



PROVA DE ANÁLISE DE DADOS TREINAMENTO 3

SELEÇÃO DAS EQUIPES BRASILEIRAS OLIMPIADAS INTERNACIONAIS DE 2023

Instruções Gerais

1. Escreva seu NÚMERO DE IDENTIFICAÇÃO em TODAS as folhas de resposta que serão escaneadas.
2. Escreva o número de cada questão na folha de resposta, bem como o número da página.
3. Essa prova é de aplicação única. NÃO HAVERÁ SEGUNDA CHAMADA.
4. A duração da prova é de 2 horas (2h) para resolução mais o tempo de 20 (vinte) minutos para *scan*, sem possibilidade de tempo adicional, a não ser em casos de imprevistos.
5. A prova é composta por 2 questões (totalizando 150 pontos).
6. A prova é individual e sem consultas. Informações relevantes para a Prova de Análise de Dados estão disponibilizadas nas páginas 2 e 3.
7. O uso de calculadoras é permitido, desde que não sejam programáveis/gráficas/com acesso a internet.
8. As resoluções das questões podem ser feitas a lápis (bem escuro) ou caneta e devem ser apresentadas de forma clara, concisa e completa. Faça um retângulo ao redor da resposta de cada item. Sempre organize seus dados em tabelas. Recomendamos o uso de borracha, régua e compasso.
9. Você pode utilizar folhas de rascunho para auxiliar no processo de resolução da prova, mas elas não devem ser escaneadas.

Instruções Específicas

1. Após o término da prova, os alunos deverão escanear suas soluções com um aparelho celular para enviarem suas provas pelo Gradescope.
2. Só serão aceitos arquivos em pdf. Em caso de dúvidas, leia o passo a passo da OBA de como escanear suas soluções.
3. O uso de aparelhos celulares ou câmeras fotográficas só é permitido enquanto o aluno realiza o scan de suas soluções.
4. Para questões em branco, faça *upload* de uma folha escrito apenas “Esta questão foi pulada”.

Formulário

Incerteza e Propagação de Erros

Para encontrar a incerteza σ_ω de uma função genérica $\omega = \omega(x_1, x_2, \dots, x_n)$, a fórmula geral é:

$$\sigma_\omega^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \dots + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_{x_n}^2$$

A tabela abaixo mostra algumas aplicações clássicas disso:

Função $\omega = \omega(x, y, \dots)$	Expressões para σ_ω
$\omega = x \pm y \pm \dots$	$\sigma_\omega^2 = \sigma_x^2 + \sigma_y^2 + \dots$
$\omega = x^m$	$\sigma_\omega = mx^{m-1} \sigma_x$ ou $\left \frac{\sigma_\omega}{\omega}\right = \left m \frac{\sigma_x}{x}\right $
$\omega = ax$	$\sigma_\omega = a \sigma_x$ ou $\left \frac{\sigma_\omega}{\omega}\right = \left \frac{\sigma_x}{x}\right $
$\omega = ax + b$	$\sigma_\omega = a \sigma_x$
$\omega = axy$	$\sigma_\omega^2 = (ay)^2 \sigma_x^2 + (ax)^2 \sigma_y^2$ ou $\left(\frac{\sigma_\omega}{\omega}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2$
$\omega = a \frac{x}{y}$	$\sigma_\omega^2 = \left(\frac{a}{y}\right)^2 \sigma_x^2 + \left(\frac{ax}{y^2}\right)^2 \sigma_y^2$ ou $\left(\frac{\sigma_\omega}{\omega}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2$
$\omega = ax^p y^q$	$\sigma_\omega^2 = (apx^{p-1}y^q)^2 \sigma_x^2 + (ax^p qy^{q-1})^2 \sigma_y^2$ ou $\left(\frac{\sigma_\omega}{\omega}\right)^2 = \left(p \frac{\sigma_x}{x}\right)^2 + \left(q \frac{\sigma_y}{y}\right)^2$
$\omega = a \sin(bx)$	$\sigma_\omega = ab \cos(bx) \sigma_x$
$\omega = b \log_a x$	$\sigma_\omega = \left \frac{b}{\ln a}\right \frac{\sigma_x}{x}$

Método dos Mínimos Quadrados

- Pelo método dos mínimos quadrados, a melhor estimativa para os coeficientes da reta $y = A + Bx$ são:

$$A = \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta}$$

$$B = \frac{N \sum xy - \sum x \sum y}{\Delta}$$

Onde:

$$\Delta = N \sum x^2 - (\sum x)^2$$

- As incertezas dos coeficientes são dadas por:

$$\sigma_A = \sigma_y \sqrt{\frac{\sum x^2}{\Delta}}$$

$$\sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}$$

Em que:

$$\sigma_y = \sqrt{\frac{1}{N-2} \sum_{i=1}^N (y_i - A - Bx_i)^2}$$

Há também outras fórmulas que levam aos mesmos resultados, e são mais rápidas:

$$\sigma_A = \sigma_B \sqrt{\frac{\sum x^2}{N}}$$

$$\sigma_B = |B| \sqrt{\frac{\frac{1}{r^2} - 1}{N-2}}$$

Em que r é o coeficiente de correlação linear, A e B são os coeficientes linear e angular da reta de melhor ajuste, respectivamente, e N é o número de pontos.

Derivadas e Regras Úteis

- Derivada da Função Exponencial:

$$\frac{d}{dx} e^x = e^x$$

- Derivada do Logaritmo Natural:

$$\frac{d}{dx} \ln(x) = \frac{1}{x}$$

- Regra da Cadeia:

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$$

1. Viés de Malmquist - Olimpíada Iraniana (60 pontos)

Compreendendo o viés de Malmquist

O viés de Malmquist é um efeito na astronomia observacional que leva à detecção preferencial de objetos intrinsecamente brilhantes. Na estatística, esse fenômeno é chamado de viés de seleção ou censura de dados. Se tivermos um certo limite na observação de estrelas ou outros objetos (como a magnitude aparente limítrofe de um detector), não observaremos uma composição homogênea de todas as estrelas do céu, e a maioria das estrelas observadas serão aquelas que possuem maior brilho intrínseco. Por esse motivo, se precisarmos usar os valores médios de luminosidade de objetos como estrelas da mesma categoria ou espectro, teremos um grande erro. Para entender melhor esse fenômeno, preste atenção no gráfico 1. Os dados referentes a esse gráfico também são apresentados na tabela (1). Este diagrama de dispersão mostra a distância das estrelas com categoria espectral G. Assumimos que a distribuição da magnitude absoluta das estrelas pertencentes a esta categoria espectral segue uma distribuição normal com média $M_0 = 4,75$ e desvio padrão $\sigma = 0.4$.

- (a) **(15 pontos)** Considerando a utilização de um telescópio de magnitude limite $m_{lim_1} = 14$, desenhe uma curva no gráfico 1 que separa as estrelas observáveis das inobserváveis. Desenhe uma curva análoga para o caso em que $m_{lim_2} = 16$. Com essa nova magnitude limite, qual será o aumento (em porcentagem) do número de estrelas observáveis? O número total de estrelas da amostra é 200.
- (b) **(7 pontos)** Calcule a magnitude absoluta média das estrelas observáveis M para o caso em que $m_{lim} = 14$. Compare seu resultado com a média real (M_0) calculando $\Delta M = M - M_0$ e escreva sua conclusão a partir dessa comparação.

Método de Malmquist para a correção do fenômeno

No início da década de 1920, Malmquist derivou um método para corrigir esse fenômeno considerando algumas suposições, sendo essas:

- Não há absorção interestelar (ie. o material no espaço entre as estrelas, como gás e poeira, não está afetando a luz e absorvendo parte dela);
- A função de luminosidade (ϕ) é independente da distância (r). Isso basicamente significa que o universo é o mesmo em todos os lugares e que as estrelas estão distribuídas de maneira semelhante em qualquer outro lugar;
- A densidade numérica das estrelas é independente da linha de visão e é apenas uma função da distância do observador;
- A função de distribuição da magnitude absoluta de estrelas de uma categoria espectral segue uma função normal (Gaussiana).

Observe que a função de distribuição normal com média M_0 e desvio padrão σ é a seguinte:

$$\phi(M) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(M-M_0)^2}{2\sigma^2}}$$

Para entender o método matemático desenvolvido por Malmquist, seguiremos os seguintes passos:

- (c) **(13 pontos)** A partir da função de distribuição ϕ , é possível encontrar uma função $a(m)$ da abundância de estrelas com magnitude aparente m . Isto é, $a(m) \cdot dm$ representa o número de estrelas da classe analisada que podem ser encontradas com magnitude aparente entre $m \pm dm/2$ ao observar uma região do céu no interior de um ângulo sólido Ω . É possível mostrar que $a(m)$ pode ser representado da seguinte maneira:

$$a(m) = \int_0^{\infty} f(r, m) dr$$

Mostre que ϕ pode ser escrito como função das variáveis m e r e, por fim, encontre $f(r,m)$ em função de $\phi(m,r)$, da densidade numérica de estrelas $n(r)$, do ângulo sólido observado Ω e da variável de distância r .

Utilizando a função $a(m)$, podemos encontrar o valor médio da magnitude absoluta \overline{M} das estrelas com uma dada magnitude aparente m do seguinte modo:

$$\overline{M}_m a(m) = \int_0^\infty M f(r, m) dr$$

(d) **(15 pontos)** Mostre que:

$$\Delta M_m = \overline{M}_m - M_0 = -\sigma^2 \frac{d(\ln a(m))}{dm}$$

Dica: Pode ser útil utilizar a seguinte propriedade:

$$\frac{da(m)}{dm} = \int_0^\infty \frac{df(r, m)}{dm} dr$$

Ainda a partir da função $a(m)$, também podemos encontrar a função $A(m) = \int_{-\infty}^m a(m) dm$ que nos dá o número total de estrelas com magnitude menor que m no ângulo sólido observado. Assim, podemos encontrar a magnitude absoluta média \overline{M} de todas as estrelas observáveis - com magnitude aparente menor que m - do seguinte modo:

$$\overline{M} A(m) = \int_{-\infty}^m \overline{M}_m a(m) dm$$

Com alguns passos matemáticos, podemos então encontrar a seguinte relação:

$$\Delta M = \overline{M} - M_0 = -\sigma^2 \frac{d(\ln A(m))}{dm}$$

(e) **(10 pontos)** Mostre que para o caso em que a densidade numérica n é constante, a relação acima pode ser simplificada para:

$$\Delta M = \overline{M} - M_0 = -1,382\sigma^2$$

A partir dos dados fornecidos, calcule também o valor de ΔM para o problema em questão. Qual a diferença percentual com o resultado obtido no item b)?

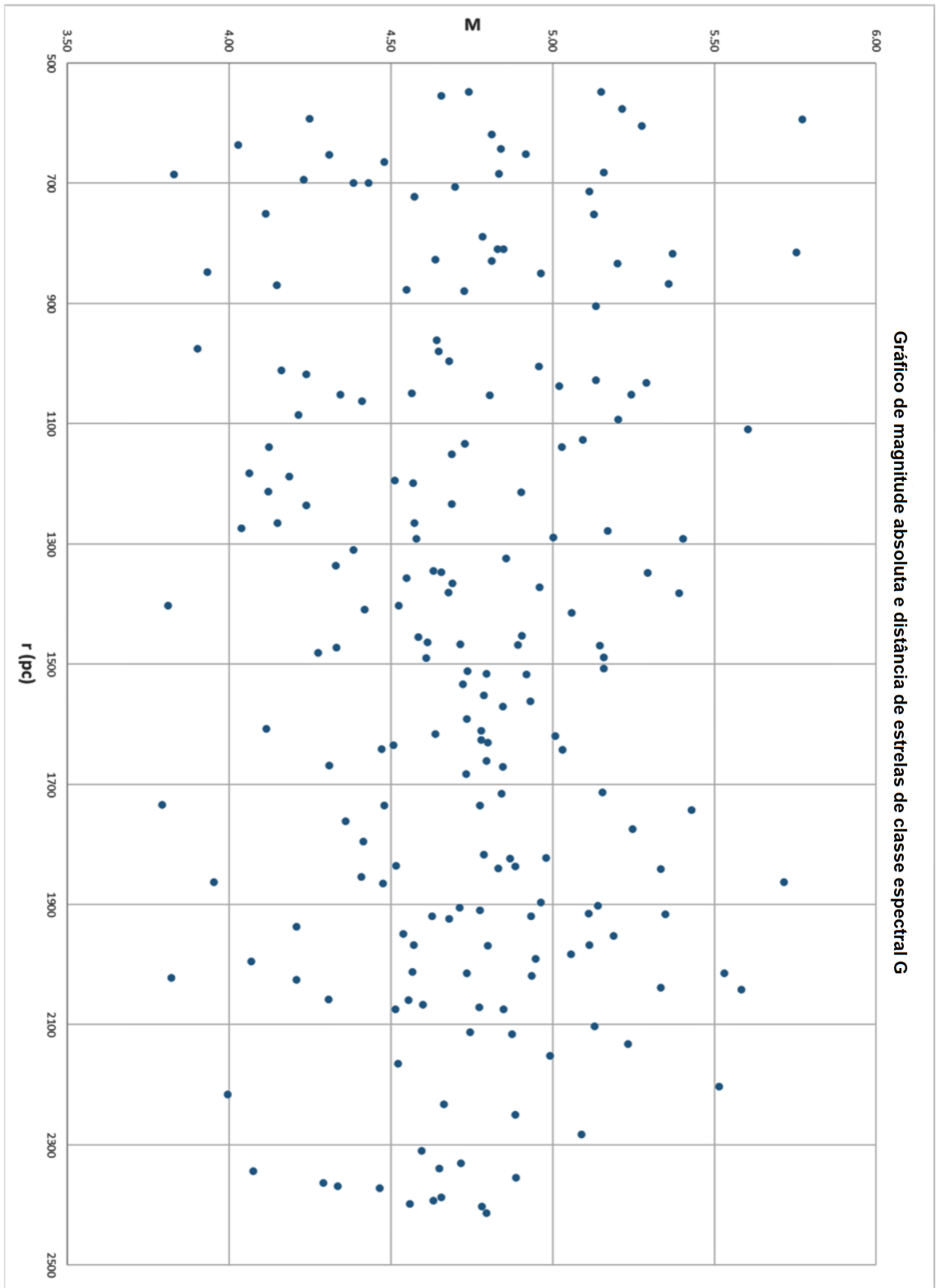


Tabela 1: Dados do gráfico

$r(pc)$	M	$r(pc)$	M	$r(pc)$	M	$r(pc)$	M
548	4.74	1052	5.24	1489	5.16	1920	4.93
548	5.15	1052	4.35	1490	4.61	1920	4.63
554	4.66	1053	4.81	1508	5.16	1924	4.68
577	5.21	1062	4.41	1512	4.74	1938	4.21
593	4.25	1086	4.22	1516	4.80	1949	4.54
594	5.77	1093	5.20	1517	4.92	1953	5.19
604	5.28	1109	5.60	1533	4.72	1968	4.57
618	4.81	1127	5.09	1553	4.79	1968	5.11
636	4.03	1134	4.73	1562	4.93	1969	4.80
643	4.84	1139	5.03	1571	4.85	1982	5.06
652	4.92	1139	4.12	1591	4.73	1991	4.95
653	4.31	1151	4.69	1608	4.12	1995	4.07
665	4.48	1182	4.06	1611	4.78	2013	4.57
683	5.16	1188	4.19	1617	4.64	2015	4.74
684	4.83	1195	4.51	1620	5.01	2015	5.53
685	3.83	1199	4.57	1627	4.78	2019	4.94
694	4.23	1213	4.12	1631	4.80	2022	3.82
700	4.43	1214	4.90	1635	4.51	2025	4.21
700	4.38	1234	4.69	1641	4.47	2038	5.33
707	4.70	1235	4.24	1643	5.03	2042	5.58
714	5.11	1266	4.15	1661	4.80	2059	4.31
722	4.57	1266	4.57	1669	4.31	2059	4.56
750	4.11	1274	4.04	1671	4.85	2067	4.60
752	5.13	1279	5.17	1683	4.73	2071	4.77
789	4.78	1290	5.00	1713	5.15	2075	4.52
810	4.83	1291	4.58	1715	4.84	2075	4.85
810	4.85	1292	5.40	1735	3.79	2103	5.13
816	5.75	1310	4.39	1735	4.78	2113	4.75
817	5.37	1324	4.86	1735	4.48	2116	4.87
827	4.64	1336	4.33	1743	5.43	2132	5.23
829	4.81	1345	4.63	1762	4.36	2151	4.99
834	5.20	1348	4.66	1775	5.25	2165	4.52
847	3.93	1348	5.29	1796	4.42	2204	5.51
850	4.96	1357	4.55	1817	4.79	2216	4.00
867	5.36	1366	4.69	1823	4.98	2233	4.66
870	4.15	1372	4.96	1824	4.87	2250	4.89
878	4.55	1381	4.68	1836	4.52	2283	5.09
880	4.73	1382	5.39	1837	4.89	2310	4.59
905	5.13	1403	3.81	1841	4.83	2331	4.72
962	4.64	1403	4.52	1842	5.34	2339	4.65
975	3.90	1410	4.42	1854	4.41	2344	4.07
980	4.65	1414	5.06	1863	3.95	2355	4.89
996	4.68	1453	4.91	1863	5.72	2364	4.29
1005	4.96	1455	4.59	1865	4.48	2369	4.34
1012	4.16	1464	4.61	1897	4.97	2372	4.47
1018	4.24	1468	4.72	1902	5.14	2388	4.66
1028	5.13	1468	4.89	1905	4.71	2393	4.63
1032	5.29	1469	5.15	1910	4.78	2398	4.56
1037	5.02	1473	4.33	1915	5.11	2403	4.78
1050	4.57	1481	4.28	1916	5.35	2413	4.80

2. A Relação Massa-Luminosidade - SAO 2018 Adaptada (90 pontos)

Atenção: O cálculo de incertezas só é necessário quando explicitamente requisitado.

A relação massa-luminosidade para estrelas da sequência principal foi proposta pela primeira vez em um artigo do astrônomo Jakob Karl Ernst Halm em 1911. Essa relação empírica tem a forma funcional

$$L \propto M^\alpha \Rightarrow \frac{L}{L_\odot} = k \left(\frac{M}{M_\odot} \right)^\alpha$$

onde L é a luminosidade da estrela, M é a massa da estrela, L_\odot é a luminosidade solar, M_\odot é a massa solar, k é uma constante de proporcionalidade que depende da massa da estrela, e α é o expoente da lei de potência. Pesquisas de estrelas próximas e binárias na sequência principal mostram que uma relação única $L \propto M^\alpha$ se desvia significativamente dos valores observados para estrelas de massas variadas. Assim, buscaremos neste problema encontrar uma relação massa-luminosidade por partes, a qual fornece um ajuste melhor para estrelas em diferentes faixas de massa. **Uma tabela com dados de 53 estrelas dentro da faixa $0,38 M_\odot \leq M < 23 M_\odot$ é fornecida ao fim da questão, na Página 9.**

- (20 pontos) Faça um gráfico de $\log(M/L)$ versus M para estrelas na faixa $M < 7 M_\odot$. Chame-o de **Gráfico 1**.
- (8 pontos) Em seu **Gráfico 1**, você deve conseguir observar duas quebras distintas no comportamento dos dados, as quais dividem a amostra em três distribuições lineares. Identifique essas quebras no gráfico por meio de duas linhas verticais cheias, e indique os valores de massa M_1 e M_2 nas quais elas ocorrem, em M_\odot .
- (15 pontos) Faça um gráfico de $\log(M/L)$ versus M para estrelas na faixa $M > 3 M_\odot$. Chame-o de **Gráfico 2**.
- (4 pontos) Similarmente, em seu **Gráfico 2**, você deve conseguir observar uma quebra adicional no comportamento dos dados, a qual divide a amostra em duas distribuições lineares. Identifique-a por meio de uma linha vertical cheia, e indique o valor de massa M_3 na qual ela ocorre, em M_\odot .

A relação massa-luminosidade por partes para essas quatro distribuições de massa de estrelas (muito baixas, baixas, intermediárias, altas) podem ser expressas como

$$\frac{L}{L_\odot} = \begin{cases} 1,062 \left(\frac{M}{M_\odot} \right)^{4,841}, & 0,38 M_\odot < M \leq M_1 \\ k_2 \left(\frac{M}{M_\odot} \right)^{\alpha_2}, & M_1 < M \leq M_2 \\ k_3 \left(\frac{M}{M_\odot} \right)^{\alpha_3}, & M_2 < M \leq M_3 \\ k_4 \left(\frac{M}{M_\odot} \right)^{\alpha_4}, & M_3 < M \leq 23 M_\odot \end{cases}$$

- (30 pontos) Para estrelas na faixa $M_1 < M < 23 M_\odot$, faça um gráfico apropriado que lhe permita obter os valores de k_2 , k_3 , k_4 , α_2 , α_3 , e α_4 . Chame-o de **Gráfico 3**. Trace nele as três retas de melhor ajuste associadas aos dados dos intervalos recém-definidos, e obtenha k_2 , k_3 , k_4 , α_2 , α_3 , e α_4 .
- (5 pontos) Usando a relação apropriada, calcule a magnitude absoluta M_V na banda V esperada para a estrela *Formalhaut*, dada a sua massa $M = 1,92 M_\odot$ e sua correção bolométrica $BC = -0,07$ mag.
- (8 pontos) Estime a incerteza dos coeficientes α e k da relação apropriada escolhida, bem como a incerteza associada à M_V .

Nome da estrela	M/M_{\odot}	$\log(M/M_{\odot})$	$\log(L/L_{\odot})$	$\log(M/L)$
2MASS J23143816+0339493 B	0.383	-0.417	-1.824	-1.407
CU Cnc A	0.427	-0.370	-1.770	-1.400
2MASS J23143816+0339493 A	0.469	-0.329	-1.602	-1.273
2MASS J10305521+0334265 A	0.499	-0.302	-1.444	-1.142
2MASSJ07431157+0316220 B	0.544	-0.264	-1.398	-1.134
2MASSJ04480963+0317480 A	0.567	-0.246	-1.187	-0.941
TYC 5998-1918-1 B	0.687	-0.163	-0.910	-0.747
TYC 5998-1918-1 A	0.703	-0.153	-0.810	-0.657
V568 B	0.827	-0.082	-0.516	-0.434
V636 B	0.854	-0.069	-0.408	-0.339
KIC 6131659 A	0.922	-0.035	-0.146	-0.111
ZZ UMa B	0.960	-0.018	-0.090	-0.072
V565 A	0.995	-0.002	0.029	0.031
V568 A	1.075	0.031	0.258	0.227
NGC188 KR V12 A	1.102	0.042	0.344	0.302
VZ Hya A	1.271	0.104	0.481	0.377
IT Cas A	1.330	0.124	0.607	0.483
DM Vir A	1.454	0.163	0.698	0.535
TV Nor B	1.661	0.220	0.907	0.687
XY Cet A	1.773	0.249	1.084	0.835
V335 B	1.844	0.266	1.157	0.891
V392 Car A	1.900	0.279	1.162	0.883
AI Hya A	1.974	0.295	1.241	0.946
TV Nor A	2.048	0.311	1.328	1.017
SW CMa B	2.105	0.323	1.382	1.059
FT Ori A	2.168	0.336	1.426	1.090
SW CMa A	2.240	0.350	1.567	1.217
V624 A	2.277	0.357	1.560	1.203
V396 Cas A	2.397	0.380	1.640	1.260
Zet Phe B	2.545	0.406	1.793	1.387
V451 A	2.769	0.442	1.932	1.490
Psi Cen A	3.114	0.493	2.150	1.657
V906 B	3.246	0.511	2.163	1.652
Zet Phe A	3.922	0.594	2.515	1.921
GG Lup A	4.106	0.613	2.381	1.768
V760 B	4.610	0.664	2.650	1.986
EP Cru B	4.830	0.684	2.790	2.106
V760 A	4.969	0.696	2.827	2.131
V399 Vul B	5.440	0.736	3.088	2.352
CV Vel B	5.977	0.776	3.138	2.362
CV Vel A	6.076	0.784	3.206	2.422
V379 Cep B	6.233	0.795	3.149	2.354
V1388 Ori A	7.421	0.870	3.697	2.827
V399 Vul A	7.550	0.878	3.694	2.816
HI Mon B	9.864	0.994	4.137	3.143
V379 Cep A	10.873	1.036	4.130	3.094
HI Mon A	11.426	1.058	4.220	3.162
CW Cep A	11.797	1.072	4.245	3.173
AH Cep B	13.714	1.137	4.315	3.178
V453 Cyg A	14.377	1.158	4.518	3.360
AH Cep A	15.565	1.192	4.466	3.274
EM Car B	21.376	1.330	4.922	3.592
EM Car A	22.833	1.359	5.021	3.662