

Perfil dos estudantes do Brasil em 2022 - Análise entre cursos presenciais e a distância

Alunos:

Alexandre Alves

Heitor Dias

João Capelasso

Matheus Fernando

Professora Melina Garcia



Introdução

EAD cresce mais de 420% nos últimos 10 anos

Por **Papo de Primeira**

15/08/22 às 15H00 atualizado em 15/08/22 às 15H00



Inep), o Brasil fechou o ano com quase 9 milhões de matrículas em 2.574 instituições de ensino superior. Do total de matrículas, 6,9 milhões (76,8%) pertenciam à rede privada. Considerando somente o volume de ingressantes, seis em cada dez novos alunos do ensino superior (62,8%) já optaram por cursos EaD. Se o recorte considerar apenas a adesão no ensino privado, a taxa de preferência é de sete a cada dez estudantes.

Evasão e Educação Superior: 2,3 milhões abandonaram curso superior em 2021



Ingressos e Conclusões na Educação Superior Pública e Privada



■ Ingressantes Pública ■ Concluintes Pública ■ Ingressantes Privada ■ Concluintes Privada

Elaboração do SoU_Ciência com dados das Sinopses dos Censos da Educação Superior do MEC, dos anos de 2017 a 2021, para matrículas totais (cursos presenciais e a distância).

Objetivos

- Entender a demografia dos estudantes da modalidade **EAD/Presencial**.
- Compreender a adesão ao ensino superior no ano pós-pandemia (2022).
- Embasado na análise de dados fazer possíveis sugestões as modalidades de ensino.

Coleta e organização de dados

Coleta de dados

- **INEP** (Instituto Nacional de Estudos e Pesquisas Educacionais).
- Proveniente do Censo da Educação Superior de 2022.
- Utilizamos a base de dados referente a todos os cursos ministrados em território nacional e regulamentados pelo MEC, totalizando **573.019 cursos(linhas)**.



Logística de tratamento de dados

- Dado que o Excel não suporta grandes bases de dados, (**573.019 linhas**) o tratamento inicial foi feito em **Python**.
- Em Python foi feita a organização, filtragem e posteriormente os gráficos.
- Em um segundo momento os dados agrupados foram importados para o Excel onde foram feitas as planilhas e os cálculos.

Limitações e Escolhas

Apesar da animação inicial de possuir uma base de dados tão completa para trabalhar, o fato é que parte significativa das linhas possuíam valores 0 ou nulos. A inclusão dos mesmos causaria um **ruído estatístico** significativo, o qual iria prejudicar a qualidade dos cálculos e a análise dos resultados.

Uma decisão teve de ser tomada e para preservar a coesão de dados ,seleccionamos apenas os cursos que possuíam pelo menos **1 Interessante, 1 Matriculado e 1 Concluente**.

Tal decisão levou o total de dados iniciais de **573.019** para **1537**.

Calculo de Amostra

$$N = (1,96^2 * 0,5 (1 - 0.5)) / 0,05^2$$

$$N = 3,8416 * 0,25 / 0,0025$$

$$N = 384,16$$

$$n = 384,16 / 1 + ((384,16 - 1/573.019)$$

$$n = 384,16 / 1,000669$$

$$n \approx 384$$

Dito isso, dado que possuímos uma amostra de **1517** dados estatisticamente confiáveis, iremos utilizar todos. Uma maior amostra aumenta a precisão o que consequentemente **diminui a margem de erro**.

Organização de Variáveis

- Entre as variáveis disponíveis, separamos em 3 grupos principais:

Ingressantes (Alunos que acabaram de ingressar no curso)

Matriculados (Alunos ingressantes + 'veteranos')

Concluintes (Alunos que acabaram o curso)

- Destes 3 grupos, subdividimos entre 2 principais características:

Idade e Gênero.

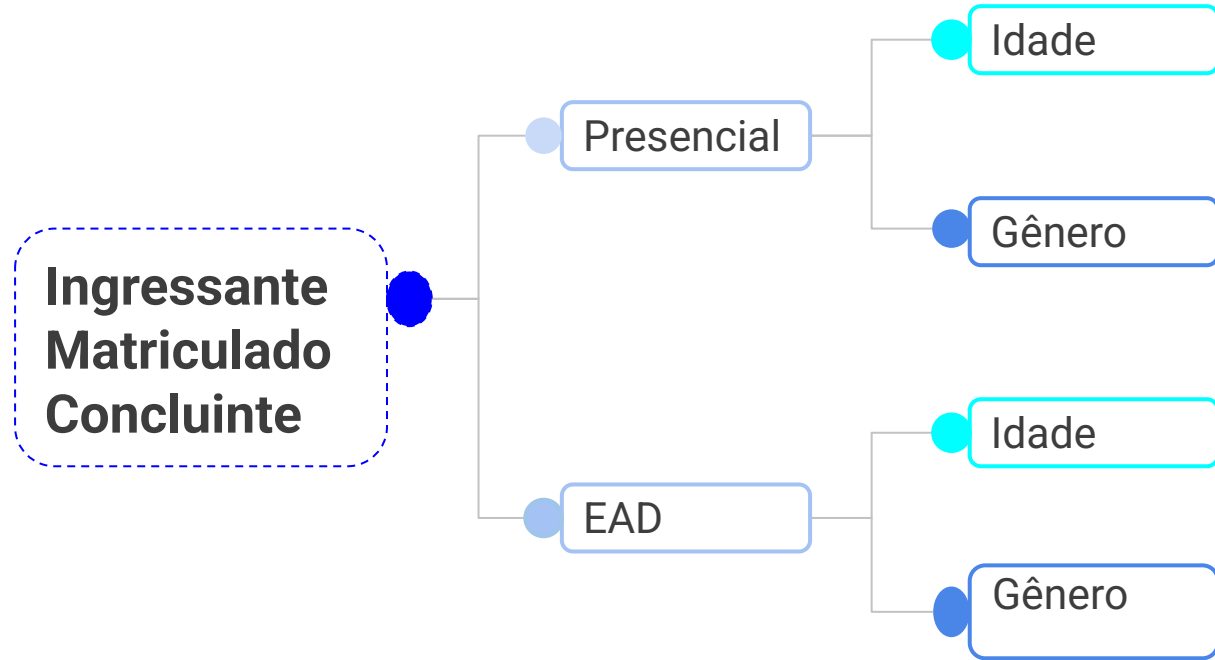
- Todas as 31 variáveis são do tipo quantitativa discreta, pois são valores inteiros, e representam a quantidade de alunos pertencente a cada característica, daquele curso:

Exemplo : **TP_MODALIDADE_ENSINO** : Indica se o curso é presencial (1) ou EAD (2).

Exemplo : **QT_ING_18_24** : Quantidade de ingressantes daquele curso, de faixa etária 18-24

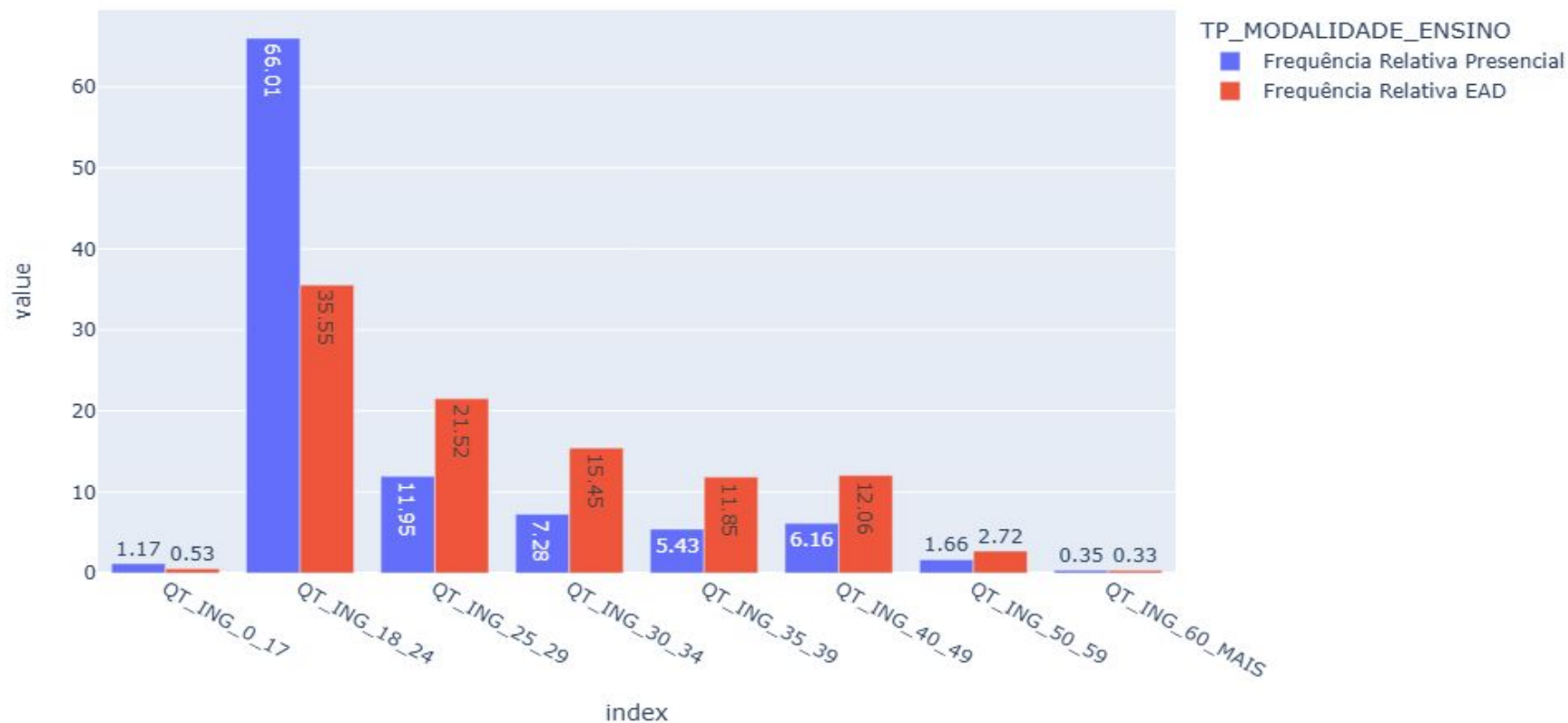
Exemplo : **QT_CONC_MASC** : Quantidade de concluintes daquele curso, do sexo masculino

Diagrama de Organização dos dados

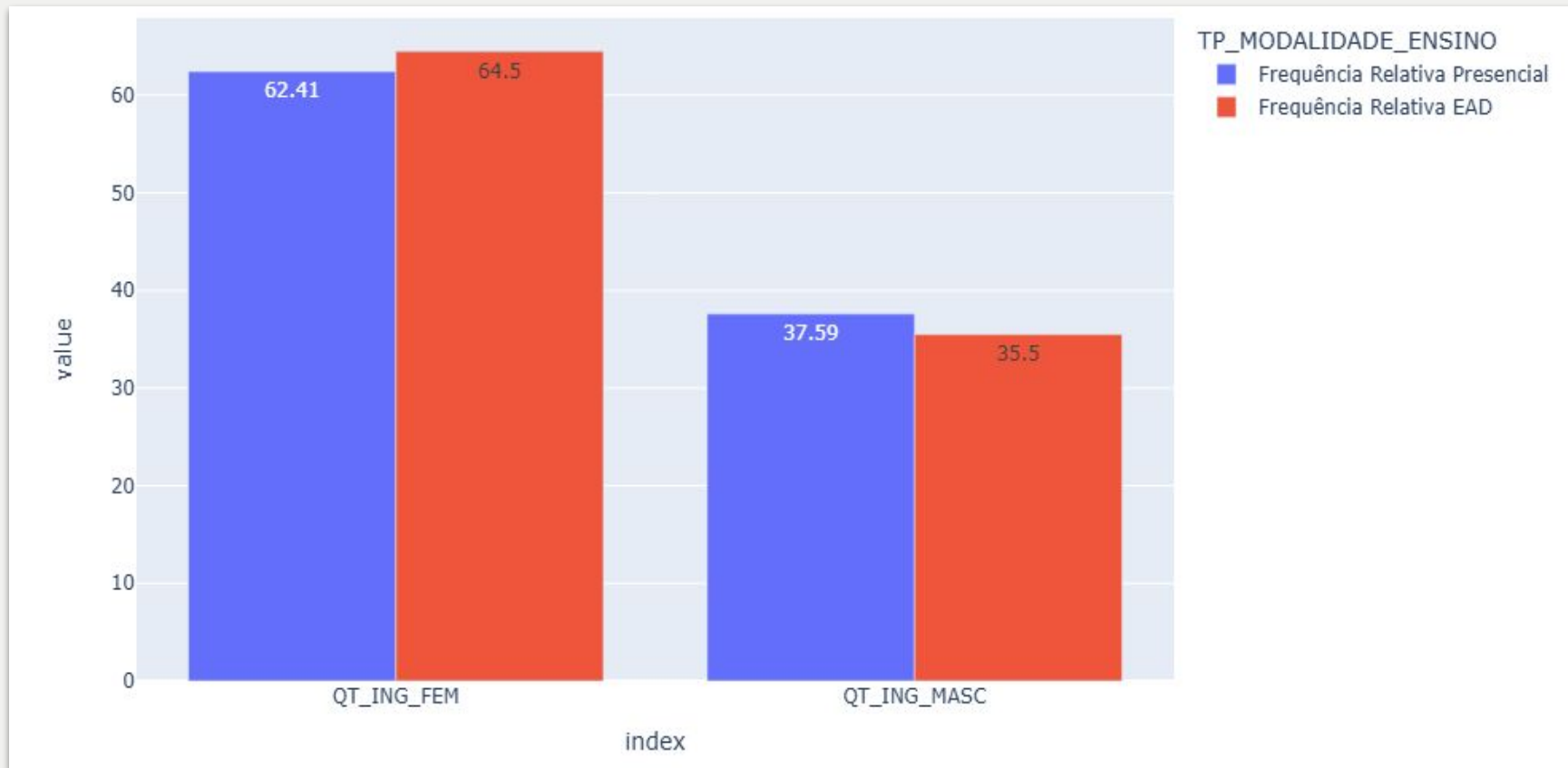


Apresentação de Dados - Graficos

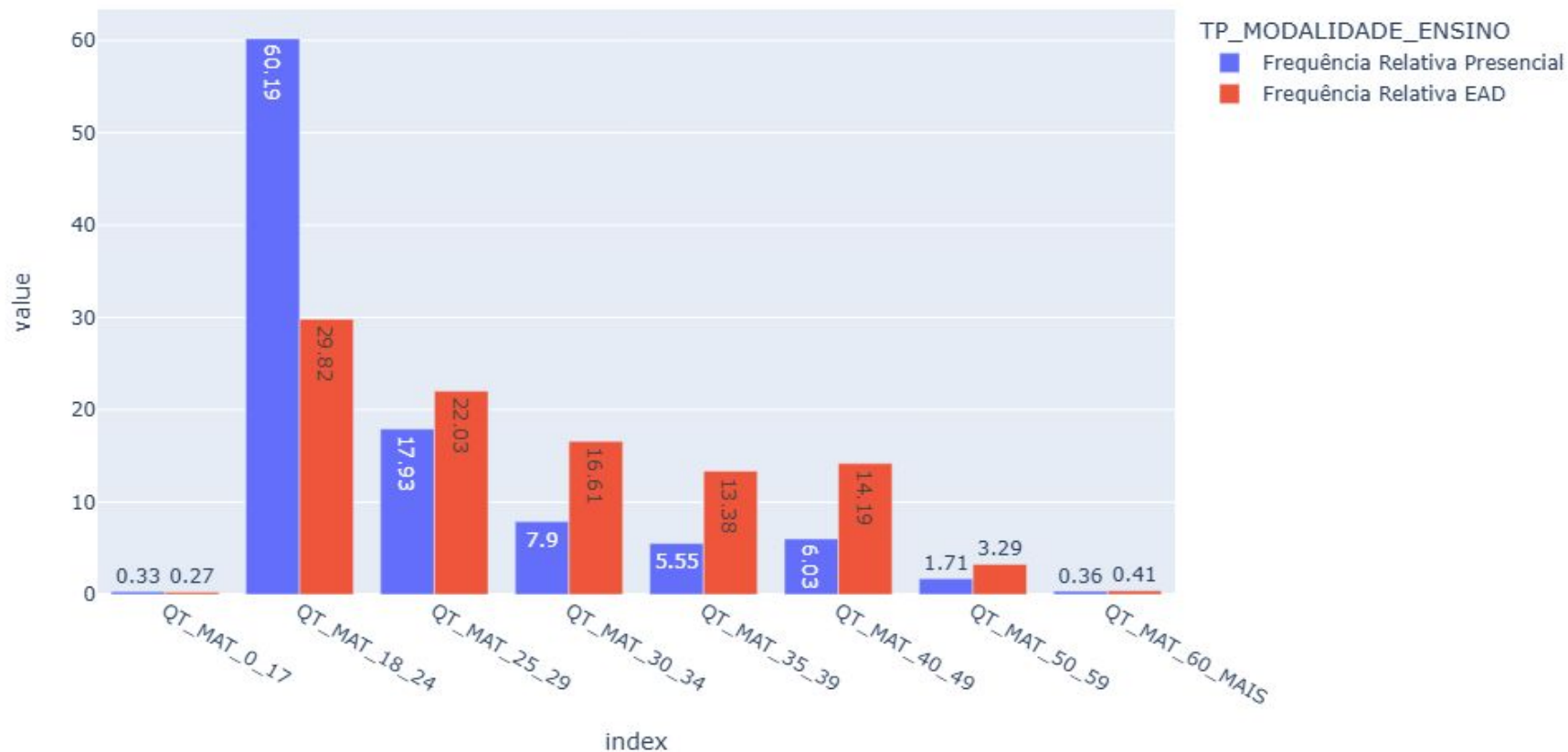
Percentual de Ingressantes por faixa etaria - Presencial x EAD



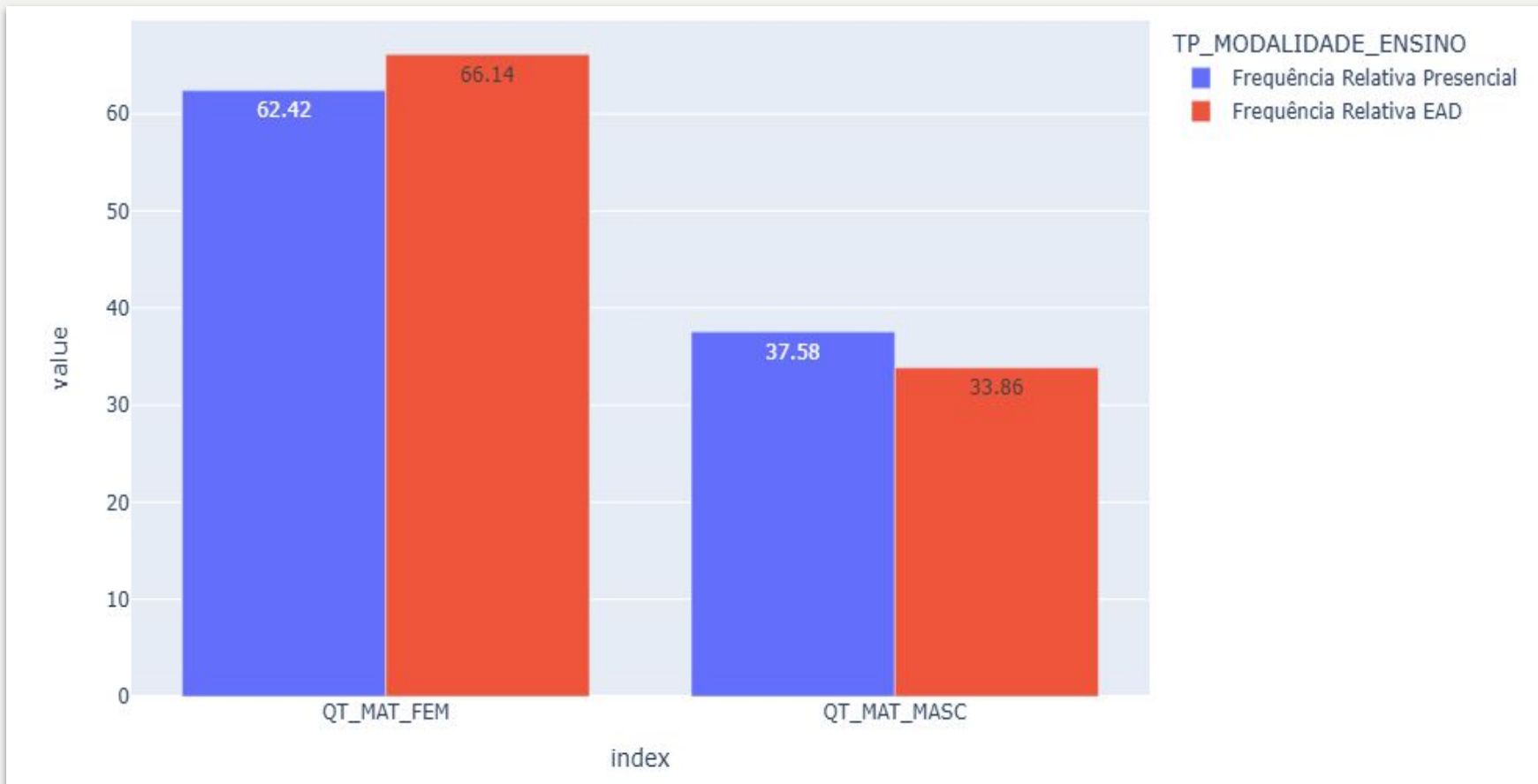
Percentual de Ingressantes por Genero - Presencial x EAD



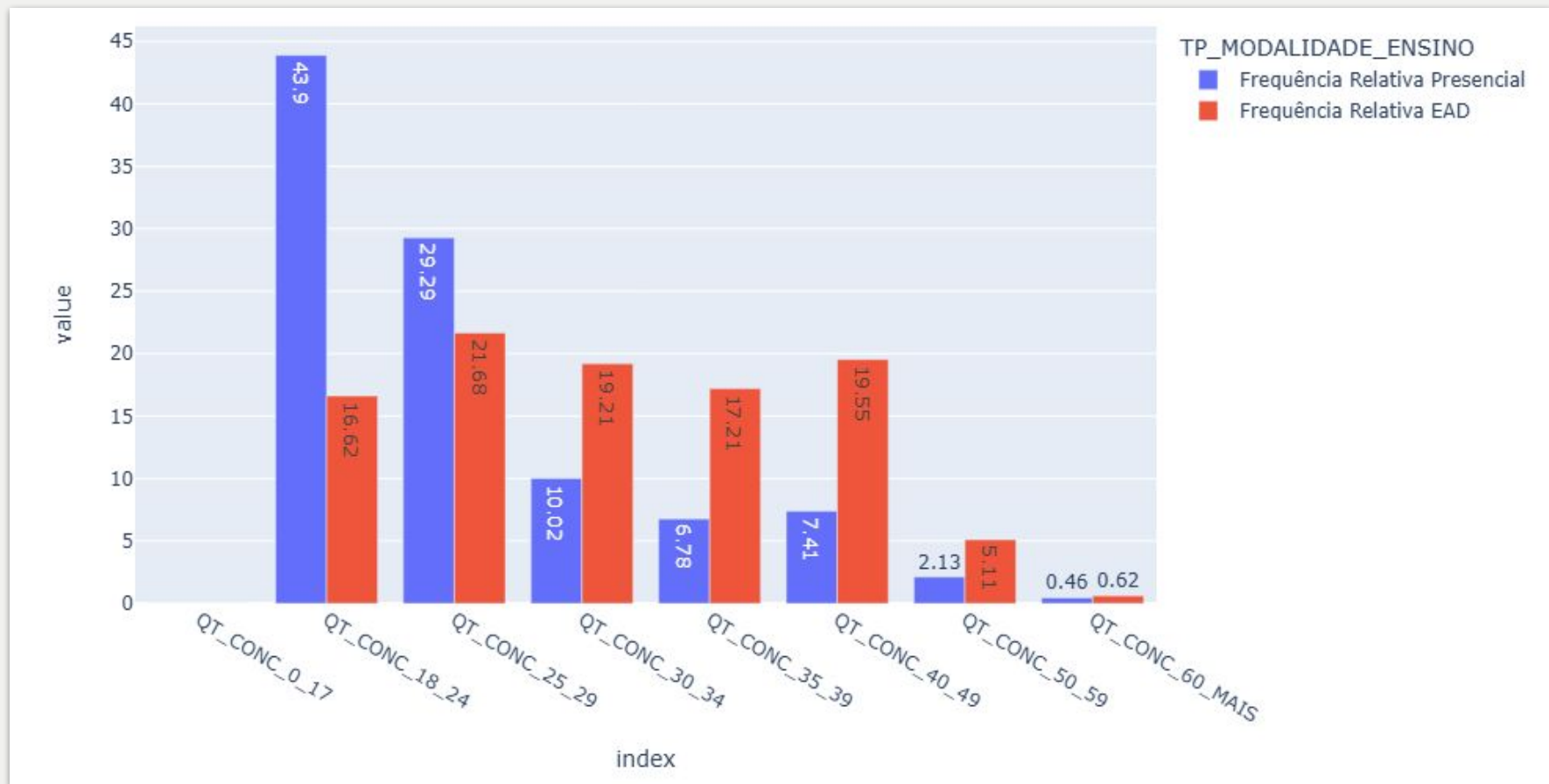
Percentual de Matriculados por faixa etária - Presencial x EAD



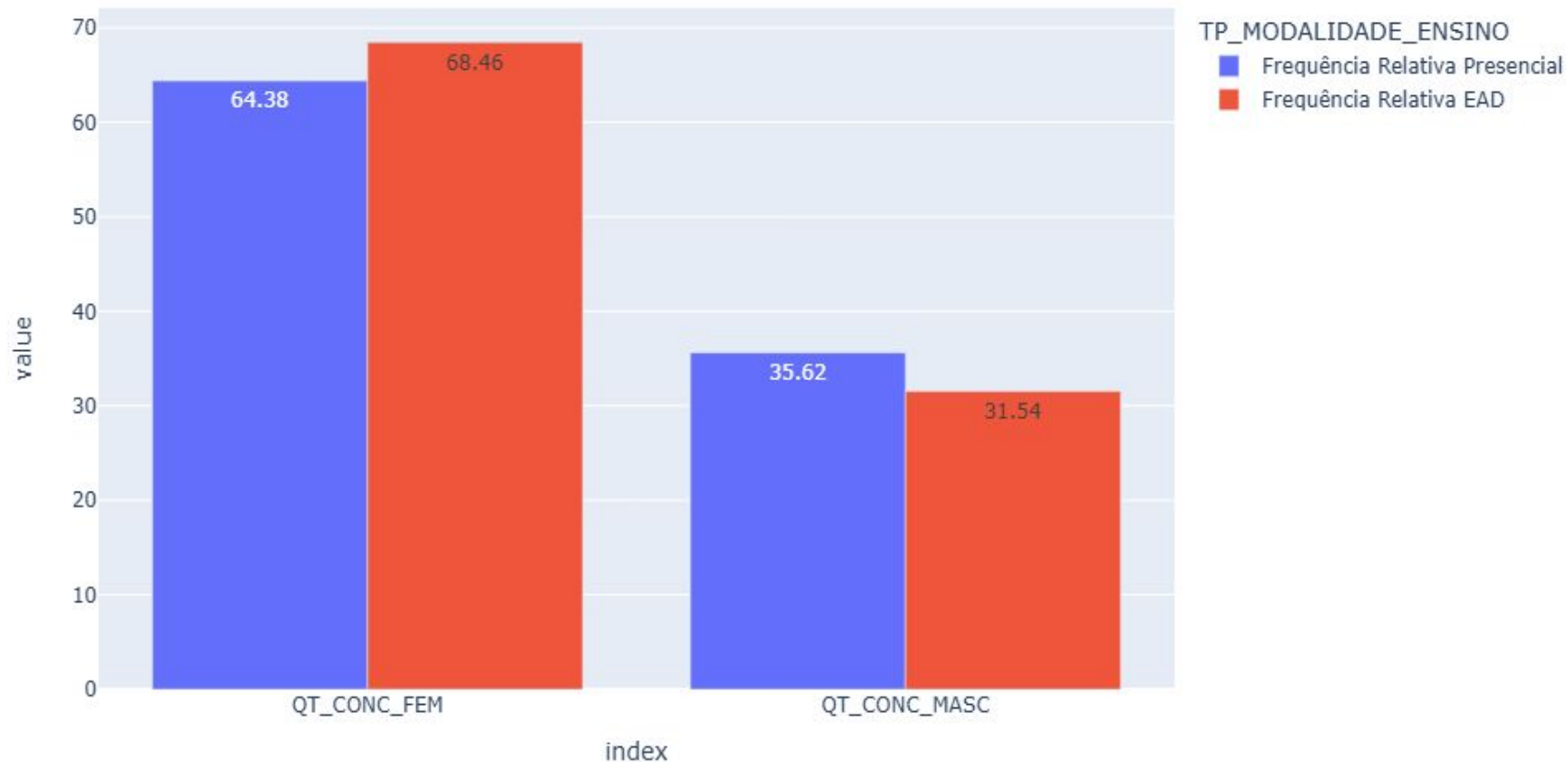
Percentual de Matriculados por Genero - Presencial x EAD



Percentual de Concluintes por faixa etária - Presencial x EAD



Percentual de Concluintes por sexo - Presencial x EAD



Apresentação de Dados - Calculos

Análise de dados

Fórmula utilizada para achar as modas:

Fórmula de Czuber

Procedimento:

- Identifica-se a classe modal (aquela que possuir maior frequência) – CLASSE(Mo).
- Utiliza-se a fórmula:

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$$

em que:

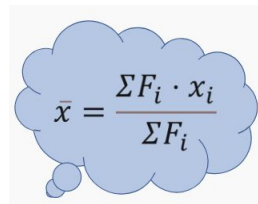
l_i = limite inferior da classe modal

$$\Delta_1 = F_i - F_{i,ant}$$

$$\Delta_2 = F_i - F_{i,post}$$

h = amplitude da classe modal

Fórmula utilizada para achar as médias:


$$\bar{x} = \frac{\sum F_i \cdot x_i}{\sum F_i}$$

Fórmula utilizada para achar as medianas:

Mediana para Dados Agrupados Com Intervalo de Classe

Procedimento:

- Calcula-se a posição da mediana: $POS(Md) = \frac{n}{2}$
- Pela F_{ac} identifica-se a classe que contém o valor da mediana - CLASSE(Md)
- Utiliza-se a fórmula: $Md = l_i + \frac{POS(Md) - F_{ac,ant}}{F_i} \cdot h$

Onde:

l_i = Limite inferior da classe mediana

n = Tamanho da amostra ou número de elementos

$F_{ac,ant}$ = Frequência acumulada anterior à classe mediana

h = Amplitude da classe mediana

Análise de dados

Fórmula utilizada para achar os desvios:

3ª Situação dados agrupados com classe

$$Dp = \sqrt{\frac{\sum (xi - \bar{x})^2 \cdot Fi}{\sum Fi}}$$

$$\bar{x} = \frac{3610}{58} = 62,24$$

ESCORES	ALUNOS				
	F_i	xi	$(xi - \text{média})^2$	$(xi - \text{média})^2 \cdot Fi$	
35 - 45	5	40	494,62	2473,10	
45 - 55	12	50	149,82	1797,84	
55 - 65	18	60	5,02	90,36	
65 - 75	14	70	60,22	843,08	
75 - 85	6	80	315,42	1892,52	
85 - 95	3	90	770,62	2311,86	
TOTAL	58	-	-	9408,76	

$$Dp = \sqrt{\frac{9408,76}{58}} = \sqrt{162,22} = 12,73$$

Fórmula utilizada para os coeficientes de variação:

COEFICIENTE DE VARIAÇÃO

Trata-se de uma média relativa à dispersão, útil para a comparação e observação em termos relativos do grau de concentração em torno da média de séries distintas. E dada por:

$$cv = \frac{d_p}{\bar{x}} \cdot 100$$

Classificação da distribuição quanto à dispersão:

- DISPERSÃO BAIXA: $CV \leq 15\%$
- DISPERSÃO MÉDIA: $15\% < CV < 30\%$
- DISPERSÃO ALTA: $CV \geq 30\%$

Análise de dados

Fórmula utilizada para achar os coeficientes de correlação:

$$r = \frac{n \cdot \sum x_i \cdot y_i - (\sum x_i) \cdot (\sum y_i)}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}}$$

Fórmula utilizada para achar os coeficientes de regressão:

$$b = \frac{\sum x \cdot y - n \cdot \bar{x} \cdot \bar{y}}{\sum x^2 - n \cdot \bar{x}^2}$$

$$\bar{x} = \frac{\sum x}{n} \quad \bar{y} = \frac{\sum y}{n}$$

O **a** é calculado como

$$a = \overline{Y} - b \cdot \overline{X}$$

Ingressantes Presencial

FAIXA ETÁRIA (IDADE)	QT_ALUNOS (Fi)	xi	xiFi	fac	(xi - média) ²	(xi - média) ² *fi
QT_ING_0_17	2336	8.5	19856	2336	281.40	657353
QT_ING_18_24	131708	21	2765868	134044	18.28	2407065
QT_ING_25_29	23837	27	643599	157881	2.98	70929
QT_ING_30_34	14533	32	465056	172414	45.23	657261
QT_ING_35_39	10825	37	400525	183239	137.48	1488169
QT_ING_40_49	12289	44.5	546860.5	195528	369.60	4542014
QT_ING_50_59	3317	54.5	180776.5	198845	854.10	2833049
QT_ING_60 MAIS	692	30	20760	199537	22.33	15449
Total	199537	-	5043301	-	-	12671289

- Média = 25 anos
- Moda = 21 anos
- Médiana 18 anos
- Desvio padrão = 8 anos
- Coef. Variação = 31%

Faixa etária 18-24 é mais frequente, perfil típico universitário.
 Faixa etária 60+ menos frequente, Melhor idade.
 O Coef de 31% mostra que há uma dispersão significativa; um grande número de ingressantes fora da média (25 anos).
 Composição assimétrica; maior concentração 18 até 29 anos e depois uma queda progressiva, dados heterogeneos.

Ingressantes EAD

FAIXA ETÁRIA (IDADE)	QT_ALUNOS (Fi)	xi	xiFi	fac	(xi - média) ²	(xi - média) ² *fi
QT_ING_0_17	633	8.5	5380.5	633	444.96	281657
QT_ING_18_24	42556	21	893676	43189	73.86	3143024
QT_ING_25_29	25763	27	695601	68952	6.73	173350
QT_ING_30_34	18497	32	591904	87449	5.79	107079
QT_ING_35_39	14184	37	524808	101633	54.85	777984
QT_ING_40_49	14437	44.5	642446.5	116070	222.19	3207756
QT_ING_50_59	3252	54.5	177234	119322	620.31	2017250
QT_ING_60 MAIS	396	30	11880	119718	0.16	65
Total	119718	-	3542930	-	-	9708166

- Média = 30 anos
- Moda = 22 anos
- Médiana = 18 anos
- Desvio padrão = 9 anos
- Coef. Variação = 30%

Perfil mais comum ainda é 18-24.
 A média é de 5 anos a mais que o presencial.
 O desvio padrão mostra que a variação de idade é ainda maior que a da modalidade presencial, ou seja, o perfil EAD é mais diverso, é uma modalidade mais “favorável” à pessoas mais velhas que conciliam estudo com outras responsabilidades.

Matriculados Presencial

FAIXA ETÁRIA (IDADE)	QT_ALUNOS (Fi)	xi	xiFi	fac	(xi - média) ²	(xi - média) ² *fi
QT_MAT_0_17	2133	8.5	18130.5	2133	299.80	639479
QT_MAT_18_24	391394	21	8219274	393527	23.18	9073447
QT_MAT_25_29	116567	27	3147309	510094	1.40	163739
QT_MAT_30_34	51383	32	1644256	561477	38.26	1965738
QT_MAT_35_39	36097	37	1335589	597574	125.11	4516041
QT_MAT_40_49	39201	44.5	1744444.5	636775	349.14	13686494
QT_MAT_50_59	11131	54.5	606639.5	647906	822.84	9159034
QT_MAT_60 MAIS	2372	30	71160	650278	17.52	41548
Total	650278	-	16786802.5	-	-	39245521

- *Média = 26 anos*
- *Moda = 22 anos*
- *Médiana = 18 anos*
- *Desvio padrão = 8 anos*
- *Coef. Variação = 30%*

A faixa etária 18-24 ainda se mantém muito dominante.
As outras tendências e padrões se mantêm semelhantes aos dados anteriores.

Matriculados EAD

FAIXA ETÁRIA (IDADE)	QT_ALUNOS (Fi)	xi	xiFi	fac	(xi - média) ²	(xi - média) ² *fi
QT_MAT_0_17	553	8,5	4700,5	553	494,10	273236
QT_MAT_18_24	60802	21	1276842	61355	94,64	5754280
QT_MAT_25_29	44902	27	1212354	106257	13,90	624145
QT_MAT_30_34	33869	32	1083808	140126	1,62	54774
QT_MAT_35_39	27272	37	1009064	167398	39,33	1072726
QT_MAT_40_49	28924	44,5	1287118	196322	189,66	5485725
QT_MAT_50_59	6704	54,5	365368	203026	565,09	3788391
QT_MAT_60 MAIS	841	30	25230	203867	0,53	446
Total	203867	-	6264484,5	-	-	17053723

- Média = 31 anos
- Moda = 23 anos
- Médiana = 18 anos
- Desvio padrão = 9 anos
- Coef. Variação = 30%

Apesar de ainda ser a mais frequente a faixa 18-24 já não destoa tanto de outras demografias, principalmente das 25-29 e 30-34, isso corrobora o padrão de maior diversidade da modalidade EAD.

Assimetria positiva entre a moda e a media demonstrando que faixa etárias mais altas estão influenciando a media.

Concluintes Presencial

FAIXA ETÁRIA (IDADE)	QT_ALUNOS (Fi)	xi	xiFi	fac	(xi - média) ²	(xi - média) ² *fi
QT_CONC_18_24	45007	21	945147	45007	41,50	1867932
QT_CONC_25_29	30029	27	810783	75036	0,20	5874
QT_CONC_30_34	10268	32	328576	85304	20,77	213294
QT_CONC_35_39	6954	37	257298	92258	91,35	635246
QT_CONC_40_49	7601	44,5	338244,5	99859	290,97	2211628
QT_CONC_50_59	2184	54,5	119028	102043	732,12	1598949
QT_CONC_60 MAIS	476	30	14280	102519	6,54	3114
Total	102519	-	2813356,5	-	-	6536037

- Média = 27 anos
- Moda = 23 anos
- Médiana = 18 anos
- Desvio padrão = 8 anos
- Coef. Variação = 29%

Aproximadamente 44% do total de alunos entre 18-24.
A porcentagem de formandos 50-60+ é de 2,8%, indicando a idade como fator inversamente proporcional ao número de concluintes., confirma a tendência de representatividade diminuta da melhor idade no Ensino Superior presencial, principalmente como concluintes.

Concluintes EAD

FAIXA ETÁRIA (IDADE)	QT_ALUNOS (Fi)	xi	xiFi	fac	(xi - média) ²	(xi - média) ² *fi
QT_CONC_18_24	4566	21	95886	4566	157,00	716853
QT_CONC_25_29	5957	27	160839	10523	42,64	254003
QT_CONC_30_34	5279	32	168928	15802	2,34	12356
QT_CONC_35_39	4729	37	174973	20531	12,04	56945
QT_CONC_40_49	5370	44,5	238965	25901	120,34	646245
QT_CONC_50_59	1405	54,5	76572,5	27306	439,75	617843
QT_CONC_60 MAIS	169	30	5070	27475	12,46	2106
Total	27475	-	921233,5	-	-	2306349

- Média = 34 anos
- Moda = 28 anos
- Médiana = 22 anos
- Desvio padrão = 9 anos
- Coef. Variação = 27%

A média e moda do EAD são consideravelmente maiores que a do presencial, coeso com padrões anteriores.
A porcentagem de concluintes 50-60+ é de 5,7%, ainda é relativamente pequena porém mais que o dobro da presencial.

Correlação Ingressantes x Concluintes - Presencial

indice	qt_ingressantes_presencial(xi)	qt_conc_presencial(yi)	xi * yi	xi²	yi²
9616	250	181	45250	62500	32761
9649	123	87	10701	15129	7569
9650	127	87	11049	16129	7569
9688	102	56	5712	10404	3136
9736	90	60	5400	8100	3600
9752	162	118	19116	26244	13924
9798	323	129	41667	104329	16641
9824	115	95	10925	13225	9025
9837	421	66	27786	177241	4356
9839	272	81	22032	73984	6561
9850	893	489	436677	797449	239121
9865	215	93	19995	46225	8649
9866	453	130	58890	205209	16900
9868	649	214	138886	421201	45796
9891	93	28	2604	8649	784
9917	62	19	1178	3844	361
9984	572	141	80652	327184	19881
9988	150	58	8700	22500	3364
10032	99	41	4059	9801	1681
10055	167	121	20207	27889	14641
10066	146	102	14892	21316	10404
10067	189	121	22869	35721	14641
10069	102	62	6324	10404	3844
10071	127	83	10541	16129	6889

Correlação : 0,75 (Positiva forte)

Correlação Ingressantes x Concluintes - EAD

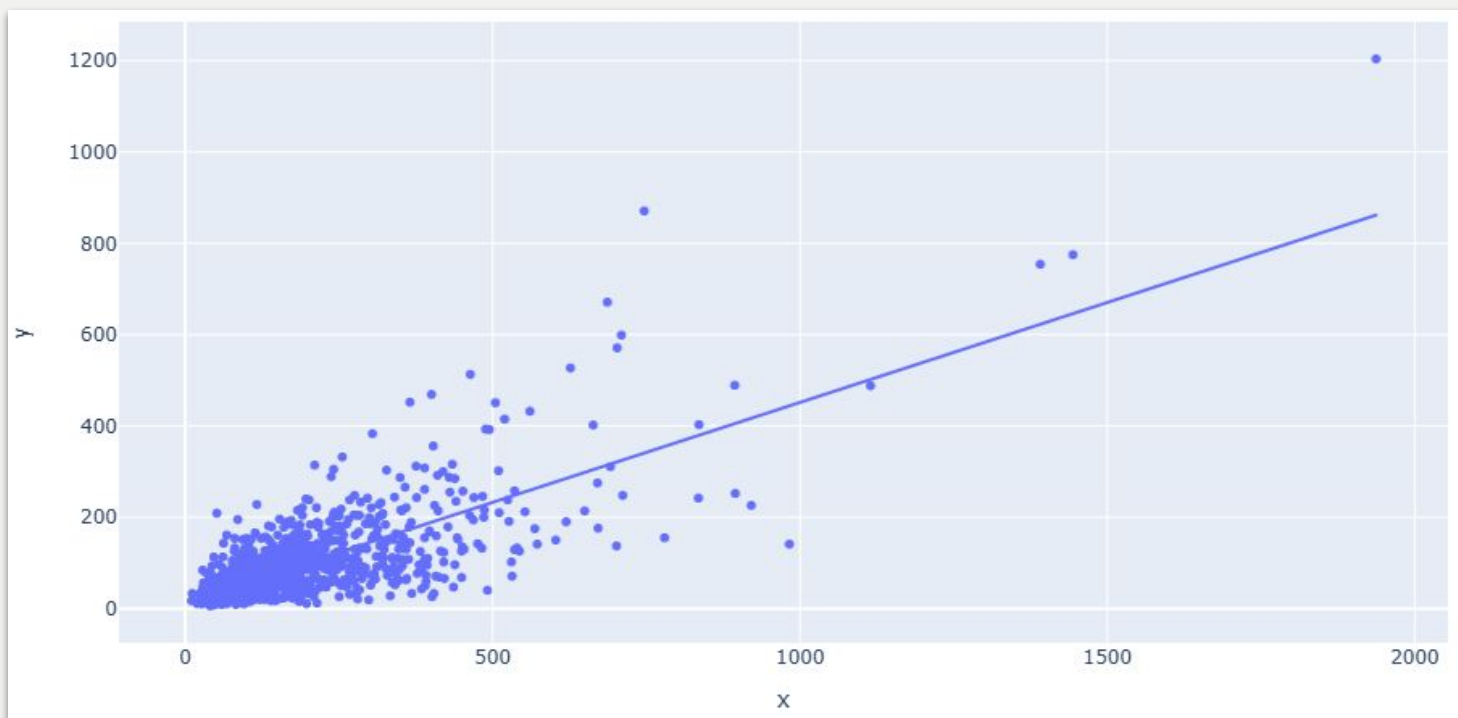
indice	qt_ingressantes_presenci	qt_conc_presenci	$x_i * y_i$	x_i^2	y_i^2
10479	54	21	1134	2916	441
10519	696	113	78648	484416	12769
10525	62	25	1550	3844	625
10576	335	73	24455	112225	5329
10634	588	47	27636	345744	2209
10637	743	72	53496	552049	5184
10662	325	52	16900	105625	2704
10696	206	24	4944	42436	576
10715	1118	222	248196	1249924	49284
10730	97	25	2425	9409	625
10736	248	28	6944	61504	784
10737	486	23	11178	236196	529
10783	356	119	42364	126736	14161
11120	124	19	2356	15376	361
11141	98	18	1764	9604	324
11651	154	40	6160	23716	1600
11666	77	54	4158	5929	2916
11723	541	132	71412	292681	17424
11795	297	48	14256	88209	2304
11933	184	7	1288	33856	49
12058	65	22	1430	4225	484
12104	54	19	1026	2916	361
12180	313	167	52271	97969	27889
12895	320	67	21440	102400	4489

Correlação : 0,81 (Positiva forte)

Análise de dados

Regressão - Ingressantes x Concluintes Presencial

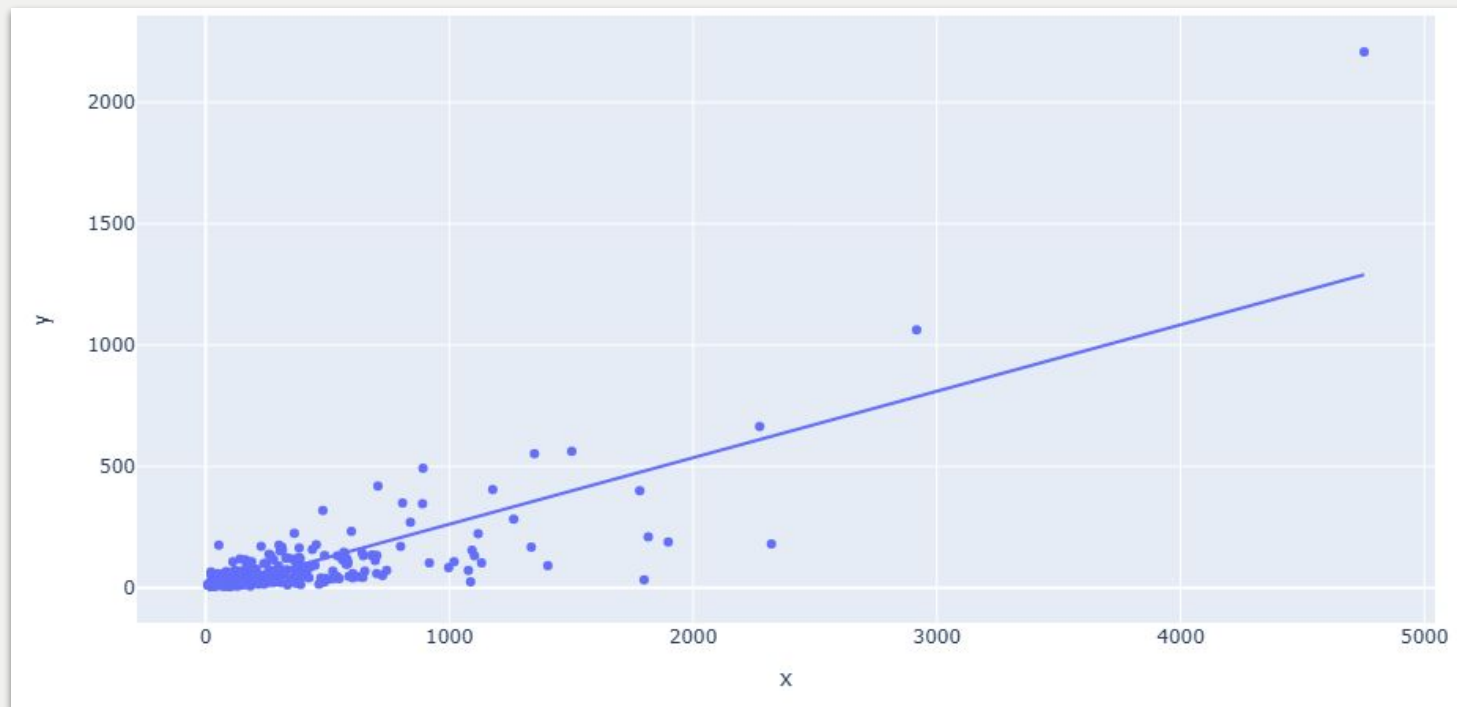
$$Y = 12,3 + 0,27x$$



Análise de dados

Regressão - Ingressantes x Concluintes EAD

$$Y = 13,8 + 0,44x$$



Considerações Finais

Em resumo

- A faixa etária 18-24 é mais frequente em ambas modalidades, mas completamente dominante em Ingressantes/Matriculados no ensino presencial.
- O Ensino EAD possui uma maior e mais significativa diversidade de faixas etárias.
- A modalidade EAD possui uma maior taxa de aprovação.
- As mulheres são maioria no ensino superior em ambas as modalidades, representando quase dois terços dos estudantes.
- A demografia 50-60+ é percentualmente pequena em ambas modalidades.
- A faixa etária de 50-60+ anos ingressa significativamente mais no ensino presencial, mas acaba se formando em maior número (mais que o dobro) no ensino EAD.

Racional de tomada de decisão

A análise de dados delineou uma alta taxa de evasão entre indivíduos da melhor idade (50-60+) no ensino superior, especialmente na modalidade presencial.

Primeiramente, estudos apontam que a participação de indivíduos dessa faixa etária em programas de educação superior tem impacto direto na **redução de sintomas depressivos e aumento do senso de realização pessoal tal qual a integração social** (Souza, Pamplona e Fixina , 2016)

Dado que projeções demográficas apontam que até 2050 o Brasil terá 97 milhões de pessoas com 50 anos ou mais (Exame.com), cerca de **44% (quase metade do País)**, se torna claro que o bem estar da população 50+ se tornará uma questão estratégica de interesse nacional no longo prazo.

Portanto, é essencial que exista um maior foco e alocação de recursos voltados a ações afirmativas para a melhor idade.

Sugestão Final - Ações Afirmativas para a Melhor Idade



CE debate reserva de vaga para idosos em universidades e institutos nesta sexta

Da Agência Senado | 18/04/2024, 09h19





A Câmara Legislativa celebrou os oito anos da Universidade do Envelhecer (**Uniser**), programa de extensão da Universidade de Brasília (UnB), na tarde desta segunda-feira (10), em sessão solene no voltado aos idosos, suscita emoções como fraternidade, amor, alegria e dignidade no processo de "saber envelhecer". Ao citar as ações sociais da Uniser, a parlamentar opinou que a instituição é uma prova que "a sociedade unida e organizada faz a diferença e a transformação na vida de todos".

OBRIGADO PELA SUA ATENÇÃO!