



ANÁLISE DE DADOS  
SELEÇÃO DAS EQUIPES BRASILEIRAS  
OLIMPÍADAS INTERNACIONAIS DE 2025

---

## Instruções Gerais

1. Identifique seu grupo em **TODAS** as folhas de respostas. Não coloque mais nenhum meio de identificação pessoal;
2. Escreva o número de cada questão nas folhas de respostas;
3. Enumere as folhas de resposta em ordem crescente com o número das questões. A enumeração não deve reiniciar a cada questão;
4. Se não responder a uma ou mais questões, escreva uma folha declarando os números das questões não resolvidas, p. ex., “não respondi à Q1 e à Q2”;
5. A duração da prova é de 4 horas;
6. Essa prova é composta por 5 questões valendo um total de 300 pontos (4 questões valendo 50 pontos e questão 1 valendo 100);
7. O uso de calculadoras é permitido, desde que não sejam programáveis/gráficas;
8. Não é permitido o uso de celulares ou similares, nem calculadoras de celulares;
9. Todo o desenvolvimento, cálculos e respostas das questões devem ser feitos nas folhas de respostas. Serão desconsideradas as respostas que requererem, mas não apresentarem, as devidas explicações e desenvolvimentos matemáticos.
10. Ao final da prova, devolva as somente as folhas utilizadas para resolução.
11. Um formulário com informações relevantes para análise de dados está disponibilizado.

## Formulário

- Para uma função real  $y(x)$  cuja taxa de variação é diretamente proporcional ao valor da função, é válido que:

$$\frac{\Delta y}{\Delta x} = \alpha \cdot y \rightarrow y = y(0) \cdot e^{\alpha x}$$

- Podemos reescrever  $(1+x)^n$  como um polinômio de grau tendendo ao infinito:

$$(1+x)^n = 1 + nx + n(n-1)\frac{x^2}{2} + \dots$$

Se desprezarmos os termos de ordem superior em  $x$ , obtemos:

$$(1+x)^n \approx 1 + nx$$

- Podemos reescrever  $e^x$  como um polinômio de grau tendendo ao infinito:

$$e^x = 1 + x + \frac{x^2}{2} + \dots$$

Se desprezarmos os termos de ordem superior em  $x$ , obtemos:

$$e^x \approx 1 + x$$

- O desvio padrão de uma amostra de dados  $x_i$  é calculado da seguinte forma:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

Em que  $\bar{x}$  é o valor médio (média aritmética simples) da amostra, e  $N$  o tamanho do conjunto amostral.

- As seguintes constantes físicas podem vir a serem úteis:

Constante da gravitação universal:  $G = 6,67 \cdot 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$

Velocidade da luz no vácuo:  $c = 3,00 \cdot 10^8 \text{ m s}^{-1}$

Massa solar:  $M_{\odot} = 1,99 \cdot 10^{30} \text{ kg}$

Parsec:  $1 \text{ pc} = 3,09 \cdot 10^{16} \text{ m}$

Tabela 1: Tabela de escore-z. A informação na linha representa o escore-z até a primeira casa depois da vírgula, e a informação na coluna representa a segunda casa. Por exemplo, o p-valor associado ao escore-z de 0,42 é 0,66276

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5279	0,53188	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,71566	0,71904	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,9803	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,983	0,98341	0,98382	0,98422	0,98461	0,985	0,98537	0,98574
2,2	0,9861	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,9901	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,9918	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,9943	0,99446	0,99461	0,99477	0,99492	0,99506	0,9952
2,6	0,99534	0,99547	0,9956	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,9972	0,99728	0,99736
2,8	0,99744	0,99752	0,9976	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,999
3,1	0,99903	0,99906	0,9991	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,9994	0,99942	0,99944	0,99946	0,99948	0,9995
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,9996	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,9997	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,9998	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,9999	0,9999	0,9999	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

**1. DAora (60 pontos)**

Determinar o período de fenômenos astronômicos periódicos é essencial para compreender suas propriedades físicas. Imagine que você esteja estudando a estrutura de pequenos corpos próximos da Terra. Usando seu telescópio, você observa variações periódicas na magnitude de um desses objetos — variações que refletem a rotação do corpo, já que a luz solar é refletida de forma diferente conforme ele gira. Conhecer esse período de rotação é importante para modelar as forças centrífugas internas que atuam sobre o objeto, o que, por sua vez, permite inferir aspectos estruturais da sua composição. No entanto, há desafios. Um deles é a limitação da frequência de coleta de dados, por culpa, por exemplo, da quantidade de objetos sendo estudados, ou mais inevitavelmente, do ciclo de dia e noite. A tabela a seguir fornece dados fictícios, que poderiam ser provenientes de um estudo com coleta de dados a cada 24h.

- (a) **(10 pontos)** No papel milimetrado, esboce a curva expressa pela seguinte equação:

$$y(t) = \cos\left(\frac{2\pi}{15} \cdot t\right)$$

de  $t = 0$  até  $t = 150$ . Esse é um modelo simplificado de fenômeno periódico.

- (b) **(5 pontos)** Complete a tabela a seguir com os valores de  $y$  para valores de  $t$  de 24 em 24h.

$t$	$y(t)$
0	
24	
48	
72	
96	
120	
144	

- (c) **(20 pontos)** Encontre os parâmetros  $A$  e  $B$  para o seguinte ajuste de curva:

$$y(t) = A \cdot \cos\left(\frac{2\pi}{60} \cdot t\right) + B \cdot \sin\left(\frac{2\pi}{60} \cdot t\right)$$

Se necessário, utilize a solução do método dos mínimos quadrados para duas variáveis:

$$y = \alpha_1 x_1 + \alpha_2 x_2 \Rightarrow \begin{cases} \alpha_1 \cdot \sum x_1^2 + \alpha_2 \cdot \sum x_1 x_2 = \sum x_1 y \\ \alpha_1 \cdot \sum x_1 x_2 + \alpha_2 \cdot \sum x_2^2 = \sum x_2 y \end{cases}$$

- (d) **(10 pontos)** Esboce a curva encontrada na folha milimetrada. Ela parece se encaixar com os dados?
- (e) **(15 pontos)** Determinações experimentais de periodicidade e frequência podem estar submetidas ao fenômeno de *aliasing*, que ocorre quando diversos ajustes periódicos são passíveis de regredir os dados coletados. Para cada um dos fenômenos a seguir, responda se o *aliasing* é uma preocupação prática.
- Determinação do período de um pulsar, utilizando um equipamento com coleta a cada 100 microssegundos.
  - Determinação do período de um pulsar, utilizando um equipamento com coleta a cada meio segundo.
  - Determinação do período de translação de um exoplaneta, utilizando o telescópio Kepler.
  - Determinação do período de uma variável cefeida, utilizando um telescópio terrestre com uma coleta a cada 12h.

**Solução:**

- (a) Imagem ao final
- (b) Segue a tabela com os valores calculados:

$t$	$y(t)$
0	1,000
24	-0,809
48	0,309
72	0,309
96	-0,809
120	1,000
144	-0,809

- (c) A aplicação do método dos mínimos quadrados gera o seguinte sistema:

$$\begin{cases} A \cdot \sum \cos^2 \left( \frac{2\pi}{60} \cdot t \right) + B \cdot \sum \cos \left( \frac{2\pi}{60} \cdot t \right) \cdot \sin \left( \frac{2\pi}{60} \cdot t \right) = \sum \cos \left( \frac{2\pi}{60} \cdot t \right) \cdot y \\ A \cdot \sum \cos \left( \frac{2\pi}{60} \cdot t \right) \cdot \sin \left( \frac{2\pi}{60} \cdot t \right) + B \cdot \sum \sin^2 \left( \frac{2\pi}{60} \cdot t \right) = \sum \sin \left( \frac{2\pi}{60} \cdot t \right) \cdot y \end{cases}$$

$$\Rightarrow \begin{cases} 4,154 \cdot A - 0,476 \cdot B = 4,154 \\ -0,476 \cdot A + 2,845 \cdot B = -0,476 \end{cases}$$

$$\Rightarrow \begin{cases} A = 1,000 \\ B = 0,000 \end{cases}$$

- (d) Imagem ao final
- (e) A preocupação com *aliasing* surge quando a frequência de amostragem é menor ou comparável à frequência de periodicidade do fenômeno. Sendo assim,

$$\begin{cases} f_{amostragem} \gg f_{fenomeno} \Rightarrow \text{Não} \\ f_{amostragem} < f_{fenomeno} \quad \text{ou} \quad f_{amostragem} \approx f_{fenomeno} \Rightarrow \text{Sim} \end{cases}$$

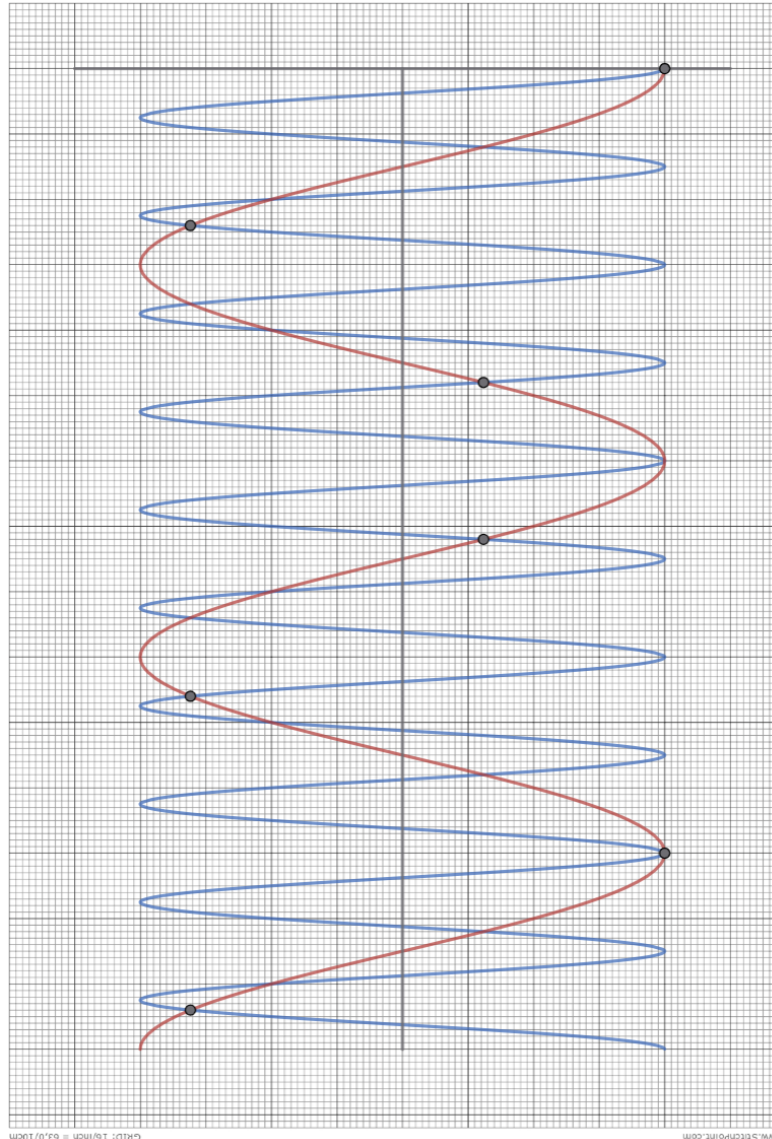
Normalmente, utiliza-se a regra de que a frequência de amostragem deve ser, pelo menos, o dobro (regra de Nyquist).

Sendo assim, uma comparação de ordem de grandeza responde a cada um dos itens:

- (i) Um pulsar dificilmente tem um período menor que um milissegundo. Sendo assim, 100 microssegundos é uma frequência suficientemente alta de amostragem. Portanto, não.
- (ii) Muitos pulsares têm períodos menores do que um segundo, sendo assim, não podemos garantir que a frequência de amostragem seja maior que a do fenômeno. Portanto, sim.
- (iii) A translação de um exoplaneta tem período de dias ou anos, enquanto o telescópio Kepler tem taxa de amostragem na ordem de minuto. A figura na questão 2 ilustra muito bem porque não precisamos nos preocupar com *aliasing* nesse caso. Portanto, não.

- (iv) Variáveis cefeidas costumam ter períodos de dias, o que é bem maior que o tempo entre as amostragens. Portanto, não.

A imagem a seguir representa a folha de papel milimetrado ao final da resolução. A linha em azul representa a curva do item a; a linha em vermelho, a do item d:



## 2. Exoplanetas (90 pontos)

O telescópio espacial Kepler foi lançado pela NASA em 2009 com intuito de descobrir exoplanetas pelo método de trânsito. Considere o estudo de uma estrela similar ao Sol, de magnitude 12 na banda K. O número de contagens na detecção nominal dessa estrela pode ser modelado por uma distribuição de Poisson com variância  $\sigma^2$ . Sabe-se que o telescópio Kepler é equipado com um espelho primário de 1,4 m de diâmetro.

- (a) (20 pontos) Suponha que essa estrela seja orbitada por um planeta similar à Terra. No evento de trânsito, a ocultação parcial da estrela gera uma queda  $\Delta N$  no número médio de contagens.

Para detectar esse planeta, precisamos que essa variação a seja comparável ou maior que o ruído. Determine  $\Delta N$  e o tempo de exposição  $\Delta t$  para que  $\sigma \approx \Delta N$ .

**Dados:**

- Raio do Sol:  $6,96 \cdot 10^8$  m
  - Raio da Terra:  $6,36 \cdot 10^6$  m
  - Comprimento de onda central da banda K:  $2,16 \mu\text{m}$
  - Constante de Planck:  $h = 6,62 \cdot 10^{-34}$  J.s
  - Velocidade da luz:  $c = 3,00 \cdot 10^8$  m
  - Fluxo de um objeto de magnitude zero na banda K:  $1,12 \cdot 10^{-10}$  W.m<sup>-2</sup>
- (b) **(20 pontos)** Em um dado momento, observou-se uma queda significativa no brilho da estrela. Durante 250 medições, o brilho se manteve, em média, 1000 partes por milhão abaixo do valor nominal. Estime a probabilidade de que a média de 250 medições seja menor ou igual a esse valor por mero ruído estatístico. Considere que a missão Kepler detectou exoplanetas em cerca de 0,5% das estrelas estudadas.
- (c) **(15 pontos)** A curva de luz a seguir remete a Kepler 52-c. Como podemos observar, a queda de luminosidade não demonstra periodicidade a longo prazo, o que nos impede de incluir esse sistema como um caso confirmado de exoplaneta. Falsos-positivos formaram cerca de 10% do total de positivos da missão Kepler. Considerando o item anterior, estime qual deveria ser a frequência esperada de falsos-positivos, e interprete brevemente seu resultado.

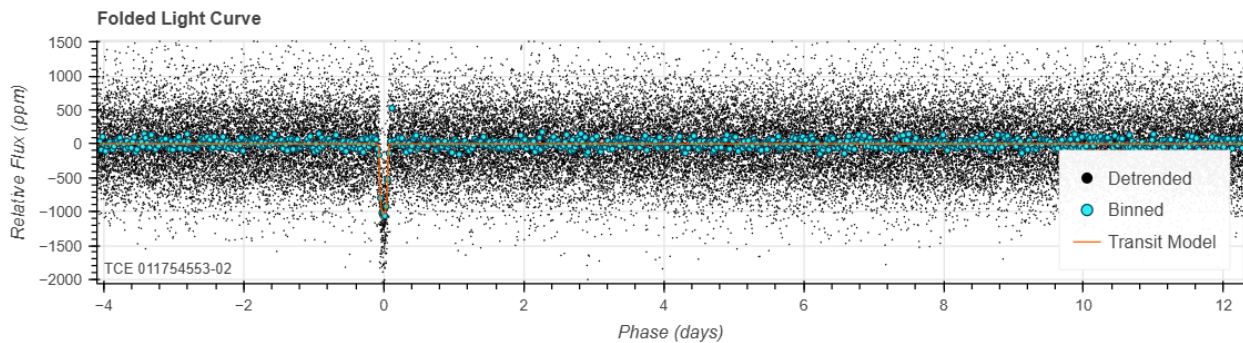


Figura 1: Curva de luz de Kepler 52-c. Imagem retirada do arquivo Mikulski de telescópios espaciais.

- (d) **(15 pontos)** Um dos problemas do método de trânsito é sua limitação ao estudo de sistemas *edge-on*. Sejam  $R_P$ ,  $R_S$  e  $a$  o raio de um exoplaneta, de sua estrela hospedeira, e seu semi-eixo maior, respectivamente, determine a probabilidade de que esse exoplaneta produza qualquer tipo de trânsito sobre sua estrela.
- (e) **(20 pontos)** O raio de um exoplaneta pode ser inferido da própria curva de luz. A estimativa da massa, no entanto, depende de outros métodos, como a análise do efeito Doppler da translação da estrela em torno do centro de massa comum. Entre os 5921 exoplanetas descobertos pelas missões Kepler, K2 e TESS, em 4424 foi possível realizar estimativa de raio, e em 1891, estimativa de massa. As tabelas abaixo mostram as respectivas frequências.

Raio	Frequência	Massa	Frequência
$R \leq 1,25 R_{\oplus}$	542	$M \leq 3 M_{\oplus}$	99
$1,25 < R \leq 2 R_{\oplus}$	1094	$3 < M \leq 10 M_{\oplus}$	315
$2 < R \leq 6 R_{\oplus}$	1894	$10 < M \leq 30 M_{\oplus}$	218
$6 < R \leq 15 R_{\oplus}$	681	$30 < M \leq 100 M_{\oplus}$	194
$15 R_{\oplus} < R$	213	$100 < M \leq 300 M_{\oplus}$	341
		$300 M_{\oplus} < M$	724

Construa dois histogramas com os dados relatados e explique sucintamente **(i)** como os resultados dos itens anteriores enviesam a distribuição de raios, e **(ii)** como se explica a aparente contradição entre dois picos de frequência para as massas e apenas um pico de frequência para os raios.

**Solução:**

(a) O fluxo da estrela pode ser calculado por:

$$F = F_0 \cdot 10^{-m/2,5}$$

$$F = 1,12 \cdot 10^{-10} \cdot 10^{-12/2,5} \text{ W.m}^{-2}$$

$$F = 1,78 \cdot 10^{-15} \text{ W.m}^{-2}$$

A potência coletada pelo telescópio:

$$P = 1,78 \cdot 10^{-15} \cdot \pi \cdot \left(\frac{1,4}{2}\right)^2$$

$$P = 2,73 \cdot 10^{-15} \text{ W}$$

A energia de um fóton é dada por:

$$E = \frac{hc}{\lambda}$$

$$E = 9,19 \cdot 10^{-20} \text{ J}$$

A quantidade de contagens, portanto, é:

$$N = \frac{P \cdot \Delta t}{E} = 30000 \cdot \Delta t$$

A variação na quantidade de fótons será:

$$\Delta N = -\frac{r_{\text{planeta}}^2}{r_{\text{estrela}}^2} \cdot N$$

$$\Delta N = -2,48 \cdot \Delta t$$

O desvio-padrão é dado por uma distribuição de Poisson:

$$\sigma = \sqrt{N} = 170 \cdot \sqrt{\Delta t}$$

A condição de contorno é  $\sigma = -\Delta N$ :

$$170 \cdot \sqrt{\Delta t} = 2,48 \cdot \Delta t$$

$$\Delta t = 4800 \text{ s}$$

$$\Delta t = 1,3 \text{ h} \Rightarrow \Delta N = 12000$$

Esse tempo revela uma dificuldade experimental extrema: não só a matriz CCD já estaria completamente saturada, como posteriormente, assumiremos 250 medições, cada uma teria que ter 1,3 h, e o trânsito, portanto, deveria durar mais de 300 h. Todavia, o trânsito da Terra sobre o Sol ocorre num ângulo tão pequeno da órbita que dura só meio dia. Ou seja, um sistema planetário como a Terra ao redor do Sol já seria dificilmente mensurável pela missão Kepler nessa magnitude. Para que o sistema fosse detectável dentro de 30 segundos de exposição, seria necessário que fosse 156 vezes mais luminoso. Ainda, perceba que o resultado assumiu que a luminosidade da banda  $K$  é igual à magnitude bolométrica, o que por si só subestima o resultado

- (b) Desvio padrão da média:

$$\bar{\sigma} = \frac{\sigma}{\sqrt{N}}$$

O escore- $z$  é calculado por:

$$z = \frac{\Delta N}{\bar{\sigma}} = -\sqrt{N} = -16$$

Percebemos que esse valor é tão alto, em módulo, que sequer aparece na tabela de escore- $z$ . Podemos afirmar que a probabilidade do evento é praticamente. À título de curiosidade, ele está na casa de  $10^{-57}$ .

- (c) A quantidade de falsos positivos por mero ruído estatístico pode ser calculada por Bayes:

$$P(\text{falso se positivo}) = \frac{P(\text{positivo se falso}) \cdot P(\text{falso})}{P(\text{positivo})}$$

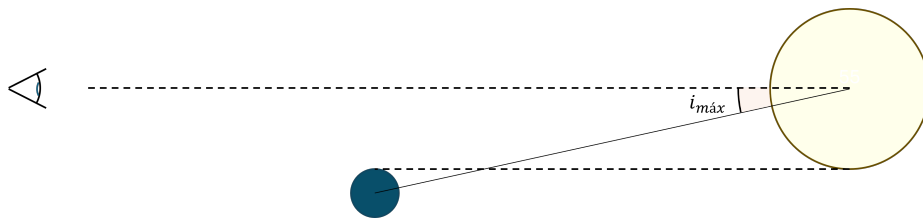
Em primeira aproximação, vamos desprezar os positivos que não foram nem confirmados nem rejeitados como verdadeiros. Nesse caso, se 10% dos positivos são falsos e 0,5% dos planetas são positivos verdadeiros, então a probabilidade de positivo é 0,556%.

A priori, a probabilidade de falso depende da proporção real de estrelas com exoplanetas. No entanto, não queremos acertar que uma estrela tem exoplanetas “no chute”. Sendo assim, estamos interessados não nos falsos totais, mas nos falsos dentro dos nossos limites de medição. Uma primeira estimativa seria a quantidade de estrelas que não apresentaram exoplanetas, i.e., 99,5%.

A probabilidade de apontar positivo se falso é virtualmente zero. Esse é o termo que domina a multiplicação. Como discutido anteriormente, existem alguns problemas com estarmos assumindo um sistema Terra-Sol, como a exposição muito mais alta que o real. No entanto, como mostra a imagem, os fenômenos de vale em casos de falso positivo são suficientemente bem determinados para descartarmos ruído estatístico.

Os falsos positivos observados ocorrem por culpa de fenômenos como a passagem de uma nuvem de gás ou poeira em frente à estrela, ou se as variações, ao invés de serem um exoplaneta, forem um sistema binário, por exemplo.

- (d) Para produzir algum tipo de trânsito, existe uma limitação no ângulo entre o observador e o planeta, visto a partir da estrela:



$$i_{max} = \text{sen}^{-1} \left( \frac{R_P + R_S}{a} \right)$$

A região em que pode estar o momento angular da órbita, portanto, é uma pequena faixa em volta do plano perpendicular à linha de visada.

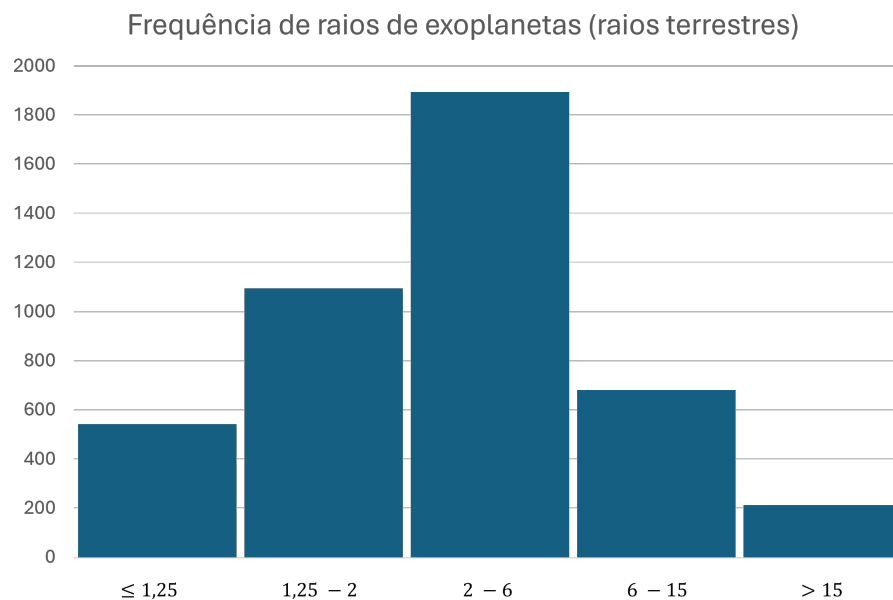


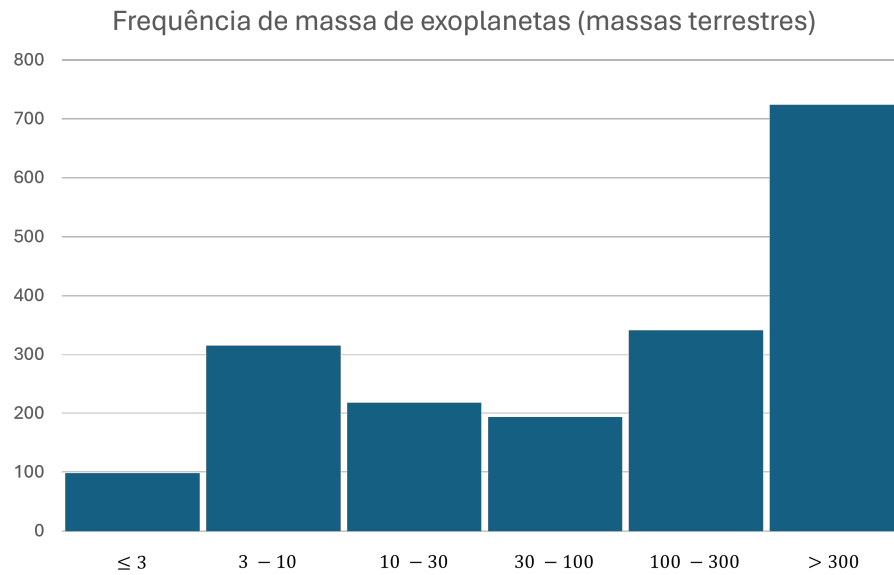
Sendo assim, por isotropia, a probabilidade é calculada pelo ângulo sólido:

$$P = \frac{\Omega}{4\pi} = \text{sen}(i_{max}) = \frac{R_P + R_S}{a}$$

Isso significa que o método tende a encontrar exoplanetas próximos de sua estrela hospedeira.

(e) Seguem os histogramas solicitados:





(i) Por um lado, o método tende a favorecer planetas de raios maiores, pois têm mais chance de ocultar sua estrela hospedeira, e quando o fazem, causam uma variação maior no número de contagens. Por outro lado, ao favorecer a detecção de planetas próximos à estrela, pois são os com maior probabilidade de trânsito, o método também tem um pequeno favorecimento de planetas rochosos e de raio menor, que costumam ser mais internos. No geral, esses dois vieses não se balanceiam, e o método favorece planetas de raios maiores.

(ii) O pico de massas menores é provavelmente o que melhor representa o pico estatístico. O pico de massas maiores é um provável fruto do viés de medição, que favorece planetas de alta massa, pois são os que podem gerar redshifts consideráveis em suas estrelas hospedeiras.