**Laporan Tugas Analisis Univariat dan Bivariat Menggunakan SPSS**



**Disusun oleh:**

**· Fadilah Ratu Azzahra (5232811015)**

**· Lathif Ramadhan (5231811022)**

**PROGRAM STUDI SAINS DATA**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS TEKNOLOGI YOGYAKARTA**

**YOGYAKARTA**

**2024**

# Pendahuluan

Analisis data memiliki peran penting dalam memahami pola dan karakteristik tersembunyi dari suatu fenomena, termasuk dalam peristiwa bersejarah seperti tenggelamnya kapal Titanic yang cukup terknal. Pada tugas kali ini, kami melakukan analisis univariat dan bivariat terhadap data penumpang Titanic menggunakan perangkat lunak SPSS untuk menemukan faktor-faktor yang mungkin berhubungan dengan tingkat kelangsungan hidup penumpang. Titanic merupakan salah satu dataset yang menarik untuk dianalisis karena mengandung data rinci mengenai penumpang, seperti kelas tiket, usia, jenis kelamin, serta apakah penumpang tersebut berhasil selamat atau tidak.

Dalam analisis ini, kami menggunakan dua kolom, yaitu **Pclass** dan **Survived**, untuk menggali informasi lebih lanjut mengenai hubungan antara kelas tiket dengan peluang selamat. Kolom **Pclass** mengindikasikan kelas tiket yang dimiliki penumpang, di mana kelas ini dapat mencerminkan kondisi atau fasilitas yang dimiliki setiap kelompok penumpang di kapal tersebut. Sedangkan kolom **Survived** menunjukkan apakah seorang penumpang selamat (1) atau tidak (0) dalam peristiwa tragis ini.

Pendekatan univariat pada masing-masing kolom akan memberikan gambaran distribusi penumpang berdasarkan kelas tiket serta proporsi kelangsungan hidup. Sementara itu, analisis bivariat akan membantu kami memahami apakah ada hubungan antara kelas tiket dan tingkat kelangsungan hidup, yang dapat diukur melalui uji statistik tertentu. Dengan memahami data ini, kita dapat memperoleh wawasan mengenai bagaimana kelas sosial atau jenis tiket berpengaruh terhadap kesempatan bertahan hidup pada peristiwa tenggelamnya Titanic.

Analisis ini diharapkan dapat menjadi contoh penerapan metode statistik sederhana dalam menjawab pertanyaan-pertanyaan kompleks dari data historis yang nyata.

Untuk datasetnya kami peroleh dari situs Kaggle.com, berikut link sumber datasetnya: <https://www.kaggle.com/competitions/titanic/data>

# Daftar Isi

[Pendahuluan 2](#_Toc181278922)

[Daftar Isi 3](#_Toc181278923)

[1. Analysis Univariat 4](#_Toc181278924)

[A. Analysis Univariat untuk kolom Passenger Class 4](#_Toc181278925)

[B. Analysis Univariat untuk kolom Survived 8](#_Toc181278926)

[2. Analisis Bivariat 11](#_Toc181278927)

[A. Menggunakan Crosstabs 11](#_Toc181278928)

[1. Tabel Crosstab (Passenger Class \* Survived Crosstabulation): 15](#_Toc181278929)

[2. Uji Chi-Square (Chi-Square Tests): 16](#_Toc181278930)

[Kesimpulan 16](#_Toc181278931)

[B. Analisis Bivariat dengan Mengubah Kedua Kolom Pasangger Class dan Survived yang Bertipe Kategorikal ke Bentuk Numerik 17](#_Toc181278932)

[1. Interpretasi Korelasi Pearson 21](#_Toc181278933)

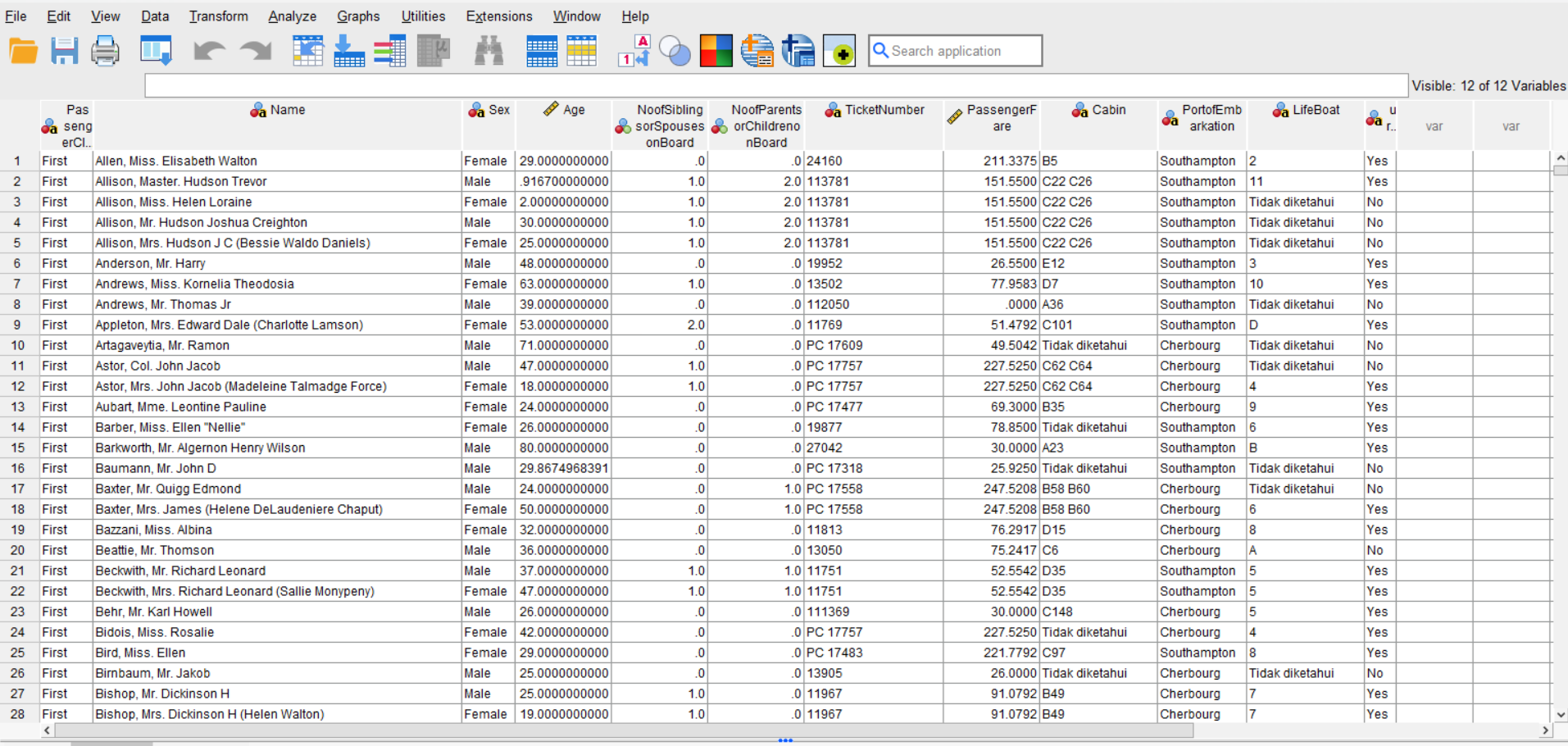
[2. Interpretasi Korelasi Spearman 22](#_Toc181278934)

[Catatan Tambahan 22](#_Toc181278935)

[C. Menggunakan Grafik Scatter Plot 23](#_Toc181278936)

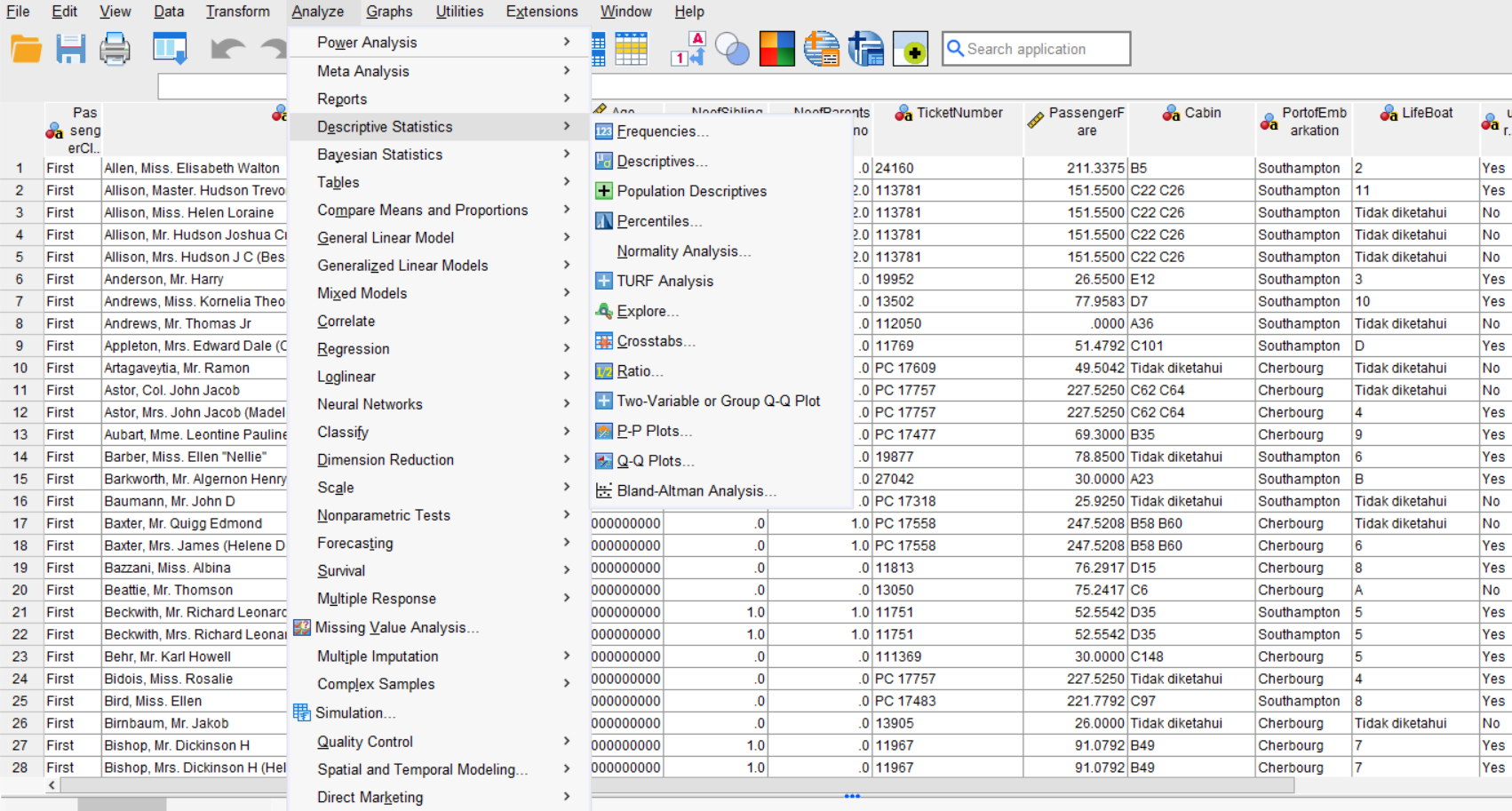
# Analysis Univariat

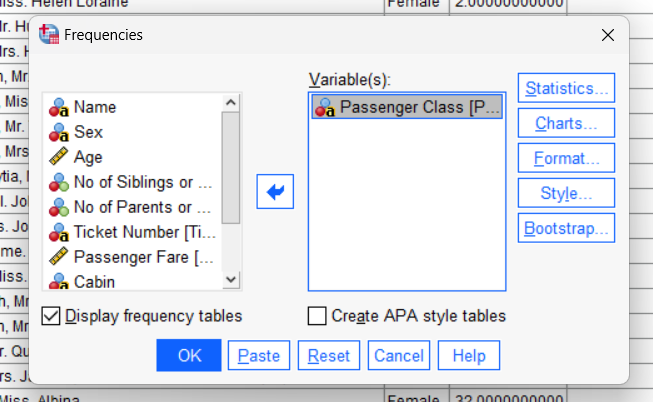
Analisis univariate bertujuan untuk menjelaskan atau mendeskripsikan karakteristik setiap variabel penelitian. Pada umumnya dalam analisis univariate hanya menghasilkan distribusi, frekuensi dan persentase dari setiap variabel. Di sini saya akan melakukan analisis pada data titanic untuk kolom passenger class dan survived.

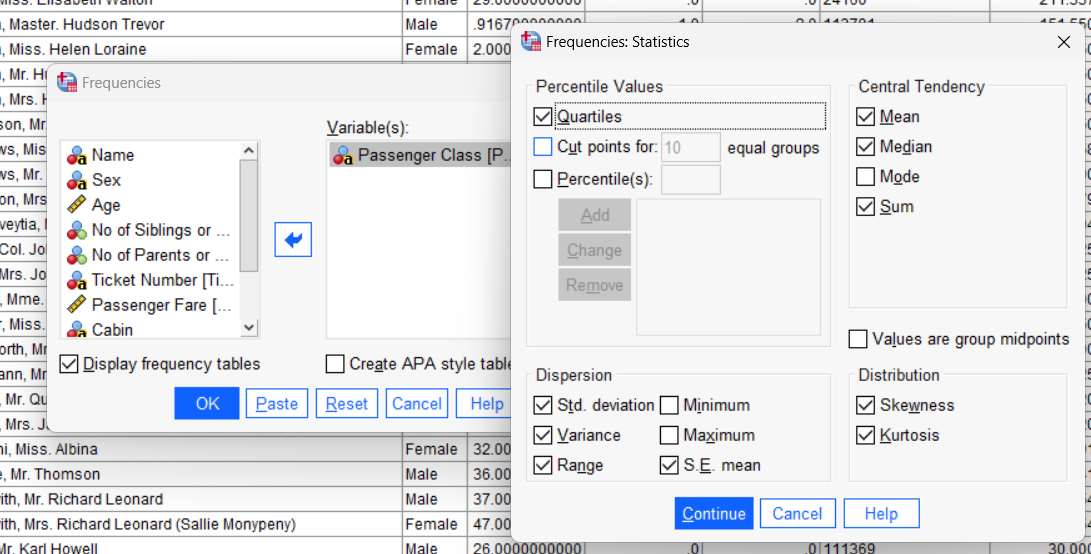


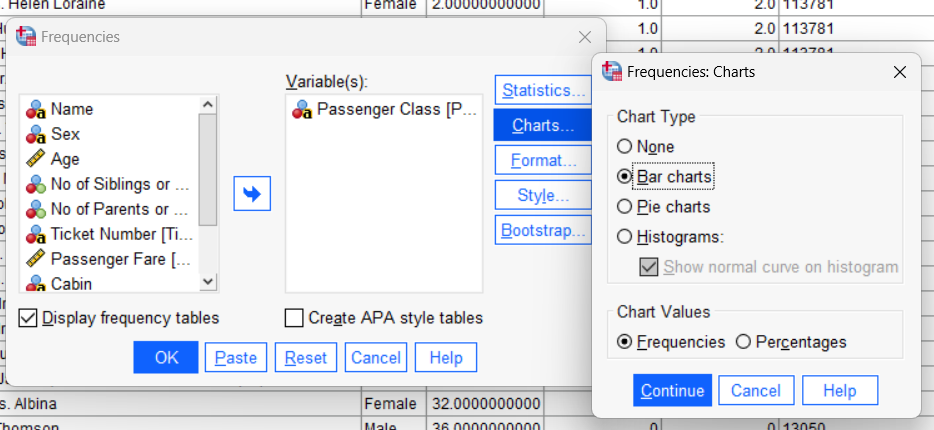
Melakukan analisis univariat:

## A. Analysis Univariat untuk kolom Passenger Class

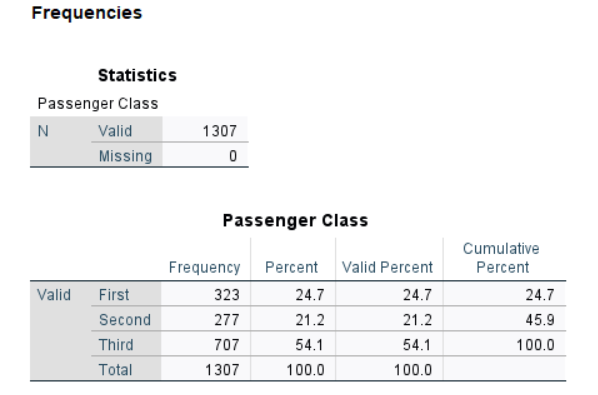


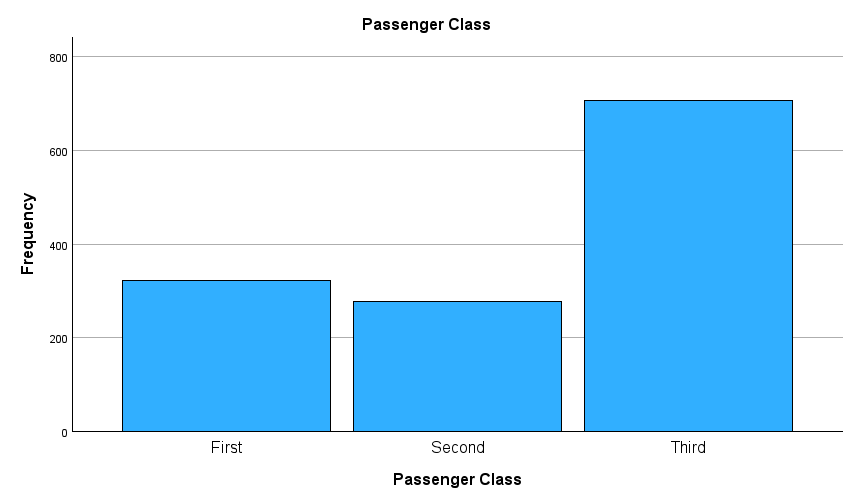






**Output:**

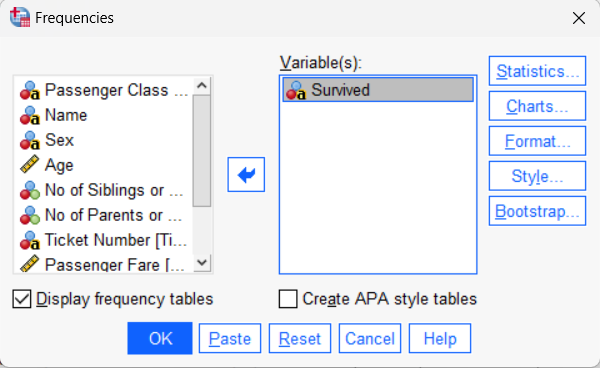


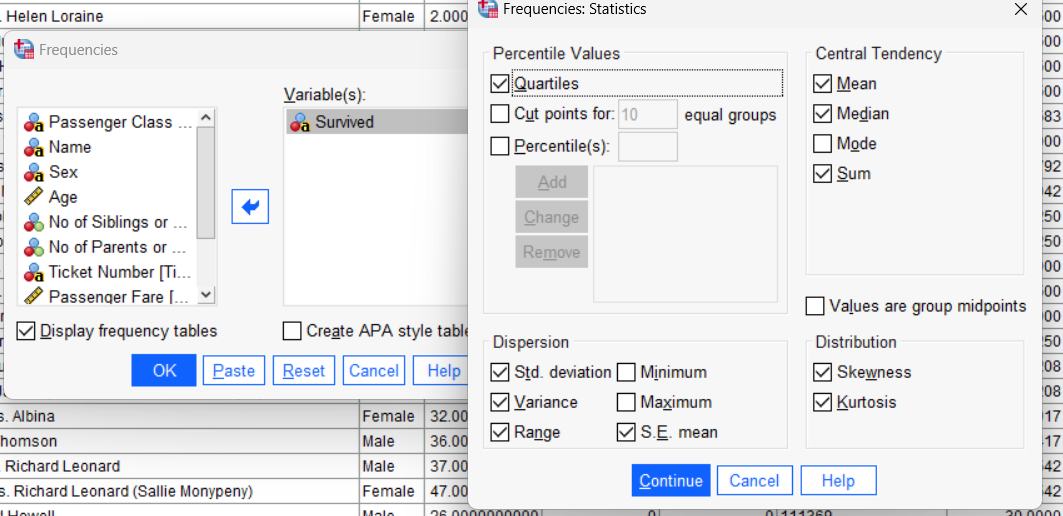


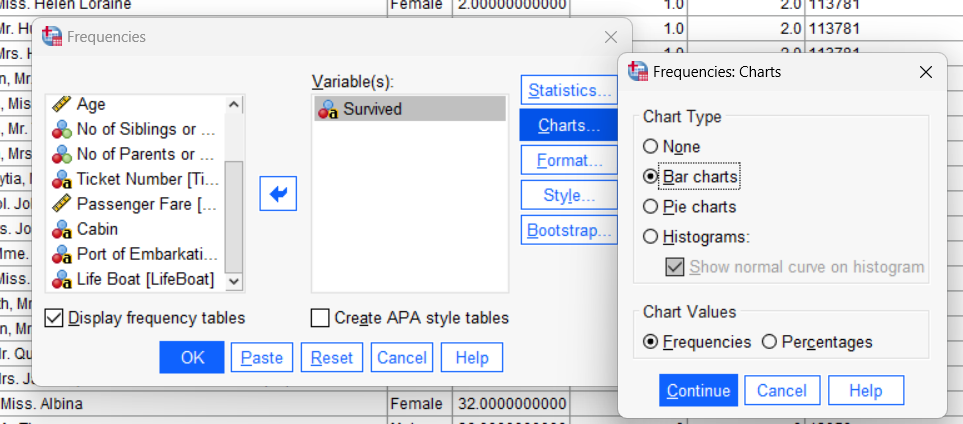
Dari hasil analisis univariat ini, dapat dilihat distribusi kelas penumpang dalam sebuah dataset Titanic :

1. **Jumlah Observasi**:
   * Terdapat total **1,307** data penumpang yang valid, dan tidak ada data yang hilang (Missing = 0).
2. **Distribusi Kelas Penumpang**:
   * **First Class**: Sebanyak 323 penumpang atau 24.7% dari total.
   * **Second Class**: Sebanyak 277 penumpang atau 21.2%.
   * **Third Class**: Sebanyak 707 penumpang atau 54.1%.
3. **Presentase Kumulatif**:
   * Persentase kumulatif menunjukkan total persentase dari kategori yang telah dihitung secara bertahap:
     + First Class menyumbang 24.7% dari total penumpang.
     + Setelah menambahkan Second Class, kumulatifnya menjadi 45.9%.
     + Dengan menambahkan Third Class, kumulatif mencapai 100%, yang menunjukkan semua data penumpang tercakup.
4. **Interpretasi Umum**:
   * Mayoritas penumpang berada di **Third Class** (54.1%), diikuti oleh First Class (24.7%) dan Second Class (21.2%).
   * Informasi ini menunjukkan dominasi jumlah penumpang di kelas ketiga, yang mungkin merepresentasikan karakteristik sosial-ekonomi dari populasi penumpang dalam dataset ini.

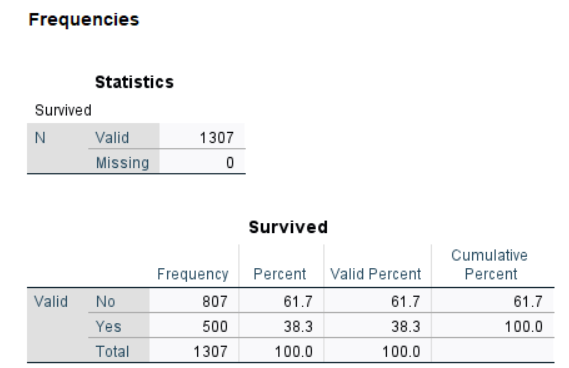
## B. Analysis Univariat untuk kolom Survived

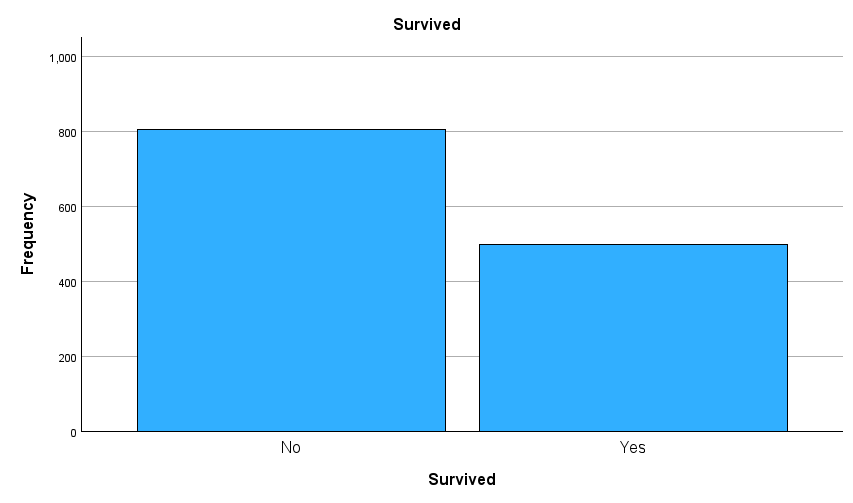






**Output:**





1. **Jumlah Sampel:** Terdapat 1307 data yang valid untuk variabel "Survived". Ini berarti ada 1307 individu dalam dataset ini.
2. **Distribusi:**
   1. Sebanyak 807 individu (sekitar 61.7%) tidak bertahan hidup.
   2. Sebanyak 500 individu (sekitar 38.3%) berhasil bertahan hidup.

**Interpretasi:**

Dari hasil di atas, kita dapat menarik kesimpulan awal sebagai berikut:

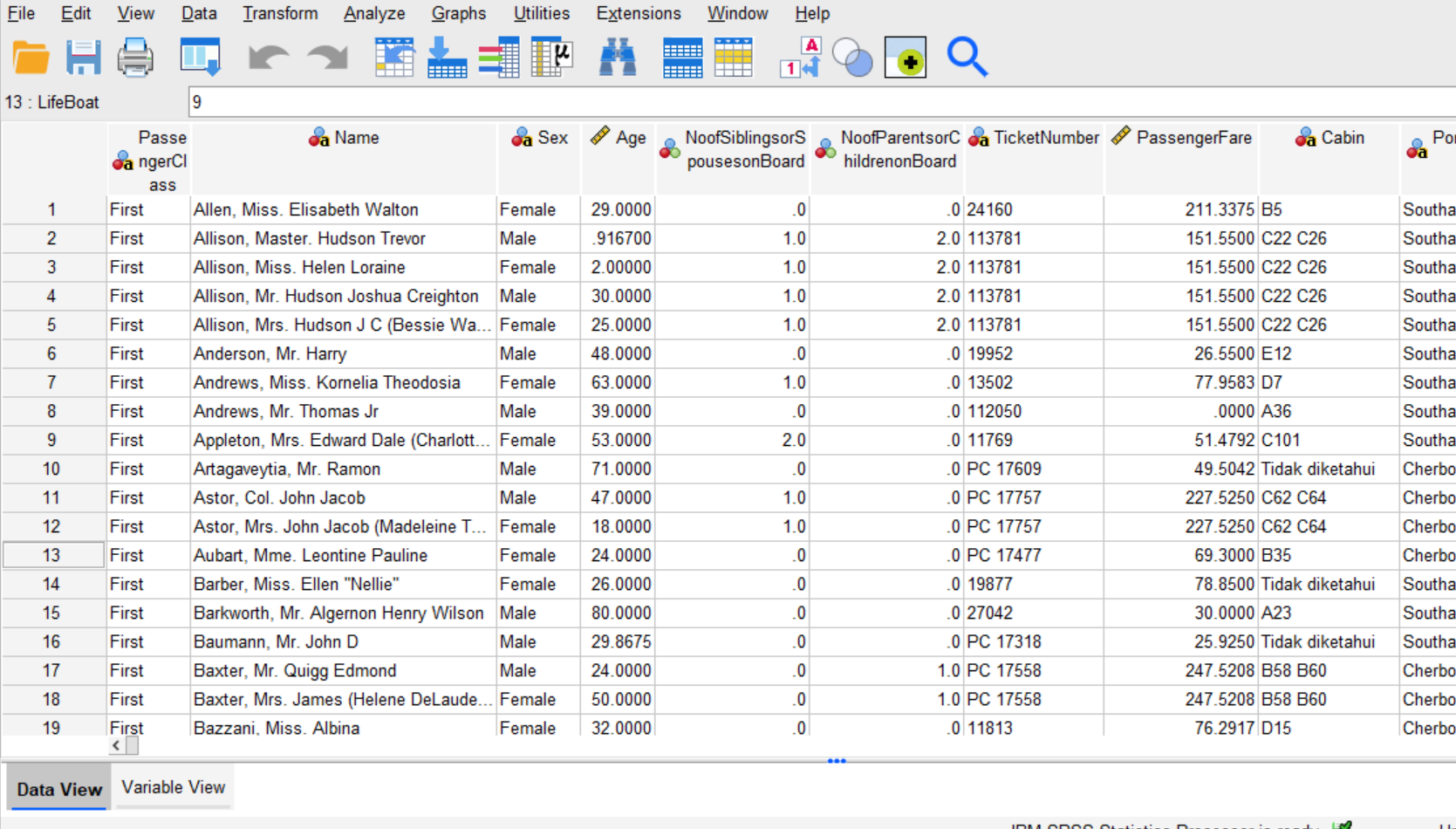
* Sebagian besar individu dalam dataset ini tidak berhasil bertahan hidup. Ini mengindikasikan adanya peristiwa atau kondisi yang menyebabkan tingkat kematian yang cukup tinggi.
* Perbandingan antara yang tidak bertahan hidup dan yang bertahan hidup adalah sekitar 1.61:1. Artinya, untuk setiap 2 individu, sekitar 1 individu yang berhasil bertahan hidup.

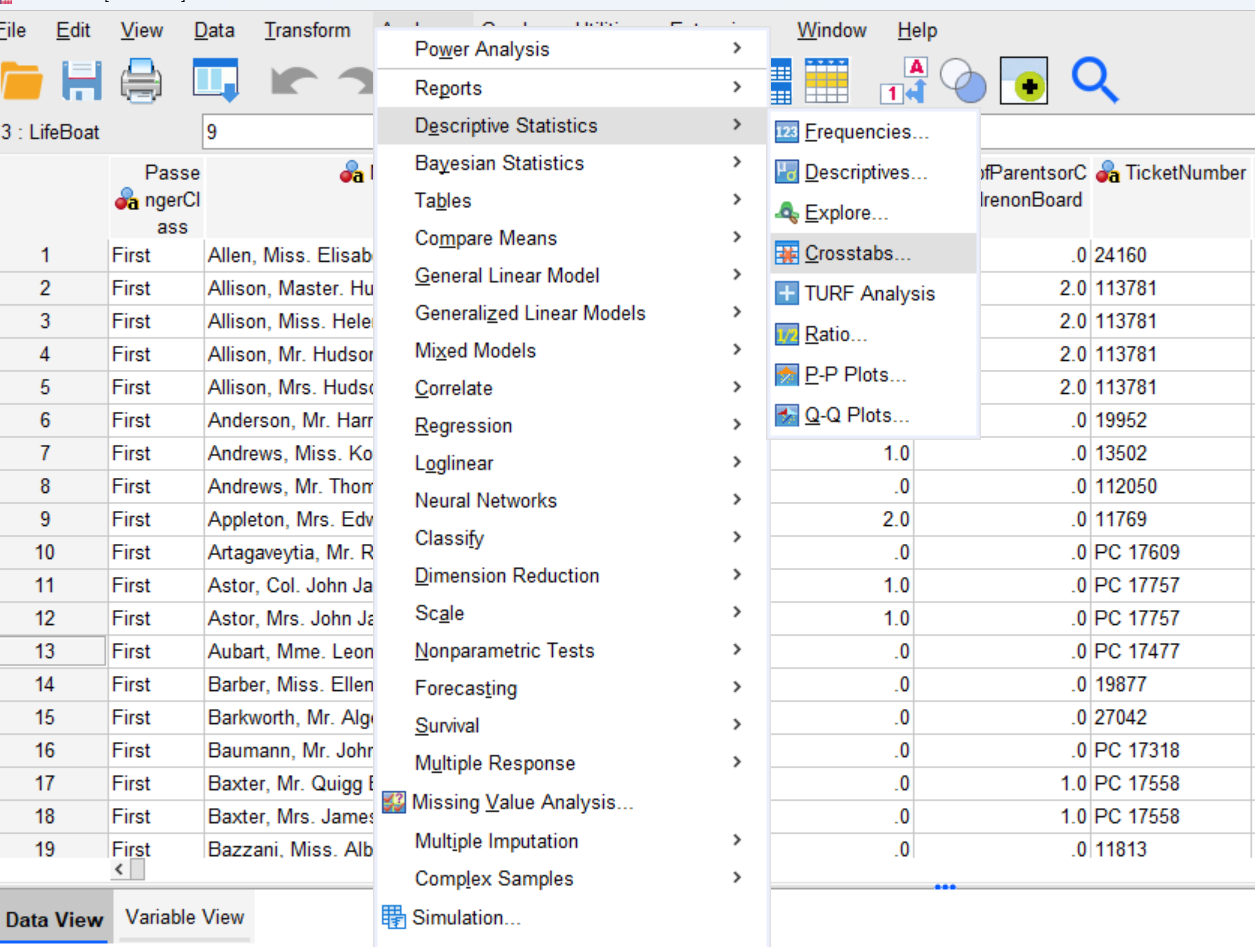
# 2. Analisis Bivariat

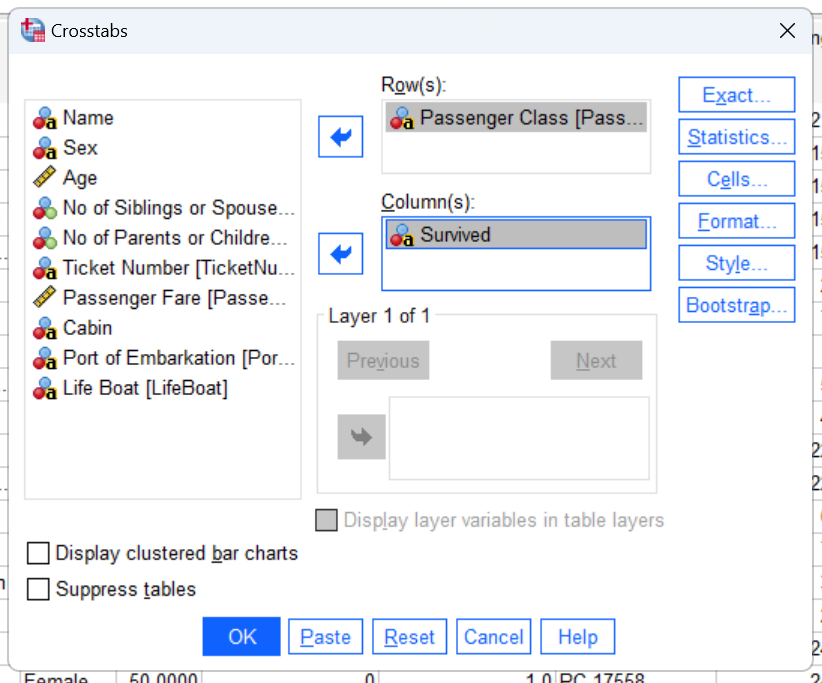
Pada laporan ini, analisis bivariat digunakan untuk mengeksplorasi hubungan antara dua variabel bertipe kategorikal, yaitu **Pclass** (kelas tiket penumpang) dan **Survived** (status kelangsungan hidup). Analisis ini bertujuan untuk mengetahui apakah terdapat hubungan signifikan antara kelas sosial penumpang di Titanic dengan kemungkinan mereka selamat atau tidak dalam peristiwa tenggelamnya kapal tersebut.

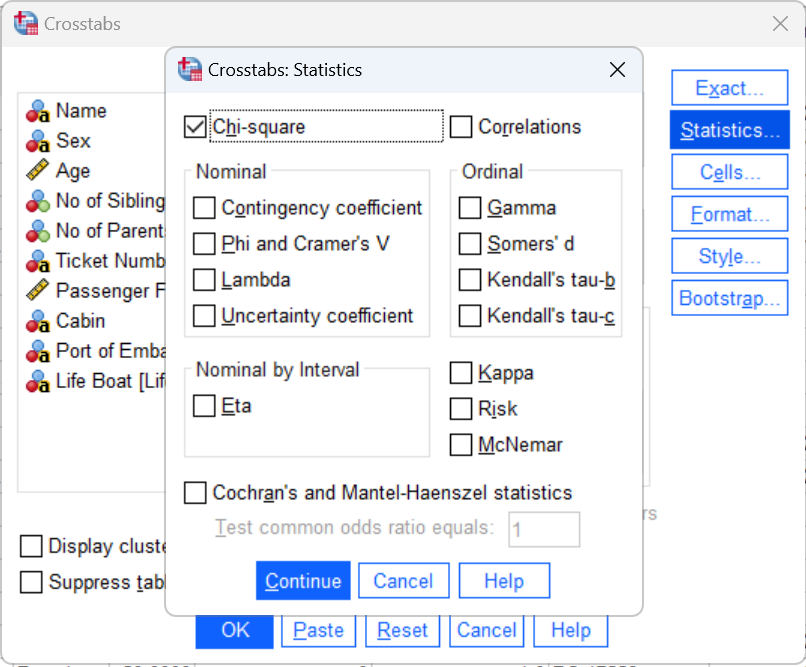
## A. Menggunakan Crosstabs

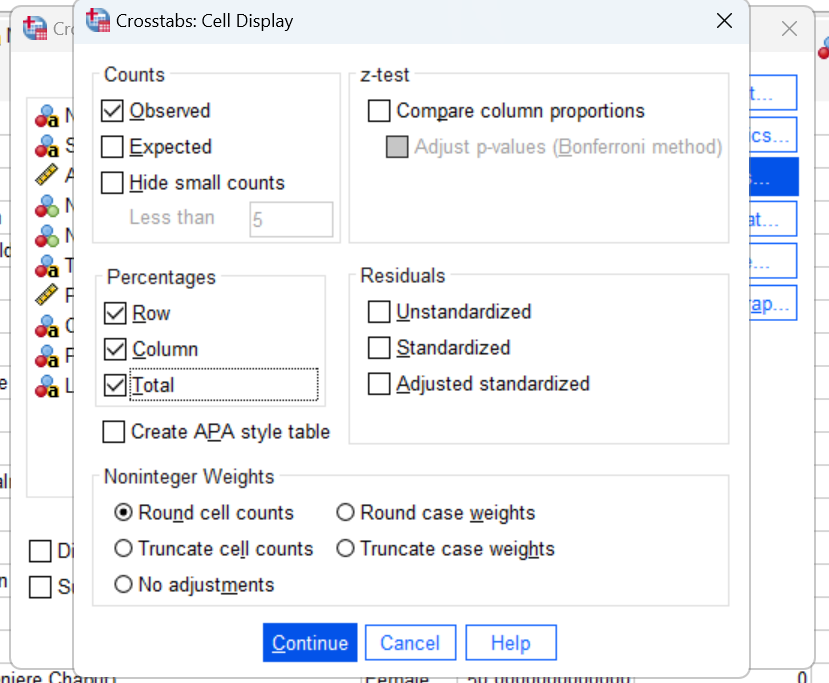
Dengan menggunakan tabel silang (*Crosstabs*) dan uji *Chi-Square*, distribusi kelangsungan hidup dalam setiap kelas tiket akan dianalisis. Uji *Chi-Square* akan menguji apakah pola kelangsungan hidup yang berbeda antar kelas tiket muncul secara signifikan atau hanya kebetulan saja. Hasil analisis ini diharapkan dapat memberikan wawasan apakah kelas tiket, yang mungkin mencerminkan akses penumpang terhadap fasilitas dan lokasi di kapal, mempengaruhi tingkat kelangsungan hidup mereka.



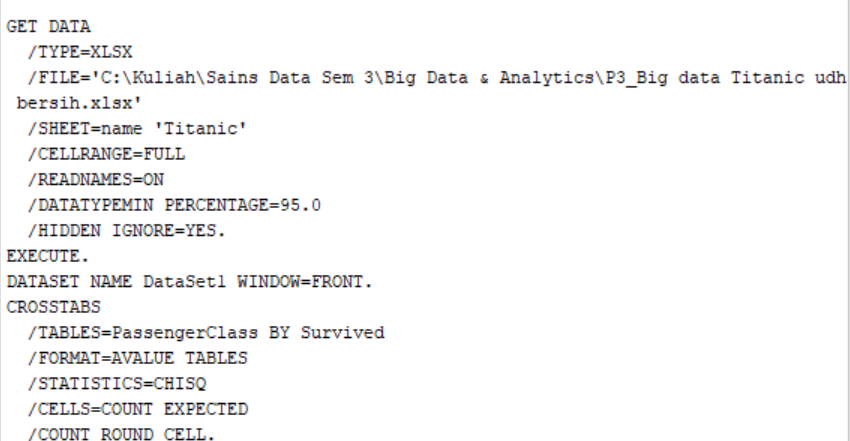


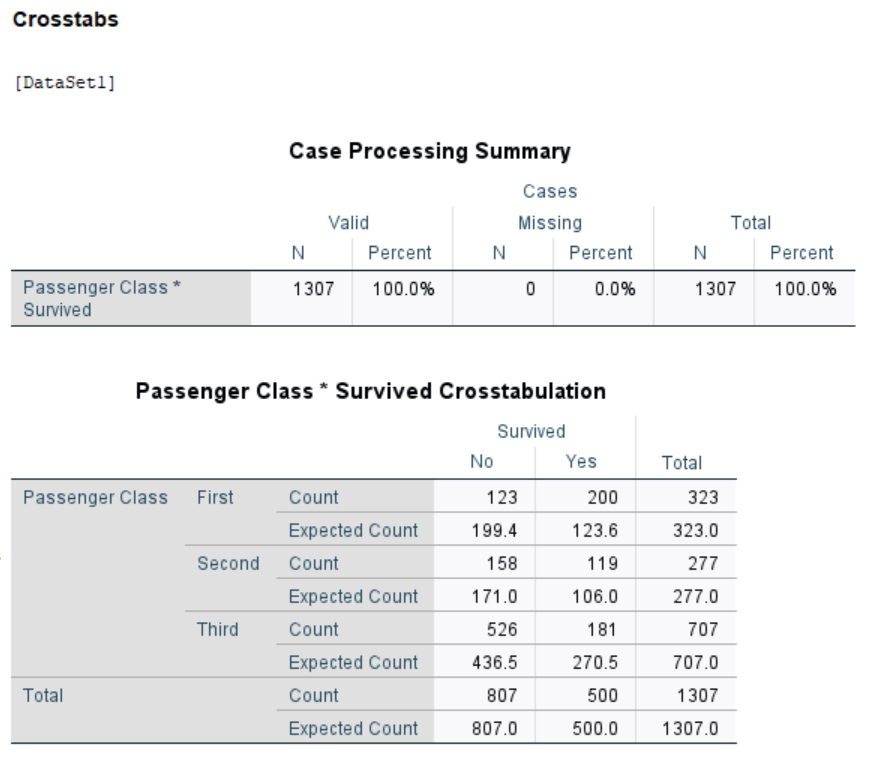


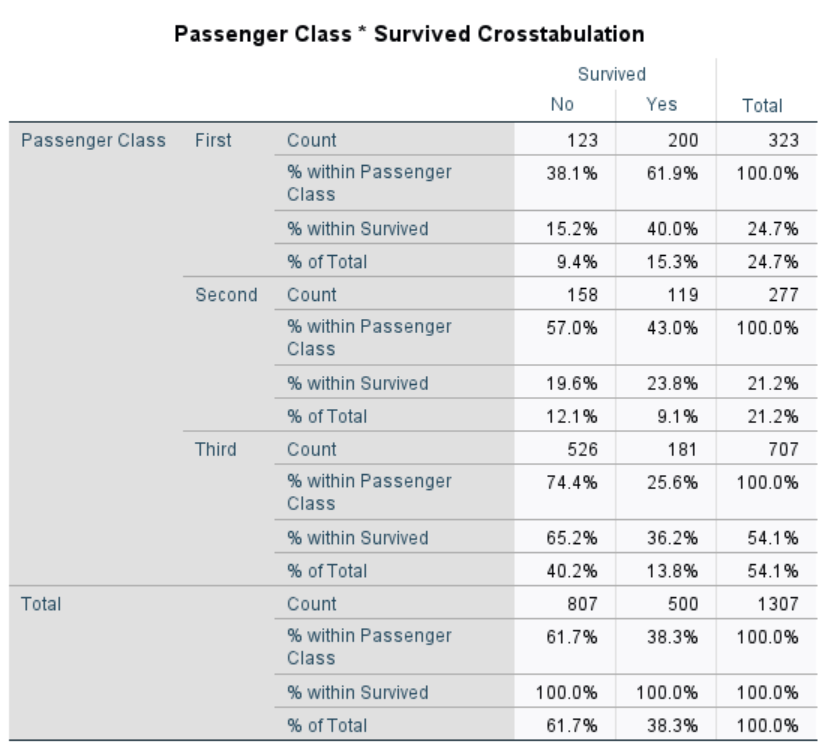


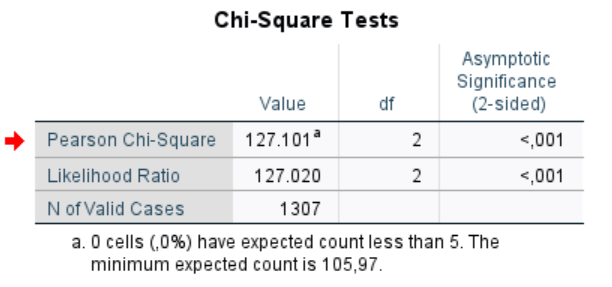


**Ouput:**









Dari hasil analisis tabel *Crosstabs* dan uji *Chi-Square* yang Anda lakukan di SPSS, berikut adalah interpretasinya:

### **1. Tabel Crosstab (Passenger Class \* Survived Crosstabulation):**

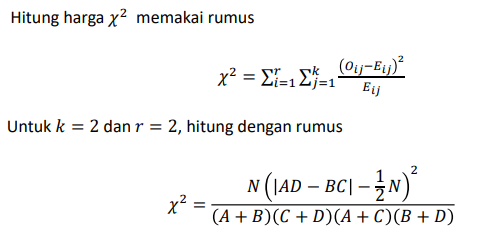
* Tabel ini menunjukkan distribusi kelangsungan hidup penumpang berdasarkan kelas tiket mereka yakni First, Second, dan Third.
* Di kolom **No** atau Tidak Selamat, terlihat bahwa sebagian besar penumpang dari kelas ketiga/Third Class tidak selamat, yakni 526 orang, sementara kelas pertama/First Class memiliki jumlah yang jauh lebih rendah, yakni 123 orang.
* Di kolom **Yes** atau Selamat, penumpang di kelas pertama memiliki angka kelangsungan hidup yang relatif tinggi, yakni 200 orang, dibandingkan dengan kelas ketiga ada 181 orang.
* Hal ini menunjukkan adanya perbedaan yang cukup signifikan dalam tingkat kelangsungan hidup berdasarkan kelas tiket.

**Distribusi Berdasarkan Kelas Tiket:**

* **Penumpang Kelas Pertama (First Class):** Sebanyak 38.1% tidak selamat, sementara 61.9% berhasil selamat.
* **Penumpang Kelas Kedua (Second Class):** 57.0% tidak selamat, dan 43.0% selamat.
* **Penumpang Kelas Ketiga (Third Class):** Sebanyak 74.4% tidak selamat, dan hanya 25.6% yang selamat.

### **2. Uji Chi-Square (Chi-Square Tests):**

**Rumus Uji Chi Square 2 Sampel:**

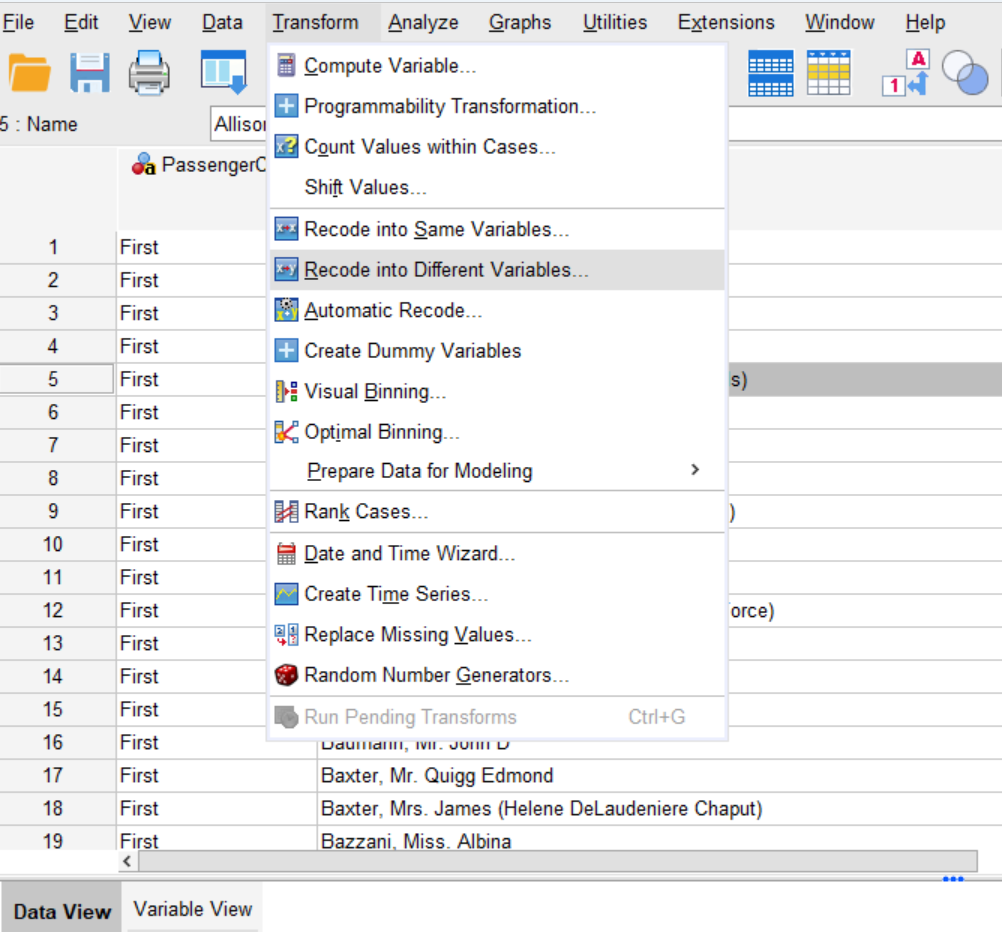


* Pada uji *Chi-Square*, nilai **Pearson Chi-Square** adalah 127.101 dengan tingkat signifikansi (<0.001).
* Karena nilai signifikansi lebih kecil dari 0.05, hasil ini menunjukkan bahwa terdapat hubungan yang signifikan antara kelas tiket penumpang dan kelangsungan hidup.
* Dengan kata lain, kelas tiket atau Pclass berpengaruh terhadap kemungkinan selamat atau tidaknya penumpang (Survived) dalam bencana tenggelamnya kapal Titanic ini.

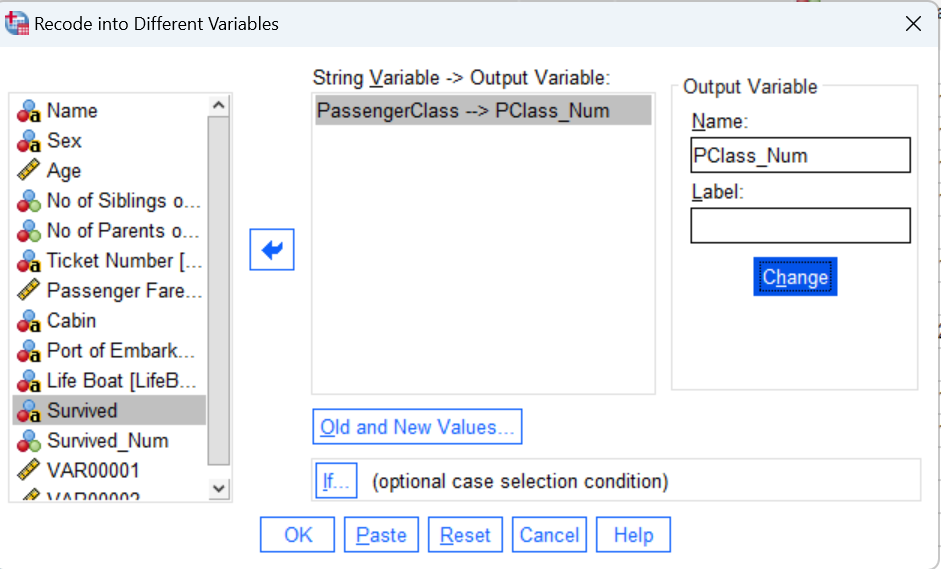
### **Kesimpulan**

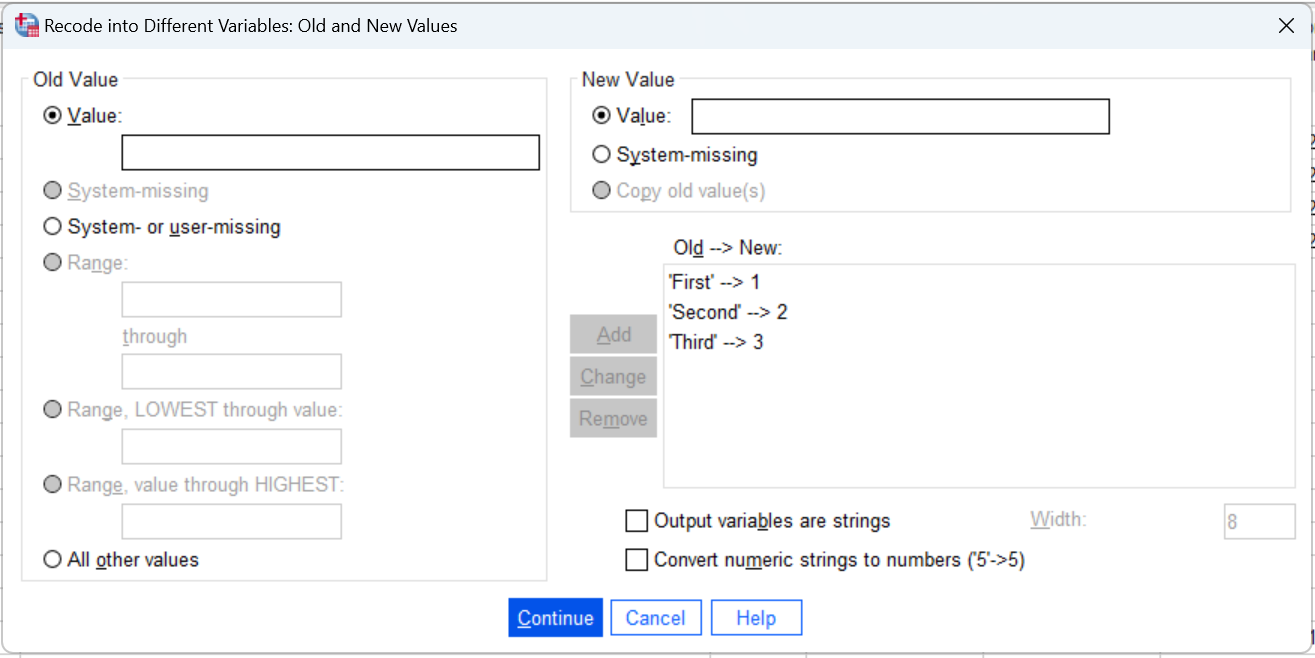
Berdasarkan analisis ini, dapat disimpulkan bahwa terdapat hubungan yang signifikan antara kelas tiket dan kelangsungan hidup penumpang Titanic. Penumpang di kelas pertama memiliki tingkat kelangsungan hidup yang lebih tinggi dibandingkan dengan kelas yang lebih rendah.

## B. Analisis Bivariat dengan Mengubah Kedua Kolom Pasangger Class dan Survived yang Bertipe Kategorikal ke Bentuk Numerik

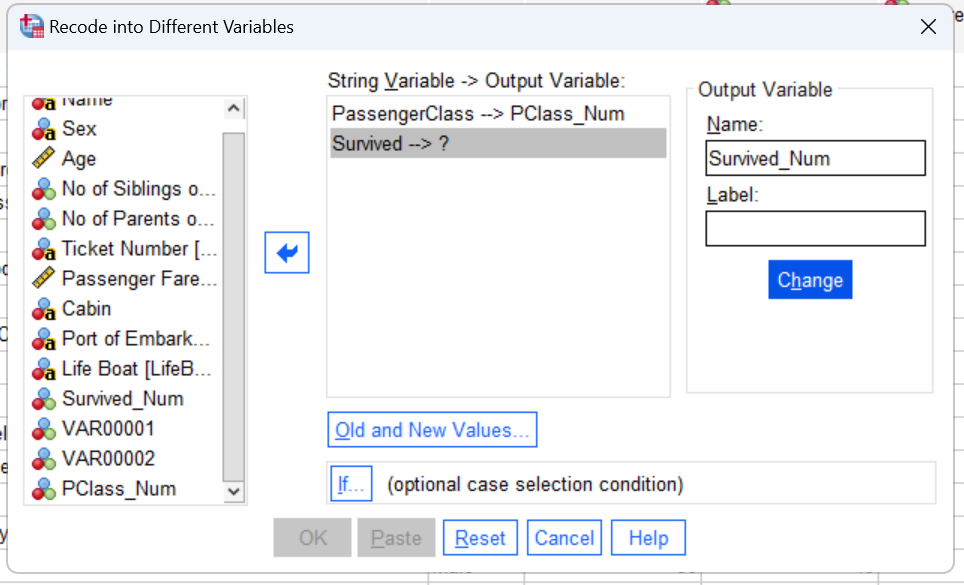


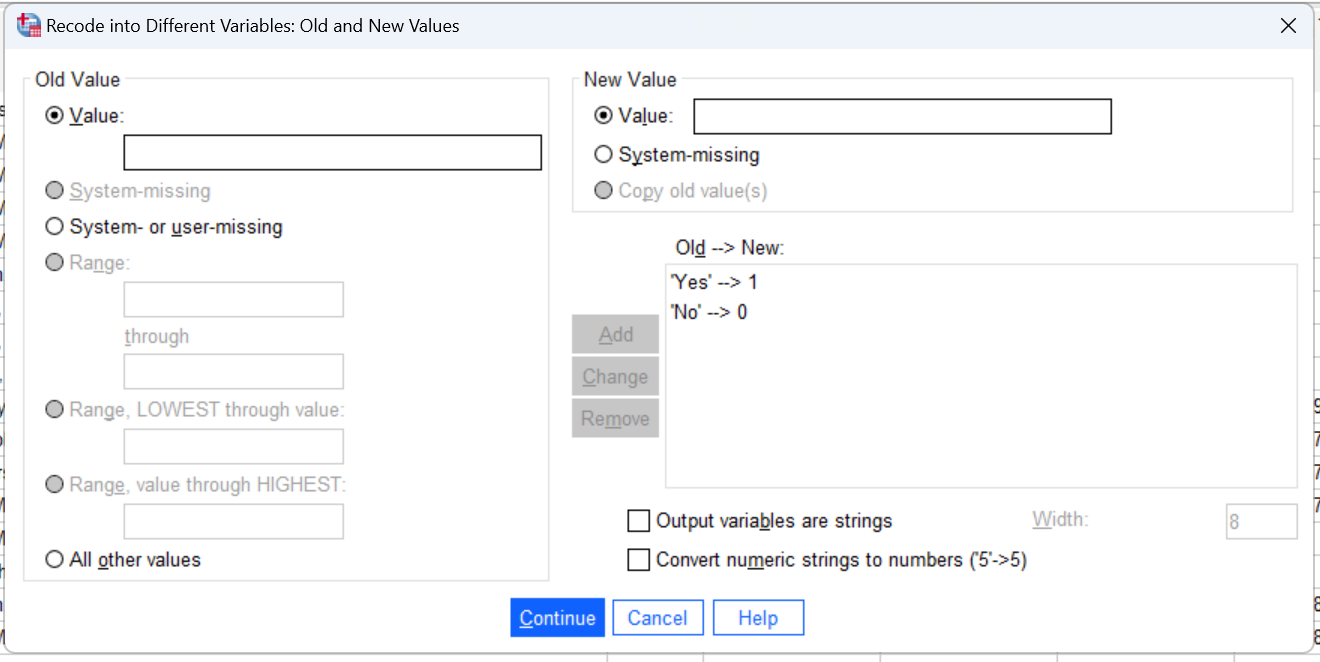
1. **Kolom Pclass dari tipe kategorikal ke numerik (First=1, Second=2, Third=3)**



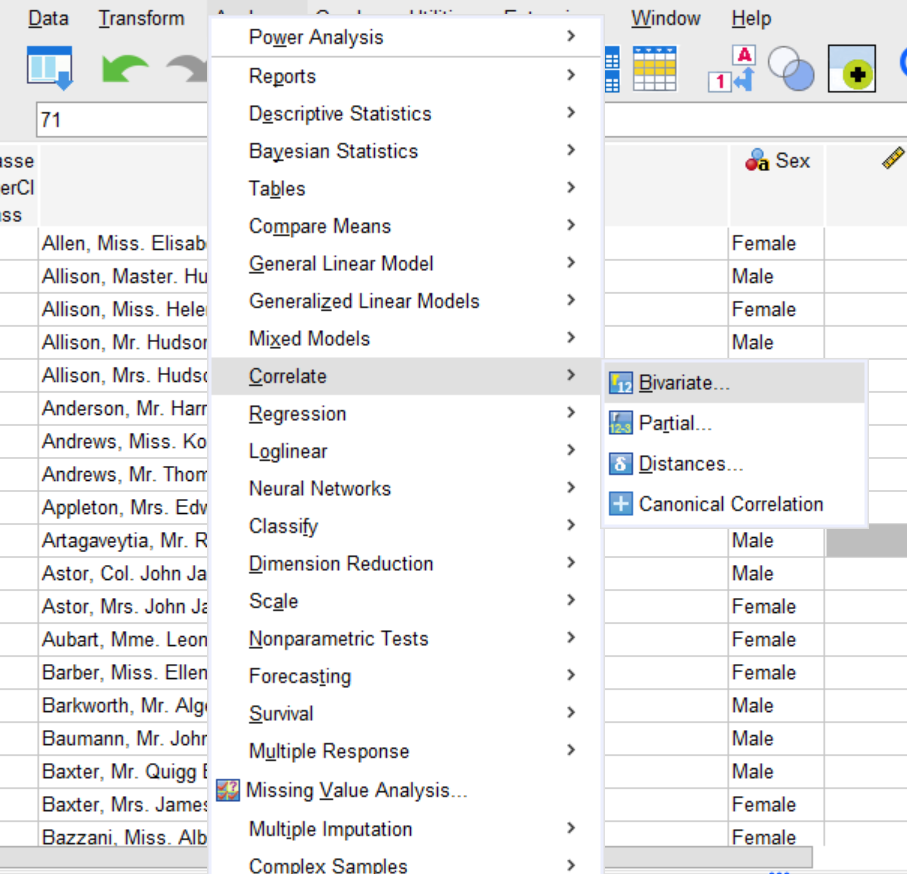


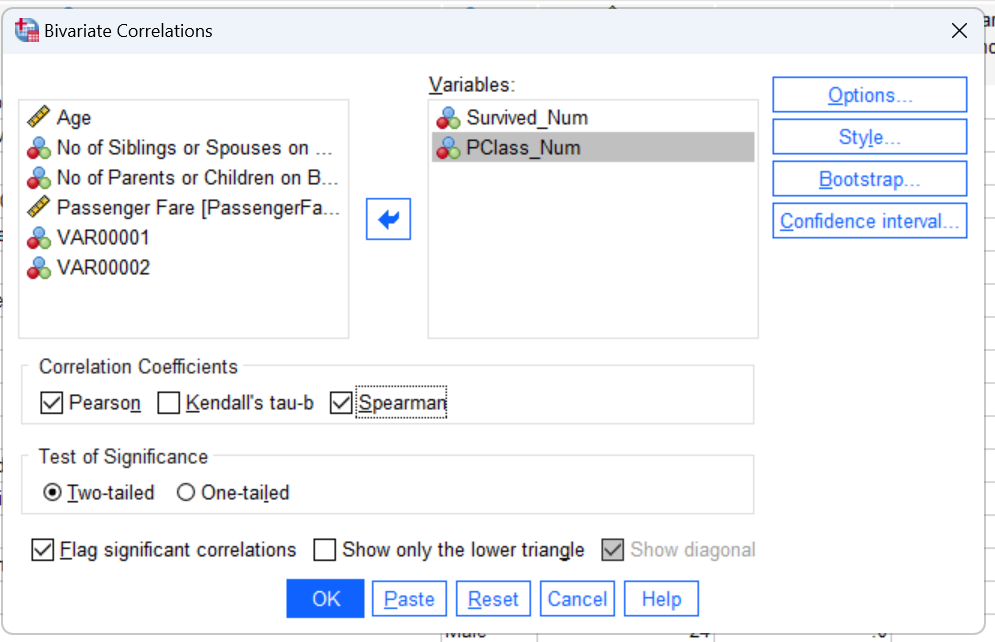
1. **Kolom Survived dari tipe kategorikal ke numerik (Yes=1, No=0)**





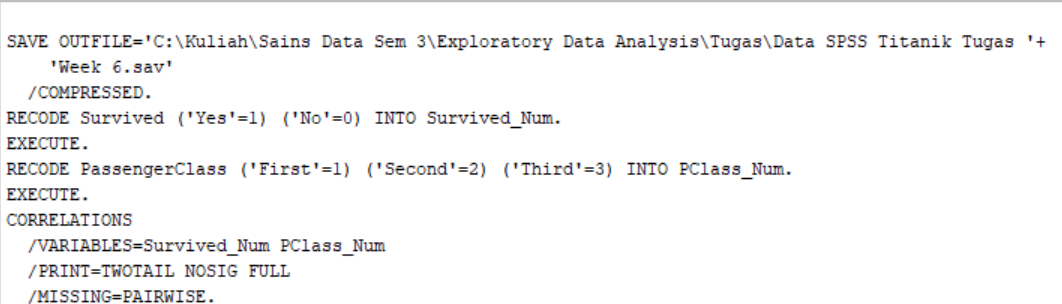
**Analisis Bivariat Kolom Pclass\_Num dan Survived\_Num**

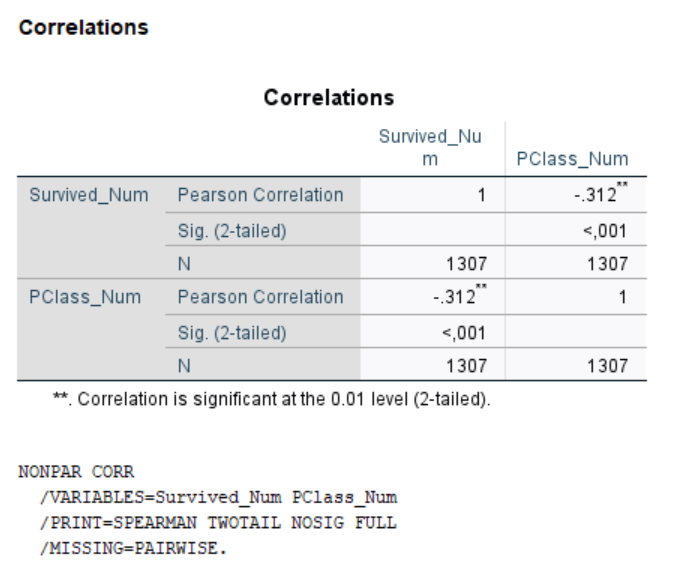




**Output:**

1. **Korelasi Pearson**





### **1. Interpretasi Korelasi Pearson**

Pada tabel pertama, terlihat bahwa hasil penggunaan **korelasi Pearson** untuk melihat hubungan antara variabel Survived\_Num dan PClass\_Num.

* **Pearson Correlation**

Nilai korelasi antara Survived\_Num dan PClass\_Num adalah **-0.312**. Ini menunjukkan adanya korelasi negatif yang lemah, tetapi signifikan. Korelasi negatif berarti ketika nilai PClass\_Num meningkat (artinya kelasnya lebih rendah, misalnya dari kelas 1 ke kelas 3), maka kecenderungan untuk bertahan hidup (Survived\_Num) menurun.

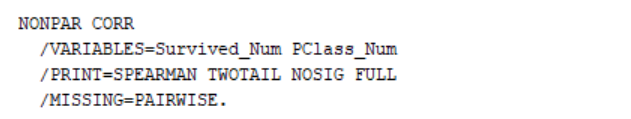
* **Signifikansi (Sig.)**

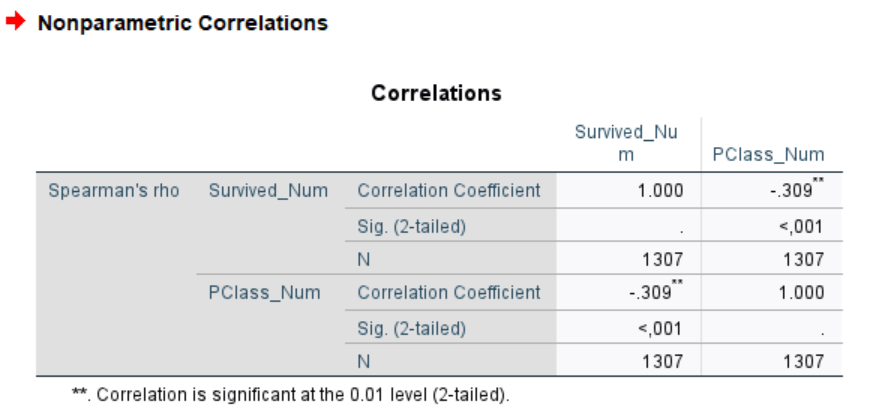
Nilai **p-value < 0.001**, yang artinya hasil ini signifikan pada level kepercayaan 99% (0.01 atau 1%). Dengan kata lain, kita bisa yakin bahwa hubungan yang terdeteksi bukanlah kebetulan.

* **Kesimpulan Pearson**

Dari hasil ini, kita bisa menyimpulkan bahwa kelas penumpang (PClass) memang mempengaruhi kemungkinan selamat atau tidaknya seorang penumpang di Titanic. Penumpang kelas yang lebih tinggi (kelas 1) lebih mungkin untuk selamat dibandingkan dengan kelas yang lebih rendah (kelas 3).

1. **Korelasi Spearman**





### **2. Interpretasi Korelasi Spearman**

Pada tabel kedua, kita menggunakan **korelasi Spearman** untuk variabel yang sama. Korelasi Spearman sering digunakan untuk data yang tidak berdistribusi normal atau untuk variabel kategori yang telah dikodekan menjadi angka.

* **Spearman's rho**

Nilai korelasi Spearman antara Survived\_Num dan PClass\_Num adalah **-0.309**, yang hampir sama dengan hasil Pearson. Ini juga menunjukkan korelasi negatif yang lemah, namun signifikan.

* **Signifikansi (Sig.)**

Hasil p-value juga di bawah 0.001, sehingga dapat disimpulkan bahwa hasil ini signifikan pada level kepercayaan 99%.

* **Kesimpulan Spearman**

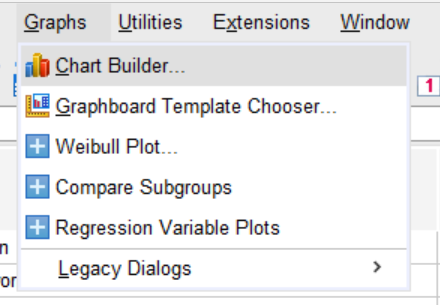
Korelasi Spearman memperkuat kesimpulan dari Pearson, yaitu adanya hubungan negatif antara kelas penumpang dengan kemungkinan selamat. Korelasi negatif menunjukkan bahwa semakin rendah kelas penumpang, semakin kecil kemungkinannya untuk selamat.

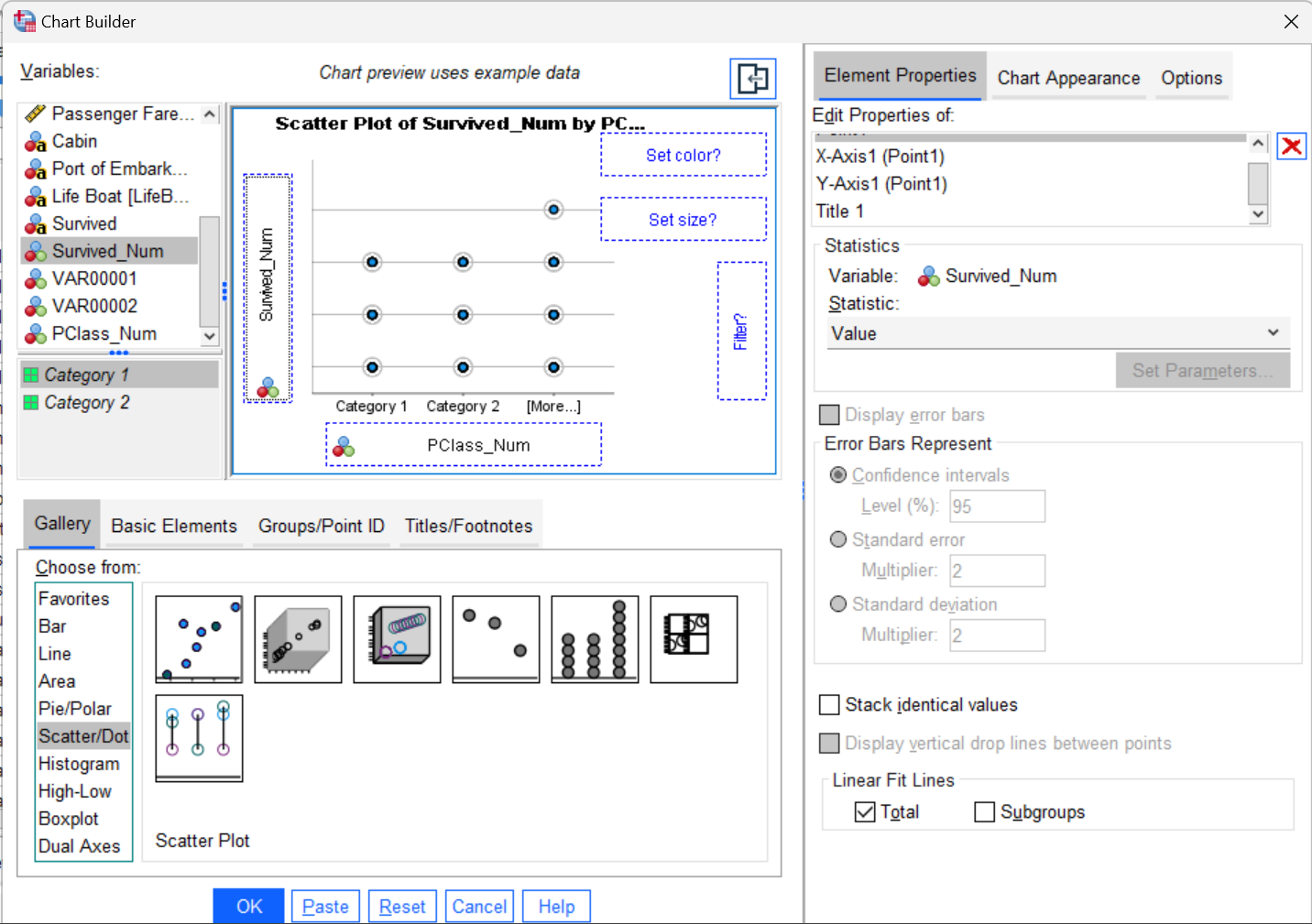
### **Catatan Tambahan**

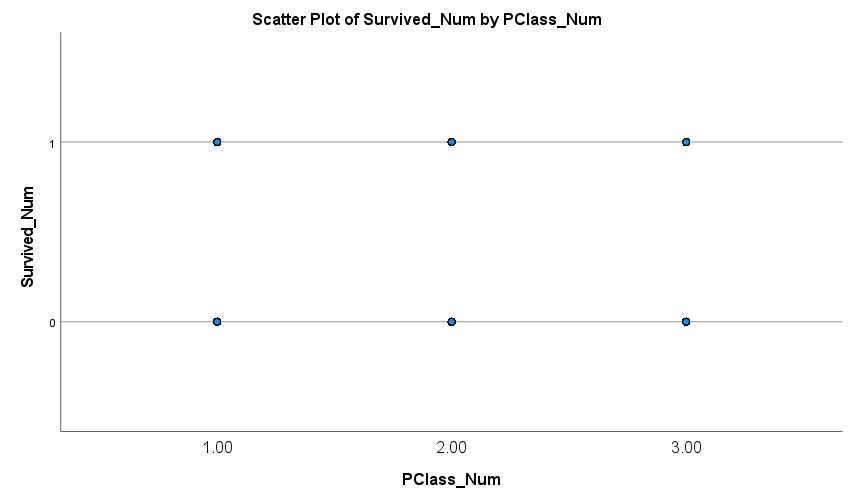
Baik Pearson maupun Spearman memberikan hasil yang konsisten dalam analisis ini, yaitu korelasi negatif yang signifikan antara Survived\_Num dan PClass\_Num. Namun, meskipun kita mengonversi variabel kategori menjadi numerik, korelasi ini masih perlu diinterpretasikan dengan hati-hati. Sebagai contoh, PClass dan Survived adalah data kategori asli, sehingga hubungan yang ditunjukkan oleh korelasi ini lebih mencerminkan perbedaan proporsi daripada hubungan linier.

Kesimpulannya, dari analisis ini kita bisa melihat bahwa kelas penumpang memiliki pengaruh terhadap tingkat keselamatan di Titanic, di mana penumpang kelas satu lebih mungkin untuk selamat dibandingkan penumpang kelas tiga.

## C. Menggunakan Grafik Scatter Plot







**Penjelasan:**

1. **Sumbu X (Horizontal)**

Ini menunjukkan PClass\_Num, yaitu kelas penumpang, dengan nilai 1, 2, dan 3. Nilai ini mewakili kelas 1 (kelas atas), kelas 2 (kelas menengah), dan kelas 3 (kelas bawah).

1. **Sumbu Y (Vertical)**

Ini menunjukkan Survived\_Num, yaitu status selamat, dengan nilai 0 dan 1. Nilai 0 artinya "Tidak Selamat," sementara nilai 1 artinya "Selamat".

1. **Titik-titik**

Setiap titik pada plot ini menunjukkan kombinasi dari PClass\_Num dan Survived\_Num. Karena variabel PClass\_Num dan Survived\_Num adalah data kategori yang telah diubah menjadi numerik, hasilnya adalah titik-titik pada nilai-nilai diskret (tidak ada variasi data kontinu di antara nilai-nilai tersebut).

1. **Interpretasi Awal**:
   * Kita mungkin akan melihat bahwa titik-titik terkumpul di posisi tertentu saja (misalnya di (1, 1), (2, 1), atau (3, 0)). Ini karena data kategori hanya bisa berada di nilai spesifik seperti ini.
   * Scatter plot ini menunjukkan bahwa kemungkinan ada pola tertentu, seperti penumpang dari kelas tertentu yang cenderung lebih besar peluangnya untuk selamat atau tidak selamat. Untuk melihat pola lebih jelas, kita bisa mencoba bar chart atau stacked bar chart yang membandingkan proporsi penumpang selamat per kelas.

Scatter plot ini memberikan visualisasi awal tentang distribusi data, tapi untuk analisis mendalam, kita mungkin perlu jenis grafik lain yang lebih cocok untuk data kategori, seperti bar chart pada analisis univariat sebelumnya yang menunjukkan jumlah atau proporsi penumpang selamat dan tidak selamat di setiap kelas.