



Analisis Survival

Analisis Survival *Turnover* Karyawan

06211940000007 Intan Citra Phonskaningtyas

06211940000021 Nadhirah Adzani Khairunnisa

06211940000036 Sekar Puspita Sari



Table of Contents



**Latar
Belakang**

Metodologi

**Pembahasan
dan
Kesimpulan**



Latar Belakang

Turnover karyawan adalah peristiwa saat karyawan yang diterima berpindah perusahaan.

- Dalam melakukan seleksi karyawan, perusahaan menyiapkan budget yang digunakan untuk seleksi, pelatihan dan persiapan karyawan dalam bekerja.
- Dengan turnover karyawan yang tinggi, maka perusahaan akan kehilangan budget yang telah dikeluarkan dan harus mengeluarkan budget lainnya untuk rekrutment.
- Oleh karena itu, diperlukan analisis survival waktu turnover karyawan agar perusahaan dapat mendeteksi masalah berdasarkan faktor – faktor yang ditentukan sebelumnya.



Metodologi



Sumber Data

Data digunakan adalah data sekunder bernama 'Turnover Survival Analysis' yang diperoleh dari website Kaggle yang diakses pada 17 Desember 2022 pada URL sebagai berikut.

www.kaggle.com/datasets/aman2114/turnover-survival-analysis/

Variabel Penelitian

Nama Variabel	Keterangan	Skala
Stage	Waktu yang diperlukan hingga terjadi turnover	Rasio
Event	Terjadi turnover atau tidak	Nominal (1/0)
Industry	Industri yang diobservasi	Nominal (Bank, Retail, IT, Manufaktur)
Age	Usia karyawan	Rasio
Gender	Gender Karyawan	Gender Karyawan (F/M)



Metodologi

Stage	Event	Age	Industry	Gender
Waktu 1	Event 1	Usia 1	Industri 1	Gender 1
Waktu 2	Event 2	Usia 2	Industri 2	Gender 2
...
Waktu 670	Event 670	Usia 670	Industri 670	Gender 670

Struktur Data

Metodologi

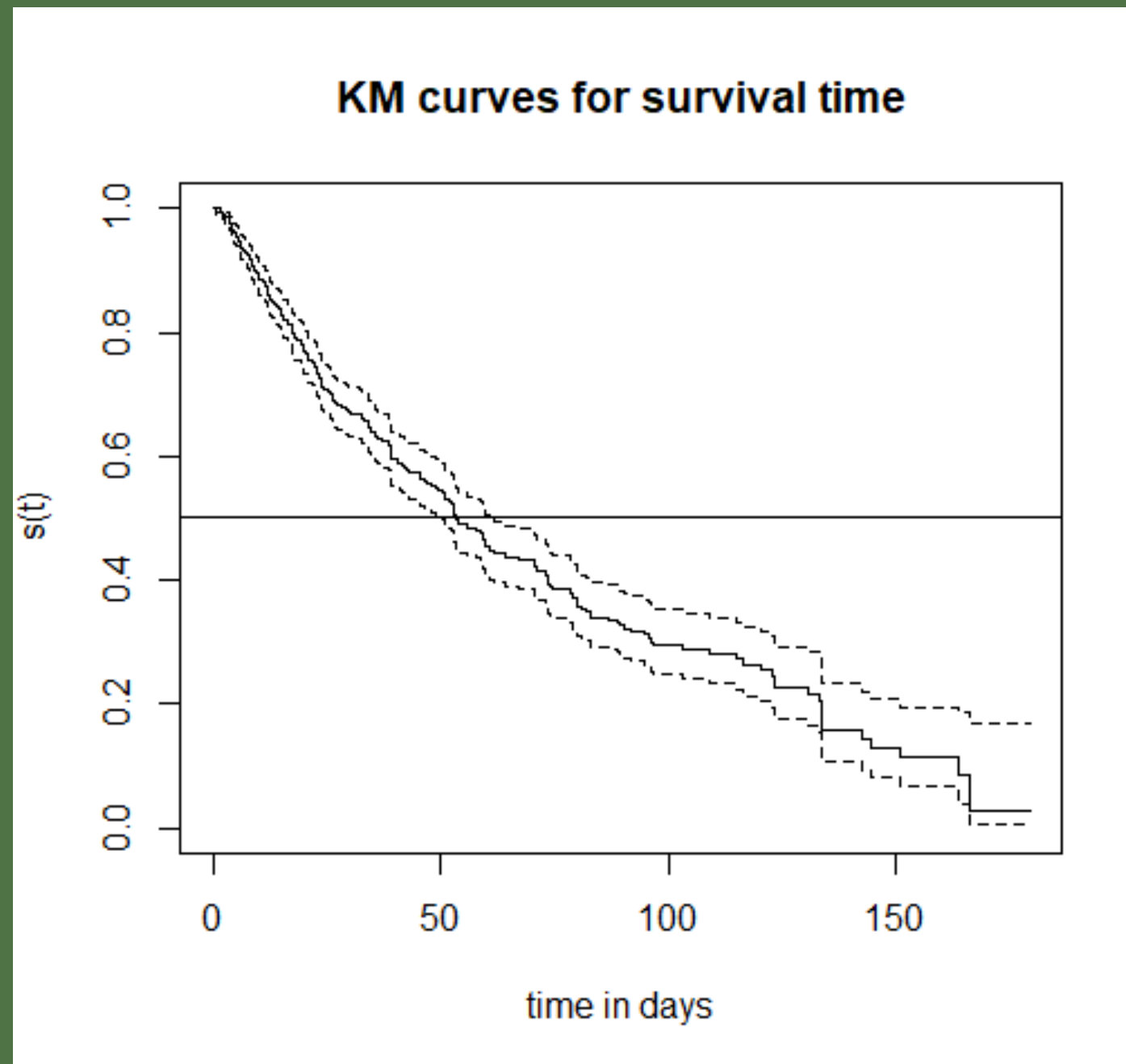


Langkah Analisis

1. Mengumpulkan data
2. Mengidentifikasi variabel yang digunakan
3. Membuat kurva survival Kaplan Meier untuk keseluruhan data
4. Membuat kurva survival Kaplan Meier untuk variabel industry dan gender
5. Menguji kurva survival dengan uji log rank
6. Membuat model Cox Proportional Hazard
7. Cek Asumsi PH pada model
8. Seleksi model terbaik (dengan interaksi atau tanpa interaksi)
9. Uji signifikansi parameter
10. Membuat Kurva Adjusted Survival
11. Menarik kesimpulan dan saran



Kurva Kaplan Meier



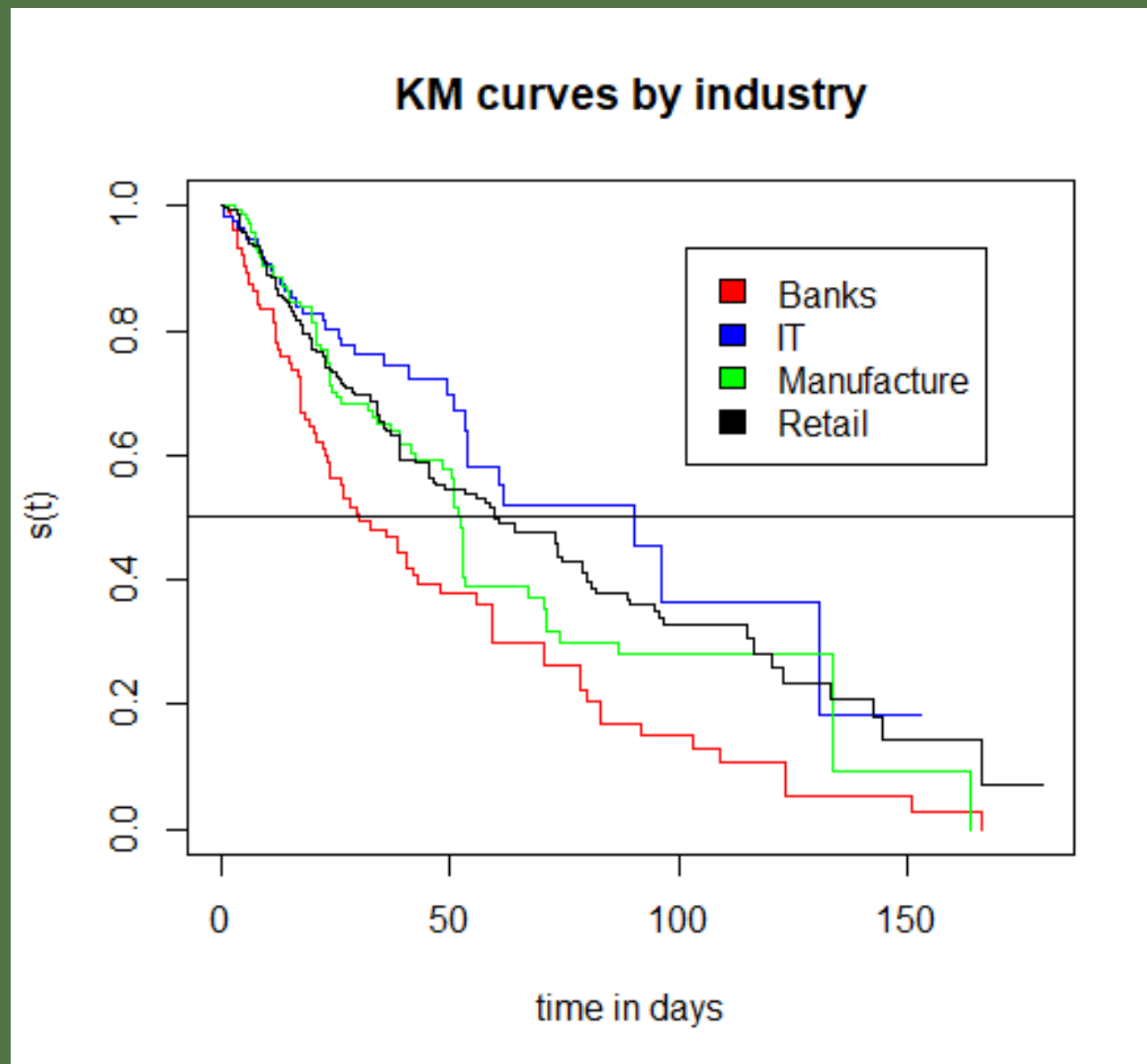
```
> surv
```

```
Call: survfit(formula = Surv(turnover$stag,  
turnover$event == 1) ~ 1)
```

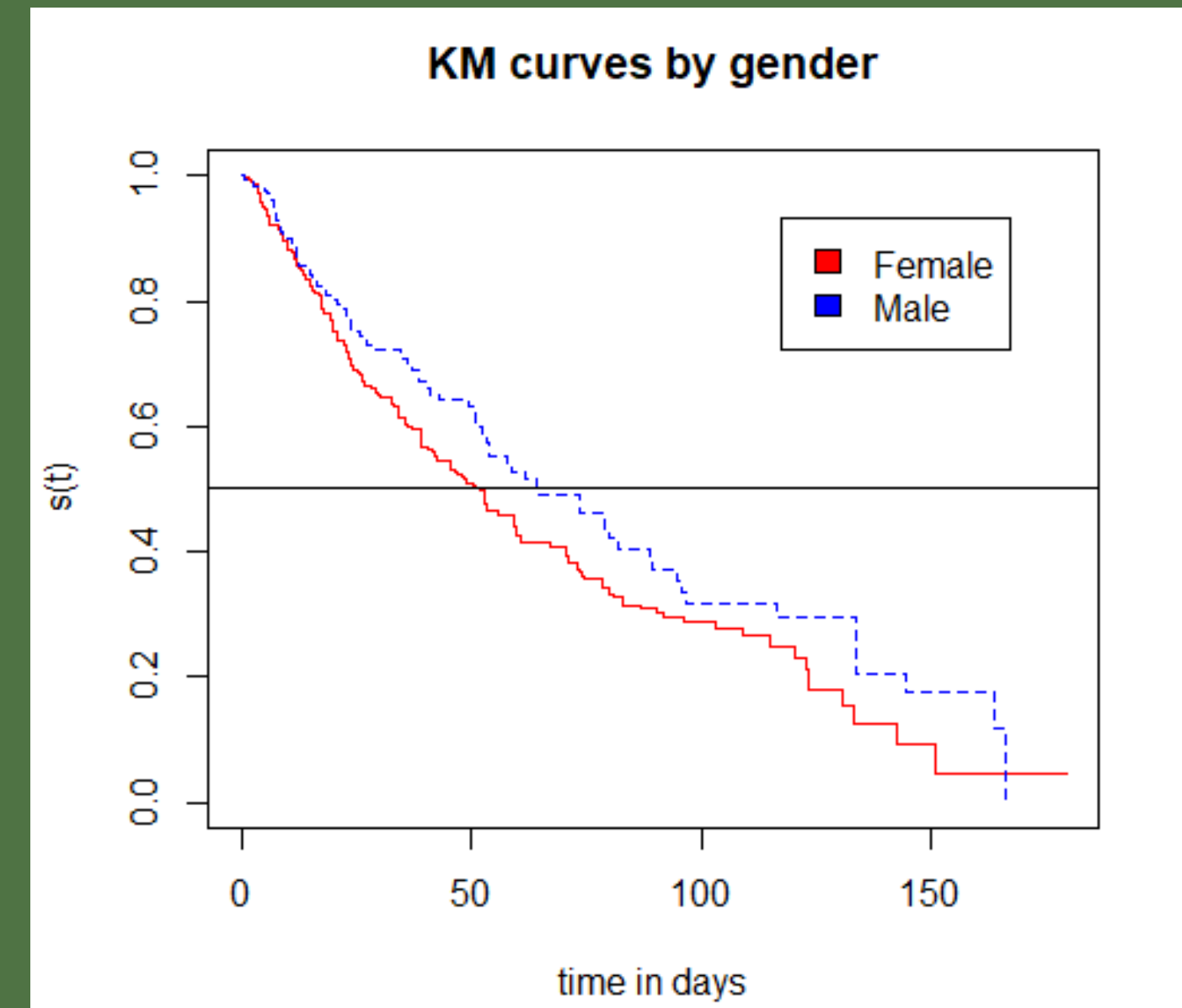
n	events	median	0.95LCL	0.95UCI
670	315	53.5	50.6	60.9

Karyawan mampu bertahan selama sekitar 53 hari pada saat survival time = 0.5 (median).

Kurva Kaplan Meier



Dari ke-4 kategori industri yang ada, Banks memiliki tingkat survival time terendah diikuti oleh Manufacture, Retail, lalu IT.



Gender male memiliki tingkat survival time lebih tinggi dibandingkan female.

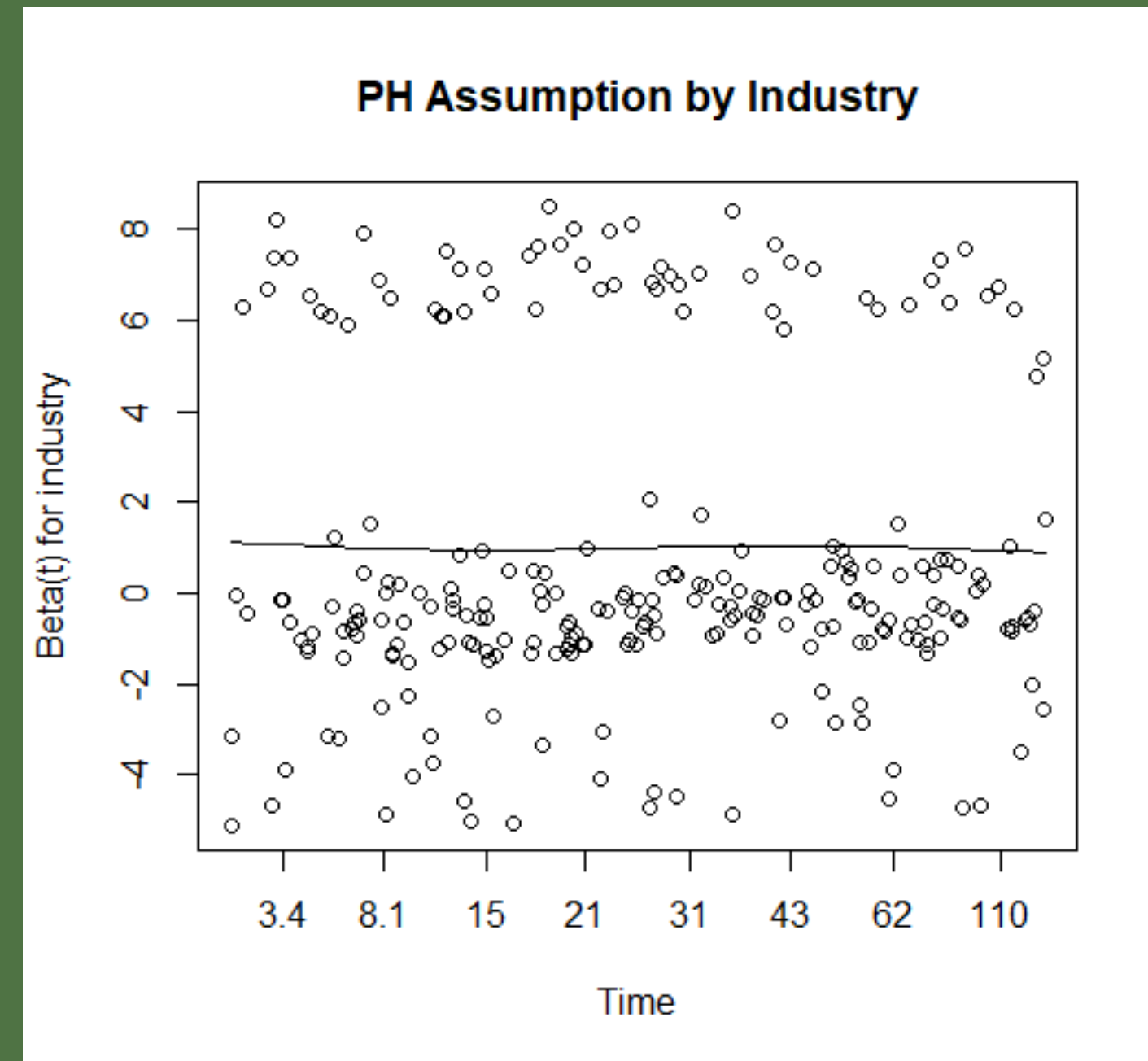
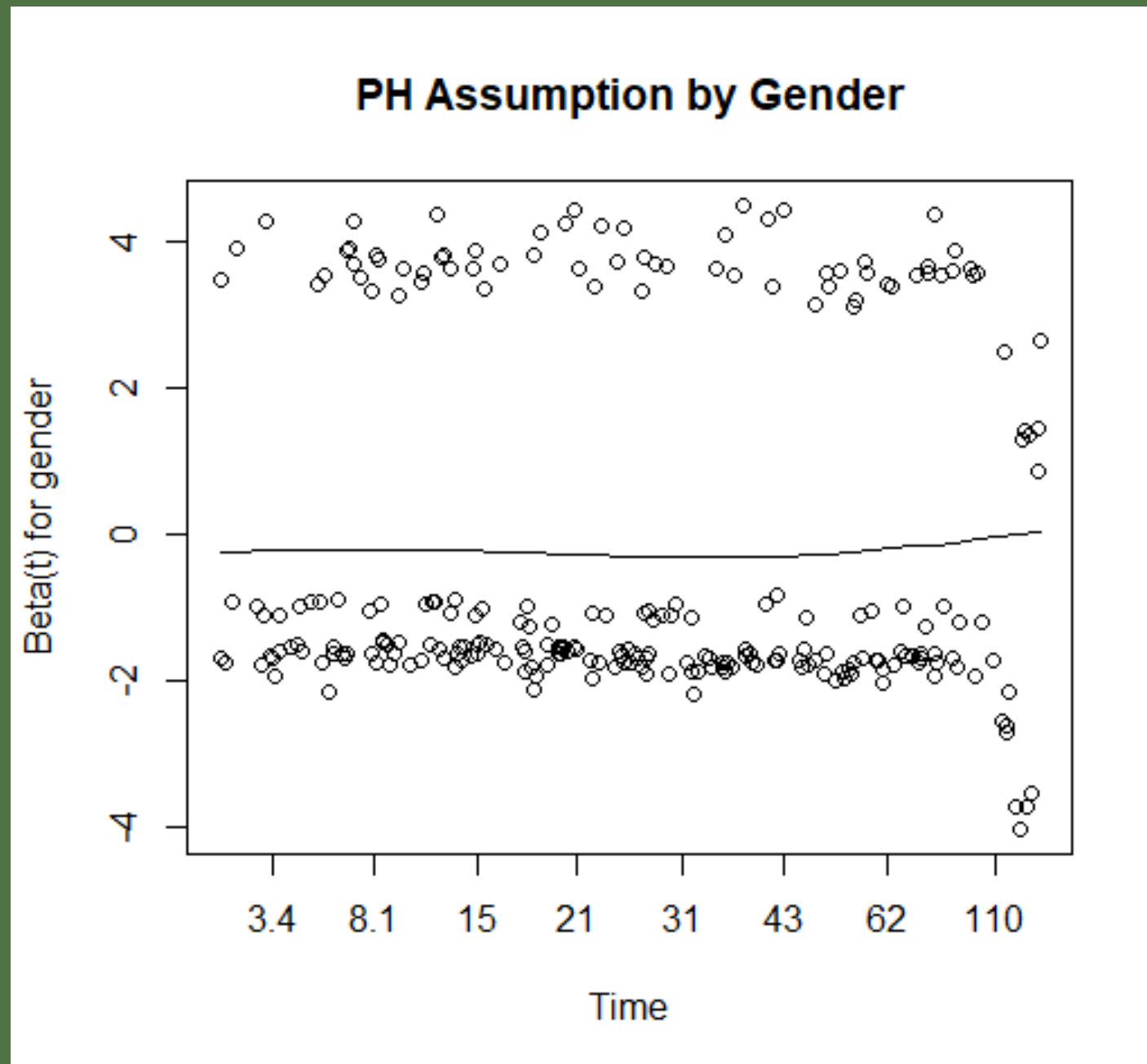
Asumsi Proportional Hazard

$H_0 : \rho = 0$ (Asumsi PH terpenuhi)

$H_1 : \rho \neq 0$ (Asumsi PH tidak terpenuhi)

	Chisq	df	p
industry	2.493	3	0.48
age	0.928	1	0.34
gender	0.081	1	0.78

Dengan menggunakan taraf signifikan sebesar 10%, dapat disimpulkan bahwa variabel industry, age, dan gender telah memenuhi asumsi PH karena nilai p-value > alpha.



Cek Asumsi PH

Interpretasi: Karena garis pada kedua plot horizontal, maka dapat dikatakan bahwa variabel independen terhadap waktu survival, sehingga kedua variabelnya memenuhi asumsi PH.

Model Cox PH

	coef	exp(coef)	se(coef)	z	p
industryIT	-0.911746	0.401822	0.212004	-4.301	1.70e-05
industrymanufacture	-0.614876	0.540708	0.170491	-3.606	0.00031
industryRetail	-0.622515	0.536593	0.145227	-4.286	1.82e-05
age	0.041802	1.042688	0.008399	4.977	6.45e-07
genderm	-0.222169	0.800780	0.131786	-1.686	0.09183

Likelihood ratio test = 46.83 on 5 df, p = 6.158e-09

n = 670, number of events = 315

Dengan taraf signifikansi 10%, dapat disimpulkan bahwa semua variabel telah signifikan karena nilai p-value < alpha. Sehingga model persamaannya adalah sebagai berikut.

$$h(t, X) = h_0(t) \exp(-0.912 \text{IndustryIT} - 0.615 \text{IndustryManufacture} - 0.622 \text{IndustryRetail} + 0.042 \text{Age} - 0.222 \text{Gender})$$

Model Cox PH dengan Interaksi



	coef	exp(coef)	se(coef)	z	p
industryIT	-0.89817	0.40731	0.26987	-3.328	0.000874
industrymanufacture	-0.62918	0.53303	0.19756	-3.185	0.001449
industryRetail	-0.57829	0.56086	0.16484	-3.508	0.000451
age	0.03911	1.03989	0.01012	3.865	0.000111
genderm	-0.49151	0.61170	0.59910	-0.820	0.411986
industryIT*genderm	-0.10361	0.90158	0.45586	-0.227	0.820207
industrymanufacture*genderm	0.08411	1.08775	0.39957	0.211	0.833274
industryRetail*genderm	-0.18193	0.83366	0.34759	-0.523	0.600690
age*genderm	0.01102	1.01108	0.01781	0.619	0.536097

Likelihood ratio test=47.91 on 9 df, p=2.651e-07

n= 670, number of events= 315

Dengan menggunakan taraf signifikansi 10%, maka variabel yang signifikan adalah IndustriIT, IndustryManufacture, IndustryRetail, dan Age. Sehingga model persamaannya adalah:

$$h(t;X)=h_0(t)\exp(-0.898\text{IndustryIT}-0.629\text{IndustryManufacture}-0.578\text{IndustryRetail}+0.039\text{Age})$$

Model Terbaik



Uji Log Rank

H_0 : Model tanpa interaksi adalah model terbaik

H_1 : Model dengan interaksi adalah model terbaik

```
> lr=(-2)*(mod1$loglik[2]-mod2$loglik[2])
> lr
[1] 1.08526
> Pvalue = 1 - pchisq(lr, 2)
> Pvalue
[1] 0.5812176
```

Dengan menggunakan taraf signifikansi 10%, dapat disimpulkan bahwa model tanpa interaksi adalah model terbaik karena diperoleh keputusan gagal tolak H_0 dimana $LR < x_{2;0.1}^2$ atau p-value > alpha.

Uji Serentak

	exp(coef)	exp(-coef)	lower .95	upper .95
industryIT	0.4018	2.4887	0.2652	0.6088
industrymanufacture	0.5407	1.8494	0.3871	0.7552
industryRetail	0.5366	1.8636	0.4037	0.7133
age	1.0427	0.9591	1.0257	1.0600
genderm	0.8008	1.2488	0.6185	1.0368

Concordance = 0.598 (se = 0.019)

Likelihood ratio test= 46.83 on 5 df, p=6e-09

Wald test = 48.39 on 5 df, p=3e-09

Score (logrank) test = 49.25 on 5 df, p=2e-09

Dengan menggunakan taraf signifikansi 10%, dapat disimpulkan bahwa model telah signifikan secara serentak karena nilai p-value < alpha.

Uji Parsial

```

              coef      exp(coef)      se(coef)      z      Pr(>|z|)
industryIT   -0.911746    0.401822    0.212004   -4.301   1.70e-05***
industrymanufacture -0.614876    0.540708    0.170491   -3.606   0.00031***
industryRetail -0.622515    0.536593    0.145227   -4.286   1.82e-05***
age           0.041802    1.042688    0.008399    4.977   6.45e-07***
genderm      -0.222169    0.800780    0.131786   -1.686   0.09183
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Dengan menggunakan taraf signifikansi 10%, dapat disimpulkan bahwa model telah signifikan secara parsial karena nilai $p\text{-value} < \alpha$.

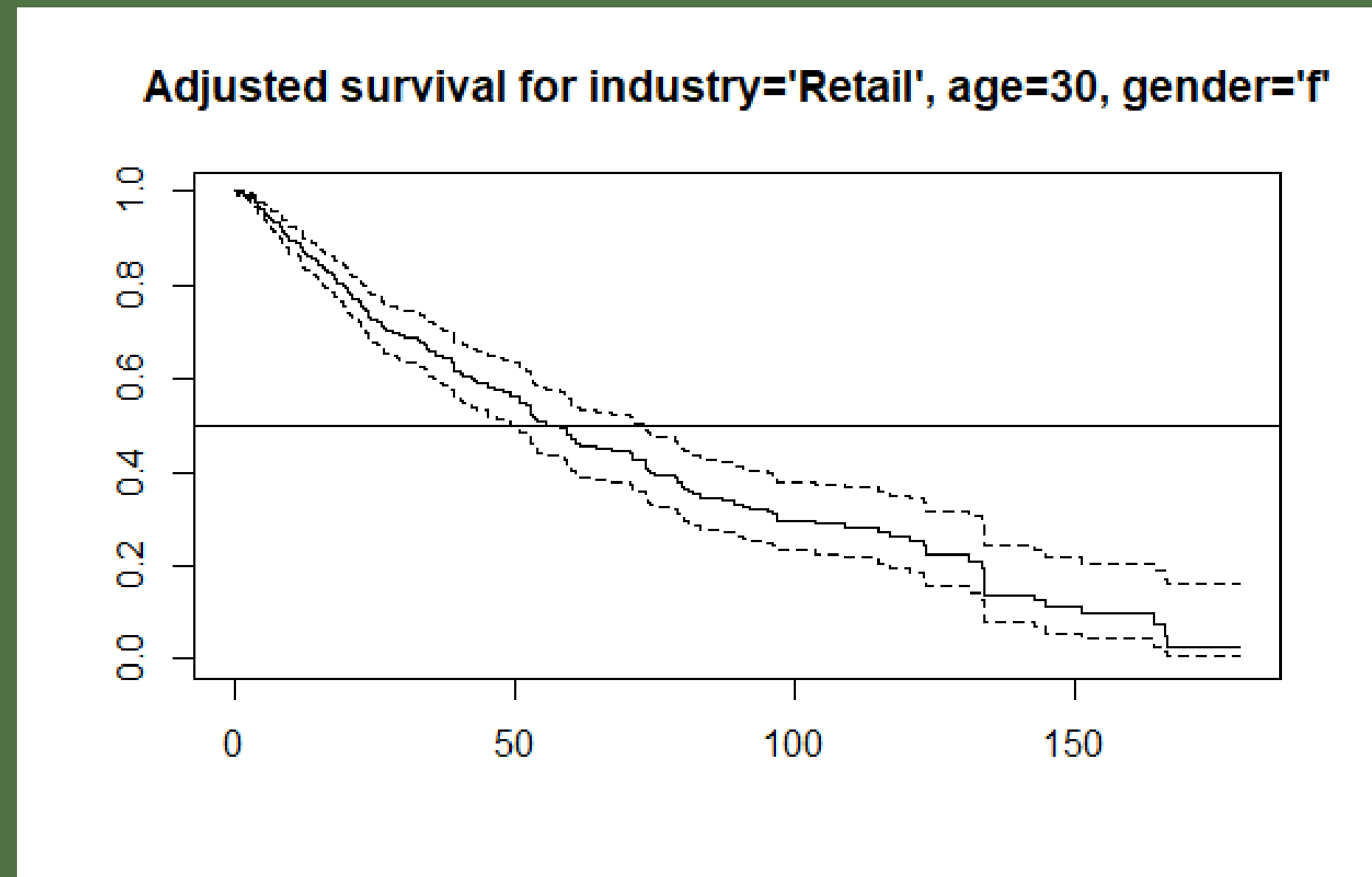
Adjusted Survival

Menggunakan modus dan median variabel sebagai berikut:

- Industry = Retail
- Age = 30
- Gender = Female

Interpretasi:

Pada pemusatan data di masing-masing variabel, didapatkan median dari waktu survivalnya yaitu pada 57.8 hari.



Kesimpulan

1

Variabel industry, age, dan gender telah memenuhi asumsi PH.

2

Model tanpa interaksi adalah model terbaik.

$$h(t,X)=h_0(t)\exp(-0.912\text{IndustryIT} \\ -0.615\text{IndustryManufacture} \\ -0.622\text{IndustryRetail}+0.042\text{Age} \\ -0.222\text{Gender})$$

3

Model telah signifikan secara serentak dan parsial.



Kesimpulan

1. Turnover Karyawan dipengaruhi oleh (1) Perbedaan industri; (2) Usia; (3) Jenis Kelamin
2. Karyawan yang bekerja di industri Bank, memiliki probabilitas risiko mengalami turnover dibanding 3 industri lainnya. Jika diurutkan berdasarkan probabilitas risiko tertinggi, maka urutannya adalah Bank, Manufaktur, Retail, kemudian IT.
3. Setiap penambahan usia sebanyak 1 tahun, maka meningkatkan probabilitas risiko turnover sebesar 1.04 kali.
4. Perempuan memiliki probabilitas risiko turnover yang lebih besar dibanding laki-laki, yaitu sebesar 1,24 kali.
5. Karyawan rata-rata bekerja selama 53 hari sebelum akhirnya mengalami turnover



THANK YOU!



syntax:

www.github.com/intancitra/turnover