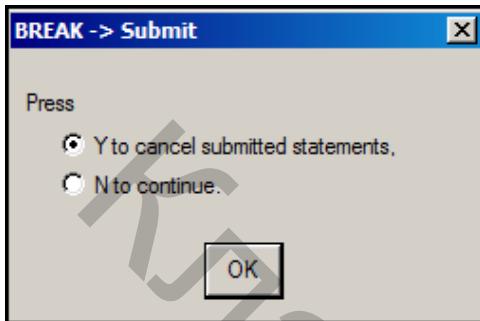
- e. Для очистки содержимого журнала, сделайте активным окно Log и нажмите (New) или выберите в меню **Edit** ⇒ **Clear All**.
- f. Вернитесь в программу. Обратите внимание на сообщение “DATA STEP running” в заголовке окна редактора. Это сообщение появилось потому, что оператор RUN рассматривается как часть символьной константы, а не как граница шага.



- g. Чтобы остановить работу шага DATA, нажмите (Break) или клавиши Ctrl и Break.
- h. Выберите из предложенных вариантов 1. Cancel Submitted Statements в окне Tasking Manager. Нажмите **OK**.



- i. Нажмите **Y to Cancel submitted statements.** . Нажмите **OK**.



- j. Исправьте программу, добавив закрывающую кавычку в параметр DLM= оператора INFILE.

```
infile "&path\newemps.csv" dlm=' , ' ;
```

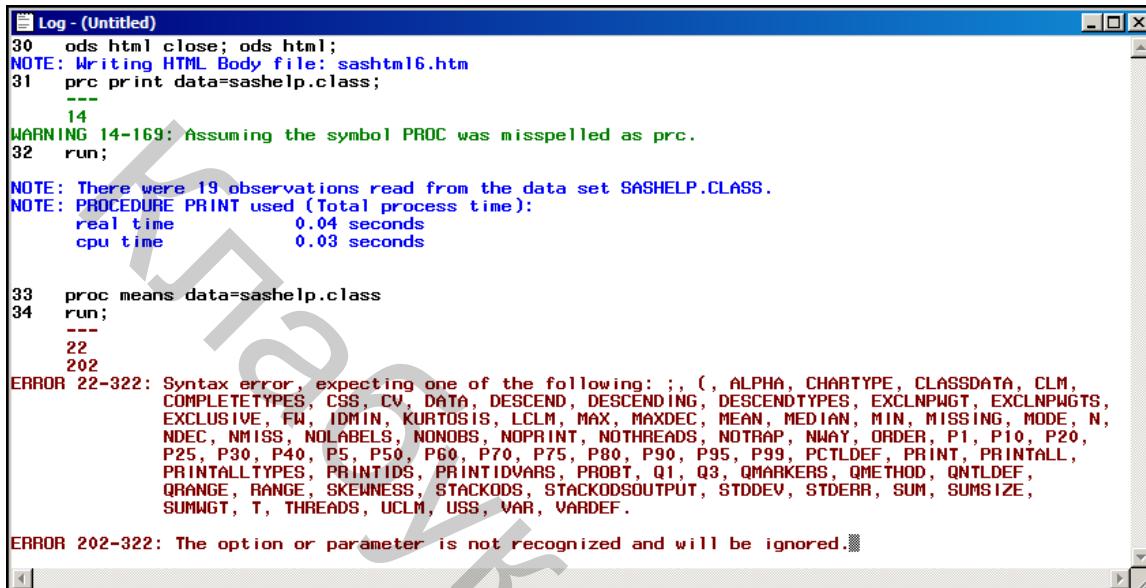
- k. Запустите программу, нажмите (Submit) или клавишу F3.
- l. В окне Results Viewer просмотрите вывод PROC PRINT и PROC MEANS.
- m. В окне Log проверьте отсутствие ошибок или предупреждений.
- n. Для очистки содержимое окна Log нажмите (New) или выберите в меню **Edit**  $\Rightarrow$  **Clear All**.
- o. Сохраните программу, нажав (Save) в окне редактора.

#### Диагностика ошибок в оконном интерфейсе SAS

- a. Откройте программу, нажав (Open) при активном окне Editor или выберите **File**  $\Rightarrow$  **Open Program**.
- b. В окне Open перейдите в место расположения программы **MyProgram**. Нажмите **Open**.
- c. Добавьте в эту программу две ошибки. Удалите букву **o** из первого шага PROC и удалите точку с запятой после последнего названия набора данных **sashelp.class**.

```
ods html close; ods html;
proc print data=sashelp.class;
run;
proc means data=sashelp.class 
```

- d. Выполните программу, нажав (Submit) или клавишу F3.
- e. Обратите внимание, что в окнах Results Viewer и Output содержится вывод PROC PRINT, но отсутствует вывод PROC MEANS.
- f. В окне Log прокрутите журнал. Примечания (Note) выделяются голубым цветом, предупреждения (warning) - зелёным, а ошибки (error) - красным.



```

Log - (Untitled)
30 ods html close; ods html;
NOTE: Writing HTML Body file: sashmt16.htm
31 proc print data=sashelp.class;
   ---
14
WARNING 14-169: Assuming the symbol PROC was misspelled as prc.
32 run;

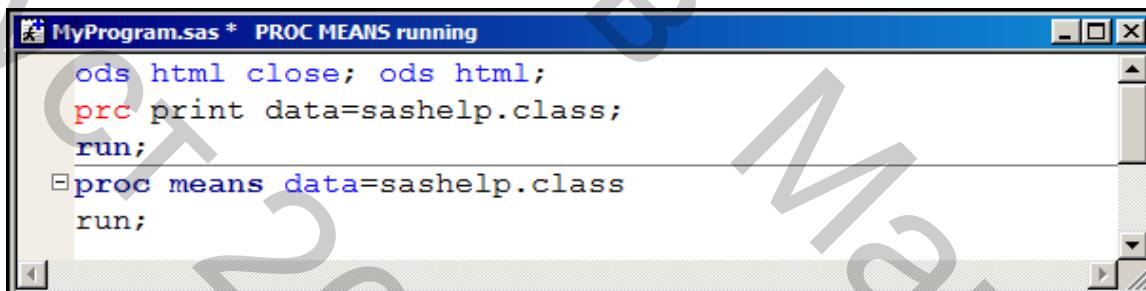
NOTE: There were 19 observations read from the data set SASHHELP.CLASS.
NOTE: PROCEDURE PRINT used (Total process time):
      real time          0.04 seconds
      cpu time           0.03 seconds

33 proc means data=sashelp.class
34 run;
   ---
22
202
ERROR 22-322: Syntax error, expecting one of the following: ;, (, ALPHA, CHARTYPE, CLASSDATA, CLM,
COMPLETETYPES, CSS, CV, DATA, DESCEND, DESCENDING, DESCENDTYPES, EXCLNPWGT, EXCLNPWGTS,
EXCLUSIVE, FW, IDMIN, KURTOSIS, LCLM, MAX, MAXDEC, MEAN, MEDIAN, MIN, MISSING, MODE, N,
NDEC, NMISS, NOLABELS, NONOBS, NOPRINT, NOTREADS, NOTRAP, NWAY, ORDER, P1, P10, P20,
P25, P30, P40, P5, P50, P60, P70, P75, P80, P90, P95, P99, PCTLDEF, PRINT, PRINTALL,
PRINTALLTYPES, PRINTIDS, PRINTIDVARS, PROBT, Q1, Q3, QMARKERS, QMETHOD, QNTLDEF,
QRANGE, RANGE, SKEWNESS, STACKODS, STACKODSOUTPUT, STDDEV, STDERR, SUM, SUMSIZE,
SUMWGT, T, THREADS, UCLM, USS, VAR, VARDEF.

ERROR 202-322: The option or parameter is not recognized and will be ignored.

```

- g. Вернитесь в окно редактора. Обратите внимание на заголовок окна и надпись “PROC MEANS running.”

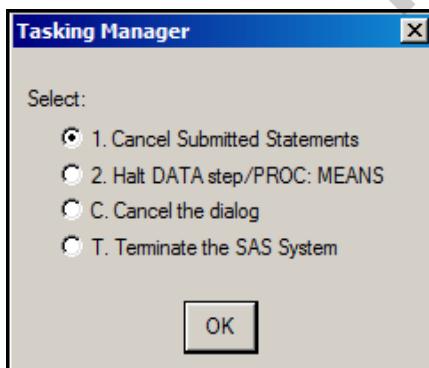


```

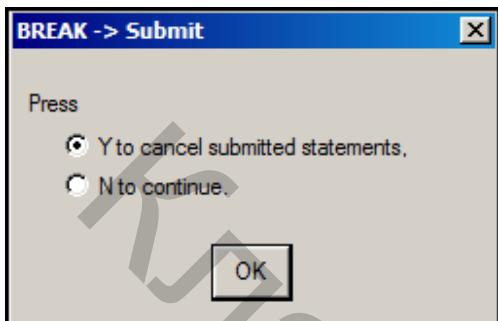
MyProgram.sas * PROC MEANS running
ods html close; ods html;
proc print data=sashelp.class;
run;
proc means data=sashelp.class
run;

```

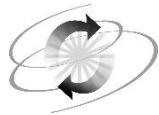
- h. Остановите выполнение шага, нажав  (Break) или клавиши Ctrl и Break.
- i. Выберите из предложенных вариантов 1. Cancel Submitted Statements в окне Tasking Manager. Нажмите OK.



- j. Нажмите Y to cancel submitted statements. Нажмите OK.



- k. Программа больше не работает, поэтому можно исправить обе ошибки. Очистите журнал, перезапустите программу и проверьте, что журнал не содержит никаких ошибок или предупреждений.



## Упражнения. Изучение компоненты данных набора данных SAS

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными. После этого также будет задано значение макропеременной **path**.

### Изучение компоненты данных набора данных SAS

- Откройте программу **p103e01**.
- После шага PROC CONTENTS допишите шаг PROC PRINT, выводящий из набора данных `work.donations` все наблюдения, все переменные и столбец с номером наблюдения (Obs).
- Выполните программу, создающую отчёт, который показан ниже. Результаты содержат 124 наблюдения.

Фрагмент вывода PROC PRINT

| Obs | Employee_ID | Qtr1 | Qtr2 | Qtr3 | Qtr4 | Total |
|-----|-------------|------|------|------|------|-------|
| 1   | 120265      | .    | .    | .    | 25   | 25    |
| 2   | 120267      | 15   | 15   | 15   | 15   | 60    |
| 3   | 120269      | 20   | 20   | 20   | 20   | 80    |
| ... |             |      |      |      |      |       |
| 123 | 121145      | 35   | 35   | 35   | 35   | 140   |
| 124 | 121147      | 10   | 10   | 10   | 10   | 40    |

## Просмотр библиотек SAS при помощи программы



### libname, p103d03

11. Откройте программу **libname** и добавьте в неё оператор LIBNAME, показанный ниже.

```
%let path=s:/workshop;
libname orion "s:/workshop";
```

Оператор %LET создаёт макропеременную с названием **path** и задаёт для неё значение, содержащее полный путь к директории с данными курса. Оператор LIBNAME создаёт библиотеку **orion** для данных, располагающихся по этому же пути.

 Программа **libname** создаётся программой **cre8data**. В коде используются знаки "прямого слэша" чтобы обеспечить переносимость между различными операционными средами. В UNIX и Linux для этого требуется прямой слэш. Windows принимает прямой слэш и может преобразовывать его в "обратный слэш" (\).

12. Выполните программу.
13. Проверьте журнал, убедитесь, что библиотека **orion** была успешно создана. Физическое имя директории (physical name) отражает расположение файлов с данными в формате SAS. Обратите внимание, что ваши файлы могут находиться в другой директории.

```
43      %let path=s:/workshop;
44      libname orion "s:/workshop";
NOTE: Libref ORION was successfully assigned as follows:
      Engine:          V9
      Physical Name:  s:\workshop
```

14. Сохраните программу **libname**. При этом её предыдущая версия будет перезаписана.
-  Когда вы работаете с данными и подготовленными программами для курса, запускайте программу **libname** в начале каждого сеанса SAS. Она создаёт макропеременную **path** и определяет библиотеку **orion**.

15. Воспользуйтесь процедурой CONTENTS для изучения содержимого библиотеки **orion** программным способом. Откройте и запустите программу **p103d03**, создающую список членов библиотеки. Фрагмент вывода показан ниже.

```
proc contents data=orion._all_nods;
run;
```

The CONTENTS Procedure

#### Directory

|               |             |
|---------------|-------------|
| Libref        | ORION       |
| Engine        | V9          |
| Physical Name | s:\workshop |
| Filename      | s:\workshop |

| Member | File |
|--------|------|
|--------|------|

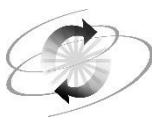
| # | Name        | Type  | Size  | Last Modified    |
|---|-------------|-------|-------|------------------|
| 1 | CHARITIES   | DATA  | 9216  | 23Aug12:15:58:39 |
| 2 | CONSULTANTS | DATA  | 5120  | 23Aug12:15:58:39 |
| 3 | COUNTRY     | DATA  | 17408 | 13Oct10:19:04:39 |
|   | COUNTRY     | INDEX | 17408 | 13Oct10:19:04:39 |
| 4 | CUSTOMER    | DATA  | 33792 | 04Nov11:09:52:27 |

16. Откройте новую программу. Введите и выполните следующий оператор, который удаляет ссылочное имя библиотеки Orion (НЕ её содержимое!) из сеанса SAS.

```
libname orion clear;
```

17. Проверьте журнал, убедитесь, что ссылочное имя библиотеки **orion** было удалено.

```
15      libname orion clear;
NOTE: Libref ORION has been deassigned.
```



## Упражнения. Создание и просмотр содержимого библиотеки

Используя выбранный вами интерфейс, выполните упражнение 4 и одно дополнительное упражнение. Для выполнения этих упражнений просто следуйте инструкциям.

### Создание и просмотр содержимого библиотеки при помощи программы

- Откройте программу **libname** и добавьте в неё оператор LIBNAME, показанный ниже.

```
%let path=s:/workshop;
libname orion "s:/workshop";
```



Программа **libname** создаётся программой **cre8data**. В коде используются знаки "прямого слэша", чтобы обеспечить переносимость между различными операционными средами. В UNIX и Linux для этого требуется прямой слэш. Windows принимает прямой слэш и может преобразовывать его в "обратный слэш" (\).

- Выполните программу.
- Проверьте журнал, убедитесь, что библиотека **orion** была успешно создана. Физическое имя директории (physical name) отражает расположение файлов с данными в формате SAS. Обратите внимание, что ваши файлы могут находиться в другой директории.

```
43      %let path=s:/workshop;
44      libname orion "s:/workshop";
NOTE: Libref ORION was successfully assigned as follows:
      Engine:      V9
      Physical Name: s:\workshop
```

- Сохраните программу **libname**. При этом её предыдущая версия будет перезаписана.
- Воспользуйтесь процедурой CONTENTS для изучения содержимого библиотеки **orion** программным способом. Создайте новую программу. Введите следующий шаг PROC CONTENTS и оператор LIBNAME, который удаляет ссылочное имя библиотеки :

```
proc contents data=orion._all_nods;
run;

libname orion clear;
```

- Выполните программу. Фрагмент вывода показан ниже.

The CONTENTS Procedure

#### Directory

|               |             |
|---------------|-------------|
| Libref        | ORION       |
| Engine        | V9          |
| Physical Name | s:\workshop |
| Filename      | s:\workshop |

| # | Name | Member | File |               |
|---|------|--------|------|---------------|
|   |      | Type   | Size | Last Modified |

|   |             |       |       |                  |
|---|-------------|-------|-------|------------------|
| 1 | CHARITIES   | DATA  | 9216  | 23Aug12:15:58:39 |
| 2 | CONSULTANTS | DATA  | 5120  | 23Aug12:15:58:39 |
| 3 | COUNTRY     | DATA  | 17408 | 13Oct10:19:04:39 |
|   | COUNTRY     | INDEX | 17408 | 13Oct10:19:04:39 |
| 4 | CUSTOMER    | DATA  | 33792 | 04Nov11:09:52:27 |

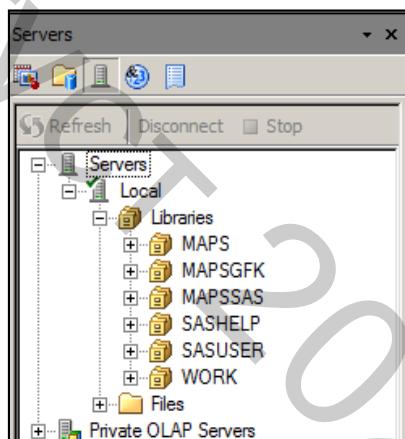
- g. Проверьте журнал, убедитесь, что ссылочное имя библиотеки **orion** было удалено.

```
15      libname orion clear;
NOTE: Libref ORION has been deassigned.
```

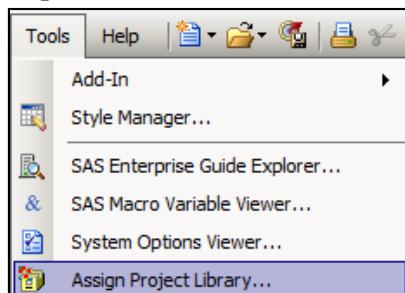
 Когда вы работаете с данными курса и программами, запускайте программу **libname** в начале каждого сеанса SAS. Она создаёт макропеременную **path** и определяет библиотеку **orion**.

### Создание и просмотр библиотеки в интерактивном режиме в SAS Enterprise Guide

- a. В области Servers (Серверы) выберите **Servers** ⇒ **Local** ⇒ **Libraries** (Серверы ⇒ Локальный ⇒ Библиотеки). Изучите библиотеки, которые автоматически назначаются в сеансе SAS Enterprise Guide.

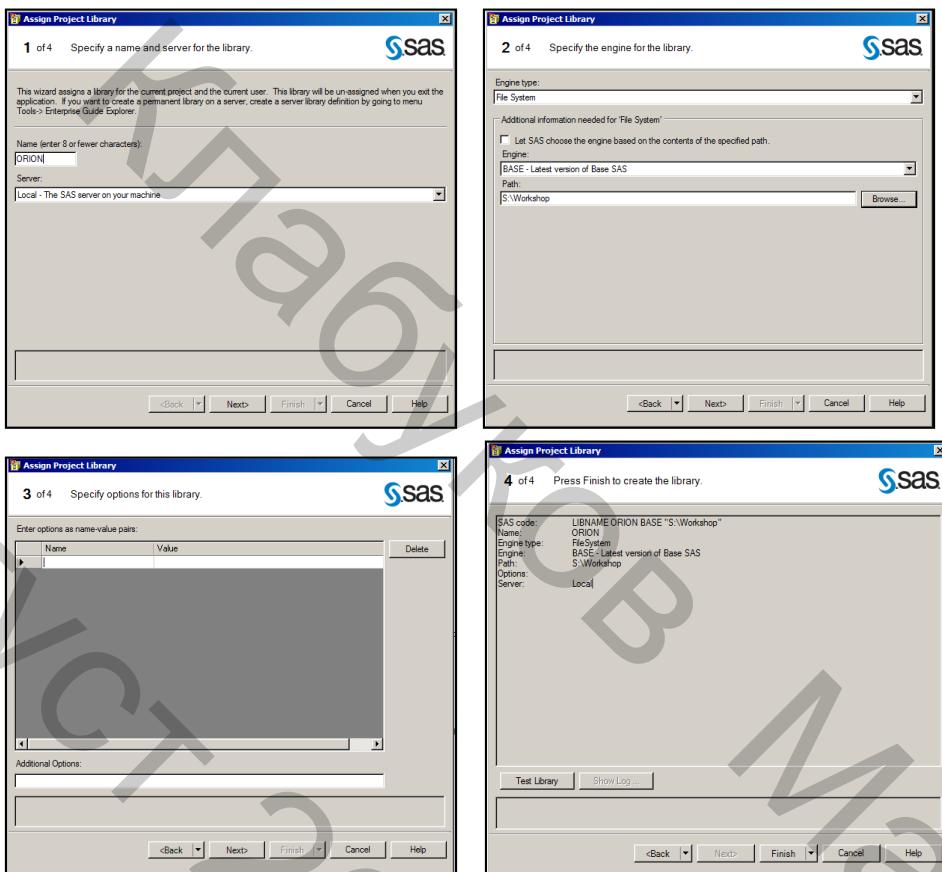


- b. Чтобы создать новую библиотеку, выберите в меню **Tools** ⇒ **Assign Project Library** (Сервис ⇒ Назначить библиотеку проекта).

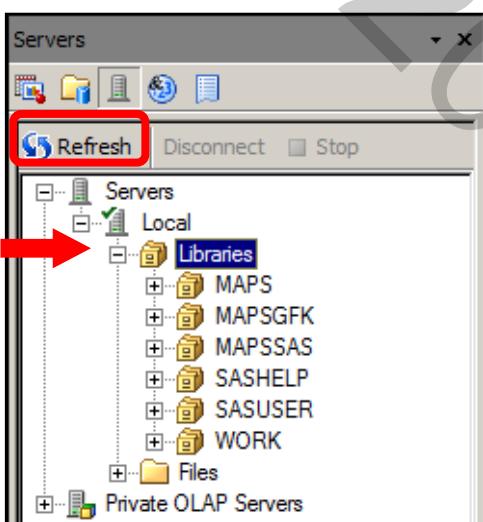


- c. В окне мастера настроек выполните следующие действия:
- Шаг 1: Введите **orion** в поле **Name** (Имя) и нажмите **Next** (Вперёд).
  - Шаг 2: Оставьте значения по умолчанию для механизма и его типа. Нажмите кнопку **Browse** (Обзор) и укажите место расположения данных курса (например, **s:\workshop**). Нажмите **OK**. Выводится выбранный вами путь. Нажмите **Next** (Вперёд).

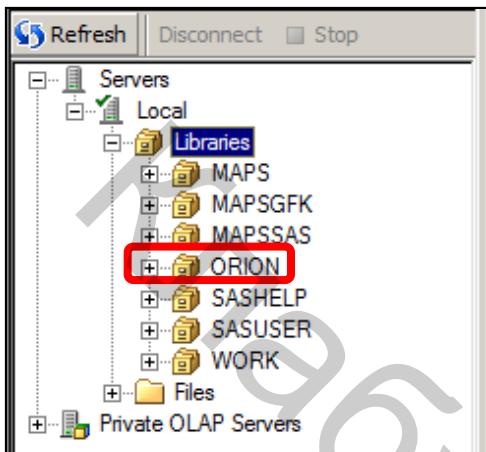
- Шаг 3: Нажмите **Next** (Вперёд).
- Шаг 4: Нажмите **Finish** (Готово).



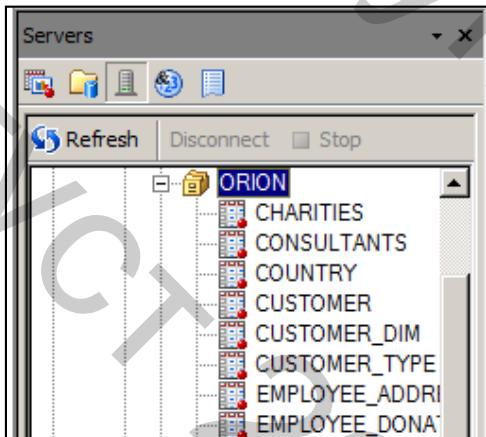
- d. В области Servers (Серверы) выберите **Libraries** (Библиотеки) и нажмите кнопку **Refresh** (Обновить), или нажмите правой кнопкой мыши на **Libraries** (Библиотеки) и выберите **Refresh** (Обновить).



- e. Проверьте, что библиотека **orion** содержится в списке активных библиотек.



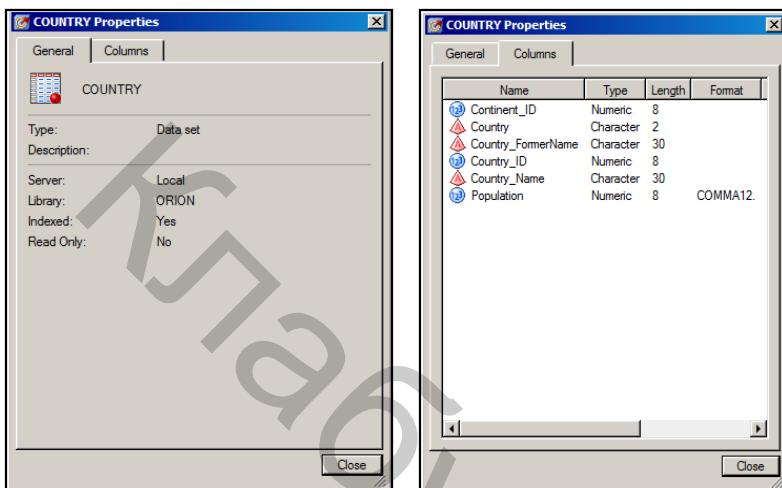
- f. Раскройте библиотеку **orion**.



- g. Откройте набор данных **COUNTRY**, дважды щёлкнув на него в библиотеке **orion**. Данные выводятся в сетке данных, которая открывается в рабочей области интерфейса. Наведите указатель мыши на название столбца и просмотрите его свойства. Нажмите **x**, чтобы закрыть сетку данных.

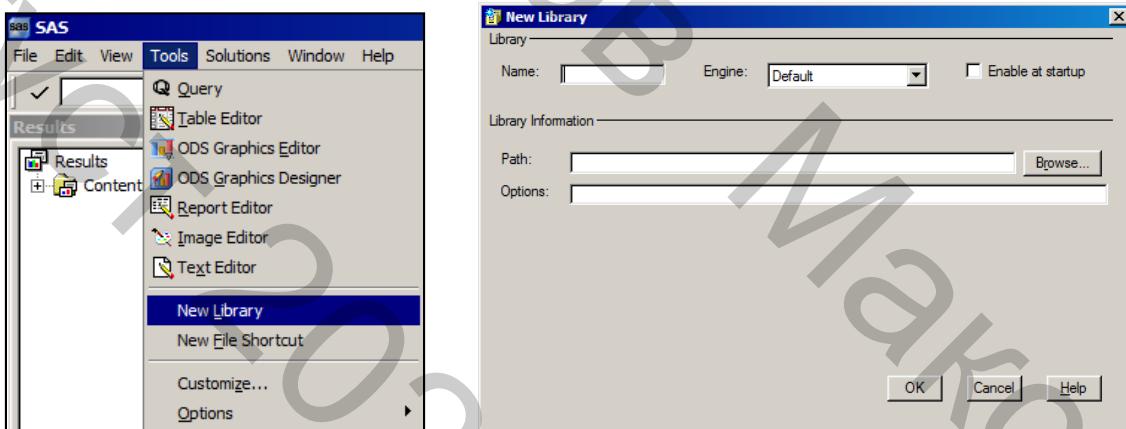
|   | Country | Country_Name  | Population  | Continent | Continent_ID | Country_ForeignName |
|---|---------|---------------|-------------|-----------|--------------|---------------------|
| 1 | AU      | Australia     | 20,000,000  |           | 96           |                     |
| 2 | CA      | Canada        |             |           | 91           |                     |
| 3 | DE      | Germany       | 80,000,000  | 394       | 93           | East/West Germany   |
| 4 | IL      | Israel        | 5,000,000   | 475       | 95           |                     |
| 5 | TR      | Turkey        | 70,000,000  | 905       | 95           |                     |
| 6 | US      | United States | 280,000,000 | 926       | 91           |                     |
| 7 | ZA      | South Africa  | 43,000,000  | 801       | 94           |                     |

- h. Нажмите правой кнопкой мыши на набор данных **COUNTRY** в библиотеке **orion** и выберите в меню пункт **Properties** (Свойства). Открывается вкладка **General** (Общие). Для просмотра свойств столбцов откройте вкладку **Columns** (Столбцы). Нажмите **Close** (Закрыть), чтобы закрыть диалог.

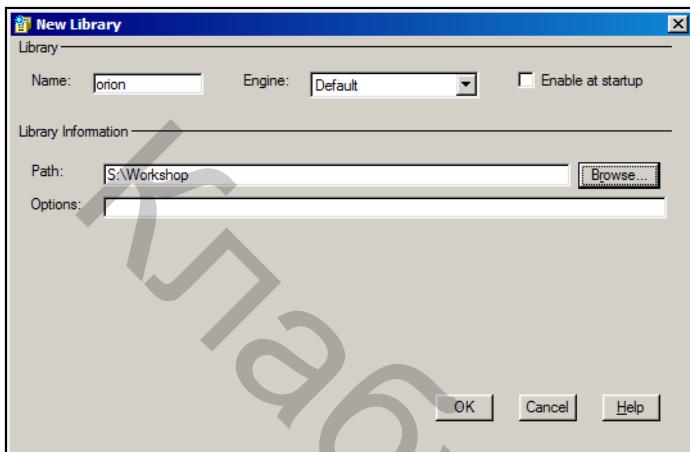


### Создание и просмотр библиотеки в интерактивном режиме в оконном интерфейсе SAS

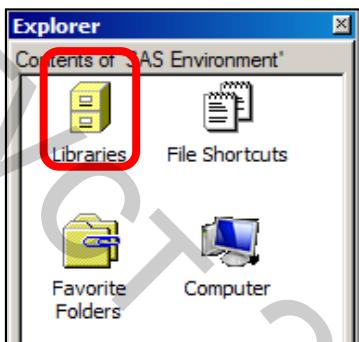
- a. Выберите в меню Tools  $\Rightarrow$  New Library или нажмите (Add New Library). Выводится диалоговое окно New Library.



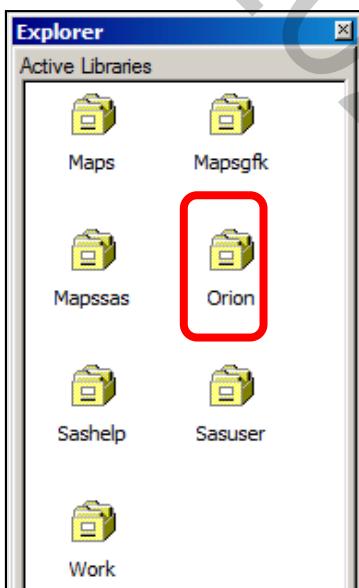
- b. Введите **orion** в поле Name. Нажмите кнопку **Browse** и укажите место расположения данных курса (например, **s:\workshop**). Нажмите **OK**. Путь к данным появляется в поле Path. Установка флашка **Enable at startup** позволяет пересоздавать ссылочное имя библиотеки в начале каждого сеанса SAS.



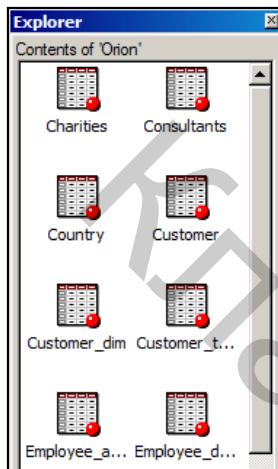
- c. Нажмите **OK**.
- d. В окне Explorer двойным щелчком откройте список библиотек, **Libraries**.



- e. Проверьте, что библиотека **Orion** выводится в списке активных библиотек.



- f. Дважды щёлкните библиотеку **Orion**. Проверьте, что выводятся наборы данных, содержащиеся в библиотеке **Orion**.



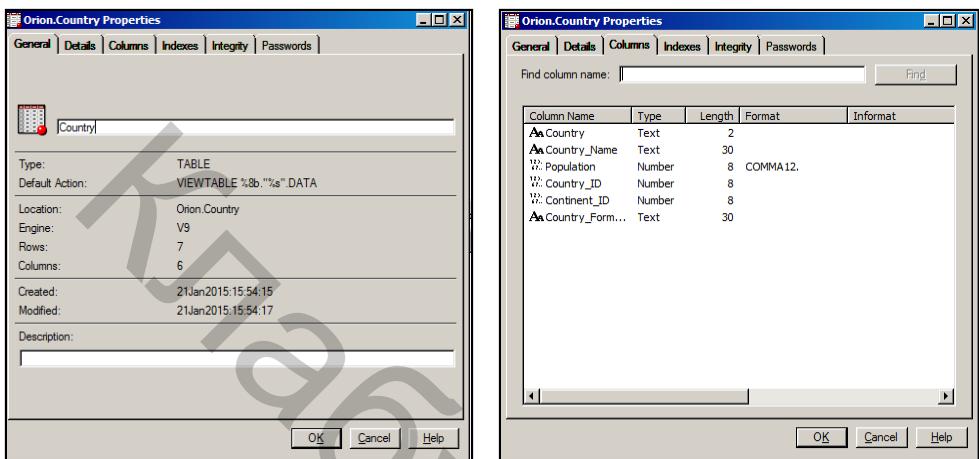
- g. Дважды щёлкните набор данных **Country** в библиотеке **Orion** чтобы открыть его. Данные отображаются в окне viewtable. Ярлыки столбцов выводятся в заголовках по умолчанию.

|   | Country Abbreviation | Current Name of Country | Population (approx.) | Country ID | Numeric Rep. for Continent | Former Name of Country |
|---|----------------------|-------------------------|----------------------|------------|----------------------------|------------------------|
| 1 | AU                   | Australia               | 20,000,000           | 160        | 96                         |                        |
| 2 | CA                   | Canada                  | .                    | 260        | 91                         |                        |
| 3 | DE                   | Germany                 | 80,000,000           | 394        | 93                         | East/West Germany      |
| 4 | IL                   | Israel                  | 5,000,000            | 475        | 95                         |                        |
| 5 | TR                   | Turkey                  | 70,000,000           | 905        | 95                         |                        |
| 6 | US                   | United States           | 280,000,000          | 926        | 91                         |                        |
| 7 | ZA                   | South Africa            | 43,000,000           | 801        | 94                         |                        |

- h. Выберите в меню **View** ⇒ **Column Names** для вывода в заголовке имён столбцов вместо ярлыков. Чтобы закрыть окно с таблицей, нажмите .

|   | Country | Country_Name  | Population  | Country_ID | Continent_ID | Country_ForeName  |
|---|---------|---------------|-------------|------------|--------------|-------------------|
| 1 | AU      | Australia     | 20,000,000  | 160        | 96           |                   |
| 2 | CA      | Canada        | .           | 260        | 91           |                   |
| 3 | DE      | Germany       | 80,000,000  | 394        | 93           | East/West Germany |
| 4 | IL      | Israel        | 5,000,000   | 475        | 95           |                   |
| 5 | TR      | Turkey        | 70,000,000  | 905        | 95           |                   |
| 6 | US      | United States | 280,000,000 | 926        | 91           |                   |
| 7 | ZA      | South Africa  | 43,000,000  | 801        | 94           |                   |

- i. Нажмите правой кнопкой мыши **на** значок таблицы **Country** в библиотеке **Orion** и выберите **Properties**. На вкладке **General** выводятся общие свойства набора данных. Откройте вкладку **Columns** для просмотра свойств столбцов. Чтобы закрыть диалог, нажмите  или **OK**.





## Упражнения. Создание набора данных SAS

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными.

### Создание набора данных SAS

- Откройте и запустите программу **p106e01**.

Какое имя у переменной, содержащей пол клиента? \_\_\_\_\_

Какие два значения принимает эта переменная? \_\_\_\_\_

- Добавьте шаг DATA перед шагом PROC PRINT для создания нового набора данных **work.youngadult**. В качестве входного набора данных используйте **orion.customer\_dim**. Добавьте оператор WHERE, чтобы из всех клиентов выбрать только женщин.

Выполните программу и убедитесь, что набор данных **work.youngadult** содержит 30 наблюдений и 11 переменных.

- Измените программу, выбрав клиентов женского пола с возрастом от 18 до 36 лет. Выполните программу и убедитесь, что набор данных **work.youngadult** содержит 15 наблюдений и 11 переменных.
- Измените эту программу, чтобы выбрать клиентов женского пола с возрастом от 18 до 36 лет, у которых в значениях **Customer\_Group** содержится слово *Gold*. Выполните программу и убедитесь, что **work.youngadult** содержит 5 наблюдений и 11 переменных.
- Добавьте оператор присваивания в шаг DATA для создания новой переменной **Discount**, и присвойте ей значение 0.25.
- Измените шаг PROC PRINT, чтобы распечатать новый набор данных как показано ниже. Используйте оператор ID, чтобы вывести столбец **Customer\_ID** вместо столбца Obs. Результаты должны содержать пять наблюдений.

| Customer_ID | Customer_Name      | Age | Customer_Gender | Customer_Group          | Discount |
|-------------|--------------------|-----|-----------------|-------------------------|----------|
| 5           | Sandrina Stephano  | 28  | F               | Orion Club Gold members | 0.25     |
| 9           | Cornelia Krahl     | 33  | F               | Orion Club Gold members | 0.25     |
| 45          | Dianne Patchin     | 28  | F               | Orion Club Gold members | 0.25     |
| 49          | Annmarie Leveille  | 23  | F               | Orion Club Gold members | 0.25     |
| 2550        | Sanelisiwe Collier | 19  | F               | Orion Club Gold members | 0.25     |



## Упражнения. Выбор наблюдения по двум условиям

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными.

### Выбор наблюдений по двум условиям

- Откройте программу **p106e04**.
- Измените шаг DATA для выбора только наблюдений, содержащих значения **Emp\_Hire\_Date** начиная с 1 июля 2010 года (включительно). Выберите эти наблюдения при заполнении PDV.
- На шаге DATA напишите другой оператор, выбирающий только наблюдения со значениями переменной **Increase**, превышающими 3000.
- Новый набор данных должен содержать только следующие переменные: **Employee\_ID**, **Emp\_Hire\_Date**, **Salary**, **Increase** и **NewSalary**.
- Добавьте постоянные ярлыки для **Employee\_ID**, **Emp\_Hire\_Date** и **NewSalary**, как показано в отчёте ниже.
- Добавьте постоянные форматы для вывода **Salary** и **NewSalary** со знаком доллара, запятыми и двумя десятичными разрядами, и переменной **Increase** с запятыми и без десятичных разрядов.
- Выполните шаг PROC CONTENTS и проверьте, что ярлыки и форматы хранятся в дескрипторе нового набора данных **work.increase**.

#### Фрагмент вывода PROC CONTENTS

| Alphabetic List of Variables and Attributes |               |      |     |            |          |                        |
|---------------------------------------------|---------------|------|-----|------------|----------|------------------------|
| #                                           | Variable      | Type | Len | Format     | Informat | Label                  |
| 3                                           | Emp_Hire_Date | Num  | 8   | DATE9.     | DATE9.   | Hire Date              |
| 1                                           | Employee_ID   | Num  | 8   | 12.        |          | Employee ID            |
| 4                                           | Increase      | Num  | 8   | COMMA5.    |          |                        |
| 5                                           | NewSalary     | Num  | 8   | DOLLAR10.2 |          | New Annual Salary      |
| 2                                           | Salary        | Num  | 8   | DOLLAR10.2 |          | Employee Annual Salary |

- Для некоторых переменных в дескрипторе появились ярлыки и форматы, которые не были определены в этой программе. Откуда они были взяты?
- Выполните программу и создайте отчёт PROC PRINT, показанный ниже. Разбейте ярлыки на несколько строк. Результаты должны содержать 10 наблюдений.

| Employee |             |               | New       |          |               |
|----------|-------------|---------------|-----------|----------|---------------|
| Obs      | Employee ID | Annual Salary | Hire Date | Increase | Annual Salary |
| 1        | 120128      | \$30,890.00   | 01NOV2010 | 3,089    | \$33,979.00   |
| 2        | 120144      | \$30,265.00   | 01OCT2010 | 3,027    | \$33,291.50   |
| 3        | 120161      | \$30,785.00   | 01OCT2010 | 3,079    | \$33,863.50   |
| ...      |             |               |           |          |               |
| 9        | 121085      | \$32,235.00   | 01JAN2011 | 3,224    | \$35,458.50   |
| 10       | 121107      | \$31,380.00   | 01JUL2010 | 3,138    | \$34,518.00   |



## Упражнения. Создание новых переменных

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

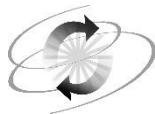
### Создание новых переменных

- Откройте программу **p109e01**.
- В шаге DATA создайте три новых переменных:
  - Increase**, равную **Salary**, умноженной на 0.10
  - NewSalary**, равную **Salary**, сложенной с **Increase**
  - BdayQtr**, равную кварталу от даты рождения сотрудника
- Новый набор данных должен включать только **Employee\_ID**, **Salary**, **Birth\_Date** и три новые переменные.
- В дескриптор набора данных сохраните форматы для вывода переменных **Salary**, **Increase** и **NewSalary** с запятыми, отделяющими в этих числах по три разряда.
- Измените программу для создания отчёта, показанного ниже, в том числе указанных ярлыков. Результаты должны содержать 424 наблюдения.

Фрагмент вывода PROC PRINT

| Obs | Employee    |                  | Employee<br>Birth Date | Increase | NewSalary | Bday<br>Qtr |
|-----|-------------|------------------|------------------------|----------|-----------|-------------|
|     | Employee ID | Annual<br>Salary |                        |          |           |             |
| 1   | 120101      | 163,040          | 18AUG1980              | 16,304   | 179,344   | 3           |
| 2   | 120102      | 108,255          | 11AUG1973              | 10,826   | 119,081   | 3           |
| 3   | 120103      | 87,975           | 22JAN1953              | 8,798    | 96,773    | 1           |
| 4   | 120104      | 46,230           | 11MAY1958              | 4,623    | 50,853    | 2           |

## Упражнения. Использование условной обработки



Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### Использование условной обработки

- Откройте программу **p109e04**.
- В шаге DATA создайте новую переменную **Method**, присвойте ей одно из значений в зависимости от **Order\_Type**.

Если переменная **Order\_Type** равна 1, **Method** должен быть равен *Retail*.

Если переменная **Order\_Type** равна 2, **Method** должен быть равен *Catalog*.

Если переменная **Order\_Type** равна 3, **Method** должен быть равен *Internet*.

Для любых других значений **Order\_Type**, **Method** должен принимать значение *Unknown*.

- Измените шаг PROC PRINT для вывода отчёта, показанного ниже. Результаты должны содержать 490 наблюдений.

Фрагмент вывода PROC PRINT

| Obs | Order_ID   | Order_Type | Method  |
|-----|------------|------------|---------|
| 1   | 1230058123 | 1          | Retail  |
| 2   | 1230080101 | 2          | Catalog |
| 3   | 1230106883 | 2          | Catalog |
| 4   | 1230147441 | 1          | Retail  |
| 5   | 1230315085 | 1          | Retail  |

### Использование условной обработки с группами DO

- Откройте программу **p109e05**.
- Измените шаг DATA для создания трёх новых переменных: **Discount**, **DiscountType** и **Region**. Присвойте значения новым переменным на основании значений **Country**.

Если переменная **Country** равна *CA* или *US*, тогда указанные переменные должны принимать значения:

**Discount** значение 0.10.

**DiscountType** значение *Required*.

**Region** значение *North America*.

Если **Country** имеет любое другое значение, то эти переменные должны принимать значения:

**Discount** значение 0.05.

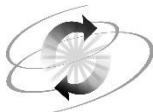
**DiscountType** значение *Optional*.

**Region** значение *Not North America*.

- c. Новый набор данных должен включать только переменные **Supplier\_Name**, **Country**, **Discount**, **DiscountType** и **Region**.
- d. Выполните программу и создайте отчёт, показанный ниже. Результаты должны содержать 52 наблюдения.

Фрагмент вывода PROC PRINT

| Obs | Supplier_Name             | Country | Region            | Discount | Discount Type |
|-----|---------------------------|---------|-------------------|----------|---------------|
| 1   | Scandinavian Clothing A/S | NO      | Not North America | 0.05     | Optional      |
| 2   | Petterson AB              | SE      | Not North America | 0.05     | Optional      |
| 3   | Prime Sports Ltd          | GB      | Not North America | 0.05     | Optional      |



## Упражнения. Подсчет уровней переменной

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### Уровень 1

#### 1. Подсчёт уровней переменной при помощи PROC FREQ

- Откройте программу **p111e01**.
- Запустите исходную версию программы для анализа переменных **Customer\_ID** и **Employee\_ID** в наборе данных **orion.orders**. Как вы думаете, будут ли выведены единичные частоты для идентификаторов клиентов (**Customer\_ID**) и сотрудников (**Employee\_ID**)?
- Измените программу так, чтобы создать два отдельных отчёта.
  - Выведите количество различных уровней **Customer\_ID** и **Employee\_ID** для розничных заказов.
    - Используйте оператор WHERE, чтобы ограничить информацию в отчёте только розничными продажами (**Order\_Type=1**).
    - Отключите вывод таблиц с частотами.
    - Задайте для отчёта заголовок: **Unique Customers and Salespersons for Retail Sales**.
    - Выполните программу и получите следующий отчёт:

#### PROC FREQ Output

| Unique Customers and Salespersons for Retail Sales |                    |        |
|----------------------------------------------------|--------------------|--------|
| The FREQ Procedure                                 |                    |        |
| Number of Variable Levels                          |                    |        |
| Variable                                           | Label              | Levels |
| <b>Customer_ID</b>                                 | <b>Customer ID</b> | 31     |
| <b>Employee_ID</b>                                 | <b>Employee ID</b> | 100    |

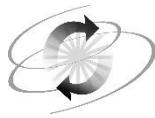
- Выведите количество различных уровней **Customer\_ID** для заказов по каталогу и через интернет.
  - Используйте оператор WHERE, чтобы проанализировать в отчёте эти заказы, отвечающие любым значениям **Order\_Type**, кроме 1.
  - Укажите параметр для вывода результатов, упорядоченных по убыванию частот.
  - Укажите параметр, чтобы отключить вывод кумулятивных статистик.
  - Выведите в отчёте заголовок: **Catalog and Internet Customers**.

- e) Выполните программу и получите следующий отчёт:

Фрагмент вывода PROC FREQ

| Catalog and Internet Customers |           |         |
|--------------------------------|-----------|---------|
| The FREQ Procedure             |           |         |
| Customer ID                    |           |         |
| Customer_ID                    | Frequency | Percent |
| 16                             | 15        | 6.52    |
| 29                             | 9         | 3.91    |
| 5                              | 8         | 3.48    |
| ...                            |           |         |
| 26148                          | 1         | 0.43    |

## Упражнения. Создание сводного отчета



Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### 2. Создание сводного отчёта при помощи PROC MEANS

- Откройте программу **p111e06**.
- Выведите в отчёт только статистику SUM для переменной **Total\_Retail\_Price**.
- Добавьте оператор CLASS для вывода значений этой статистики для каждой комбинации переменных **Order\_Date** и **Order\_Type** по отдельности.
- Примените формат ORDERTYPES, чтобы описание типов заказов отображалось в виде текста. Примените формат YEAR4. для вывода дат заказов в виде года.
- Выполните программу и получите следующий отчёт:

Фрагмент вывода PROC MEANS

| Revenue from All Orders                                                    |               |          |          |
|----------------------------------------------------------------------------|---------------|----------|----------|
| The MEANS Procedure                                                        |               |          |          |
| Analysis Variable : Total_Retail_Price Total Retail Price for This Product |               |          |          |
| Date<br>Order<br>was<br>placed<br>by<br>Customer                           | Order<br>Type | N<br>Obs | Sum      |
| 2007                                                                       | Retail        | 53       | 7938.80  |
|                                                                            | Catalog       | 52       | 10668.08 |
|                                                                            | Internet      | 23       | 4124.05  |
|                                                                            | Retail        | 63       | 9012.22  |
|                                                                            | Catalog       | 23       | 3494.60  |
|                                                                            | Internet      | 22       | 3275.70  |

## Приложение (День 2)

Для решения указанных ниже задач используйте папку additional. При необходимости воспользуйтесь справочной документацией (<https://support.sas.com/en/documentation.html>)

### Задача 1.

Изучите набор данных **products**.

Распечатайте ID, сгруппированные по годам, используя встроенный формат SAS и процедуру PROC PRINT. Предполагаем, что набор данных отсортирован по переменной Order\_Date.

### Задача 2\*.

Переменная Recipients в наборе данных **donations** содержит названия благотворительных организаций и процент отчисления сотрудником пожертвований:

| Employee_ID | Recipients                                            |
|-------------|-------------------------------------------------------|
| 120265      | Mittleid International 90%, Save the Baby Animals 10% |
| 120267      | Disaster Assist, Inc. 80%, Cancer Cures, Inc. 20%     |
| 120269      | Cancer Cures, Inc. 10%, Cuidadores Ltd. 90%           |
| 120270      | AquaMissions International 10%, Child Survivors 90%   |
| 120271      | Cuidadores Ltd. 80%, Mittleid International 20%       |
| 120272      | AquaMissions International 10%, Child Survivors 90%   |
| 120275      | AquaMissions International 60%, Child Survivors 40%   |

Используя функции SAS BASE, создайте следующие переменные:

|          |                                                            |
|----------|------------------------------------------------------------|
| Charity1 | Название первой благотворительной организации              |
| Percent1 | Процент пожертвований первой благотворительной организации |
| Charity2 | Название второй благотворительной организации              |
| Percent2 | Процент пожертвований второй благотворительной организации |

Преобразуйте переменные Percent1 и Percent2 в числовой тип и примените соответствующий формат.

Итоговый набор данных должен выглядеть следующим образом:

| Employee_ID | Recipients                                            | Charity1                   | Percent1 | Charity2               | Percent2 |
|-------------|-------------------------------------------------------|----------------------------|----------|------------------------|----------|
| 120265      | Mittleid International 90%, Save the Baby Animals 10% | Mittleid International     | 90%      | Save the Baby Animals  | 10%      |
| 120267      | Disaster Assist, Inc. 80%, Cancer Cures, Inc. 20%     | Disaster Assist, Inc.      | 80%      | Cancer Cures, Inc.     | 20%      |
| 120269      | Cancer Cures, Inc. 10%, Cuidadores Ltd. 90%           | Cancer Cures, Inc.         | 10%      | Cuidadores Ltd.        | 90%      |
| 120270      | AquaMissions International 10%, Child Survivors 90%   | AquaMissions International | 10%      | Child Survivors        | 90%      |
| 120271      | Cuidadores Ltd. 80%, Mittleid International 20%       | Cuidadores Ltd.            | 80%      | Mittleid International | 20%      |
| 120272      | AquaMissions International 10%, Child Survivors 90%   | AquaMissions International | 10%      | Child Survivors        | 90%      |
| 120275      | AquaMissions International 60%, Child Survivors 40%   | AquaMissions International | 60%      | Child Survivors        | 40%      |

Фрагмент дескриптора выходного набора данных:

| Alphabetic List of Variables and Attributes |             |      |       |          |             |
|---------------------------------------------|-------------|------|-------|----------|-------------|
| #                                           | Variable    | Type | Len   | Format   | Label       |
| 3                                           | Charity1    | Char | 65    |          |             |
| 5                                           | Charity2    | Char | 65    |          |             |
| 1                                           | Employee_ID | Num  | 8 12. |          | Employee ID |
| 4                                           | Percent1    | Num  | 8     | PERCENT. |             |
| 6                                           | Percent2    | Num  | 8     | PERCENT. |             |
| 2                                           | Recipients  | Char | 65    |          |             |

### Задача 3.

Набор данных orders содержит информацию по заказам за 2010-2014 годы. Используя указанный набор данных, создайте новые наборы по каждому году, включающие заказы со сроком доставки от 1 дня.

### Задача 4.

Таблица **activeemployees** содержит информацию о текущих сотрудниках Orion Star. Пакет досрочного выхода на пенсию будет предложен отдельным сотрудникам, сумма возраста и срока службы (в годах) которых на 1 января 2015 года превышает значение 90 или равняется ему.

Создайте таблицу с именем **Retirement** и включите в результаты столбцы **Employee\_ID**, **Name** и **Department**. Также вычислите три новых столбца:

- Столбец **Age**, в котором вычисляется возраст сотрудников в годах на 1 января 2015 года
- Столбец **YearsOfService**, в котором вычисляется число лет их службы на 1 января 2015 года
- Столбец **RetirementGroup**, значения в котором определяются на основании суммы этих двух значений^
  - Eligible ( $\geq 90$ )
  - Non-eligible ( $< 90$ )

 Столбцы **Age** и **YearsOfService** должны содержать целые значения. Например, если возраст сотрудника составляет 34.5, столбец должен содержать число 34.

Подсказка. Для расчета разницы в годах между двумя датами можно использовать функцию YRDIF.

Подсказка. Для возврата целой части числа можно использовать функцию INT.

### Задача 5.

Используя таблицу **employee\_addresses**, создайте отчет и выведите столбцы Employee\_Name, Street\_Number, Street\_Name, City и Postal\_Code. Для каждой строки вместо её номера укажите значение столбца Employee\_ID. Измените свойства столбцов так, чтобы в отчете отображались следующие подписи столбцов: Employee ID, Name, Street Number, Street Name, City и Postal Code.

### Задача 6.

Используя набор данных **employee\_addresses** создайте отчет, содержащий всех сотрудников из Австралии (в разных регистрах). Выведите столбцы Employee\_Name, Street\_Number, Street\_Name, City и Postal\_Code. Для каждой строки вместо её номера укажите значение столбца Employee\_ID.

### Задача 7.

Используя таблицу **employee\_payroll**, создайте новый набор данных, содержащий в себе столбцы **Employee\_ID** и **Employee\_Hire\_Date**. А также новый столбец **Salary\_Range**, который основан на значениях столбца **Salary**.

|                       |                 |
|-----------------------|-----------------|
| 0 – 25000 (искл.)     | Below \$25K     |
| 25000 – 50000 (искл.) | \$25K to \$50K  |
| 50000 – 100000(искл.) | \$50K to \$100K |
| 100000 и больше       | Over \$100K     |

### Задача 8\*.

Столбец **Customer\_Type** в таблице **Customer\_dim** содержит информацию об участии в программе Orion Club и уровне активности. Используйте его для распределения клиентов по группам High, Medium и Low. Создайте новую таблицу под названием **ActivityLevel**. Включите в новую таблицу столбцы **Customer\_ID**, **Customer\_Name** и **Customer\_Type**, а также новый столбец с именем **ActivityLevel**.

В столбце **ActivityLevel** должно содержаться значение **High**, если в столбце **Customer\_Type** присутствует слово «high», значение **Medium**, если в столбце **Customer\_Type** есть слово «medium», значение **Low**, если в столбце **Customer\_Type** указано слово «low»; в противном случае значение в этом столбце должно отсутствовать.

Клабуков Максим  
Август 2020

## День 3. SAS(R) Base: Data Preparation



### Изучение ошибок данных

p108d04

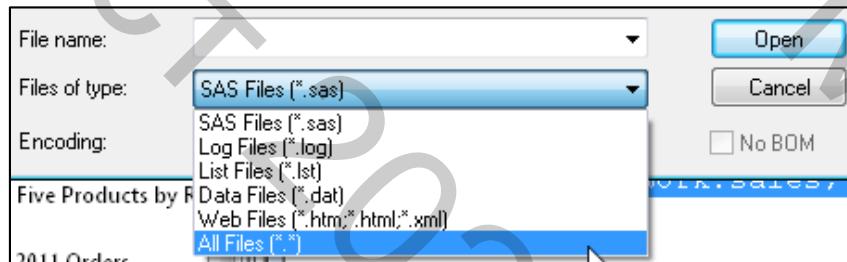
```
data work.sales;
  infile "&path\sales3inv.csv" dlm=',';
  input Employee_ID First $ Last $ 
        Job_Title $ Salary Country $;
run;

proc print data=work.sales;
run;
```

18. Откройте текстовый файл с данными **sales3inv.csv**.

В SAS Enterprise Guide выберите **File** ⇒ **Open** ⇒ **Data** (Файл ⇒ Открыть ⇒ Данные).  
Перейдите в директорию с данными.

В оконном интерфейсе SAS выберите **File** ⇒ **Open Program**. Выберите **All Files (\*.\*)** в списке типов файлов (File of type).



Выберите **sales3inv.csv** из списка файлов.

19. Просмотрите значения **Salary**. Он содержит три недопустимых значения и одно пропущенное.
20. Откройте и выполните программу **p108d04**.
21. Изучите журнал. Шаг DATA создаёт выходной набор данных **work.sales**, содержащий 50 наблюдений.  
Три недопустимых значения **Salary** были обозначены в журнале. Соответствующие им наблюдения содержат пропущенные значения в переменной **Salary**. Пропущенное значение **Salary** не вызвало появление ошибки данных, поскольку *пропущенное* значение в SAS является допустимым.
22. Изучите вывод. Обратите внимание, что имеются четыре наблюдения с пропущенными значениями переменной **Salary**.

## Упражнения. Чтение текстового файла с разделителем



Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### Чтение текстового файла с разделителем - запятой

- a. Откройте программу **p108e01**. Добавьте необходимые операторы LENGTH, INFILE и INPUT для чтения файла с разделителем - запятой, имя которого указано ниже:

|               |                           |
|---------------|---------------------------|
| Windows       | "&path\newemps.csv"       |
| UNIX          | "&path/newemps.csv"       |
| z/OS (OS/390) | "&path..rawdata(newemps)" |

Фрагмент исходного текстового файла с данными

```
Satyakam,Denny,Sales Rep. II,26780
Monica,Kletschkus,Sales Rep. IV,30890
Kevin,Lyon,Sales Rep. I,26955
Petrea,Soltau,Sales Rep. II,27440
Marina,Iyengar,Sales Rep. III,29715
```

- b. Прочитайте следующие поля из текстового файла с данными:

| Название | Тип       | Длина |
|----------|-----------|-------|
| First    | Текстовый | 12    |
| Last     | Текстовый | 18    |
| Title    | Текстовый | 25    |
| Salary   | Числовой  | 8     |

- c. Выполните программу и создайте отчёт, показанный ниже. Результаты должны содержать 71 наблюдение.

Фрагмент вывода PROC PRINT

| Obs | First    | Last       | Title          | Salary |
|-----|----------|------------|----------------|--------|
| 1   | Satyakam | Denny      | Sales Rep. II  | 26780  |
| 2   | Monica   | Kletschkus | Sales Rep. IV  | 30890  |
| 3   | Kevin    | Lyon       | Sales Rep. I   | 26955  |
| 4   | Petrea   | Soltau     | Sales Rep. II  | 27440  |
| 5   | Marina   | Iyengar    | Sales Rep. III | 29715  |



## Использование списочного ввода: важность модификатора "двоеточие"

### p108a02

- Откройте программу **p108a02** и изучите оператор INPUT.
  - Оператор INFILE не содержит параметр DLM= так как поля в файле разделены пробелами.
  - HireDate** и **Salary** содержат нестандартные числовые данные, и для их чтения необходимо указать формат ввода (informat).
- В оконном окружении SAS выберите **File ⇄ Open Program**, измените значение в списке **Files of type** на **Data Files (\*.dat)**, и выберите файл **salary.dat**.
 

Файлы с расширением .dat можно импортировать в SAS Enterprise Guide, но не просматривать во встроенным редакторе.

Фрагмент файла salary.dat

```
Donny 5MAY2008 25 FL $43,132.50
Margaret 20FEB2008 43 NC 65,150
Dan 1JUN2008 27 FL $40,000.00
Subash 2FEB2008 45 NC 75,750
Antonio 25MAY2008 35 FL $43,500.50
```

- Выполните первую часть (Part 1) и изучите журнал и вывод. Ожидаемый вывод показан ниже.

```
/* Part 1 - using colon format modifiers*/
data work.salaries;
  infile "&path\salary.dat";
  input Name $ HireDate :date. Age State $ Salary :comma.;
run;

proc print data=work.salaries;
run;
```

Фрагмент вывода PROC PRINT

| Obs | Name     | Hire Date | Age | State | Salary  |
|-----|----------|-----------|-----|-------|---------|
| 1   | Donny    | 17657     | 25  | FL    | 43132.5 |
| 2   | Margaret | 17582     | 43  | NC    | 65150.0 |
| 3   | Dan      | 17684     | 27  | FL    | 40000.0 |
| 4   | Subash   | 17564     | 45  | NC    | 75750.0 |
| 5   | Antonio  | 17677     | 35  | FL    | 43500.5 |

- Проверьте, что будет, если пропустить двоеточие при указании формата ввода для переменной **Salary**. Выполните часть 2 (Part 2).

```
/* Part 2 - omit the colon format modifier for Salary */
data work.salaries;
  infile "&path\salary.dat";
  input Name $ HireDate :date. Age State $ Salary comma.;
```

```

run;

proc print data=work.salaries;
run;

```

5. Изучите журнал. Он не содержит ни ошибок, ни предупреждений.

```

923 /* Part 2 \ omit the colon format modifier for Salary */
924 data work.salaries;
925     infile "&path\salary.dat";
926     input Name $ HireDate :date. Age State $ Salary comma.;
927 run;

NOTE: The infile "s:\workshop\salary.dat" is:
      Filename=s:\workshop\salary.dat,
NOTE: 8 records were read from the infile "s:\workshop\salary.dat".
NOTE: The data set WORK.SALARIES has 8 observations and 5 variables.

```

6. Изучите вывод.

| Obs | Name     | Date  | Age | State | Salary |
|-----|----------|-------|-----|-------|--------|
| 1   | Donny    | 17657 | 25  | FL    | .      |
| 2   | Margaret | 17582 | 43  | NC    | 6      |
| 3   | Dan      | 17684 | 27  | FL    | .      |
| 4   | Subash   | 17564 | 45  | NC    | 7      |
| 5   | Antonio  | 17677 | 35  | FL    | .      |

Значения переменной **Salary** считаны неверно. Почему эти значения или пропущены или содержат всего одну цифру?

- Формат ввода COMMA. имеет ширину, равную 1 по умолчанию, поэтому SAS считывает один столбец из входного файла.  
Таким образом, считывается только первый символ из значения переменной **Salary**.
- Когда первый столбец содержит символ доллара, числовой переменной присваивается пропущенное значение. Это не вызывает ошибку данных поскольку формат ввода COMMA. удаляет из входных данных нечисловые символы, в том числе символ доллара. Когда первый столбец содержит цифру, именно она и присваивается переменной.



В некоторых случаях пропуск "двоеточия" вызывает появление пропущенных или недопустимых значений. В других случаях - ошибки данных.



## Упражнения

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### Чтение нестандартных данных из файла с разделителем - запятой

- a. Откройте программу **p108e04**. Добавьте необходимые операторы LENGTH, INFILE и INPUT для чтения текстового файла с разделителем - запятой, имя которого указано ниже:

|               |                          |
|---------------|--------------------------|
| Windows       | "&path\custca.csv"       |
| UNIX          | "&path/custca.csv"       |
| z/OS (OS/390) | "&path..rawdata(custca)" |

Фрагмент исходного текстового файла с данными

```
Bill,Cuddy,11171,M,16/10/1986,21,15-30 years
Susan,Krasowski,17023,F,09/07/1959,48,46-60 years
Andreas,Rennie,26148,M,18/07/1934,73,61-75 years
Lauren,Krasowski,46966,F,24/10/1986,21,15-30 years
Lauren,Marx,54655,F,18/08/1969,38,31-45 years
```

Прочитайте из файла поля в соответствии со списком ниже:

| Название  | Тип       | Длина |
|-----------|-----------|-------|
| First     | Текстовый | 20    |
| Last      | Текстовый | 20    |
| ID        | Числовой  | 8     |
| Gender    | Текстовый | 1     |
| BirthDate | Числовой  | 8     |
| Age       | Числовой  | 8     |
| AgeGroup  | Текстовый | 12    |

- b. Используйте операторы FORMAT и DROP на шаге DATA, чтобы создать набор данных, на основании которого при помощи PROC PRINT создаётся отчёт, показанный ниже. Добавьте в отчёт требуемый заголовок. Результаты должны содержать 15 наблюдений.

Фрагмент вывода PROC PRINT

| Canadian Customers |         |           |        |             |            |
|--------------------|---------|-----------|--------|-------------|------------|
| Obs                | First   | Last      | Gender | AgeGroup    | Birth Date |
| 1                  | Bill    | Cuddy     | M      | 15-30 years | OCT 1986   |
| 2                  | Susan   | Krasowski | F      | 46-60 years | JUL 1959   |
| 3                  | Andreas | Rennie    | M      | 61-75 years | JUL 1934   |
| 4                  | Lauren  | Krasowski | F      | 15-30 years | OCT 1986   |
| 5                  | Lauren  | Marx      | F      | 31-45 years | AUG 1969   |



## Упражнения. Чтение текстового файла с пропущенными значениями

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### Чтение текстового файла с разделителем - запятой с пропущенными значениями

- a. Откройте программу **p108e07**. Добавьте в неё операторы INFILE и INPUT, чтобы считать данные из текстового файла с разделителем - запятой, имя которого указано ниже:

|               |                            |
|---------------|----------------------------|
| Windows       | "&path\donation.csv"       |
| UNIX          | "&path/donation.csv"       |
| z/OS (OS/390) | "&path..rawdata(donation)" |

Фрагмент исходного текстового файла с данными

```
120265, , , ,25
120267,15,15,15,15
120269,20,20,20,20
120270,20,10,5
120271,20,20,20,20
```

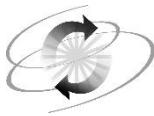
- b. В середине или в конце записи могут быть пропущенные данные. Прочитайте следующие поля из текстового файла с данными:

| Название | Тип      |
|----------|----------|
| EmpID    | Числовой |
| Q1       | Числовой |
| Q2       | Числовой |
| Q3       | Числовой |
| Q4       | Числовой |

- c. Выполните шаг PROC PRINT для создания отчёта, показанного ниже. Результаты должны содержать 124 наблюдения.

| Obs | EmpID  | Q1 | Q2 | Q3 | Q4 |
|-----|--------|----|----|----|----|
| 1   | 120265 | .  | .  | .  | 25 |
| 2   | 120267 | 15 | 15 | 15 | 15 |
| 3   | 120269 | 20 | 20 | 20 | 20 |
| 4   | 120270 | 20 | 10 | 5  | .  |
| 5   | 120271 | 20 | 20 | 20 | 20 |

## Упражнения. Конкатенация наборов данных.



Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### Конкатенация наборов данных с похожими структурами

- a. Напишите и выполните шаг DATA, в котором объединяются **orion.mnth7\_2011**, **orion.mnth8\_2011** и **orion.mnth9\_2011**, и создаётся новый набор данных **work.thirdqtr**.

Сколько наблюдений добавлено в **work.thirdqtr** из **orion.mnth7\_2011**? \_\_\_\_\_

Сколько наблюдений добавлено в **work.thirdqtr** из **orion.mnth8\_2011**? \_\_\_\_\_

Сколько наблюдений добавлено в **work.thirdqtr** из **orion.mnth9\_2011**? \_\_\_\_\_

- b. Напишите шаг PROC PRINT, создающий отчёт ниже. Результаты должны содержать 32 наблюдения.

Фрагмент вывода PROC PRINT

| Obs | Order_ID   | Type | Employee_ID | Customer_ID | Order_Date | Delivery_Date |
|-----|------------|------|-------------|-------------|------------|---------------|
| 1   | 1242691897 | 2    | 99999999    | 90          | 02JUL2011  | 04JUL2011     |
| 2   | 1242736731 | 1    | 121107      | 10          | 07JUL2011  | 07JUL2011     |
| 3   | 1242773202 | 3    | 99999999    | 24          | 11JUL2011  | 14JUL2011     |
| 4   | 1242782701 | 3    | 99999999    | 27          | 12JUL2011  | 17JUL2011     |
| 5   | 1242827683 | 1    | 121105      | 10          | 17JUL2011  | 17JUL2011     |

### Конкатенация данных с непохожими структурами

Откройте программу **p110e02**. Выполните два шага PROC CONTENTS или исследуйте указанные в них наборы данных интерактивно. Сравните переменные в наборах данных. Как называются две переменные, названия которых отличаются в этих наборах данных?

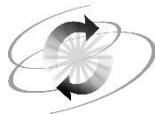
**orion.sales**

**orion.nonsales**

- a. После PROC CONTENTS добавьте шаг DATA, в котором данные из **orion.sales** и **orion.nonsales** объединяются в новый набор данных **work.allemployees**.
- Для изменения названий отличающихся переменных используйте параметр набора данных RENAME= в **orion.nonsales**.
- Новый набор данных должен содержать только **Employee\_ID**, **First\_Name**, **Last\_Name**, **Job\_Title** и **Salary**.
- b. Добавьте в программу шаг PROC PRINT, чтобы создать отчёт, указанный ниже. Результаты должны содержать 400 наблюдений.

## Фрагмент вывода PROC PRINT

| Obs | Employee_ID | First_Name | Last_Name | Salary | Job_Title     |
|-----|-------------|------------|-----------|--------|---------------|
| 1   | 120102      | Tom        | Zhou      | 108255 | Sales Manager |
| 2   | 120103      | Wilson     | Dawes     | 87975  | Sales Manager |
| 3   | 120121      | Irenie     | Elvish    | 26600  | Sales Rep. II |
| 4   | 120122      | Christina  | Ngan      | 27475  | Sales Rep. II |
| 5   | 120123      | Kimiko     | Hotstone  | 26190  | Sales Rep. I  |



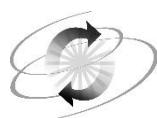
## Упражнения. Слияние наборов данных.

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### Слияние двух отсортированных наборов данных типа "один ко многим"

- Откройте программу **p110e04**.
- Выполните два шага PROC CONTENTS или исследуйте указанные в них наборы данных интерактивно, чтобы определить общую переменную, которая будет использоваться для их слияния.
- Добавьте шаг DATA после двух шагов PROC CONTENTS для слияния **orion.orders** и **orion.order\_item** по общей переменной в новый набор данных **work.allorders**. Предварительная сортировка исходных наборов данных не требуется, поскольку они уже сортированы по общей переменной.
- Выполните программу и убедитесь, что набор данных **work.allorders** имеет 732 наблюдений и 12 переменных.
- Добавьте оператор для выбора переменных. Новый набор данных должен содержать шесть переменных: **Order\_ID**, **Order\_Item\_Num**, **Order\_Type**, **Order\_Date**, **Quantity** и **Total\_Retail\_Price**.
- Напишите шаг PROC PRINT для создания отчёта. Выведите только такие наблюдения, где значения **Order\_Date** относятся к четвёртому кварталу 2011 года. Результаты должны содержать 35 наблюдений.

| Order_ID   | Order_Type | Order_Date | Order_Item_Num | Quantity | Total_Retail_Price |
|------------|------------|------------|----------------|----------|--------------------|
| 1243515588 | 1          | 01OCT2011  | 1              | 1        | \$251.80           |
| 1243515588 | 1          | 01OCT2011  | 2              | 1        | \$114.20           |
| 1243568955 | 1          | 07OCT2011  | 1              | 1        | \$172.50           |
| 1243643970 | 1          | 16OCT2011  | 1              | 1        | \$101.50           |
| 1243644877 | 3          | 16OCT2011  | 1              | 1        | \$14.60            |



## Упражнения. Использование параметра IN=

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу **libname**, которая находится в директории с данными курса.

### Слияние с использованием параметра IN=

- Откройте программу **p110e07**. Добавьте в неё шаг DATA после шага PROC SORT и проведите слияние **work.product** и **orion.supplier** по переменной **Supplier\_ID**, создав новый набор данных **work.prodsup**.
- Выполните программу и убедитесь, что **work.prodsup** содержит 556 наблюдений и 10 переменных.
- Измените шаг DATA, чтобы выводились только такие наблюдения, куда даёт вклад **work.product**, и не даёт вклад **orion.supplier**. В него нужно добавить выборочный оператор IF, в котором проверяются значения переменных, создающихся параметром IN= в операторе MERGE.
- Выполните программу и убедитесь, что **work.prodsup** содержит 75 наблюдений и 10 переменных.
- Выполните шаг PROC PRINT для создания отчёта, показанного ниже. Результаты должны содержать 75 наблюдений.



Информация о поставщиках в выводе будет содержать пропущенные значения.

| Obs | Product_ID   | Product_Name         | Supplier_ID | Supplier_Name |
|-----|--------------|----------------------|-------------|---------------|
| 1   | 210000000000 | Children             | .           | .             |
| 2   | 210100000000 | Children Outdoors    | .           | .             |
| 3   | 210100100000 | Outdoor things, Kids | .           | .             |
| 4   | 210200000000 | Children Sports      | .           | .             |
| 5   | 210200100000 | A-Team, Kids         | .           | .             |

## Упражнения. Выполнение вычислений в циклах DO

Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу libname.sas, которую можно найти в директории с данными курса.

### Выполнение вычислений в циклах DO

Отделу начисления заработной платы в Orion Star нужно запланировать суммарные затраты на сотрудников (включая заработную плату, пенсии, затраты на медицинское обслуживание) на несколько лет вперёд, учитывая предполагаемые коэффициенты по их увеличению.

- a. Откройте файл **p207e01** и внесите следующие изменения:

- Добавьте цикл DO, содержащий операторы для вычисления значений переменных **Wages**, **Retire** и **Medical**.
  - Используйте **Start** и **Stop** как начальное и конечное значение индексной переменной **Year**.
- Предположим, что значения коэффициентов для ежегодного увеличения переменных задаются следующей таблицей. Например, чтобы вычислить **Wages**, нужно использовать формулу:

$$\text{wages} = \text{wages} * 1.06;$$

| Переменная     | Текущее значение | Оценка ежегодного роста |
|----------------|------------------|-------------------------|
| <b>Wages</b>   | \$12,874,000     | 6.0%                    |
| <b>Retire</b>  | 1,765,000        | 1.4%                    |
| <b>Medical</b> | 649,000          | 9.5%                    |

- Создайте другую переменную, **Total\_Cost**, равную сумме значений **Wages**, **Retire** и **Medical** за определённый год.
- Выведите по одному наблюдению на каждый год. Ваш отчёт должен содержать 10 наблюдений.

- b. Распечатайте и проверьте ваши результаты.

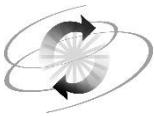
Фрагмент вывода PROC PRINT (Всего 10 наблюдений)

| Obs | Year | Wages         | Retire       | Medical      | Total_Cost    |
|-----|------|---------------|--------------|--------------|---------------|
| 1   | 2013 | 13,646,440.00 | 1,789,710.00 | 710,655.00   | 16,146,805.00 |
| 2   | 2014 | 14,465,226.40 | 1,814,765.94 | 778,167.23   | 17,058,159.57 |
| 3   | 2015 | 15,333,139.98 | 1,840,172.66 | 852,093.11   | 18,025,405.76 |
| 4   | 2016 | 16,253,128.38 | 1,865,935.08 | 933,041.96   | 19,052,105.42 |
| 5   | 2017 | 17,228,316.09 | 1,892,058.17 | 1,021,680.94 | 20,142,055.20 |



Указанные выше результаты созданы 14 февраля 2012 года. В ваших результатах значения столбца **Year** будут другими.

## Упражнения. Вычисления в циклах DO с условием



Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу libname.sas, которую можно найти в директории с данными курса.

### Вычисления в циклах DO с условием

Отделу начисления заработной платы в Orion Star нужно запланировать суммарные затраты на сотрудников (включая заработную плату, пенсии, затраты на медицинское обслуживание) на несколько лет вперёд, учитывая их предполагаемый рост.

- a. Доход корпорации за прошлый год составил \$50 000 000. Рост дохода предполагается на уровне одного процента в год.

Измените программу p207e02, чтобы цикл DO остановился, когда ежегодные расходы превысят ежегодный доход.

- b. Распечатайте столбцы Year, Income и Total\_Cost. Проверьте, что совокупные расходы превысят доход в 26 наблюдении.

Вывод PROC PRINT

| Obs | Year | Income        | Total_Cost    |
|-----|------|---------------|---------------|
| 1   | 2013 | 50,500,000.00 | 16,146,805.00 |
| 2   | 2014 | 51,005,000.00 | 17,058,159.57 |
| 3   | 2015 | 51,515,050.00 | 18,025,405.76 |
| 4   | 2016 | 52,030,200.50 | 19,052,105.42 |
|     | ...  |               |               |
| 26  | 2038 | 64,762,815.75 | 67,973,189.29 |



Указанные выше результаты созданы 14 февраля 2012 года. В ваших результатах значения столбца Year будут другими.

## Упражнения. Использование массивов

 Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу libname.sas, которую можно найти в директории с данными курса.

### Использование массивов для повторяющихся вычислений

В наборе данных **orion.orders\_midyear** хранится информация о сумме покупок для каждого клиента, сгруппированная по месяцам первой половины года. Менеджер по продажам Orion Star хочет на основании этих данных оценить, сколько будут тратить клиенты с учётом 5% скидки на все товары. Требуется создать таблицу с прогнозами расходов для каждого клиента в разбивке по месяцам.

Фрагмент **orion.orders\_midyear**

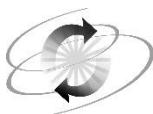
| Obs | Customer_ID | Month1  | Month2 | Month3 | Month4  | Month5  | Month6  |
|-----|-------------|---------|--------|--------|---------|---------|---------|
| 1   | 5           | 213.10  | .      | 478.0  | 525.80  | 394.35  | 191.79  |
| 2   | 10          | 188.10  | 414.09 | 2876.9 | 3164.59 | 2373.44 | 169.29  |
| 3   | 11          | 78.20   | .      | .      | .       | .       | 70.38   |
| 4   | 12          | 135.60  | .      | 117.6  | 129.36  | 97.02   | 122.04  |
| 5   | 18          | .       | .      | 29.4   | 32.34   | 24.26   | .       |
| 6   | 24          | 93.00   | 265.80 | .      | .       | .       | 83.70   |
| 7   | 27          | 310.70  | 782.90 | .      | .       | .       | 279.63  |
| 8   | 31          | 1484.30 | 293.30 | .      | .       | .       | 1335.87 |
| 9   | 34          | 642.50  | .      | 86.3   | 94.93   | 71.20   | 578.25  |
| 10  | 41          | 134.00  | 119.20 | 313.0  | 344.30  | 258.23  | 120.60  |
| 11  | 45          | 443.88  | 216.20 | 40.2   | 44.22   | 33.17   | 399.49  |
| 12  | 49          | 24.80   | .      | .      | .       | .       | 22.32   |

- c. Создайте набор данных **discount\_sales**, учитывающий 5% скидку.
  - Создайте массив **Mon** для доступа к переменным с **Month1** по **Month6**.
  - Используйте цикл DO для пересчёта старых значений в новые. Уменьшите каждое из них на 5%.
- d. Распечатайте полученный набор данных и проверьте ваши результаты.
  - Добавьте подходящий заголовок.
  - Используйте формат DOLLAR для вывода сумм ежемесячных покупок.

Фрагмент вывода PROC PRINT (Всего 24 наблюдения)

| Monthly Sales with 5% Discount |          |          |            |            |            |          |
|--------------------------------|----------|----------|------------|------------|------------|----------|
| Customer_ID                    | Month1   | Month2   | Month3     | Month4     | Month5     | Month6   |
| 5                              | \$202.45 | .        | \$454.10   | \$499.51   | \$374.63   | \$182.20 |
| 10                             | \$178.70 | \$393.39 | \$2,733.06 | \$3,006.36 | \$2,254.77 | \$160.83 |
| 11                             | \$74.29  | .        | .          | .          | .          | \$66.86  |
| 12                             | \$128.82 | .        | \$111.72   | \$122.89   | \$92.17    | \$115.94 |
| 18                             | .        | .        | \$27.93    | \$30.72    | \$23.04    | .        |

## Упражнения. Массивы для табличного поиска



Если вы перезапускали сеанс SAS после выполнения предыдущих упражнений, откройте и выполните программу libname.sas, которую можно найти в директории с данными курса.

### Использование массива для табличного поиска

Менеджер отдела продаж хочет определить подходящих клиентов для будущей промо-акции. Используйте **orion.orders\_midyear** и таблицу для поиска, чтобы создать новый набор данных **preferred\_cust**.

Фрагмент **orion.orders\_midyear**

| Customer_ID | Month1 | Month2 | Month3 | Month4  | Month5  | Month6 |
|-------------|--------|--------|--------|---------|---------|--------|
| 5           | 213.10 | .      | 478.0  | 525.80  | 394.35  | 191.79 |
| 10          | 188.10 | 414.09 | 2876.9 | 3164.59 | 2373.44 | 169.29 |
| 11          | 78.20  | .      | .      | .       | .       | 70.38  |
| 12          | 135.60 | .      | 117.6  | 129.36  | 97.02   | 122.04 |
| 18          | .      | .      | 29.4   | 32.34   | 24.26   | .      |

a. Откройте файл **p207e08** и внесите следующие изменения:

- Создайте временную таблицу для поиска, **Target**, которая содержит целевые значения суммы покупок за каждый месяц:  
200, 400, 300, 100, 100, 200
- Создайте новые переменные, **Over1**, **Over2**, ..., **Over6**, которые содержат разницу между суммой покупок клиента в этом месяце и целевым значением для соответствующего месяца.
- Примените цикл DO и вычислите значения переменных **Over1** ... **Over6**, если сумма покупок клиента в этом месяце превышает соответствующее целевое значение.



Если сумма покупок не превышает целевое значение, не выполняйте этот расчёт.

- Сохраните сумму **Over1** ... **Over6** в другой новой переменной, **Total\_Over**.
- Выполните наблюдение только если значение **Total\_Over** больше 500.
- Новый набор данных должен содержать только **Customer\_ID**, **Over1** ... **Over6** и **Total\_Over**.

b. Распечатайте новый набор данных и проверьте свои результаты.

Фрагмент вывода PROC PRINT (Всего 9 наблюдений)

| Customer_ID | Over1  | Over2  | Over3  | Over4   | Over5   | Over6   | Total_Over |
|-------------|--------|--------|--------|---------|---------|---------|------------|
| 5           | 13.1   | .      | 178.0  | 425.80  | 294.35  | .       | 911.25     |
| 10          | .      | 14.09  | 2576.9 | 3064.59 | 2273.44 | .       | 7929.02    |
| 27          | 110.7  | 382.90 | .      | .       | .       | 79.63   | 573.23     |
| 31          | 1284.3 | .      | .      | .       | .       | 1135.87 | 2420.17    |
| 34          | 442.5  | .      | .      | .       | .       | 378.25  | 820.75     |

Клабуков Максим  
Август 2020

## День 4. SAS(R) Macro Language and SQL: Essentials



### Exercises. Writing Text to the SAS Log

#### Writing Text to the SAS Log with the %PUT Statement

Submit a %PUT statement that writes your name to the SAS log.

#### Writing NOTE, WARNING, and ERROR Messages to the SAS Log with the %PUT Statement

- Open the program **m101e02** shown below into the Editor window.

```
%put NOTE: Is this a SAS note?;  
%put WARNING: Is this a SAS warning?;  
%put ERROR: Is this a SAS error?;
```

- Submit the program and review the results in the SAS log. What is unusual about the results?

---

---

- Replace the colon so that a hyphen follows the keywords NOTE, WARNING, and ERROR. Submit the program and review the results in the SAS log. How do the results change?

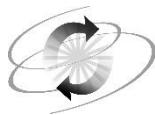
---

---

- Modify the program so that the keywords NOTE, WARNING, and ERROR are in lowercase. Submit the program and review the results in the SAS log. Did the change affect the results?

---

---



## Exercises. Automatic Macro Variables

---

Submit a LIBNAME statement to assign the **orion** libref to the course SAS library according to instructions provided by the instructor.

**libname orion ' \_\_\_\_\_';**

### Displaying Automatic Macro Variables

a. Use the %PUT statement to list all automatic macro variables in the SAS log.

b. What are the values of the following automatic macro variables?

- **SYSLAST** \_\_\_\_\_
- **SYSUSERID** \_\_\_\_\_
- **SYSTIME** \_\_\_\_\_
- **SYSDATE9** \_\_\_\_\_

c. Are the values for the automatic macro variables **SYSTIME** and **SYSDATE9** accurate?

\_\_\_\_\_



## Exercises. Defining and Using Macro Variables

### Defining and Using Macro Variables for Character Substitution

- a. Open the **m102e05** program shown below into the Editor window. Submit the program and examine the output that it creates.

```
proc print data=orion.customer_dim;
  var Customer_Name Customer_Gender Customer_Age ;
  where Customer_Group contains 'Gold';
  title 'Gold Customers';
run;
```

- b. Precede the program with a %LET statement to assign the value *Gold* to **type**. Modify the program so that the two occurrences of *Gold* are replaced by references to the macro variable **type**. Submit the program. It produces the same output as before.
- c. Include the appropriate system option to display resolved values of macro variables in the SAS log. Resubmit the program and examine the log.
- d. Modify the value of **type** to *Internet*. Resubmit the program and examine the log.
- e. Turn off the system option from part c above.



## Exercises. Consecutive Macro Variable References

### Consecutive Macro Variable References

- f. Open the **m102e08** program shown below into the Editor window. Submit the program and examine the output that it creates.

```
proc print data=orion.organization_dim;
  where Employee_Hire_Date='01AUG2006'd;
  id Employee_ID;
  var Employee_Name Employee_Country Employee_Hire_Date;
  title 'Personal Information for Employees Hired in AUG 2006';
run;
```

- g. Precede the program with %LET statements to assign the value *AUG* to **month** and the value *2006* to **year**. Modify the program so that the two occurrences of *AUG* and *2006* are replaced by references to the macro variables **month** and **year**, respectively. Submit the program. It produces the same output as before.
- h. Modify the value of **month** to *JUL* and **year** to *2003*. Resubmit the program.



## Exercises. Macro Functions

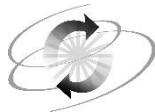
---

### Using the %SUBSTR and %SCAN Functions

- i. Submit a %LET statement to assign the value *Anthony Miller* to a macro variable named **fullname**.
- j. Extract the first initial and last name, putting them together into a new macro variable as *A. Miller*. Use the %PUT statement to display the results.

### Using the %SYSFUNC Function

Use the %PUT statement and the %SYSFUNC function to display the current date and time. Format the date with the MMDDYY10. format and the time with the TIMEAMPM. format.



## Exercises

---

### Defining and Calling a Macro

- a. Open the **m103e01** program shown below into the Editor window.

```
proc print data=orion.customer_dim;
  var Customer_Group Customer_Name Customer_Gender Customer_Age;
  where Customer_Group contains "&type";
  title "&type Customers";
run;
```

- b. Convert the program into a macro named **Customers**. Set the appropriate system option to display a note in the SAS log when a macro definition has compiled. Submit the macro definition and examine the log.
- c. Submit a %LET statement to assign the value *Gold* to the macro variable **type**. Call the macro and examine the log.
- d. Change the value of **type** to *Internet*.
- e. Activate the appropriate system option to display source code received by the SAS compiler. Call the macro again and examine the log.

## Macros with Positional Parameters



### m103d06a

Use positional parameters to specify a range of dates and TABLE statement options for the FREQ procedure.

```
%macro count(opts, start, stop);
proc freq data=orion.orders;
  where order_date between "&start"d and "&stop"d;
  table order_type / &opts;
  title1 "Orders from &start to &stop";
run;
%mend count;
options mprint;
%count(nocum,01jan2008,31dec2008)
%count(,01jul2008,31dec2008)
```

A null value is passed to the OPTS parameter in the second call.

#### Partial SAS Log

```
50  %count(nocum,01jan2008,31dec2008)
MPRINT(COUNT): proc freq data=orion.orders;
MPRINT(COUNT):   where order_date between "01jan2008"d and "31dec2008"d;
MPRINT(COUNT):   table order_type / nocum;
MPRINT(COUNT):   title1 "Orders from 01jan2008 to 31dec2008";
MPRINT(COUNT):   run;
NOTE: There were 87 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01JAN2008'D and order_date<='31DEC2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.28 seconds
      cpu time           0.11 seconds

51  %count(,01jul2008,31dec2008)
MPRINT(COUNT): proc freq data=orion.orders;
MPRINT(COUNT):   where order_date between "01jul2008"d and "31dec2008"d;
MPRINT(COUNT):   table order_type / ;
MPRINT(COUNT):   title1 "Orders from 01jul2008 to 31dec2008";
MPRINT(COUNT):   run;
NOTE: There were 40 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01JUL2008'D and order_date<='31DEC2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.01 seconds
      cpu time           0.01 seconds
```



## Macros with Keyword Parameters

---

### m103d06b

Alter the previous macro by using keyword parameters. Issue various calls to the macro.

```
%macro count(opts=,start=01jan08,stop=31dec08);
  proc freq data=orion.orders;
    where order_date between "&start"d and "&stop"d;
    table order_type / &opts;
    title1 "Orders from &start to &stop";
  run;
%mend count;
options mprint;
%count(opts=nocum)
%count(stop=01jul08,opts=nocum nopercent)
%count()
```

Partial SAS Log

```
64  %count(opts=nocum)
MPRINT(COUNT):  proc freq data=orion.orders;
MPRINT(COUNT):  where order_date between "01jan08"d and "31dec08"d;
MPRINT(COUNT):  table order_type / nocum;
MPRINT(COUNT):  title1 "Orders from 01jan08 to 31dec08";
MPRINT(COUNT):  run;
NOTE: There were 87 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01JAN2008'D and order_date<='31DEC2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.03 seconds
      cpu time          0.03 seconds

65  %count(stop=01jul08,opts=nocum nopercent)
MPRINT(COUNT):  proc freq data=orion.orders;
MPRINT(COUNT):  where order_date between "01jan08"d and "01jul08"d;
MPRINT(COUNT):  table order_type / nocum nopercent;
MPRINT(COUNT):  title1 "Orders from 01jan08 to 01jul08";
MPRINT(COUNT):  run;
NOTE: There were 47 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01JAN2008'D and order_date<='01JUL2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.01 seconds
      cpu time          0.01 seconds

66  %count()
MPRINT(COUNT):  proc freq data=orion.orders;
MPRINT(COUNT):  where order_date between "01jan08"d and "31dec08"d;
MPRINT(COUNT):  table order_type / ;
MPRINT(COUNT):  title1 "Orders from 01jan08 to 31dec08";
MPRINT(COUNT):  run;
NOTE: There were 87 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01JAN2008'D and order_date<='31DEC2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.01 seconds
      cpu time          0.01 seconds
```

## Macros with Mixed Parameter Lists



### m103d06c

Alter the previous macro by using a mixed parameter list. Issue various calls to the macro.

```
%macro count(opts,start=01jan08,stop=31dec08);
proc freq data=orion.orders;
  where order_date between "&start"d and "&stop"d;
  table order_type / &opts;
  title1 "Orders from &start to &stop";
run;
%mend count;
options mprint;
%count(nocum)
%count(stop=30jun08,start=01apr08)
%count(nocum nopercent,stop=30jun08)
%count()
```

### Partial SAS Log

```
76  %count(nocum)
MPRINT(COUNT): proc freq data=orion.orders;
MPRINT(COUNT):   where order_date between "01jan08"d and "31dec08"d;
MPRINT(COUNT):   table order_type / nocum;
MPRINT(COUNT):   title1 "Orders from 01jan08 to 31dec08";
MPRINT(COUNT):   run;
NOTE: There were 87 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01JAN2008'D and order_date<='31DEC2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.04 seconds
      cpu time          0.03 seconds

77  %count(stop=30jun08,start=01apr08)
MPRINT(COUNT): proc freq data=orion.orders;
MPRINT(COUNT):   where order_date between "01apr08"d and "30jun08"d;
MPRINT(COUNT):   table order_type / ;
MPRINT(COUNT):   title1 "Orders from 01apr08 to 30jun08";
MPRINT(COUNT):   run;
NOTE: There were 28 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01APR2008'D and order_date<='30JUN2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.01 seconds
      cpu time          0.01 seconds
```

```
78  %count(nocum nopercent,stop=30jun08)
MPRINT(COUNT):  proc freq data=orion.orders;
MPRINT(COUNT):    where order_date between "01jan08"d and "30jun08"d;
MPRINT(COUNT):    table order_type / nocum nopercent;
MPRINT(COUNT):    title1 "Orders from 01jan08 to 30jun08";
MPRINT(COUNT):    run;
NOTE: There were 47 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01JAN2008'D and order_date<='30JUN2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.01 seconds
      cpu time          0.01 seconds

79  %count()
MPRINT(COUNT):  proc freq data=orion.orders;
MPRINT(COUNT):    where order_date between "01jan08"d and "31dec08"d;
MPRINT(COUNT):    table order_type / ;
MPRINT(COUNT):    title1 "Orders from 01jan08 to 31dec08";
MPRINT(COUNT):    run;
NOTE: There were 87 observations read from the data set ORION.ORDERS.
      WHERE (order_date>='01JAN2008'D and order_date<='31DEC2008'D);
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.03 seconds
      cpu time          0.03 seconds
```



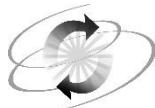
## Exercises. Defining and Using Macro Parameters

### Defining and Using Macro Parameters

- a. Open the m103e04 program shown below into the Editor window.

```
%macro customers;
  proc print data=orion.customer_dim;
    var Customer_Name Customer_Gender Customer_Age;
    where Customer_Group contains "&type";
    title "&type Customers";
  run;
%mend customers;
```

- b. Convert this macro without parameters into a macro with one positional parameter. Name the parameter based on macro variable references within the macro. Set the appropriate system option to display a note in the SAS log when a macro definition has compiled. Submit the macro definition to compile the macro.
- c. Call the macro defined in the previous step with a value of *Gold* for the parameter.
- d. Call the macro again, but with a parameter value of *Catalog*.
- e. Change the positional parameter to a keyword parameter with a default value of *Club*. Submit the revised macro definition to compile the macro.
- f. Call the macro defined in the previous step with a value of *Internet* for the keyword parameter.
- g. Call the macro again, but enable the macro to use its default parameter value.



## Exercises. Quering a Table

If you restarted your SAS session since the last exercise, open and submit the **libname.sas** program found in the data folder.

### Querying a Table

- a. Write a query that displays all rows and all columns from the **orion.employee\_payroll** table.

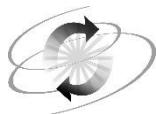
Partial PROC SQL Output

| Employee_ID | Employee_Gender | Birth_Date | Hire_Date | Employee_Term_Date | Marital_Status | Dependents |
|-------------|-----------------|------------|-----------|--------------------|----------------|------------|
| 120101      | M               | 163040     | 7535      | 17384              | .              | S          |
| 120102      | M               | 108255     | 4971      | 12205              | .              | O          |
| 120103      | M               | 87975      | -2535     | 6575               | .              | M          |
| 120104      | F               | 46230      | -600      | 9132               | .              | M          |
| 120105      | F               | 27110      | 6929      | 15826              | .              | S          |

- b. Modify the previous query so that only the columns for **Employee\_ID**, **Employee\_Gender**, **Marital\_Status**, and **Salary** are displayed.

Partial PROC SQL Output

| Employee_ID | Employee_Gender | Marital_Status | Salary |
|-------------|-----------------|----------------|--------|
| 120101      | M               | S              | 163040 |
| 120102      | M               | O              | 108255 |
| 120103      | M               | M              | 87975  |
| 120104      | F               | M              | 46230  |
| 120105      | F               | S              | 27110  |



## Exercises. Eliminating Duplicates

If you restarted your SAS session since the last exercise, open and submit the **libname.sas** program found in the data folder.

### Eliminating Duplicates

Write a query that generates a report that displays the cities where the Orion Star employees reside.

The report should do the following:

- display one unique row per **City**
- use the **orion.employee\_addresses** table

PROC SQL Output

| City         |
|--------------|
| Melbourne    |
| Miami-Dade   |
| Philadelphia |
| San Diego    |
| Sydney       |