

# **FINAL PROJECT ANALISIS MULTIVARIAT**

**ATA 2023/2024**

Analisis Korespondensi Sederhana



Disusun oleh:

Kayla Zahira Amadya (2206053890)

Analisis Multivariat (B)

Nomor Absen: 27

**PROGRAM STUDI STATISTIKA**

**UNIVERSITAS INDONESIA**

**2024**

# Bagian 1. Pendahuluan

**Masalah yang Dibahas:** Project ini bertujuan untuk memahami hubungan antara pekerjaan orang tua (ayah dan ibu) dengan kinerja akademis siswa dalam mata pelajaran matematika di sekolah menengah. Melalui analisis korespondensi sederhana, kita dapat mengetahui bagaimana kategori pekerjaan orang tua berhubungan dengan kategori nilai akhir siswa.

**Sumber Data:** Data yang digunakan berasal dari dataset "Student Performance in Secondary Education" yang tersedia di Kaggle. Dataset ini mencakup informasi mengenai kinerja siswa dalam mata pelajaran matematika serta berbagai atribut demografis lainnya. Dataset dapat diakses melalui tautan berikut: [Kaggle - Student Performance Data](#).

## Ukuran Data:

- Jumlah observasi: 395 siswa
- Jumlah pengukuran (kolom data): 33 atribut yang meliputi informasi pribadi, sosial, dan akademis siswa.

## Skala/Tipe Data:

1. **Kualitatif Nominal:** school, sex, address, famsize, Pstatus, Mjob, Fjob, reason, guardian, schoolsup, famsup, paid, activities, nursery, higher, internet, romantic, G3\_cat
2. **Kuantitatif Diskrit:** age, Medu, Fedu, traveltime, studytime, failures, famrel, freetime, goout, Dalc, Walc, health, absences, G1, G2, G3

## Arti/Maksud Pengukuran:

- school: Sekolah tempat siswa belajar (GP atau MS)
- sex: Jenis kelamin siswa (M atau F)
- age: Usia siswa
- address: Tipe tempat tinggal siswa (U: urban, R: rural)
- famsize: Ukuran keluarga siswa (LE3:  $\leq 3$ , GT3:  $> 3$ )
- Pstatus: Status hidup bersama orang tua (T: tinggal bersama, A: terpisah)
- Medu, Fedu: Tingkat pendidikan ibu dan ayah

- Mjob, Fjob: Pekerjaan ibu dan ayah
- reason: Alasan memilih sekolah
- guardian: Wali siswa
- traveltime: Waktu perjalanan ke sekolah
- studytime: Waktu belajar mingguan
- failures: Jumlah kegagalan kelas sebelumnya
- schoolsup: Dukungan sekolah tambahan (yes atau no)
- famsup: Dukungan keluarga (yes atau no)
- paid: Kursus berbayar tambahan (yes atau no)
- activities: Kegiatan ekstrakurikuler (yes atau no)
- nursery: Menghadiri tempat penitipan anak (yes atau no)
- higher: Keinginan untuk pendidikan tinggi (yes atau no)
- internet: Akses internet di rumah (yes atau no)
- romantic: Hubungan romantis (yes atau no)
- famrel: Kualitas hubungan keluarga (skala 1-5)
- freetime: Waktu luang setelah sekolah (skala 1-5)
- goout: Waktu bersosialisasi dengan teman (skala 1-5)
- Dalc: Konsumsi alkohol harian (skala 1-5)
- Walc: Konsumsi alkohol akhir pekan (skala 1-5)
- health: Kesehatan saat ini (skala 1-5)
- absences: Jumlah ketidakhadiran
- G1, G2, G3: Nilai periode pertama, kedua, dan akhir matematika

## Bagian 2: Pre-processing dan Analisis Deskriptif untuk Visualisasi Data

### o Langkah-langkah Pre-processing Data

1. Mengimpor Data dari CSV: Data diimpor dari file CSV untuk mendapatkan akses ke dataset yang akan dianalisis.

```
# Mengimpor data dari file CSV
data <- read.csv("C:/Users/Kayla/Downloads/kuliah/smt 4/anmul/student-mat.csv", sep=';')
```

2. Menghapus Nilai yang Hilang: Data yang mengandung nilai hilang dihapus untuk memastikan analisis tidak terganggu oleh ketidaksempurnaan data.

```
> # Identifikasi dan Penanganan Missing Values
> missing_values <- colSums(is.na(data))
> print(missing_values)
```

|          |           |        |          |            |            |           |
|----------|-----------|--------|----------|------------|------------|-----------|
| school   | sex       | age    | address  | famsize    | Pstatus    | Medu      |
| 0        | 0         | 0      | 0        | 0          | 0          | 0         |
| Fedu     | Mjob      | Fjob   | reason   | guardian   | traveltime | studytime |
| 0        | 0         | 0      | 0        | 0          | 0          | 0         |
| failures | schoolsup | famsup | paid     | activities | nursery    | higher    |
| 0        | 0         | 0      | 0        | 0          | 0          | 0         |
| internet | romantic  | famrel | freetime | goout      | Dalc       | Walc      |
| 0        | 0         | 0      | 0        | 0          | 0          | 0         |
| health   | absences  | G1     | G2       | G3         |            |           |
| 0        | 0         | 0      | 0        | 0          |            |           |

3. Mengubah Kategori Nilai Akhir (G3) menjadi Faktor: Mengkategorikan nilai akhir menjadi tiga grup, yaitu:

```
> # Mengkategorikan G3 ke dalam tiga grup: Low, Medium, High
> data$G3_cat <- cut(data$G3, breaks=c(0, 10, 14, 20), labels=c('Low', 'Medium', 'High'),
include.lowest=TRUE)
```

a) Low (0-10):

Kategori ini mencakup siswa dengan nilai akhir yang sangat rendah hingga cukup rendah. Rentang 0-10 dipilih untuk merepresentasikan siswa yang mungkin memiliki kesulitan besar dalam pelajaran matematika atau mengalami hambatan signifikan dalam mencapai hasil yang lebih baik.

b) Medium (11-14):

Kategori ini mencakup siswa dengan nilai akhir sedang. Rentang 11-14 dipilih untuk merepresentasikan siswa yang berada di tengah-tengah distribusi nilai, yang menunjukkan performa yang memadai namun tidak luar biasa.

c) High (15-20):

Kategori ini mencakup siswa dengan nilai akhir yang tinggi. Rentang 15-20 dipilih untuk merepresentasikan siswa yang menunjukkan performa yang baik hingga sangat baik dalam pelajaran matematika.

Pemilihan batasan ini dilakukan dengan pertimbangan bahwa kategori harus cukup luas untuk mencakup variasi dalam performa siswa tetapi juga cukup spesifik untuk memberikan informasi yang bermakna dalam analisis korespondensi sederhana.

4. Mengubah Variabel Kualitatif Menjadi Faktor: Variabel pekerjaan orang tua (Fjob dan Mjob) serta beberapa variabel kualitatif lainnya diubah menjadi faktor untuk analisis korespondensi.

```
# Mengubah variabel Fjob dan Mjob menjadi faktor
data$Fjob <- as.factor(data$Fjob)
data$Mjob <- as.factor(data$Mjob)
```

5. Memilih Variabel: Memilih variabel yang akan digunakan, yaitu Fjob, Mjob, G3\_cat

```
# Memilih variabel yang relevan
data <- data %>% select(Fjob, Mjob, G3_cat)
```

#### ○ Hasil dari Langkah Pre-processing

Langkah-langkah ini menghasilkan dataset yang bersih tanpa nilai yang hilang dan dengan kategori yang sesuai untuk analisis lebih lanjut. Transformasi variabel kuantitatif menjadi faktor agar variabel bisa digunakan untuk analisis korespondensi sederhana.

Proses pre-processing ini saling berkaitan karena semua langkah ini bertujuan untuk memastikan data siap dianalisis. Menghapus nilai yang hilang dan mengubah tipe data menjadi faktor memastikan integritas data dan kesesuaian dengan metode analisis korespondensi yang akan digunakan.

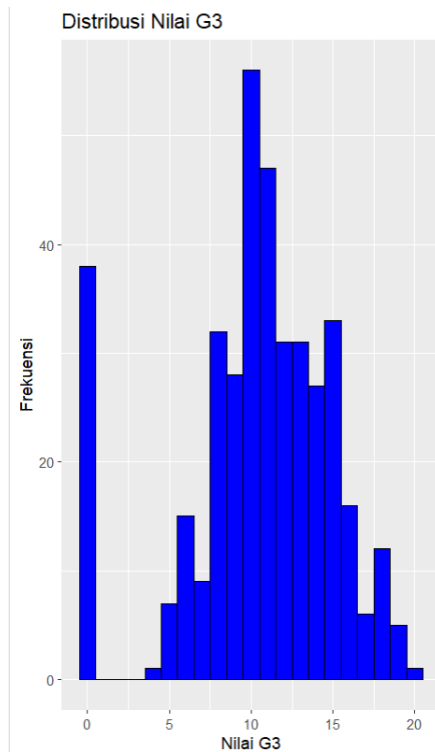
|    | Fjob     | Mjob     | G3_cat |
|----|----------|----------|--------|
| 1  | teacher  | at_home  | Low    |
| 2  | other    | at_home  | Low    |
| 3  | other    | at_home  | Low    |
| 4  | services | health   | High   |
| 5  | other    | other    | Low    |
| 6  | other    | services | High   |
| 7  | other    | other    | Medium |
| 8  | teacher  | other    | Low    |
| 9  | other    | services | High   |
| 10 | other    | other    | High   |
| 11 | health   | teacher  | Low    |
| 12 | other    | services | Medium |
| 13 | services | health   | Medium |
| 14 | other    | teacher  | Medium |
| 15 | other    | other    | High   |
| 16 | other    | health   | Medium |
| 17 | services | services | Medium |
| 18 | other    | other    | Low    |
| 19 | services | services | Low    |
| 20 | other    | health   | Low    |
| 21 | other    | teacher  | High   |
| 22 | health   | health   | High   |
| 23 | other    | teacher  | High   |
| 24 | other    | other    | Medium |
| 25 | health   | services | Low    |
| 26 | services | services | Low    |

Showing 1 to 26 of 395 entries, 3 total columns

## ○ Eksplorasi dan Visualisasi Data

### 1) Histogram Nilai G3

Visualisasi ini menunjukkan distribusi nilai akhir siswa. Histogram ini menunjukkan frekuensi nilai G3 dari 0 hingga 20. Dari histogram, terlihat bahwa sebagian besar siswa memiliki nilai di sekitar 10 hingga 15.



## 2) Statistik Deskriptif untuk Nilai G3

```
> # Descriptive Statistics
> descriptive_stats <- data %>%
+   summarise(
+     mean_G3 = mean(G3, na.rm = TRUE),
+     median_G3 = median(G3, na.rm = TRUE),
+     sd_G3 = sd(G3, na.rm = TRUE),
+     min_G3 = min(G3, na.rm = TRUE),
+     max_G3 = max(G3, na.rm = TRUE)
+   )
> print(descriptive_stats)
```

|   | mean_G3  | median_G3 | sd_G3    | min_G3 |
|---|----------|-----------|----------|--------|
| 1 | 10.41519 | 11        | 4.581443 | 0      |

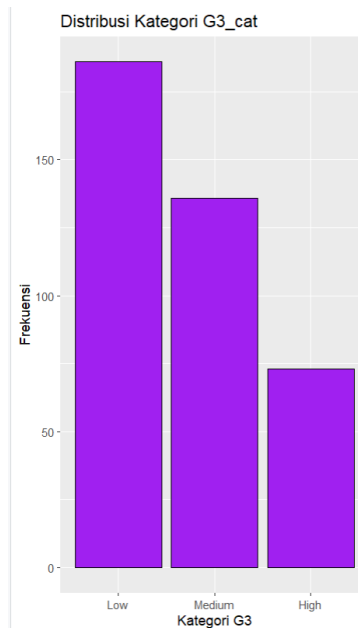
```
max_G3
1      20
```

Hasil analisis menunjukkan:

- Rata-rata (mean) nilai G3 adalah 10.415.
- Median nilai G3 adalah 11.
- Standar deviasi (sd) nilai G3 adalah 4.581.
- Nilai G3 minimum adalah 0.
- Nilai G3 maksimum adalah 20.

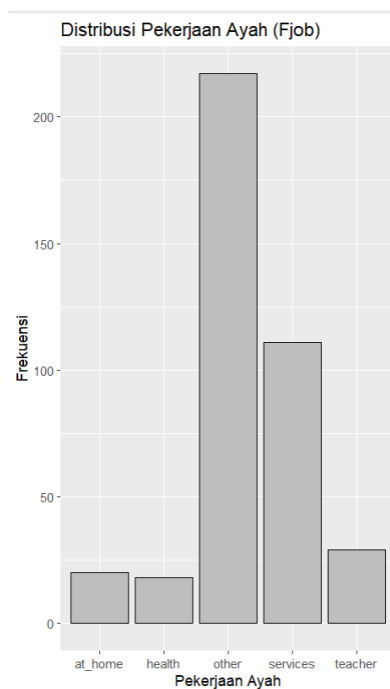
## 3) Bar Plot Kategori G3\_cat

Visualisasi ini menunjukkan distribusi kategori nilai akhir siswa.



Bar plot digunakan untuk menunjukkan distribusi kategori nilai akhir siswa yang telah dikategorikan ke dalam Low, Medium, dan High. Dari bar plot, terlihat bahwa mayoritas siswa berada di kategori Low, diikuti oleh Medium, dan sisanya di kategori High.

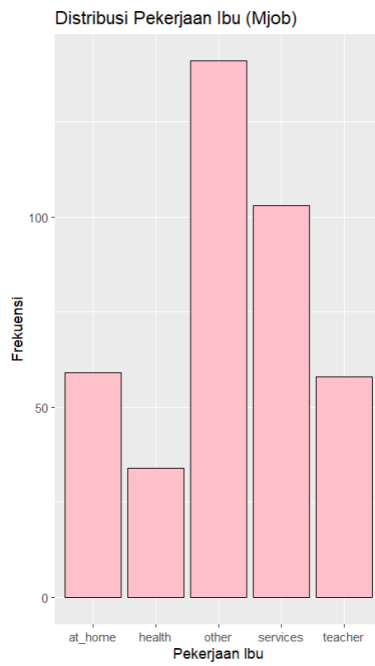
#### 4) Bar Plot untuk Variabel Pekerjaan Ayah/Father



Bar plot ini menunjukkan distribusi pekerjaan ayah siswa. Kategori pekerjaan termasuk at\_home, health, other, services, dan teacher. Distribusi menunjukkan bahwa sebagian besar ayah siswa bekerja di kategori "other" dan "services".

#### 5) Bar Plot untuk Variabel Pekerjaan Ibu/Mother





Bar plot ini menunjukkan distribusi pekerjaan ibu siswa. Seperti pada pekerjaan ayah, kategori pekerjaan ibu juga termasuk at\_home, health, other, services, dan teacher. Mayoritas ibu siswa bekerja di kategori other dan services.

## Bagian 3: Metode

Dalam project ini, metode yang digunakan adalah Analisis Korespondensi Sederhana (Simple Correspondence Analysis/SCA). Analisis Korespondensi merupakan teknik analisis data multivariat yang digunakan untuk menggambarkan hubungan antara dua variabel kategorikal dalam bentuk grafis dua dimensi. Teknik ini sangat berguna untuk memahami bagaimana kategori dari satu variabel berhubungan dengan kategori dari variabel lain dalam sebuah tabel kontingensi.

Analisis Korespondensi Sederhana beroperasi dengan mengkonversi frekuensi dalam tabel kontingensi menjadi skor dalam dua atau lebih dimensi (biasanya dua untuk kemudahan visualisasi). Skor ini dihitung sedemikian rupa sehingga jarak antara titik-titik pada plot mencerminkan kedekatan atau kemiripan antara kategori. Metode ini menggunakan dekomposisi nilai singular (Singular Value Decomposition - SVD) pada matriks chi-squared yang dinormalisasi, yang berasal dari tabel kontingensi.

### ○ Langkah-Langkah dalam Analisis Korespondensi

#### 1. Pembuatan Tabel Kontingensi

Data dikategorikan dan disusun dalam bentuk tabel kontingensi yang menunjukkan frekuensi gabungan antar-kategori variabel.

#### 2. Normalisasi Data

Matriks frekuensi diubah menjadi matriks profil yang menunjukkan proporsi relatif dari setiap kategori.

#### 3. Penghitungan Chi-Squared dan SVD:

Matriks chi-squared dihitung untuk menilai tingkat asosiasi antara baris dan kolom. SVD dari matriks ini menghasilkan nilai eigen dan vektor eigen yang digunakan untuk menghitung skor untuk visualisasi.

#### 4. Visualisasi

5. Skor yang dihasilkan digambarkan dalam plot dua dimensi, di mana jarak antar titik menggambarkan tingkat asosiasi antara kategori-kategori tersebut.

- **Alasan Menggunakan Metode Ini**

1. Visualisasi Data

Visualisasi dari SCA mudah untuk diinterpretasikan, maka dari itu sangat memungkinkan hubungan kompleks antara variabel kategorikal dapat dipahami dengan lebih mudah.

2. Interpretasi Mudah

Dengan mengurangi kompleksitas data menjadi dua dimensi, SCA memudahkan pengguna untuk melihat hubungan dan membuat interpretasi yang bermakna.

3. Tidak Memerlukan Asumsi Distribusi

SCA tidak memerlukan asumsi distribusi normal dari data, menjadikannya cocok untuk data kategorikal dan ordinal.

## Bagian 4: Pengolahan Data dan Analisis Hasil

### A. Analisis Korespondensi untuk Mjob vs Fjob

#### Matriks Korespondensi (P)

```
# Analisis Korespondensi untuk Mjob vs Fjob
table_Mjob_Fjob <- table(data$Mjob, data$Fjob)
matrix_p <- prop.table(table_Mjob_Fjob)
print("Matriks P (Korespondensi) untuk Mjob vs Fjob")

## [1] "Matriks P (Korespondensi) untuk Mjob vs Fjob"

print(matrix_p)

##
##          at_home    health    other    services    teacher
## at_home 0.017721519 0.005063291 0.083544304 0.037974684 0.005063291
## health  0.000000000 0.015189873 0.043037975 0.025316456 0.002531646
## other   0.012658228 0.005063291 0.263291139 0.060759494 0.015189873
## services 0.015189873 0.010126582 0.106329114 0.108860759 0.020253165
## teacher 0.005063291 0.010126582 0.053164557 0.048101266 0.030379747
```

Matriks ini menampilkan distribusi proporsional antara pekerjaan ibu dan ayah. Contohnya, proporsi ibu yang bekerja di sektor 'health' lebih tinggi relatif terhadap ayah yang bekerja sebagai 'teacher'. Ini mengindikasikan adanya pola tertentu dalam pemilihan pekerjaan antara ibu dan ayah.

#### Matriks Profil Baris (R) dan Kolom (C)

```
matrix_r <- prop.table(matrix_p, 1)
print("Matriks R (Profil Baris) untuk Mjob vs Fjob")

## [1] "Matriks R (Profil Baris) untuk Mjob vs Fjob"

print(matrix_r)

##
##          at_home    health    other    services    teacher
## at_home 0.11864407 0.03389831 0.55932203 0.25423729 0.03389831
## health  0.00000000 0.17647059 0.50000000 0.29411765 0.02941176
## other   0.03546099 0.01418440 0.73758865 0.17021277 0.04255319
## services 0.05825243 0.03883495 0.40776699 0.41747573 0.07766990
## teacher 0.03448276 0.06896552 0.36206897 0.32758621 0.20689655

matrix_c <- prop.table(matrix_p, 2)
print("Matriks C (Profil Kolom) untuk Mjob vs Fjob")

## [1] "Matriks C (Profil Kolom) untuk Mjob vs Fjob"

print(matrix_c)

##
##          at_home    health    other    services    teacher
## at_home 0.35000000 0.11111111 0.15207373 0.13513514 0.06896552
## health  0.00000000 0.33333333 0.07834101 0.09009009 0.03448276
## other   0.25000000 0.11111111 0.47926267 0.21621622 0.20689655
## services 0.30000000 0.22222222 0.19354839 0.38738739 0.27586207
## teacher 0.10000000 0.22222222 0.09677419 0.17117117 0.41379310
```

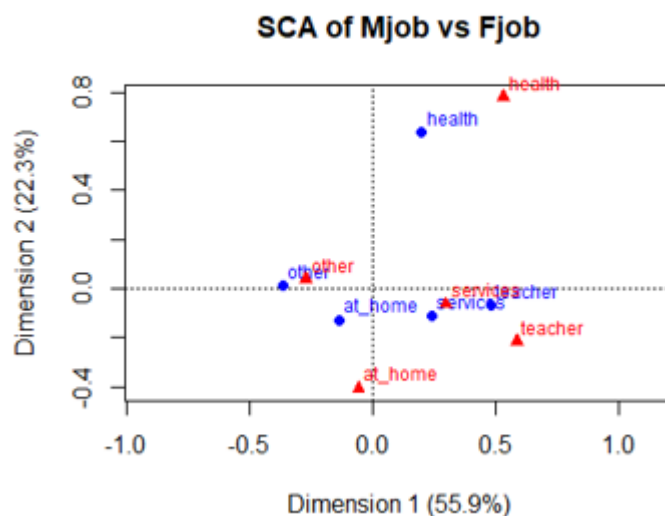
- Matriks Profil Baris menunjukkan distribusi persentase pekerjaan ibu terhadap pekerjaan ayah secara kondisional. Sebagai contoh, jika ibu bekerja di sektor 'health', distribusi pekerjaan ayah cenderung di sektor 'health' dan 'services'.

- Matriks Profil Kolom memperlihatkan seberapa signifikan kontribusi setiap pekerjaan ibu terhadap distribusi pekerjaan ayah.

## Visualisasi dengan Plot SCA

```
fit_Mjob_Fjob <- ca(table_Mjob_Fjob)
print(summary(fit_Mjob_Fjob))

##
## Principal inertias (eigenvalues):
##
## dim    value    % cum%    scree plot
## 1      0.103859 55.9 55.9 *****
## 2      0.041468 22.3 78.2 *****
## 3      0.030760 16.6 94.8 *****
## 4      0.009688  5.2 100.0 *
##
## Total: 0.185774 100.0
##
## Rows:
##   name    mass    qlt    inn    k=1 cor ctr    k=2 cor ctr
## 1 | at_h | 149 298 95 | -137 158 27 | -129 148 60 |
## 2 | hlth | 86 971 212 | 199 87 33 | 636 884 840 |
## 3 | othr | 357 913 283 | -367 912 462 | 12 1 1 |
## 4 | srvc | 261 678 148 | 241 553 146 | -114 125 82 |
## 5 | tchr | 147 723 262 | 485 709 332 | -67 14 16 |
##
## Columns:
##   name    mass    qlt    inn    k=1 cor ctr    k=2 cor ctr
## 1 | at_h | 51 417 111 | -58 8 2 | -409 409 204 |
## 2 | hlth | 46 955 229 | 533 305 125 | 779 650 667 |
## 3 | othr | 549 967 228 | -271 951 387 | 36 16 17 |
## 4 | srvc | 281 770 185 | 300 734 243 | -66 35 29 |
## 5 | tchr | 73 625 246 | 586 550 243 | -216 75 83 |
##
plot(fit_Mjob_Fjob, main = "SCA of Mjob vs Fjob")
```



Plot Analisis Korespondensi Sederhana (SCA) untuk Mjob vs Fjob menunjukkan hubungan antara jenis pekerjaan ibu (Mjob) dan ayah (Fjob) dalam data. Pada visualisasi ini, kita dapat melihat bagaimana distribusi pekerjaan antara ibu dan ayah berkelompok berdasarkan kesamaan pekerjaan:

- Titik "health" untuk Mjob dan Fjob yang saling berdekatan menunjukkan bahwa pasangan di mana ibu bekerja di sektor kesehatan cenderung memiliki pasangan yang juga bekerja di sektor yang sama.
- Titik "teacher" untuk kedua variabel yang berdekatan menunjukkan bahwa banyak ibu dan ayah bekerja sebagai guru, mencerminkan tren pasangan dalam profesi pendidikan yang sama.
- Titik "services" dan "at\_home" menyebar lebih luas antar dimensi, menandakan bahwa pekerjaan di sektor layanan dan pekerjaan rumahan cenderung tidak terkonsentrasi pada satu jenis pekerjaan pasangan, mengindikasikan variasi yang lebih luas dalam pilihan karier antara ibu dan ayah.
- "other" mencakup variasi pekerjaan lain yang tidak spesifik ke satu sektor, yang ditunjukkan oleh posisi yang lebih terpusat dan tersebar, mencerminkan heterogenitas dalam jenis pekerjaan antara ibu dan ayah.

## Uji Chi-squared

```
# Uji Chi-squared untuk Mjob vs Fjob
chi_sq_Mjob_Fjob <- chisq.test(table_Mjob_Fjob)

## Warning in chisq.test(table_Mjob_Fjob): Chi-squared approximation may be
## incorrect

print(chi_sq_Mjob_Fjob)

##
## Pearson's Chi-squared test
##
## data: table_Mjob_Fjob
## X-squared = 73.381, df = 16, p-value = 2.534e-09
```

Uji Chi-squared menghasilkan p-value yang sangat kecil ( $2.534e-09$ ), yang mengindikasikan adanya hubungan yang signifikan secara statistik antara jenis pekerjaan ibu dan ayah. Walaupun terdapat peringatan bahwa aproksimasi Chi-squared mungkin tidak tepat, nilai p-value yang sangat rendah ini menunjukkan bahwa ketergantungan antara variabel cukup kuat.

## Uji Fisher's Exact Test

```
# Uji Fisher's Exact Test untuk Mjob vs Fjob dengan simulasi
fisher_Mjob_Fjob <- fisher.test(table_Mjob_Fjob, simulate.p.value = TRUE, B = 10000)
print(fisher_Mjob_Fjob)

##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 10000 replicates)
##
## data: table_Mjob_Fjob
## p-value = 9.999e-05
## alternative hypothesis: two.sided
```

Uji Fisher's Exact Test mengkonfirmasi hasil uji Chi-squared dengan memberikan p-value yang juga sangat rendah, memperkuat bukti bahwa pekerjaan ibu dan ayah tidak independen. Penggunaan uji ini tepat untuk tabel kontingensi dengan jumlah sampel yang lebih kecil atau distribusi yang tidak merata.

Dalam Analisis Korespondensi Sederhana (SCA), hasil yang diperoleh bisa menjadi bias jika asumsi independensi tidak terpenuhi. Ketika uji Chi-squared menunjukkan p-value  $< 0.05$ , ini menunjukkan adanya ketergantungan yang signifikan antara variabel-variabel tersebut, yang bisa mempengaruhi hasil analisis dan interpretasinya.

Namun, dalam konteks SCA, meskipun uji independensi menunjukkan ketergantungan antara variabel 'Mjob' dan 'Fjob', Analisis Korespondensi Sederhana (SCA) masih diaplikasikan karena analisis masih dapat memberikan wawasan berharga tentang hubungan antar kategori ini. Tujuan utama SCA adalah untuk memvisualisasikan dan mengeksplorasi pola dalam data kategorikal. Meski demikian, hasil harus ditafsirkan dengan hati-hati dan pertimbangan terhadap potensi bias ini harus disampaikan secara jelas.

## B. Analisis Korespondensi untuk Mjob vs G3\_cat

### Matriks Korespondensi (P)

```
# Analisis Korespondensi untuk Mjob vs G3_cat
table_Mjob_G3 <- table(data$Mjob, data$G3_cat)
matrix_p <- prop.table(table_Mjob_G3)
print("Matriks P (Korespondensi) untuk Mjob vs G3_cat")

## [1] "Matriks P (Korespondensi) untuk Mjob vs G3_cat"

print(matrix_p)

##
##           Low      Medium      High
## at_home  0.09873418 0.03544304 0.01518987
## health   0.02531646 0.03291139 0.02784810
## other    0.17468354 0.14177215 0.04050633
## services 0.10126582 0.09367089 0.06582278
## teacher  0.07088608 0.04050633 0.03544304
```

Matriks korespondensi menampilkan distribusi proporsional antara pekerjaan ibu dan kategori nilai G3 siswa. Proporsi tinggi ibu yang bekerja di sektor 'health' berhubungan dengan nilai G3 yang tinggi, menunjukkan pola dimana pekerjaan ibu di sektor tertentu dapat berpengaruh atau memiliki korelasi dengan performa akademik anak.

### Matriks Profil Baris (R) dan Kolom (C)

```
matrix_r <- prop.table(matrix_p, 1)
print("Matriks R (Profil Baris) untuk Mjob vs G3_cat")

## [1] "Matriks R (Profil Baris) untuk Mjob vs G3_cat"

print(matrix_r)

##
##           Low      Medium      High
## at_home  0.6610169 0.2372881 0.1016949
## health   0.2941176 0.3823529 0.3235294
## other    0.4893617 0.3971631 0.1134752
## services 0.3883495 0.3592233 0.2524272
## teacher  0.4827586 0.2758621 0.2413793

matrix_c <- prop.table(matrix_p, 2)
print("Matriks C (Profil Kolom) untuk Mjob vs G3_cat")

## [1] "Matriks C (Profil Kolom) untuk Mjob vs G3_cat"

print(matrix_c)

##
##           Low      Medium      High
## at_home  0.20967742 0.10294118 0.08219178
## health   0.05376344 0.09558824 0.15068493
## other    0.37096774 0.41176471 0.21917808
## services 0.21505376 0.27205882 0.35616438
## teacher  0.15053763 0.11764706 0.19178082
```

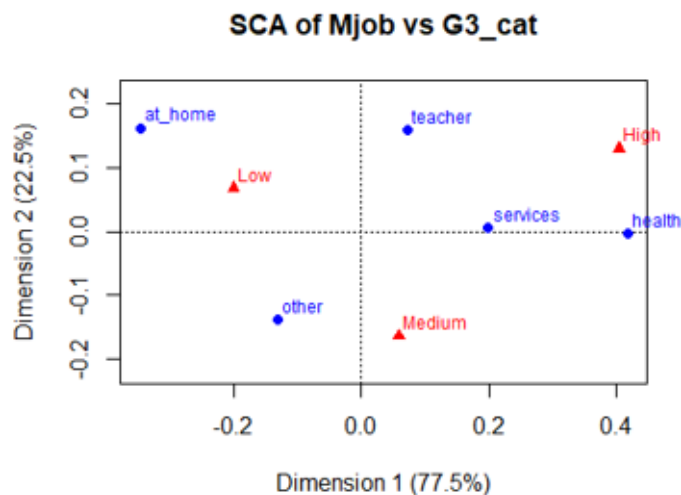
- Matriks Profil Baris menunjukkan bagaimana distribusi kategori nilai G3 terkondisikan pada pekerjaan ibu. Misalnya, jika ibu bekerja di 'health', maka distribusi nilai G3 cenderung lebih tinggi.



- Matriks Profil Kolom menggambarkan kontribusi relatif dari setiap jenis pekerjaan ibu terhadap kategori nilai G3, memberikan wawasan tentang seberapa signifikan setiap pekerjaan ibu dalam menentukan distribusi nilai G3.

## Visualisasi dengan Plot SCA

```
|fit_Mjob_G3 <- ca(table_Mjob_G3)
print(summary(fit_Mjob_G3))
## Principal inertias (eigenvalues):
##
## dim      value      % cum%      scree plot
## 1      0.050419  77.5  77.5 *****
## 2      0.014596  22.5 100.0 *****
## -----
## Total: 0.065015 100.0
##
## Rows:
##   name  mass  qlt  inr  k=1 cor ctr  k=2 cor ctr
## 1 | at_h | 149 1000 339 | -348 820 358 | 163 180 272 |
## 2 | hith |  86 1000 231 |  418 1000 298 |  -2   0   0 |
## 3 | othr | 357 1000 200 | -131 474 122 | -138 526 468 |
## 4 | srvc | 261 1000 160 |  200 999 206 |   6   1   1 |
## 5 | tchr | 147 1000  71 |   74 173  16 | 161 827 260 |
##
## Columns:
##   name  mass  qlt  inr  k=1 cor ctr  k=2 cor ctr
## 1 | Low | 471 1000 327 | -201 896 378 |  69 104 151 |
## 2 | Medm | 344 1000 160 |   58 111  23 | -164 889 633 |
## 3 | High | 185 1000 513 |  404 906 599 | 131  94 216 |
plot(fit_Mjob_G3, main = "SCA of Mjob vs G3_cat")
```



- Titik yang mewakili 'teacher' berada dekat dengan 'High', menandakan bahwa anak-anak dari ibu yang bekerja sebagai guru cenderung memiliki nilai akademik yang tinggi.
- Titik 'health' juga berada dekat dengan 'High', menunjukkan bahwa pekerjaan ibu di sektor kesehatan mungkin memiliki hubungan dengan pencapaian akademik yang lebih baik.

- Titik 'at\_home' dan 'other' lebih dekat ke 'Low', menunjukkan bahwa ibu yang tidak bekerja atau bekerja di luar sektor formal cenderung memiliki anak-anak dengan nilai akademik lebih rendah.

Hal ini mengindikasikan bahwa pekerjaan ibu di sektor formal seperti pendidikan dan kesehatan mungkin memberi kondisi yang lebih mendukung bagi keberhasilan akademik anak-anak mereka. Plot ini memberikan pandangan yang berguna tentang bagaimana faktor sosioekonomi keluarga, seperti jenis pekerjaan ibu, dapat berpengaruh terhadap hasil pendidikan anak.

## Uji Chi-squared

```
# Uji Chi-squared untuk Mjob vs G3_cat
chi_sq_Mjob_G3 <- chisq.test(table_Mjob_G3)
print(chi_sq_Mjob_G3)
## Pearson's Chi-squared test
##
## data:  table_Mjob_G3
## X-squared = 25.681, df = 8, p-value = 0.001191
```

Memberikan p-value yang sangat kecil, mengindikasikan adanya hubungan yang signifikan secara statistik antara jenis pekerjaan ibu dan kategori nilai G3. Meskipun ada peringatan bahwa aproksimasi mungkin tidak tepat, nilai p-value ini menegaskan ketergantungan yang kuat antara variabel.

## Uji Fisher's Exact Test

```
# Uji Fisher's Exact Test untuk Mjob vs G3_cat dengan simulasi
fisher_Mjob_G3 <- fisher.test(table_Mjob_G3, simulate.p.value = TRUE, B = 10000)
print(fisher_Mjob_G3)
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 10000 replicates)
##
## data:  table_Mjob_G3
## p-value = 0.0009999
## alternative hypothesis: two.sided
```

P-value yang sangat rendah, memperkuat bukti bahwa distribusi pekerjaan ibu dan kategori nilai G3 tidak independen. Uji ini penting untuk validasi hasil Chi-squared.

Meskipun uji independensi menunjukkan ketergantungan antara variabel 'Mjob' dan 'G3\_cat', Analisis Korespondensi Sederhana (SCA) masih diaplikasikan. Hal ini karena SCA bertujuan untuk memvisualisasikan dan mengeksplorasi hubungan antara

kategori. Alasan utama penggunaan SCA dalam konteks ini adalah untuk memahami bagaimana pekerjaan ibu berhubungan dengan kinerja akademis siswa.

### C. Analisis Korespondensi untuk Fjob vs G3\_cat

#### Matriks Korespondensi (P)

```
# Analisis Korespondensi untuk G3_cat vs Fjob
table_G3_Fjob <- table(data$G3_cat, data$Fjob)
matrix_p <- prop.table(table_G3_Fjob)
print("Matriks P (Korespondensi) untuk G3_cat vs Fjob")

## [1] "Matriks P (Korespondensi) untuk G3_cat vs Fjob"

print(matrix_p)

##
##          at_home    health    other    services    teacher
## Low    0.02278481 0.02025316 0.27088608 0.13417722 0.02278481
## Medium 0.01772152 0.01518987 0.19493671 0.09620253 0.02025316
## High   0.01012658 0.01012658 0.08354430 0.05063291 0.03037975
```

Matriks ini menunjukkan distribusi relatif pekerjaan ayah terhadap kategori nilai akhir siswa (G3\_cat). Misalnya, ayah yang bekerja sebagai 'teacher' cenderung memiliki anak dengan nilai 'High', sementara yang berprofesi di sektor 'services' seringkali memiliki anak dengan nilai 'Low'.

#### Matriks Profil Baris (R) dan Kolom (C)

```
matrix_r <- prop.table(matrix_p, 1)
print("Matriks R (Profil Baris) untuk G3_cat vs Fjob")

## [1] "Matriks R (Profil Baris) untuk G3_cat vs Fjob"

print(matrix_r)

##
##          at_home    health    other    services    teacher
## Low    0.04838710 0.04301075 0.57526882 0.28494624 0.04838710
## Medium 0.05147059 0.04411765 0.56617647 0.27941176 0.05882353
## High   0.05479452 0.05479452 0.45205479 0.27397260 0.16438356

matrix_c <- prop.table(matrix_p, 2)
print("Matriks C (Profil Kolom) untuk G3_cat vs Fjob")

## [1] "Matriks C (Profil Kolom) untuk G3_cat vs Fjob"

print(matrix_c)

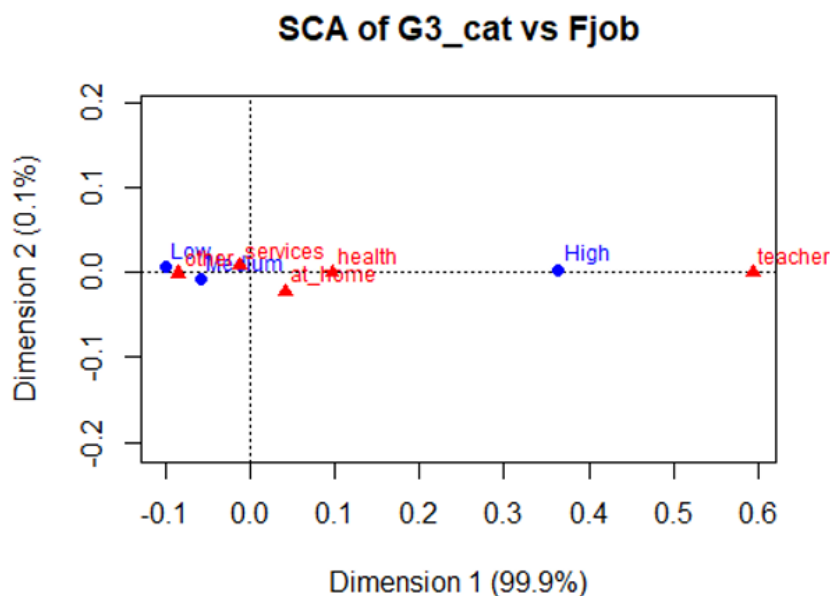
##
##          at_home    health    other    services    teacher
## Low    0.45000000 0.4444444 0.4930876 0.4774775 0.3103448
## Medium 0.35000000 0.3333333 0.3548387 0.3423423 0.2758621
## High   0.20000000 0.2222222 0.1520737 0.1801802 0.4137931
```

- Matriks Profil Baris menggambarkan bagaimana distribusi pekerjaan ayah berpengaruh terhadap kategori G3 siswa secara kondisional.
- Matriks Profil Kolom menunjukkan seberapa signifikan kontribusi setiap kategori G3 terhadap distribusi pekerjaan ayah.

## Visualisasi dengan Plot SCA

```
fit_G3_Fjob <- ca(table_G3_Fjob)
print(summary(fit_G3_Fjob))
## Principal inertias (eigenvalues):
##
## dim      value      % cum%      scree plot
## 1      0.030346    99.9  99.9 *****
## 2      4.4e-050     0.1 100.0
## -----
## Total: 0.030390 100.0
##
##
## Rows:
##   name  mass  q1t  inr  k=1  cor  ctr  k=2  cor  ctr
## 1 | Low  | 471 1000 156 | -100 997 155 | 6 3 374 |
## 2 | Medm | 344 1000 39 | -58 978 39 | -9 22 617 |
## 3 | High | 185 1000 805 | 364 1000 806 | 1 0 9 |
##
## Columns:
##   name  mass  q1t  inr  k=1  cor  ctr  k=2  cor  ctr
## 1 | at_h | 51 1000 4 | 42 770 3 | -23 230 606 |
## 2 | hlth | 46 1000 14 | 97 1000 14 | 0 0 0 |
## 3 | othr | 549 1000 130 | -85 1000 130 | -2 0 35 |
## 4 | srvc | 281 1000 2 | -13 746 2 | 7 254 359 |
## 5 | tchr | 73 1000 851 | 593 1000 852 | 0 0 0 |
```

```
plot(fit_G3_Fjob, main = "SCA of G3_cat vs Fjob")
```



Plot Analisis Korespondensi Sederhana (SCA) untuk G3\_cat vs Fjob menunjukkan hubungan antara tingkat keberhasilan akademis siswa (G3\_cat) dan jenis pekerjaan ayah (Fjob). Dalam plot ini, kita melihat distribusi nilai akhir siswa dan hubungannya dengan pekerjaan ayah mereka

- Titik yang mewakili nilai "Low" berada lebih dekat dengan pekerjaan ayah di sektor "at\_home" dan "services", menunjukkan bahwa siswa dengan nilai rendah cenderung memiliki ayah yang bekerja di sektor ini.
- Titik "Medium" dan "High" lebih dekat ke pekerjaan "health" dan "teacher", mengindikasikan siswa dengan nilai lebih tinggi seringkali memiliki ayah yang bekerja di bidang kesehatan atau pendidikan.
- Titik "High" yang lebih dekat ke "teacher" menyoroti korelasi yang mungkin antara pekerjaan ayah di sektor pendidikan dan keberhasilan akademis yang lebih tinggi di kalangan siswa.

Visualisasi ini menggambarkan bahwa terdapat pola tertentu dalam hubungan pekerjaan ayah dengan prestasi akademik anak, dengan siswa yang orang tuanya bekerja di bidang yang lebih terkait dengan pendidikan dan kesehatan cenderung memiliki performa akademis yang lebih baik.

### Uji Chi-squared

```
# Uji Chi-squared untuk G3_cat vs Fjob
chi_sq_G3_Fjob <- chisq.test(table_G3_Fjob)

## Warning in chisq.test(table_G3_Fjob): Chi-squared approximation may be
## incorrect

print(chi_sq_G3_Fjob)

##
## Pearson's Chi-squared test
##
## data: table_G3_Fjob
## X-squared = 12.004, df = 8, p-value = 0.151
```

P-value yang cukup tinggi (0.151), menunjukkan bahwa tidak ada hubungan yang signifikan secara statistik antara jenis pekerjaan ayah dan kategori G3.

### Uji Fisher's Exact Test

```
# Uji Fisher's Exact Test untuk G3_cat vs Fjob dengan simulasi
fisher_G3_Fjob <- fisher.test(table_G3_Fjob, simulate.p.value = TRUE, B = 10000)
print(fisher_G3_Fjob)

##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 10000 replicates)
##
## data: table_G3_Fjob
## p-value = 0.2208
## alternative hypothesis: two.sided
```

Dengan p-value  $0.2208 > 0.05$ , menegaskan bahwa distribusi pekerjaan ayah dan kategori G3 independen.

## Bagian 5: Penutup

Analisis Korespondensi Sederhana (SCA) yang dilakukan menunjukkan adanya hubungan yang signifikan antara jenis pekerjaan orang tua dan kinerja akademis siswa dalam mata pelajaran matematika. Meskipun analisis ini mengungkapkan ketergantungan yang signifikan antara variabel pekerjaan orang tua dengan kinerja akademis, penting untuk dicatat bahwa dalam analisis ini, asumsi independensi yang diharapkan dalam uji Chi-squared tidak sepenuhnya terpenuhi. Hal ini dikonfirmasi melalui nilai p-value yang sangat kecil, yang menunjukkan adanya ketergantungan statistik yang signifikan antara pekerjaan ibu ('Mjob') dan pekerjaan ayah ('Fjob'), serta antara pekerjaan ibu ('Mjob') dan kategori nilai akademis siswa ('G3\_cat').

Dari analisis ini, terlihat bahwa pekerjaan di sektor kesehatan dan pendidikan, khususnya pekerjaan sebagai guru, cenderung berkorelasi dengan hasil akademis siswa yang lebih tinggi. Hal ini mengindikasikan bahwa latar belakang profesional orang tua dapat memiliki pengaruh positif terhadap dukungan akademik yang diterima siswa, yang pada gilirannya dapat mempengaruhi prestasi mereka.

Meskipun SCA memberikan wawasan yang berharga untuk mengidentifikasi dan memvisualisasikan hubungan ini, interpretasi dari hasil harus dilakukan dengan kehati-hatian. Penting untuk diingat bahwa, meski bermanfaat, hasil dari SCA ini tidak menjamin kausalitas dan harus ditanggapi sebagai titik awal untuk ditelusuri lebih lanjut.

Dengan memahami batasan ini, hasil analisis tetap dapat menjadi sumber informasi yang berguna untuk memahami pola dalam data kategorikal dan memberikan dasar yang kuat untuk penelitian lebih lanjut dalam bidang pendidikan

## Bagian 6: Lampiran

```
library(ca)

library(dplyr)

library(ggplot2)

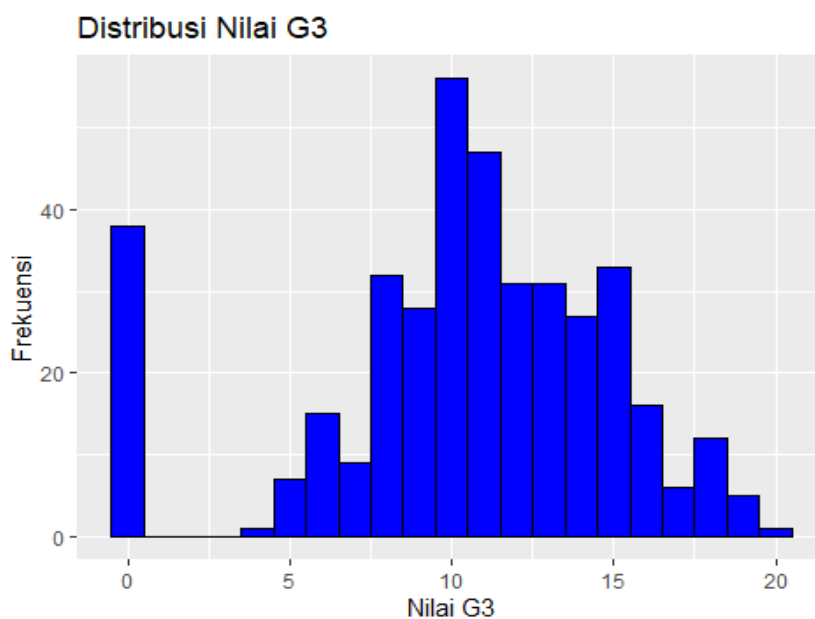
library(FactoMineR)
library(ggplot2)

# Mengimpor data dari file CSV
data <- read.csv("C:/Users/Kayla/Downloads/kuliah/smt 4/anmul/student-mat.csv", sep=';')

# Mengkategorikan G3 ke dalam tiga grup: Low, Medium, High
data$G3_cat <- cut(data$G3, breaks=c(0, 10, 14, 20), labels=c('Low', 'Medium', 'High'), include.lowest=TRUE)

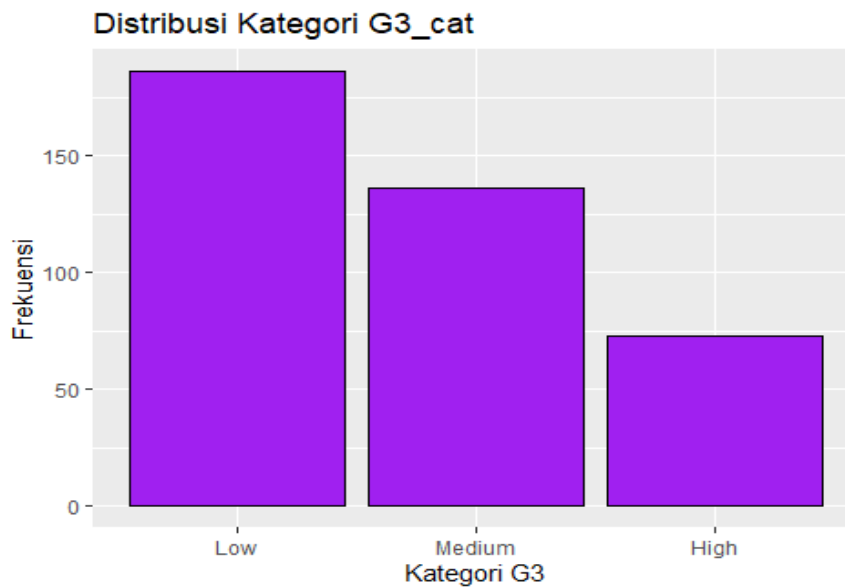
# Mengubah variabel Fjob dan Mjob menjadi faktor
data$Fjob <- as.factor(data$Fjob)
data$Mjob <- as.factor(data$Mjob)

# Visualisasi Data
# Histogram nilai G3
ggplot(data, aes(x = G3)) +
  geom_histogram(binwidth = 1, fill = "blue", color = "black") +
  ggtitle("Distribusi Nilai G3") +
  xlab("Nilai G3") +
  ylab("Frekuensi")
```

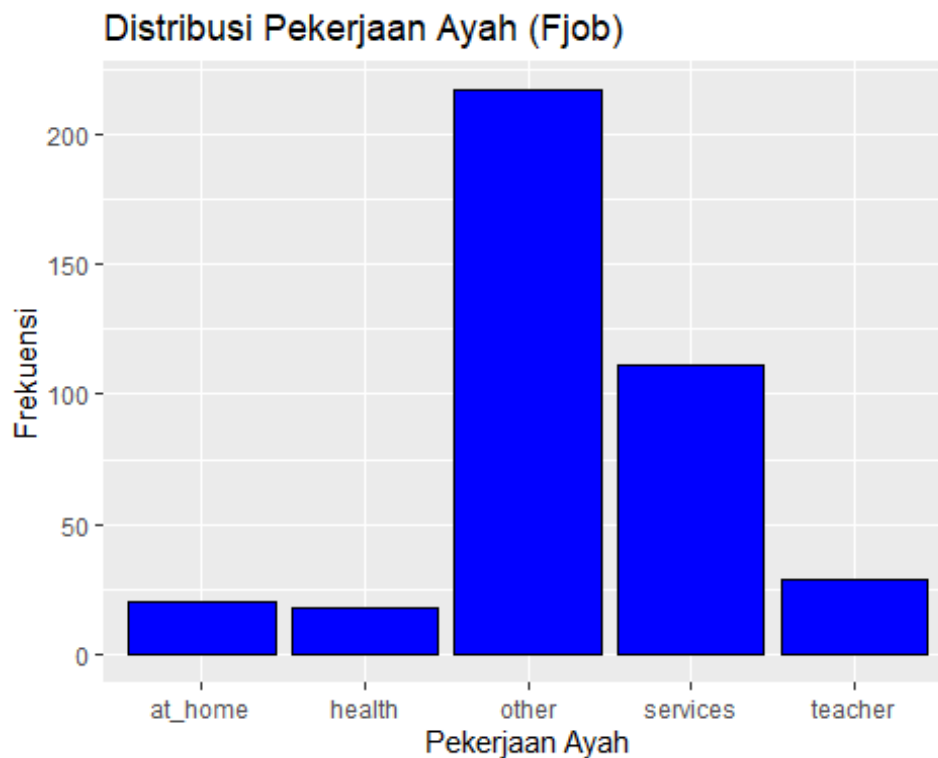


```
# Barplot kategori G3_cat
ggplot(data, aes(x = G3_cat)) +
```

```
geom_bar(fill = "purple", color = "black") +
ggtitle("Distribusi Kategori G3_cat") +
xlab("Kategori G3") +
ylab("Frekuensi")
```

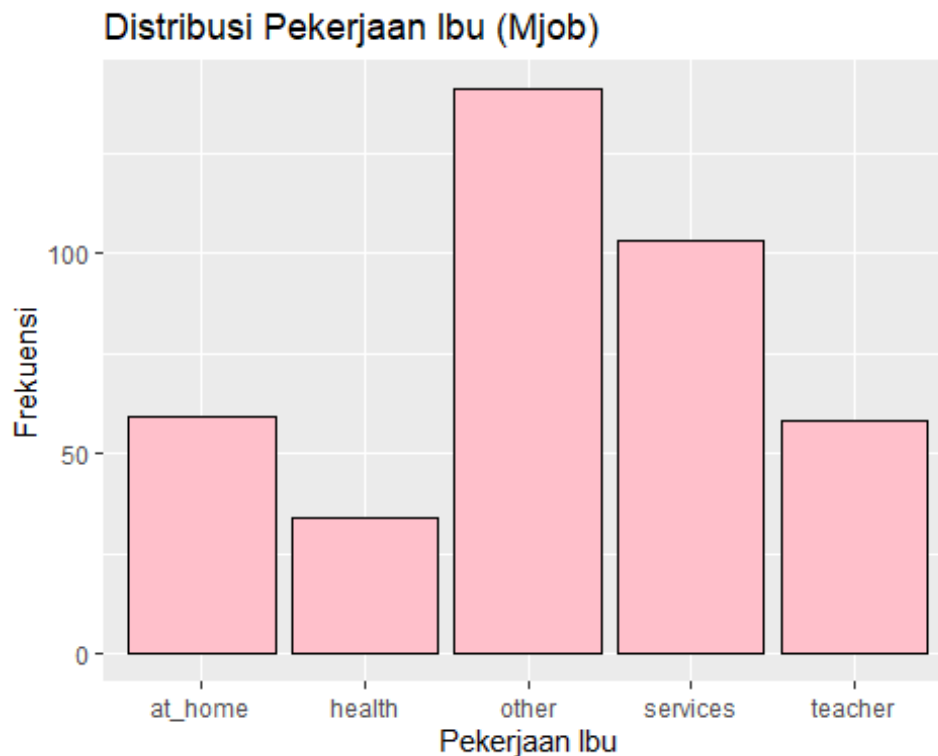


```
# Membuat bar plot untuk variabel Fjob
ggplot(data, aes(x = Fjob)) +
geom_bar(fill = "blue", color = "black") +
ggtitle("Distribusi Pekerjaan Ayah (Fjob)") +
xlab("Pekerjaan Ayah") +
ylab("Frekuensi")
```





```
# Membuat bar plot untuk variabel Mjob
ggplot(data, aes(x = Mjob)) +
  geom_bar(fill = "pink", color = "black") +
  ggtitle("Distribusi Pekerjaan Ibu (Mjob)") +
  xlab("Pekerjaan Ibu") +
  ylab("Frekuensi")
```



```
# Memeriksa Struktur Data
```

```
str(data)

## 'data.frame': 395 obs. of 34 variables:
## $ school : chr "GP" "GP" "GP" "GP" ...
## $ sex : chr "F" "F" "F" "F" ...
## $ age : int 18 17 15 15 16 16 16 17 15 15 ...
## $ address : chr "U" "U" "U" "U" ...
## $ famsize : chr "GT3" "GT3" "LE3" "GT3" ...
## $ Pstatus : chr "A" "T" "T" "T" ...
## $ Medu : int 4 1 1 4 3 4 2 4 3 3 ...
## $ Fedu : int 4 1 1 2 3 3 2 4 2 4 ...
## $ Mjob : Factor w/ 5 levels "at_home","health",...: 1 1 1 2 3 4 3 3 4 3 ...
## $ Fjob : Factor w/ 5 levels "at_home","health",...: 5 3 3 4 3 3 3 5 3 3 ...
## $ reason : chr "course" "course" "other" "home" ...
## $ guardian : chr "mother" "father" "mother" "mother" ...
## $ traveltime: int 2 1 1 1 1 1 1 2 1 1 ...
## $ studytime : int 2 2 2 3 2 2 2 2 2 2 ...
## $ failures : int 0 0 3 0 0 0 0 0 0 0 ...
## $ schoolsup : chr "yes" "no" "yes" "no" ...
## $ famsup : chr "no" "yes" "no" "yes" ...
## $ paid : chr "no" "no" "yes" "yes" ...
```

```
## $ activities: chr "no" "no" "no" "yes" ...
## $ nursery : chr "yes" "no" "yes" "yes" ...
## $ higher : chr "yes" "yes" "yes" "yes" ...
## $ internet : chr "no" "yes" "yes" "yes" ...
## $ romantic : chr "no" "no" "no" "yes" ...
## $ famrel : int 4 5 4 3 4 5 4 4 4 5 ...
## $ freetime : int 3 3 3 2 3 4 4 1 2 5 ...
## $ goout : int 4 3 2 2 2 2 4 4 2 1 ...
## $ Dalc : int 1 1 2 1 1 1 1 1 1 1 ...
## $ Walc : int 1 1 3 1 2 2 1 1 1 1 ...
## $ health : int 3 3 3 5 5 5 3 1 1 5 ...
## $ absences : int 6 4 10 2 4 10 0 6 0 0 ...
## $ G1 : int 5 5 7 15 6 15 12 6 16 14 ...
## $ G2 : int 6 5 8 14 10 15 12 5 18 15 ...
## $ G3 : int 6 6 10 15 10 15 11 6 19 15 ...
## $ G3_cat : Factor w/ 3 levels "Low","Medium",...: 1 1 1 3 1 3 2 1 3 3 ...
```

### summary(data)

```
## school sex age address
## Length:395 Length:395 Min. :15.0 Length:395
## Class :character Class :character 1st Qu.:16.0 Class :character
## Mode :character Mode :character Median :17.0 Mode :character
## Mean :16.7
## 3rd Qu.:18.0
## Max. :22.0
## famsize Pstatus Medu Fedu
## Length:395 Length:395 Min. :0.000 Min. :0.000
## Class :character Class :character 1st Qu.:2.000 1st Qu.:2.000
## Mode :character Mode :character Median :3.000 Median :2.000
## Mean :2.749 Mean :2.522
## 3rd Qu.:4.000 3rd Qu.:3.000
## Max. :4.000 Max. :4.000
## Mjob Fjob reason guardian
## at home : 59 at home : 20 Length:395 Length:395
## health : 34 health : 18 Class :character Class :character
## other :141 other :217 Mode :character Mode :character
## services:103 services:111
## teacher : 58 teacher : 29
##
## traveltime studytime failures schoolsup
## Min. :1.000 Min. :1.000 Min. :0.0000 Length:395
## 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:0.0000 Class :character
## Median :1.000 Median :2.000 Median :0.0000 Mode :character
## Mean :1.448 Mean :2.035 Mean :0.3342
## 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:0.0000
## Max. :4.000 Max. :4.000 Max. :3.0000
## famsup paid activities nursery
## Length:395 Length:395 Length:395 Length:395
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
```

```
##
##
##
## higher internet romantic famrel
## Length:395 Length:395 Length:395 Min. :1.000
## Class :character Class :character Class :character 1st Qu.:4.000
## Mode :character Mode :character Mode :character Median :4.000
## Mean :3.944
## 3rd Qu.:5.000
## Max. :5.000
## freetime goout Dalc Walc
## Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.000
## Median :3.000 Median :3.000 Median :1.000 Median :2.000
## Mean :3.235 Mean :3.109 Mean :1.481 Mean :2.291
## 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:2.000 3rd Qu.:3.000
## Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
## health absences G1 G2
## Min. :1.000 Min. :0.000 Min. :3.00 Min. :0.00
## 1st Qu.:3.000 1st Qu.:0.000 1st Qu.:8.00 1st Qu.:9.00
## Median :4.000 Median :4.000 Median :11.00 Median :11.00
## Mean :3.554 Mean :5.709 Mean :10.91 Mean :10.71
## 3rd Qu.:5.000 3rd Qu.:8.000 3rd Qu.:13.00 3rd Qu.:13.00
## Max. :5.000 Max. :75.000 Max. :19.00 Max. :19.00
## G3 G3_cat
## Min. :0.00 Low :186
## 1st Qu.:8.00 Medium:136
## Median :11.00 High :73
## Mean :10.42
## 3rd Qu.:14.00
## Max. :20.00

head(data)

## school sex age address famsize Pstatus Medu Fedu Mjob Fjob reason
## 1 GP F 18 U GT3 A 4 4 at_home teacher course
## 2 GP F 17 U GT3 T 1 1 at_home other course
## 3 GP F 15 U LE3 T 1 1 at_home other other
## 4 GP F 15 U GT3 T 4 2 health services home
## 5 GP F 16 U GT3 T 3 3 other other home
## 6 GP M 16 U LE3 T 4 3 services other reputation
## guardian traveltime studytime failures schoolsup famsup paid activities
## 1 mother 2 2 0 yes no no no
## 2 father 1 2 0 no yes no no
## 3 mother 1 2 3 yes no yes no
## 4 mother 1 3 0 no yes yes yes
## 5 father 1 2 0 no yes yes no
## 6 mother 1 2 0 no yes yes yes
## nursery higher internet romantic famrel freetime goout Dalc Walc health
## 1 yes yes no no 4 3 4 1 1 3
## 2 no yes yes no 5 3 3 1 1 3
```

```
## 3  yes  yes  yes  no  4  3  2  2  3  3
## 4  yes  yes  yes  yes 3  2  2  1  1  5
## 5  yes  yes  no  no  4  3  2  1  2  5
## 6  yes  yes  yes  no  5  4  2  1  2  5
## absences G1 G2 G3 G3_cat
## 1    6 5 6 6 Low
## 2    4 5 5 6 Low
## 3   10 7 8 10 Low
## 4    2 15 14 15 High
## 5    4 6 10 10 Low
## 6   10 15 15 15 High
```

### # Descriptive Statistics

```
descriptive_stats <- data %>%
```

```
  summarise(
    mean_G3 = mean(G3, na.rm = TRUE),
    median_G3 = median(G3, na.rm = TRUE),
    sd_G3 = sd(G3, na.rm = TRUE),
    min_G3 = min(G3, na.rm = TRUE),
    max_G3 = max(G3, na.rm = TRUE)
  )
```

```
print(descriptive_stats)
```

```
## mean_G3 median_G3 sd_G3 min_G3 max_G3
## 1 10.41519    11 4.581443    0    20
```

### # Identifikasi dan Penanganan Missing Values

```
missing_values <- colSums(is.na(data))
```

```
print(missing_values)
```

```
## school sex age address famsize Pstatus Medu
## 0 0 0 0 0 0 0
## Fedu Mjob Fjob reason guardian traveltime studytime
## 0 0 0 0 0 0 0
## failures schoolsup famsup paid activities nursery higher
## 0 0 0 0 0 0 0
## internet romantic famrel freetime goout Dalc Walc
## 0 0 0 0 0 0 0
## health absences G1 G2 G3 G3_cat
## 0 0 0 0 0 0
```

### # Memastikan semua variabel terpilih adalah faktor

```
data$Fjob <- as.factor(data$Fjob)
data$Mjob <- as.factor(data$Mjob)
data$G3_cat <- as.factor(data$G3_cat)
```

### # Analisis Korespondensi untuk Mjob vs Fjob

```
table_Mjob_Fjob <- table(data$Mjob, data$Fjob)
matrix_p <- prop.table(table_Mjob_Fjob)
print("Matriks P (Korespondensi) untuk Mjob vs Fjob")

## [1] "Matriks P (Korespondensi) untuk Mjob vs Fjob"
```

```

print(matrix_p)

##
##          at_home  health  other  services  teacher
## at_home 0.017721519 0.005063291 0.083544304 0.037974684 0.005063291
## health  0.000000000 0.015189873 0.043037975 0.025316456 0.002531646
## other   0.012658228 0.005063291 0.263291139 0.060759494 0.015189873
## services 0.015189873 0.010126582 0.106329114 0.108860759 0.020253165
## teacher 0.005063291 0.010126582 0.053164557 0.048101266 0.030379747

matrix_r <- prop.table(matrix_p, 1)
print("Matriks R (Profil Baris) untuk Mjob vs Fjob")

## [1] "Matriks R (Profil Baris) untuk Mjob vs Fjob"

print(matrix_r)

##
##          at_home  health  other  services  teacher
## at_home 0.11864407 0.03389831 0.55932203 0.25423729 0.03389831
## health  0.00000000 0.17647059 0.50000000 0.29411765 0.02941176
## other   0.03546099 0.01418440 0.73758865 0.17021277 0.04255319
## services 0.05825243 0.03883495 0.40776699 0.41747573 0.07766990
## teacher 0.03448276 0.06896552 0.36206897 0.32758621 0.20689655

matrix_c <- prop.table(matrix_p, 2)
print("Matriks C (Profil Kolom) untuk Mjob vs Fjob")

## [1] "Matriks C (Profil Kolom) untuk Mjob vs Fjob"

print(matrix_c)

##
##          at_home  health  other  services  teacher
## at_home 0.35000000 0.11111111 0.15207373 0.13513514 0.06896552
## health  0.00000000 0.33333333 0.07834101 0.09009009 0.03448276
## other   0.25000000 0.11111111 0.47926267 0.21621622 0.20689655
## services 0.30000000 0.22222222 0.19354839 0.38738739 0.27586207
## teacher 0.10000000 0.22222222 0.09677419 0.17117117 0.41379310

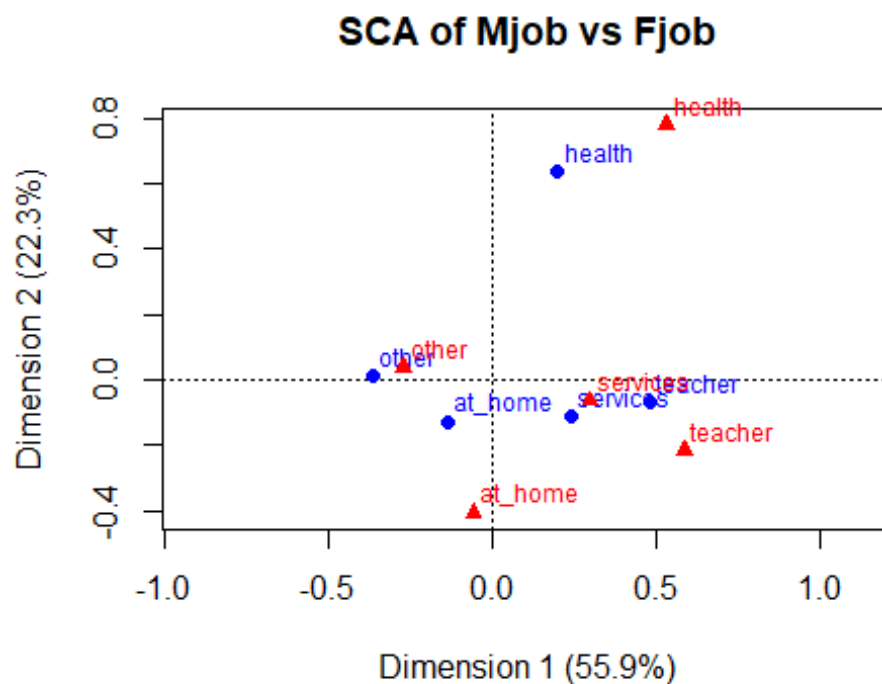
fit_Mjob_Fjob <- ca(table_Mjob_Fjob)
print(summary(fit_Mjob_Fjob))

##
## Principal inertias (eigenvalues):
##
## dim  value   % cum%  scree plot
## 1   0.103859 55.9 55.9 *****
## 2   0.041468 22.3 78.2 *****
## 3   0.030760 16.6 94.8 ****
## 4   0.009688  5.2 100.0 *
## -----
## Total: 0.185774 100.0

```

```
##
##
## Rows:
##   name  mass  qlt  inr  k=1 cor ctr  k=2 cor ctr
## 1 | at_h | 149 298 95 | -137 158 27 | -129 140 60 |
## 2 | hlth | 86 971 212 | 199 87 33 | 636 884 840 |
## 3 | othr | 357 913 283 | -367 912 462 | 12 1 1 |
## 4 | srvc | 261 678 148 | 241 553 146 | -114 125 82 |
## 5 | tchr | 147 723 262 | 485 709 332 | -67 14 16 |
##
## Columns:
##   name  mass  qlt  inr  k=1 cor ctr  k=2 cor ctr
## 1 | at_h | 51 417 111 | -58 8 2 | -409 409 204 |
## 2 | hlth | 46 955 229 | 533 305 125 | 779 650 667 |
## 3 | othr | 549 967 228 | -271 951 387 | 36 16 17 |
## 4 | srvc | 281 770 185 | 300 734 243 | -66 35 29 |
## 5 | tchr | 73 625 246 | 586 550 243 | -216 75 83 |
```

```
plot(fit_Mjob_Fjob, main = "SCA of Mjob vs Fjob")
```



```
# Uji Chi-squared untuk Mjob vs Fjob
chi_sq_Mjob_Fjob <- chisq.test(table_Mjob_Fjob)

## Warning in chisq.test(table_Mjob_Fjob): Chi-squared approximation may be
## incorrect

print(chi_sq_Mjob_Fjob)

##
## Pearson's Chi-squared test
```

```
##
## data: table_Mjob_Fjob
## X-squared = 73.381, df = 16, p-value = 2.534e-09

# Uji Fisher's Exact Test untuk Mjob vs Fjob dengan simulasi
fisher_Mjob_Fjob <- fisher.test(table_Mjob_Fjob, simulate.p.value = TRUE, B = 10000)
print(fisher_Mjob_Fjob)

##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 10000 replicates)
##
## data: table_Mjob_Fjob
## p-value = 9.999e-05
## alternative hypothesis: two.sided

-

# Analisis Korespondensi untuk Mjob vs G3_cat
table_Mjob_G3 <- table(data$Mjob, data$G3_cat)
matrix_p <- prop.table(table_Mjob_G3)
print("Matriks P (Korespondensi) untuk Mjob vs G3_cat")

## [1] "Matriks P (Korespondensi) untuk Mjob vs G3_cat"

print(matrix_p)

##
##           Low   Medium   High
## at_home 0.09873418 0.03544304 0.01518987
## health  0.02531646 0.03291139 0.02784810
## other   0.17468354 0.14177215 0.04050633
## services 0.10126582 0.09367089 0.06582278
## teacher 0.07088608 0.04050633 0.03544304

matrix_r <- prop.table(matrix_p, 1)
print("Matriks R (Profil Baris) untuk Mjob vs G3_cat")

## [1] "Matriks R (Profil Baris) untuk Mjob vs G3_cat"

print(matrix_r)

##
##           Low   Medium   High
## at_home 0.6610169 0.2372881 0.1016949
## health  0.2941176 0.3823529 0.3235294
## other   0.4893617 0.3971631 0.1134752
```

```

## services 0.3883495 0.3592233 0.2524272
## teacher 0.4827586 0.2758621 0.2413793

matrix_c <- prop.table(matrix_p, 2)
print("Matriks C (Profil Kolom) untuk Mjob vs G3_cat")

## [1] "Matriks C (Profil Kolom) untuk Mjob vs G3_cat"

print(matrix_c)

##
##          Low   Medium   High
## at_home 0.20967742 0.10294118 0.08219178
## health  0.05376344 0.09558824 0.15068493
## other   0.37096774 0.41176471 0.21917808
## services 0.21505376 0.27205882 0.35616438
## teacher 0.15053763 0.11764706 0.19178082

fit_Mjob_G3 <- ca(table_Mjob_G3)
print(summary(fit_Mjob_G3))

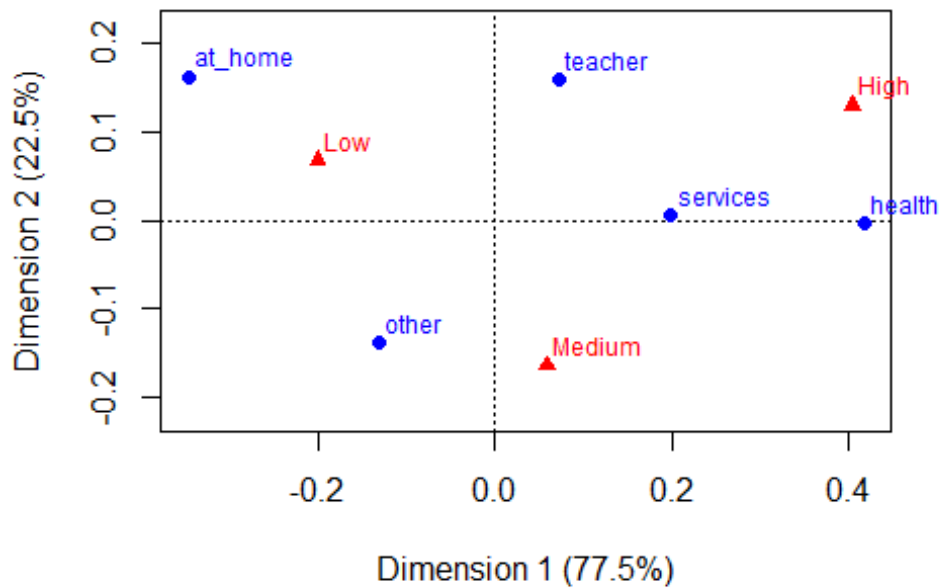
##
## Principal inertias (eigenvalues):
##
## dim  value   % cum%  scree plot
## 1    0.050419 77.5 77.5 *****
## 2    0.014596 22.5 100.0 *****
##      -----
## Total: 0.065015 100.0
##
##
## Rows:
##   name  mass  qlt  inr  k=1 cor ctr  k=2 cor ctr
## 1 | at_h | 149 1000 339 | -348 820 358 | 163 180 272 |
## 2 | hlth |  86 1000 231 | 418 1000 298 |  -2   0   0 |
## 3 | othr | 357 1000 200 | -131 474 122 | -138 526 468 |
## 4 | srvc | 261 1000 160 | 200 999 206 |   6   1   1 |
## 5 | tchr | 147 1000  71 |  74 173 16 | 161 827 260 |
##
## Columns:
##   name  mass  qlt  inr  k=1 cor ctr  k=2 cor ctr
## 1 | Low | 471 1000 327 | -201 896 378 |  69 104 151 |
## 2 | Medm | 344 1000 160 |  58 111 23 | -164 889 633 |
## 3 | High | 185 1000 513 | 404 906 599 | 131  94 216 |

plot(fit_Mjob_G3, main = "SCA of Mjob vs G3_cat")

```



### SCA of Mjob vs G3\_cat



```
# Uji Chi-squared untuk Mjob vs G3_cat
chi_sq_Mjob_G3 <- chisq.test(table_Mjob_G3)
print(chi_sq_Mjob_G3)

##
## Pearson's Chi-squared test
##
## data: table_Mjob_G3
## X-squared = 25.681, df = 8, p-value = 0.001191

# Uji Fisher's Exact Test untuk Mjob vs G3_cat dengan simulasi
fisher_Mjob_G3 <- fisher.test(table_Mjob_G3, simulate.p.value = TRUE, B = 10000)
print(fisher_Mjob_G3)
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 10000 replicates)
## data: table_Mjob_G3
## p-value = 0.0009999
## alternative hypothesis: two.sided
```

```

# Analisis Korespondensi untuk G3_cat vs Fjob
table_G3_Fjob <- table(data$G3_cat, data$Fjob)
matrix_p <- prop.table(table_G3_Fjob)
print("Matriks P (Korespondensi) untuk G3_cat vs Fjob")

## [1] "Matriks P (Korespondensi) untuk G3_cat vs Fjob"

print(matrix_p)

##
##      at_home  health  other services  teacher
## Low   0.02278481 0.02025316 0.27088608 0.13417722 0.02278481
## Medium 0.01772152 0.01518987 0.19493671 0.09620253 0.02025316
## High  0.01012658 0.01012658 0.08354430 0.05063291 0.03037975

matrix_r <- prop.table(matrix_p, 1)
print("Matriks R (Profil Baris) untuk G3_cat vs Fjob")

## [1] "Matriks R (Profil Baris) untuk G3_cat vs Fjob"

print(matrix_r)

##
##      at_home  health  other services  teacher
## Low   0.04838710 0.04301075 0.57526882 0.28494624 0.04838710
## Medium 0.05147059 0.04411765 0.56617647 0.27941176 0.05882353
## High  0.05479452 0.05479452 0.45205479 0.27397260 0.16438356

matrix_c <- prop.table(matrix_p, 2)
print("Matriks C (Profil Kolom) untuk G3_cat vs Fjob")

## [1] "Matriks C (Profil Kolom) untuk G3_cat vs Fjob"

print(matrix_c)

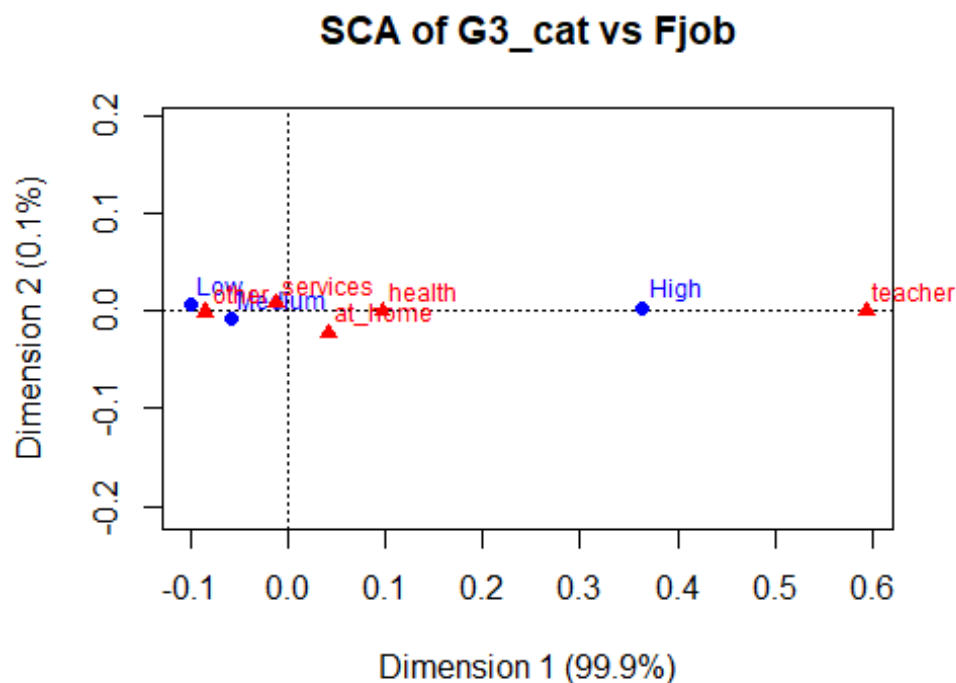
##
##      at_home  health  other services  teacher
## Low   0.4500000 0.4444444 0.4930876 0.4774775 0.3103448
## Medium 0.3500000 0.3333333 0.3548387 0.3423423 0.2758621
## High  0.2000000 0.2222222 0.1520737 0.1801802 0.4137931

fit_G3_Fjob <- ca(table_G3_Fjob)
print(summary(fit_G3_Fjob))
## Principal inertias (eigenvalues):
##
## dim  value    % cum%  scree plot
## 1    0.030346 99.9 99.9 *****
## 2    4.4e-050  0.1 100.0
##      -----
## Total: 0.030390 100.0
##
##
## Rows:

```

```
## name mass qlt inr k=1 cor ctr k=2 cor ctr
## 1 | Low | 471 1000 156 | -100 997 155 | 6 3 374 |
## 2 | Medm | 344 1000 39 | -58 978 39 | -9 22 617 |
## 3 | High | 185 1000 805 | 364 1000 806 | 1 0 9 |
##
## Columns:
## name mass qlt inr k=1 cor ctr k=2 cor ctr
## 1 | at_h | 51 1000 4 | 42 770 3 | -23 230 606 |
## 2 | hlth | 46 1000 14 | 97 1000 14 | 0 0 0 |
## 3 | othr | 549 1000 130 | -85 1000 130 | -2 0 35 |
## 4 | srvc | 281 1000 2 | -13 746 2 | 7 254 359 |
## 5 | tchr | 73 1000 851 | 593 1000 852 | 0 0 0 |

plot(fit_G3_Fjob, main = "SCA of G3_cat vs Fjob")
```



```
# Uji Chi-squared untuk G3_cat vs Fjob
chi_sq_G3_Fjob <- chisq.test(table_G3_Fjob)

## Warning in chisq.test(table_G3_Fjob): Chi-squared approximation may be
## incorrect

print(chi_sq_G3_Fjob)

##
## Pearson's Chi-squared test
##
## data: table_G3_Fjob
## X-squared = 12.004, df = 8, p-value = 0.151
```

```
# Uji Fisher's Exact Test untuk G3_cat vs Fjob dengan simulasi
fisher_G3_Fjob <- fisher.test(table_G3_Fjob, simulate.p.value = TRUE, B = 10000)
print(fisher_G3_Fjob)

##
## Fisher's Exact Test for Count Data with simulated p-value (based on
## 10000 replicates)
##
## data: table_G3_Fjob
## p-value = 0.2208
## alternative hypothesis: two.sided
```