RAE-SEA-8403

ESTUDO POPULACIONAL DO ÂNGULO DE WIBERG.

Marcos Nascimento Magalhães
Marly Grasso Nunes

- São Paulo, Fevereiro de 1984 -

SETOR DE ESTATÍSTICA APLICADA

RELATÓRIO DE ANÁLISE ESTATÍSTICA NÚMERO 03 / 84

CÓDIGO 41 / 83

TÍTULO: Estudo populacional do ângulo de Wiberg.

CONSULENTE: José Laredo Filho

INSTITUIÇÃO: Escola Paulista de Medicina

FINALIDADE: Publicação

AUTORES: Prof. Marcos Nascimento Magalhães e Profa. Marly Grasso Nunes.

REFERÊNCIA DESTE TRABALHO:

MAGALHÃES,M.N. e NUNES,M.G. - Estudo populacional do ângulo de Wiberg. São Paulo, IME-USP, 1984. 23p. (SEA. Relatório de Análise Estatística, 8403.)

FICHA TÉCNICA

BIBLIOGRAFIA:

DRAPER,N.R. and SMITH,H. - Applied regression analysis. 2.ed. New York, John Wiley, 1981. 607p. (Wiley Series in Probability and Mathematical Statistics.)

TIMM,N.H. - Multivariate analysis. USA, Brooks/Cole Series in Statistics, 1975. 689p.

BMDP - Centro de Computação Eletrônica da Universidade de São Paulo. Em Dixon,W.J. - BMDP Statistical Software. Berkeley, University of California Press, 1981.

PROGRAMA DE COMPUTAÇÃO: BMDP

DADOS E TÉCNICAS ESTATÍSTICAS UTILIZADAS:

Foi realizada uma análise de Regressão sobre os ângulos de Wiberg obtidos de radiografias de 1091 pacientes do Hospital São Paulo, divididos em idades de 1 a 81 anos. (Teste de falta de ajuste e comparação de duas curvas de regressão (sexo masculino x sexo feminino).)

1. Objetivo

O ângulo de Wiberg constitui um ótimo parâmetro para o estudo e avaliação das displasias acetabulares nas diversas formas de patologia dos quadris. Pretende-se conseguir um modelo que nos dê o comportamento do "ângulo de Wiberg" segundo a variação das idades, permitindo discernir entre indivíduos portadores ou não de displasias.

2. Descrição dos Dados

Foram obtidas radiografias dos quadris de pacientes do Hospital São Paulo, e após exame clínico ortopédico foram afastadas patologias do sistema ósseo articular. O grupo de indivíduos resultante entre as idades de 1 a 81 anos foi dividido em faixas cujos limites estarão dentro de 6 meses. Os ângulos de Wiberg foram medidos pelo método classicamente descrito pelo autor.

3. Análise Estatística

3.1. Preliminares

(a) As considerações feitas abaixo foram tomadas baseadas na média dos ângulos (direito-esquerdo) uma vez que ficou constatado não haver diferença significativa no comportamento do modelo quando tomado isoladamente para cada ângulo.

(b) A variável raça não foi levada em conta tendo em vista as seguintes dificuldades:

(i) a subjetividade da classificação

(ii) a discrepância acentuada entre o número de pessoas presentes na amostra classificadas como da raça branca (903), o número de pessoas classificadas como pretas (104) ou mulatas (84).

A tentativa de ajustar um modelo para a raça preta ou mulata redundou num absurdo que é explicável pelo número muito reduzido de observações das respectivas raças.

(c) Da análise puramente descritiva dos dados (médias e desvios) observou-se que até 17 anos a velocidade de crescimento do ângulo é maior do que em idades posteriores. (Ver gráfico 1).

(d) As estimativas obtidas neste trabalho foram encontradas mediante a utilização do Pacote BMDP-2R implantado no Centro de Computação Eletrônica da USP.

3.2. Modelo

Tradicionalmente, a busca de um modelo parte de um mais simples para os mais complexos. No nosso caso, o modelo mais simples que é de regressão linear, satisfez plenamente nossas exigências de acuidade. O modelo proposto inicialmente foi:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

onde

Y = ângulo (média entre direito e esquerdo).

X_1: variação das idades até 17 anos (sendo igual a 17 a partir de 17 anos).

X_2: descreve o salto em 17 anos.

X_3: variação das idades a partir de 17 anos (sendo igual a zero antes dos 17 anos).

Constatamos, entretanto, que o β_2 correspondente ao salto foi irrelevante para o modelo. Isto mostra que a escolha do ponto mudança em 17 anos foi a melhor possível. Isto é, podemos considerar como modelo definitivo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \varepsilon$$

temos então

β_0: intercepto (ângulo previsto no nascimento).

β_1: velocidade de crescimento do ângulo até 17 anos.

β_2: velocidade de crescimento a partir dos 17 anos.

ε : erro.

Suposições necessárias para os testes efetuados são:

(1) ε obedece a distribuição normal com média zero e variância σ^2.

(2) As observações são independentes.

As estimativas obtidas separadamente para cada sexo:

Sexo FEMININO:

$$Y = 25{,}51 + 0{,}71X_1 + 0{,}07X_3.$$

A seguir apresentamos a tabela de Análise de Variância que dá informações sobre o ajuste do modelo.

F.V.	g.l.	S.Q.	QM	F
Reg.	2	5245,36	2622,68	64,18
Res.	626	25579,87	40,86	

$R^2 = 0{,}17$

onde: F.V.: Fonte de Variação.

Reg.: Regressão.

Res.: Resíduo.

g.l.: graus de liberdade.

S.Q.: Soma de Quadrados.

QM: quadrado médio.

F: valor da estatística F de Snedecor.

R^2: coeficiente de ajuste do modelo.

Sexo MASCULINO:

$$Y = 29,04 + 0,48X_1 + 0,10X_3.$$

ANOVA

F.V.	g.l.	S.Q.	QM	F
Reg.	2	2475,73	1237,87	28,40
Res.	459	20009,76	43,59	

$R^2 = 0,11$

Foi feito um teste para verificar a diferença ou não dos dois modelos (ver gráfico 2). A descrição do teste está no apêndice 1. Concluímos não haver diferença entre os 2 modelos. Isto é, estatisticamente as retas que descrevem o ângulo para cada sexo, podem ser consideradas como a mesma. Assim, com base nesse resultado, estimou-se novamente os parâmetros do modelo usando o conjunto total de dados. Obtemos então:

$$Y = 26,99 + 0,62X_1 + 0,08X_3$$

ANOVA

F.V.	g.l.	S.Q.	QM	F
Reg.	2	7625,47	3812,74	90,74
Res.	1088	45713,70	42,02	

$R^2 = 0,14$

Como podemos observar o R^2, coeficiente de explicação do modelo, é relativamente baixo o que poderia indicar uma inadequação deste modelo ao conjunto de dados (apesar da análise de resíduos aparentemente não apresentar problemas); porém, como os dados continham várias observações para cada idade, foi possível realizar um teste de falta de ajuste (ver apêndice 2), que redundou na aceitação do modelo. Portanto, o baixo valor de R^2 pode ser explicado pela variabilidade intrínseca da medida, ou seja, os dados são por natureza muito dispersos. Lembremos que não houve uma seleção à priori das medidas incluídas no conjunto de dados.

3.3. Intervalo de Confiança para Modelo

Tendo em vista que o modelo ajustado estabelece um único valor de ângulo (o valor médio) para cada idade, e como esse modelo tem como uma das finalidades discernir entre pacientes afetados ou não pela "displasia óssea" é necessário estabelecer uma região ao redor desse modelo em que indivíduos com medidas angulares pertencentes a essa região são considerados não afetados.

Como o número de observações difere bastante de idade para idade, não nos pareceu conveniente utilizar essas observações para construir, isoladamente em cada idade, um intervalo de confiança. Obteríamos precisões diferentes para idades diferentes. Nossa opção foi então, o uso global das observações construindo o intervalo da seguinte forma:

$$\hat{Y} \pm 1{,}96.\sqrt{QMEP}$$

onde

$\hat{Y}$: valor estimado pelo modelo para uma certa idade.

QMEP : quadrado médio de erro puro (ver apêndice 2).

A partir das suposições do modelo constatamos que Y tem distribuição Normal, logo o valor 1,96 é obtido da tabela Normal com o intuito de fazer com que o intervalo construído tenha a confiança de 95%.

O QMEP é uma estimativa da variância, que segundo as suposições do modelo, é constante em todas as idades.

O valor numérico encontrado foi:

$$\hat{Y} \pm 12,67.$$

4. Ajuste de modelo para a população até 60 anos

Havia interesse do pesquisador em verificar se o comportamento do modelo se alteraria substancialmente com a exclusão das observações correspondentes às idades superiores a 60 anos.

O mesmo modelo descrito em 3.2 foi ajustado agora, desprezando-se as idades superiores a 60 anos. Obtemos:

$$Y = 26,20 + 0,66X_1 + 0,08X_3.$$

ANOVA

F.V.	g.l.	S.Q.	QM	F
Reg.	2	6122,80	3061,40	76,17
Res.	836	33601,26	40,19	

$R^2 = 0,15$

Os procedimentos utilizados em 3.2 para verificar a adequação do modelo, foram repetidos neste caso e a conclusão adotada indica a adequação do modelo proposto.

Partimos então, para a comparação dos 2 modelos: o obtido em 3.2 com todas as observações e o obtido nesta seção. Uma dificuldade para uma comparação estatística mais acurada vem do fato de que as estimativas dos modelos estão baseadas no "mesmo" conjunto de dados. Isto é, para o 2º modelo utilizamos uma parte ponderável das observações que construiram o 1º, isto acarreta dependência entre as estimativas dos dois modelos.

A comparação foi feita então de modo descritivo:

1º MODELO:

$$Y = 26{,}99 + 0{,}62X_1 + 0{,}08X_3$$

$R^2 = 0{,}14.$

2º MODELO:

$$Y = 26{,}20 + 0{,}66X_1 + 0{,}08X_3$$

$R^2 = 0{,}15.$

Percebemos que a diferença entre as estimativas não é muito grande. O R^2 do 2º modelo é praticamente igual ao do 1º. Construindo o intervalo de confiança para o 2º modelo nos moldes do descrito em 3.3 e fazendo uma comparação entre as duas regiões de confiança obtidas, notamos uma área comum bastante grande (vide gráfico 3). Acreditamos portanto, não haver diferenças significantes entre os dois modelos.

5. Conclusão

Com as considerações feitas anteriormente e os dados disponíveis optamos por recomendar o 1º modelo, isto é:

$$Y = 26{,}99 + 0{,}62X_1 + 0{,}08X_3$$

como explicativo do comportamento do ângulo em função da idade. Sen do assim para qualquer idade, se o ângulo médio medido pertencer a região de confiança (vide gráfico 4), o paciente é considerado sadio.

A escolha do modelo acima fundamenta-se no fato de que para esse modelo utilizou-se todas as observações disponíveis; o truncamento discutido na seção 4 não apresentou ganhos relevantes de ajuste quando comparado com o modelo escolhido e tem a desvantagem de não servir como bom previsor para idades superiores a 60 anos.

MNM/MGN/ls - 24.02.84

GRĀFICOS

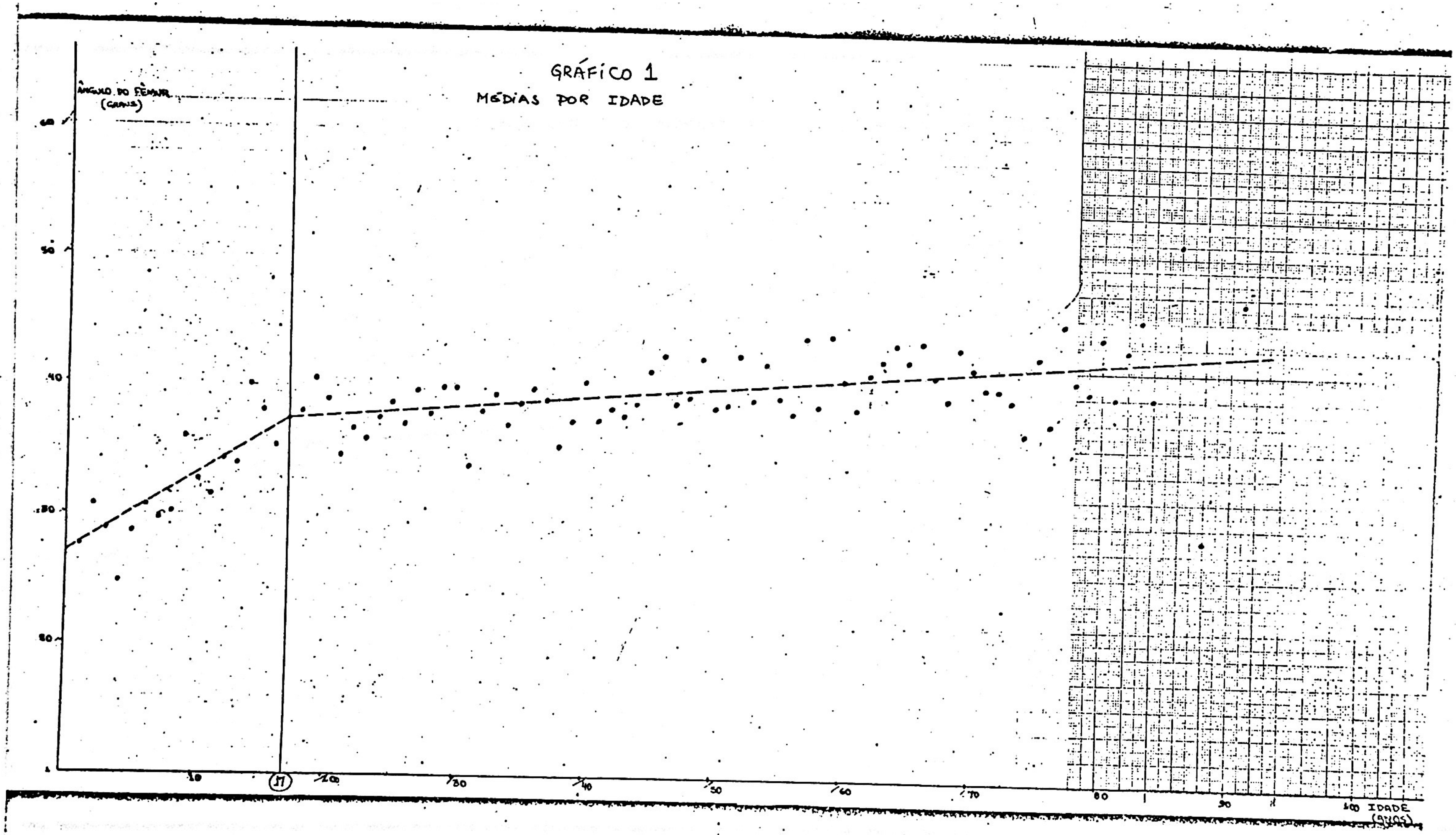


GRÁFICO 1
MÉDIAS POR IDADE
ÂNGULO DO FÊMUR (GRAUS)
60
50
40
30
20
10
17
20
30
40
50
60
70
80
90
100 IDADE (anos)

GRÁFICO 2

COMPARAÇÃO DOS MODELOS: MASCULINO, FEMININO e TOTAL

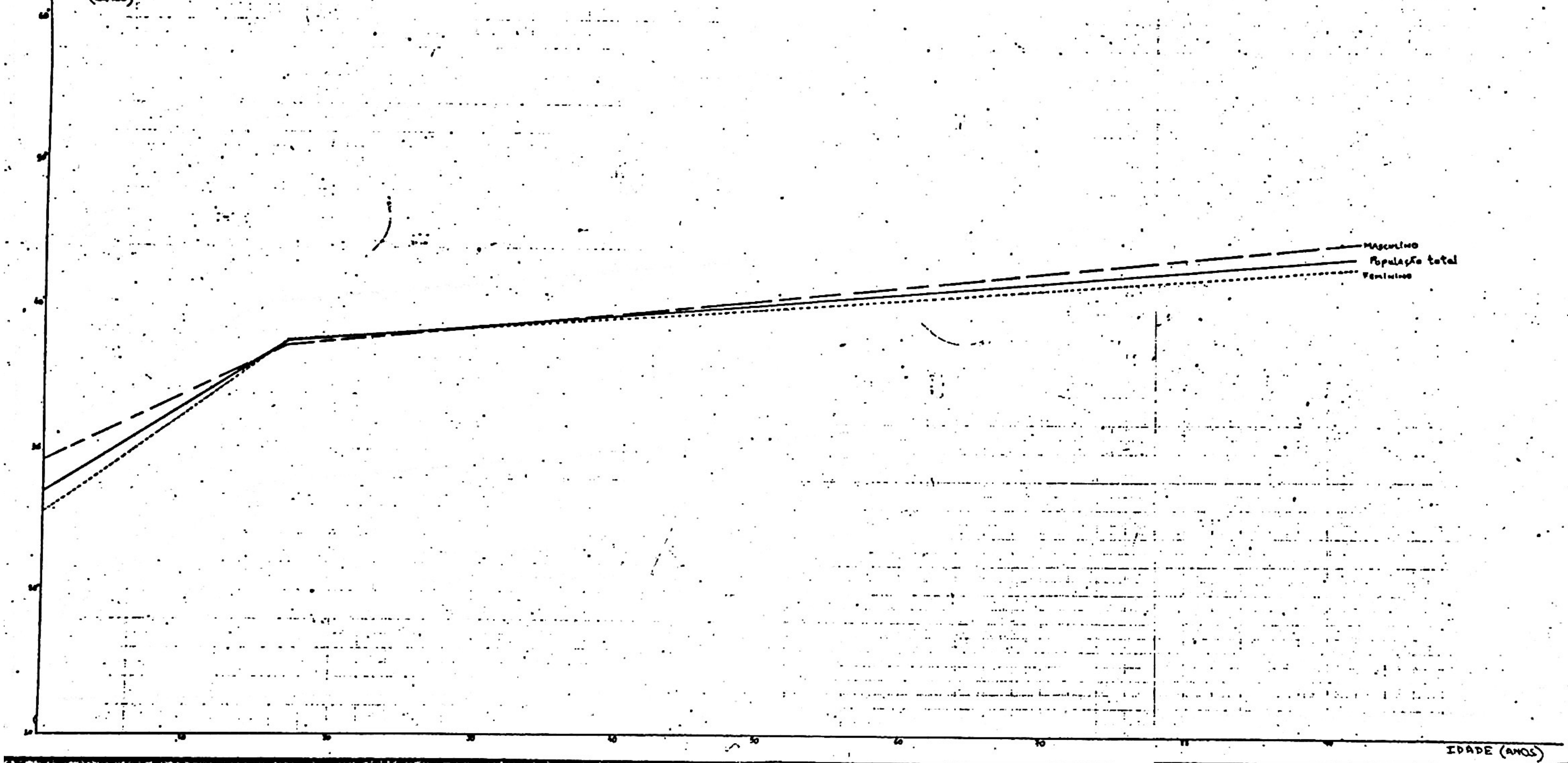


GRÁFICO 3

COMPARAÇÃO ENTRE O MODELO TRUNCADO E O TOTAL

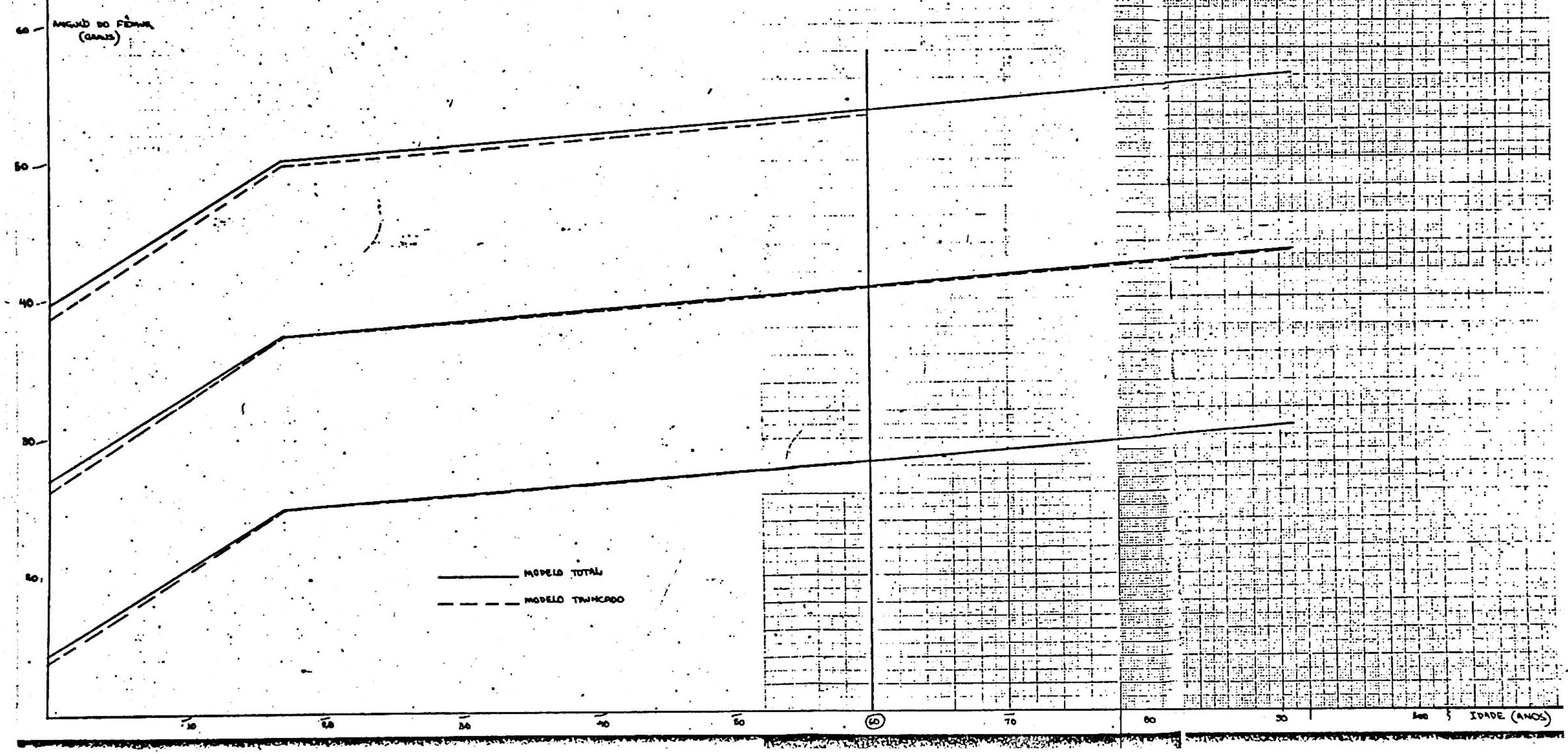
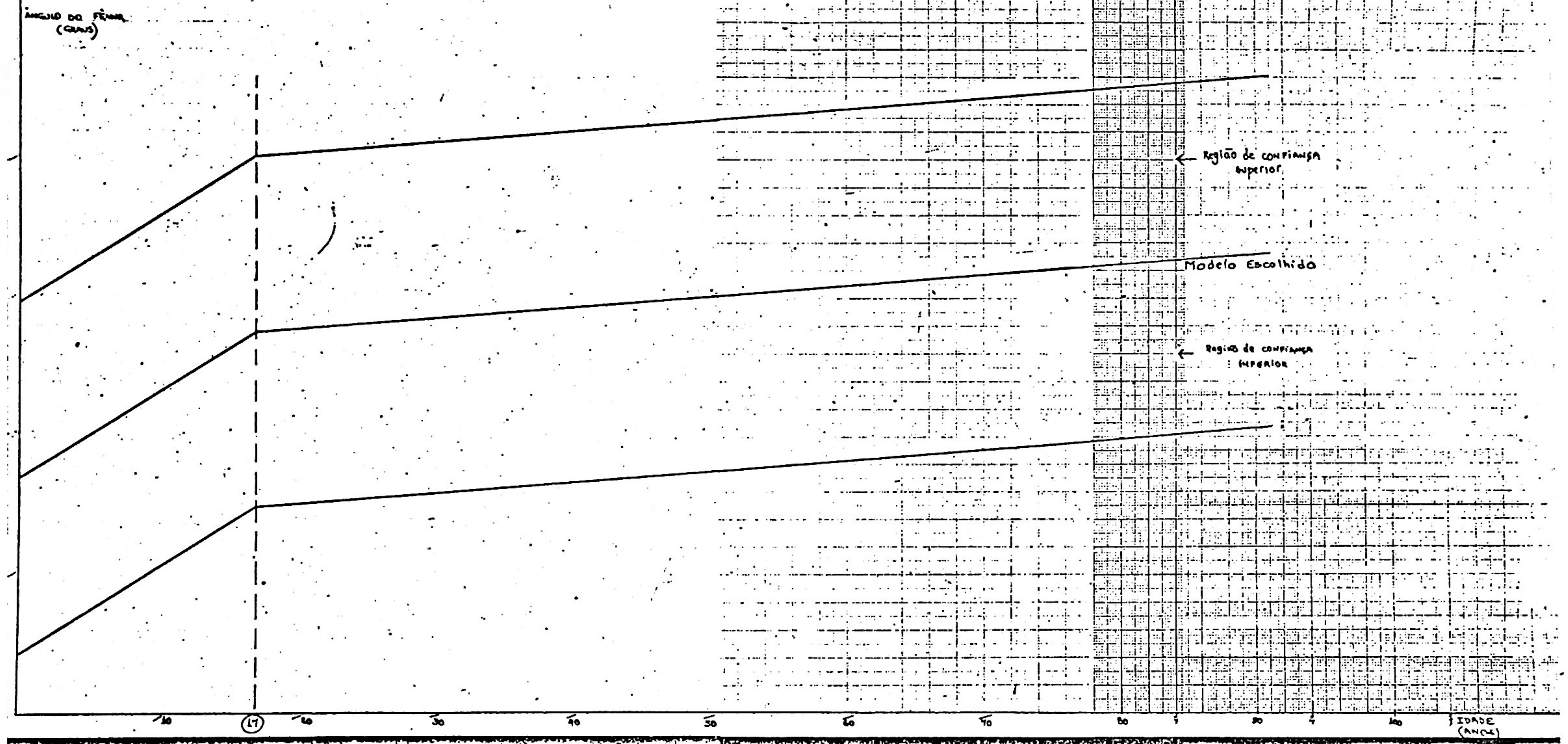


GRÁFICO 4

MODELO ESCOLHIDO E REGIÃO DE CONFIANÇA



APÊNDICES

APÊNDICE 1

Comparação dos ajustes obtidos para os dois sexos

Os modelos encontrados foram:

Sexo Fem.: $\hat{Y} = 25,51 + 0,71X_1 + 0,07X_3$

Sexo Masc.: $\hat{Y} = 29,04 + 0,48X_1 + 0,10X_3$

Hipótese a ser testada:

$$H_o: \underset{\sim}{\beta}_f = \underset{\sim}{\beta}_M$$

onde

$\underset{\sim}{\beta}_f = (\beta_{of}, \beta_{1f}, \beta_{2f})$ e

$\underset{\sim}{\beta}_M = (\beta_{oM}, \beta_{1M}, \beta_{2M})$.

O teste feito requer as mesmas hipóteses já assinaladas em 3.2. A estatística utilizada será:

$$F_o = \frac{SQR_{H_o} - SQR}{SQR} \cdot \frac{n-q}{k}$$

de tal forma que, considerando H_o verdadeira, F_o tem distribuição Fisher-Snedecor com k e n-q graus de liberdade, com

$SQR_{H_o} = SQR + (\hat{\underset{\sim}{\beta}}_f - \hat{\underset{\sim}{\beta}}_M)'((\underset{\sim}{X}_f'\underset{\sim}{X}_f)^{-1} + (\underset{\sim}{X}_M'\underset{\sim}{X}_M)^{-1})^{-1}(\hat{\underset{\sim}{\beta}}_f - \hat{\underset{\sim}{\beta}}_M)$

$SQR = SQR_f + SQR_M$

SQR_f = soma de quadrados residual do modelo ajustado para o sexo feminino.

SQR_M = soma de quadrados residual do modelo ajustado para o sexo masculino.

$\hat{\underset{\sim}{\beta}}_f$ = estimativas encontradas para $\underset{\sim}{\beta}_f$.

$\hat{\underset{\sim}{\beta}}_M$ = estimativas encontradas para $\underset{\sim}{\beta}_M$.

$\underset{\sim}{X}_f$ = matriz de planejamento do modelo para o sexo feminino.

onde

$$\underset{\sim}{X}_f = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 16 & 0 \\ 1 & 17 & 0 \\ 1 & 17 & 1 \\ 1 & 17 & 2 \\ \vdots & \vdots & \vdots \\ 1 & 17 & 73 \end{bmatrix}$$

$\underset{\sim}{X}_M$ = idem ao anterior para o sexo masculino.

k = nº de parâmetros em cada modelo.

n-q = soma dos graus de liberdade residual do modelo para o sexo masculino.

Os valores estimados foram:

$$\left[(\underset{\sim}{X}_f'\underset{\sim}{X}_f)^{-1} + (\underset{\sim}{X}_M'\underset{\sim}{X}_M)^{-1}\right]^{-1} = \begin{bmatrix} -53,505 & -1087,926 & -3834,493 \\ -1087,926 & -19343,380 & -64079,127 \\ -3834,493 & -64079,127 & -105360,86 \end{bmatrix}$$

SQR = 45589,63

SQR_{H_o} -SQR = 53,91

k = 3

n-q = 1085

Assim, o valor de F_o encontrado foi 0,43 que, com k-3 e n-q=1085, nos indica que nada leva a crer que H_o deva ser rejeitada.

APÊNDICE 2

Teste para verificação de falta de ajuste

Lembremos que:

$$R^2 = \frac{SQReg}{SQT} \quad \text{e} \quad SQReg = SQT - SQRes.$$

Notemos ainda que, no nosso caso, temos várias observações para uma mesma idade, e portanto é possível expressar SQRes da seguinte forma:

$$SQRes = SQEP + SQFA$$

onde

SQEP : soma dos quadrados de erro puro.

SQFA : soma dos quadrados de falta de ajuste linear.

Para calcular SQEP e SQFA temos

$$SQEP = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{.j})^2$$

$$SQFA = \sum_{j=1}^{k} n_j \cdot (y_{.j} - \hat{Y}_j)^2$$

com

k : número de idades diferentes presentes na amostra.

n_j: número de observações na idade j.

$\bar{Y}_{.j}$: média dos ângulos na idade j.

Y_{ij}: ângulo da i-ésima observação na idade j.

$\hat{Y}_j$: valor estimado pelo modelo na idade j.

A hipótese a ser testada será:

$$H_o : EY = \beta_o + \beta_1 X_1 + \beta_3 X_3$$

ou seja queremos verificar se Y pode ser expresso através de um modelo linear.

As suposições necessárias para a realização do teste são as mesmas já referidas em 3.2.

A estatística utilizada será:

$$F_o = \frac{SQFA/k-2}{SQEP/\sum_{i=1}^{k} n_j - k}$$

que sob a condição de H_o ser verdadeira tem distribuição F de Snedecor com k-2 e $\sum_{j=1}^{k} n_j - k$ graus de liberdade.

Os resultados encontrados foram:

$$SQEP = 41927{,}81$$

$$SQFA = 3785{,}89$$

$$k-2 = 85$$

$$\sum_{j=1}^{k} n_j - k = 1004$$

$$F_o = 1{,}07.$$

Concluímos pela aceitação de H_o, ou seja, o modelo linear é um bom ajuste para os dados.

ANEXO

TITLE IS 'DISPERSAO POR IDADE'

CRITERION VARIABLE MEDIA

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
IDADE	42.		346.5000	38.5000	4.7236	22.3125	(9)
IDADE	43.		532.5000	38.0357	7.6973	59.2486	(14)
IDADE	44.		662.5000	38.9706	8.1230	65.9835	(17)
IDADE	45.		787.5000	41.4474	7.7044	59.3582	(19)
IDADE	46.		341.5000	42.6875	7.7871	60.6384	(8)
IDADE	47.		507.5000	39.0385	6.0016	36.0192	(13)
IDADE	48.		473.0000	39.4167	5.2606	27.6742	(12)
IDADE	49.		637.0000	42.4667	4.6578	21.6952	(15)
IDADE	50.		892.0000	38.7826	6.0525	36.6324	(23)
IDADE	51.		625.0000	39.0625	6.4727	41.8958	(16)
IDADE	52.		557.0000	42.8462	6.6845	44.6827	(13)
IDADE	53.		354.0000	39.3333	4.7893	22.9375	(9)
IDADE	54.		632.5000	42.1667	4.8132	23.1667	(15)
IDADE	55.		516.0000	39.6923	7.7528	60.1058	(13)
IDADE	56.		498.0000	38.3077	5.8186	33.8558	(13)
IDADE	57.		531.0000	44.2500	6.4084	41.0682	(12)
IDADE	58.		701.5000	38.9722	5.1220	26.3374	(18)
IDADE	59.		311.0000	44.4286	6.7233	45.2024	(7)
IDADE	60.		1026.0000	41.0400	6.3195	39.9358	(25)
IDADE	61.		659.0000	38.7647	7.0203	49.2849	(17)
IDADE	62.		662.0000	41.4375	5.0526	25.5292	(16)
IDADE	63.		553.5000	42.5769	5.9796	35.1603	(13)
IDADE	64.		877.5000	43.8750	7.5809	57.4704	(20)
IDADE	65.		895.0000	42.6190	5.4724	29.9476	(21)
IDADE	66.		616.5000	44.0357	4.8416	23.4409	(14)
IDADE	67.		454.5000	41.3182	7.9758	63.6136	(11)
IDADE	68.		595.0000	39.6667	6.5210	42.5238	(15)
IDADE	69.		435.5000	43.5500	6.5847	43.3583	(10)
IDADE	70.		884.0000	42.0952	6.8312	46.6655	(21)
IDADE	71.		364.0000	40.4444	7.4559	55.5903	(9)
IDADE	72.		565.0000	40.3571	6.5909	43.4396	(14)
IDADE	73.		317.5000	39.6875	6.0234	36.2813	(8)
IDADE	74.		370.5000	37.0500	6.8935	47.5250	(10)
IDADE	75.		558.0000	42.9231	5.9401	35.2853	(13)
IDADE	76.		227.0000	37.8333	6.4317	41.3667	(6)
IDADE	77.		137.0000	45.6667	6.0277	36.3333	(3)
IDADE	78.		331.0000	41.3750	6.4754	41.9821	(8)
IDADE	79.		161.5000	40.3750	6.6254	43.8958	(4)
IDADE	80.		312.0000	44.5714	7.3848	54.5357	(7)
IDADE	81.		40.0000	40.0000	0.0000	0.0000	(1)
IDADE	82.		87.5000	43.7500	5.3033	28.1250	(2)
IDADE	83.		46.0000	46.0000	0.0000	0.0000	(1)
IDADE	84.		40.0000	40.0000	0.0000	0.0000	(1)
IDADE	86.		50.0000	50.0000	0.0000	0.0000	(1)
IDADE	88.		29.0000	29.0000	0.0000	0.0000	(1)
IDADE	91.		47.5000	47.5000	0.0000	0.0000	(1)

TOTAL CASES = 1091

TITLE IS 'DISPERSAO POR IDADE'

FILE NONAME (CREATION DATE = 07/12/83) 07/12/83 PAGE 2

D E S C R I P T I O N O F S U B P O P U L A T I O N S

CRITERION VARIABLE MEDIA
BROKEN DOWN BY IDADE

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			42568.0000	39.0174	6.9954	48.9350	(1091)
IDADE	1.		165.0000	27.5000	3.8210	14.6000	(6)
IDADE	2.		184.0000	30.6667	8.8015	77.4667	(6)
IDADE	3.		86.5000	28.8333	3.4034	11.5833	(3)
IDADE	4.		49.5000	24.7500	3.8891	15.1250	(2)
IDADE	5.		200.0000	28.5714	2.8347	8.0357	(7)
IDADE	6.		214.0000	30.5714	5.8197	33.8650	(7)
IDADE	7.		236.5000	29.5625	7.3700	54.3170	(8)
IDADE	8.		150.5000	30.1000	5.4245	29.4250	(5)
IDADE	9.		432.5000	36.0417	7.5602	57.1572	(12)
IDADE	10.		294.0000	32.6667	8.2272	67.6875	(9)
IDADE	11.		188.5000	31.4167	2.0104	4.0417	(6)
IDADE	12.		480.0000	34.2857	4.1031	16.8352	(14)
IDADE	13.		375.0000	34.0909	7.1162	50.6409	(11)
IDADE	14.		321.0000	40.1250	8.8186	77.7679	(8)
IDADE	15.		229.0000	38.1667	4.5350	20.5667	(6)
IDADE	16.		353.5000	35.3500	6.9524	48.3361	(10)
IDADE	17.		300.0000	37.5000	6.4087	41.0714	(8)
IDADE	18.		343.0000	38.1111	7.7533	60.7361	(9)
IDADE	19.		284.0000	40.5714	6.2479	39.0357	(7)
IDADE	20.		819.5000	39.0238	6.3629	40.4869	(21)
IDADE	21.		243.0000	34.7143	5.5891	31.2381	(7)
IDADE	22.		515.5000	36.8214	5.0559	25.5618	(14)
IDADE	23.		722.5000	36.1250	5.2862	27.9441	(20)
IDADE	24.		829.5000	37.7045	6.4229	41.2538	(22)
IDADE	25.		738.0000	38.8421	7.8706	61.9459	(19)
IDADE	26.		632.5000	37.2059	6.1924	38.3456	(17)
IDADE	27.		595.5000	39.7000	6.3914	40.8500	(15)
IDADE	28.		721.5000	37.9737	6.4560	41.6758	(19)
IDADE	29.		919.0000	39.9565	6.3352	40.1344	(23)
IDADE	30.		961.5000	40.0625	7.5575	57.1155	(24)
IDADE	31.		345.0000	34.5000	4.6963	22.0556	(10)
IDADE	32.		765.5000	38.2750	5.7617	33.1967	(20)
IDADE	33.		1221.5000	39.4032	7.3285	53.7070	(31)
IDADE	34.		631.5000	37.1471	6.5878	43.3989	(17)
IDADE	35.		854.5000	38.6513	6.6434	44.1354	(23)
IDADE	36.		1078.5000	39.9444	7.3253	53.6603	(27)
IDADE	37.		627.0000	39.1875	5.6976	32.4625	(16)
IDADE	38.		571.0000	35.6875	6.0852	37.0252	(16)
IDADE	39.		525.5000	37.5357	5.8719	34.4754	(14)
IDADE	40.		1055.5000	40.5962	5.9950	35.9404	(26)
IDADE	41.		1019.0000	37.7407	6.9826	48.7571	(27)