

IBGE/DIRETORIA TÉCNICA
SUPERINTENDÊNCIA DE ESTATÍSTICAS PRIMÁRIAS
DEPARTAMENTO DE COORDENAÇÃO DE MÉTODOS

METODOLOGIA DO
PLANO DE CONTROLE DE QUALIDADE DA CODIFICAÇÃO DO
CENSO AGROPECUÁRIO DE 1980

O DECME tem desenvolvido e aplicado planos de inspeção, por amostragem, da qualidade dos processos de preparo de dados, em diversas pesquisas da SUESP.

Para o controle da qualidade da codificação do Censo Agropecuário de 80, foi elaborado um plano mais adaptado às peculiaridades dessa pesquisa, e que se beneficia da experiência recente adquirida pelos técnicos do DECME/DICAV-S.1 durante o processo de codificação do Censo Demográfico de 80.

S U M Á R I O

	Pág.
1. Introdução	1
2. Considerações Gerais	2
3. Definições	3
4. Desenho da Amostra	3
5. Teste de Hipótese	5
5.1. Teste	5
5.2. Curva Característica de Operação	6
5.3. Curva de QMS (Qualidade Média de Saída)	7
6. Critérios para Classificação dos Revisores	7
7. Descrição do Funcionamento do Plano	8
ANEXO I: Curvas Características de operação dos dois níveis de inspeção, ambas para N= 200 e M= 80.	
ANEXO II-A: Curva de QMS para N= 200 e M= 80, da categoria A.	
ANEXO II-B: Curva de QMS para N= 200 e M= 80, da categoria B.	
ANEXO III: Fluxograma de funcionamento do plano.	
ANEXO IV: Folha de transcrição.	
ANEXO V: Manual de instruções do Controle de Qualidade da Codificação do Censo Agropecuário de 1980.	

1. INTRODUÇÃO

Visando garantir uma boa qualidade do material codificado e acelerar o processo de codificação, foi feito um controle estatístico de qualidade da codificação do Questionário Geral (CA 2.01) do Censo Agropecuário de 1980, através da implantação de um plano de inspeção por amostragem, por atributos.

O atual processo de codificação dos questionários do Censo Agropecuário de 80 consta de três fases: codificação, revisão e verificação. Essas fases são identificadas pelos códigos 11, 12 e 31, respectivamente.

Na fase de codificação (fase 11) são preenchidos os quadros ditos de controle (controle de página e controle do quadro) como também é feita uma crítica de valor e quantidade nas informações sobre produção. Além disso, algumas informações são totalizadas e transcritas para campos específicos.

Na fase de revisão (fase 21) são feitas, de forma visual, as mesmas tarefas da fase anterior (fase 11), corrigindo-se os erros encontrados.

Na fase de verificação (fase 31) é feita uma correção dos erros cometidos na fase anterior (fase 21), da mesma forma que os revisores corrigiram os erros cometidos na fase 11.

A alocação de um operador em uma das 3 fases (codificação, revisão ou verificação) é determinada tanto pela sua produção como pela qualidade do seu serviço.

O Plano de Controle de Qualidade do CA-80 foi elaborado de maneira a ser implantado na fase 31, ou seja, serão submetidos ao Plano os questionários que já tenham passado pelas fases 11 e 21.

Será feito um acompanhamento do Plano de forma a possibilitar considerações de algumas hipóteses feitas durante o seu desenvolvimento, e um posterior aprimoramento do mesmo.

2. CONSIDERAÇÕES GERAIS

Para efeito do Controle de Qualidade, os itens do questionário CA 2.01, a serem controlados, serão denominados de CAMPOS e classificados em 8 grupos. Cada grupo se caracteriza pela quantidade de campos por quadro. Na tabela a seguir estão os grupos com seus respectivos quadros:

GRUPO (CARACTERÍSTICA)	QUADRO(S)
I - Dois campos	03 - 05 - 07 - 08 - 09 - 10 11 - 12 - 13 - 14 - 16 - 17 18 - 19 - 20 - 22 - 23 - 25 26 - 27 - 32 - 33 - 35 - 36 37 - 41 - 42 - 44 - 46 - 47
II - Três campos	15 - 24 - 43 - 45
III - Cinco campos	29
IV - Seis campos	34
V - Oito campos	21 - 40
VI - Dez campos	02
VII - Quinze campos	06
VIII - Quadros com características especiais	28 - 30 - 31 - 38 - 39

Os quadros com características especiais (Grupo VIII) serão controlados de uma maneira diferente dos demais. Neles o controle será feito por produto, isto é, será computado um erro quando pelo menos uma informação referente ao produto estiver errada. A quantidade de produtos em cada um desses quadros depende do particular estabelecimento de onde provém o questionário, podendo, inclusive, ultrapassar a quantidade prevista no questionário.

Uma vez que os questionários são colocados em pastas, e como cada pasta é revisada por um único revisor, é conveniente considerar a unidade a ser submetida ao Plano como sendo uma pasta. Desta forma serão inspecionados os códigos dentro de cada pasta. É importante observar que uma pasta pode conter questionários de vários setores, embora estes sejam, necessariamente, de um mesmo município.

Considerando a qualidade de uma pasta em função da proporção de códigos errados contidos nela, será abordado o problema de decidir se uma pasta, depois de ter sido codificada, e revisada, é de BOA ou MÁ qualidade. Considera-se uma pasta como sendo de boa qualidade quando essa proporção é no máximo 1%. Um código será considerado errado quando o verificador fizer qualquer alteração no mesmo.

Para estimar a proporção de códigos errados na pasta, será selecionada uma amostra aleatória de campos preenchidos.

3. DEFINIÇÕES

Na pasta a ser submetida a inspeção, foram definidas as seguintes variáveis:

S = quantidade de setores contidos na pasta;

N_s = total de questionários do s -ésimo setor $1 < s < S$;

$N = \sum N_s$ = total de questionários contidos na pasta;

M = média de campos preenchidos por questionários, correspondente à Unidade da Federação a que a pasta pertence;

$$X_{ij} = \begin{cases} 1, & \text{se o } j\text{-ésimo campo do } i\text{-ésimo questionário está errado. } 1 < j < M \quad ; \quad 1 < i < n \\ 0, & \text{caso contrário.} \end{cases}$$

$X = \sum_i \sum_j X_{ij}$ = total de campos preenchidos errados;

$p = \frac{X}{NM}$ = proporção de campos preenchidos errados.

4. DESENHO DA AMOSTRA

4.1. Desenho da amostra

Para calcular o tamanho de amostra necessário para estimar a proporção de campos preenchidos errados em uma dada pasta, é necessário que se tenha o total de campos preenchidos em cada questionário da pasta. Para se obter essa informação seria preciso que se fizesse uma contagem exaustiva de todos os campos com informações para cada questionário, o que seria um processo pouco viável, dado o seu custo e o fator operacional. Uma vez que os questionários da pasta são de estabelecimentos geograficamente próximos, e baseando-se no fato de que as informações dos estabelecimentos de uma mesma Unidade da Federação não diferem em tipo e quantidade, atribuiu-se um mesmo valor para os totais de campos preenchidos nos questionários pertencentes a uma mesma Unidade da Federação. Este valor foi estimado com base em publicações do Censo Agropecuário de 1975, tomando-se, para cada Unidade da Federação, a média de campos preenchidos por questionários.

A fim de se proteger da ocorrência de alguns tipos de erros ditos sistemáticos, deseja-se que todos os setores contidos na pasta estejam representados na amostra. Esses erros são introduzidos, na maioria das vezes, pelo recenseador de cada setor, e ocorrem com maior frequência nos quadros do grupo VIII (quadros com características especiais). Entre os principais erros sistemáticos estão os de omissão de vírgula, colocação indevida de vírgula e erros em unidades de medida.

Considerando cada questionário como sendo um conglomerado de compos fez-se um estudo com os questionários já codificados e verificou-se que o coeficiente de correlação intraclasse não diferia de zero de maneira significativa. Isto faz com que a variância de \bar{p} (o estimador de p), numa amostra por conglomerados, seja igual a variância do mesmo estimador, quando é selecionada uma amostra aleatória simples.

Levando-se em conta o fator operacional, bem como todas as considerações feitas acima, decidiu-se utilizar uma amostra sistemática de questionários, onde cada setor da pasta fosse representado de maneira tal que, quanto maior a quantidade de questionários, do setor na pasta, maior a quantidade na amostra.

De uma amostra formada por n questionários, o quadrado do coeficiente de variação de \bar{p} é dado por:

$$(CV(\bar{p}))^2 = \frac{Np(1-p) \left(1 - \frac{n}{N}\right)}{(NM - 1) np^2}$$

Tendo em vista que uma pasta é considerada de qualidade aceitável quando contêm, no máximo, 1% de campos preenchidos errados, considerou-se $p = 0,01$. Admitindo-se um CV de 50% temos:

$$n = \frac{0,99 N}{0,0025 (NM-1) + 0,99}$$

Para ilustrar, na tabela 1 encontram-se os números médios de questionários selecionados por pasta para quatro UF's (PA, SC, PI e CE).

TAB.1 - Número de questionários selecionados por pasta

UF	M	n-MÉDIO
PA	86	5
SC	99	4
PI	87	5
CE	89	5

A seleção da amostra é feita via computador, segundo o seguinte esquema:

1 - Ordena-se $N_{(1)} < \dots < N_{(S)}$, define-se

$$K = \frac{N}{n} \text{ e } m_0 = N_{(1)};$$

2 - Se $k \leq m_0$, seleciona-se aleatoriamente um inteiro i_0 entre 1 e k , e forma-se a amostra com os questionários $i_0, k+i_0, \dots, (N-1)k+i_0$;

3 - Se $k > m_0$, seleciona-se aleatoriamente um questionário do setor que tem m_0 questionários e desprezam-se os outros questionários desse setor;

3 - Recalcula-se n , para a "nova pasta" que tem $(N - m_0)$ questionários de $(S - 1)$ setores e, volta-se para o passo 1.

5. TESTE DE HIPÓTESE

5.1. Teste

Sob o ponto de vista de inferência estatística, a hipótese de nulidade a ser formulada é $H_0: p = p_0$ contra a hipótese alternativa $H_1: p < p_0$. A hipótese assim formulada torna possível controlar o risco de quem utiliza os questionários codificados, e, por conseguinte, garante sua proteção.

A partir da fixação do nível de significância (em 5% e em 1%), que considerada a formulação do teste, corresponde a probabilidade de aceitação de material de má qualidade, foram calculadas várias curvas características de operação e verificou-se que para $p_0 = 0,02$ eram atingidas as melhores condições de viabilidade operacional.

O plano de inspeção de qualidade adotado aceita a pasta como sendo de boa qualidade quando a hipótese nula é rejeitada, pois a rejeição de H_0 significa que a pasta é de qualidade aceitável.

Na amostra de n questionários (que correspondia a nM campos) considera-se Y como sendo o total de campos preenchidos errados. A distribuição de probabilidade da variável aleatória Y é dada por:

$$P(y=k) = \begin{cases} \frac{\binom{nM}{k} \binom{NM - pNM}{nM - k}}{\binom{NM}{nM}}, & \text{se } 0 \leq k \leq \min(nM, pNM) \\ 0, & \text{caso contrário} \end{cases}$$

A média e a variância de y são dadas por:

$$\mu = nMp$$

$$\sigma^2 = \frac{nM^2 p(1-p)(N-n)}{(NM-1)}$$

Utilizando-se o teorema Central do limite, já que as condições desse teorema são satisfeitas pelas variáveis envolvidas, obtêm-se as seguintes aproximações:

$$P(T \leq c(\alpha)) = \Phi\left(Z_{c(\alpha)} + \frac{1}{2\sigma}\right)$$

e

$$c(\alpha) = \sigma\left(\Phi^{-1}(\alpha) - \frac{1}{2\sigma}\right) + \mu$$

onde:

$\Phi(\cdot)$ = função distribuição de uma $N(0,1)$

$$Z_{c(\alpha)} = \frac{c(\alpha) - \mu}{\sigma}$$

α = nível de significância adotado.

Utilizando-se os dados da Tab. 1, e considerando uma pasta contendo 200 questionários, os pontos críticos (ou pontos de rejeição) são dados na Tab. 2. Esses pontos foram calculados para $\alpha = 0,1$.

TAB.2 - PONTOS CRÍTICOS PARA $\alpha=0,1$ e $N=200$

UF	M	$c(0,10)$
PA	86	5
SC	99	4
PI	87	5
CE	89	5

Com o objetivo de acelerar o processo de codificação como um todo, é interessante que as pastas revistas por bons operadores sejam de forma diferente das demais. Pensando assim, adotou-se a política de inspecionar cada pasta em um dos dois níveis, 10% ou 5%. Se uma pasta tiver sido revisada por um bom operador, será inspecionada ao nível de 10%, caso contrário o nível de inspeção será de 5%.

Para efeito do presente Plano, os revisores considerados bons operadores são ditos pertencer à categoria A, os demais, à categoria B. Os critérios usados para essa classificação são descritos no item 6.

5.2. Curva Característica de Operação

Essa curva nos fornece, neste caso, a probabilidade de rejeitar a hipótese nula, dando que a proporção de erros na pasta era p , isto é, $P(p) = P(\text{Rejeitar } H_0/P)$.

Relembrando que uma rejeição da hipótese nula implica numa aceitação da pasta, a curva característica de operação fornece a probabilidade de aceitar uma pasta contendo $100p\%$ de campos preenchidos errados.

Como o controle vai ser composto por dois níveis de inspeção (correspondentes às categorias A e B) é preciso que se tenha uma Curva Característica para cada um.

Deve-se observar que cada Curva Característica depende do total de questionários na pasta, da UF de onde a pasta provém, bem como do nível de inspeção adotado. Por isto será feita uma Curva Característica de Operação para cada grupo de cruzamento das variáveis M , N e α .

No anexo I encontram-se as Curvas Características de Operação dos dois níveis de inspeção, ambas para $N=200$ e $M=80$.

5.3. Curva de QMS (Qualidade Média de Saída)

A curva de QMS (Qualidade Média de Saída) representa, como o próprio nome sugere, a qualidade média de saída das pastas que são submetidas ao Plano. Se uma pasta é submetida ao Plano, e contém $100p\%$ de campos errados, sua qualidade de saída é, em média, dada por $P(p)$.

Nos anexos II-A e II-B encontram-se as curvas de QMS para os níveis de inspeção que correspondem às duas categorias A e B. As duas curvas também foram feitas para $N = 200$ e $M = 80$.

6. CRITÉRIOS PARA CLASSIFICAÇÃO DOS REVISORES

A classificação nas categorias A ou B será feita quinzenalmente, considerando toda a produção até aquela data. Inicialmente todos os revisores devem ser classificados na categoria B (nível de inspeção = 5%).

Haverá mudança da categoria B para a categoria A (nível de inspeção=1%) se as seis últimas pastas de primeira ordem do revisor tiverem sido aceitas (isto é, sem considerar pastas de 2ª ordem).

A troca de categoria A para B será feita nos seguintes casos:

- se a produção do revisor na quinzena tiver sido menor do que 10 pastas, será feita uma consulta às dez últimas pastas de 1ª ordem revisadas até então. Se entre essas dez pastas, duas foram rejeitadas numa primeira inspeção, a categoria da próxima quinzena será B.

- Se a produção da quinzena tiver sido maior ou igual a dez pastas de 1ª ordem, a categoria será B se o número de pastas rejeitadas for maior ou igual a 40% do total das pastas.

Será desenvolvido um sistema de acompanhamento da produção que, entre outras coisas, manterá um cadastro da produção de cada revisor de forma a poder classificá-lo, ao final de cada quinzena, como pertencente à categoria A ou à categoria B.

7. DESCRIÇÃO DO FUNCIONAMENTO DO PLANO

A implantação do presente Plano de Controle de Qualidade requer a criação de novos grupos de trabalho (GI, GII, GIII), que hão de compor as fases de COLOCAÇÃO DAS ETIQUETAS, VERIFICAÇÃO POR AMOSTRAGEM e VERIFICAÇÃO TOTAL. Quando uma pasta é submetida ao plano, deve seguir o fluxo mostrado no Anexo III, que é explicado no que se segue.

Uma pasta é dita ser de 2ª ordem se ela já passou pelo plano e está sendo submetida novamente. A identificação deve ser feita na fase de COLOCAÇÃO DAS ETIQUETAS. As etiquetas foram confeccionadas na DICAV, com base no cadastro do empastamento, fornecido pela DI.

DESCRIÇÃO DAS NOVAS FASES

7.1. COLOCAÇÃO DAS ETIQUETAS

- 7.1.1. Se a pasta NÃO for de 2ª ordem, uma etiqueta deve ser colocada na capa da pasta. Essa etiqueta contém o número da pasta, os pontos críticos (C_A e C_B) e os números dos questionários que foram selecionados para a amostra.
- 7.1.2. Se a pasta for de 2ª ordem, será colocada uma nova etiqueta contendo no vos pontos críticos e novos números dos questionários da amostra.
- 7.1.3. Será consultada uma listagem contendo a categoria de cada revisor. De acordo com a classificação de quem fez a revisão da pasta, como sendo de categoria A ou da categoria B, deve ser assinalado um dos dois valores, C_A ou C_B.

7.2. VERIFICAÇÃO POR AMOSTRAGEM

- 7.2.1. Os questionários contidos na etiqueta colocada na fase anterior devem ser verificados exaustivamente.
- 7.2.2. Será preenchida uma folha de transcrição idêntica à do Anexo IV.
- 7.2.3. Se a quantidade de campos errados for maior do que o valor assinalado em 7.1.3, a pasta deve ser rejeitada e seguir para a fase de VERIFICAÇÃO TOTAL. Caso contrário aceita-se a pasta.

7.3. VERIFICAÇÃO TOTAL

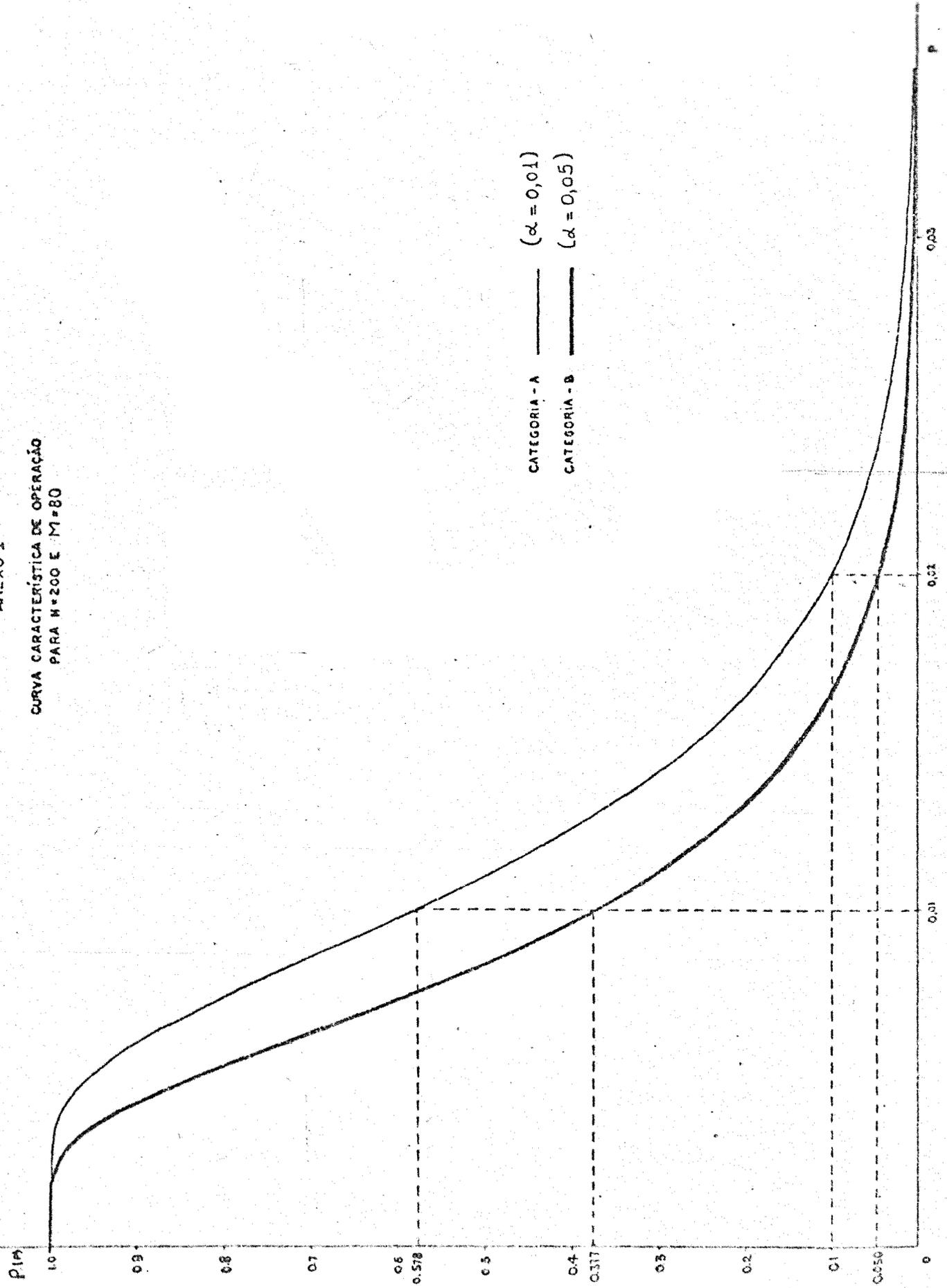
7.3.1. Se a pasta rejeitada não for de 2.^a ordem, todos os questionários devem ser verificados exhaustivamente. Em seguida será preenchida uma folha de transcrição idêntica à do Anexo IV e submeter a pasta novamente ao Controle de Qualidade

7.3.2. Se a pasta rejeitada for de 2.^a ordem, deverá seguir para um Processo Especial, a ser decidido pelo DEAGRO, e não ser mais submetida ao plano.

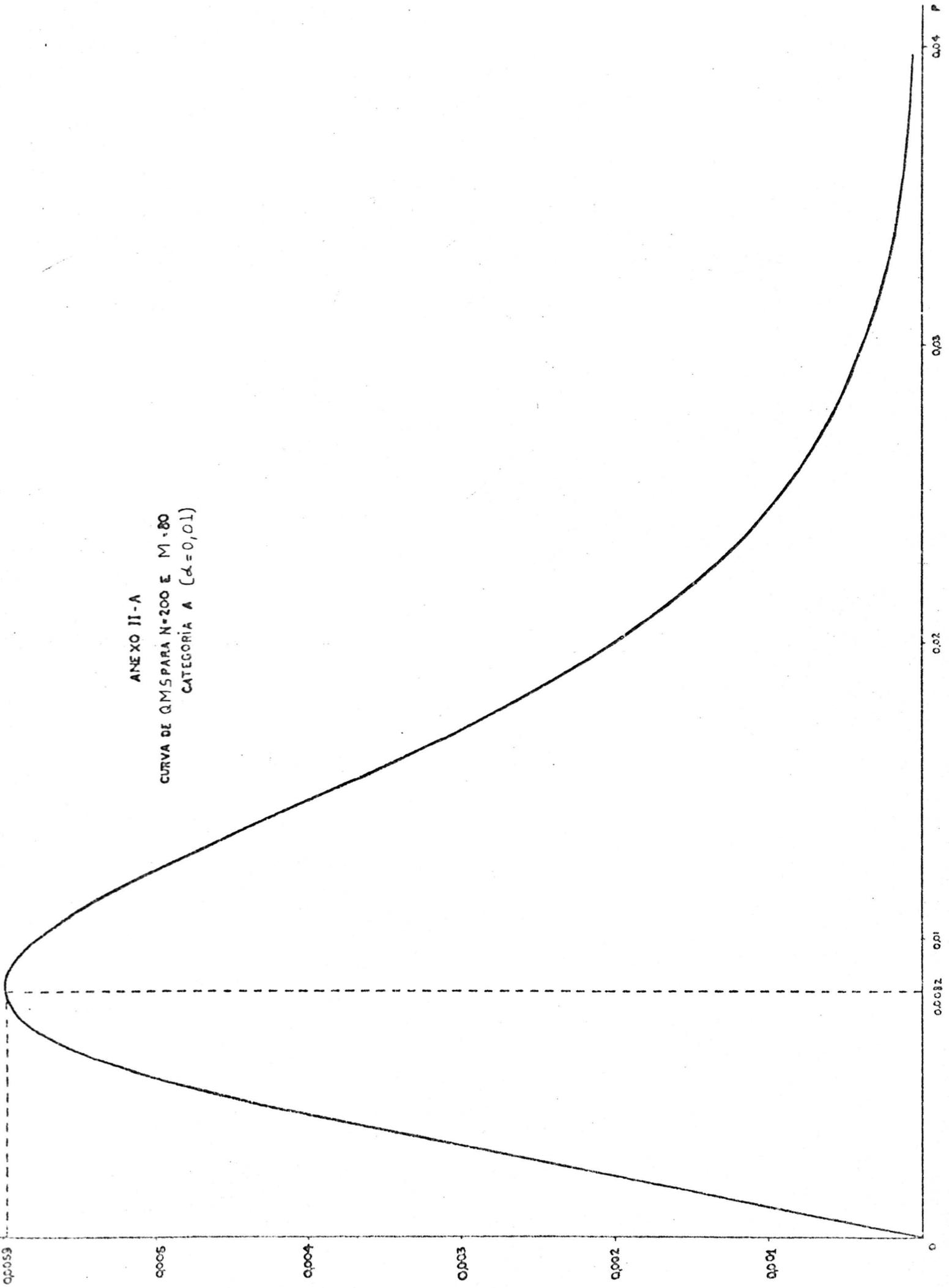
Como as consequências de erros no quadro 2 são, em geral, mais graves do que as dos demais quadros, estabeleceu-se que uma pasta mesmo sendo aceita, mas que contenha algum erro no quadro 2, vai para uma verificação 100%, no quadro 2.

ANEXO I

CURVA CARACTERÍSTICA DE OPERAÇÃO
PARA $M=200$ E $M=80$

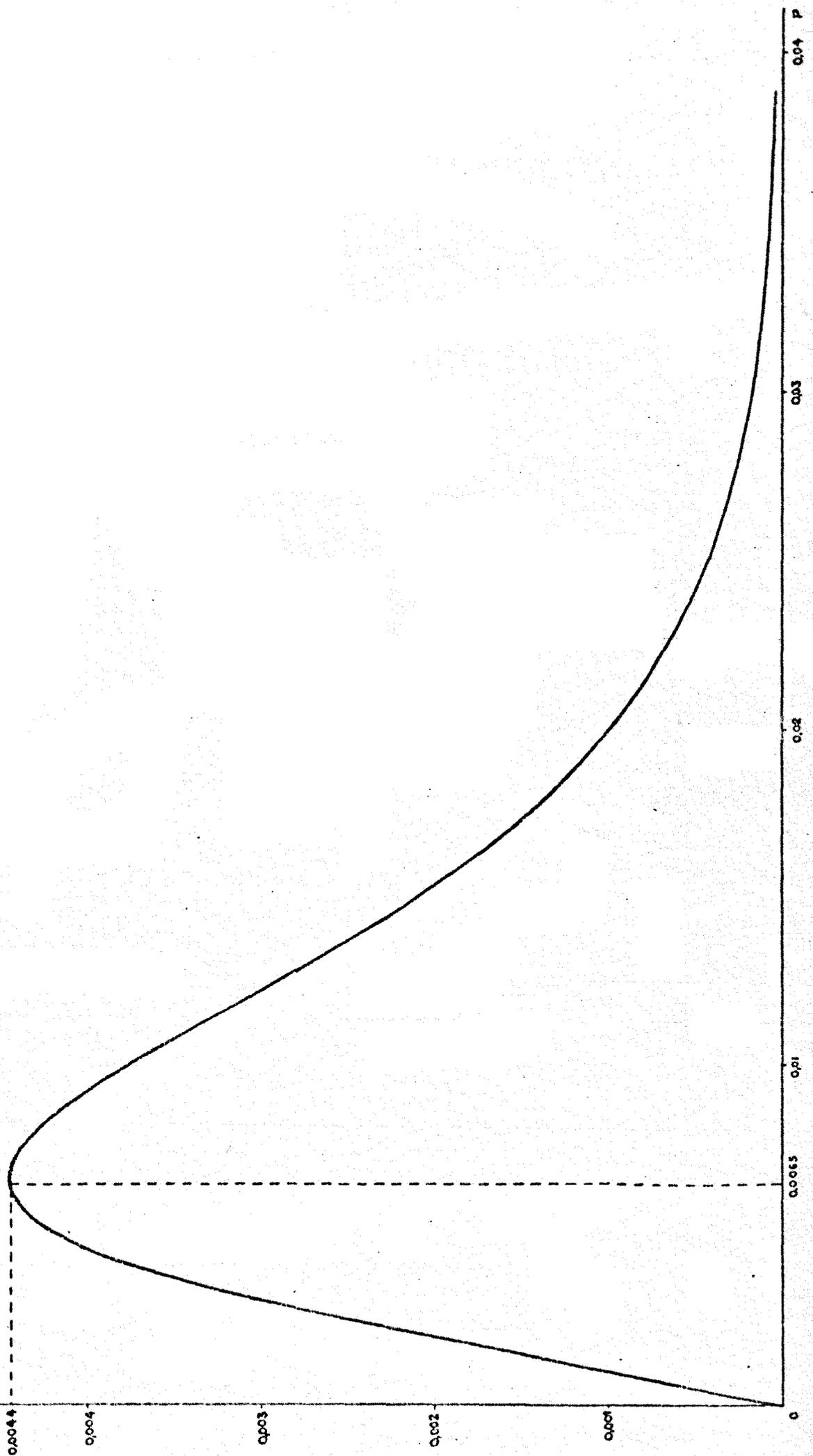


ANEXO II-A
CURVA DE QMS PARA N=200 E M=80
CATEGORIA A ($\alpha=0,01$)



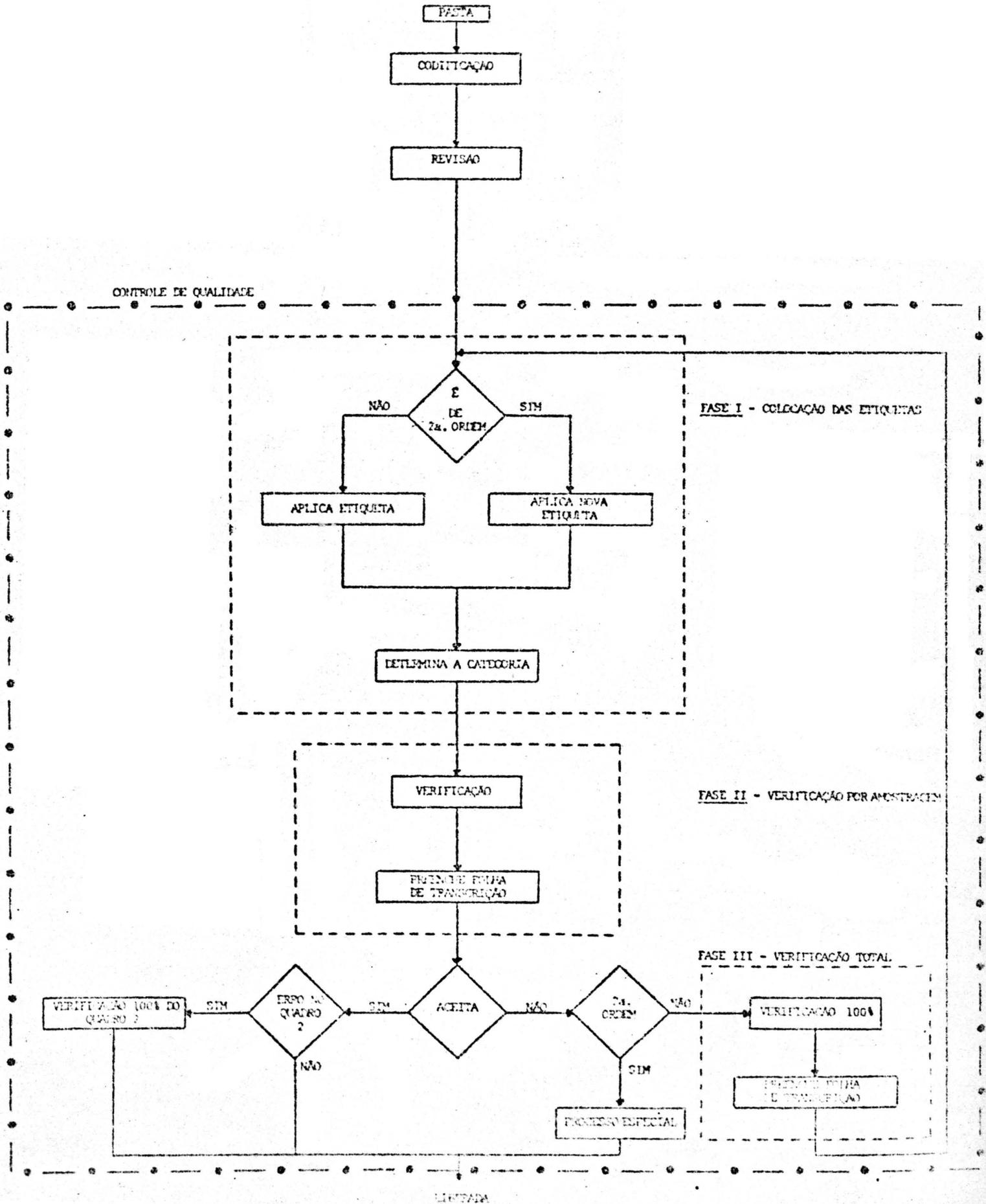
QMS

ANEXO II-B
CURVA DE QMS PARA N=200 E M=80
CATEGORIA B ($\alpha=0.05$)



ANEXO III

FLUXOGRAMA DE FUNCIONAMENTO DO PLANO



ANEXO V

DIRETORIA TÉCNICA
SUPERINTENDÊNCIA DE ESTATÍSTICAS PRIMÁRIAS
DEPARTAMENTO DE COORDENAÇÃO DE MÉTODOS

MANUAL DE INSTRUÇÕES
DO CONTROLE DE QUALIDADE DA CODIFICAÇÃO
DO CENSO AGROPECUÁRIO DE 1980

MAIO/1982
DICA-V-S.1

INTRODUÇÃO

O Plano de Controle de Qualidade da Codificação do CA-80, doravante referido como Controle, prevê a criação de 3 grupos de trabalho que executam as atividades das fases distintas; Colocação das Etiquetas, Verificação por Amostragem e Verificação Total.

Pretende-se, neste documento, expor de forma direta as atividades inerentes a cada um desses 3 grupos, como também explicar o funcionamento do Controle, sem se preocupar, entretanto, com os conceitos teóricos por ele envolvidos.

DESCRIÇÃO

Uma pasta, quando é submetida ao Controle, percorre o fluxo mostrado no Anexo III. As atividades foram separadas por grupos, que compoem as 3 fases do controle. Essas atividades serão aplicadas, segundo a ordem cronológica.

FASE I - Colocação das Etiquetas

1.a) Se a pasta não for de 2.^a ordem (ainda não tem a etiqueta do Controle), será colocada, na sua capa, a etiqueta da 1.^a Amostra, semelhante à da fig. 1. Estando as 18 posições dessa etiqueta ocupadas, o operador deve observar a etiqueta seguinte e colocá-la abaixo da anterior, caso esta seja da mesma pasta.

1.b) Se a pasta for de 2.^a ordem, ou seja, se já tem etiqueta(s) do Controle, a(s) etiqueta(s) da 2.^a Amostra deve(m) ser sobreposta(s) sobre a(s) existente(s). No caso da 1.^a amostra ter mais etiquetas do que a 2.^a amostra, deverão ser anuladas as excedentes.

PASTA - 179 103			CA- 6	CB- 5	
10	59	108	157	185	

FIG. 1 - Etiqueta do Plano de Controle de Qualidade

2) Será consultada uma listagem contendo a categoria em que se encontra cada Revisor. Deverá ser assinalado um dos valores, CA ou CB, conforme o revisor da pasta esteja na categoria A ou B, respectivamente.

FASE II - Verificação por Amostragem

Nesta fase, serão verificados, de forma exaustiva, os questionários cujos números estão na etiqueta do Controle, corrigindo-se os erros encontrados. Se o total de erros encontrados for maior do que CA ou CB, conforme o valor que foi assinalado na FASE I, a pasta será REJEITADA, caso contrário ela será ACEITA.

Se a pasta for aceita e tiver pelo menos um erro no quadro 2, deverá ser submetida a uma verificação total desse quadro, e ser liberada em seguida.

As pastas rejeitadas, sendo de 2^a ordem, deverão seguir para um proceso especial, a ser definido pelo DEAGRO. Caso sejam de 1^a ordem irão, para a FASE III, que é a fase de Verificação Total.

Para cada pasta, de 1^a ou 2^a ordem, deve ser preenchida uma linha da Folha de Transcrição (Anexo IV), com as seguintes informações:

- Código do verificador
- Código do revisor
- Código do codificador
- Nº da pasta
- Ordem
- Categoria do revisor
- Nº de erros na pasta
- Total
- No quadro 02
- Condição da pasta

A contagem de erros, para efeito do preenchimento da folha de transcrição, deve ser feita segundo o esquema descrito a seguir.

1) Nos quadros 02 e 03, bem como, nos quadros de controle (06, 21, 29, 34 e 40), deverá ser computado um erro para cada correção feita pelo verificador.

2) Nos quadros 15, 24, 43 e 45 os erros deverão ser computadas da seguinte forma:

2.1 - Computa-se um erro se for feita qualquer correção no campo de controle, como também no quesito 01;

2.2 - Computa-se um erro se pelo menos uma correção for feita no restante do quadro.

3) Nos quadros 28, 30, 31, 38 e 39 os erros deverão ser computados da seguinte forma:

3.1 - Computa-se um erro se for feita alguma correção no campo de controle;

3.2 - Computa-se um erro se, nas informações de um produto, for feita pelo menos uma correção.

4) Nos demais quadros, o cômputo dos erros é feito como segue:

4.1 - Computa-se um erro se for feita alguma correção no campo de controle;

4.2 - Computa-se um erro se pelo menos uma alteração for feita no restante do quadro.

FASE III - Verificação Total

Nesta fase serão verificados, de forma exaustiva, todos os questionários da pasta, e será preenchida, para cada pasta, uma linha de transcrição. Para efeito do preenchimento da folha de transcrição, os erros devem ser computados na mesma forma que na FASE II. Em seguida, a pasta deve ser submetida novamente ao Controle.