



UNIVERSIDADE DE SÃO PAULO
Faculdade de Zootecnia e Engenharia de Alimentos

ZAB1111 – Estatística Básica

Prof. César Gonçalves de Lima cegdlima@usp.br

Aula 1

Noções de estatística descritiva

Gráficos e tabelas

1. NOÇÕES DE ESTATÍSTICA DESCRITIVA

1.1. Introdução

Em uma pesquisa, o pesquisador estará interessado em **analisar** e **entender** um conjunto de dados importante ao seu particular objeto de estudos.

O pesquisador precisará **resumir** os seus dados para que estes sejam **mais informativos** e possa, posteriormente,

- **Comparar** os seus resultados com outros já obtidos.
- **Verificar a adequação** desses dados a algum modelo teórico.

Objetivo:

Estudar técnicas usadas para **resumir** um conjunto de dados (**Análise Exploratória dos Dados**) que indiquem possíveis modelos a serem utilizados numa fase final, chamada **Inferência Estatística**.

Análise Exploratória dos Dados:

- Gráficos e tabelas.
- Medidas de tendência central (média, moda e mediana).
- Medidas de dispersão (erro médio, variância, desvio padrão e coeficiente de variação).
- Medidas de achatamento e simetria.

1.2. Definições iniciais

- **População (ou universo):** conjunto de indivíduos ou objetos sobre os quais desejamos desenvolver algum estudo e que têm alguma característica em comum. Pode ser finita ou infinita.
- **Amostra:** qualquer subconjunto de elementos retirados de uma população.

Amostra representativa da população (situação ideal) \Rightarrow as conclusões tiradas a partir desta amostra podem ser inferidas (expandidas, induzidas) para toda a população.

A cada elemento investigado associamos os resultados de uma ou mais **variáveis**.

As variáveis podem ser: **qualitativas** ou **quantitativas**.

- **Variável qualitativa:** apresenta como possíveis realizações um **atributo** ou **qualidade** do indivíduo pesquisado. Podem ser classificadas de acordo com a sua **escala de medida**, como por exemplo:
 - **Escala nominal:** não permite uma ordenação dos seus resultados.
Exemplo: Sexo, raça, cidade de origem, cor da pele etc.
 - **Escala ordinal:** permite uma ordenação dos seus resultados.
Exemplo: Conceito final em disciplina de pós-graduação, nota em prova de degustação, classe social etc.

- **Variável quantitativa:** quando suas realizações são expressas em números resultantes de uma **contagem** ou de uma **mensuração**. Podem ser classificadas como:

- **Discreta:** quando as suas realizações formam um conjunto finito ou enumerável de valores sendo, frequentemente, resultantes de uma contagem.

Exemplo: Idade (anos), número de filhos, número de acessos à internet por dia, número de compras à prazo etc.

- **Contínua:** quando os seus possíveis valores formam um intervalo de números reais resultantes, geralmente, de uma mensuração.

Exemplo: Peso, altura, salário etc.

Importante: Existem ferramentas estatísticas específicas para tratar os dados de diferentes tipos de variáveis.

Notação: As variáveis são denotadas por letras latinas maiúsculas e os seus resultados, por letras latinas minúsculas com ou sem índices.

Exemplo: Os pesos (X) de seis frangos ao abate podem ser escritos, genericamente, como:

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6\}.$$

Esta notação é bastante utilizada quando precisamos escrever as fórmulas de cálculo de **estatísticas** comuns, como por exemplo:

$$\text{Média: } \bar{x} = \frac{1}{6} \sum_{i=1}^6 x_i \quad \text{Desvio padrão: } s = \sqrt{\frac{1}{5} \sum_{i=1}^6 (x_i - \bar{x})^2}$$

Exemplo 1. Seja a variável X que descreve o peso corporal (gramas) de frangos de corte aos 42 dias de vida e que está assumindo os seguintes valores para uma amostra de $n = 6$ frangos:

$$X = \{1900, 2050, 1950, 2100, 1950, 2050\}$$

Então:

- $\sum_{i=1}^6 x_i = 1900 + \dots + 2050 = 12000$ gramas é o peso total dos 6 frangos
- $\bar{x} = \frac{\sum_{i=1}^n x_i}{6} = \frac{12000}{6} = 2000$ gramas é o peso médio dos frangos
- $s = \sqrt{\frac{30000}{5}} = 77,5$ gramas é o desvio padrão amostral dos pesos

PROPRIEDADES IMPORTANTES DO SOMATÓRIO

1) Se $k \in \mathbb{R}$ (k é uma constante real) então $\sum_{i=1}^n k = nk$

2) $\sum_{i=1}^n kx_i = kx_1 + kx_2 + \dots + kx_n = k(x_1 + x_2 + \dots + x_n) = k \sum_{i=1}^n x_i$

3) $\sum_{i=1}^n (x_i - k) = (x_1 - k) + (x_2 - k) + \dots + (x_n - k) = \sum_{i=1}^n x_i - nk$

Se $k = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \Rightarrow \sum_{i=1}^n (x_i - \bar{x}) = 0$, ou seja, a soma dos desvios dos dados em relação à média aritmética é nula [Prove!!!]

4) $\sum_{i=1}^n x_i^2 = x_1^2 + x_2^2 + \dots + x_n^2$ *(Soma de quadrados)*

5) $(\sum_{i=1}^n x_i)^2 = (x_1 + x_2 + \dots + x_n)^2$ *(Quadrado da soma)*

$$\begin{aligned}6) \sum_{i=1}^n (x_i - k)^2 &= (x_1 - k)^2 + \cdots + (x_n - k)^2 \\ &= \sum_{i=1}^n x_i^2 - 2k \sum_{i=1}^n x_i + nk^2\end{aligned}$$

(Soma dos quadrados dos desvios em relação a uma constante k)

Caso particular: Se $k = \bar{x} \Rightarrow s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - k)^2$ é chamado de variância amostral.

$$7) \sum_{i=1}^n x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

(Soma de duplos produtos)

1.3. Estatística descritiva

Etapas importantes do planejamento de uma pesquisa:

- Definição do problema a ser estudado.
- Estabelecimento da forma pela qual os dados serão coletados.
- Definição de um cronograma das atividades e dos custos envolvidos (planejamento).
- Exame das informações disponíveis sobre o assunto.
- Delineamento da amostra.
- Coleta de dados.
- Revisão crítica dos dados, suprimindo os valores reconhecidamente estranhos ao levantamento.

- Organização dos dados de maneira prática e racional (tabelas e gráficos) para o melhor entendimento do fenômeno estudado.
- Escolha da melhor técnica de análise estatística dos dados.
- Realização da análise, com a verificação dos seus pressupostos.
- Elaboração de um relatório com os principais resultados, incluindo tabelas e gráficos que auxiliem na sua apresentação e discussão.

Após a coleta dos dados, iniciamos uma análise exploratória visando conhecer mais sobre algumas características importantes das respostas.

A seguir conheceremos algumas técnicas simples usadas para **resumir** os dados, como:

- Apresentação dos dados em tabelas e em gráficos
- Cálculo de medidas de posição, de dispersão, de assimetria e de curtose.

1.3.1. APRESENTAÇÃO TABULAR DOS DADOS

Tabela 1. Evolução do quadro de docentes na USP por categoria - 1988/1996.

Ano	Categoria docente (*)					Total
	MS-1	MS-2	MS-3	MS-5	MS-6	
1988	614	1.253	2.140	855	527	5.389
1989	593	1.271	2.267	864	550	5.545
1990	516	1.258	2.347	828	630	5.579
1991	383	1.191	2.392	817	637	5.420
1992	320	1.129	2.396	782	682	5.309
1993	254	1.047	2.484	788	703	5.276
1994	175	947	2.553	797	720	5.192
1995	113	804	2.522	831	673	4.943
1996	107	771	2.510	804	664	4.856

*) MS-1: Auxiliar de Ensino; MS-2: Assistente; MS-3: Doutor; MS-5: Associado; MS-6: Titular.

Fonte: Anuário Estatístico da USP – 1996

Tabela 2. Evolução do quadro de docentes da FZEA por categoria – 2010/2021

Ano	Categoria docente (*)			Total
	MS-3	MS-5	MS-6	
2010	56	30	10	96
2011	60	31	11	102
2012	65	35	11	111
2013	61	40	10	111
2014	59	41	10	110
2015	53	47	10	110
2016	50	50	9	109
2017	46	53	10	109
2018	49	49	13	111
2019	45	49	13	107
2020	45	49	13	107
2021	35	55	13	103

☉MS-3: Doutor; MS-5: Associado; MS-6: Titular

Fonte: Anuários Estatísticos da USP – 2013, 2018, 2022

As tabelas são constituídas, basicamente, de:

- **Corpo:** é a parte estrutural da tabela na qual se inscrevem os dados, sendo constituído de linhas e colunas onde cada cruzamento linha-coluna é chamado de *casela* ou *célula*.
- **Título:** é a definição clara e concisa da natureza do fato estatístico, o espaço geográfico abrangido e a referência no tempo a que fato corresponde.
- **Cabeçalho:** especifica a informação apresentada em cada coluna.
- **Coluna indicadora:** identifica o tipo de informação contida em cada linha.

- **Fonte:** é a identificação da entidade responsável pelo levantamento dos dados, registros estatísticos ou por séries estatísticas derivadas. Evita fraudes e plágios.
- **Notas:** são informações de natureza geral, destinadas a conceituar ou esclarecer o conteúdo das tabelas, a indicar o critério adotado no levantamento ou o método de elaboração das estatísticas derivadas.
- **Chamadas:** são informações específicas relativas a uma parte da tabela, destinadas a complementar as informações sobre os dados numéricos.

Importante: Toda tabela deve ter título, cabeçalho e corpo.

Título

Coluna indicadora

Tabela 1. Evolução do quadro de docentes na USP por categoria - 1988/ 1996.

Ano	Categoria docente (*)						Total
	MS-1	MS-2	MS-3	MS-5	MS-6		
1988	614	1.253	2.140	855	527	5.389	
1989	593	1.271	2.267	864	550	5.545	
1990	516	1.258	2.347	828	630	5.579	
1991	383	1.191	2.392	817	637	5.420	
1992	320	1.129	2.396	782	682	5.309	
1993	254	1.047	2.484	788	703	5.276	
1994	175	947	2.553	797	720	5.192	
1995	113	804	2.522	831	673	4.943	
1996	107	771	2.510	804	664	4.856	

Cabeçalho

Corpo

*) MS-1: Auxiliar de Ensino; MS-2: Assistente; MS-3: Doutor; MS-5: Associado; MS-6: Titular.

Fonte: Anuário Estatístico da USP - 1996.

Fonte

Nota

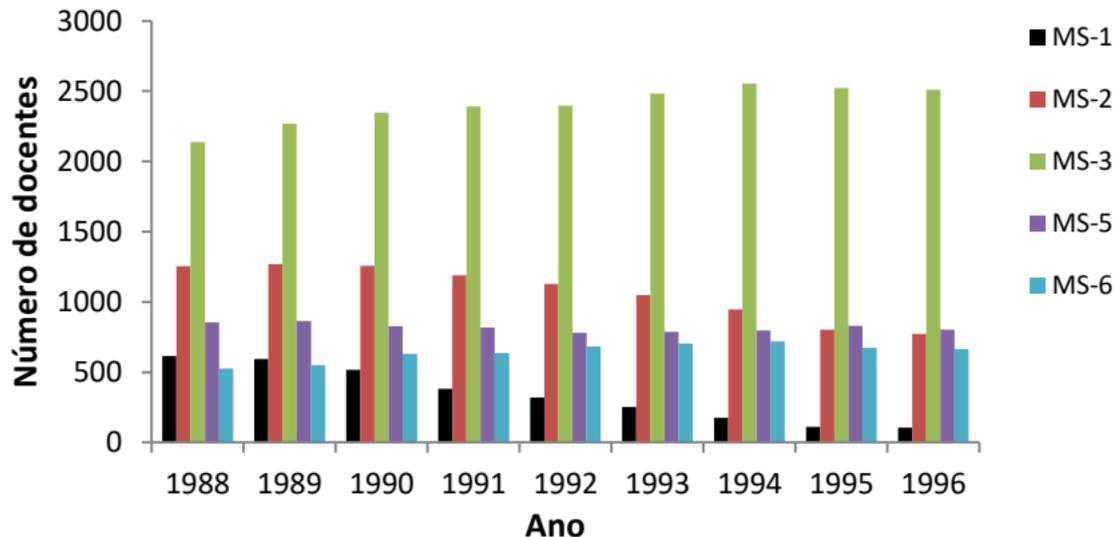
1.3.2. APRESENTAÇÃO GRÁFICA DOS DADOS

Finalidade: Dar uma ideia, a mais imediata possível, dos resultados obtidos numa pesquisa, permitindo chegar-se a conclusões sobre a evolução do fenômeno ou como se relacionam os valores da série.

A escolha do tipo de gráfico mais apropriado para representar uma série estatística ficará a critério do pesquisador.

São aspectos imprescindíveis na construção de um gráfico: a *simplicidade*, a *clareza* e a *veracidade*.

A seguir apresentaremos algumas regras básicas de construção de gráficos simples e situações onde podem/devem ser utilizados.



MS-1: Auxiliar de Ensino; MS-2: Assistente; MS-3: Doutor; MS-5: Associado; MS-6: Titular.

Fonte: Anuário Estatístico da USP – 1996.

Figura 1. Evolução do quadro de docentes da USP nas diversas categorias, de 1988 a 1996.

Lembrar sempre que:

- a) O gráfico deve ter **título** e **escala** para que possa ser interpretado sem que haja necessidade de esclarecimentos adicionais no texto.
- b) No eixo das abscissas a escala cresce da esquerda para a direita e é escrita sob o eixo; no eixo das ordenadas a escala cresce de baixo para cima e é escrita à esquerda do eixo.
- c) Podem ser feitas setas para indicar a direção dos eixos.
- d) As variáveis representadas em cada eixo devem ser claramente identificadas, bem como, suas unidades de medidas.
- e) Quando os valores iniciais dos dados são muito altos, deve ser feita uma interrupção no eixo, com indicação clara da posição da origem.
- f) As **legendas** explicativas devem ser colocadas à direita do gráfico.
- g) Indicar sempre a **fonte** de onde os dados foram obtidos.

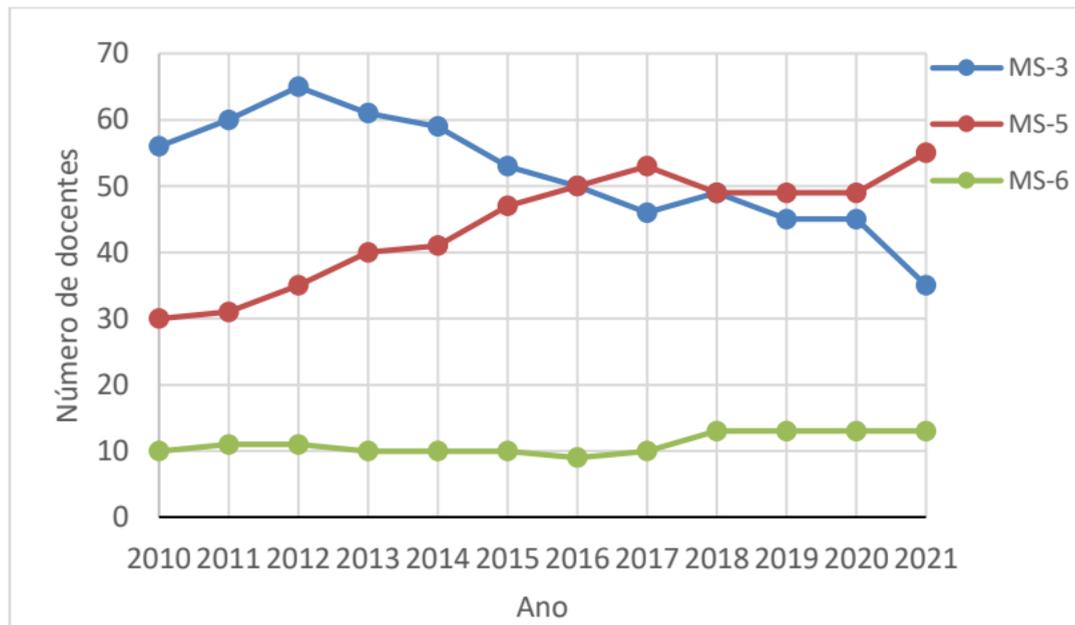
Gráficos mais comuns

Gráfico de Linhas: usado, principalmente, para apresentar séries cronológicas.

Gráfico de Barras ou Colunas: usado para apresentar séries cronológicas, geográficas e categóricas ou classificatórias. As barras podem ser construídas na posição horizontal quando as categorias são identificadas por nomes muito extensos.

Gráficos de Áreas ou de Setores: são usados para comparar proporções ou evidenciar a composição percentual de uma parte dos dados.

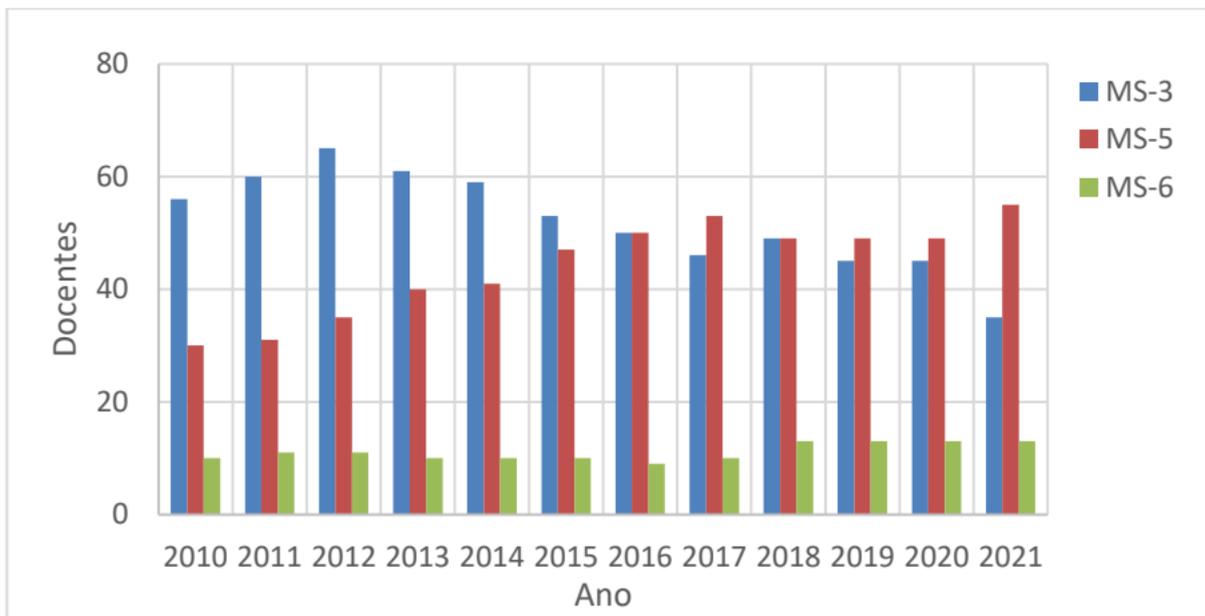
O Gráfico Polar (ou Radar) é usado para representar dados que variam ao longo de um intervalo de tempo limitado.



☛ MS-3: Doutor; MS-5: Associado; MS-6: Titular.

Fonte: Anuário Estatístico da USP – 1996.

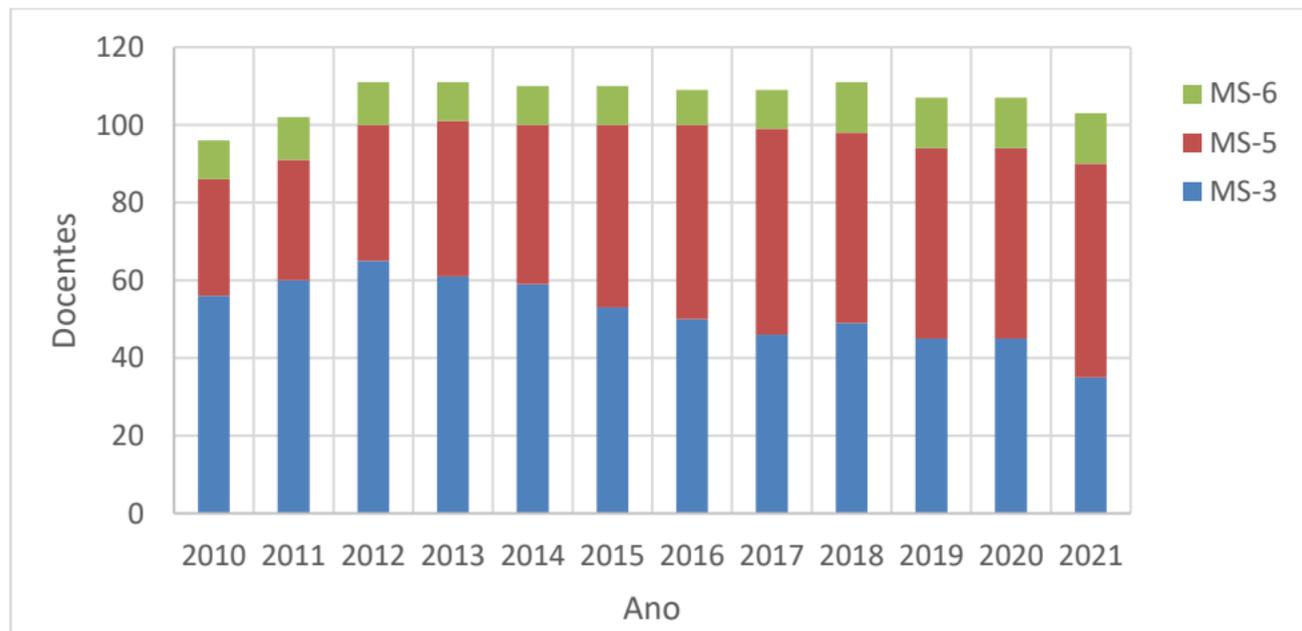
Figura 2. Evolução do quadro de docentes da FZEA por categoria, de 2010 a 2021.



♥ MS-3: Doutor; MS-5: Associado; MS-6: Titular.

Fonte: Anuário Estatístico da USP – 1996.

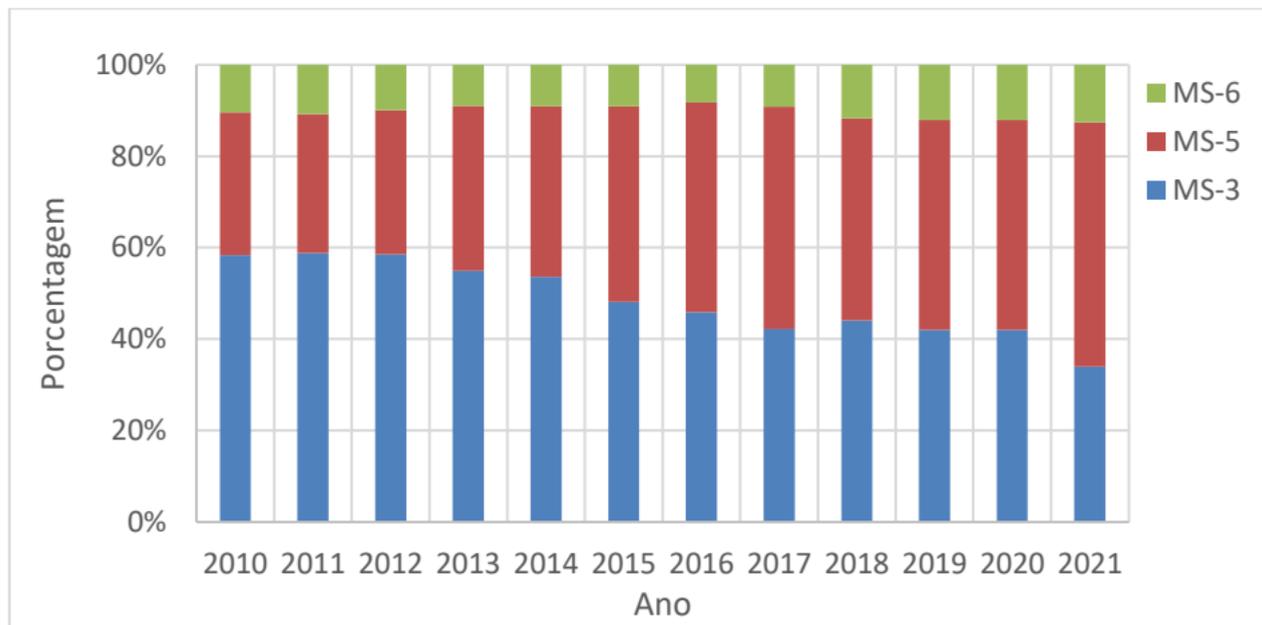
Figura 3. Evolução do quadro de docentes da FZEA por categoria, de 2010 a 2021.



♥ MS-3: Doutor; MS-5: Associado; MS-6: Titular.

Fonte: Anuário Estatístico da USP – 1996.

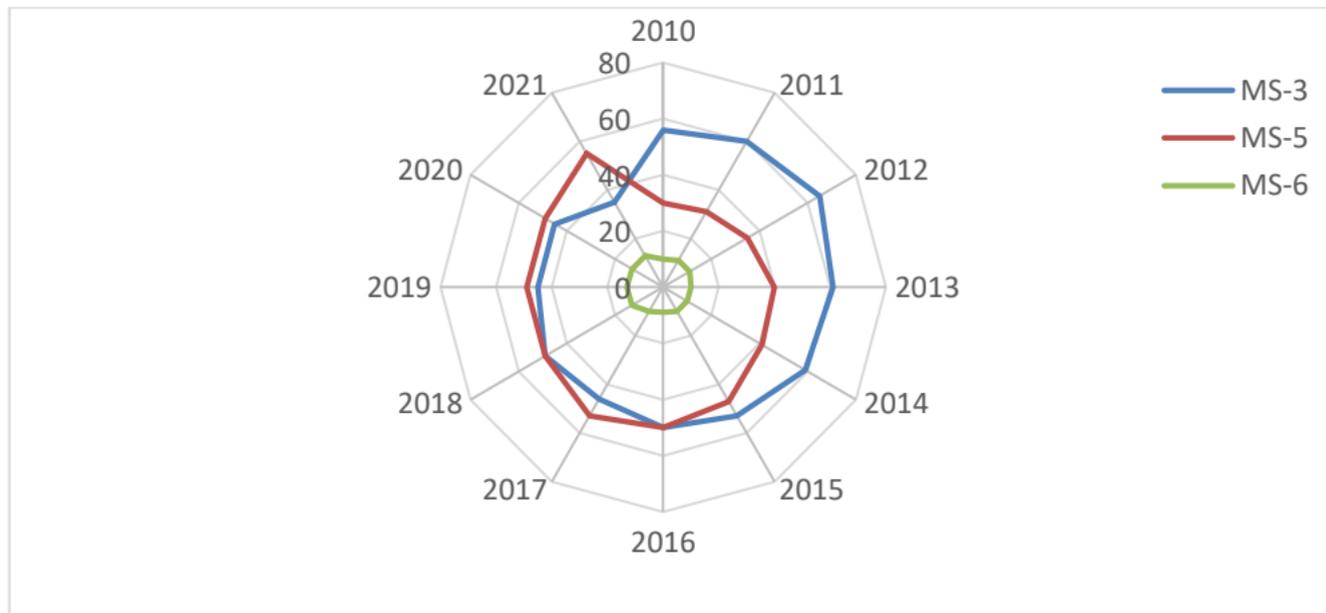
Figura 4. Evolução do quadro de docentes da FZEA por categoria, de 2010 a 2021.



☺ MS-3: Doutor; MS-5: Associado; MS-6: Titular.

Fonte: Anuário Estatístico da USP – 1996.

Figura 5. Evolução do quadro de docentes da FZEA por categoria, de 2010 a 2021.



♥ MS-3: Doutor; MS-5: Associado; MS-6: Titular.

Fonte: Anuário Estatístico da USP – 1996.

Figura 6. Evolução do quadro de docentes da FZEA por categoria, de 2010 a 2021.