

Thực hành phân tích số liệu với phần mềm SPSS

ThS.BS. Trần Thế Trung
Bộ môn Nội Tiết
Đại học Y Dược TP.HCM

Các mục tiêu thực hành

- Nhập và gán các đặc tính cho các biến số định tính (như giới), định lượng (tuổi, chiều cao, cân nặng, chỉ số HbA1c)
- Thực hiện tính toán các biến số mới: chỉ số BMI, phân độ cân nặng dựa trên BMI, phân loại mức độ đường huyết dựa trên HbA1c
- Xử lý **thống kê mô tả**: tỉ lệ (giới tính), trung bình, trung vị, độ lệch chuẩn (của tuổi, chiều cao, cân nặng, HbA1c, BMI)
- So sánh hai trung bình bằng **t-test**: HbA1c giữa hai nhóm theo biến số giới tính (nam, nữ)

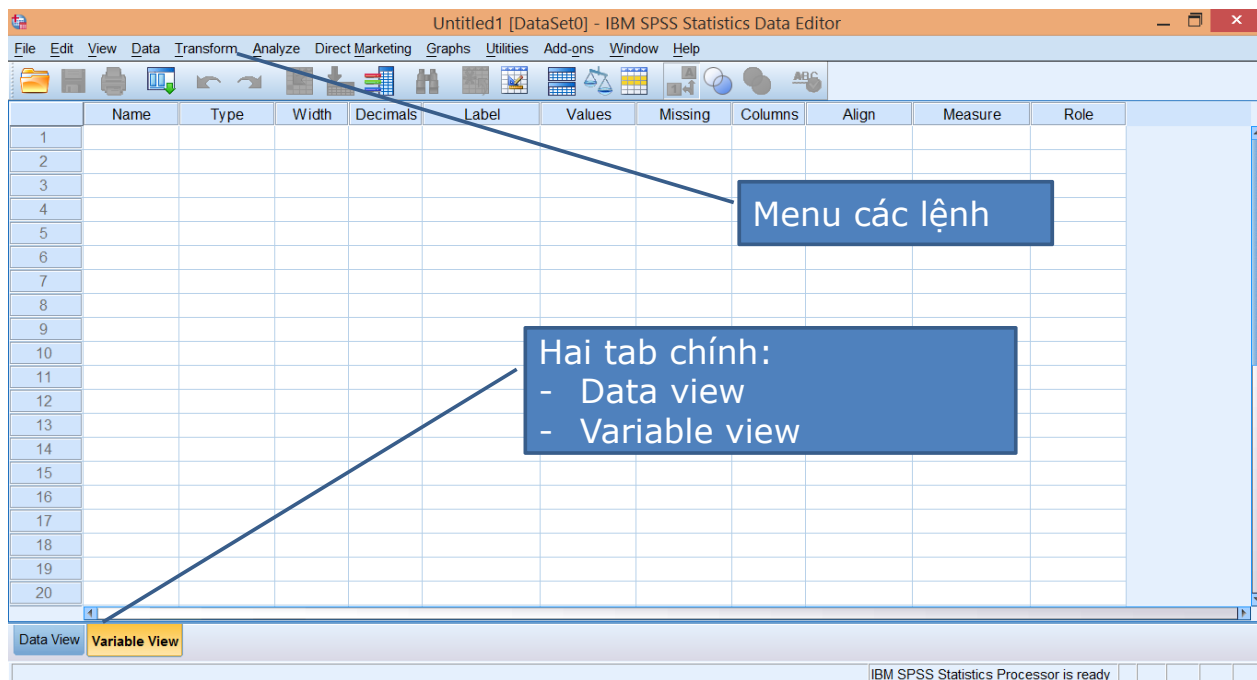
Các mục tiêu thực hành

- Thực hiện phép kiểm **chi-bình phương**: so sánh tỉ lệ thừa cân, tỉ lệ kiểm soát tốt HbA1c theo giới tính
- Phân tích **tương quan Pearson và Spearman**: giữa BMI và HbA1c, giữa tuổi và HbA1c, giữa giới tính và HbA1c
- **Hồi qui logistic** với biến số phụ thuộc là (HbA1c ≤ 7) hoặc (HbA1c > 7), biến số độc lập bao gồm: tuổi, giới và BMI

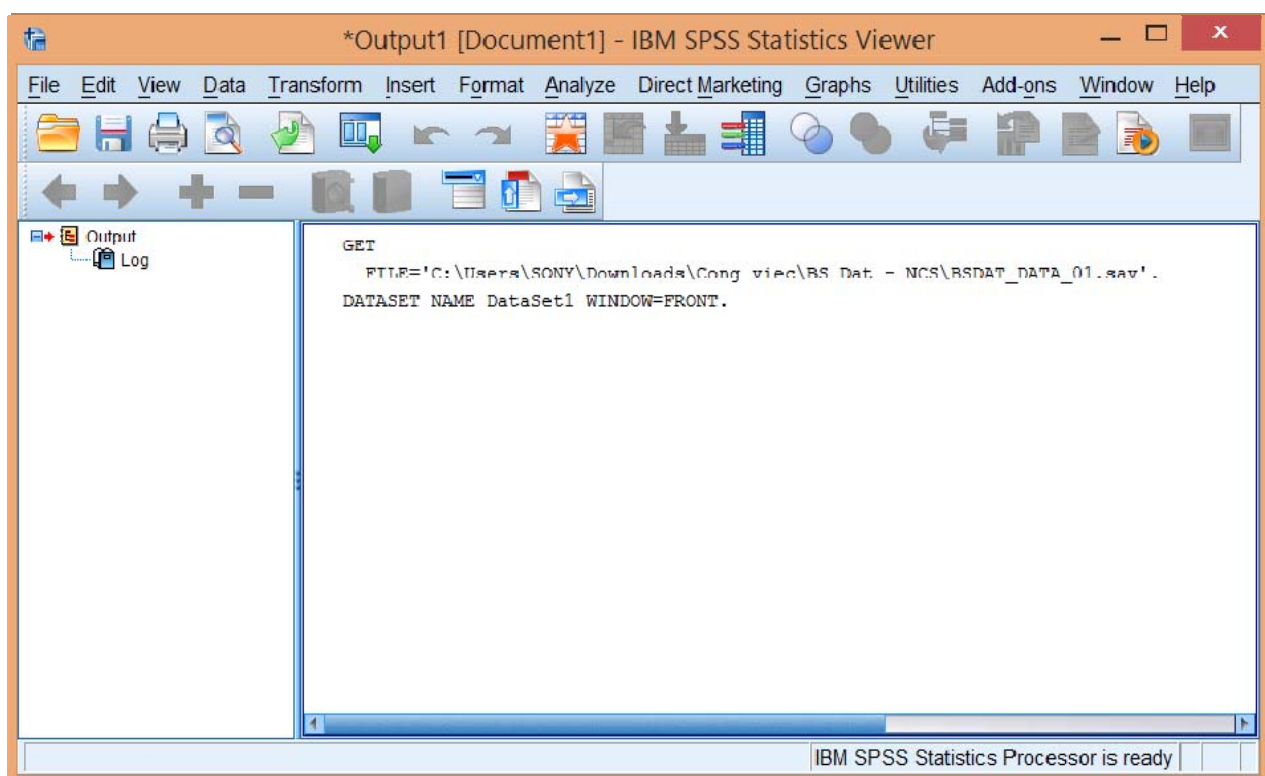
Nội dung

1. Mở ứng dụng SPSS: làm quen các cửa sổ
2. Chuẩn bị số liệu
3. Nhập số liệu: trực tiếp hoặc từ tập tin Excel. Định dạng các đặc tính của biến số. Tính toán các biến số mới: BMI, mức HbA1c
4. Phân tích mô tả: tỉ lệ, trung bình, độ lệch chuẩn
5. So sánh trung bình: t-test
6. So sánh tỉ lệ: chi-bình phương
7. Phân tích tương quan: Pearson và Spearman
8. Hồi qui logistic và chỉ số OR

1. Cửa sổ chính SPSS



Cửa sổ Đầu ra (Output)



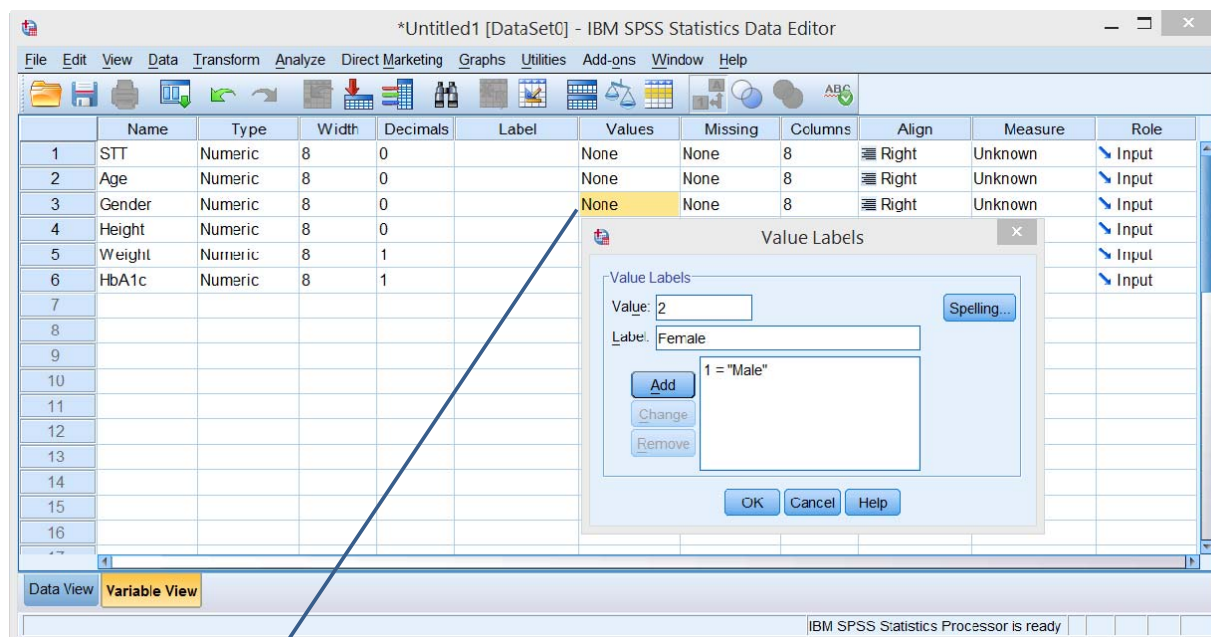
2. Chuẩn bị số liệu: ví dụ

- Các biến số định tính:
 - Giới tính: nam (M) hoặc nữ (F)
 - Định nghĩa: nam = 1, nữ = 2
- Các biến số định lượng:
 - Tuổi: làm tròn số nguyên, không có số thập phân
 - Chiều cao: đơn vị cm, số nguyên
 - Cân nặng: kg, một số thập phân
 - BMI: tính từ cân nặng và chiều cao
 - HbA1c: một số thập phân

Ví dụ bảng số liệu

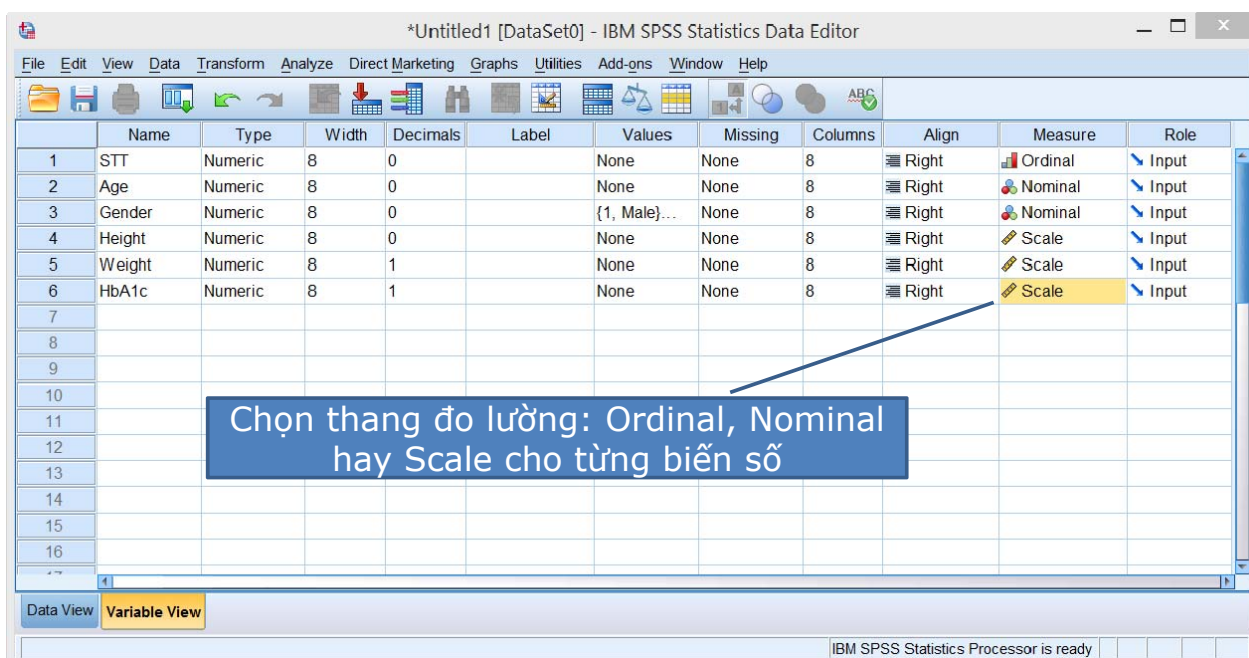
| STT | Họ tên | Tuổi | Giới | Ccao | Cnang | BMI | HbA1c |
|-----|--------|------|------|------|-------|-----|-------|
| 1 | NVT | 43 | M | 165 | 66 | | 7,6 |
| 2 | TBT | 56 | M | 159 | 69 | | 6,5 |
| 3 | TTNY | 19 | M | 148 | 45 | | 6,9 |
| 4 | HNC | 68 | M | 178 | 76 | | 8,1 |
| 5 | BMN | 71 | F | 161 | 78 | | 7,2 |
| 6 | TTKO | 40 | F | 149 | 55 | | 5,8 |
| 7 | TTKH | 59 | F | 154 | 61 | | 6,6 |
| 8 | NTKK | 28 | F | 158 | 70 | | 6,4 |
| 9 | TMH | 49 | M | 165 | 49 | | 7,8 |
| 10 | LTNH | 68 | M | 150 | 62 | | 8,4 |
| 11 | LTTT | 36 | M | 172 | 58 | | 8,0 |
| 12 | TRR | 51 | F | 174 | 66 | | 7,9 |

3. Nhập số liệu trực tiếp Gán giá trị cho biến định tính



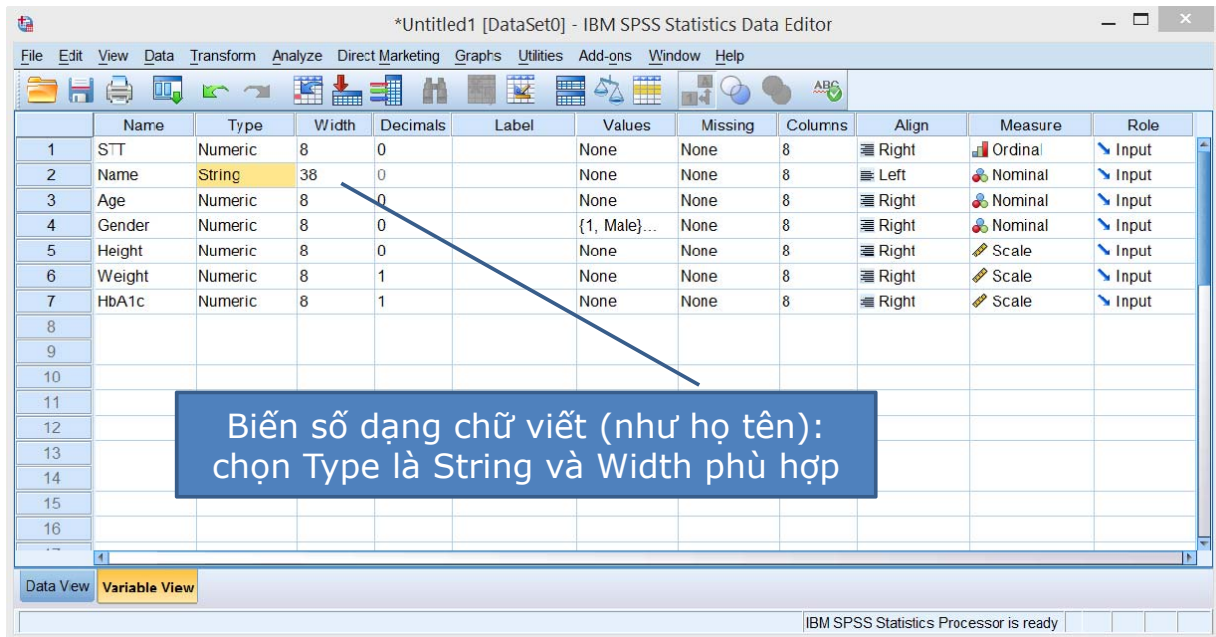
Chọn Values của biến số Gender:
Gán định nghĩa 1 = Male, 2 = Female

3. Nhập số liệu trực tiếp Chọn loại thang đo

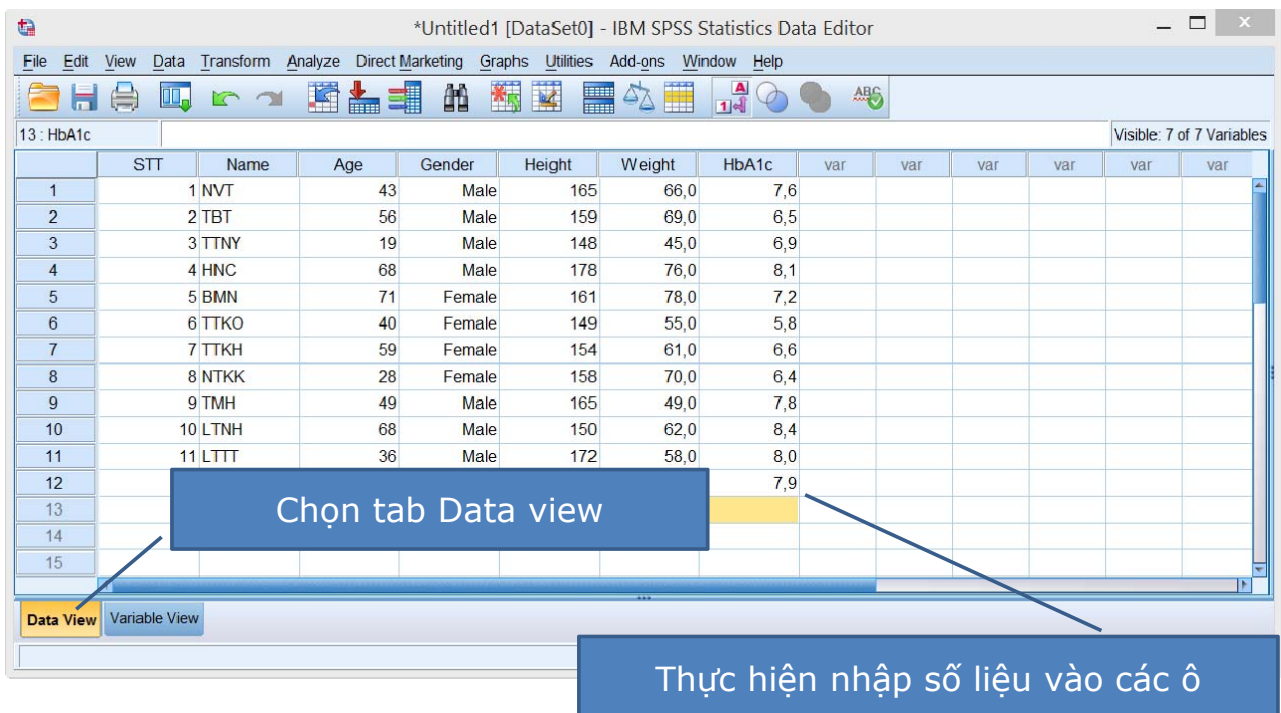


Chọn thang đo lường: Ordinal, Nominal
hay Scale cho từng biến số

3. Nhập số liệu trực tiếp Biến số dạng văn bản (họ tên)



3. Nhập số liệu trực tiếp Nhập số liệu vào Data view



3. Nhập số liệu từ Excel

Chọn File -> Open -> Data ...

Chọn Files of type: Excel

3. Nhập số liệu từ Excel

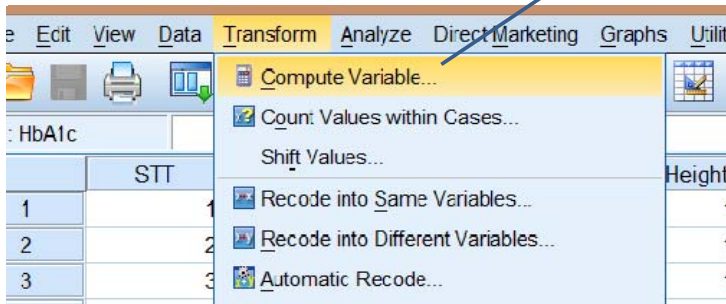
Sau đó, chọn Variable view để chỉnh sửa các đặc tính của biến số

| STT | Htên | Tui | Ccao | Cnang | BMI | HbA1c |
|-----|---------|------|------|-------|-----|-------|
| 1 | 1 NVT | 43 M | 165 | 66 | . | 7,6 |
| 2 | 2 TBT | 56 M | 159 | 69 | . | 6,5 |
| 3 | 3 TTNY | 19 M | 148 | 45 | . | 6,9 |
| 4 | 4 HNC | 68 M | 178 | 76 | . | 8,1 |
| 5 | 5 BMN | 71 F | 161 | 78 | . | 7,2 |
| 6 | 6 TTKO | 40 F | 149 | 55 | . | 5,8 |
| 7 | 7 TTKH | 59 F | 154 | 61 | . | 6,6 |
| 8 | 8 NTKK | 28 F | 158 | 70 | . | 6,4 |
| 9 | 9 TMH | 49 M | 165 | 49 | . | 7,8 |
| 10 | 10 LTNH | 68 M | 150 | 62 | . | 8,4 |
| 11 | 11 LTTT | 36 M | 172 | 58 | . | 8,0 |
| 12 | 12 TRR | 51 F | 174 | 66 | . | 7,9 |

3. Nhập số liệu Tính toán biến số mới

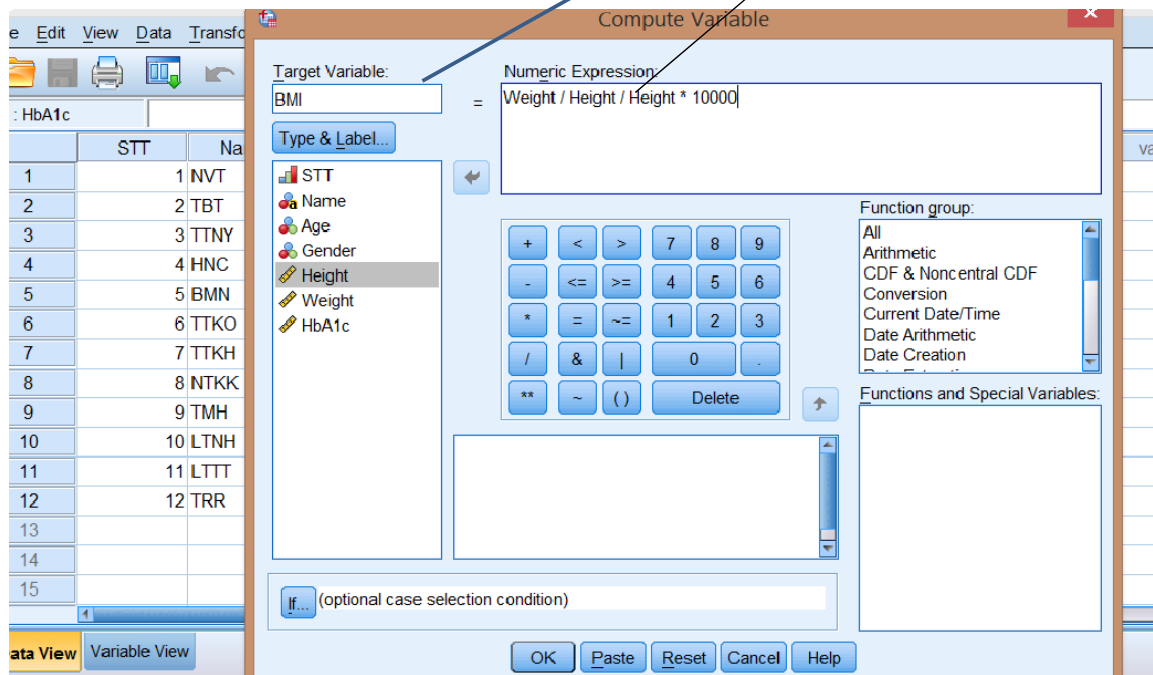
- Tạo biến số mới:
 - BMI = (cân nặng-kg)/(chiều cao-m)^2
 - A1c_7: có 2 giá trị
 - = 1: nếu HbA1c <= 7%
 - = 0: nếu HbA1c > 7%

Transform
-> Compute Variable...



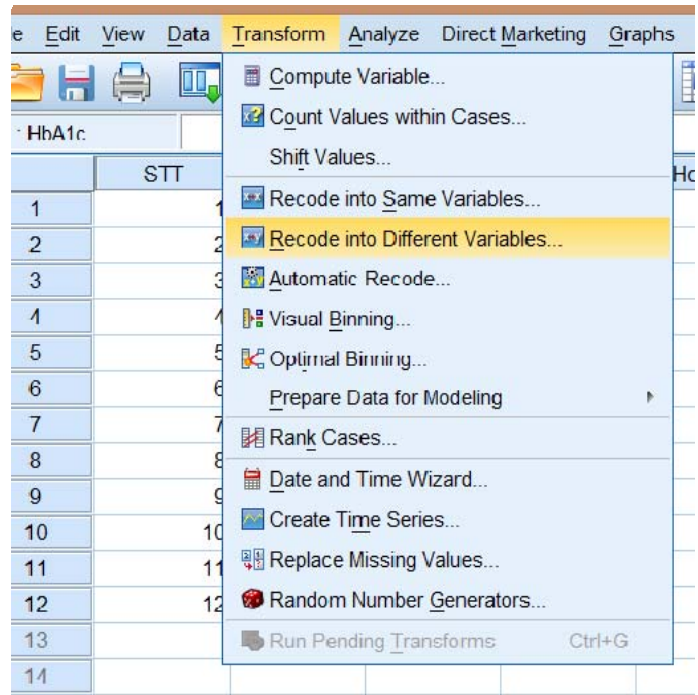
3. Nhập số liệu Tính biến số mới

Nhập tên biến số
Xây dựng biểu thức tính



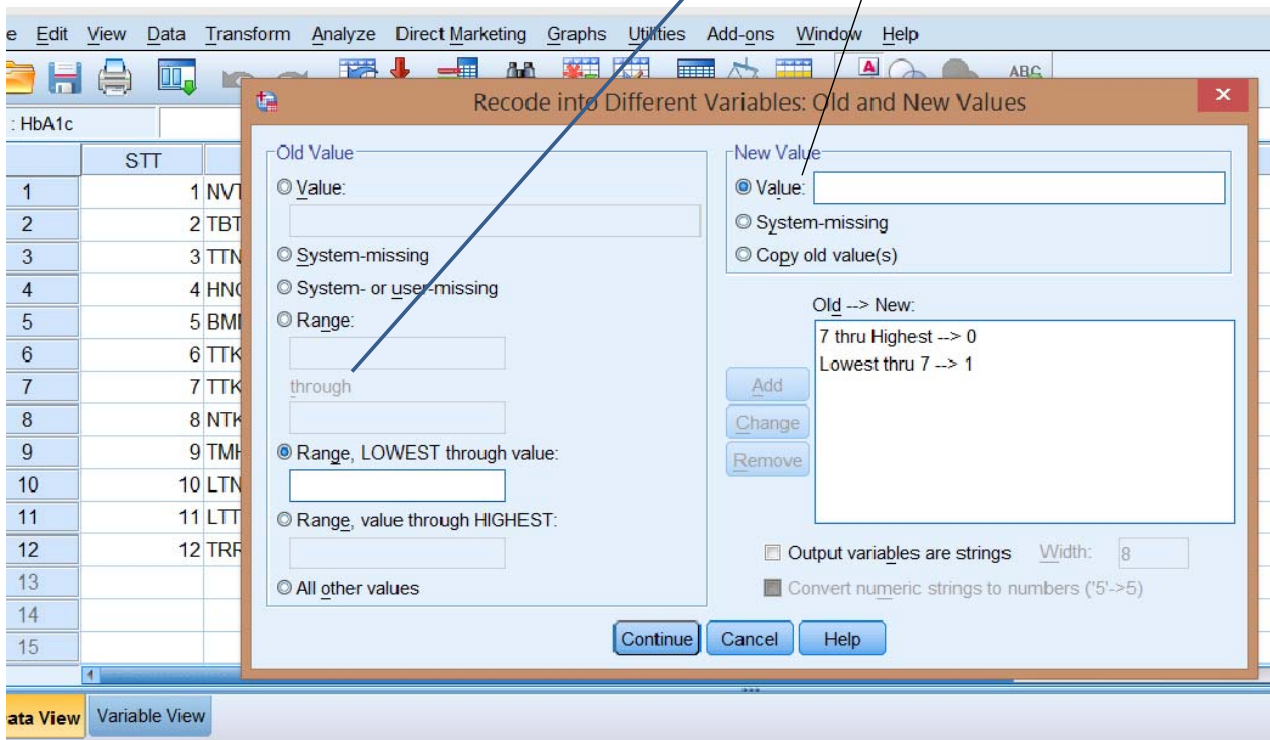
3. Nhập số liệu Tính biến số mới: A1c_7

- Tên biến số: A1c_7
- Hai giá trị 0 và 1
 - 0 : nếu HbA1c > 7%
 - 1 : nếu HbA1c <=7%
- Thao tác: Transform
- > Recode into Different Variable...



3. Nhập số liệu Tính biến số mới

Nhập các giá trị biến số cũ và mới.



3. Nhập số liệu

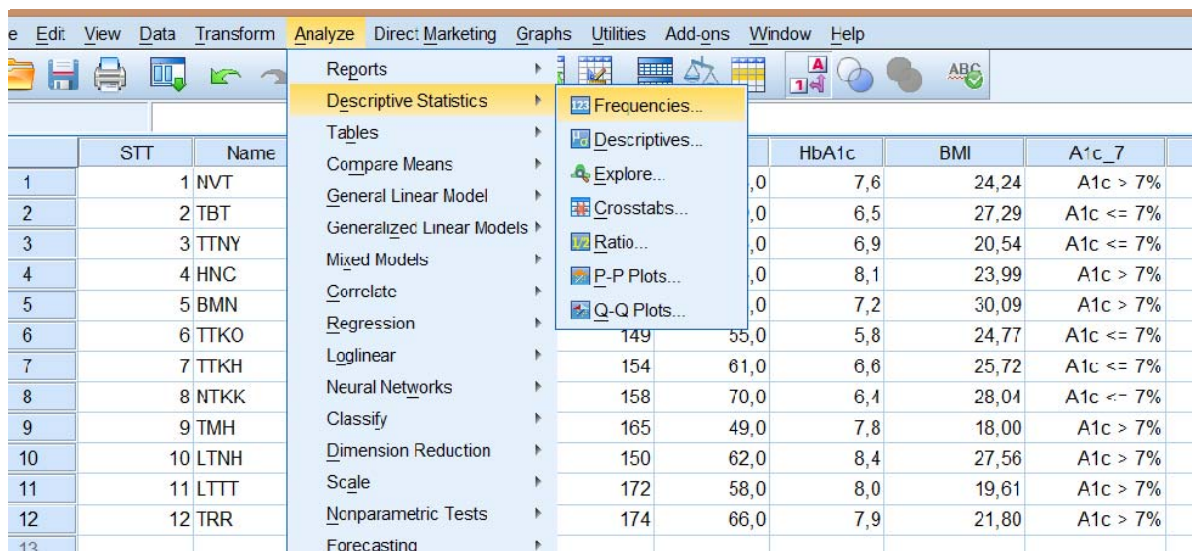
Tính biến số mới

- Tự thực hành: tạo biến số mới
- Tên biến số: Weight_grade
- Các giá trị:
 - 1: nếu BMI \leq 18,5
 - 2: nếu BMI $>$ 18,5 và \leq 23
 - 3: nếu BMI $>$ 23
- Hướng dẫn: chọn Transform -> Recode into Different Variable...

4. Phân tích mô tả

Frequencies

- Analyze -> Descriptive Statistics -> Frequencies...



The screenshot shows the SPSS software interface. The 'Analyze' menu is open, and 'Frequencies...' is selected under 'Descriptive Statistics'. A data table is visible in the background with columns 'SST', 'Name', 'HbA1c', 'BMI', and 'A1c_7'. The 'A1c_7' column contains values like 'A1c > 7%' and 'A1c <= 7%'.

| SST | Name | HbA1c | BMI | A1c_7 |
|-----|---------|-------|-------|-----------|
| 1 | 1 NVT | 7,6 | 24,24 | A1c > 7% |
| 2 | 2 TBT | 6,5 | 27,29 | A1c <= 7% |
| 3 | 3 TTNY | 6,9 | 20,54 | A1c <= 7% |
| 4 | 4 HNC | 8,1 | 23,99 | A1c > 7% |
| 5 | 5 BMN | 7,2 | 30,09 | A1c > 7% |
| 6 | 6 TTKO | 5,8 | 24,77 | A1c <= 7% |
| 7 | 7 TTKH | 6,6 | 25,72 | A1c <= 7% |
| 8 | 8 NTKK | 6,4 | 28,04 | A1c <= 7% |
| 9 | 9 TMH | 7,8 | 18,00 | A1c > 7% |
| 10 | 10 LTNH | 8,4 | 27,56 | A1c > 7% |
| 11 | 11 LTTT | 8,0 | 19,61 | A1c > 7% |
| 12 | 12 TRR | 7,9 | 21,80 | A1c > 7% |

4. Phân tích mô tả Frequencies

Chọn các biến số từ danh sách

The screenshot shows the IBM SPSS Statistics interface with the 'Frequencies' dialog box open. The dialog box has a list of variables on the left, including 'STT', 'Name', 'Age', 'Height', 'Weight', 'HbA1c', and 'BMI'. The 'Variable(s):' list on the right contains 'Gender' and 'A1c_7'. The 'Display frequency tables' checkbox is checked. The background shows a data view with columns: Name, Age, Gender, Height, Weight, HbA1c, BMI, A1c_7, and var. The A1c_7 column contains values like 'A1c > 7%' and 'A1c <= 7%'.

4. Phân tích mô tả Frequencies: kết quả

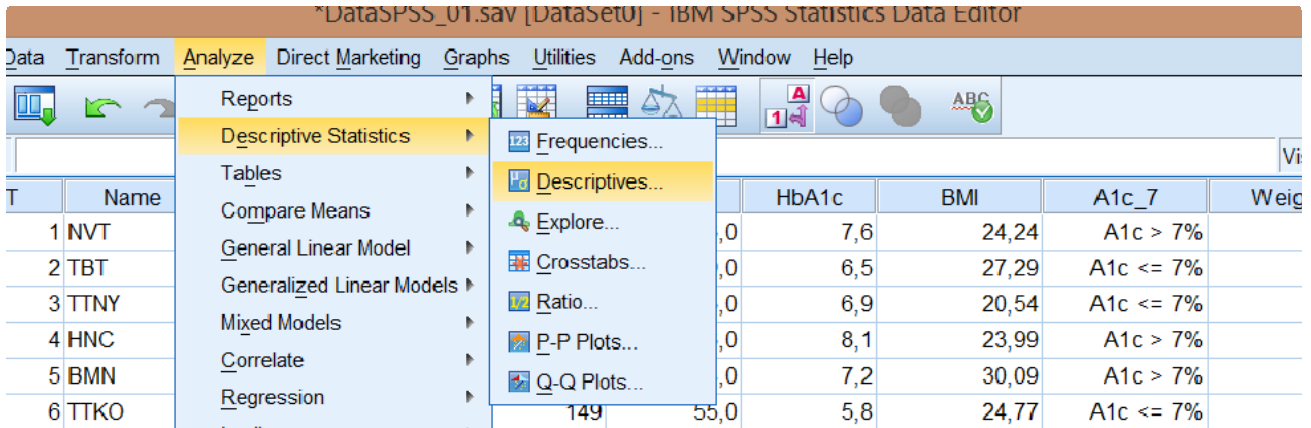
The screenshot shows the IBM SPSS Statistics Viewer displaying the output of the Frequencies procedure. The output includes a 'Frequency Table' for 'Gender' and 'A1c_7'. The 'Gender' table shows 7 males (58.3%) and 5 females (41.7%). The 'A1c_7' table shows 7 cases for 'A1c > 7%' (58.3%) and 5 cases for 'A1c <= 7%' (41.7%).

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|--------|-----------|---------|---------------|--------------------|
| Valid | Male | 7 | 58,3 | 58,3 | 58,3 |
| | Female | 5 | 41,7 | 41,7 | 100,0 |
| | Total | 12 | 100,0 | 100,0 | |

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-----------|-----------|---------|---------------|--------------------|
| Valid | A1c > 7% | 7 | 58,3 | 58,3 | 58,3 |
| | A1c <= 7% | 5 | 41,7 | 41,7 | 100,0 |
| | Total | 12 | 100,0 | 100,0 | |

4. Phân tích mô tả Descriptive

- Analyze -> Descriptive Statistics -> Descriptives...



4. Phân tích mô tả Descriptive

Chọn các biến số từ danh sách

Vào Options... để chọn

The image shows the 'Descriptives' dialog box in SPSS. The 'Variable(s):' list contains 'Height', 'Weight', 'HbA1c', and 'BMI'. The 'Options...' button is highlighted. A blue callout box points to the 'Options...' button with the text 'Vào Options... để chọn'. Another blue callout box points to the 'Descriptives: Options' dialog box with the text 'Chọn các biến số từ danh sách'. The 'Descriptives: Options' dialog box shows the following options:

- Mean Sum
- Dispersion
 - Std. deviation Minimum
 - Variance Maximum
 - Range S.E. mean
- Distribution
 - Kurtosis Skewness
- Display Order
 - Variable list
 - Alphabetic
 - Ascending means
 - Descending means

Buttons: Continue, Cancel, Help

4. Phân tích mô tả

The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays the following SPSS syntax:

```
RECODE BMI (Lowest thru 18.5=1) (18.5 thru 23=2) (23 thru Highest=3) INTO Weigh  
t_grade.  
EXECUTE.  
DESCRIPTIVES VARIABLES=Height Weight HbA1c BMI  
/STATISTICS=MEAN STDDEV MIN MAX.
```

Below the syntax, the 'Descriptives' section is expanded, showing a table of Descriptive Statistics:

| | N | Minimum | Maximum | Mean | Std. Deviation |
|--------------------|----|---------|---------|---------|----------------|
| Height | 12 | 148 | 178 | 161,08 | 10,031 |
| Weight | 12 | 45,0 | 78,0 | 62,917 | 10,0495 |
| HbA1c | 12 | 5,8 | 8,4 | 7,267 | ,8195 |
| BMI | 12 | 18,00 | 30,09 | 24,3043 | 3,71174 |
| Valid N (listwise) | 12 | | | | |

5. So sánh trung bình: t-test

- Nhiệm vụ: so sánh trung bình HbA1c giữa nam và nữ
- Analyze -> Compare Means -> Independent-Samples T Test...

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Compare Means' -> 'Independent-Samples T Test...' is highlighted. The data editor shows a table with columns for Name, BMI, A1c_7, and Weigh.

| T | Name | BMI | A1c_7 | Weigh |
|---|------|-------|-----------|-------|
| 1 | NVT | 24,24 | A1c > 7% | |
| 2 | TBT | 27,29 | A1c <= 7% | |
| 3 | TTNY | 20,54 | A1c <= 7% | |
| 4 | HNC | 23,99 | A1c > 7% | |
| 5 | BMN | 30,09 | A1c > 7% | |
| 6 | TTKO | 24,77 | A1c <= 7% | |

5. So sánh trung bình: t-test

Chọn Gender vào Grouping Variable

Chọn HbA1c vào Test Variable(s)

Vào Define Groups... để chọn nhóm so sánh

5. So sánh trung bình: t-test

Các thông số về hai nhóm có thể xem ở đây (n, trung bình, độ lệch chuẩn)

| Gender | N | Mean | Std. Deviation | Std. Error Mean |
|--------------|---|-------|----------------|-----------------|
| HbA1c Male | 7 | 7,614 | ,6817 | ,2577 |
| HbA1c Female | 5 | 6,780 | ,8012 | ,3583 |

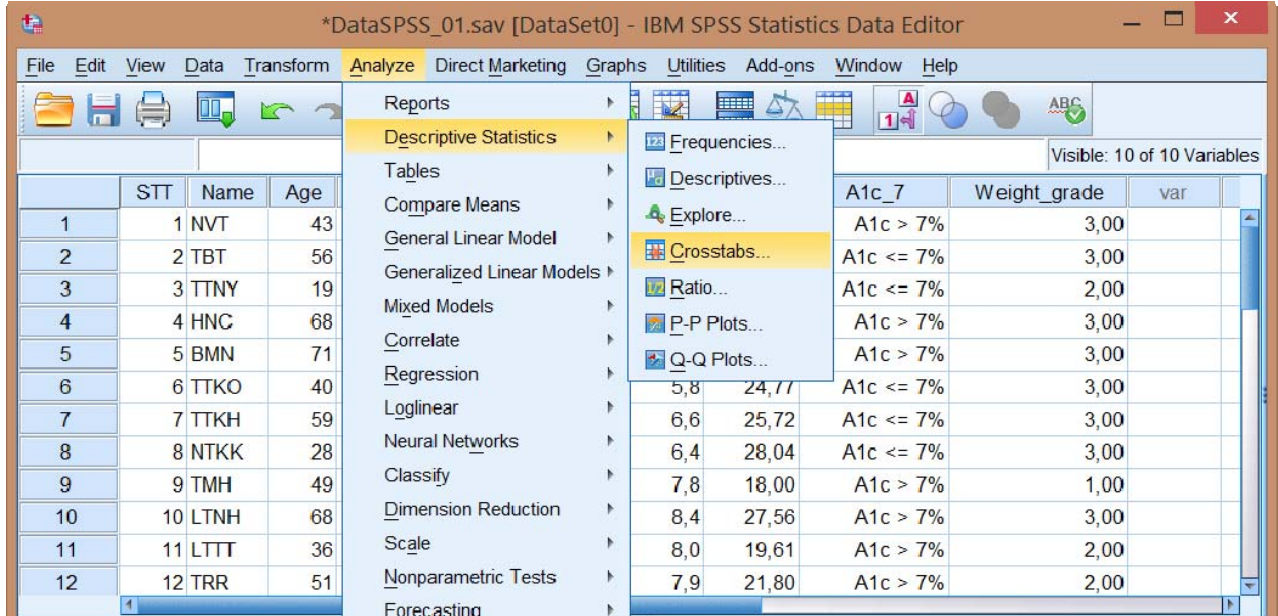
| | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | | |
|-------|---|------|------------------------------|-------|-----------------|-----------------|-----------------------|---|--------|--------|
| | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | | |
| | | | | | | | | Lower | Upper | |
| HbA1c | Equal variances assumed | ,154 | ,703 | 1,947 | 10 | ,080 | ,8343 | ,4285 | -,1206 | 1,7892 |
| | Equal variances not assumed | | | 1,890 | 7,913 | ,096 | ,8343 | ,4414 | -,1877 | 1,8563 |

P-value (2 đuôi)

Mức khác biệt giữa 2 nhóm

6. So sánh tỉ lệ: Phép kiểm Chi-bình phương

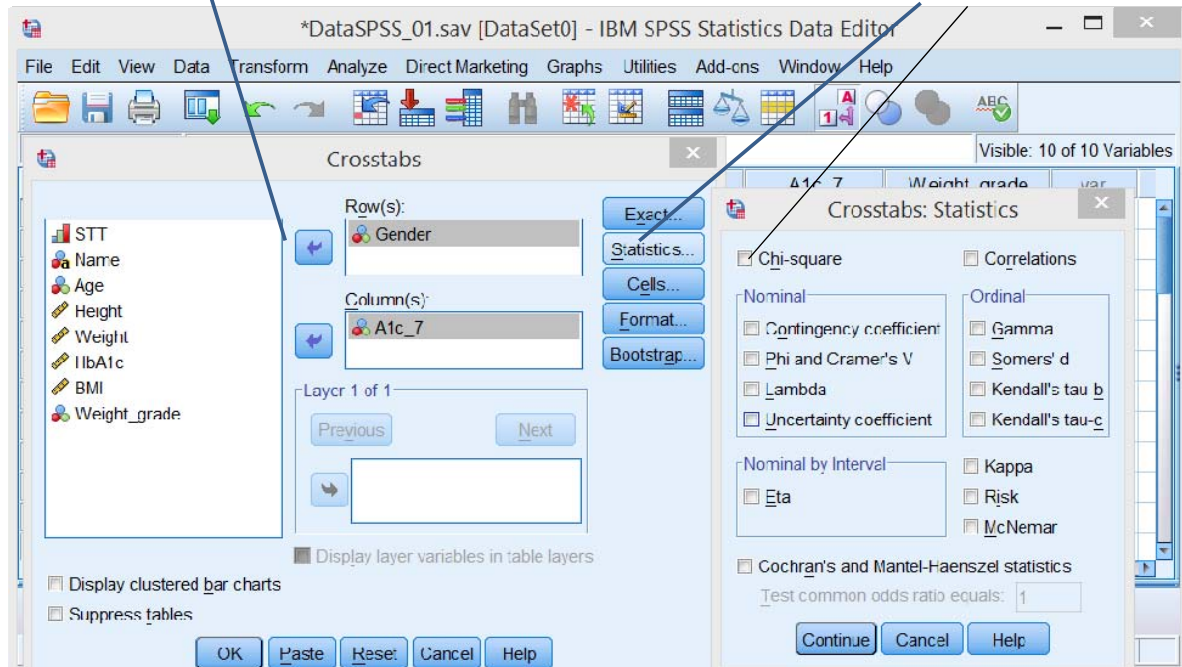
- Mục tiêu: so sánh tỉ lệ đạt HbA1c $\leq 7\%$ của hai giới, nam và nữ
- Analyze -> Descriptive Statistics...-> Crosstabs...



6. So sánh tỉ lệ: Phép kiểm Chi-bình phương

Chọn các biến số đưa vào ô Row(s) và Column(s)

Nhấn vào nút Statistics... để mở cửa sổ và chọn Chi-square



6. So sánh tỉ lệ: Phép kiểm Chi-bình phương

Gender * A1c_7 Crosstabulation

| Count | | A1c_7 | | Total |
|--------|--------|----------|-----------|-------|
| | | A1c > 7% | A1c <= 7% | |
| Gender | Male | 5 | 2 | 7 |
| | Female | 2 | 3 | 5 |
| Total | | 7 | 5 | 12 |

Chi-Square Tests

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|------------------------------------|--------------------|----|-----------------------|----------------------|----------------------|
| Pearson Chi-Square | 1,185 ^a | 1 | ,276 | | |
| Continuity Correction ^b | ,245 | 1 | ,621 | | |
| Likelihood Ratio | 1,195 | 1 | ,274 | | |
| Fisher's Exact Test | | | | ,558 | ,311 |
| Linear-by-Linear Association | 1,087 | 1 | ,297 | | |
| N of Valid Cases | 12 | | | | |

a. 4 cells (100,0%) have expected count less than 5. The minimum expected count is 2,08.
b. Computed only for a 2x2 table

Giá trị p-value của phép kiểm Chi-square

7. Phân tích tương quan hệ số tương quan r-Pearson và Spearman

- Nhiệm vụ: Tính hệ số tương quan r-Pearson giữa BMI và HbA1c
- Analyze -> Correlate -> Bivariate...

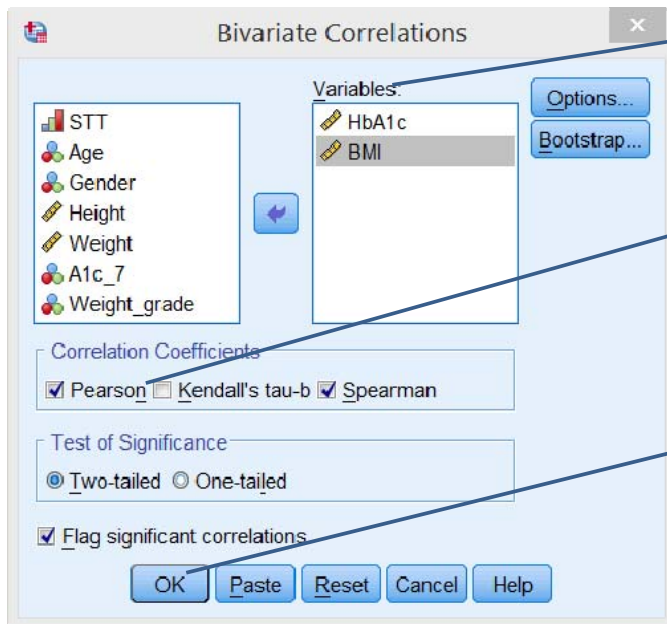
*DataSPSS_01.sav [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 10 of 10 Variables

| STT | Name | Age | HbA1c | BMI | A1c_7 | Weight_grade | var |
|-----|-------|-----|-------|-------|-----------|--------------|------|
| 1 | NVT | 43 | 7,6 | 24,24 | A1c > 7% | | 3,00 |
| 2 | TBT | 56 | 6,5 | 27,29 | A1c <= 7% | | 3,00 |
| 3 | TTNY | 19 | 6,9 | 20,54 | A1c <= 7% | | 2,00 |
| 4 | IINC | 68 | 8,1 | 23,00 | A1c > 7% | | 3,00 |
| 5 | BMN | 71 | 8,1 | 23,00 | A1c > 7% | | 3,00 |
| 6 | TTKO | 40 | 6,5 | 27,29 | A1c <= 7% | | 3,00 |
| 7 | TTKII | 59 | 6,9 | 20,54 | A1c <= 7% | | 3,00 |
| 8 | NTKK | 28 | 6,4 | 20,04 | A1c <= 7% | | 3,00 |

7. Phân tích tương quan hệ số tương quan r-Pearson và Spearman

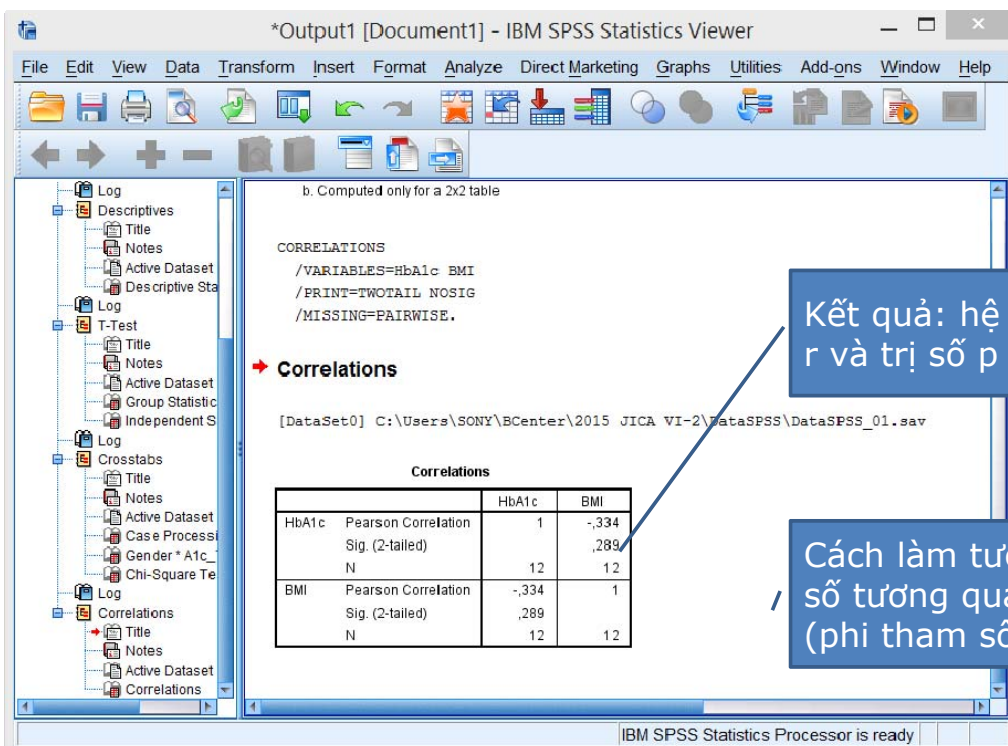


Chọn các biến (HbA1c, BMI) đưa vào ô Variables

Chọn hệ số tương quan phù hợp (trường hợp này là Pearson)

Bấm vào nút OK sau khi chọn xong

7. Phân tích tương quan hệ số tương quan r-Pearson và Spearman



Kết quả: hệ số tương quan r và trị số p (2 đuôi)

Cách làm tương tự với hệ số tương quan Spearman (phi tham số)

7. Phân tích tương quan hệ số tương quan r-Pearson và Spearman

Tự thực hành:

- Tính hệ số tương quan giữa:
 - Tuổi và HbA1c
 - Giới và HbA1c

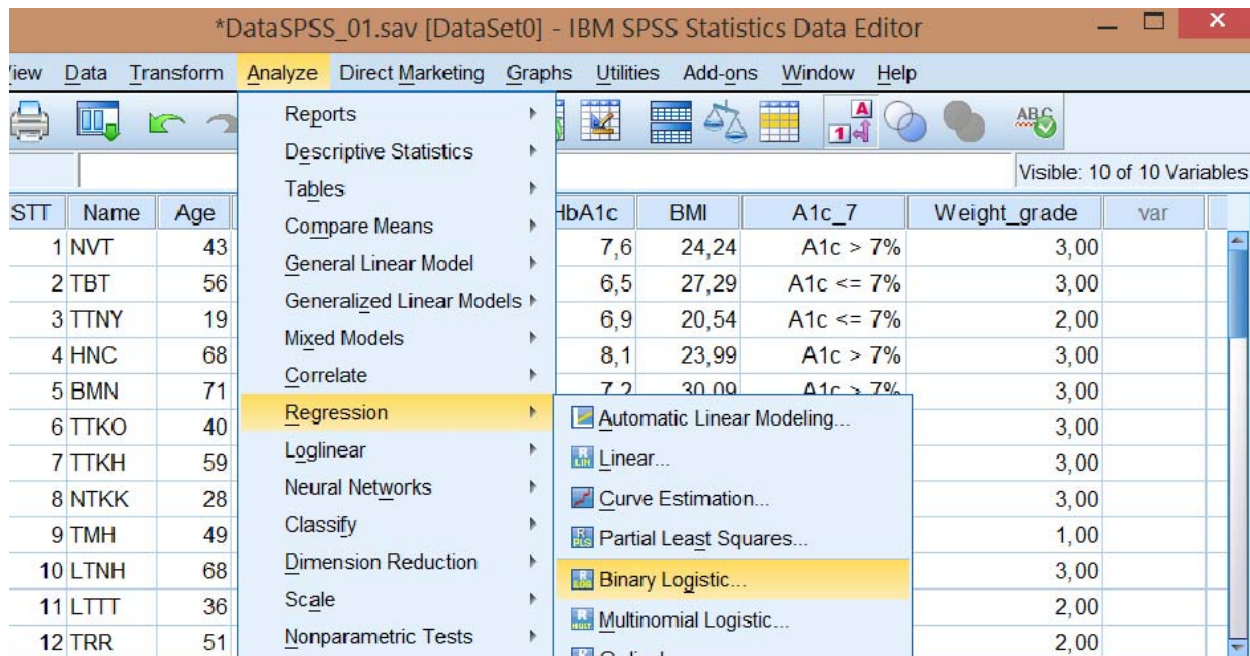
8. Phân tích hồi qui logistic Các điều kiện và khái niệm

- Biến số phụ thuộc: phải có hai giá trị 0 và 1
 - A1c_7: gồm có 0 (HbA1c > 7%) và 1 (HbA1c ≤ 7%)
- Biến độc lập:
 - Giới (biến định danh)
 - Tuổi
 - BMI
- Nhiệm vụ:
 - Phân tích hồi qui logistic đơn biến với từng biến độc lập
 - Phân tích hồi qui logistic đa biến

8. Phân tích hồi qui logistic

Phân tích đơn biến

- Analyze -> Regression -> Binary Logistic...



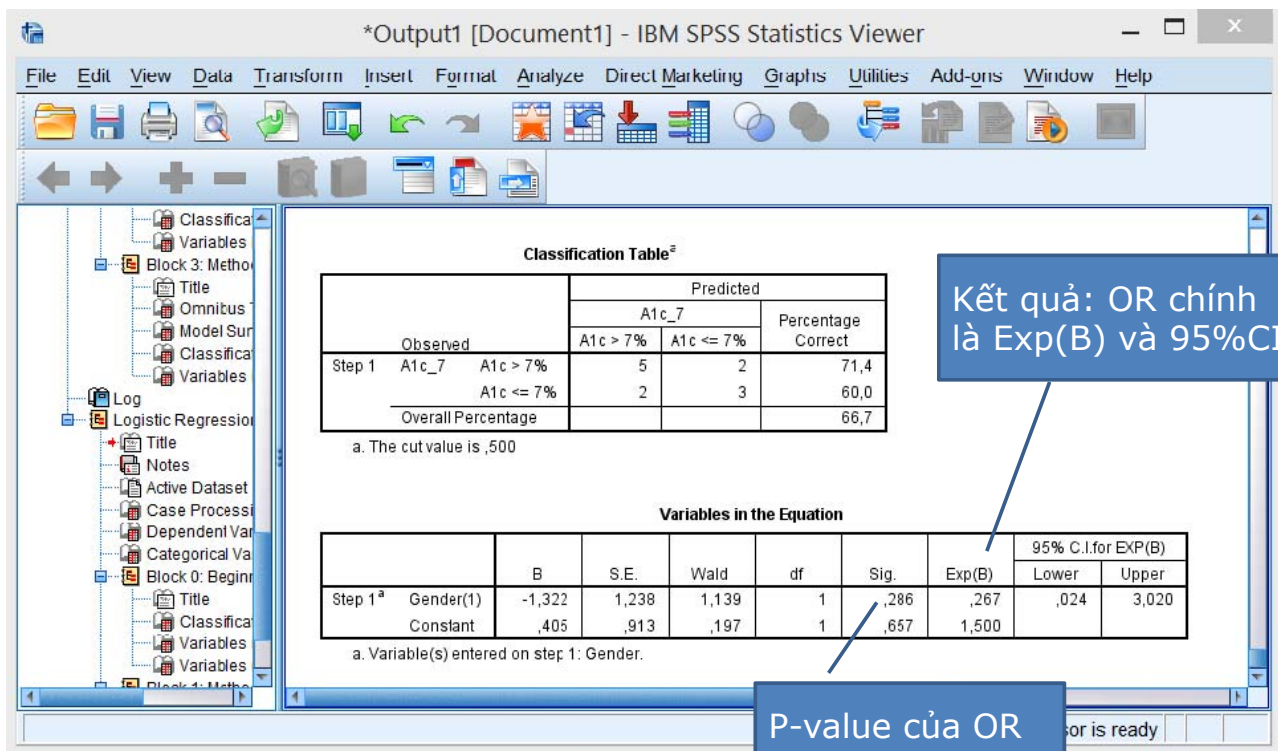
8. Phân tích hồi qui logistic

Phân tích đơn biến

The image shows the 'Logistic Regression' dialog box and its 'Define Categorical Variables' sub-dialog. The main dialog has 'A1c_7' in the 'Dependent' field and 'Gender' in the 'Covariates' field. The sub-dialog has 'Gender(Indicator)' in the 'Categorical Covariates' field. Numbered annotations explain the steps:

1. Chọn biến số phụ thuộc (A1c_7)
2. Chọn biến số độc lập (Gender)
3. Chọn Categorical... để gán đặc tính cho biến Gender
4. Vào Options để chọn 95%CI cho OR

8. Phân tích hồi qui logistic kết quả đơn biến OR và 95%CI



8. Phân tích hồi qui logistic Phân tích đơn biến

Tự thực hành:

- Phân tích đơn biến với hai biến số còn lại:
 - BMI
 - Tuổi

8. Phân tích hồi qui logistic

Phân tích đa biến: chú ý bước (2)

1. Chọn biến số phụ thuộc (A1c_7)

2. Chọn các biến số độc lập (Gender, Age, BMI)

3. Chọn Categorical... để gán đặc tính cho biến Gender

4. Vào Options để chọn 95%CI cho OR

8. Phân tích hồi qui logistic

Kết quả đa biến OR và 95%CI

P-value của OR của từng biến số

Kết quả: OR = Exp(B) của các biến số và 95%CI

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|-------------------------------|--------|-------|-------|----|------|--------|---------------------|--------|
| | | | | | | | Lower | Upper |
| Step 1 ^a Gender(1) | -2,016 | 2,387 | ,713 | 1 | ,398 | ,133 | ,001 | 14,323 |
| Age | -,251 | ,175 | 2,052 | 1 | ,152 | ,778 | ,552 | 1,097 |
| BMI | ,879 | ,630 | 1,946 | 1 | ,163 | 2,409 | ,700 | 8,283 |
| Constant | -8,388 | 9,165 | ,838 | 1 | ,360 | ,000 | | |

a. Variable(s) entered on step 1: Gender, Age, BMI.

Q & A

Câu hỏi và Thảo luận