

# As questões irrelevantes e a categorização das estratégias de Teste de Hipóteses acerca dos outros: Um (novo) enviesamento do investigador?

## – Exemplo de aplicação do coeficiente *Kappa*

ANA CRISTINA MARTINS (\*)

### INTRODUÇÃO

Este artigo consiste na proposta de um procedimento alternativo ao usualmente adoptado na categorização das questões formuladas nos estudos sobre a forma como procuramos informação acerca dos outros ao pretendermos saber se estes possuem ou não determinadas características. Sendo este processo comumente designado, na literatura, por teste de hipóteses, convém, desde já, sublinhar que o uso deste termo ao longo do presente artigo nada tem a ver com o igualmente usual teste de hipóteses estatístico.

Os domínios em que estes trabalhos têm sido

desenvolvidos são variados, indo desde o julgamento social, até à investigação científica, passando pelo raciocínio lógico (*e.g.*, Klayman & Ha, 1987, 1989) e pelas questões intergrupais (*e.g.*, Leyens & Yzerbyt, 1992). Dado o objectivo do presente artigo, são, aqui, referidos, somente, alguns dos que respeitam ao teste de hipóteses acerca dos outros, os quais têm incidido sobre os traços de personalidade.

### O TESTE DE HIPÓTESES SOBRE A PERSONALIDADE

Testar uma hipótese relativa à existência de determinado traço numa pessoa implica a recolha e interpretação de informação a seu respeito. Para tal, o indivíduo que testa a hipótese coloca, ao sujeito alvo, uma série de questões às quais este responde. É, então, com base na infor-

---

(\*) Instituto Superior de Psicologia Aplicada. Membro da Unidade de Investigação em Psicologia Social, ISPA.

mação assim adquirida, que concluirá acerca da sua personalidade (Zuckerman, Knee, Hodgins & Miyake, 1995). Um dos estudos pioneiros nesta área foi desenvolvido por Snyder e Swann (1978). Os autores colocaram uma hipótese sobre um traço de personalidade de determinada pessoa (introversão ou extroversão) aos sujeitos que integraram a sua investigação, tendo-lhes pedido, após a disponibilização da descrição do perfil de uma pessoa com um traço e com o outro, que seleccionassem 12 questões de uma lista, no sentido de testarem tal hipótese. Esta lista continha 26 questões sobre crenças, sentimentos e acções das pessoas, as quais haviam sido previamente classificadas por 9 juízes (com base na regra da maioria): 11 perguntavam sobre características da introversão (por exemplo, «Que factores lhe dificultam o expôr-se aos outros?»), 10 sobre características da extroversão (por exemplo, «Que tipo de situações procura se pretende conhecer novas pessoas?») e 5 eram neutras relativamente à dimensão em causa, isto é, tratavam-se de questões em relação às quais não houve consenso entre os codificadores ou que foram por estes classificadas, maioritariamente, como irrelevantes tanto para a introversão como para a extroversão (por exemplo, «Quais são os seus objectivos em termos de carreira?»).

Snyder e Swann verificaram que os sujeitos seleccionavam diferentemente as questões disponíveis, consoante a hipótese que lhes havia sido dada a testar. Assim, enquanto que aqueles que testaram a hipótese da introversão seleccionaram, em maior número, questões que analisavam a introversão, aqueles que testaram a hipótese da extroversão escolheram mais questões que exploravam esta característica. As questões neutras foram escolhidas com igual frequência pelos sujeitos que testavam uma ou outra das hipóteses.

Os autores concluíram, então, que os sujeitos haviam adoptado uma estratégia de confirmação da hipótese, a qual se verificou mesmo quando receberam informação adicional de que era improvável que o alvo possuísse o traço hipotetizado e quando lhes foi prometida uma recompensa monetária pela exactidão na realização da tarefa proposta (Snyder & Swann, 1978, investigações 3 e 4).

Com o intuito de responderem a algumas das críticas tecidas ao trabalho de Snyder e Swann,

concretamente, 1) o facto de os seus estudos não se debruçarem sobre as estratégias adoptadas espontaneamente, uma vez que foi fornecida, aos sujeitos, uma lista de questões já formuladas e 2) o facto de os resultados poderem não ser generalizáveis a outras dimensões para além da introversão-extroversão, Swann e Giuliano (1987) realizaram um estudo no qual os sujeitos tinham de preparar as suas próprias questões no sentido de verificarem se o respondente correspondia a um perfil previamente fornecido. Fizeram-no para a dimensão dominância-submissão e para a introversão-extroversão. As questões assim geradas foram, posteriormente, codificadas por dois juízes, tendo sido calculada a percentagem de situações relativamente às quais estes estiveram de acordo (93%). Nos casos de empate, os autores socorreram-se de uma terceira pessoa.

As categorias resultantes foram as seguintes:

1. *Questões confirmatórias* – Questões às quais uma resposta afirmativa indicaria a presença do traço hipotetizado. Por exemplo, «Você assume, usualmente, a liderança num grupo de pessoas?»
2. *Questões infirmativas* – Questões às quais uma resposta afirmativa indicaria a presença do traço oposto ao hipotetizado. Por exemplo, «Você é vacilante?».
3. *Questões de igual oportunidade* – Questões que davam, aos respondentes, a oportunidade de se descreverem em termos do traço hipotetizado ou do seu oposto. Por exemplo, «Você prefere passar uma noite calma em casa ou ir a um clube ou a uma festa?».
4. *Questões irrelevantes* – Questões sem relação óbvia com a hipótese nem com a alternativa. Por exemplo, «O que considera ser a família ideal?».

Os resultados obtidos foram indicativos da formulação, por parte dos sujeitos, de um maior número de questões confirmatórias e infirmativas do que de igual oportunidade e irrelevantes. Para ambas as dimensões, os participantes solicitaram mais evidência sobre o traço hipotetizado, tendo sustentado os estudos de Snyder e Swann.

Apesar de o presente artigo não consistir nu-

ma revisão de literatura sobre este tipo de investigações, é de referir que os estudos que se seguiram aos descritos passaram a usar as questões formuladas espontaneamente pelos sujeitos, tendo sido encontradas estratégias diferentes.

Assim, Trope e Bassok (1982) encontraram uma estratégia diagnóstica, ou seja, «(...) to test the hypothesis that a target person belongs to one trait-category rather than to another, the information gatherer asks about features that maximally discriminate between the two categories» (p. 30), posteriormente corroborada pelo trabalho de Trope, Bassok e Alon (1984).

Mais tarde, Devine, Hirt e Gehrke (1990) verificaram que as estratégias isoladas até então não eram mutuamente exclusivas.

Tal como Swann e Giuliano (1987), estes últimos encontraram, igualmente, questões irrelevantes, ou seja, questões sem qualquer relação óbvia com o traço hipotetizado nem com o traço alternativo. A sua codificação foi feita por dois juízes, tendo sido calculada a percentagem de classificações em que estes estiveram de acordo (85%) e feito o desempate por um terceiro elemento.

De acrescentar que, tal como referido pelos mesmos, algumas das questões de resposta sim/não encontradas eram ambíguas, não sendo claro se se relacionavam com o traço hipotetizado ou com o alternativo.

Parece, pois, estar-se perante um procedimento de classificação que pode conduzir a alguns enviesamentos e perda de informação eventualmente pertinente, dado os sujeitos formularem questões que, por um lado, embora possam ser pertinentes para si, acabam por ser categorizadas como neutras ou irrelevantes e que, por outro, levantam dúvidas quanto à característica do alvo sobre a qual incidem. Considera-se que este problema ganhou especial relevância quando, num estudo realizado por Martins (1996) sobre o teste de hipóteses em selecção de pessoal, os codificadores encontraram uma maioria de questões irrelevantes, relativamente às quais manifestaram 100% de acordo.

#### NA SENDA DE UMA ALTERNATIVA

Tendo partido do pressuposto de que todas as questões geradas pelos sujeitos aquando da

procura de informação acerca dos outros têm um objectivo para os mesmos, possivelmente decorrente das suas teorias implícitas (Bruner & Tagiuri, 1954), ainda que esse não seja claro para os codificadores, Martins (1995) utilizou um procedimento alternativo que consistiu na categorização de cada questão com base no seu objectivo, explicitado pelo sujeito que a colocou.

O seu estudo incidiu sobre a selecção de uma candidata a recepcionista de hotel e envolveu as seguintes categorias, isoladas pela investigadora a partir da análise de conteúdo por si efectuada às 824 questões formuladas pelos sujeitos.

1. *Questões de resposta sim/não acerca da adequação da candidata à função:* Questões às quais a candidata podia responder «sim» ou «não» e que, perante uma resposta afirmativa, levavam a concluir que determinado aspecto da candidata era adequado à função. Exemplos: Gosta de se dar com pessoas, de conviver? Já trabalhou no ramo hoteleiro?
2. *Questões de resposta sim/não acerca da inadequação da candidata à função:* Questões às quais a candidata podia responder «sim» ou «não» e que, perante uma resposta afirmativa, levavam a concluir que determinado aspecto da candidata não era adequado à função. Exemplos: É casada? Cederia a pedidos ou «chantagens», mesmo que pusessem em causa o seu emprego, apenas por pena ou sentimentos de culpa?
3. *Questões de resposta sim/não acerca da adequação da candidata à função, formuladas pela negativa:* Questões às quais a candidata podia responder «sim» ou «não» e que, perante uma resposta negativa, levavam a concluir que determinado aspecto da candidata era adequado à função. Exemplos: Não se irrita facilmente, mesmo que às vezes as pessoas sejam um pouco inoportunas? Não se importa de trabalhar por turnos?
4. *Questões de resposta sim/não acerca da inadequação da candidata à função formuladas pela negativa:* Questões às quais a candidata podia responder «sim» ou «não» e que, perante uma resposta negativa, levavam a concluir que determinado

- aspecto da candidata não era adequado à função. Exemplos: Não frequenta ou não pretende frequentar um curso de outra língua estrangeira? Não pretende adquirir mais nenhum grau académico?
5. *Questões bidireccionais*: Questões que solicitavam uma de duas alternativas de resposta em si contidas, uma a favor da adequação da candidata à função e outra a favor da sua não adequação. Exemplos: Gostaria de executar este trabalho ou é apenas pela recompensa monetária? É uma pessoa calma ou aborrece-se facilmente?
  6. *Questões abertas*: Questões que impediam uma resposta sim/não, levando a candidata a explicar sobre algo. Exemplos: Porque escolheu vir trabalhar como recepcionista de hotel? Quais são as suas expectativas?
  7. *Afirmações imperativas*: Afirmações que solicitavam à candidata que falasse sobre algo. Exemplos: Fale das experiências de trabalho anteriores. Defina a sua personalidade.
  8. *Questões abertas enviesadas a favor da adequação da candidata à função*: Questões abertas que assentavam no pressuposto da adequação de determinados aspectos da candidata à função ou que geravam uma sobre-representação destes aspectos. Exemplos: Quais são os seus pontos fortes? Quais as aptidões que possui e que estão adequadas à função?
  9. *Questões abertas enviesadas a favor da inadequação da candidata à função*: Questões abertas que assentavam no pressuposto da não adequação de determinados aspectos da candidata à função ou que geravam uma sobre-representação destes aspectos. Exemplos: Quais as suas maiores dificuldades que considera importantes e relevantes? Que problemas teve ao longo da sua experiência profissional?
  10. *Questões mistas acerca da adequação da candidata à função*: Questões que se iniciavam em formato aberto, indicando sugestões de resposta sim/não, às quais uma resposta afirmativa levava a concluir que determinado aspecto da candidata era adequado à função. Exemplo: Qual o seu objectivo ao concorrer a esta vaga, mais experiência neste campo específico e um emprego com estabilidade?
  11. *Questões mistas acerca da inadequação da candidata à função*: Questões que se iniciavam em formato aberto, indicando sugestões de resposta sim/não, às quais uma resposta afirmativa levava a concluir que determinado aspecto da candidata não era adequado à função. Exemplo: Porque motivo pretende mudar de emprego, não gosta da presente função?
  12. *Questões mistas bidireccionais*: Questões que se iniciavam em formato aberto, indicando duas alternativas de resposta, uma a favor da adequação da candidata à função e outra a favor da sua não adequação. Exemplo: O que representa para si um cliente, só mais um para atender ou um novo desafio?
  13. *Afirmações de adequação da candidata à função*: Afirmações feitas à candidata a respeito da adequação de determinados aspectos seus à função e que induziam um comentário a tal respeito. Exemplo: Pela sua experiência profissional, parece-me que gosta de contactar com os outros.
  14. *Afirmações de inadequação da candidata à função*: Afirmações feitas à candidata a respeito da não adequação de determinados aspectos seus à função e que induziam um comentário a tal respeito. Exemplo: Como é uma pessoa tensa, rígida e introvertida deve ter dificuldades em lidar com pessoas que não conhece.
- A categorização das questões foi feita por 2 juízes. A fim de aumentar a sua exactidão, e com base no trabalho de Rosenberg e Levy (1955), os protocolos e o sistema de categorias previamente criados foram apresentados conjuntamente com uma série de instruções escritas que se pretendiam precisas e facilitadoras de uma codificação o mais padronizada possível. Estas consistiram, fundamentalmente, na descrição da tarefa que havia sido dada aos sujeitos e na definição de cada uma das categorias, tendo-se alertado os juízes para o facto de que a sua classificação deveria seguir os objectivos estipulados pelos sujeitos para cada uma das questões e não aquilo que eles próprios consideravam ser procurado pelas mesmas.

No final da codificação efectuada pelos juízes, foi-lhes perguntado se se haviam sentido «confortáveis» com as categorias apresentadas ou se, pelo contrário, estas se haviam revelado insuficientes, não mutuamente exclusivas ou inúteis. A sua avaliação das mesmas foi favorável.

Para cálculo do acordo inter-juízes foi utilizado o indicador *Kappa* (ver, e.g., Siegel & Castellan, Jr., 1989). Recorde-se que aquilo que lhes havia sido solicitado fora a atribuição de cada questão a cada uma das categorias disponíveis. Pretendia-se, agora, saber qual o grau de acordo entre os juízes quanto à categoria a que pertenceria cada uma das questões. Tratavam-se, pois, de classificações nominais, das quais a estatística *Kappa* é uma das medidas de acordo possíveis, aplicável a um número arbitrário de juízes.

Para o efeito, foi atribuído um número a cada uma das categorias, tendo sido, então, possível construir uma tabela semelhante à Tabela 1, em que *N* é o número total de questões e *C<sub>i</sub>* traduz o número da categoria em que cada um dos dois juízes colocou cada uma das questões.

TABELA 1  
*Classificações atribuídas por dois juízes a N questões*

QUESTÕES	CATEGORIAS	
	Juiz 1	Juiz 2
1	<i>C<sub>i</sub></i>	<i>C<sub>i</sub></i>
2	<i>C<sub>i</sub></i>	<i>C<sub>i</sub></i>
3	<i>C<sub>i</sub></i>	<i>C<sub>i</sub></i>
•	•	•
•	•	•
•	•	•
<i>N</i>	<i>C<sub>i</sub></i>	<i>C<sub>i</sub></i>

Os dados foram tratados com base no *package* informático SPSS/PC<sup>+</sup>, versão 4.0, tendo-se obtido um *K*=.92.

Os pontos de desacordo foram desempatados

por um terceiro juiz, ou seja, foi-lhe dada a mesma tarefa, tendo sido a sua classificação aquela que prevaleceu.

Na impossibilidade da utilização deste *package* informático, ou de algum alternativo que proceda ao cálculo desta estatística, Siegel e Castellan, Jr. (1989) apresentam o modo de o fazer recorrendo-se à respectiva fórmula:

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

em que

*P(A)* = proporção de vezes em que os *k* juízes estiveram de acordo

*P(E)* = proporção de vezes que seria de esperar que os *k* juízes estivessem de acordo devido ao acaso

O coeficiente *K* é, pois, a razão do número de vezes que os juízes acordaram (corrigido para o acordo devido ao acaso) para o número máximo de vezes que poderiam estar de acordo (igualmente corrigido para o acaso).

Em caso de acordo total entre os juízes obter-se-á um *K*=1; pelo contrário, se não houver qualquer acordo entre eles, exceptuando aquele que seria de esperar pelo acaso, obter-se-á um *K*=0.

Vejam os como proceder, tomando como exemplo o caso em questão.

1. Cálculo do número de vezes em que os juízes atribuíram cada questão a cada categoria, sendo *N* o número de questões (824), *m* o número de categorias (14) e *k* o número de juízes (2).

Construir, para o efeito, uma tabela *N x m*, em que o total de cada linha deve corresponder ao número total de juízes (Tabela 2).

Note-se que, relativamente aos dados constantes da Tabela 2, os juízes estão de acordo quanto a todos os objectos, à excepção da questão 823, dado que um juiz a incluiu na categoria 1 e o outro juiz a integrou na categoria 6.

TABELA 2

QUESTÕES	CATEGORIAS													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	2	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	2	-	-	-	-	-	-	-	-
3	2	-	-	-	-	-	-	-	-	-	-	-	-	-
4	2	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	2	-	-	-	-	-	-	-	-
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
819	-	-	-	2	-	-	-	-	-	-	-	-	-	-
820	-	-	-	-	-	-	2	-	-	-	-	-	-	-
821	-	-	-	-	-	2	-	-	-	-	-	-	-	-
822	-	-	-	-	-	2	-	-	-	-	-	-	-	-
823	1	-	-	-	-	1	-	-	-	-	-	-	-	-
824	-	-	-	-	-	2	-	-	-	-	-	-	-	-
$C_j$	430	63	15	8	33	1017	58	6	8	3	2	1	2	2
$P_j$	.261	.038	.009	.005	.020	.617	.035	.004	.005	.002	.001	.001	.001	.001

2. Cálculo do número de vezes que foram atribuídas questões a cada categoria ( $C_j$ ), o qual consiste na soma das frequências de cada coluna.

3. Cálculo da proporção de atribuições feitas a cada uma das categorias ( $P_j$ ), com base na seguinte fórmula:

$$P_j = \frac{C_j}{Nk}$$

Assim, para a Categoria 1 temos:

$$P_1 = \frac{C_1}{Nk} = \frac{430}{824 \times 2} = .261$$

Os restantes valores, assim calculados, encontram-se na Tabela 2.

4. Cálculo do acordo inter-juízes esperado se a sua classificação das questões houvesse sido feita ao acaso ( $P(E)$ ), sendo que:

$$P(E) = \sum_{j=1}^m p_j^2$$

Para o exemplo em questão temos:

$$P(E) = .2612 + .0382 + .0092 + .0052 + .0202 + .6172 + .0352 + .0042 + .0052 + .0022 + .0012 + .0012 + .0012 + .0012 = .451$$

5. Cálculo da proporção geral de acordo  $P(A)$ :

$$P(A) = \left[ \frac{1}{Nk(k-1)} \sum_{i=1}^N \sum_{j=1}^m n_{ij}^2 \right] - \frac{1}{k-1}$$

No caso presente será:

$$P(A) = \frac{1}{(824(2)1)} (2^2 + 2^2 + 2^2 + \dots + 2^2 + 1^2 + 1^2 + \dots)$$

$$+2^2) - \frac{1}{2-1} = \frac{1}{1648} \times 3224 - 1 = .956$$

6. Finalmente, procede-se ao cálculo do K:

$$K = \frac{P(A) - P(E)}{1 - P(E)} = \frac{.956 - .451}{1 - .451} = .92$$

Resta, ainda, testar a significância do valor K obtido, ou seja, testar a hipótese  $H_0: K=0$  contra a  $H_1: K>0$ . Para tal, dever-se-á proceder ao:

7. Cálculo da variância de K:

$$var(K) = \frac{2}{Nk(k-1)} \times \frac{P(E) - (2k-3)[P(E)]^2 + 2(k-2)\sum p_j^3}{[1 - P(E)]^2}$$

$$\sum p_j^3 = .261^3 + .038^3 + .009^3 + .005^3 + .005^3 + .020^3 + .617^3 + .035^3 + .004^3 + .005^3 + .002^3 + .001^3 + .001^3 + .001^3 + .001^3 = .253$$

$$var(K) = \frac{2}{824(2)(1)} \times \frac{.451 - [2(2)-3] (.451)^2 + 2(2-2)(.253)}{(1 - .451)^2}$$

$$= \frac{2}{1648} \times \frac{.248}{.301} = .0009998$$

8. Cálculo do valor z correspondente:

$$z = \frac{K}{\sqrt{var(K)}} = \frac{.92}{\sqrt{.0009998}} = 29.114$$

Se o valor z obtido exceder o valor z crítico (consultar a Tabela A apresentada em anexo pelos autores), como é o caso ( $29.114 > 2.326$ , para um  $\alpha = .01$ ), rejeita-se  $H_0$ , concluindo-se, então, que os juízes apresentam um grau de acordo significativo quanto às codificações por si efectuadas.

De referir, ainda, que Siegel e Castellan, Jr. (1989) descrevem as estatísticas adequadas ao

cálculo do acordo inter-juízes quando estão em causa variáveis de outra natureza que não a categorial, como seja o caso das variáveis ordinais, em que é solicitado aos juízes, não a atribuição de  $x$  objectos a  $y$  categorias mas, sim, a sua ordenação com base em determinado critério.

Parece, pois, que este procedimento se revelou mais eficaz, na medida em que, para além do elevado consenso obtido entre os codificadores, não surgiram questões irrelevantes nem dúvidas quanto à favorabilidade da informação procurada, com as quais já se haviam confrontado Devine *et al.* (1990). Por exemplo, uma das questões encontradas foi «Em que medida se considera uma pessoa ambiciosa?». Enquanto que para uns o traço «ambiciosa» tem uma conotação positiva, por a ele associarem uma motivação para o sucesso, para outros tem uma conotação negativa ao considerarem que alguém que possui este atributo é capaz de tudo fazer para atingir aquilo que pretende. Apenas a explicitação do objectivo do sujeito que formulou esta questão permitiu perceber qual o sentido que ele dava a este traço e, portanto, qual a valência da informação que pretendia obter junto da candidata.

Espera-se, pois, que esta alternativa possa contribuir para contornar alguns enviesamentos inerentes aos investigadores os quais, infelizmente, ocorrem com alguma frequência (ver, *e.g.*, Greenwald, Pratkanis, Leippe & Baumgardner, 1986; Innes & Fraser, 1971).

#### NÃO HÁ BELA SEM «SE NÃO...»

Não se poderá ser tão ingénuo, por mais forte que, às vezes, seja esse o nosso desejo, ao ponto de considerar que a «solução foi encontrada». Para desespero dos intolerantes à ambiguidade, as ciências sociais e humanas há muito que nos tentam habituar à ideia de que, para abordar empiricamente o seu objecto de estudo, temos, frequentemente, de nos confrontar com «paus de dois bicos». Recordemos, por exemplo, as vantagens e desvantagens inerentes às questões abertas e às questões fechadas, as quais há que colocar nos pratos da balança para, tendo em conta a investigação em causa, conseguir optar

pelo formato mais adequado, um pouco, por vezes, com base no critério «do mal, o menos».

Pretende-se, com isto, dizer que, a alternativa aqui proposta é, também ela, alvo de críticas, podendo, em última instância, produzir outro tipo de enviesamentos. Assim, quem nos garante que, ao solicitar o objectivo de cada questão aos sujeitos, eles não enveredarão por uma racionalização da forma como estão a procurar a informação que, talvez, não ocorresse se esse pedido não lhes fosse feito? Basta, tão somente, pensarmos na dicotomia entre o processamento automático e controlado (Schneider & Shiffrin, 1977) para que este problema se nos coloque.

Para além disso, poder-se-á evocar, aqui, a mesma crítica que sempre foi colocada à escala de atitudes de Thurstone (ver, *e.g.*, Lima, 1993): até que ponto os juizes se conseguem descentrar das suas posições, codificando cada questão de acordo com o objectivo explicitado pelo sujeito e não com base naquilo que cada um deles considera que a questão, efectivamente, procura?

Não tendo este artigo o objectivo de induzir ansiedade naqueles que buscam «a solução correcta», parece estar na altura de terminar.

Que continue a discussão, ainda que dela possa vir a não nascer a luz. Até porque, como alternativa, só nos resta cruzar os braços, descansando à sombra da velha expressão popular «... que venha o diabo e escolha!».

#### REFERÊNCIAS

- Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (vol. 2, pp. 634-654). Cambridge, Mass: Addison-Wesley.
- Devine, P. G., Hirt, E. R., & Gehrke, E. M. (1990). Diagnostic and confirmation strategies in trait hypothesis testing. *Journal of Personality and Social Psychology*, 58 (6), 952-963.
- Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, 93 (2), 216-229.
- Innes, J. M., & Fraser, C. (1971). Experimenter bias and other possible biases in psychological research. *European Journal of Social Psychology*, 1 (3), 297-310.
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94 (2), 211-228.
- Klayman, J., & Ha, Y.-W. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15 (4), 596-604.
- Leyens, J.-P., & Yzerbyt, V. Y. (1992). The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology*, 22, 549-569.
- Lima, M. (1993). Atitudes. In J. Vala & M. Monteiro (Eds.), *Psicologia social* (cap. VIII, pp. 167-199). Lisboa: Fundação Calouste Gulbenkian.
- Martins, A. (1995). *A influência da informação acerca dos candidatos sobre o desempenho de leigos, iniciados e especialistas em selecção de pessoal*. Tese de Mestrado em Comportamento Organizacional, Instituto Superior de Psicologia Aplicada, Lisboa.
- Martins, A. (1996). *A influência da raça dos candidatos na entrevista de selecção de pessoal*. Artigo a submeter a publicação.
- Rosenberg, S., & Levy, B. I. (1955). A proposed method for increasing accuracy of judges' classification of verbal material. *Journal of Abnormal and Social Psychology*, 50 (3), 357-362.
- Schneider, W., & Shiffrin, R. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84 (1), 1-66.
- Siegel, S., & Castellan, N. Jr. (1989). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Snyder, M., & Swann, W. B., Jr. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, 36 (11), 1202-1212.
- Swann, W. B., Jr., & Giuliano, T. (1987). Confirmatory search strategies in social interaction: How, when, why, and with what consequences. *Journal of Social and Clinical Psychology*, 5 (4), 511-524.
- Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology*, 43 (1), 22-34.
- Trope, Y., Bassok, M., & Alon, E. (1984). The questions lay interviewers ask. *Journal of Personality*, 52 (1), 90-106.
- Zuckerman, M., Knee, C. Hodgins, H., & Miyake, K. (1995). Hypothesis confirmation: The joint effect of positive test strategy and acquiescence response set. *Journal of Personality and Social Psychology*, 68 (1), 52-60.

#### RESUMO

Este artigo consiste na proposta de um procedimento alternativo ao usualmente adoptado na categori-

zação das questões formuladas nos estudos de teste de hipóteses acerca dos outros. Em alguns deles (e.g., Devine, Hirt & Gehrke, 1990; Martins, 1996; Snyder & Swann, 1978; Swann & Giuliano, 1987), tal categorização tem oferecido dificuldades aos juízes codificadores, cuja solução adoptada se considera poder constituir uma fonte de enviesamentos. O procedimento utilizado por Martins (1995) é apresentado como uma alternativa aos das investigações anteriores, ao mesmo tempo que se discutem as suas desvantagens.

*Palavras-chave:* Coeficiente Kappa, Acordo inter-juízes, Acordo inter-codificadores, Codificação de dados qualitativos, Análise de Conteúdo, Teste de Hipóteses.

## ABSTRACT

This article consists in the proposal of an alternative procedure to the usually adopted in the categorization of the questions formulated in the studies of hypothesis testing about the others. In some of them (e.g., Devine, Hirt & Gehrke, 1990; Martins, 1996; Snyder & Swann, 1978; Swann & Giuliano, 1987), that categorization has been offering difficulties to the judges, whose solution adopted is, here, considered as a possible source of biases. The procedure used by Martins (1995) is presented as an alternative to those of the antecedent investigations and his disadvantages are discussed.

*Key words:* Kappa coefficient, Inter-judges agreement, Inter-raters agreement, Codification of qualitative data, Qualitative analysis, Hypothesis testing.