

**SENTIDO PROVÁVEL DE DECISÃO RELATIVO AOS VALORES DOS INDICADORES DE QUALIDADE DE SERVIÇO VERIFICADOS PELOS CTT, EM 2016**

**RESPOSTA DOS CTT**

Lisboa, 29 de setembro de 2017

## 1. INTRODUÇÃO

O Conselho de Administração da ANACOM aprovou, em 15.09.2017, o Sentido Provável de Decisão (“SPD”) relativo aos valores dos indicadores de qualidade de serviço verificados pelos CTT – Correios de Portugal, S.A. (“CTT”) em 2016.

Através de Ofício com a Ref.ª ANACOM-S023039/2017, de 15.09.2017, recebido na mesma data, a ANACOM notificou os CTT para, em sede de audiência prévia, apresentarem a sua pronúncia, nos termos e para os efeitos previstos nos artigos 121.º e seguintes do Código do Procedimento Administrativo (doravante “CPA”). Para o procedimento foi definido um prazo de pronúncia de 10 de dias úteis, terminando assim o prazo de resposta em 29.09.2017.

A questão central no presente SPD é a da não-aceitação, pela ANACOM, da fórmula de cálculo alternativa<sup>1</sup> apresentada pelos CTT para o IQS4 e para o IQS5. A apresentação de uma fórmula de cálculo alternativa pelos CTT para os referidos IQS decorre, essencialmente, da entrada em laboração de um novo sistema de medição dos IQS (sistema de medição independente, efetuado pela PricewaterhouseCoopers – “PwC”) em 01.10.2016, o qual esteve sujeito a limitações de funcionamento intrínsecas à fase de arranque de um novo sistema que, no entender dos CTT, não foram devidamente ponderadas pela ANACOM.

No entender da ANACOM, expresso agora no SPD, os CTT não cumpriram o valor mínimo definido para o IQS4 no ano de 2016. Deste facto decorre a aplicação de um mecanismo de compensação para os utilizadores, correspondente a uma dedução de 0,03 p.p. na variação média ponderada dos preços do cabaz de serviços de correspondências, encomendas e correio editorial verificada no ano de 2017.

Dado existirem interpretações diferentes entre a ANACOM e os CTT quanto às fórmulas de cálculo dos IQS4 e IQS5 para o ano de 2016, os CTT solicitaram ao Professor Doutor António Pacheco Pires, Professor Catedrático de Probabilidades e Estatística do Instituto Superior Técnico um parecer técnico (doravante “Parecer Técnico”) sobre análise estatística dos procedimentos usados, resultados obtidos e interpretação dos resultados dos IQS4 e IQS5 em 2016. Este Parecer Técnico integra a pronúncia dos CTT ao SPD e consta em anexo ao mesmo.

---

<sup>1</sup> À formula de cálculo definida pela ANACOM em 02.03.2017.

## 2. Da análise da ANACOM à fórmula de cálculo do IQS4 e do IQS5 apresentada pelos CTT

Os comentários seguintes dos CTT referem-se especificamente ao ponto 2.2 do SPD em que a ANACOM analisa, ponto por ponto, a fundamentação apresentada pelos CTT para a fórmula de cálculo alternativa dos IQS4 e IQS5.

### 2.1. Ponto 5.2.2 da especificação técnica TS 14773:2004

Tal como já tinha sido referido, os CTT manifestaram preocupação relativamente à medição dos IQS4 e IQS5 no 4.º trimestre de 2016, efetuada pelo sistema independente, devido às elevadas margens de erro estatísticas face aos valores objetivo estabelecidos para estes IQS (de 2,8‰ para um objetivo de 1,4‰ no IQS4 e de 2,1‰ para um objetivo de 1,5‰ no IQS5).

Sobre esta questão, a ANACOM refere no SPD (cf. pp. 8 a 10) que a especificação técnica desenvolvida pelo Comité Europeu de Normalização (CEN) TS 14773:2004 (“Especificação Técnica”) prevê, no seu ponto 5.2.2, que quando o valor absoluto da margem de erro for superior a 40% da estimativa para extravios ou demoras longas, devido ao valor da estimativa ser muito baixo ou devido à dimensão da amostra ser muito baixa, então deverá ser reportado o valor do limite superior do seu intervalo de confiança (para um nível de confiança de 95%).

De acordo com a interpretação da ANACOM, mesmo com menores níveis de precisão devido a um valor baixo da estimativa ou devido a uma amostra de dimensão reduzida, a Especificação Técnica indica que ainda assim deve ser reportado um valor para a estimativa do valor dos extravios ou demoras longas (neste caso, o valor do limite máximo), ou seja, que nesta situação deveria ser reportado, como resultado do estimador do IQS4 e do IQS5, o limite superior do intervalo de confiança.

Os CTT discordam da interpretação efetuada pela ANACOM do ponto 5.2.2 da Especificação Técnica, pelos motivos que a seguir se indicam.

*Por um lado*, no ponto 5.2.1 da Especificação Técnica é definido  $\hat{P}_1$  enquanto o estimador para a probabilidade verdadeira de perda ou demora excessiva no correio doméstico P1. Aplicando ao caso concreto, P1 representa o IQS4 (no caso do correio normal) / IQS5 (no caso do correio azul).

No ponto 5.2.2 é indicado que, se a margem de erro for superior a  $0,4\hat{P}_1$ , então o valor do  $\hat{P}_1$  (estimador do IQS4/IQS5) não deve ser reportado explicitamente, devendo ser antes reportado o limite superior para P1 (com 95% de confiança).

Assim sendo, nesta situação, não devem ser reportados estimadores para os IQS4 e IQS5 (correio normal (azul) não entregue até 15 (10) dias úteis) mas apenas indicar que P1 (% real de correio normal (azul) não entregue até 15 (10) dias úteis pelos CTT) não excede o limite superior, com 95% de confiança.

*Por outro lado, ainda no que se refere ao ponto 5.2.2. da Especificação Técnica, o Parecer Técnico junto a esta pronúncia indica o seguinte, que aproveitamos aqui para salientar: "(...) deve entender-se do referido na Especificação Técnica 14773:2004 é que: em situações em que o rácio da margem de erro do IC (intervalo de confiança) a aproximadamente 95% para um IQS populacional sobre a estimativa desse mesmo IQS populacional seja superior a 40%, é preferível usar o IC unilateral superior associado, ou seja (0, UCL95%), em substituição do correspondente IC bilateral para o mesmo indicador."*

Assim, a ANACOM não pode nestes casos assumir como estimador do IQS o limite superior do intervalo de confiança, pelo que a análise efetuada a diversos cenários não é válida em termos de análise estatística, partindo assim de um pressuposto manifestamente errado que deve ser alterado na decisão final.

Refere ainda a ANACOM que comparando os resultados do 4º trimestre de 2016, apurados pelo sistema da PwC e pelo sistema dos CTT, verifica-se que as margens de erro obtidas pelo sistema PwC, sendo superiores em valor absoluto às obtidas com o sistema dos CTT, apresentam, tanto no caso do IQS4 como no caso do IQS5, um desvio, face ao valor realizado, inferior ao obtido com o sistema dos CTT. No caso do IQS4 o desvio obtido é inferior ao limite de 40%, situação que não ocorre nem com o IQS4 nem com o IQS5 com o sistema dos CTT.

Todavia, importa salientar que a referida situação deriva da ordem de grandeza dos resultados apurados por ambos os sistemas.

O estimador do IQS4 apurado pelo sistema de medição dos CTT é de 0,07%, pelo que, para a margem de erro não ser superior a  $0,4\hat{P}_1$ , esta teria de ser inferior ou igual a 0,028%, ou seja, uma

margem de erro extremamente reduzida, sendo para tal necessário uma dimensão de amostra muito elevada, com custos não razoáveis de operação. No sistema de medição independente da PwC, com um estimador para o IQS4 bastante mais elevado (0,74%), mesmo com margem de erro significativamente maior (até 0,296%) já é possível ficar dentro do limiar dos 40%. Análise idêntica se pode fazer para o IQS5.

Atendendo a que os estimadores obtidos não têm ordens de grandeza substancialmente diferentes (menores que 1%), independentemente de margem erro ser, ou não, superior a  $0,4\hat{P}_1$ , importa referir que margens de erro maiores implicam uma maior amplitude do intervalo de confiança, aumentando a incerteza relativamente ao valor da verdadeira proporção da população.

Face ao exposto, ao contrário do que a ANACOM conclui no SPD, a análise efetuada com base nos limites superiores dos intervalos não válida a metodologia de cálculo definida em 02.03.2017, pelo contrário. A reforçar este aspeto, e como mais à frente se irá referir, existem outros aspetos considerados fulcrais para a fiabilidade da estimação dos IQS4 e IQS5 (proporções populacionais com valores próximos de zero), nomeadamente as dimensões das amostras.

## **2.2. Sobre os constrangimentos admitidos pela PwC, nomeadamente no que concerne à elevada rotatividade do painel**

No período inicial de arranque (outubro a dezembro 2016) o sistema de medição da PwC sofreu vários constrangimentos, nomeadamente a elevada taxa de rotatividade do painel, com os consequentes impactos em termos de fiabilidade da informação. A elevada rotatividade é um potenciador do erro humano, na medida em que evidencia que muitos dos painelistas não estavam realmente comprometidos com o sistema de medição ou não conseguiam cumprir as condições para serem painelistas (nomeadamente, disponibilidade de tempo, disciplina).

A ANACOM refere no SPD que o nível de rotatividade do painel não é um aspeto específico dos IQS4 e IQS5, afetando também os restantes IQS (com exceção do IQS 10, que sendo relativo à medição do tempo em fila de espera, não recorre a painelistas), não apresentando, só por si, motivo para serem tratados de forma distinta os IQS4 e IQS5, face aos restantes.

No entanto, de acordo com a metodologia de medição dos IQS (com exceção do IQS 10) as datas de expedição e receção dos envios são informadas pelos painelistas, pessoas dispersas por todo o



país (Continente e Ilhas) que, a par com todas as suas outras atividades (profissionais e pessoais), expedem e/ou rececionam objetos de teste, informam sobre datas de expedição e receção e providenciam a devolução dos respetivos objetos / informação à entidade que efetua a medição.

Ora, no contexto desta metodologia, não se pode manifestamente afirmar que o nível de rotatividade do painel não é um aspeto relevante para a fiabilidade da medição dos IQS. Pelo contrário, facilmente se compreende que possíveis falhas nos dados comunicados e possíveis não devoluções dos objetos rececionados, terão impacto (relevante e sério) nos resultados apurados.

Efetivamente, e tal como a ANACOM indica, a rotatividade do painel afetou não só os IQS4 e IQS5 mas também os outros IQS. No entanto o impacto nos resultados é mais crítico nestes dois últimos indicadores, pelos seguintes motivos:

- (i) Nos IQS relativos a demoras de encaminhamento, os erros nas datas de expedição e datas de receção reportadas (que para serem garantidas obriga à verificação, por parte do painelista, do seu recetáculo postal todos os dias úteis após as 18h00, devendo todas as situações de exceção serem devidamente informadas à entidade responsável pela medição), têm um impacto mais significativo e relevante nos resultados apurados do que nos outros IQS.

Para o cálculo dos resultados dos IQS4 e IQS5, para além do acima referido, têm também impacto objetos rececionados pelos painelistas que não tenham sido devolvidos à entidade que efetua a medição (exceto quando a devolução tiver sido assumida pelo painelista), bem como objetos não tenham sido expedidos, em que não seja possível confirmar tal situação. Mais uma vez, por aqui fica claro que a rotatividade do painel pode ter impacto relevante nos resultados apurados.

- (ii) Os IQS 4 e 5 estimam proporções populacionais com valores perto de zero, sendo que neste caso em específico o impacto de erros de medição (no caso em apreço derivados de erros humanos) é muito relevante e mais impactante do que noutros IQS. Este impacto é analisado no ponto 2, alínea b) do Parecer Técnico, onde se salienta o seguinte:

*“O impacto de erros de medição na amostragem, em inferências sobre parâmetros populacionais, é o alvo da investigação realizada no seguinte artigo: SCHWARTZ, JOSEPH. E. (1985). The Neglected Problem of Measurement Error in Categorical Data. Sociological Methods &*

*Research 13:435-466. Este artigo conclui que o impacto em questão é muito relevante, especialmente para proporções populacionais com valores próximos de zero, como é o caso das associadas aos IQS amostrais 4 e 5 no estudo em análise, conduzindo à obtenção de IQS amostrais com enviesamento que tendem a sobrestimar os verdadeiros valores dos respetivos IQS populacionais.”*

*É referido ainda neste âmbito que: “(...) caso não sejam tidos em conta em inferências sobre os IQS populacionais, erros de medição na amostragem causados por painelistas conduzem ao aumento da variabilidade de estimadores (IQS amostrais) do tipo dos IQS 4 e 5 usados no estudo, para além do seu enviesamento por sobrestimação, sendo que é o efeito de enviesamento o que provoca consequências mais negativas em termos de estimação e posteriores inferências estatísticas; adianta-se que enviesamentos menores e de sentido contrário tendem a afetar os IQS amostrais 1, 2 e 3.”*

Neste contexto, se analisarmos o impacto de erros de medição (decorrentes de erros humanos) nos estimadores do IQS4 (correio normal não entregue até 15 dias úteis) e do IQS1 (correio normal entregue até 3 dias), que estimam proporções populacionais com valores completamente diferentes, constata-se que:

- (i) No caso do IQS1, que tem um objetivo definido de 96,3% com um mínimo de 95,5%, resulta uma margem de 4,5 pontos percentuais (tendo em consideração o valor mínimo), para acomodar entregas acima do padrão de serviço definido e erros de medição;
- (ii) No entanto, o IQS4, com objetivo definido de 0,14% com um máximo desejável de 0,23%, tem apenas uma margem máxima de 0,23 pontos percentuais para acomodar demoras excessivas e não entregas e erros de medição.

É indicado no Anexo B da Especificação Técnica que, caso seja possível, deve ser estimado até que ponto o nível de demoras excessivas e não entregas é afetado pelos erros de reporte que permanecem na amostra. Sendo este exercício de difícil concretização, uma análise possível pode ser efetuada utilizando-se a informação sobre reclamações neste âmbito.

Da análise das reclamações no 4º trimestre de 2016, a evolução do número de reclamações por correspondência nacional não registada não entregue ou com demora de entrega superior ao esperado registou um decréscimo face ao período homólogo (situação também identificada pela ANACOM). Este facto também indicia um enviesamento dos IQS4 e IQS5 amostrais do 4º trimestre de 2016 apurados pelo sistema de medição independente, por sobrestimação.

Note-se, ainda, que a preocupação com o impacto do erro humano na medição de extravios e demoras excessivas está referido na própria Especificação Técnica (mais concretamente no Anexo E – “*Experience of Measuring Loss*”), estando indicado que se este erro não estiver rigorosamente circunscrito coloca em causa a utilização da metodologia de medição.

No referido Anexo E da Especificação Técnica são descritas as experiências com medição de não entregas / demoras excessivas que tenham sido reportadas ao CEN bem como comentários sobre a implementação desta Especificação Técnica. Para além do caso português, apenas é indicada a implementação de medição de não entregas, através de objetos de teste, na Alemanha, sendo que neste último caso a medição era efetuada pelos colaboradores da própria entidade reguladora, que desempenhavam as tarefas de painelistas expedidores e recetores como parte das suas tarefas diárias, sendo assumidos os resultados obtidos como fiáveis. Acresce ainda que a dimensão elevada da amostra no caso alemão (280.000 itens ano) contribui para a diminuição das margens erro.

Este anexo inclui também diversos comentários sobre as limitações da medição destas variáveis, tais como:

- (i) Reino Unido: informa que o seu sistema de medição E2E mede apenas demoras de encaminhamento não sendo apropriado para a medição de não entregas, concluindo que mesmo se efetuassem melhorias ao sistema, os resultados ainda seriam afetados por erros humanos;
- (ii) *International Post Corporation* (“IPC”): refere que mesmo para painalista com taxa de resposta muito elevada (99,5%) ainda existira espaço de 1% (considerando o expedidor e o recetor) de *deficit* de resposta, acrescentando ainda que raramente a taxa de resposta é tão elevada e que em muitos casos pode ser difícil de não impossível distinguir entre uma não devolução, por parte do painalista, e uma não entrega por parte do operador postal. Acrescenta ainda que, para taxas de não entrega muito baixas (na ordem de 1 por 1000) é duvidoso que a taxa de resposta possa ser melhorada ao ponto de ser possível medir não entregas;



(iii) França: a AFNOR considerou a aplicação da metodologia em apreço para medir não entregas e demoras excessivas, tendo recebido evidências da *La Poste* e de perita em estatística do ASSEC Paris que apontam para problemas na sua aplicação, nomeadamente:

- Num sistema de medição com recurso a painel existem perdas de eficiência e existe um número significativo de não respostas e de respostas não fiáveis por parte do painel;
- Mesmo que implementados todos os procedimentos de validação recomendados na Especificação Técnica, não é possível eliminar o efeito do erro humano dos resultados apurados. Neste âmbito são descritas duas experiências efetuadas com o objetivo de aferir a fiabilidade da informação reportada pelos painelistas, que apontam para níveis de erro frequentemente acima de 1% (de referir que os limites máximos considerados desejáveis para os IQS4 e IQS5 são de respetivamente de 0,23% e 0,25%, ou seja bastante inferiores ao potencial efeito do erro humano na medição).

Nas conclusões finais (ponto E.7 do Anexo E da Especificação Técnica) é indicado que existem dois fatores limitadores desta metodologia: a variância amostral e os erros de reporte (que não podem ser minimizados com um aumento da amostra). Indica também que o CEN recolher mais informação sobre experiências na medição de não entregas e demoras excessivas, em particular se com a implementação desta metodologia, acrescentado que esta informação seria valiosa para determinar como, se possível, se esta especificação técnica poderia ser desenvolvida para passar a ser uma norma.

Tendo por base a experiência Portuguesa com a medição de não entregas e demoras excessivas e o vertido na própria Especificação Técnica, consideram os CTT que a medição do correio não entregue com recurso a informação de painel não é a mais adequada devido ao impacto muito relevante dos erros de reporte na medição. Em particular, no caso na medição dos IQS4 e IQS5 no 4.º trimestre de 2016, efetuada pelo sistema independente da PwC, este aspeto foi especialmente agravado dada a elevada rotatividade do painel (com efeito negativo na fiabilidade da informação reportada) e dado que, em muitas situações, é impossível diferenciar uma não entrega e uma falha de informação do painel.

Como tal, atendendo ao estabelecido na Especificação Técnica, e considerando que na deliberação da ANACOM de 30.12.2014 é determinado que a medição dos IQS deve atender à referida Especificação Técnica, entendem os CTT que o impacto do erro humano na medição de extravios e demoras excessivas não pode deixar de ser considerado pela ANACOM na produção da decisão final.

### **2.3. Sobre a inexistência de uma rotina de controlo das não devoluções de informação/objetos de teste pelos painelistas**

Relativamente a este constrangimento, que ocorreu no período inicial de arranque do sistema PwC e potenciou a dificuldade de destriçar entre a não devolução dos objetos teste por parte do painel e as não entregas CTT, bem como entre objetos não expedidos e objetos não entregues pelos CTT, a ANACOM refere no SPD que a especificação técnica relevante para efeitos de cálculo dos IQS4 e IQS5 prevê que, quando haja dúvidas sobre a atuação dos painelistas em conformidade com os procedimentos estipulados, ou quando haja dúvidas sobre a demora de encaminhamento dos envios de teste ou sobre se os mesmos efetivamente não foram entregues, estes devem ser retirados do painel, ou parte deles quando esteja em causa a atuação do painalista durante determinado período de tempo específico.

Refere ainda a ANACOM que a não implementação atempada de rotinas de controlo poderá potenciar, ou ter potenciado, as situações de dúvida e, assim, as situações de envios de teste que foram anulados, não contando para o cálculo do valor dos IQS.

Efetivamente, a Especificação Técnica prevê mecanismos de validação para mitigar os erros de reporte, sendo claro da leitura deste documento (ver anexo E e anexo B, ponto 3) que é não possível eliminá-los totalmente. Se num sistema estabilizado é muitas vezes difícil fazer a destriça entre uma não entrega e erros de reporte, num sistema com elevada rotatividade de painel e com rotinas de validação imberbes esta situação é, em muito, potenciada, como a ANACOM não pode deixar de reconhecer.

Por este motivo não se pode descurar a importância do impacto dos erros de reporte na medição dos IQS4 e IQS5, sendo ainda mais preocupante à luz dos reduzidos valores máximos desejáveis (de 0,23% e 0,25%, respetivamente), valores que têm de acomodar as demoras excessivas e as não entregas e os erros de reporte do painel.

A ANACOM refere ainda que os CTT tinham conhecimento anterior da necessidade de implementação de sistema de medição independente, não tendo encetado mais cedo esta implementação, o que poderia estar na origem dos constrangimentos detetados.

Sobre este aspeto importa começar por referir que uma implementação mais cedo da externalização da medição não colmataria, de *per si*, os constrangimentos iniciais do período de arranque, que seriam sempre de esperar. Ainda assim, reforça-se a ideia de que os CTT lançaram um concurso público em 05.11.2015 para adjudicação deste serviço, tendo a entidade independente começado a montar o sistema de medição em abril de 2016, pelo que seria expectável, de acordo com um critério de razoabilidade, ter o sistema a funcionar adequadamente a 01.10.2016 (o que não se veio a verificar, como já foi apontado à ANACOM).

Em agosto de 2016, considerando que o processo de implementação já deveria estar mais concretizado, os CTT alertaram a ANACOM para a possibilidade (à data ainda teórica) de ocorrência de limitações de funcionamento inicial de um novo sistema de medição, tendo informado em 30.09.2016 do ponto de situação da implementação, ainda com dificuldades no recrutamento do painel.

As limitações de funcionamento do novo sistema de medição foram comunicadas regularmente à ANACOM, tendo a ANACOM, de resto, previsto na decisão de 02.03.2017 que *“...Caso efetivamente se venham a verificar sérias limitações de funcionamento inicial do sistema independente, que coloquem em causa a fiabilidade dos resultados obtidos com esse sistema no período em causa (4º trimestre 2016), estará a ANACOM disponível para analisar a situação concreta, após oportuna comunicação dos CTT, devidamente fundamentada.”*

Assim, atendendo às efetivas limitações verificadas (e reconhecidas pela entidade externa) não parece adequado que a ANACOM as desconsidere com o argumento de que os CTT deveriam ter acatado atempadamente a situação. Ora, se assim fosse, não teria sentido que a ANACOM se tivesse disponibilizado para avaliar limitações ao funcionamento do sistema na decisão de 02.03.2017.

De resto, os planos aqui em análise são diferentes, pelo que entendem os CTT que não se devem confundir e, sobretudo, importa garantir que um não influencia negativamente o outro. Um plano diz efetivamente respeito àquilo que os CTT fizeram para implementar um sistema de medição independente da qualidade de serviço; outro plano diz respeito à fiabilidade e robustez dos dados

obtidos pelo referido sistema. Estes planos são independentes e a exatidão e força dos dados apurados não pode ser afirmada apenas e só pelo facto de que os CTT poderiam (deveriam) ter feito mais para acautelar.

O que interessa neste SPD é, na perspetiva dos CTT, assegurar que as regras determinadas pela ANACOM são aplicadas sobre factos verdadeiros, substanciais e corretamente apurados e não sobre factos formais que não refletem a realidade material. O que interessa então é saber se os factos apurados e comunicados pelo sistema independente são ou não fiáveis e robustos.

Ora, a verdade é que os CTT sempre alertaram a ANACOM para as limitações do sistema iniciado pela PwC, tendo explicado e fundamentado estes aspetos, nomeadamente por cartas de 15 de março e 3 de maio. De resto, nada na decisão de 02.03.2017 indica que os CTT tinham um prazo determinado para comunicar estas limitações à ANACOM. Também o Parecer Técnico agora junto vem apontar falhas de análise que não podem ser desconsideradas.

Assim, entendem os CTT que o impacto negativo de erros de medição dos IQS, o qual tende a ser mais sentido na fase de arranque inicial de implementação de amostragem por painel, devido à falta de experiência dos painelistas e controladores envolvidos no processo de amostragem, não foi devidamente considerado pela ANACOM, ao contrário do que estava estabelecido na própria decisão de 02.03.2017.

### **3. Outros aspetos relevantes a ter em conta no cálculo dos IQS4 e IQS5 de 2016**

No contexto da medição dos IQS de 2016, efetuada através dois sistemas de medição, permitimo-nos referir seguidamente alguns dos aspetos relevantes indicados no Parecer Técnico:

#### **3.1. Investigação sobre IQS populacionais dos CTT recorrendo a amostragem por painel**

No ponto 2 do Parecer Técnico (p. 3), sobre a possibilidade de ocorrência de erros de medição da demora de entrega de objetos de teste ou de avaliação do prazo de entrega, decorrentes da amostragem por painel, é referido no ponto 2 a) o seguinte: ***“(…) É de realçar a grande importância reconhecida pelos CTT e pela EI em relação à implementação, teste e monitorização dos esquemas de amostragem usados, tendo a EI identificado, em relatório de março de 2017,***



que "... o esforço de formação e de alinhamento com os processos acordados não se pode considerar terminado", a par dos seguintes três impactos fundamentais dos desafios com o arranque do sistema (de amostragem) da EI:

- **Redução da dimensão da amostra sobre a qual incide o estudo** no que concerne à não devolução de informação/objetos de teste por parte dos painelistas.
- **Elevada rotatividade do painel, que se verificou no período de outubro a dezembro de 2016, teve impacto na fiabilidade dos resultados, em especial nos IQS4 e IQS5.**
- **Redução do número de controlos sobre a atuação de parte dos painelistas e sobre a validação de informação.**

Uma redução da dimensão da amostra conduz a maior incerteza (variabilidade) dos estimadores usados, e uma elevada rotatividade do painel e a redução do número de controlos sobre a atuação de parte dos painelistas e sobre a validação de informação são potenciadores de produzir enviesamento dos estimadores (IQS amostrais) em relação aos parâmetros populacionais que estimam (IQS populacionais).

Contudo, uma vez que os valores dos IQS populacionais dos CTT em análise são desconhecidos, não é possível quantificar o efeito concreto que os desafios referidos com o arranque do sistema da EI poderão ter eventualmente causado em termos de desvio dos IQS amostrais referentes ao 4º trimestre de 2016 produzidos pela EI em relação aos valores dos respetivos IQS populacionais para o mesmo período."

### **3.2. Análise da adequabilidade do modo de cálculo definido pela ANACOM e proposta de metodologia correta do ponto de vista estatístico**

No ponto 3 a) do Parecer Técnico (p. 5) é referido que "(...) a decisão de 02.03.2017 da ANACOM referente ao modo de cálculo dos IQS 4 e 5 não é adequada do ponto de vista estatístico, em virtude de essa proposta necessitar que a proporção de objetos de teste de correio normal do 4º trimestre de 2016 não entregue pelos CTT até 15 dias úteis na amostra da EI, bem como a proporção de objetos de teste de correio azul do 4º trimestre de 2016 não entregue pelos CTT até 10 dias úteis na amostra da EI, tivessem distribuição aproximadamente normal."

Refere ainda o Parecer Técnico neste ponto (p. 5) que, para efeitos de inferências sobre os IQS populacionais 4 e 5, tendo em conta os objetivos fixados desses indicadores (1,4‰ e 1,5‰, respetivamente), “... *chega-se à conclusão que deve exigir-se para dimensão da amostra de objetos de teste do 4º trimestre de 2016 pelo menos um valor na ordem de 10 000, o qual é substancialmente maior que os valores amostrais de 3663 (3717) do número de objetos de teste de correio normal (azul) do 4º trimestre de 2016 na amostra do sistema da EI para o IQS4 (IQS5) nesse período.*”

Prosseguindo na alínea b) da mesma página por salientar que “(...) *neste âmbito, conclui-se que já a consideração de todos os objetos de teste do 4º trimestre de 2016, incorporando tanto os objetos de teste do sistema da EI [PwC] como os do sistema dos CTT para esse período, conduziria a um número de objetos de teste de correio normal (azul) de 14 496 (18 117), valores que são bem superiores a 10 000 em ambos os casos.*”

Concluindo depois na alínea c) da p. 5 da seguinte forma “*Pelo referido nas duas alíneas anteriores, considera-se que, tendo por referencial uma avaliação reportada a uma data passada característica da fase de delineamento amostrai do estudo em análise, a opção metodologicamente mais apropriada para ser adotada no cálculo dos IQS amostrais 4 e 5 é a seguinte:*

$$IQSk_{2016} = \frac{9}{12} \times IQSk_{1T-3T}^{CTT} + \frac{3}{12} \times IQSk_{4T}^{CTT+EI}$$

com  $IQSk_{2016}$  denotando o estimador do  $IQSk$  populacional para o ano de 2016 proposto, para  $k=4,5$ , com valores em permilagem:

- $IQS4_{1T-3T}^{CTT}$  denotando a proporção de cartas de teste de correio normal dos três primeiros trimestres de 2016 na amostra de teste do sistema dos CTT não entregues até 15 dias úteis, em permilagem;
  - $IQS4_{4T}^{CTT+EI}$  denotando a proporção de cartas de teste de correio normal do 4º trimestre de 2016 das amostras de teste dos sistemas dos CTT e da EI não entregues até 15 dias úteis, em permilagem;
- sendo  $IQS5_{1T-3T}^{CTT}$  e  $IQS5_{4T}^{CTT+EI}$  definidos analogamente alterando “correio normal” para “correio azul” e “15 dias úteis” para “10 dias úteis”.

### 3.3. Interpretação da ANACOM dos resultados amostrais obtidos

No seu ponto 5, o Parecer Técnico considera que a ANACOM não interpretou corretamente os resultados amostrais obtidos ao considerar, no que se refere ao IQS4 (proporção de correio normal de 2016 não entregue pelos CTT até 15 dias úteis) que os CTT não cumpriram o valor máximo desejável de 2,3‰ definido para o IQS4 populacional, pelas seguintes razões:

- (i) Alínea a): *“Efetivamente, os resultados obtidos não corroboram a conclusão referida estabelecida pela ANACOM, uma vez que, tendo por base o valor do IQS4 amostra para o ano de 2016 calculado de acordo com a decisão da ANACOM de 02.03.2017, o teste da hipótese nula  $H_0: IQS4_{2016} \leq 2,3\%$  contra a hipótese alternativa  $H_1: IQS4_{2016} > 2,3\%$  conduz a uma valor-p de aproximadamente 14%, ou seja, à não rejeição da hipótese nula (correspondente à proporção de correio normal de 2016 não entregue pelos CTT até 15 dias úteis ser inferior ao limite máximo desejável de 2,3‰, fixado para o ano de 2016) a qualquer dos níveis usuais de significância (1%, 5% 10%).”*;
- (ii) Alínea g): *“(…) ligando à alínea inicial, refira-se que, com maioria de razão, conclusões idênticas, de não rejeição da hipótese  $H_0: IQS4_{2016} \leq 2,3\%$  a favor de  $H_1: IQS4_{2016} > 2,3\%$ , são obtidas se forem usados os estimadores alternativos de  $IQS4_{2016}$  obtidos considerando apenas os dados referentes ao sistema dos CTT para o ano de 2016 ou os dados agregados referentes aos três primeiros trimestres de 2016 obtidos pelo sistema dos CTT a par dos dados do 4º trimestre de 2016 obtidos pelo sistema da EI, sem ponderação; para ambas as situações os valor-p obtidos são superiores a 0,5 porque os valores observados desses dois estimadores alternativos de  $IQS4_{2016}$  são inferiores ao valor máximo desejável de 2,3‰ para esse indicador em 2016.”*

Embora o Parecer Técnico se pronuncie sobre aspetos que vão além da matéria que consta do SPD, a verdade é que questiona, fortemente, a posição da ANACOM em relação à medição dos IQS4 e IQS5, nomeadamente por não existir evidência estatística de que os CTT não tenham cumprido o IQS4, não existindo assim base sólida para a decisão da ANACOM.

### 3.4. Da compatibilidade da metodologia proposta pelos CTT com o quadro legal aplicável

Importa ainda salientar que a metodologia alternativa proposta pelos CTT à ANACOM, no sentido de que o valor do ano de 2016 deve assentar nos valores apurados pelos CTT (primeiros três trimestres de 2016) e os valores apurados pela PwC (último trimestre de 2016), sem ponderação, é compatível com o quadro legal aplicável.

Faz-se notar aqui que apesar do artigo 13.º, n.º 3, da Lei Postal impor a medição dos níveis de qualidade de serviço por uma entidade externa, nada nesta disposição determina que a ANACOM esteja obrigada a considerar *exclusivamente* os resultados produzidos por este sistema de medição. O 4.º parágrafo do artigo 16.º da Diretiva Postal depõe exatamente no mesmo sentido.

De igual modo, é compatível com o quadro legal aplicável a metodologia indicada no Parecer Técnico como sendo a mais adequada para ser adotada no cálculo dos IQS 4 e IQS5, em que considera que o valor do último trimestre de 2016 deve assentar na amostra de ambos os sistemas de medição.

Quer isto dizer que, perante um conjunto de circunstâncias especiais, como sucede no presente caso, a ANACOM pode considerar uma metodologia de cálculo que considere valores obtidos por outros sistemas de medição, mesmo para os mesmos períodos, desde que isso se destine a robustecer a análise e fiabilidade das medições realizadas. Aliás, foi neste sentido que a ANACOM veio aprovar a decisão de 02.03.2017, embora para o último trimestre de 2016 os CTT desde sempre tivessem manifestado a necessidade de reforçar a análise, sob pena de se chegar a um resultado desfasado da realidade.

Tal é o que se verifica no caso concreto, quer pela existência de erros amostrais, que pela verificação de constrangimentos e dificuldades no funcionamento inicial do sistema. Estes aspetos foram corroborados e reforçados pelo Parecer Técnico junto à pronúncia dos CTT e que mostra a necessidade de corrigir o sentido de decisão na decisão final.

Assim, a proposta metodológica dos CTT é, não apenas compatível com o quadro regulatório aplicável, mas também a única que se mostra capaz de evitar que os resultados apurados no



4.º trimestre de 2016 sejam inconsistentes, errados e pouco robustos, dando origem a decisões manifestamente desfasadas da realidade.

#### **4. Dos valores dos IQS em 2016 e da aplicação do mecanismo de compensação**

Face ao exposto, os CTT não aceitam as conclusões da ANACOM nas pp. 16 e 17 do SPD a propósito dos valores do IQS4 e do IQS5, nomeadamente a conclusão de que o IQS4 não atingiu o valor objetivo nem o valor mínimo definidos para 2016. Consideram os CTT, entre outros aspetos, que não existem evidências de que este indicador não foi cumprido, baseando-se a decisão da ANACOM em pressupostos manifestamente errados.

Como consequência, os CTT contestam a aplicação do mecanismo de compensação, considerando não existir fundamento nem evidências que o suportem. Em todo o caso, sem conceder, mesmo que houvesse fundamento para aplicar o mecanismo de compensação, os CTT não podem deixar de referir que nos termos do artigo 47.º da Lei Postal o mecanismo de compensação deve ser aplicado em conformidade com os princípios da proporcionalidade e da adequação, princípios que não parecem ter sido ponderados neste caso concreto:

- (i) *Por um lado*, a ANACOM desconsidera, no momento de aplicação do mecanismo de compensação, todos os constrangimentos e dificuldades reconhecidamente verificadas na medição dos IQS no 4.º trimestre de 2016. Por outras palavras, ainda que a ANACOM não considerasse estes aspetos para efeitos de apuramento dos valores dos IQS para o ano de 2016, entendem os CTT que deveria ponderar estas situações ao nível da aplicação do mecanismo de compensação;
- (ii) *Por outro lado*, entendem os CTT que a ANACOM não valorizou adequadamente na aplicação do mecanismo de compensação a circunstância de os CTT terem superado os valores objetivo definidos para o ano de 2016 em vários indicadores, nomeadamente nos IQS3, 6, 8, 10 e 11, estando assim bem acima dos valores mínimos anuais nestes indicadores.

Embora a Lei Postal não obrigue a uma ponderação global dos resultados dos IQS em cada ano, à luz dos princípios da proporcionalidade e adequação, bem como dos princípios regulatórios aplicáveis, creem os CTT que haveria espaço para uma ponderação global e menos isolada dos

IQS, sobretudo, e em especial, em relação a um ano de intensas mudanças ao nível dos procedimentos utilizados, como sucedeu em 2016.

## 5. Conclusão

Face ao exposto e não obstante o facto de a anterior proposta metodológica dos CTT ser compatível com o quadro regulatório aplicável, tendo em consideração o Parecer Técnico agora junto, os CTT requerem que a ANACOM aceite a metodologia de cálculo dos IQS para o ano de 2016 descrita no Parecer Técnico, recalculando-se os IQS 4 e IQS5 à luz da metodologia proposta no ponto 3 c) do Parecer Técnico, apurando-se também as margens de erro para os IQS, de acordo com a fórmula constante no Anexo 1 do Parecer Técnico.



**Anexo**

# PARECER TÉCNICO-CIENTÍFICO

## 1. Enquadramento

Este parecer técnico-científico diz respeito à análise estatística dos procedimentos usados e resultados obtidos para os níveis anuais de qualidade do serviço prestado pelos CTT em 2016 e a considerações sobre a leitura dos mesmos resultados refletida no “Sentido provável de decisão relativo aos valores dos indicadores de qualidade de serviço verificados pelos CTT, em 2016”, constantes de carta de 15.SET.2017, com Ref. ANACOM-S023039/2017, enviada pela ANACOM aos CTT.

Nos termos do artigo 13º, nº 3, da Lei Postal, “Os prestadores de serviço universal devem dispor de um sistema de medição dos níveis de qualidade de serviço efetivamente oferecidos, o qual deve respeitar as normas aplicáveis à medição da qualidade do serviço universal, nomeadamente aos serviços intracomunitários, devendo efetuar a medição dos níveis de qualidade do serviço pelo menos uma vez por ano, através do recurso a uma entidade externa independente.” Nesse âmbito os CTT reportaram à ANACOM estimativas dos níveis anuais de qualidade do serviço prestado pelos CTT em 2016 associados aos seguintes indicadores:

- IQS1: Demora de encaminhamento de correio normal (D+3).
- IQS2: Demora de encaminhamento de correio azul – Continente (D+1).
- IQS3: Demora de encaminhamento de correio azul – CAM (D+2).
- IQS4: Correio normal não entregue até 15 dias úteis (por cada mil cartas).
- IQS5: Correio azul não entregue até 10 dias úteis (por cada mil cartas).
- IQS6: Demora de encaminhamento de jornais e publicações periódicas (D+3).
- IQS7: Demora de encaminhamento de correio transfronteiriço intracomunitário (D+3).
- IQS8: Demora de encaminhamento de correio transfronteiriço intracomunitário (D+5).
- IQS9: Demora de encaminhamento de encomenda normal (D+3).
- IQS10: Tempo em fila de espera no atendimento (% de eventos até 10 minutos).
- IQS11: Demora de encaminhamento de correio registado (D+1).

No âmbito da atividade reguladora, são fixados mínimos anuais para a qualidade do serviço prestado pelos CTT associada aos indicadores referidos, estando previstas compensações por parte dos CTT em caso de incumprimento de algum desses valores mínimos anuais.

Importa referir que os dados para o cálculo dos IQS 1, 2, 3, 4, 5, 6, 9, 10 e 11, que vinham sendo produzidos pelos CTT, passaram a ser produzidos por uma entidade independente (EI) a partir de 01.10.2016. Contudo, os CTT mantiveram paralelamente no 4º trimestre de 2016 a produção de dados para o cálculo dos IQS mencionados. O método de produção dos dados para o cálculo dos IQS 7 e 8 não sofreu alterações em 2016.

A questão relativa ao procedimento de cálculo dos valores dos IQS 1, 2, 3, 4, 5, 6, 9, 10 e 11 para o ano de 2016 foi motivo de discordância entre os CTT e a ANACOM, não tendo as propostas apresentadas pelos CTT, as quais contemplavam a incorporação de informação



sobre os dados produzidos pelos CTT no 4º trimestre de 2016 [a par da informação sobre os dados produzidos pela EI no mesmo período], no cálculo dos valores dos IQS para o ano de 2016, sido acolhidas pela ANACOM.

Por decisão de 02.03.2017, a ANACOM determinou que "o valor anual de 2016 de cada um dos IQS referentes a correio nacional (IQS1, 2, 3, 4, 5, 6, 9 e 11) e ao tempo em fila de espera (IQS10) deve corresponder à média ponderada do valor dos três primeiros trimestres de 2016 (período em que vigorou o sistema CTT) e do valor do último trimestre de 2016 (período durante o qual vigora o sistema independente) ponderando o primeiro pelo fator 9/12 e o segundo pelo fator 3/12, ou seja, ponderando o primeiro pelo período decorrido desde o início do ano até ao do dia anterior ao início do sistema independente, e ponderando o segundo pelo período de tempo, até ao final do ano, em que o sistema independente esteve em vigor."

Na referida decisão, a ANACOM determinou ainda que "Caso efetivamente se venham a verificar sérias limitações de funcionamento inicial do sistema independente, que coloquem em causa a fiabilidade dos resultados obtidos com esse sistema no período em causa [último trimestre de 2016], estará a ANACOM disponível para analisar a situação concreta, após oportuna comunicação dos CTT, devidamente fundamentada."


## 2. Análise Estatística

As componentes principais a ter em conta em estudos estatísticos conducentes a inferências sobre índices de qualidade de serviço (IQS) por recurso a amostragem do tipo do analisado neste parecer são as seguintes:

- (a) Definição da população em estudo.
- (b) Definição dos IQS populacionais a investigar.
- (c) Definição do(s) esquema(s) de amostragem a usar e sua implementação, teste e monitorização.
- (d) Definição das fórmulas de cálculo dos IQS amostrais a usar como estimadores dos IQS populacionais.
- (e) Obtenção das distribuições dos IQS amostrais definidos em (d).
- (f) Realização e interpretação de inferências sobre os IQS populacionais.

Tendo por referencial estas componentes, as quais estão interligadas, apresentam-se a seguir os pontos considerados mais importantes relativamente à avaliação da metodologia estatística seguida para a investigação de IQS dos CTT em 2016. A fim de facilitar a identificação dos aspectos mais relevantes mencionados neste parecer, recorrer-se-á ao uso de negrito para realçar os mesmos.

**PONTO 1: A importância da investigação de IQS dos CTT é bem reconhecida e devidamente valorizada por todas os intervenientes [CTT, EI e ANACOM] na produção, análise e interpretação dos resultados obtidos para os IQS amostrais dos CTT para o ano de 2016.**



**PONTO 2:** A investigação sobre IQS populacionais dos CTT é feita recorrendo a amostragem por painel, com recurso a objetos de teste (cartas e encomendas) e objetos reais (jornais e publicações periódicas), consoante o IQS em questão, podendo ocorrer erros de medição da demora de entrega de objetos de teste ou de avaliação se a sua entrega ocorre num determinado prazo. Neste quadro:

(a) É de realçar a grande importância reconhecida pelos CTT e pela EI em relação à implementação, teste e monitorização dos esquemas de amostragem usados, tendo a EI identificado, em relatório de março de 2017, que "... o esforço de formação e de alinhamento com os processos acordados não se pode considerar terminado", a par dos seguintes três impactos fundamentais dos desafios com o arranque do sistema (de amostragem) da EI:

- **Redução da dimensão da amostra sobre a qual incide o estudo** no que concerne à não devolução de informação/objetos de teste por parte dos painelistas.
- **Elevada rotatividade do painel**, que se verificou no período de outubro a dezembro de 2016, teve impacto na fiabilidade dos resultados, em especial nos IQS4 e IQS5.
- **Redução do número de controlos sobre a actuação de parte dos painelistas e sobre a validação de informação.**

Uma redução da dimensão da amostra conduz a maior incerteza (variabilidade) dos estimadores usados, e uma elevada rotatividade do painel e a **redução do número de controlos sobre a actuação de parte dos painelistas e sobre a validação de informação** são potenciadores de produzir enviesamento dos estimadores (IQS amostrais) em relação aos parâmetros populacionais que estimam (IQS populacionais).

Contudo, uma vez que os valores dos IQS populacionais dos CTT em análise são desconhecidos, **não é possível quantificar o efeito concreto que os desafios referidos com o arranque do sistema da EI poderão ter eventualmente causado em termos de desvio dos IQS amostrais referentes ao 4º trimestre de 2016 produzidos pela EI em relação aos valores dos respetivos IQS populacionais para o mesmo período.**

(b) **O impacto de erros de medição na amostragem, em inferências sobre parâmetros populacionais**, é o alvo da investigação realizada no seguinte artigo: SCHWARTZ, JOSEPH. E. (1985). The Neglected Problem of Measurement Error in Categorical Data. Sociological Methods & Research 13:435-466. Este artigo conclui que o impacto em questão é **muito relevante, especialmente para proporções populacionais com valores próximos de zero**, como é o caso das associadas aos IQS amostrais 4 e 5 no estudo em análise, conduzindo à obtenção de IQS amostrais com enviesamento que tendem a sobrestimar os verdadeiros valores dos respetivos IQS populacionais.

O facto acabado de mencionar relativamente aos IQS amostrais 4 e 5 decorre de ser possível objetos (cartas) que não sejam enviadas por painelistas encarregues de o fazer ou objetos recebidas sem atraso significativo pelo painalista receptor e por este não devolvidos sejam declarados como não tendo sido entregues pelos CTT, caso tal não seja identificado por controladores, contribuindo erradamente para aumentar a contabilização de objectos da amostra não entregues pelos CTT em 15 (10) dias úteis para correio normal (azul).



Conclui-se assim que, caso não sejam tidos em conta em inferências sobre os IQS populacionais, erros de medição na amostragem causados por painelistas conduzem ao aumento da variabilidade de estimadores (IQS amostrais) do tipo dos IQS 4 e 5 usados no estudo, para além do seu enviesamento por sobrestimação, sendo que é o efeito de enviesamento o que provoca consequências mais negativas em termos de estimação e posteriores inferências estatísticas; adianta-se que enviesamentos menores e de sentido contrário tendem a afetar os IQS amostrais 1, 2 e 3.

Neste âmbito, é especialmente relevante o facto de os enviesamentos referidos não serem reduzido por aumento da dimensão da amostra, conforme mencionado no artigo atrás referido, porque (essencialmente) as observações efetivamente recolhidas de amostras de teste não possuem a mesma distribuição que os dados reais (correspondentes a correio não pertencente às amostras de teste) que se pretende replicar através de amostras de teste.

Da leitura da troca de correspondência entre a ANACOM e os CTT relativamente às divergências destas entidades quanto às fórmulas de cálculo dos IQS amostrais 4 e 5 para o ano de 2016, parece transparecer que o impacto negativo de erros de medição na amostragem atrás referido, o qual tende a ser mais sentido na fase inicial de implementação de amostragem por painel, face a provável falta de experiência dos painelistas e controladores envolvidos no processo de amostragem, não tem sido devidamente considerado pela ANACOM.

- (c) O atrás referido fornece indícios de que pode revelar-se difícil para a ANACOM a obtenção de conclusões fiáveis sobre o cumprimento/incumprimento por parte dos CTT de valores máximos muito baixos (próximos de zero) para IQS populacionais de teor negativo (como é o caso dos IQS 4 e 5) recorrendo ao tipo de delineamento amostral atualmente em uso, tendo em conta que não se afigura fácil à ANACOM tirar conclusões sobre a magnitude dos erros de medição na amostragem causados por falhas humanas por parte de painelistas e controladores.
- (d) Relativamente a inferências sobre extravios de correspondência, uma abordagem alternativa cujo uso poderá ser equacionado futuramente é a de inferir, por extrapolação, proporções de extravios a partir de dados de reclamações e de informação adicional sobre a propensão dos utilizadores a apresentarem reclamações decorrentes de extravio de correspondência, abordagem essa que poderá não implicar elevados custos de implementação.
- (e) Como nota final deste ponto, faz-se notar que não se considera apropriado o uso da terminologia de "valor mínimo" (presente nos documentos analisados) para os IQS 4 e 5 (com valores de 2.3‰ e 2.5‰ para os IQS 4 e 5 no ano de 2016, respetivamente) pois se entende efetivamente que esses são os valores máximos desejáveis para os IQS populacionais 4 e 5 para o ano de 2016: Correio normal (azul) não entregue até 15 (10) dias úteis, por cada mil cartas, em 2016. Nesse sentido, passar-se-á a usar a partir daqui a terminologia "valor máximo desejável" no referente aos IQS 4 e 5.

**PONTO 3.** A componente (e), referente à obtenção da distribuição dos IQS amostrais usados, o qual é de importância capital para realização e interpretação de inferências sobre os IQS populacionais, não foi adequadamente tida em conta nos aspectos a seguir apresentados:

- (a) No pressuposto de uso de uma metodologia de cálculo de intervalos de confiança para valores de IQS populacionais baseados no uso de normalidade (ou normalidade aproximada) dos IQS amostrais usados<sup>1</sup>, a qual parece ter sido tacitamente assumida pela ANACOM e pelos CTT, a **decisão de 02.03.2017 da ANACOM referente ao modo de cálculo dos IQS 4 e 5 não é adequada do ponto de vista estatístico**, em virtude de essa proposta necessitar que a proporção de objetos de teste de correio normal do 4º trimestre de 2016 não entregue pelos CTT até 15 dias úteis na amostra da EI, bem como a proporção de objetos de teste de correio azul do 4º trimestre de 2016 não entregue pelos CTT até 10 dias úteis na amostra da EI, tivessem distribuição aproximadamente normal.

**Está em causa a razoabilidade de uso da aproximação à normal da distribuição binomial para as amostras da EI referentes ao 4º trimestre de 2016 para efeitos de inferências sobre os IQS populacionais 4 e 5**, como explicado a seguir. Por associação, como os objetivos fixados para os valores populacionais desses indicadores são respetivamente 1.4‰ e 1.5‰, a metodologia a usar deve ser apropriada, pelo menos, para valores populacionais desses indicadores na ordem de 1‰, para assegurar que a metodologia estatística usada funciona razoavelmente para valores da ordem de grandeza dos valores objectivo mas um pouco inferiores a esses.

Usando como critério prático para uso da aproximação à distribuição normal da distribuição Binomial( $n, p$ )<sup>2</sup>, para efeitos de cálculo de valores aproximados da respetiva função de distribuição, a condição  $np(1-p) \geq 10$  (cf. <sup>3</sup>) **chega-se à conclusão que deve exigir-se para dimensão da amostra de objetos de teste do 4º trimestre de 2016 pelo menos um valor na ordem de 10 000, o qual é substancialmente maior que os valores amostrais de 3663 (3717) do número de objetos de teste de correio normal (azul) do 4º trimestre de 2016 na amostra do sistema da EI para o IQS4 (IQS5) nesse período.**

- (b) Neste âmbito, conclui-se que já a **consideração de todos os objetos de teste do 4º trimestre de 2016, incorporando tanto os objetos de teste do sistema da EI como os do sistema dos CTT para esse período, conduziria a um número de objetos de teste de correio normal (azul) de 14 501 (18 132), valores que são bem superiores a 10 000 em ambos os casos.**
- (c) Pelo referido nas duas alíneas anteriores, considera-se que, tendo por referencial uma avaliação reportada a uma data passada característica da fase de delineamento amostral do estudo em análise, a **opção metodologicamente mais apropriada para ser adoptada no cálculo dos IQS amostrais 4 e 5 é a seguinte:**

$$IQSk_{2016} = \frac{9}{12} \times IQSk_{1T-3T}^{CTT} + \frac{3}{12} \times IQSk_{4T}^{CTT+EI}$$

<sup>1</sup>Veja-se a este respeito o desenvolvimento metodológico apresentado no Anexo 1 para o cálculo da variância de estimadores que sejam médias ponderadas de outros estimadores (como é o caso de alguns dos IQS amostrais para o ano de 2016 calculados através das fórmulas de cálculo fixadas por decisão da ANACOM de 02.03.2017), pressupondo normalidade ou normalidade aproximada desses últimos estimadores. É de notar que as inferências que recorram a distribuições (essencialmente) exatas dos IQS baseadas no modelo binomial são necessariamente bastante complicadas do ponto de vista estatístico.

<sup>2</sup>Por distribuição Binomial( $n, p$ ), entenda-se a distribuição binomial com número de provas  $n$  e probabilidade de sucesso de cada prova igual a  $p$ .

<sup>3</sup>Veja-se, e.g., [S.M. ROSS (1987). Introduction to Probability and Statistics for Engineers and Scientists. Wiley, New York: p. 107].



com  $IQSk_{2016}$  denotando o estimador do IQSk populacional para o ano de 2016 proposto, para  $k=4,5$ , com valores em permilagem;

- $IQS4_{1T-3T}^{CTT}$  denotando a proporção de cartas de teste de correio normal dos três primeiros trimestres de 2016 na amostra de teste do sistema dos CTT não entregues até 15 dias úteis, em permilagem;
- $IQS4_{4T}^{CTT+EI}$  denotando a proporção de cartas de teste de correio normal do 4º trimestre de 2016 das amostras de teste dos sistemas dos CTT e da EI não entregues até 15 dias úteis, em permilagem;

sendo  $IQS5_{1T-3T}^{CTT}$  e  $IQS5_{4T}^{CTT+EI}$  definidos analogamente alterando "correio normal" para "correio azul" e "15 dias úteis" para "10 dias úteis".

**As principais razões porque esta proposta de cálculo dos IQS amostrais 4 e 5 para o ano de 2016 é metodologicamente adequada são:**

- Manter o uso, aceite pela ANACOM, dos pesos  $\frac{9}{12}$  e  $\frac{3}{12}$ , respetivamente para ponderação da informação associada aos três primeiros trimestre de 2016 e ao 4º trimestres de 2016, o que é apropriado por **seguir o princípio da amostragem estratificada de ponderar os estimadores usados em cada estrato** [neste caso os estratos devem ser entendidos como intervalos de tempo do ano de 2016] **com peso proporcional à dimensão do estrato** [viz. duração dos intervalos de tempo dos três primeiros trimestres de 2016 e duração do 4º trimestre de 2016, respetivamente].
- **Seguir o princípio de usar todas as observações amostrais recolhidas contendo informação relevante para o IQS populacional em causa.**
- **Garantir a possibilidade de realizar inferências sobre os IQS populacionais IQS4 e IQS5.**

Note-se que **esta proposta não dá igual importância a cada uma das observações amostrais usadas no cálculo do IQS amostral** [princípio que é seguido em amostragem aleatória simples] **por se considerar mais importante a satisfação do princípio alternativo enunciado no primeiro dos três itens anteriores, tendo em conta que se pretende estimar um IQS populacional referente à globalidade do ano de 2016 no seu conjunto.** Nesse sentido, esta proposta difere claramente da última proposta apresentada à ANACOM pelos CTT para o cálculo dos IQS amostrais 4 e 5, a qual usa toda as observações amostrais recolhidas em 2016 contendo informação relevante para o IQS populacional em causa, sem recurso a ponderação de estimadores.

- (d) **A fórmula implementada pelos CTT para cálculo da variância dos IQS amostrais 1, 2, 3, 4, 5, 6, 9, 10 e 11** [definidos segundo a decisão da ANACOM de 02.03.2017<sup>4</sup>] **não está correta.** Para clarificação, o modo correto de cálculo dessas variâncias é apresentado no Anexo 1, sendo que **os valores corretos dessas variâncias são inferiores aos calculados pelos CTT e, em conformidade, conduzem a estimativas inferiores para as margens de erro de intervalos de confiança para os respetivos IQS populacionais.**

<sup>4</sup>Segundo essa decisão os IQS amostrais 1, 2, 3, 4, 5, 6, 9, 10 e 11 são calculados através de uma média ponderada do valor dos três primeiros trimestres de 2016 (referente ao sistema dos CTT) e do valor do último trimestre de 2016 referente ao sistema da EI



A título de exemplo, as margens de erro corrigidas para os intervalos de confiança a 95% para IQS4 e IQS5 baseadas nos IQS amostrais 4 e 5, calculados pela fórmula fixada por decisão da ANACOM de 02.03.2017, são respetivamente 0,74‰ e 0,57‰ em vez dos valores 1,0‰ e 0,8‰ reportados pelos CTT.

**PONTO 4:** Os desafios com o arranque do sistema da EI referidos no PONTO 2 podem ajudar, em parte, a explicar a falta de consistência observada entre resultados referentes ao 4º trimestre de 2016 obtidos pelo sistema dos CTT e pelo sistema da EI no que diz respeito aos IQS 4 e 5.

- (a) Recorrendo ao teste exacto de Fisher-Irwin<sup>5</sup> para testar a igualdade dos valores esperados dos IQS4 (IQS5) amostrais referentes ao 4º trimestre de 2016 do sistema dos CTT e do sistema da EI contra a hipótese contrária (de os valores esperados referidos serem diferentes) somos conduzidos a um valor-p de aproximadamente zero tanto no teste referente ao IQS4 como no referente ao IQS5.

Conclui-se assim que estes dois testes fornecem evidência estatística clara de que os IQS4 (IQS5) amostrais referentes ao 4º trimestre de 2016 obtidos a partir do sistema CTT e os correspondentemente IQS amostrais obtidos a partir do sistema da EI resultam de populações Bernoulli (porque as observações associados são variáveis binárias) com distintas probabilidades de sucesso.

- (b) Do referido no alínea anterior, pode concluir-se, por maioria de razão, que existe evidência estatística clara de que um ou ambos dos IQS calculados pelos CTT e pela EI para o 4º trimestre de 2016 resultam de estimadores enviesados do valor populacional (proporção de correio normal do 4º trimestre de 2016 não entregue pelos CTT até 15 dias úteis). Esta conclusão pode não parecer estranha, tendo em conta o anteriormente referido no Ponto 2 (b).

**PONTO 5:** A ANACOM não interpretou correctamente os resultados amostrais obtidos ao considerar (essencialmente) que no que diz respeito ao IQS4 (proporção de correio normal de 2016 não entregue pelos CTT até 15 dias úteis) os CTT não cumpriram o valor máximo desejável de 2,3‰ definido para o ano de 2016 para o IQS4 populacional, valor que no resto deste ponto é designado por  $IQS4_{2016}$ , conforme explicado em seguida:

- (a) Efetivamente, os resultados obtidos não corroboram a conclusão referida estabelecida pela ANACOM, uma vez que, tendo por base o valor do IQS4 amostral para o ano de 2016 calculado de acordo com a decisão da ANACOM de 02.03.2017, o teste da hipótese nula  $H_0: IQS4_{2016} \leq 2,3‰$  contra a hipótese alternativa  $H_1: IQS4_{2016} > 2,3‰$  conduz a uma valor-p de aproximadamente 14%<sup>(6)</sup>, ou seja, à não

<sup>5</sup>O teste de Fisher-Irwin para comparação de parâmetros de duas populações de Bernoulli é descrito, e.g., em [S.M. ROSS (1987). Introduction to Probability and Statistics for Engineers and Scientists. Wiley, New York: pp.232-233].

<sup>6</sup>Para a realização do teste referido, usaram-se para valor observado do IQS4 amostral e da estimativa do respetivo desvio padrão os correspondentes valores exactos arredondados a duas casas decimais, respetivamente 2,71‰ e 0,38‰ (fazendo com que o valor usado para valor observado do IQS amostral, 2,71‰, não coincida com aquele que foi reportado pelos CTT à ANACOM, 2,8‰), a par do pressuposto de que o estimador IQS4 amostral para o ano de 2016 calculado de acordo com a decisão da ANACOM de 02.03.2017, o qual vamos designar por  $IQS4_{2016}^{ANACOM}$ , é centrado para  $IQS4_{2016}$ . Nessa situação, tirando partido do facto de este estimador possuir distribuição aproximadamente normal, pode usar-se para estatística de teste

rejeição da hipótese nula (correspondente à proporção de correio normal de 2016 não entregue pelos CTT até 15 dias úteis ser inferior ao limite máximo desejável de 2.3‰, fixado para o ano de 2016) a qualquer dos níveis usuais de significância (1%, 5%, 10%).

- (b) A conclusão anterior não é de surpreender, uma vez que a mesma poderia ser parcialmente antecipada a partir da informação que foi reportada à ANACOM pelos CTT. Note-se que, tendo em conta os valores reportados pelos CTT à ANACOM para estimativa de  $IQS4_{2016}$  e da margem de erro do intervalo de confiança a aproximadamente 95% para o mesmo indicador [2,8‰ e 1,0‰, respetivamente] o intervalo de confiança a aproximadamente 95% para  $IQS4_{2016}$  resultante seria (1,8‰; 3,8‰), o que garantiria desde logo a não rejeição de  $H_0: IQS4_{2016} \leq 2,3‰$  a favor de  $H_1: IQS4_{2016} > 2,3‰$  para níveis de significância superiores ou iguais a 5%, atendendo ao facto de o intervalo de confiança a aproximadamente 95% para  $IQS4_{2016}$  calculado conter o respetivo valor máximo desejável de 2.3‰ para esse indicador em 2016.
- (c) O facto de a ANACOM não ter realizado que a informação que lhe foi reportada pelos CTT implicar que o intervalo de confiança a aproximadamente 95% para  $IQS4_{2016}$ <sup>7</sup> contém o valor máximo desejável de 2.3‰, fixado para este IQS, fazem-me antecipar a conveniência de a informação a enviar pelos CTT à ANACOM, bem como a informação a disponibilizar ao público em geral e à comunicação social passe a incluir preferencialmente intervalos de confiança a aproximadamente 95% para os IQS dos CTT em vez de estimativas pontuais dos mesmos.
- (d) No mesmo âmbito, sendo extremamente importante que a informação disponibilizada ao público em geral, seja clara e estatisticamente rigorosa, é essencial que a informação disponibilizada publicamente não seja passível de levantar dúvidas aos utentes dos CTT e a membros da comunicação social que os IQS dos CTT que são alvo da atenção deste parecer são obtidos por amostragem e constituem, por isso, apenas estimativas de IQS efetivamente prestados pelos CTT.
- (e) Exemplificando, para o tipo de estudo em análise neste parecer não faz qualquer sentido reportar como única estatística associada a um dado IQS populacional o limite superior de um intervalo de confiança (IC) a aproximadamente 95%,  $UCL_{95\%}$ , para esse IQS populacional, por esse valor ser uma estimativa de um quantil elevado da distribuição do IQS populacional; a interpretação correcta do significado da estimativa  $UCL_{95\%}$  do IQS populacional seria necessariamente muito difícil para uma grande fatia dos utentes dos CTT e também dos membros da comunicação social portuguesa.

Neste âmbito, o que deve entender-se do referido na Especificação Técnica 14773:2004 é que: em situações em que o rácio da margem de erro do IC a aproximadamente

$T = (IQS4_{2016}^{ANACOM} - 2,3‰) / \sqrt{Var[IQS4_{2016}^{ANACOM}]}$ , onde  $Var[IQS4_{2016}^{ANACOM}]$  é o estimador da variância amostral de  $IQS4_{2016}^{ANACOM}$  calculado com a teoria apresentada no Anexo I. Como se deve rejeitar  $H_0$  para valores grandes da estatística de teste, tirando da normalidade aproximada da distribuição da estatística de teste caso  $IQS4_{2016} = 2,3‰$  e do facto de o valor observado da estatística de teste ser  $t = (2,71‰ - 2,3‰) / 0,38‰ \approx 1,08$ , decorre que o valor-p observado para o teste é aproximadamente  $1 - \Phi(1,08)$ , ou seja aproximadamente 14%.

<sup>7</sup>Os valores reportados pelos CTT no relatório enviado à ANACOM conduzem a um intervalo de confiança a aproximadamente 95% para  $IQS4_{2016}$  de (1,8‰; 3,8‰), o qual possui maior amplitude que o referido neste parecer, devido ao problema de sobreestimação da variância do IQS4 amostral para o ano de 2016, calculado de acordo com a decisão da ANACOM de 02.03.2017, decorrente do referido no Ponto 3 (b) deste parecer.



95% para um IQS populacional sobre a estimativa desse mesmo IQS populacional seja superior a 40%, é preferível usar o IC unilateral superior associado, ou seja  $(0, UCL_{95\%})$ , em substituição do correspondente IC bilateral para o mesmo indicador.

- (f) Ainda neste quadro, enquanto entidades prestadoras de serviço público, os CTT e a ANACOM têm o dever e a obrigação de contribuírem ao aumento da literacia estatística dos portugueses e dos residentes em Portugal.

Por outro lado, não deve ser menosprezado o impacto que interpretações estatísticas menos rigorosas podem ter junto da opinião pública, como prova o grande impacto que a informação constante do "Sentido provável de decisão relativo aos valores dos indicadores de qualidade de serviço verificados pelos CTT, em 2016", Ref. ANACOM-S023039/2017, já teve junto da comunicação social, a qual divulgou numerosas notícias com títulos como "Os CTT não cumpriram em 2016 o indicador relativo ao correio normal" e outros semelhantes, sem que o corpo de tais notícias ajudasse a clarificar a interpretação dos resultados relativos a IQS dos CTT, para já não falar da metodologia seguida na produção desses mesmos indicadores.

- (g) Para terminar este ponto e ligando à alínea inicial, refira-se que, com maioria de razão, conclusões idênticas, de não rejeição da hipótese  $H_0: IQS4_{2016} \leq 2,3\%$  a favor de  $H_1: IQS4_{2016} > 2,3\%$ , são obtidas se forem usados os estimadores alternativos de  $IQS4_{2016}$  obtidos considerando apenas os dados referentes ao sistema dos CTT para o ano de 2016 ou os dados agregados referentes aos três primeiros trimestres de 2016 obtidos pelo sistema dos CTT a par dos dados do 4º trimestre de 2016 obtidos pelo sistema da EI, sem ponderação; para ambas as situações os valor-p obtidos são superiores a 0,5 porque os valores observados desses dois estimadores alternativos de  $IQS4_{2016}$  são inferiores ao valor máximo desejável de 2,3% para esse indicador em 2016.

Lisboa, 28 de setembro de 2017.



(António Pacheco Pires)

(Professor Catedrático de Probabilidades e Estatística do Instituto Superior Técnico)

AK

## Anexo 1

Neste anexo apresenta-se a fórmula de cálculo do valor esperado ( $E[Y]$ ) e da variância ( $\text{Var}[Y]$ ) de um estimador  $Y$ , de um dado parâmetro populacional de interesse, da forma

$$Y = \beta Y_1 + (1 - \beta) Y_2$$

com  $0 < \beta < 1$  e  $Y_1$  e  $Y_2$  sendo estatísticas independentes ( $Y_1$  e  $Y_2$  não necessitam de individualmente serem estimadores naturais do parâmetro populacional atrás referido). É ainda apresentada a obtenção de intervalos de confiança para o valor esperado de  $Y$ , caso  $Y_1$  e  $Y_2$  possuam distribuições normais ou aproximadamente normais.

Da linearidade do valor esperado resulta que:

$$E[Y] = \beta E[Y_1] + (1 - \beta) E[Y_2]. \quad (1)$$

Por outro lado, das propriedades da variância, usando o facto de  $Y_1$  e  $Y_2$  serem estatísticas independentes, conclui-se que:

$$\text{Var}[Y] = \beta^2 \text{Var}[Y_1] + (1 - \beta)^2 \text{Var}[Y_2]. \quad (2)$$

Se  $Y_1$  e  $Y_2$  possuírem distribuição normal, então

$$IC_{1-\alpha}(E[Y]) = \left( Y - z_{\alpha/2} \sqrt{\text{Var}[Y]}, Y + z_{\alpha/2} \sqrt{\text{Var}[Y]} \right) \quad (3)$$

é um intervalo de confiança a  $(1 - \alpha) \times 100\%$  para  $E[Y]$ , onde  $z_p$  denota o quantil de probabilidade  $(1 - p)$  da distribuição normal padrão, e.g.,  $z_{0.025} \approx 1.96$  é o valor a ser usado para construir um IC a 95% para  $E[Y]$ .

Por outro lado, em virtude de (2), conclui-se então que  $IC_{1-\alpha}(E[Y])$  é dado por:

$$\left( Y - z_{\alpha/2} \sqrt{\beta^2 \text{Var}[Y_1] + (1 - \beta)^2 \text{Var}[Y_2]}, Y + z_{\alpha/2} \sqrt{\beta^2 \text{Var}[Y_1] + (1 - \beta)^2 \text{Var}[Y_2]} \right). \quad (4)$$

Se, em alternativa, apenas se puder garantir que  $Y_1$  e  $Y_2$  possuem distribuições aproximadamente normais, então o intervalo aleatório anterior é um intervalo de confiança a aproximadamente  $(1 - \alpha) \times 100\%$  para  $E[Y]$ .

Ainda na mesma situação, caso  $\text{Var}[Y_1]$  e  $\text{Var}[Y_2]$  sejam desconhecidas, mas se possuam estimadores consistentes de  $\text{Var}[Y_1]$  e  $\text{Var}[Y_2]$ ,  $\widehat{\text{Var}}[Y_1]$  e  $\widehat{\text{Var}}[Y_2]$ , e as dimensões das amostras independentes donde são obtidos esses estimadores sejam grandes, pode concluir-se usando o Teorema de Slutsky que:

$$\left( Y - z_{\alpha/2} \sqrt{\beta^2 \widehat{\text{Var}}[Y_1] + (1 - \beta)^2 \widehat{\text{Var}}[Y_2]}, Y + z_{\alpha/2} \sqrt{\beta^2 \widehat{\text{Var}}[Y_1] + (1 - \beta)^2 \widehat{\text{Var}}[Y_2]} \right) \quad (5)$$

é um intervalo de confiança a aproximadamente  $(1 - \alpha) \times 100\%$  para  $E[Y]$ .

É importante realçar que caso os estimadores  $Y_1$  ou  $Y_2$  não possuam distribuições aproximadamente normais, não é lícito usar o procedimento anterior, baseado no uso da distribuição normal, para a construção de intervalos de confiança para  $E[Y]$ .