

Survival Analysis of Penile Cancer Patients Using Cox-Proportional Hazard and GPT-4

Wiwik Wiyanti*

* Department of Statistics, Matana University, Tangerang, Indonesia

Email: wiwik.chae@gmail.com

Abstract

Today, in 2024, the evolution of the evolution of artificial intelligence (AI) is very impressive. An AI product that continues to develop is ChatGPT. The threat of ChatGPT to some job professions is real. Beside that, ChatGPT also helps many professions make their work easier. In this research, we analyze the survival analysis of penile cancer patients. The design of this research is mixed-method (quantitative-qualitative). The quantitative and qualitative methods that were used are Cox-proportional hazards and GPT-4, respectively. Based on quantitative analysis, it can be concluded that the Survival and Hazard model for study penile cancer are $S(t) = S_0(t)e^{2,385(Stage_2)}$ and $(t) = H_0(t)e^{2,385(stage_2)}$, respectively. The predicted chances of survival and death for the patient in stage 2, 14 days after surgery, are 40% and 60%, respectively. Based on qualitative analysis, GPT-4 cannot quantitatively obtain the cox proportional hazard number. However, even so, GPT4 can read the cox proportional hazard output image, providing Python and R survival analysis coding quite perfectly. Thus, the qualitative method's conclusion is that GPT-4 can greatly assist data analysts in their analysis of survival data.

Keywords: survival, analysis, cancer, gpt4, proportional, hazard

1. INTRODUCTION AND PRELIMINARIES

Humans are living creatures that can get sick or die. As intelligent living creatures, humans can estimate or predict when they will get sick, relapse from illness, or die using science. Statistically, the science that can study predictions can be studied in forecasting methodology or survival analysis. I used survival analysis, just like in this study, to find out how long penile cancer patients will live. In the field of medicine, the word "survival" is frequently employed. For instance, a cardiac patient's prognosis during monitoring following surgery dictates whether or not the patient lives until a specific time frame. Following chemotherapy, cancer patients are monitored for a predetermined amount of time to see whether they pass away or recover from their illness. It could even be a virus, even if the subject is not exclusively human. Survival analysis is the phrase used to describe the analysis of data in the form of times from a well-defined time origin until the occurrence of some particular event or end point [2]. There are many models of survival analysis methods, including the Kaplan-Meier, log rank, cox-proportional hazard, etc.



The research conducted by Farishi and Khambri [6] who researched survival analysis about young breast cancer patients using Kaplan Meier and Log Rank. The result provide information that young breast cancer is associated with high progression, a tendency for recurrence and a worse prognosis than older breast cancer. Pradana and Sofro's research [14] applied Log rank and Kaplan Meier for survival analysis in multiple myeloma patients with the following research results showing that there was no significant influence between the patient's chances of survival based on gender or protein. Next is the research of Dewi, Dwidayati & Agoestanto [4] applying Cox regression with the MLE method for diabetes mellitus sufferers, with research results on blood sugar level factors that significantly affect the life time of diabetes mellitus sufferers at Tugurejo Hospital. The research of Abimayu & Makful [1] using Kaplan Meier, log rank and Cox proportional hazards obtained research results that type II diabetes sufferers who have comorbidities have a lower cumulative survival probability (CSP) and have a higher risk of death compared to those who do not have comorbidities, especially in sufferers who have inappropriate comorbidities (CSP=85.6%; 95% CI=84.4-86.7%; aHR=1.64; 95%; CI=1.150-1.390). Mufarida et al.'s research [13] used six variables/factors, namely age, stage, type of treatment, comorbidities, complications, and anemia status to analyze the survival of cervical cancer patients using the Kaplan Meier and log rank tests.

Given these contexts, I would like to use the Cox proportional hazards to survival analysis testing for cancer of the penis using penile cancer data. Using GPT-4 and Cox proportional hazards in data analysis techniques. The shortcomings of the earlier GPT-3, which was unable to interpret images or perform autocorrelation analysis, were evaluated using GPT-4 [15]. In addition, Melantika & Wiyanti [12] conducted study utilizing GP-4 and found that GP-4 can model forecasts using Box-Jenkin's technique. However, because GPT-4 calculations do not mathematically match manual calculations, as per Evania & Wiyanti's research [5], this study uses GPT-4 to demonstrate whether it is genuinely capable of performing accurate calculations.

The purpose of this study was to implement the Cox proportional-hazard and GPT-4 methods for survival analysis of penile cancer patients through penile cancer patient data. The benefits of this study are to determine the implementation of the Cox proportional-hazard and GPT-4 methods on penile cancer data so that it can be used as a reference for further researchers to determine whether they can use the Cox proportional-hazard and GPT-4 methods in data research related to survival. In addition, this study can be used as scientific evidence to determine whether GPT-4 can be used for survival data analysis.

2. REVIEW OF THE LITERATURE

Basic mathematical terminology and notation for survival analysis of Cox Proportional-Hazard method could be read in [3]. The researcher in this research denote T the random variable for a person's survival time. The small letter t any specific value of interest for the random variable capital T . The Greek letter δ denote a (0,1) random variable indicating either failure or censored, $\delta = 0$ for censored data and $\delta = 1$ for failure if the event occur during the study period. Two quantitative terms considered in any survival analysis are survivor function which mean the probability that a person survives longer than some specified time t , denoted by $S(t)$ and the hazard function, denoted by $h(t)$ is given by the formula the limit as Δt approaches zero, divided by Δt , where Δt denotes a small interval of time. implementation for the purposes of data analysis, these terms can be arranged in Table 2.1.

Table 2.1. General data

Individu id	t	δ	X_1	...	X_n
-------------	-----	----------	-------	-----	-------

1	t_1	δ_1	X_{11}		X_{n1}
2	t_2	δ_2	X_{12}		X_{n2}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	t_n	δ_n	X_{1n}		X_{nn}

2.1. Survival $S(t)$ and Hazard $h(t)$ function

$S(t)$ and $h(t)$ are the survivor function. The survivor function $S(t)$ gives the probability that a person survives longer than some specified time t and the hazard function $h(t)$ is instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t . In fact, the relationship between $S(t)$ and $h(t)$ can be expressed equivalently in either of two calculus formula [11]. See in equation (2.1) and (2.2). If we know one, we determine the other.

$$S(t) = e^{-\int_0^t h(u)du} \quad (2.1)$$

$$h(t) = -\frac{\frac{dS(t)}{dt}}{S(t)} \quad (2.2)$$

In any actual data analysis a computer program can make the numerical transformation from $S(t)$ to $h(t)$, or vice versa, without the user ever having to use either formula. The point here is simply that if you know either $S(t)$ or $h(t)$, you can get the other directly [11]. Describe this statement could be read in Figure 2.1. The flowchart to analysis survival could be read in Figure 3.2.



Figure 2.1. Relationship between $S(t)$ and $h(t)$

2.2. Cox Proportional Hazard

Equation (2.3) is the general formula for the likelihood of KM survival at failure time $t_{(j)}$. Then the formula for the Cox proportional-hazard (PH), is equation (2.4), where $\mathbf{X} = (X_1, X_2, \dots, X_n)$ explanatory/predictor variable.

$$\hat{S}_{t_{(j)}} = \hat{S}_{t_{(j)}} \cdot \widehat{Pr}(T > t_{(j)} | T \geq t_{(j)}) \quad (2.3)$$

$$h(t, \mathbf{X}) = h_0(t) e^{\sum_{i=1}^p \beta_i X_i} \quad (2.4)$$

Despite the lack of a baseline hazard in this study, reasonably reasonable estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be derived for a wide range of data conditions using Cox proportional hazard.

2.3. GPT-4

The ChatGPT used in this research is GPT-4, which is the latest and paid version of GPT. The reason for using GPT-4 is one of the claims from GPT that GPT-4 is better than the previous version and is even claimed to be able to read images [7], while GPT-3 based on previous research is not yet able to read images [15]. It should also be noted that in survival analysis, the output of the Kaplan Meier and log rank data processing is also an image. From the output results in the form of this image, it will prove whether using GPT-4 is capable of reading and analyzing the output results correctly. For this reasons, this research will also be used to prove the claims of GPT-4 whether it is indeed better than GPT-3.

3. RESEARCH METHODOLOGY

The data used in this study is penile cancer patient data obtained from Kagle [8]. The design of this research is a mixed method (quantitative-qualitative), based on Jhon & Creswell [10]. The quantitative and qualitative methods are Cox proportional-hazard and GPT-4, respectively. The quantitative method used in this research is the Cox proportional-hazard method, as referenced in [3,11]. Qualitative method using GPT-4. The research design figure could be seen in Figure 1.1. The full explanation of this design can be read in [10,15]. STATA software was utilized in this study for data processing and analysis.

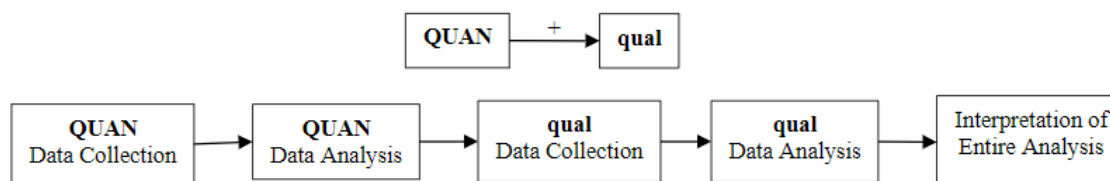


Figure 3.1. Design mixed method sequential explanatory

Flowchart analysis data survival and line the research, could be seen in Figure 3.2 and Figure 3.3, respectively.

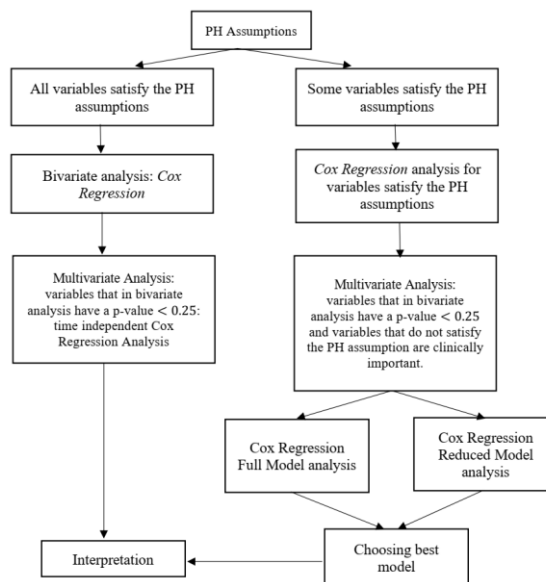


Figure 3.2. Flowchart of survival data analysis.

The explanation of the flow chart in Figure 2.3 is begin with, determine whether all of the variables that will be used for the analysis match the PH assumptions, or whether only a portion of the variables do. Furthermore, if all variables meet the PH assumptions, then data analysis uses bivariate analysis: Cox Regression. In order for multivariate analysis to be applied and evaluated. However, if only some of variables meet the requirements, then use Cox Regression analysis. Next, analyze it with multivariate analysis using a full model and a partial model, and then choosing best model and interpret it.

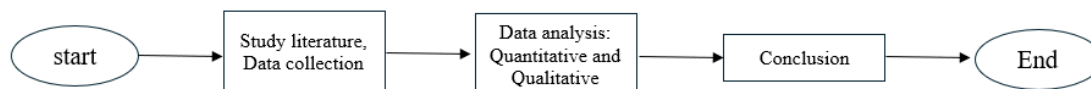


Figure 3.3. Flowchart this research

The explanation in figure 3.3. is that the research begins with a literature study and collecting data. Furthermore, the data obtained is analyzed using quantitative methods using Cox PH and qualitative methods using GPT-4. From this analysis, a conclusion is obtained at the end of the research.

4. MAIN RESULTS

4.1. Quantitative Analysis

International penile cancer statistic were used in this study. Data is sourced from [8] Kaggle. The secondary data on penile cancer (also known as testicular cancer) that was used included 123 observational data, 85 data occurrences, and 38 censored data points. The patient's cancer recovery (the cancer virus's demise) is the occurrence in this case. Age, stage, history, period, and status are the factors that were used in this study. Figure 2.1 shows an example of the sample (the first ten samples) used in this investigation. Based on WHO recommendations [9], the age category division in this study was created. Figure 4.1 displays the research's flowchart. STATA software is the program used to analyze the data for this study.

	id	status	Age	Stage	History	time
1	TCGA-2G-AAEW	Recurred	<=40	IS	No	4.7
2	TCGA-2G-AAEX	VirusDie	<=40	IS	No	41.36
3	TCGA-2G-AAF1	VirusDie	<=40	IS	No	46.09
4	TCGA-2G-AAF4	VirusDie	<=40	StageI	No	76.05
5	TCGA-2G-AAF6	Recurred	<=40	IS	No	26.61
6	TCGA-2G-AAF8	Recurred	<=40	StageIA	Yes	106.24
7	TCGA-2G-AAFE	Recurred	<=40	StageIIA	No	140.28
8	TCGA-2G-AAFG	Recurred	<=40	IS	Yes	191.43
9	TCGA-2G-AAFG	Recurred	<=40	IS	Yes	191.43
10	TCGA-2G-AAFH	VirusDie	<=40	StageIB	No	45.17

Figure 4.1. Screenshot, ten sample data for data analysis from STATA.

As per the flowchart 3.2, testing the proportional-hazard (PH) assumption is the initial stage in data analysis. The global test, log-log plot, and kaplan meier curve are used to perform the PH assumption test [3, 11]. The output outcomes by STATA that were attained are displayed in Table 4.1. From checking the PH assumptions, the variables that satisfy the PH assumptions are the Age and Stage variables. With the result, then bivariate analysis was carried out for these two variables. The bivariate test outputs are as shown in Figure 4.2.

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Wiwik Wiyanti

To facilitate understanding of the results of survival data analysis, this study uses the following labeling terms:

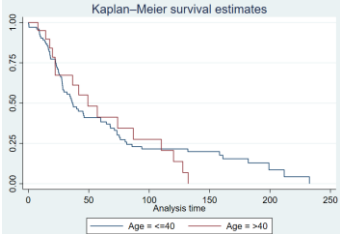
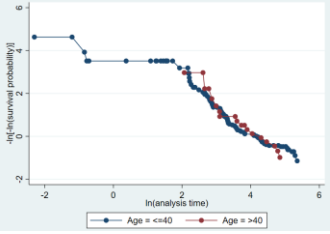
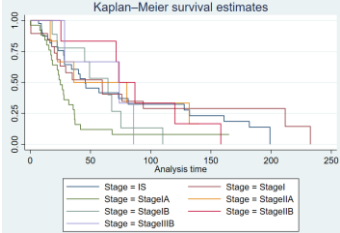
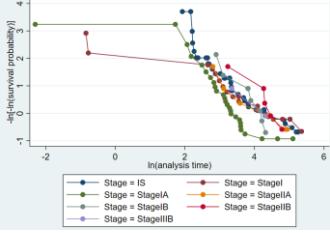
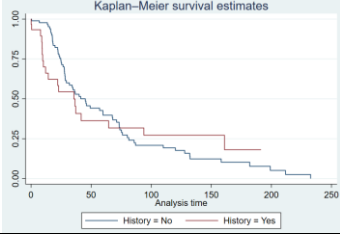
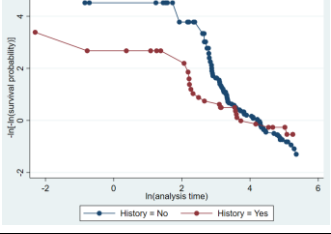
time = survival time of penile cancer patients recorded from the start of surgery until the last visit to the hospital.

status = 0 and 1 (0 = virus is still present or patient has not recovered, 1 = virus is dead or patient has recovered).

Age = age group (0 = patients aged up to 40 years, 1 = patients aged over 40 years).

Stage = Tumor stage (0 = IS, 1 = Stage, 2 = Stage, 3 = StageIIA, 4 = StageIB, 5 = tageIIB, 6 = StageIIIB, 7 = StageIIIC).

Table 4.1. summarizing PH assumption by STATA

Variable	Kaplan Meier	$-\ln(-\ln(\text{survival}))$	Global test								
Age			<pre>. stphtest</pre> Test of proportional-hazards assumption Time function: Analysis time <table border="1"> <thead> <tr> <th></th> <th>chi2</th> <th>df</th> <th>Prob>chi2</th> </tr> </thead> <tbody> <tr> <td>Global test</td> <td>2.89</td> <td>1</td> <td>0.0891</td> </tr> </tbody> </table>		chi2	df	Prob>chi2	Global test	2.89	1	0.0891
	chi2	df	Prob>chi2								
Global test	2.89	1	0.0891								
Stage			<pre>. stphtest</pre> Test of proportional-hazards assumption Time function: Analysis time <table border="1"> <thead> <tr> <th></th> <th>chi2</th> <th>df</th> <th>Prob>chi2</th> </tr> </thead> <tbody> <tr> <td>Global test</td> <td>3.63</td> <td>1</td> <td>0.0567</td> </tr> </tbody> </table>		chi2	df	Prob>chi2	Global test	3.63	1	0.0567
	chi2	df	Prob>chi2								
Global test	3.63	1	0.0567								
History			<pre>. stphtest</pre> Test of proportional-hazards assumption Time function: Analysis time <table border="1"> <thead> <tr> <th></th> <th>chi2</th> <th>df</th> <th>Prob>chi2</th> </tr> </thead> <tbody> <tr> <td>Global test</td> <td>4.45</td> <td>1</td> <td>0.0348</td> </tr> </tbody> </table>		chi2	df	Prob>chi2	Global test	4.45	1	0.0348
	chi2	df	Prob>chi2								
Global test	4.45	1	0.0348								

Next, from output in the Figure 4.2, It can be seen that the P -value is significant only at Stage_2, namely $P = 0.005 < 0.05$ with a 95% confidence level in the interval $[1.28 - 4.08]$ with a Hazard Ratio (HR) is 2.288.

Cox regression with no ties

```

No. of subjects = 123
No. of failures = 85
Time at risk = 6,083.96

Log likelihood = -320.83148

Number of obs = 123
LR chi2(1) = 0.09
Prob > chi2 = 0.7703

```

_t	Haz. ratio	Std. err.	z	P> z	[95% conf. interval]
_IAge_1	1.086144	.3048471	0.29	0.768	.6265868 1.882753

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Wiwik Wiyanti

(a) Age Variable

Cox regression with no ties

No. of subjects = 123	Number of obs = 123
No. of failures = 85	
Time at risk = 6,083.96	
Log likelihood = -315.57824	LR chi2(6) = 10.59
	Prob > chi2 = 0.1018

_t	Haz. ratio	Std. err.	z	P> z	[95% conf. interval]	
_IStage_1	.8384522	.288834	-0.51	0.609	.4268274	1.647041
_IStage_2	2.288062	.6756498	2.80	0.005	1.282666	4.081521
_IStage_3	.9807833	.4836483	-0.04	0.969	.3730995	2.578228
_IStage_4	1.220581	.5030226	0.48	0.629	.5442165	2.737546
_IStage_5	.9265433	.4245772	-0.17	0.868	.3774096	2.274671
_IStage_6	1.151785	.7089871	0.23	0.818	.3446751	3.848867

(b) Stage Variable

Figure 4.2. (a) and (b) are bivariate analysis output.

Next is Multivariate analysis. The results of the multivariate analysis are as shown in Figure 4.3. Meanwhile, the results of Baseline Hazard and Baseline Survival are respectively in Figure 4.4.

Cox regression with no ties

No. of subjects = 123	Number of obs = 123
No. of failures = 85	
Time at risk = 6,083.96	
Log likelihood = -313.02493	LR chi2(12) = 15.70
	Prob > chi2 = 0.2054

_t	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
_IAge_1	-.3114487	.6203105	-0.50	0.616	-1.527235	.9043376
_IStage_1	-.4334025	.3854298	-1.12	0.261	-1.188831	.322026
_IStage_2	.7548151	.3189735	2.37	0.018	.1296385	1.379992
_IStage_3	-.0705304	.5003138	-0.14	0.888	-1.051127	.9100665
_IStage_4	-.315538	.5032507	0.63	0.531	-.6708152	1.301891
_IStage_5	-.0101681	.5470836	-0.02	0.985	-1.082432	1.062096
_IStage_6	.1759831	.743207	0.24	0.813	-1.280676	1.632642
_IAgStag_1	1.900465	.9252307	2.05	0.040	.087046	3.713884
_IAgStag_2	.6614336	.8337451	0.79	0.428	-.9726768	2.295544
_IAgStag_4	-.0409062	.9585966	-0.04	0.966	-1.919721	1.837909
_IAgStag_5	-.0093152	1.065369	-0.01	0.993	-2.097401	2.07877
_IAgStag_6	.1018682	1.375011	0.07	0.941	-2.593103	2.79684

Figure 4.3. Output of multivariate analysis

time	Summary of baseline survivor Mean	time	Summary of cumulative baseline hazard Mean
.1	.99373487	.1	.00624301
.3	.99373487	.3	.00624301
.43	.98741453	.43	.01261033
.46	.98110838	.46	.01900405
.49	.98110838	.49	.01900405
1.45	.98110838	1.45	.01900405
2.96	.98110838	2.96	.01900405
3.45	.98110838	3.45	.01900405
3.52	.98110838	3.52	.01900405
3.98	.98110838	3.98	.01900405
4.27	.98110838	4.27	.01900405
4.57	.98110838	4.57	.01900405
4.7	.98110838	4.7	.01900405
4.83	.98110838	4.83	.01900405
5.62	.98110838	5.62	.01900405
6.87	.97425059	6.87	.0259939

Figure 4.4. Baseline Hazard and Baseline Survival

The model or formula for predicting Hazard Function and Survival Function are

Hazard function:

$$H(t) = H_0(t)e^y \quad (2.1)$$

$$H(t) = H_0(t)e^{2,385(stage_2)} \quad (2.2)$$

Survival function:

$$S(t) = S_0(t)e^y \quad (2.3)$$

$$S(t) = S_0(t)e^{2,385(Stage_2)} \quad (2.4)$$

Model predictions to patients:

For example, implementation model of (2.2) and (2.4) for Hazard and Survival functions for patients on day 14 in stage 2, respectively

$$H(t) = H_0(t)e^y = (0,006)e^{2,385(2)} = 0,6 \quad (2.5)$$

$$S(t) = S_0(t)e^{2,385(Stage)} = 0,4 \quad (2.6)$$

4.2. Qualitative Analysis

The next study design, which uses GPT-4 as a qualitative method, is an examination following quantitative analysis. Begin by questioning GPT-4 about graph survival in this study, similar to the Kaplan-Meier graphic shown in Figure 4.5. From Figure 4.5, we can see the percentage of patient survival, that is, patients still surviving from the operation carried out; more than 50% are on days 0 to around 1300 days (estimate), or if converted into months, it is from 0 months to around 49 months. Illustration screenshot the question and the answer to GPT-4 can be read and interpreted well, as can be seen in Figure 4.6

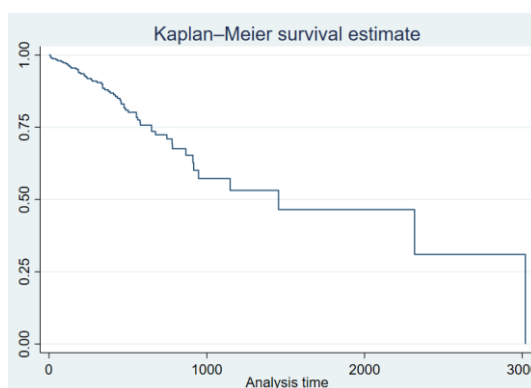
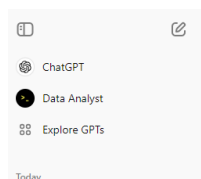


Figure 4.5. Kaplan-Meier survival estimate



Data Analyst ▾



Halo! Gambar yang Anda unggah adalah grafik estimasi bertahan hidup Kaplan-Meier. Grafik ini sering digunakan dalam analisis survival untuk memperkirakan probabilitas seseorang atau suatu objek bertahan hidup hingga waktu tertentu.

Berikut adalah interpretasi dari grafik ini:

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Wiwik Wiyanti

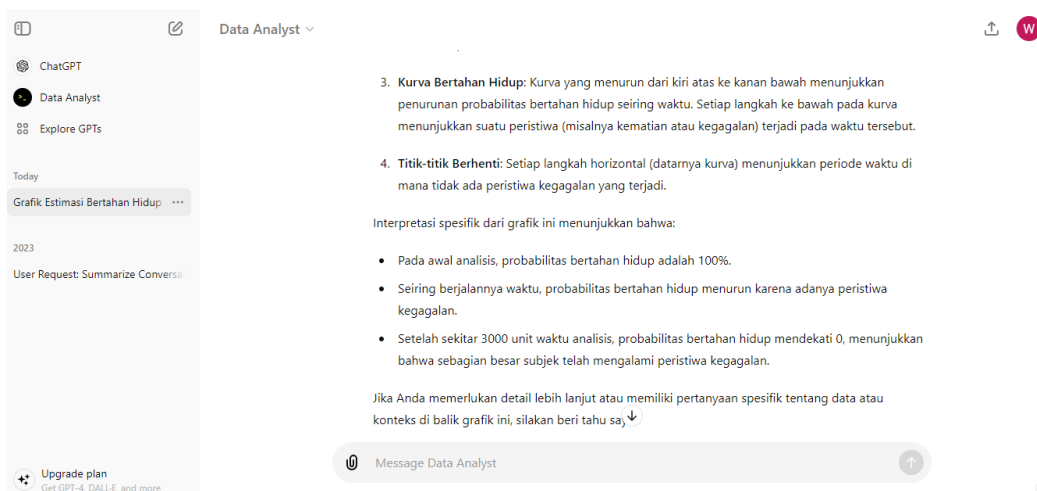


Figure 4.6. Interpretation of the picture from GPT-4.

From the GPT-4 analysis, it can be seen that it is not specifically able to read numbers, but it is able to explain correctly the survival image provided. This shows that GPT-4 makes the data analyst's job easier to interpretation the images. To make it sure, the assumption test curve is given to GPT-4 to interpret it, like Figure 4.7.

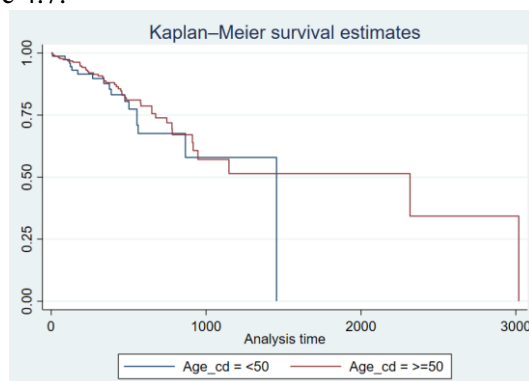
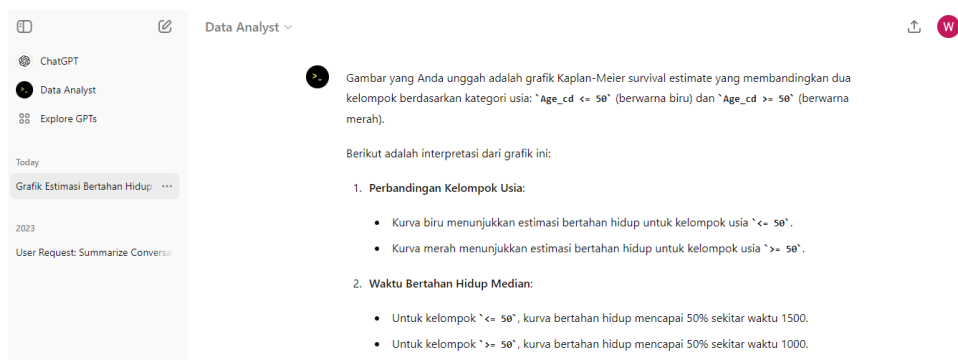


Figure 4.7. Age Kaplan Meier

From Figure 4.8., GPT-4 can explain it very well. It can be read graphic with the color very well. GPT-4 can also provide good conclusions. Overall, using GPT-4 can make the work of data analysts easier.



JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Wiwik Wiyanti

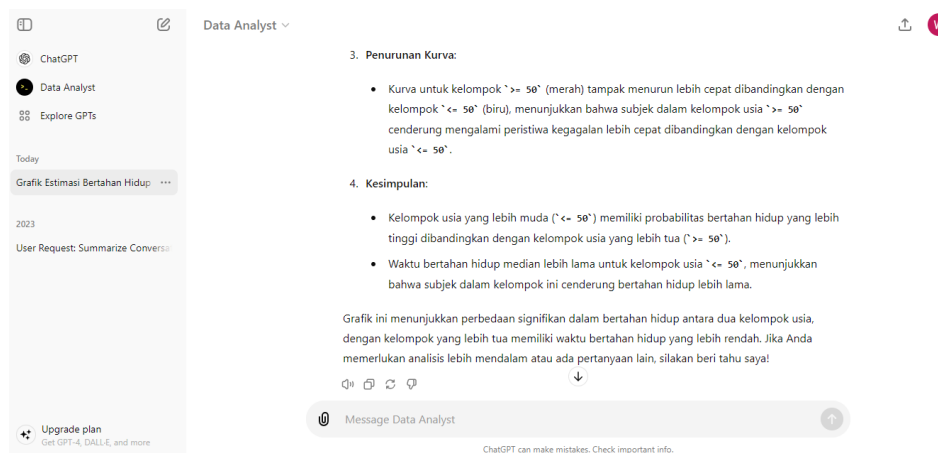


Figure 4.8. Interpretation Age Kaplan-Meier from GPT-4

In order to examine a proportional hazard model using survival data, GPT-4 is unable to determine the equation's output value. Nevertheless, GPT-4 offers recommendations for applying the recommended coding. GPT-4 can provide coding that is suitable for survival analysis algorithms. For example, the Cox Hazard assumption test, GPT-4, can provide log-log survival coding. From the description provided, GPT cannot analyze the image in detail to determine whether the PH assumption is violated or not. However, it impressively provides guidance for checking assumptions using the log log plot method, Schoenfeld residual, and provides R-Studio coding. Of course, it makes it easier for data analysts to carry out data analysis work. To prove it, the researchers used R Studio and the results could be applied correctly to the data. For an illustration of the GPT-4 answer, see Figure 4.9. By entering the coding using Python and R software, results similar to the output in Figure 4.2 are obtained.

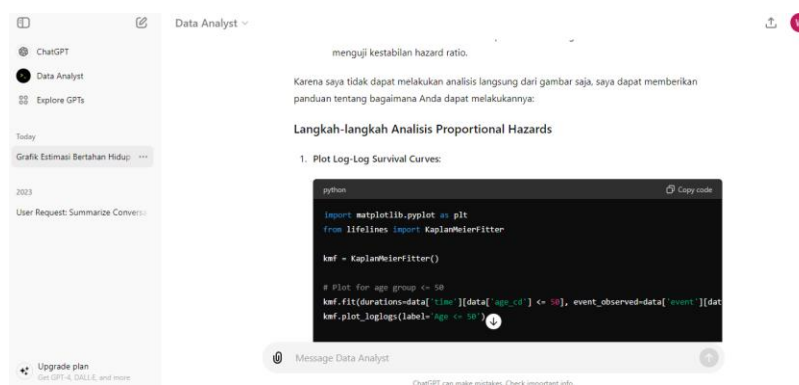


Figure 4.9. GPT-4 answer with the coding advice

3. CONCLUSION

Based on quantitative analysis, it can be concluded that the survival model and the hazard model penile cancer are $S(t) = S_0(t)e^{2,385(Stage_2)}$ and $H(t) = H_0(t)e^{2,385(stage_2)}$, respectively. The predicted chances of survival and death for the patient in stage 2, 14 days after surgery are 40% and 60%, respectively. Based on qualitative analysis, GPT-4 cannot quantitatively obtain the cox proportional hazard number. However, even so, GPT4 can read the cox proportional hazard

output image, providing Python and R survival analysis coding quite perfectly. Thus, the qualitative method's conclusion is that GPT-4 can greatly assist data analysts in their analysis of survival data.

Suggestions for further researchers are to implement other methods to find survival models for patients with different types of diseases. For example, the type of data that can be used is stratified, so the researchers can use the stratified Cox procedure. Furthermore, in order to verify the efficacy of GPT-4 in assisting with data analysis, the researchers can compare its use with manual analysis when facilitating data analysis.

ACKNOWLEDGEMENT

We extend our gratitude to our students, Calvin and Jeremy, for their efforts in collecting the data. Additionally, we thank Matana University for providing ongoing support for the lecturer's research.

CONFLICT OF INTEREST

Full results of this research is truly based on scientifically and there is no competing interests. The author is fully responsible for this matter. Thank you.

REFERENCES

- [1] Abimayu, A.T. & Makful, M.R., 2023. Pengaruh komorbid terhadap tingkat survival pasien dengan diabetes tipe ii: analisis data sampel bpjs 2015-2021. *Jurnal BIKFOKES*.166-76. DOI: <http://dx.doi.org/10.51181/bikfokes.v3i3.7124>. available at: <https://journal.fkm.ui.ac.id/bikfokes/article/view/7124>
- [2] Collet, D., 1994. *Modelling survival data in medical research*. Chapman & Hall.
- [3] Dahlan, M.S., 2009. *Analisis survival dasar-dasar teori dan aplikasi program stata*. Jakarta: Sagung Seto.
- [4] Dewi, A.Y., Dwidayati, N.K. & Agoestanto, A., 2020. Analisis survival model regresi cox dengan metode MLE untuk penderita diabetes militus. *UNNES Journal of Mathematics*. 9(1):31-0. DOI 10.15294/UJM.V9I1.30559.
- [5] Evania, C.D. & Wiyanti, W., 2024. Perbandingan Metode Holt's Linier dan GPT-4 Untuk Peramalan Jumlah Kasus Kematian Diabetes Militus Di Indonesia. *Journal of Mathematics, Computation and Statistics*, Vol. 7 No. 1 (2024): Volume 07 Nomor 01 (April 2024), 1-9. DOI: <https://doi.org/10.35580/jmathcos.v7i1.2063>
- [6] Farisyi, M.A. & Khambri, D., 2018. Analisis survival pasien kanker payudara usia muda di rsup dr m djamil padang tahun 2008-2017. *Jurnal Kesehatan Andalas*.(7):25-9. DOI: <https://doi.org/10.25077/jka.v7i0.917>.
- [7] <https://arxiv.org/abs/2303.08774v3>. <https://doi.org/10.48550/arXiv.2303.08774>
- [8] <https://www.kaggle.com/datasets/antimoni/testicular-cancer> (cited: 29 Desember 2023)
- [9] <https://www.who.int/southeastasia/health-topics/adolescent-health> (cited: 25 Mei 2024)
- [10] John W. & Creswell, 2009. *Research desain: qualitative, quantitative and mixed methods approach*. Third edition. New York: SAGE Publications, Inc.
- [11] Kleinbaum, D.G, Klein, M., 2005. *Survival analysis a self-learning text-second edition*. Second edition. Springer Science+Business Media, Inc.
- [12] Melantika, B.T.E. & Wiyanti, W., 2024. Perbandingan Box-Jenkin's Method dan ChatGPT dalam Memprediksi Jumlah Kematian Penderita Tuberkulosis di Indonesia. *Journal of Mathematics, Computation and Statistics*, Vol. 7 No. 1 (2024): Volume 07 Nomor 01 (April 2024), 10-19. DOI: <https://doi.org/10.35580/jmathcos.v7i1.1948>

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI
Wiwik Wiyanti

- [13] Mufarida, A. Saputra, N. Yuriska, R.T. & Liaunillah, Y., 2023. Analisis kurva survival kaplan meier menggunakan uji log rank pada pasien penyakit kanker serviks. *IcoPSS*. Volume2 issue 1,14-18. doi.org/10.24014/icopss.v2i1.25324
- [14] Pradana, A.N.I. & Sofro, A., 2019. Penerapan analisis survival pada pasien multiple myeloma. *Jurnal Ilmiah Matematika*. Volume 7 No.2:45-0. ISSN 2301-9115.
- [15] Wiyanti W, Siregar B., 2023. Comparison predictions of the demam berdarah dengue (dbd) using model exponential smoothing: pegel's classification and chatgpt. *Jurnal Matematika, Statistika dan Komputasi*. 20(2),351-59. <https://doi.org/10.20956/j.v20i2.32122>.