



MESTRADO EM
ECONOMIA E GESTÃO DA INOVAÇÃO

Mobile Health: a adoção e a difusão de aplicativos móveis de saúde

Ana Sofia Pereira da Silva

M

2018

*MOBILE HEALTH: A ADOÇÃO E A DIFUSÃO DE
APLICAÇÕES MÓVEIS DE SAÚDE*

Ana Sofia Pereira da Silva

Dissertação

Mestrado em Economia e Gestão da Inovação

Orientado por
José Pedro Coelho Rodrigues
Abílio Carlos Pereira Pacheco

2018

Nota Biográfica

Ana Sofia Pereira da Silva nasceu a 12 de janeiro de 1994 na cidade da Póvoa de Varzim. Licenciou-se em Gestão no ano de 2015 na Faculdade de Economia da Universidade do Porto, tendo terminado o curso com média final de 15 valores.

Em setembro do mesmo ano, ingressou o Mestrado em Economia e Gestão da Inovação também na Faculdade de Economia da Universidade do Porto. Em junho de 2016, realizou um estágio profissional (o qual se expandiu a um contrato de trabalho) numa empresa de consultoria, o que lhe permitiu desenvolver competências ao nível de várias bases de dados (nacionais e internacionais), desenvolvimento de estudos setoriais e análises de mercado.

O ano de 2018 ficou marcado pela cessação de funções nessa empresa e pela integração numa nova empresa, onde desempenha atualmente funções de Consultora Fiscal.

Agradecimentos

A elaboração da presente investigação não teria sido possível sem o apoio e constante motivação de algumas pessoas, a quem gostaria, por isso, de agradecer.

Aos professores José Coelho Rodrigues e Abílio Pereira Pacheco, pela orientação na elaboração desta dissertação e disponibilidade demonstrada, principalmente na reta final, a qual se caracteriza sempre por uma enorme pressão.

Aos meus colegas de curso, pela presença assídua e apoio em todos os momentos, pelas pausas de trabalho (que eram bem merecidas) e por todas as horas de empenho conjunto nos corredores da FEP.

Aos “meus”, por nunca desistirem de mim e por me “levantarem” quando a motivação parecia desaparecer e, principalmente, por toda a paciência. Ao Gil, à Lili, à Catarina, ao Pedro, à Raquel e à Filipa.

Por fim, ao meu maior suporte na vida: os meus pais. Obrigada por todas as oportunidades que me proporcionaram (e proporcionam) ao longo de todos os anos do meu percurso escolar e pessoal. Obrigada por estarem sempre lá para mim, por “aquele” abraço que melhora qualquer dia e por mostrarem o vosso orgulho em todas as minhas conquistas.

Resumo

Num mundo cada vez mais globalizado e em que o tempo é escasso, todas as ferramentas que permitam otimizar tarefas rotineiras são alvo de grande adoção por parte da população. O setor da saúde não foge à regra e tem sido alvo de várias transformações, principalmente ao nível da eficiência dos serviços, pelo que ao longo da última década têm sido desenvolvidos serviços de saúde baseados na mobilidade – *Mobile Health (m-Health)*.

Neste contexto, o presente estudo visa analisar a importância deste novo tipo de entrega de serviços de saúde por via de tecnologias de comunicação móveis, através da observação da adoção de aplicativos móveis de saúde por parte dos utilizadores. Características como a mobilidade, a ubiquidade, a conectividade instantânea, a conveniência e a personalização são constantemente procuradas pelos utilizadores como forma de ultrapassar barreiras geográficas, temporais e organizacionais. Ora, uma vez que os dispositivos móveis (e em particular o *smartphone*) se revestem destes elementos, podemos facilmente constatar que cada vez mais a população recorre a estas plataformas pela sua capacidade facilitadora ao nível dos serviços de saúde.

Contudo, importa também referir que o *m-health* não são apenas aspetos positivos. A verdade é que a realidade aponta para inúmeras preocupações e questões levantadas tanto pelo utilizador final como pelos profissionais da área da saúde, entre as quais a temática da segurança e da privacidade.

O ponto fulcral da dissertação assenta na compreensão dos principais fatores que influenciam a maior ou menor variabilidade de classificação no ranking atribuído às *apps*.

Assim, procedeu-se à elaboração de uma base de dados que teve como principal fonte de informação a loja *online* de aplicativos móveis da *Android - GooglePlay/PlayStore*. Com um período temporal de observação de 5 meses e com atualizações mensais dos dados, foi possível definir quais as variáveis que mais influência têm na decisão dos utilizadores finais, a qual impactua a melhor ou pior classificação das *apps*.

Os resultados obtidos através da análise da base de dados foram combinados com a criação de quatro modelos de regressão logística, de modo a termos a perceção da efetiva correlação entre as variáveis selecionadas. Neste sentido, foi possível concluir que a melhoria de classificação no ranking das *apps* é fortemente influenciada pelo número de alterações na classificação do ranking das *apps* e o número de alterações na sua versão.

Palavras-chave: Mobile-Health, Adoção, Difusão, Paradigma Tecnológico, Saúde.

Abstract

In an increasingly globalized and time-scarce world, all the tools that optimize routine tasks are widely adopted by the population. The health sector does not fall outside the norm and has undergone several transformations, mainly in terms of service efficiency, so over the past decade mobility-based health services have been developed - Mobile Health (m-Health).

In this context, the present study aims to analyze the importance of this new type of delivery of health services through mobile communication technologies, through the observation of the adoption of mobile health applications by users. Features such as mobility, ubiquity, instant connectivity, convenience and customization are constantly being sought out by users as a way to overcome geographical, temporal and organizational barriers. However, since mobile devices (and in particular the smartphone) cover these elements, we can easily see that more and more people are using these platforms for their facilitating capacity in health services.

However, it should also be noted that m-health is not just embodied with positive aspects. However, it should also be noted that m-health is not just positive aspects. Reality points to numerous concerns and issues raised by final users and healthcare professionals, including the issue of security and privacy.

The core of the dissertation is based on the understanding of the main factors that influence the higher or lower classification variability in the ranking assigned to the applications.

Thus, a database was developed whose main source of information was the online store of *Android* mobile applications - Google Play / PlayStore. With a 5-month observation period and with monthly data updates, it was possible to define which variables have the most influence on the decision of the final users, which affects the best or worse classification of the applications.

The results obtained through the analysis of the database were combined with the creation of four models of logistic regression, in order to have the perception of the effective correlation between the selected variables. In this sense, it was possible to conclude that the improvement in the ranking of the apps is strongly influenced by the number of changes in the ranking of the applications and the number of changes in its version.

Keywords: Mobile-Health, Adoption, Diffusion, Technological Paradigms, Health.

Índice

1. Introdução.....	1
2. Revisão da literatura	7
2.1. Principais conceitos e sua aplicação no mercado da saúde	7
2.2. Utilização dos <i>smartphones</i> como plataforma facilitadora dos cuidados de saúde.....	11
2.3. Principais benefícios decorrentes da utilização de aplicações móveis de saúde	12
2.4. Principais preocupações decorrentes da utilização de aplicações móveis de saúde.....	14
2.5. A componente de jogo associada à utilização de aplicações móveis de saúde .	17
2.6. O programa m-Health Economics.....	18
3. Metodologia	23
4. Resultados	30
5. Discussão dos resultados.....	70
6. Considerações finais	74
Referências bibliográficas	75
Apêndice 1.....	78
Apêndice 2.....	80
Apêndice 3.....	85
Apêndice 4.....	86

Índice de tabelas

Tabela 1. Variáveis recolhidas no Google Play	24
Tabela 2. Conjunto de variáveis auxiliares utilizadas na base de dados	80
Tabela 3. Descrição das variáveis para o modelo de regressão logística múltipla (Logit1).....	Error! Bookmark not defined.
Tabela 4. Output relativo ao modelo de regressão logística múltipla	39
Tabela 5. Teste de Independência Qui-Quadrado	Error! Bookmark not defined.
Tabela 6. Descrição das variáveis para o modelo de regressão logística múltipla (Logit2).....	Error! Bookmark not defined.
Tabela 7. Output relativo ao modelo de regressão logística múltipla	45
Tabela 8. Teste de Independência Qui-Quadrado	Error! Bookmark not defined.
Tabela 9. Descrição das variáveis para o modelo de regressão logística múltipla (Logit3).....	Error! Bookmark not defined.
Tabela 10. Output relativo ao modelo de regressão logística múltipla	50
Tabela 11. Teste de Independência Qui-Quadrado	Error! Bookmark not defined.
Tabela 12. Descrição das variáveis para o modelo de regressão logística múltipla (Logit4).....	Error! Bookmark not defined.
Tabela 13. Output relativo ao modelo de regressão logística múltipla	57
Tabela 14. Teste de Independência Qui-Quadrado	Error! Bookmark not defined.
Tabela 15. Output relativo ao modelo de regressão logística múltipla	62
Tabela 16. Teste de Independência Qui-Quadrado	Error! Bookmark not defined.
Tabela 17. Output relativo ao modelo de regressão logística múltipla	67
Tabela 18. Teste de Independência Qui-Quadrado	Error! Bookmark not defined.
Tabela 19. Número de alterações ao nível da classe de downloads	34

Índice de figuras

.....	
Figura 1. Os vários ramos do Telehealth.....	3
Figura 2. Curva de difusão em S.....	7
Figura 4. O crescimento do “móvel” a nível mundial 2020.....	11
Figura 5. Utilização e penetração dos smartphones no mundo 2013-2018.....	12
Figura 7. Estimativa do número total de downloads de aplicativos m-health por plataforma (valores em mil milhões)	20

Índice de gráficos

Gráfico 1. Frequência das alterações no ranking	30
Gráfico 2. Frequência das alterações de versão (TOTAL)	32
Gráfico 3. Frequência das alterações de versão (CM)	32
Gráfico 4. Frequência das alterações de versão (SF).....	33

1. Introdução

O setor da saúde, à semelhança de outras indústrias, está a ser profundamente alterado pelas forças da globalização (PwC, 2015: 3), o que acarreta consigo a emergência de novos desafios e a conseqüente presença ativa dos vários atores do setor no sentido de se adaptarem à nova realidade. Lentamente, os pacientes vão mudando o seu perfil ao longo dos anos, revelando-se cada vez mais ativos e capazes de influenciar a indústria da saúde, obrigando a uma redefinição dos conceitos de saúde e de bem-estar (PWC, 2015: 3).

Com o desenvolvimento cada vez mais intenso das tecnologias e de inovações tecnológicas, e com o simultâneo aumento do número de utilizadores que adota essas mesmas tecnologias, criou-se um ambiente propício a um mundo crescentemente interoperável, móvel, conectado e dinâmico. De facto, características como a mobilidade, a ubiquidade, a conectividade instantânea, a conveniência e a personalização são constantemente procuradas pelos utilizadores como forma de ultrapassar barreiras geográficas, temporais e organizacionais (Silva, Rodrigues, de la Torre Díez, López-Coronado, & Saleem, 2015).

Neste sentido, o setor da saúde tem também sido alvo de várias transformações em anos mais recentes, em muitos casos desenvolvidas pelos próprios pacientes, os quais, cansados de filas de espera em instalações de saúde ou de esquecimentos de consultas e de levantamento de exames, sentiram necessidade de ter uma resposta mais imediata e próxima. Assim, surgiram os serviços de saúde baseados na mobilidade – *Mobile Health* (a partir de agora denominado de *m-health*) (PwC, 2015: 3).

O *m-health* é considerado um segmento do *e-health* (*Eletronic Health*), o qual, por sua vez, está inserido no panorama geral da “telesaúde” (*Telehealth*) (Akter, Ray, & D'Ambra, 2013). Neste sentido, importa ter em consideração a distinção destes conceitos para uma melhor compreensão do estudo em causa.

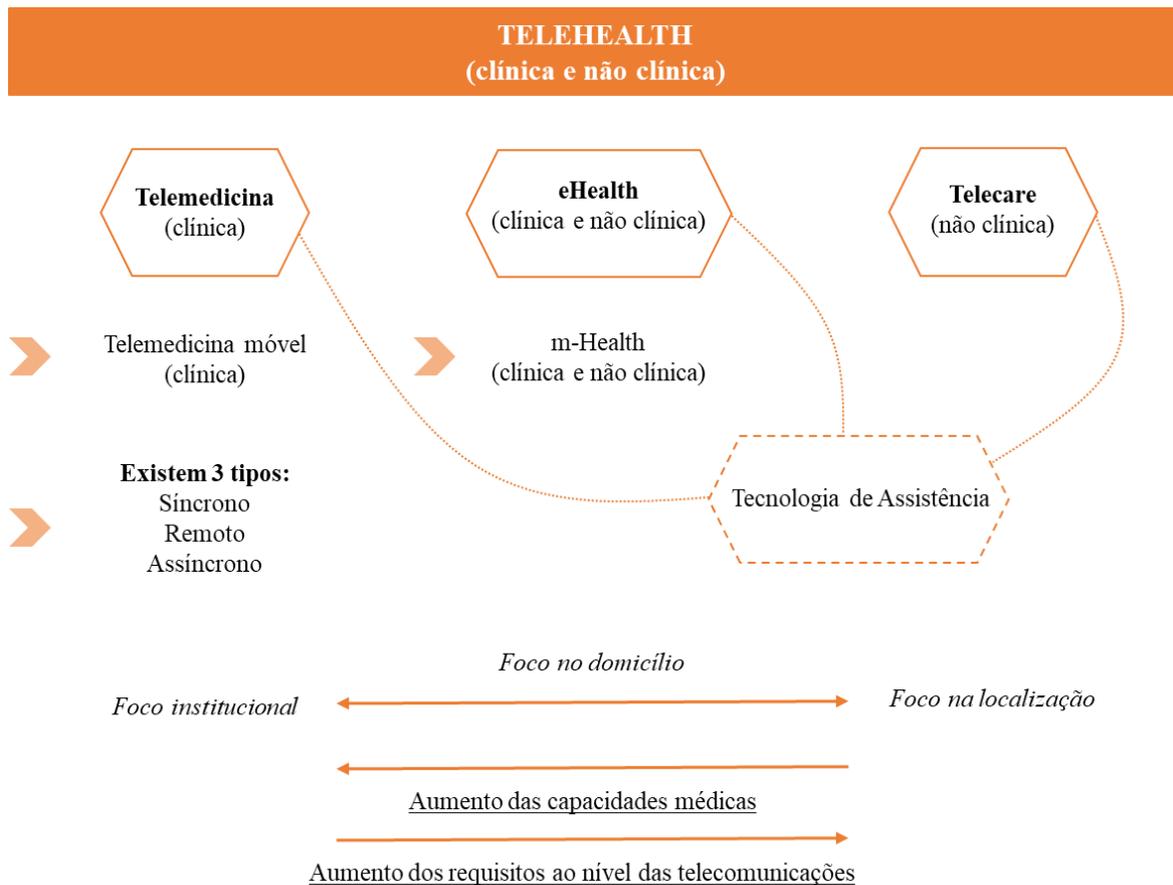
Assim, o “chapéu” que alberga todos estes serviços – *telehealth* ou telesaúde – consiste no fornecimento de serviços de saúde à distância, clínicos e não clínicos, que não exigem a presença física de um profissional de saúde (Akter et al., 2013). Já o *e-health* é um método da telesaúde para a utilização e transferência de recursos e cuidados de

saúde por meios eletrônicos, em grande parte através de tecnologias de informação e comunicação (TIC). Este segmento abrange três áreas principais: a entrega de informações a profissionais e consumidores de saúde através da Internet e das TIC; a utilização de práticas de e-commerce e e-business na gestão de sistemas de saúde; e a utilização das TIC e do e-commerce para melhorar os serviços de saúde pública. Como exemplos deste método temos o envio de emails entre pacientes e profissionais de saúde, o registo eletrónico de pacientes e os websites de informação sobre saúde (Akter et al., 2013).

Por fim, o *m-health*, uma parte do *e-health*, consiste na provisão por via eletrónica de informação e de serviços de saúde através de dispositivos de comunicação móveis, dada a sua capacidade de criar, armazenar, recuperar e transmitir dados em tempo real entre utilizadores finais, com o objetivo de melhorar quer a segurança dos mesmos, quer a qualidade dos serviços de saúde (Akter et al., 2013).

De uma forma mais ampla, o *m-health* envolve a utilização de tecnologias *wireless* que permitem a transmissão de vários conteúdos de dados e serviços, os quais são de fácil acesso para trabalhadores da área da saúde, através de dispositivos móveis, como *smartphones*, PDA's, computadores portáteis e tablets. Assim, é considerado um serviço de saúde personalizado e interativo com o objetivo de providenciar um acesso ubíquo e universal à informação e aconselhamento médico por parte de qualquer utilizador (Akter et al., 2013).

Figura 1. Os vários ramos do Telehealth



Fonte: Elaboração própria, com base em <https://doctordementia.com/2015/03/19/telehealth-terms-defined/> (consultado a 13/03/2018)

Neste contexto, o presente estudo visa analisar a importância deste novo tipo de entrega de serviços de saúde por via de tecnologias de comunicação móveis, através da observação da adoção de aplicações médicas por parte dos utilizadores, bem como da sua difusão (ou falta dela). Para isso, recorreu-se à criação de uma base de dados que teve como principal fonte de informação a loja *online* de aplicações móveis *GooglePlay*. O motivo da escolha pelo sistema *Android* ao invés do sistema *iOS*, prende-se, primeiramente, com o facto de desde o segundo trimestre de 2010 o primeiro ser o sistema operacional de *smartphones* preferencial no mundo em detrimento do segundo. Desde o seu lançamento, em 2007, as vendas de *smartphones* com o sistema *Android* têm registado um crescimento intenso, passando de uma quota de cerca de 30,5% no último trimestre de 2010 para uma quota de 81,7% no período homólogo do

ano de 2016. Por sua vez, o sistema da Apple – iOS – apesar da sua trajetória também ascendente, apresenta quotas de mercado significativamente inferiores ao sistema *Android*, fixando-se as suas quotas de mercado, e para os períodos anteriormente mencionados, em 15,8% e 17,9%, respetivamente. (Statista, 2018)

Para além disso, após uma análise inicial às duas plataformas, concluiu-se que a plataforma *Android* permite uma visualização mais acessível dos dados e disponibiliza informação mais detalhada.

A base de dados criada para este estudo inclui informações, para cada *app*, sobre o fornecedor da *app*, a data de atualização, a versão atual, a versão *Android* necessária e os preços (no caso das *apps* não gratuitas). Com base numa análise quantitativa dessa base de dados, pretende-se responder à seguinte questão de investigação: **Quais os fatores que mais influenciam a mudança de classificação no ranking das *apps*?**

A dissertação tem a seguinte estrutura: após esta Introdução, segue-se o segundo capítulo, o qual engloba a revisão da literatura relevante no âmbito das aplicações móveis para a saúde. Nele são apresentados os pontos de vista encontrados na literatura em relação aos conceitos abordados ao longo desta dissertação, tais como invenção, inovação, difusão, adoção, inovação tecnológica, paradigma tecnológico, inovação disruptiva, *telehealth*, *m-health* e *e-health*. Adicionalmente, são expostos estudos sobre a crescente utilização dos *smartphones* como plataforma móvel facilitadora dos cuidados de saúde e respetivos benefícios e desvantagens/precauções a ter na sua utilização. No terceiro capítulo é apresentada a metodologia adotada e a base de dados criada para a obtenção de resultados quanto à adoção e difusão (ou falta dela) dos aplicativos de saúde móvel. Seguidamente, no capítulo 4, apresenta-se a análise dos resultados do capítulo anterior; e, por fim, no capítulo 5 são enunciadas as principais conclusões do estudo, bem como as dificuldades e limitações encontradas e propostas e contribuições para pesquisas futuras.

2. Revisão da literatura

2.1. Principais conceitos e sua aplicação no mercado da saúde

Para uma melhor compreensão do presente estudo e da área científica em que se enquadra, torna-se quase imperativo falar do **ciclo de atividades orientadas para a inovação** (invenção – inovação – difusão - adoção). Assim, importa previamente distinguir os conceitos de invenção e inovação. Assim, “invenção é a ocorrência de uma ideia para um novo produto ou processo, enquanto a inovação é a comercialização dessa mesma ideia” (Fagerberg, 2003). Ou seja, a invenção é o primeiro passo para alcançar vantagem tecnológica e a inovação efetiva-se pela valorização e aplicação comercial da invenção, “através da combinação de diferentes tipos de conhecimento, capacidades, habilidades e recursos (regra geral, por parte de entidades empresariais) ” (Fagerberg, 2003).

Para que a inovação chegue aos utilizadores, aqueles que lhe dão valor, segue-se a fase de **difusão**, i.e., “processo através do qual as inovações são adotadas e utilizadas por pessoas e organizações” (Dogson, Gann, & Salter, 2008).

Por sua vez, Rogers segue uma linha de pensamento em que adiciona ao processo de difusão “ a sua comunicação através de certos canais longo do tempo e entre membros de um sistema social” (Rogers, 2003). Estes canais de comunicação têm como objetivo transmitir a mensagem de um indivíduo que tem conhecimento, ou experiência na utilização da inovação, para um indivíduo que ainda não tem conhecimento ou experiência com essa mesma inovação. Tais canais podem tomar a forma de *mass media* (televisão, rádio, jornais), canais interpessoais (comunicação face a face) e comunicação interativa (via Internet) (Rogers, 2003).

Entenda-se ainda como sistema social “um conjunto de unidades interrelacionadas (indivíduos, grupos informais, organizações e / ou subsistemas), envolvidas na resolução conjunta de problemas e com um objetivo comum” (Rogers, 2003).

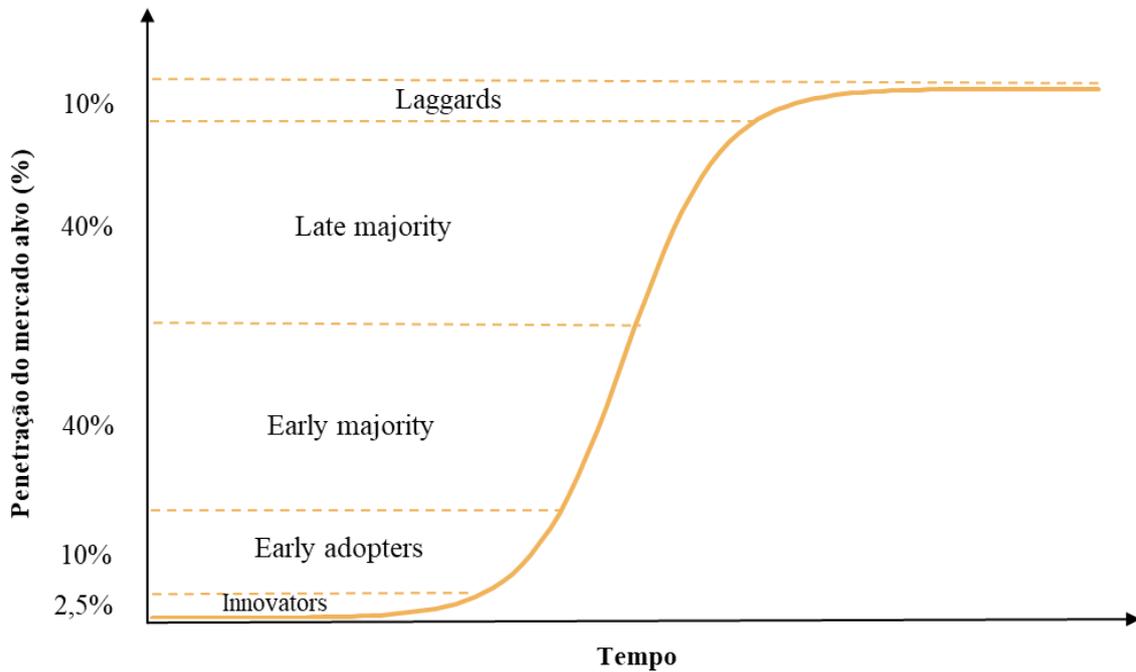
O processo de difusão assume geralmente a forma de **curva em S**, uma vez que a inovação tem impacto na eficiência dos processos e na criação de novas oportunidades na procura, gerando um impulso no crescimento. Tal impulso é considerado cíclico e não linear, dado o desfasamento verificado ao nível da adoção por cada membro do sistema social ao qual foi comunicada a inovação (Rogers, 2003).

Neste sentido, numa fase inicial, a difusão da inovação é lenta e começa por se desenvolver num pequeno número de adotantes, já que nem todos os potenciais utilizadores percebem a inovação da mesma forma, e também porque tanto o produto como o processo ainda não têm as suas características estandardizadas. Para além disso, a maior ou menor velocidade a que se propaga a inovação tem impacto na curva em S, dependendo esta velocidade dos canais de comunicação e das características da população adotante (Rogers, 2003).

À medida que a performance melhora e que os processos começam a estandardizar-se, os custos unitários começam a diminuir e é de esperar que a quota de utilizadores aumente até atingir o seu ponto de inflexão. Mais tarde, acabará por surgir uma nova inovação que tornará esta solução uma pior escolha em termos de relação performance-custo. O produto começará a tornar-se obsoleto e a curva passará a uma trajetória descendente de novos utilizadores.

Ora, para distinguir cada membro do sistema social, Rogers “categorizou-os por níveis de adoção da inovação (**Figura 2.**), tendo como critério a sua taxa de adoção, ou seja, a velocidade relativa a que cada inovação é adotada” (Rogers, 2003).

Figura 2. Curva de difusão em S



Fonte: Elaboração própria, com base em “The diffusion of innovation, InvestAura – The Art of Business Planning¹”

E como adoção entenda-se como o processo através do qual o adotante deixa de ter apenas conhecimento inicial de uma inovação e passa a ter uma opinião formada para assumir uma atitude de rejeição ou aceitação da mesma.

Na linha de pensamento da inovação, importa também distinguir dois conceitos relevantes para o estudo: o de inovação tecnológica e o de paradigma tecnológico. Enquanto uma **inovação tecnológica** compreende as atividades que contribuem para a investigação, desenvolvimento e design de novos produtos, serviços ou processos, ou para a melhoria de produtos já existentes, gerando, assim, novo conhecimento tecnológico², um **paradigma tecnológico** “diz respeito a um modelo amplamente aceite que visa a resolução de problemas tecnológicos, com base em princípios derivados das ciências naturais e das tecnologias” (Dosi, 1982). A construção de Dosi permite

¹ <http://www.business-planning-for-managers.com/main-courses/forecasting/the-diffusion-of-innovation/>, consultado a 18/02/2018

² <http://www.igi-global.com/dictionary/technological-innovation/29457>, consultado a 27/01/2017.

entender a evolução tecnológica como algo altamente seletivo, direcionado e cumulativo na aquisição de capacidades e instrumentos para resolução de problemas. “A principal função de um paradigma tecnológico passa pelo fornecimento de “puzzles” para os investigadores resolverem, bem como as ferramentas necessárias para a sua resolução” (Dosi, 1982).

De acordo com Cimoli & Dosi (1995) um paradigma tecnológico tem na sua génese três ideias fundamentais:

- 1) Sugere que qualquer descrição considerada satisfatória do conceito de tecnologia e as transformações por si desencadeadas devem incorporar a representação de formas específicas de conhecimento pelas quais uma determinada atividade se orientou. De uma forma geral, a tecnologia não deve ser reduzida a uma visão padrão de um conjunto de projetos bem definidos, até porque a tecnologia engloba atividades de resolução de problemas que envolvem formas tácitas de conhecimento incorporadas tanto em indivíduos como em procedimentos organizacionais;
- 2) Os paradigmas implicam heurísticas específicas e perspetivas sobre "como fazer as coisas" e “como melhorá-las”, perspetivas essas muitas vezes compartilhadas pelas comunidades de cada atividade. Isto é, os paradigmas envolvem quadros cognitivos compartilhados de forma coletiva (Constant, 1985);
- 3) Os paradigmas, regra geral, definem modelos básicos de artefactos e de sistemas que, ao longo do tempo, são continuamente modificados e melhorados. Estes artefactos básicos podem ainda ser descritos por meio de fundamentos tecnológicos e de características económicas.

Quando a comunidade perde a confiança na capacidade do paradigma existente em resolver problemas relevantes, segue-se a sua substituição por um novo paradigma tecnológico, o qual rompe com as práticas científicas anteriores e introduz novas regras” (Dosi, 1982).

Com o advento das comunicações móveis usando dispositivos de *smart mobile* que suportam redes móveis 3G e 4G para o transporte de dados, a computação móvel tem sido a principal atração das comunidades de pesquisa e empresariais. O *m-health* é, portanto, a nova vantagem na inovação em saúde, estando mesmo a ser considerado como sendo um novo paradigma tecnológico. Propõe-se saúde em qualquer momento

e em qualquer lugar, superando barreiras geográficas, temporais e até organizacionais. Os sistemas de *m-health* e as funcionalidades de mobilidade a eles associados têm um forte impacto em sistemas típicos de monitorização e alerta de cuidados de saúde, recolha de dados administrativos, manutenção de registos, cuidados de saúde, informação médica, deteção e sistemas de prevenção (Silva et al., 2015).

Os serviços típicos de *m-health* utilizam a Internet e serviços Web para fornecer uma interação autêntica entre médicos e pacientes. Um médico ou um paciente pode facilmente aceder ao mesmo registo a qualquer hora e em qualquer lugar através do seu computador pessoal, *tablet* ou *smartphone*. O paciente pode entrar em contacto com o profissional de saúde em caso de emergência, ou mesmo ter acesso a registos médicos ou compromissos, independentemente da hora ou lugar (Silva et al., 2015).

Ora, as mudanças constantes e simultâneas nos paradigmas tecnológicos interrompem a organização industrial existente e aumentam as oportunidades para as empresas empreendedoras (Ho & Lee, 2015). Adicionalmente, estas mudanças podem muitas vezes interromper regras, ordens, crenças e valores estabelecidos na indústria. O impacto dessa interrupção pode ser tão profundo que pode ameaçar a sobrevivência das empresas que não se conseguem adaptar (Ho & Lee, 2015). As organizações terão sucesso somente se estiverem cientes das novas condições e forem capazes de superar a inércia organizacional e abraçar a mudança. Já em 1942, Schumpeter definiu a **destruição criadora** como o processo pelo qual os empresários criam continuamente valor ao mesmo tempo que destroem valores antigos através do desenvolvimento de inovações tecnológicas disruptivas (Schumpeter, 1962). Dada a natureza complexa das tecnologias que rapidamente alteram, a rutura de paradigmas tecnológicos pode ser difícil de caracterizar e reconhecer. As empresas que desconsideram o carácter disruptivo da inovação tecnológica podem ser ultrapassadas por novos concorrentes que dominam os novos paradigmas tecnológicos (Ho & Lee, 2015). É precisamente neste sentido que o *m-health* está a ser considerado o novo paradigma que está a transformar os serviços de saúde por todo o mundo (Akter et al., 2013)

As tendências demográficas, de mercado e tecnológicas tornam a transformação digital cada vez mais crítica para o futuro dos cuidados de saúde. Assim, um estudo³ levado a

³ World Economic Forum (em colaboração com a Accenture) (2017), “*Digital Transformation Initiative: Unlocking \$100 Trillion for Business and Society from Digital Transformation*”, pp. 50-51

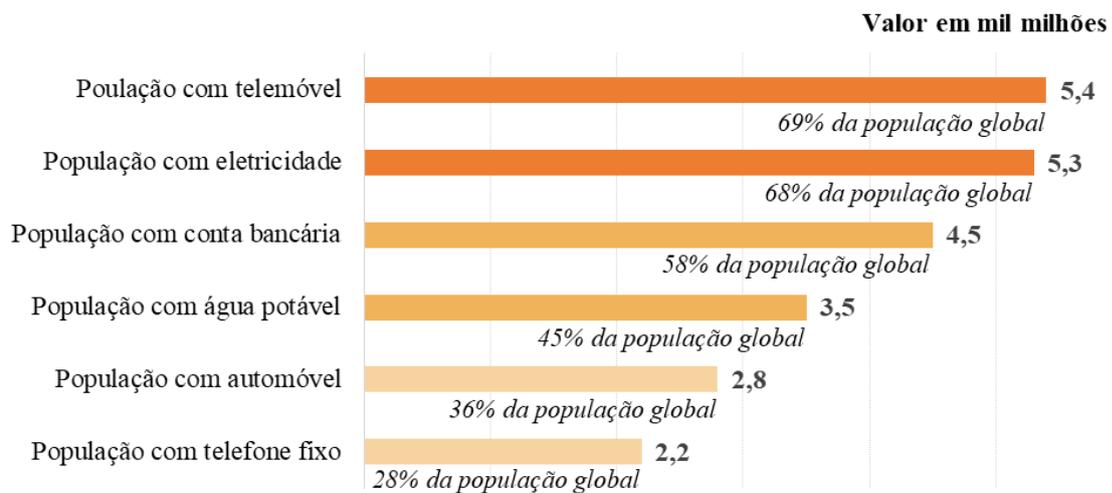
cabo pelo World Economic Forum em colaboração com a Accenture identifica quatro temas que são apontados como geradores de um novo valor para a indústria da saúde e para a sociedade em geral: **cuidados de saúde inteligentes** (avanços tecnológicos e científicos recentes impulsionaram a medicina para uma nova era de cuidados inteligentes; a medicina de precisão, a robótica e a impressão médica estão a tornar os cuidados de saúde mais inteligentes e personalizados); **cuidados de saúde em qualquer lugar** (a mudança dos cuidados de saúde para locais mais próximos dos pacientes pode ampliar o acesso aos mesmos e reduzir a pressão sobre os sistemas de saúde; a tecnologia que permitiu a proliferação da Internet of Things é a mesma que abre possibilidade a este tipo de iniciativa); **cuidados de saúde cada vez mais capacitados** (a economia digital pode agora oferecer uma gama de “serviços vivos” – serviços digitais inteligentes que respondem contextualmente às necessidades do utilizador; este tipo de serviços aumenta o envolvimento em escala do paciente, capacitando-o da gestão dos seus próprios cuidados de saúde, prevenindo assim o aparecimento de condições crónicas, como a diabetes); **empresas de saúde inteligentes** (os avanços na recolha, armazenamento e análise de dados foram acompanhados pela proliferação dos mesmos – por exemplo, a partir de sensores e de sistemas de dispositivos com informações clínicas e registos de saúde eletrónicos). Em conclusão, esse estudo aponta para que o modelo atual de prestação de cuidados de saúde é cada vez mais insustentável. Para proporcionar melhorias contínuas no setor da saúde, e em todo o mundo, o mesmo terá de ser transformado, com o digital a desempenhar um papel central (World Economic Forum, 2017).

2.2. Utilização dos *smartphones* como plataforma móvel facilitadora dos cuidados de saúde

A tecnologia móvel tem o potencial de revolucionar a forma como os médicos praticam medicina, e nos últimos anos foi crescente o número de médicos que começaram a substituir vários dispositivos (como o PDA) por um *smartphone* (Burdette, Herchline, & Oehler, 2008).

O *smartphone* é sem dúvida uma das histórias de sucesso das últimas décadas. A tecnologia móvel inteligente penetrou rapidamente na sociedade, capturando todo um espectro de subscritores, desde jovens até idosos. Esta rápida aceitação de dispositivos de comunicação móvel tem tido um forte impacto não só na forma de comunicação, como também ao nível da cultura, comunidade, identidade e relacionamento com a envolvente. (Boulos, Wheeler, Tavares, & Jones, 2011).

Figura 3. O crescimento do “móvel” a nível mundial | 2020

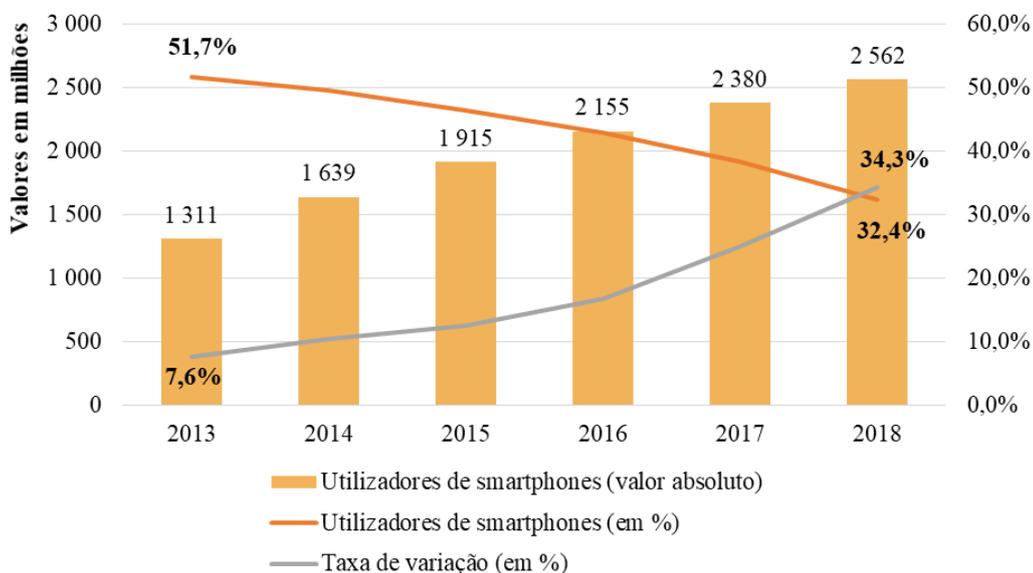


Fonte: Elaboração própria, com base em CISCO, 2016

De acordo com a líder em tecnologias de informação CISCO, a perspetiva é de que os dispositivos e conexões móveis inteligentes representem, no ano de 2020, 72% do total de dispositivos e conexões móveis e sejam responsáveis por 98% do tráfego de dados móveis. Adicionalmente, as estatísticas preveem que nesse ano a população com pelo menos um telemóvel vai exceder o número de pessoas com eletricidade (Figura 3). De uma ótica de dispositivo individual, os *smartphones* dominam o tráfego móvel e a sua proliferação está a aumentar tão rapidamente que se espera que mais pessoas tenham

smartphones (5,4 mil milhões) do que eletricidade (5,3 mil milhões), água corrente (3,5 mil milhões) e automóvel (2,8 mil milhões).

Figura 4. Utilização e penetração dos smartphones no mundo | 2013-2018



Fonte: Elaboração própria, com base em eMarketer, 2014

2.3.Principais benefícios decorrentes da utilização de aplicações móveis de saúde

Como referido anteriormente, a adoção de *smartphones* no âmbito da saúde é cada vez mais comum. A população tenta aproveitar ao máximo as facilidades disponíveis ao seu alcance para aprimorar atividades consideradas por si mais importantes e para proteger os seus ativos mais valiosos, e nenhum ativo é mais precioso do que a própria saúde. A utilização das Tecnologias da Informação e Comunicação (TIC) na prática de cuidados de saúde introduziu a noção de *e-health*, em que as telecomunicações permitem a telemedicina, os computadores processam os dados de saúde e a internet fornece a infraestrutura necessária para trocar todo o tipo de informação médica e de serviços. Quando a mobilidade emergiu, as telecomunicações começaram a ocorrer por via dos telemóveis, e os computadores a mover-se com as pessoas na forma de computadores portáteis e, em seguida, sob a forma dos mais variados dispositivos móveis (como os tablets) (Hussain et al., 2015)

A verdade é que a utilização do *smartphone* enquanto plataforma facilitadora de cuidados de saúde do utilizador apresenta um conjunto alargado de benefícios: os *smartphones* são ágeis, portáteis e podem ser usados em movimento, possuem vários recursos e permitem novas possibilidades através dos aplicativos instalados. Como se podem conectar à internet, os *smartphones* são úteis para manter os utilizadores atualizados com as últimas técnicas e avanços médicos. A conexão contínua dos *smartphones* permite ainda a partilha de dados comportamentais e de saúde com os profissionais de saúde. A crescente capacidade de os *smartphones* usarem sensores internos para inferir contexto como a localização do utilizador, movimento, emoção e compromisso social, aumentou a possibilidade de um acompanhamento contínuo e automatizado de comportamentos relacionados com a saúde e intervenções oportunas e personalizadas para contextos específicos (Hussain et al., 2015).

- **Benefícios relacionados com a agilidade/portabilidade dos dispositivos de suporte (em particular do *smartphone*):** os *smartphones* são ágeis, portáteis e podem ser usados em movimento. Esta mobilidade e portabilidade capacitam as aplicações móveis de uma comunicação oportuna, sendo, por isso, ideais para manter um diário de sintomas enquanto acompanham os utilizadores. Do ponto de vista da investigação, os aplicativos de *smartphones* permitem aos pacientes capturar dados que descrevem a sua experiência no local em tempo real e também permitem repetidas amostragens de comportamento e capturar eventos menos frequentes (Hussain et al., 2015);
- **Benefícios relacionados com os recursos dos *smartphones*:** os *smartphones* possuem vários recursos e permitem novas possibilidades através dos aplicativos instalados. Como se podem conectar à Internet, os *smartphones* são úteis para manter os clínicos atualizados com as últimas técnicas e avanços médicos. A conexão contínua dos *smartphones* permite a partilha de dados comportamentais e de saúde com profissionais de saúde. Essa habilidade também pode permitir que a telemedicina substitua visitas demoradas ao consultório médico. Para além disso, os aplicativos médicos para *smartphone* são criados especificamente para auxiliar ao nível de questões como pontos de atendimento, a dosagem de fármacos e os regimes de tratamento convencionais, agregando as informações necessárias e apresentando-as num formato

facilmente compreensível. A crescente capacidade dos *smartphones* usarem sensores internos para inferir contexto como a localização do utilizador, movimento, emoção e compromisso social, aumentou a possibilidade de um acompanhamento contínuo e automatizado de comportamentos relacionados com a saúde e intervenções oportunas e personalizadas para contextos específicos. Além de recursos de *hardware* e software, o recurso do anonimato concedido por aplicativos permite que os pacientes façam perguntas que não fariam caso estivessem perante um profissional de saúde (Hussain et al., 2015);

- **Benefícios relacionados com a penetração de mercado dos *smartphones*:** a popularidade e a ubiquidade dos *smartphones* permitem o acesso a populações que são difíceis de alcançar ou de comprometer. Para os pacientes que não podem aceder a instalações de prestação de cuidados de saúde, as aplicações médicas são especialmente benéficas (Hussain et al., 2015).

2.4.Principais preocupações/desafios decorrentes da utilização de aplicações móveis de saúde

Não obstante todos os benefícios da utilização do *smartphone* e dos aplicativos médicos, a realidade aponta para inúmeras preocupações e questões levantadas tanto pelo utilizador final como pelos profissionais da área da saúde. Neste sentido, são listados abaixo os principais desafios ao nível desta temática:

- **Qualidade dos aplicativos:** As principais questões incluem a preocupação com o baixo envolvimento de profissionais qualificados no desenvolvimento dos aplicativos, a falta de averiguação externa e a revisão por pares após a sua entrada no mercado, a falta de evidência da eficácia clínica, a falta de investigação objetiva para avaliar resultados, a ausência de regulação de conteúdo e a ausência de um quadro normativo que padroniza o processo de desenvolvimento (Hussain et al., 2015);

- **Privacidade dos pacientes/utilizadores e a segurança dos seus dados pessoais:** As *apps* e os serviços prestados são gratuitos para o indivíduo, mas a privacidade não é garantida. As informações recolhidas sobre um indivíduo que utiliza o aplicativo e os serviços a ele associados podem ser usados diretamente pela empresa para marketing direcionado ou vendidas a terceiros para marketing e desenvolvimento de produtos. Para além disso existem sempre riscos de segurança para os utilizadores menos experientes, os quais podem ser induzidos a fazer o *download* de aplicativos que contêm *malware* ou que lhes forneça informações e conselhos médicos duvidosos (Hussain et al., 2015);
- **Auto-monitorização:** a utilização de aplicativos para fins médicos por parte dos pacientes pode conduzir a quadros de depressão e pode ter mais impactos negativos do que positivos. Além disso, aplicativos que fornecem aconselhamento médico com base numa recolha própria de dados e algoritmos podem levantar preocupações desnecessárias ou falsas reafirmações. Outros desafios incluem erros e omissões em configurações de cuidados de saúde, resultante de distrações e interrupções causadas pela interação com outros aplicativos e respetivas notificações. No que respeita ao relacionamento com profissionais de saúde, a excessiva dependência na comunicação por via de aplicativos pode resultar numa diminuição do nível da comunicação verbal e da visibilidade das ações por parte destes para com os seus pacientes (Hussain et al., 2015).

Tendo em consideração os aspetos positivos e negativos mencionados anteriormente, e para melhorar a qualidade e a confiabilidade dos aplicativos, todos os utilizadores, incluindo pacientes e médicos, são aconselhados a usar os serviços de avaliação disponíveis. Um exemplo notável é a “Happtique”, uma empresa de soluções de saúde móvel que lançou o seu Programa de Certificação de *Apps* de Saúde: um programa voluntário projetado para auxiliar médicos e pacientes a identificar mais facilmente aplicativos seguros e de confiança (Hussain et al., 2015). Os pacientes são ainda recomendados a analisar cuidadosamente as revisões encontradas em lojas *online* de *apps* e a considerar fontes alternativas de aplicativos antes de proceder à sua utilização. Os médicos, por sua vez, são incentivados a trocar informações com outros médicos usando um aplicativo específico, por forma a tê-lo testado previamente antes da sua

prescrição a um paciente. Do ponto de vista da segurança, os utilizadores são aconselhados a garantir que os dados sejam criptografados e que o *smartphone* esteja protegido por um código de segurança antes de usar qualquer aplicativo através do qual as informações de saúde privadas possam ser transmitidas ou armazenadas (Hussain et al., 2015).

Helf e Hlavacs (2016) ressaltam a atual falta de utilização de teorias científicas no desenvolvimento de aplicativos, a inserção ineficiente e seletiva de uma componente de jogo e os baixos níveis de personalização e confiança neste domínio. Dada a natureza multidisciplinar deste conjunto de problemas, os autores propuseram estruturas integradas e centradas no utilizador como uma potencial solução para resolver os problemas que tanto os utilizadores como os desenvolvedores de aplicações enfrentam hoje. À medida que os *smartphones* se transformam em veículos centrais de entrega de serviços de saúde (Hovenga, Kidd, Garde, & Hullin, 2009), os criadores de aplicativos médicos passaram a incorporar uma componente de jogo, considerando-a uma ferramenta importante para aumentar o envolvimento dos utilizadores com as *apps*, tentando assim transformar tarefas desafiadoras relacionadas com a atividade física ou com a alimentação em experiências divertidas (Oh, Rizo, Enkin, & Jadad, 2005). Contudo, o rápido desenvolvimento nesta área tem criado um fosso crescente entre produtos de consumo e validação científica. Além disso, os criadores estão a competir uns contra os outros, medindo, por isso, o seu sucesso por meio de receitas financeiras, quando a medida que deveria ser utilizada era o nível de melhoria da saúde dos pacientes que utilizassem este tipo de plataforma. Ora, tal comportamento tem conduzido a um maior aparecimento de conflitos por objetivos, bem como ao crescente ceticismo entre os utilizadores em relação ao enviesamento comercial e à confiabilidade dos aplicativos disponíveis (Helf & Hlavacs, 2016).

2.5.A componente de jogo associada à utilização de aplicações móveis de saúde e a vertente de personalização

A implementação de uma componente de jogo nos aplicativos móveis relacionados com a saúde emergiu como uma ferramenta que visa aumentar o envolvimento do indivíduo e transformar obstáculos que impedem mudanças comportamentais em experiências divertidas. Originalmente colocado em prática em 2008, o processo de incluir elementos de jogo em contextos *no-gaming* foi aplicado na criação de tabelas de classificação, a fim de permitir a comparação de esforço com outros indivíduos (Deterding, Dixon, Khaled, & Nacke, 2011)

A verdade é que aplicações com uma componente de jogo podem ter impactos perceptivos e cognitivos substanciais no utilizador, aumentando a atenção seletiva ou o envolvimento emocional, o que acaba por influenciar os níveis de compromisso com a *app*. Particularmente no domínio da saúde, a utilização frequente de uma *app* pode ter efeitos diretos na mudança comportamental e, assim, impulsionar a melhoria desejada na saúde do utilizador.

No entanto, existe um aspeto bastante discutido por Helf e Hlavacs (2016): a personalização (ou falta dela) do *m-health*. No seu estudo, estes autores identificaram dois tipos de personalização – personalização “dentro de *apps*” e personalização “entre *apps*”. A personalização “dentro de *apps*” é a personalização de conteúdo e serviços com base nas previsões das necessidades do utilizador. Ou seja, estamos a falar de *apps* móveis capazes de alterar de forma dinâmica o conteúdo tendo em consideração o perfil de cada utilizador e o ambiente em que este se encontra inserido, com o intuito de proporcionar uma utilização mais efetiva e um maior compromisso do utilizador com o sistema.

Por outro lado, a personalização entre *apps* refere-se aos fatores que determinam as opções de *apps* por parte do utilizador. Dogruel et al. levaram a cabo um estudo interessante sobre os processos de decisão heurística em relação à seleção de aplicativos na plataforma *Android*. Os resultados mostraram que a maioria dos utilizadores aplica a heurística de “escolher a primeira”, isto é, utilizam regras de pesquisa simples em vez de participar em análises completas e comparativas. Dado que as classificações foram identificadas como a fonte mais importante de informação para o utilizador, pode-se

concluir que a apresentação de aplicações escolhida pelas empresas responsáveis das lojas *online* de aplicativos, como a Google e a Apple, tem um impacto considerável na escolha final dos utilizadores.

A aparente falta de adaptação a esses recursos e circunstâncias tem implicações importantes sobre como os utilizadores se comprometem atualmente com o grande número de aplicativos disponíveis: com o fim de iniciar e manter mudanças comportamentais, o utilizador tende a instalar um número significativo de aplicações (Helf & Hlavacs, 2016).

Deste modo, torna-se imperativo apostar num conjunto de potenciais soluções, entre as quais (Helf & Hlavacs, 2016):

- Reforço da colaboração entre desenvolvedores, profissionais de saúde ou entidades públicas;
- Alinhamento das metas dos desenvolvedores, para melhorar a saúde do utilizador;
- Integração de resultados de investigação empírica de tópicos relacionados com a saúde ou conhecimento de profissionais de saúde durante o desenvolvimento e uso de aplicações;
- Adaptação dinâmica das intervenções de saúde oferecidas pelos aplicativos ao indivíduo, com base nos tipos de personalidade e ambiente envolvente do utilizador;
- Inclusão de ferramentas de avaliação, de modo a criar mais evidências para o processo de personalização

2.6. O programa mHealth Economics

Segundo o relatório “mHealth App Economics 2017” elaborado pela consultora em questão, o mercado de aplicativos para dispositivos móveis está prestes a completar 10 anos de existência. Embora houvesse várias soluções de saúde digitais anteriores, foi em 2008 que a Apple possibilitou o que é agora um mercado de massas para aplicativos de saúde móveis, lançando a Apple App Store.

No setor de cuidados de saúde tradicional, 10 anos é meramente o período de tempo para um ciclo de desenvolvimento de produtos. Porém, na esfera digital, uma década é considerada uma janela temporal bastante significativa. A verdade é que apesar dos dois ambientes serem muito diferentes, a indústria digital acelerada e o setor de cuidados de saúde com ritmo lento têm colidido na saúde digital, trazendo consigo mudanças disruptivas para o mercado.

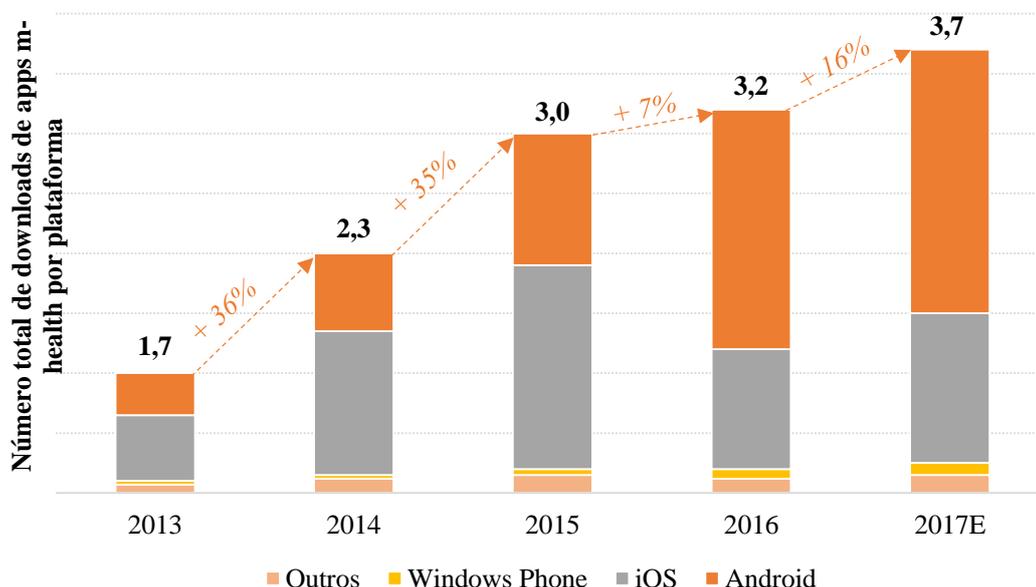
O mercado de saúde móvel tem vindo a crescer de forma constante ao longo dos últimos anos, tendo sido contabilizados, no ano de 2017, 325 mil aplicativos de saúde (divididos entre as categorias de saúde & fitness e de aplicativos médicos).

De facto, no último ano, 78 mil novos aplicativos de saúde foram adicionados às *app stores*, crescimento esse alimentado principalmente por um aumento de aplicativos *Android*.

O sistema operativo *Android* registou uma taxa de crescimento de 50% entre 2016 e 2017, em comparação com a taxa de 20% registada pelos aplicativos de saúde do iOS. O *Android* já ultrapassou o iOS como a plataforma número principal para aplicativos de saúde e possui o maior número de aplicativos de dispositivos móveis de qualquer *app store*.

Os 325.000 aplicativos de saúde acima mencionados são equivalentes a 3,7 mil milhões de *downloads* de aplicativos estimados para 2017, o que corresponde a um aumento de 16% face ao ano transato (**Figura 5**). O crescimento do número de *downloads* é impulsionado, na sua grande maioria, por *downloads* do *Android* e do iOS, desempenhando as restantes plataformas um papel marginal.

Figura 5. Estimativa do número total de downloads de aplicativos m-health por plataforma (valores em mil milhões)



Fonte: Elaboração própria, com base em Research2Guidance – *mHealth App Developer Economics* study 2017; n = 2.400

Outro fator determinante para o crescimento do mercado digital da saúde é o crescimento de empresas *startups* alimentadas por investimentos iniciais de aceleradores, incubadoras e empresas de capital de risco. Em 2016, os investidores de fase inicial investiram cerca de 5 mil milhões de dólares no mercado mundial de *startups* digitais.

Uma grande limitação para o mercado digital de saúde prende-se com as regulamentações rígidas e muitas vezes complexas que contribuem para o lento desenvolvimento das soluções digitais de saúde. A pesquisa da *mHealth Economics* revelou que 18% das partes interessadas em saúde digital são impedidas de desenvolver aplicativos devido a condições regulatórias incertas.

O setor de saúde é um dos mais sujeitos a uma regulação intensa, pelo que as soluções de saúde digitais e não digitais que podem representar um risco para a segurança do paciente devem ser esclarecidas por um órgão regulador aprovado, como é o caso do FDA (Food and Drug Administration) nos EUA.

No entanto, há sinais de grandes mudanças futuras na regulação da saúde digital. Em julho de 2017, o órgão regulador mais proeminente nos EUA - a FDA - anunciou uma nova abordagem para aprovar soluções de saúde digital – o chamado Plano de Inovação Digital de Saúde - em que ao invés de se aprovar produtos digitais individuais, empresas inteiras podem ser aprovadas e os produtos digitais lançados por essas empresas pré-selecionadas não têm que passar por um processo regulatório para cada um dos lançamentos dos seus produtos. Este desenvolvimento ainda é muito recente, mas a FDA tem impulsionado uma mudança de paradigma na regulação da saúde digital.

Ao nível europeu, a Agenda Digital para a Europa adotada pela UE leva a cabo “o plano de ação para a saúde em linha, o qual abrange domínios tão diversificados como os direitos dos doentes em matéria de cuidados de saúde transfronteiras, o financiamento de projetos de investigação e desenvolvimento avançados e a garantia de que os sistemas de registos de saúde eletrónicos são compatíveis a nível internacional. Prevê igualmente um roteiro para aumentar a capacidade de intervenção dos doentes e dos técnicos de saúde, interligar dispositivos e tecnologias e investir numa investigação direcionada para a medicina personalizada do futuro. Isto implica prestar serviços de saúde mais inteligentes, mais seguros e mais centrados no doente. Tendo em conta a rápida e crescente penetração dos tablets e dos telemóveis inteligentes, o plano de ação também dá especial atenção à «saúde móvel» (*mHealth*)” (Comissão Europeia, 2018).

Os países com melhores condições de mercado para soluções de saúde digital são os EUA e o Reino Unido. O motivo da atratividade destas duas regiões é a "dimensão de mercado atrativa", seguindo-se o "acesso aos investidores" e a "aceitação de aplicativos por médicos". Embora os editores de aplicativos tendam a escolher o seu próprio país como atraente, o país líder para ambientes favoráveis de saúde digital continua a ser os EUA.

Comparando os EUA com o Reino Unido, o seu PIB, a dimensão da população e a percentagem de gastos com saúde revelam os motivos fundamentais da sua maior atratividade.

Em suma, e de acordo com o relatório “*mHealth App Economics 2017*”, os mecanismos de pesquisa via Internet destacam o desenvolvimento do interesse pela saúde digital entre as empresas de telecomunicações (outsiders) e a indústria da saúde (insiders). Ou seja, à medida que o mercado de saúde digital se expandiu e amadureceu, foram-se

registrando cada vez menos pesquisas na internet por “saúde digital”, enquanto o interesse entre os profissionais de saúde foi fortalecido.

Neste sentido, a dissertação em causa vai ao encontro do progresso que se tem verificado ao longo da última década ao nível do *mobile health*, dando ênfase à análise da trajetória das *apps* adotadas pelos utilizadores finais. Esta análise, que teve como suporte a criação de uma base de dados com as várias alterações registadas ao longo de 6 observações mensais, teve como fim primordial compreender que tipo de fatores influenciam a alteração da classificação no ranking das aplicações móveis de saúde. Este ranking, construído pela fonte de onde foi extraída toda a informação necessária para a criação da base de dados (loja *online* do sistema operativo *Android* – Google Play/Play Store), constitui a variável em torno da qual tentamos encontrar a razão para certas *apps* se difundirem com sucesso e continuarem a ter versões atualizadas, ao contrário de outras que acabam por ter apenas uma única versão e depois desaparecem.

3. Metodologia

O propósito deste estudo é comprovar a importância crescente das aplicações de saúde na sociedade através do número de atualizações/versões ao longo de 5 meses de observação direta, bem como compreender o porquê de certas *apps* se difundirem com sucesso e continuarem a ter versões atualizadas, ao contrário de outras que acabam por ter apenas uma única versão e depois desaparecem.

Para tal, recorreu-se ao método de investigação quantitativo, por ser dedutivo, explicativo, objetivo e mensurável. Este método, de origem positivista, favorece o desenvolvimento de pesquisas orientadas para a procura de factos e/ou causas dos fenómenos (Easterby-Smith, Thorpe, & Jackson, 2012). O objetivo desta abordagem é estabelecer e provar relações entre variáveis que são definidas de forma operacional, i.e., de forma a poderem ser quantificadas direta ou indiretamente. Neste contexto, os dados obtidos numa pesquisa quantitativa são habitualmente quantitativos e estruturados, ou seja, os resultados obedecem a um padrão previamente definido, quer pela operacionalização das variáveis, quer pelos instrumentos de pesquisa (Oliveira & Ferreira, 2014).

Neste sentido, e tendo como suporte a loja *online Google Play/Play Store*, foi construída uma base de dados (ver Apêndice 2), desenvolvida propositadamente para este trabalho de investigação, uma vez que não existe um reportório de aplicativos móveis, nem foi possível obter as informações diretamente com o proprietário da loja *online* (apesar de várias tentativas com a Google). Esta base de dados é relativa a uma amostra de *apps* de *m-health* (um total de 540 *apps*), subdividida em dois segmentos “Cuidados Médicos” (CM) e “Saúde e Fitness” (SF). A partir destes segmentos, procedeu-se à divisão das mesmas em 4 grandes categorias, as quais foram definidas por autoria própria tendo como base a descrição de cada *app* na loja *Google Play/Play Store*. Subsequentemente, detalhou-se com mais profundidade a dimensão destas 4 categorias, seguindo os critérios anteriormente referidos.

No que respeita à amostra definida, de um total de 540 *apps* (270 para cada categoria), definiu-se uma proporção de 2/3 para as *apps* pagas e de 1/3 para as *apps* gratuitas. Ou seja, 180 são *apps* pagas (90 CM + 90 SF) e 360 são gratuitas (180 CM + 180 SF). Este

critério de seleção reflete o facto de as *apps* gratuitas constituírem uma realidade substancialmente maior que as *apps* pagas.

A informação foi recolhida de dezembro de 2016 a abril de 2017, tendo sido atualizada mensalmente, com as seguintes variáveis (retiradas como dados secundários⁴):

Tabela 1. Variáveis recolhidas no Google Play

<i>Variável</i>	<i>Explicação da variável</i>	<i>Observações</i>
Name	Nome da <i>app</i>	
Supplier	Nome do fornecedor	
Category	Cada segmento foi dividido em 4 grandes categorias	- Apoio/Suporte - Monitorização de saúde - Saúde feminina - Outros
Subcategory	Cada uma das 4 categorias foi subdivida em várias subcategorias	Apoio/Suporte - Didático - Recursos médicos - Organização -Outros Monitorização de saúde - Medicação - Treino físico - Alimentação Saúde feminina - Gravidez - Outros Outros - Outros
SF/CM	A amostra escolhida engloba 2 segmentos: Saúde & Fitness (SF) e Cuidados Médicos (CM)	
Free/Paid	De um total de 540 <i>apps</i> (270 para cada categoria), definiu-se uma proporção de 2/3 para as <i>apps</i> pagas e de 1/3 para as <i>apps</i> gratuitas. Ou seja, 180 são <i>apps</i> pagas (90 CM + 90 SF) e 360 são gratuitas (180 CM + 180 SF)	
O1	Observação 1 (data da recolha dos dados)	14,17, 18 dezembro 2016
O2	Observação 2 (data da recolha dos dados)	14, 16, 18 janeiro 2017

⁴ Dados que já foram recolhidos e que estão sujeitos a tratamento estatístico

O3	Observação 3 (data da recolha dos dados)	14, 17, 18 fevereiro 2017
O4	Observação 4 (data da recolha dos dados)	14, 17, 18 março 2017
O5	Observação 5 (data da recolha dos dados)	14, 17, 18 abril 2017
O6	Observação 6 (data da recolha dos dados)	14, 17, 18 maio 2017
P1	Preço 1 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 1)	[0; 35,99]
P2	Preço 2 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 2)	[0; 35,99]
P3	Preço 3 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 3)	[0; 35,99]
P4	Preço 4 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 4)	[0; 35,99]
P5	Preço 5 ((Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 5)	[0; 35,99]
P6	Preço 6 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 6)	[0; 35,99]
PI1	Produtos integrados em aplicações 1 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 1)	Refere-se apenas às <i>apps</i> gratuitas
PI2	Produtos integrados em aplicações 2 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 2)	Refere-se apenas às <i>apps</i> gratuitas
PI3	Produtos integrados em aplicações 3 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 3)	Refere-se apenas às <i>apps</i> gratuitas
PI4	Produtos integrados em aplicações 4 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 4)	Refere-se apenas às <i>apps</i> gratuitas
PI5	Produtos integrados em aplicações 5 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 5)	Refere-se apenas às <i>apps</i> gratuitas
PI6	Produtos integrados em aplicações 6 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 6)	Refere-se apenas às <i>apps</i> gratuitas

PV1	Versão atual 1 (no momento da observação 1)	
PV2	Versão atual 2 (no momento da observação 2)	
PV3	Versão atual 3 (no momento da observação 3)	
PV4	Versão atual 4 (no momento da observação 4)	
PV5	Versão atual 5 (no momento da observação 5)	
PV6	Versão atual 6 (no momento da observação 6)	
Date 1	Data de atualização 1 (data da última atualização da <i>app</i> indicada no momento da observação 1)	[23 de Maio de 2012; 27 de Janeiro de 2016]
Date 2	Data de atualização 2 (data da última atualização da <i>app</i> indicada no momento da observação 2)	[23 de Maio de 2012; 18 de Janeiro de 2017]
Date 3	Data de atualização 3 (data da última atualização da <i>app</i> indicada no momento da observação 3)	[26 de Julho de 2013; 28 de Fevereiro de 2017]
Date 4	Data de atualização 4 (data da última atualização da <i>app</i> indicada no momento da observação 4)	[26 de Julho de 2013; 19 de Março de 2017]
Date 5	Data de atualização 5 (data da última atualização da <i>app</i> indicada no momento da observação 5)	[26 de Julho de 2013; 18 de Abril de 2017]
Date 6	Data de atualização 6 (data da última atualização da <i>app</i> indicada no momento da observação 6)	[26 de Julho de 2013; 18 de Maio de 2017]
Dwnl 1	Classe de <i>downloads</i> 1 (no momento da observação 1)	D1 = 5 - 10 D2 = 10 - 50 D3 = 50 - 100 D4 = 100 - 500 D5 = 500 - 1.000 D6 = 1.000 - 5.000 D7 = 5.000 - 10.000 D8 = 10.000 - 50.000 D9 = 50.000 - 100.000 D10 = 100.000 - 500.000 D11 = 500.000 - 1.000.000 D12 = 1.000.000 - 5.000.000 D13 = 5.000.000 - 10.000.000 D14 = 10.000.000 - 50.000.000 D15 = 50.000.000 - 100.000.000

		D16 = 100.000.000 - 500.000.000
Dwnl 2	Classe de <i>downloads</i> 2 (no momento da observação 2)	<i>Idem</i>
Dwnl 3	Classe de <i>downloads</i> 3 (no momento da observação 3)	<i>Idem</i>
Dwnl 4	Classe de <i>downloads</i> 4 (no momento da observação 4)	<i>Idem</i>
Dwnl 5	Classe de <i>downloads</i> 5 (no momento da observação 5)	<i>Idem</i>
Dwnl 6	Classe de <i>downloads</i> 6 (no momento da observação 6)	<i>Idem</i>
Android Version	Versão <i>Android</i> requerida para cada <i>app</i>	[1.5; 4.4]

O passo seguinte à recolha mensal destes dados foi a criação de variáveis auxiliares, de modo a possibilitar uma melhor análise e a permitir a utilização de estatística mais avançada, como regressões logísticas.

Assim, privilegiaram-me medidas de variação, tais como *Rankchanges*, *PVchanges*, *Pchanges*, *Dwnlchanges* (ver Apêndice 2), de modo a compreender quais os elementos influenciadores para as trajetórias seguidas pelas *apps* (melhor, pior ou igual classificação entre observações).

Para testar a associação entre variáveis, ou seja, para verificar em que grau duas variáveis se comportam de forma semelhante, recorreu-se ao Teste de Independência Qui-Quadrado (χ^2). Como a maioria dos testes estatísticos, precisamos de definir uma hipótese nula⁵ (H0) e uma hipótese alternativa (H1), pelo que temos:

H0: as variáveis são independentes vs H1: as variáveis são dependentes

⁵ A hipótese nula, denotada como H0, é a negação da hipótese que o investigador considera verdadeira. A rejeição ou aceitação da hipótese nula depende do p-value e do nível de significância (α) associado às variáveis em questão: p-value > α → aceita-se H0; p-value < α → rejeita-se H0 e aceita-se a hipótese considerada correta pelo investigador (Hildebrand, Lyman Ott, & Brian Gray, 2004)

Rejeita-se a hipótese de independência (H0 ou hipótese nula) entre as variáveis quando o *p-value*⁶ associado ao valor da estatística de teste é inferior a um certo valor crítico (no caso em questão o intervalo de confiança definido foi de 0,05⁷).

Após esta análise, segue-se a construção de um modelo de regressão logística de modo a explorar possíveis explicações que permitam responder à questão “**Quais os fatores que mais influenciam a mudança de classificação no ranking das apps?**”. Para tal, recorreu-se ao programa RStudio.

Entenda-se por regressão logística, (ou modelo logit) “um processo de modelar variáveis de resultado dicotômicas. No modelo logit, o logaritmo das chances⁸ do resultado são modeladas como uma combinação linear das variáveis preditoras” (Institute for Digital Research and Education, tradução livre).

Uma vez que no decorrer da análise da base de dados se optou por dividir a variável de melhoria de classificação do ranking em duas – *Rankimp* e *Rankimp_maintain*⁹ – foram realizados quatro modelos de regressão logística múltipla¹⁰ (ver códigos do R no Apêndice 3):

- *mylogit1 <- glm(formula = Rankimp ~ Rankchanges + PVchanges + Dwnlchanges, family = "binomial", data = Logit)*
- *mylogit2 <- glm(formula = Rankimp ~ Rankups + Rankdowns + Dwnlup + Dwnldowns, family = "binomial", data = Logit2)*
- *mylogit3 <- glm(formula = Rankimp_maintain ~ Rankchanges + PVchanges + Dwnlchanges, family = "binomial", data = Logit3)*

⁶ O *p-value* é a probabilidade (assumindo H0 – hipótese nula) de um valor estatístico de teste ser igual ou superior ao valor efetivamente observado, para um dado nível de significância. Quanto mais distante da região de rejeição a estatística de teste cair, menor será o *p-value*, e mais fortes serão as evidências para rejeitar a hipótese nula e apoiar a hipótese da investigação (Hildebrand et al., 2004)

⁷ O intervalo de confiança (e respetivo nível de significância) foram definidos em 95% (e 0,05) por uma questão de generalização estatística.

⁸ Para falarmos em logaritmo das chances temos, imperativamente, que falar em chances. Uma chance é definida como o rácio entre a probabilidade de sucesso e a probabilidade de falha. Ou seja, as chances de um sucesso (por exemplo, 80% de chance de chover) têm sempre as chances da sua falha (por exemplo, 20% de chance de não chover); se tivermos em conta o exemplo refrido, o rácio das chances é $0,8/0,2 = 4$

Assim, o logaritmo das chances é feito por meio do seu simples cálculo ($\ln(4)$)

⁹ Esta divisão decorreu do facto de várias *apps* manterem a sua classificação no ranking ao longo das 6 observações

¹⁰ Neste, caso, “múltipla” apenas para relevar o facto de que foi utilizada mais do que uma variável independente, pois uma regressão logística múltipla (*mlogit*) é considerada múltipla quando utiliza mais do que duas variáveis independentes (<https://stats.idre.ucla.edu/stata/dae/multinomiallogistic-regression/>, consultado em 13/09/2018)

- `mylogit4 <- glm(formula = Rankimp_maintain ~ Rankups + Rankdowns + Dwnlup + Dwnldowns, family = "binomial", data = Logit4)`

Por fim, a partir do output das regressões logísticas, segue-se a criação de um Teste de Significância Estatística (ou teste de hipóteses), por forma a interpretar os coeficientes que nos conduzem aos resultados que nos dão resposta à questão de investigação. Este teste segue a lógica do Teste de Independência Qui-Quadrado, pelo que testa uma suposição em relação a uma amostra estatística, com o objetivo de aceitar ou rejeitar uma hipótese nula (H_0), tendo como valor de referência o *p-value*. Ou seja, consideram-se variáveis estatisticamente significativas aquelas que têm um *p-value* associado inferior ao nível de significância definido (0,05), o que permite descartar a hipótese nula e aceitar a associação entre as variáveis.

4. Resultados

Através da criação da base de dados, que serviu de suporte a toda a dissertação, foi desenvolvida uma análise de estatística descritiva¹¹ que permitiu, em primeira mão, abrir caminho para a definição das variáveis incluídas em cada modelo de regressão logística.

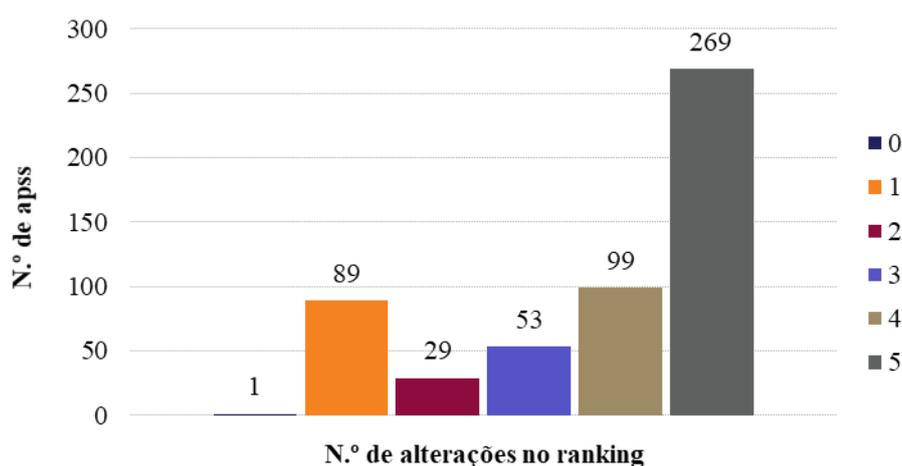
Esta estatística permitiu-nos chegar às seguintes conclusões:

- A partir da análise da variável **Rankchanges** podemos concluir que a maioria das *apps* (269) altera a sua posição no ranking a cada observação, enquanto apenas 1 não sofre qualquer alteração ao longo das 6 observações.

Importa ainda referir que as *apps* gratuitas são mais suscetíveis a alterações do que as *apps* pagas (67% das *apps* com 5 alterações ao longo das 6 observações são *apps* gratuitas) (Gráfico 1).

Ou seja, existe uma grande variabilidade no ranking das *apps*, pelo que partimos do pressuposto que esta variável é uma variável explicativa da melhoria na classificação do ranking.

Gráfico 1. Frequência das alterações no ranking



¹¹ A estatística descritiva, cujo objetivo é sintetizar uma série de valores da mesma natureza, permitindo dessa forma ter uma visão mais global da variação dos mesmos, organiza e descreve os dados de 3 formas possíveis: por meio de tabelas, de gráficos e de medidas descritivas.

- Tendo em consideração a variável **Rankimp**:
 - são 144 as *apps* que acabaram o período de observação com uma posição no ranking superior à observada no início do trabalho (65 respeitam a *apps* CM - 41 gratuitas e 24 pagas; 79 respeitam a *apps* SF - 43 gratuitas e 36 pagas);
 - por outro lado, são 396 as *apps* que viram a sua classificação igual ou pior entre a 1ª e a última observação (205 respeitam a *apps* CM - 139 gratuitas e 66 pagas; 191 respeitam a *apps* SF - 137 gratuitas e 54 pagas).

Tanto num caso como noutro, as *apps* CM destacam-se pela sua frequência face às *apps* SF, e o mesmo se pode dizer das *apps* gratuitas em detrimento das *apps* pagas.

Para além disso, podemos concluir que as *apps* cuja classificação se mantém ou deteriora estão em maioria, representando 73% do total da amostra.

- Passando a análise para a variável **Rankimp_maintain**:
 - são 132 as *apps* que acabaram o período de observação com uma posição no ranking igual ou superior à observada no início do trabalho (60 respeitam a *apps* CM - 40 gratuitas e 20 pagas; 72 respeitam a *apps* SF - 42 gratuitas e 30 pagas);
 - por outro lado, são 408 as *apps* que viram a sua classificação deteriorar-se entre a 1ª e a última observação (210 respeitam a *apps* CM - 140 gratuitas e 70 pagas; 198 respeitam a *apps* SF - 138 gratuitas e 60 pagas).

Uma vez mais, são as *apps* que descem no ranking que representam o conjunto mais significativo.

- Quanto à variável **PVchanges**, que quantifica o número de alterações de versão pelas quais cada *app* passou, a maioria das *apps* analisadas (301) não sofre qualquer alteração na sua versão ao longo das 6 observações. Destas 301, 170 respeitam a Cuidados Médicos, enquanto as restantes 131 pertencem ao segmento de Saúde & Fitness. Importa ainda salientar que, deste conjunto, as *apps* gratuitas são as mais suscetíveis a alterações face às *apps* pagas.

Seguem-se as *apps* com apenas uma alteração na versão (94), com 2 alterações (53), com 3 alterações (47), com 4 alterações (29) e, finalmente, com 5 alterações da sua versão (16) (Gráfico 2).

Quando passamos a uma análise mais fina, constatamos facilmente que as *apps* com uma maior frequência de alterações na versão (2, 3, 4 ou 5 alterações ao longo das 6 observações) correspondem a *apps* de SF, enquanto as *apps* de CM ocupam as menores frequências de alteração na versão (nenhuma alteração ou apenas uma) (Gráfico 3 e Gráfico 4). Neste caso, são também as *apps* gratuitas as que apresentam uma maior suscetibilidade para mudanças de versão.

Gráfico 2. Frequência das alterações de versão (TOTAL)

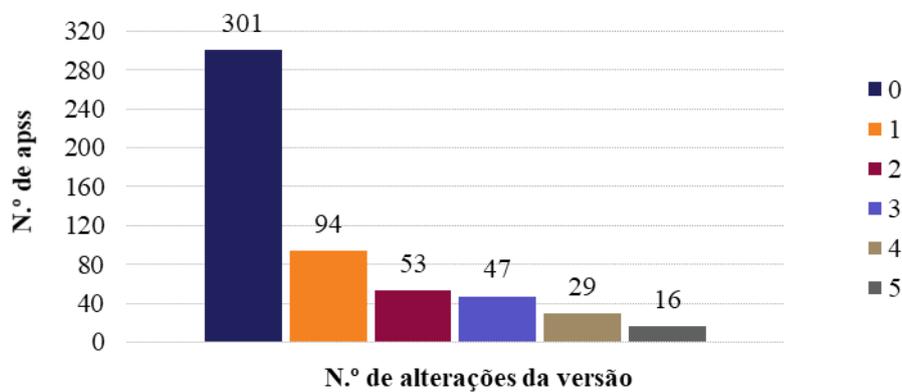


Gráfico 3. Frequência das alterações de versão (CM)

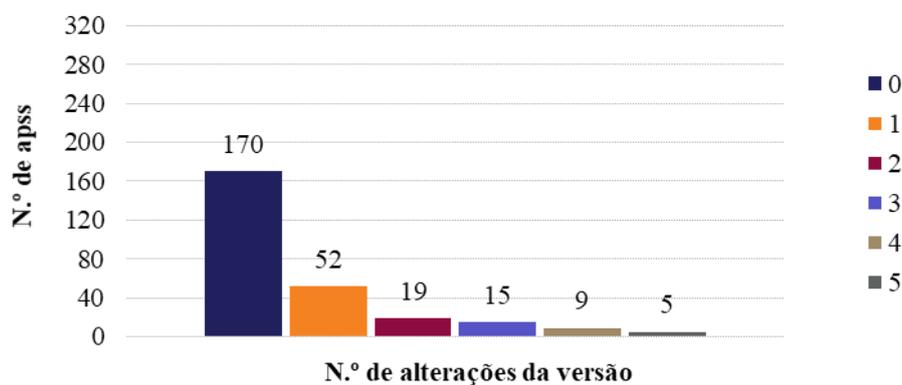
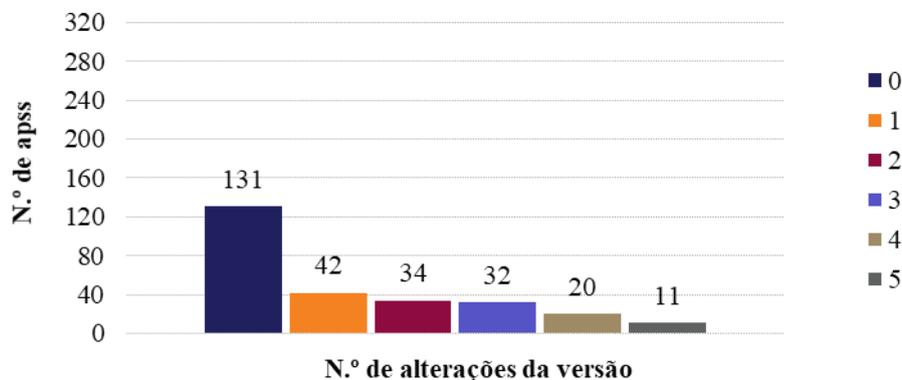


Gráfico 4. Frequência das alterações de versão (SF)



- Por fim, relativamente à variável *Dwnlchanges* (que representa o número de alterações na classe de *downloads* de cada *app*), de salientar que dada a grande distância entre cada classe, é normal que a maioria das *apps* mantenha a sua classe entre a primeira e a última observação. Na verdade, são 476 as *apps* que não registam qualquer alteração no seu número de *downloads*, sendo apenas 59 as que registam uma única alteração e 5 as que alteram duas vezes. Importa ainda referir que neste último caso, as duas alterações correspondem a uma descida de classe de *downloads*, enquanto as restantes alterações respeitam todas a um aumento no número de *downloads*.

Das 59 *apps* que sofreram uma subida na classe de *downloads*, 17 são de CM (13 gratuitas e 4 pagas) e 42 pertencem ao segmento de SF (35 gratuitas e 7 pagas).

Por sua vez, esta alteração correspondeu a uma subida de apenas uma classe para a maioria das *apps*: 52 *apps* (13 de CM e 39 de SF). No entanto, verificou-se também que 4 *apps* subiram duas classes, 1 subiu quatro classes, 1 subiu 7 classes e 1 subiu 11 classes (**Tabela 2**).

Tabela 2. Número de alterações ao nível da classe de downloads

<i>Categoria</i>	<i>Subcategoria</i>	<i>CM/SF</i>	<i>Free/Paid</i>	<i>Aumento do n.º de classes</i>
Apoio/Suporte	Didático	CM	Free	2 (D6 → D8)
Saúde feminina	Ciclo menstrual	CM	Free	2 (D6 → D8)
Outros	Relaxar	SF	Paid	2 (D4 → D6)
Monitorização de saúde	Treino físico	SF	Free	2 (D8 → D10)
Monitorização de saúde	Diabetes	CM	Paid	4 (D6 → D10)
Apoio/Suporte	Aconse./Diagn./Comum. Apoio	CM	Free	7 (D6 → D13)
Monitorização de saúde	Treino físico	SF	Free	11 (D2 → D13)

Esta análise estatística permitiu constatar que as *apps* que registam uma trajetória de descida no ranking representam o conjunto mais significativo da amostra (73%). Contudo, os fatores que influenciam tal descida não são suscetíveis de ser encontrados através deste tipo de estatística, pelo que se recorreu à construção de modelos de regressão logística.

Neste sentido, neste capítulo são ainda apresentados os quatro modelos de regressão logística criados com o objetivo de identificar as variáveis que mais impacto têm no comportamento da amostra selecionada e constituída por um total de 540 *apps*.

Tal como referido no capítulo anterior, esta análise inicia-se com um Teste de Independência Qui-Quadrado, após o qual se segue a apresentação dos modelos de regressão logística e, por fim, um Teste de Significância Estatística que nos guia para os resultados principais.

4.1. Modelo de Regressão Logística 1

Primeiramente, recorreu-se ao Teste de Independência Qui-Quadrado (χ^2) para testar a associação da variável dependente *Rankimp* com cada uma das variáveis (Tabela 3 independentes selecionadas (Tabela 4):

Tabela 3. Descrição das variáveis para o modelo de regressão logística múltipla (Logit1)

<i>Abreviatura</i>	<i>Variável</i>	<i>Unidades de medida</i>
Rankimp	<i>Variável dependente relativa ao resultado da diferença entre a 1^a e a 6^a observação: melhorou, piorou ou ficou na mesma posição do ranking (situação em que o zero está excluído)</i>	<i>0 (piorou/igual) e 1 (melhorou)</i>
Rankchanges	<i>Variável independente discreta relativa ao número de alterações na posição do ranking entre observações</i>	<i>[0; 5]</i>
PVchanges	<i>Variável independente discreta relativa ao número de alterações ocorridas na versão da app</i>	<i>[0; 5]</i>
Dwnlchanges	<i>Variável independente discreta relativa ao número de alterações ocorridas no intervalo de downloads</i>	<i>[0; 2]</i>

Tabela 4. Teste de Independência Qui-Quadrado (1)

1. Rankimp + Rankchanges

```
> Q1 <- df [c ("Rankimp", "Rankchanges", "PVchanges", "Dwnlchanges")]  
> table (Q1$Rankimp, Q1$Rankchanges)
```

		Rankchanges					
		0	1	2	3	4	5
Rankimp	0	1	0	2	6	28	107
Rankimp	1	0	89	27	47	71	162

```
> chisq.test(table (Q1$Rankimp, Q1$Rankchanges))  
Pearson's Chi-squared test  
data: table(Q1$Rankimp, Q1$Rankchanges)  
X-squared = 71.068,  
df = 5,  
p-value = 6.143e-14
```

2. Rankimp + PVchanges

```
> Q2 <- df [c ("Rankimp", "Rankchanges", "PVchanges", "Dwnlchanges")]  
> table (Q2$Rankimp, Q2$PVchanges)
```

		PVchanges					
		0	1	2	3	4	5
Rankimp	0	71	21	13	16	12	11
Rankimp	1	230	73	40	31	17	5

```
> chisq.test(table (Q2$Rankimp, Q2$PVchanges))
```

Pearson's Chi-squared test

```
data: table(Q2$Rankimp, Q2$PVchanges)  
X-squared = 21.49,  
df = 5,  
p-value  
= 0.0006543
```

3. Rankimp + Dwnlchanges

```
> Q3 <- df [c ("Rankimp", "Rankchanges", "PVchanges", "Dwnlchanges")]  
> table (Q3$Rankimp, Q3$Dwnlchanges)
```

		Dwnlchanges		
		0	1	2
Rankimp	0	125	17	2
Rankimp	1	351	42	3

```
> chisq.test(table (Q3$Rankimp, Q3$Dwnlchanges))
```

Pearson's Chi-squared test

```
data: table(Q3$Rankimp, Q3$Dwnlchanges)
```

X-squared = 0.63376,
df = 2,
p-value = 0.7284

Tendo em consideração as hipóteses *H0: as variáveis são independentes vs H1: as variáveis são dependentes*, e a partir do output supra apresentado podemos retirar as seguintes conclusões:

- Para o caso 1., o valor do χ^2 é de 71.068 e o *p-value* associado é de 6.143e-14, significativo para um intervalo de confiança de 0,05, pelo que podemos rejeitar a hipótese nula e concluir que as variáveis não são independentes ou, por outras palavras, que estão associadas uma à outra (existência de associação entre a melhoria de classificação no ranking da *app* e o número de alterações no ranking que a mesma sofre entre a 1.^a e a 6.^a observação);
- Para o caso 2., o valor do χ^2 é de 21.49 e o *p-value* é 0.0006543, inferior ao valor crítico definido, o que nos permite, também, rejeitar a hipótese nula e afirmar a existência de associação entre a melhoria de classificação no ranking da *app* e o número de alterações na versão que a mesma sofre ao longo das 6 observações;
- Por fim, o caso 3. revela-se contrário aos anteriores, uma vez que o *p-value* associado ao valor do χ^2 é de 0,7284, superior ao intervalo de confiança de 0,05, logo não significativo, pelo que a hipótese nula não é rejeitada, mostrando a independência das variáveis em causa. Ou seja, a melhoria de classificação no ranking das *apps* não está associada ao aumento ou diminuição do número de *downloads* da mesma.

Segue-se a construção efetiva do primeiro modelo de regressão logística, cujo modo de apresentação segue uma lógica sequencial, tendo em conta os passos dados no programa de suporte à análise dos resultados (RStudio). Assim, primeiramente, são apresentadas as variáveis sobre as quais a regressão incide, bem como a fórmula utilizada para o modelo *glm*. Seguidamente, é apresentado o output relativo ao modelo obtido (*summary*), que nos fornece os valores estimados dos coeficientes, valores esses que

nos elucidam das variáveis que são estatisticamente significativas para o modelo em questão.

Uma vez que o estudo em causa pretende compreender as constantes alterações de classificação no ranking dos aplicativos móveis de saúde, definiu-se como variável dependente do modelo a variável *Rankimp* (melhoria na posição do ranking entre a primeira e a última observação). Nesta regressão assumimos que uma melhor classificação no ranking implica uma maior adoção da *app* por parte dos utilizadores, mesmo que esse não seja o real significado do ranking (uma vez que existem *apps* cuja classificação se manteve inalterada, apesar de terem sido adotadas por mais utilizadores, i.e., terem subido nos patamares de números de utilizadores).

Por outro lado, temos como variáveis explicativas *Rankchanges*, *PVchanges* e *Dwnlchanges*. Esta assunção será depois complementada e corrigida com as regressões seguintes que incorporam essa estagnação – *Rankimp_maintain*.

A obtenção do modelo logístico implica que se analise previamente cada uma das variáveis selecionadas, de modo a que se garanta a sua correta classificação, bem como a não existência de valores nulos, os quais, por defeito, foram codificados como NA. Assim, procedeu-se à correta classificação de cada uma das variáveis, em caracteres e numéricas e fatores ordenados (ver Apêndice 4).

Depois de definidas todas as variáveis, a fase seguinte englobou a obtenção do modelo de regressão logística múltipla (a partir de agora denominado de *mylogit1*), através da função *glm*:

```
mylogit1 <- glm(formula = Rankimp ~ Rankchanges + PVchanges + Dwnlchanges,  
family = "binomial", data = Logit1)
```

O output relativo ao modelo obtido através do comando *summary* (*mylogit1*) encontra-se exposto na tabela seguinte:

Tabela 5. Output relativo ao modelo de regressão logística múltipla

```
> mylogit1 <- glm(formula = Rankimp ~ Rankchanges + PVchanges + Dwnlchanges, family
= "binomial", data = Logit1)
> summary(mylogit1)

Call:
glm(formula = Rankimp ~ Rankchanges + PVchanges + Dwnlchanges, family = "binomial",
data = Logit1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6676  -1.0908   0.4704   0.9456   1.3523

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   4.50459    0.55084   8.178 2.89e-16 ***
Rankchanges  -0.78627    0.11888  -6.614 3.74e-11 ***
PVchanges    -0.19511    0.07401  -2.636 0.00838 **
Dwnlchanges   0.19986    0.29153   0.686 0.49300
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 626.3 on 539 degrees of freedom
Residual deviance: 539.56 on 536 degrees of freedom
AIC: 547.56

Number of Fisher Scoring iterations: 5
```

Seguidamente, passamos à análise dos valores estimados dos coeficientes presentes no output. Para tal, recorreu-se a um teste de significância estatística (ou teste de hipóteses), o qual corresponde a uma regra decisória que nos permite rejeitar ou não rejeitar uma hipótese estatística com base nos resultados de uma amostra e para um dado nível de significância.

Temos, então, como hipóteses:

H0: as variáveis são independentes vs H1: as variáveis são dependentes,

para um nível de significância de 0,05 (e, conseqüentemente, intervalo de confiança de 95%).

A partir do output da Tabela 5. pode-se concluir que apenas as variáveis *Rankchanges* e *PVchanges* são estatisticamente significativas para o modelo em questão. O *p-value* associado a estas variáveis é inferior a um nível de significância de 0,05, o que permite descartar a hipótese nula e aceitar a associação entre as variáveis *Rankchanges* e *PVchanges* e a melhoria de classificação no ranking das *apps*.

A interpretação dos coeficientes baseia-se no seguinte raciocínio (Institute for Digital Research and Education):

- Coeficiente > 1, aumenta a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão;
- Coeficiente < 1, diminui a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão.

Estes coeficientes são ainda alvos de um cálculo ($e^{\text{coeficiente}}$), cujo resultado nos indica a maior ou menor probabilidade de o evento ocorrer, dada a variável independente que está a causar a mudança.

Assim, e *ceteris paribus*¹²:

- Quando a variável *Rankchanges* regista uma variação positiva de uma unidade, as chances de a *app* melhorar a sua classificação no ranking diminuem em $e^{-0.786} = 0.46$ (31,5%¹³).
- Por sua vez, quando a variável *PVchanges* regista uma variação positiva de uma unidade, as chances de a *app* melhorar a sua classificação no ranking diminuem em $e^{-0.195} = 0.82$ (45,1%¹⁴).

Como ambas as probabilidades são inferiores a 50%, significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que as variáveis *Rankchanges* e *PVchanges* piorem a classificação no ranking das *apps* do que a sua melhoria.

¹² Expressão latina que corresponde à assunção de que não há mudança nas restantes variáveis.

¹³ Chance = $(p/(1-p)) \rightarrow 0,46 = (p/(1-p)) \rightarrow p = 31,5\%$

¹⁴ Chance = $(p/(1-p)) \rightarrow 0,82 = (p/(1-p)) \rightarrow p = 45,1\%$

4.2. Modelo de Regressão Logística 2

Primeiramente, recorreu-se ao Teste de Independência Qui-Quadrado (χ^2) para testar a associação da variável dependente *Rankimp* com cada uma das variáveis (Tabela 6) independentes selecionadas (Tabela 7).

Tabela 6. Descrição das variáveis para o modelo de regressão logística múltipla (Logit2)

<i>Abreviatura</i>	<i>Variável</i>	<i>Unidades de medida</i>
Rankimp	<i>Variável dependente relativa ao resultado da diferença entre a 1ª e a 6ª observação: melhorou, piorou ou ficou na mesma posição do ranking (situação em que o zero está excluído)</i>	<i>0 (piorou/igual) e 1 (melhorou)</i>
Rankups	<i>Variável independente discreta relativa ao número de subidas no ranking entre observações</i>	<i>[0; 5]</i>
Rankdowns	<i>Variável independente discreta relativa ao número de descidas no ranking entre observações</i>	<i>[0; 5]</i>
Dwnlups	<i>Variável independente discreta relativa ao número de subidas ocorridas no intervalo de downloads</i>	<i>[0; 2]</i>
Dwnldowns	<i>Variável independente discreta relativa ao número de descidas ocorridas no intervalo de downloads</i>	<i>0 e 2</i>

Tabela 7. Teste de Independência Qui-Quadrado(2)

1. Rankimp + Rankups

```
> Q1 <- df [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns")]
> table (Q1$Rankimp, Q1$Rankups)
```

		Rankups					
		0	1	2	3	4	5
Rankimp	0	2	8	46	73	14	1
Rankimp	1	115	125	117	39	0	0

```
> chisq.test(table (Q1$Rankimp, Q1$Rankups))
Pearson's Chi-squared test
data: table(Q1$Rankimp, Q1$Rankups)
X-squared = 192.67,
df = 5,
p-value < 2.2e-16
```

2. Rankimp + Rankdowns

```
> Q2 <- df [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns ")]
> table (Q2$Rankimp, Q2$Rankdowns)
```

		Rankdowns					
		0	1	2	3	4	5
Rankimp	0	5	25	82	30	2	0
Rankimp	1	1	106	120	126	37	6

```
> chisq.test(table (Q2$Rankimp, Q2$Rankdowns))

Pearson's Chi-squared test

data: table(Q2$Rankimp, Q2$Rankdowns)
X-squared = 49.585,
df = 5,
p-value = 1.685e-09
```

3. Rankimp + Dwnlups

```
> Q3 <- df [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns ")]
> table (Q3$Rankimp, Q3$Dwnlups)
```

		Dwnlups		
		0	1	2
Rankimp	0	125	17	2
Rankimp	1	352	42	2

```
> chisq.test(table (Q3$Rankimp, Q3$Dwnlups))

Pearson's Chi-squared test

data: table(Q3$Rankimp, Q3$Dwnlups)
X-squared = 1.3046,
```

```
df = 2,  
p-value = 0.5209
```

4. Rankimp + Dwnldowns

```
> Q4 <- df [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns ")]  
> table (Q4$Rankimp, Q4$Dwnldowns)
```

```
          Dwnldowns  
          0    2  
Rankimp 0  144  0  
Rankimp 1  395  1
```

```
> chisq.test(table (Q4$Rankimp, Q4$Dwnldowns))
```

Pearson's Chi-squared test

```
data: table(Q4$Rankimp, Q4$Dwnldowns)  
X-squared = 1.1823e-27,  
df = 1,  
p-value = 1
```

Tendo em consideração as hipóteses *H0: as variáveis são independentes vs H1: as variáveis são dependentes*, e a partir do output supra apresentado podemos retirar as seguintes conclusões:

- Para o caso 1., o valor o χ^2 é de 192.67 e o *p-value* associado é $< 2.2e-16$, significativo para um intervalo de confiança de 0,05, pelo que podemos rejeitar a hipótese nula e concluir que as variáveis não são independentes ou, por outras palavras, que estão associadas uma à outra (existência de associação entre a melhoria de classificação no ranking da *app* e o número de subidas no ranking que a mesma sofre entre a 1.^a e a 6.^a observação);
- Para o caso 2., o valor do χ^2 é de 49.585 e o *p-value* é 1.685e-09, inferior ao valor crítico definido, o que nos permite, também, rejeitar a hipótese nula e afirmar a existência de associação entre a melhoria de classificação no ranking da *app* e o número de descidas no ranking que a mesma sofre ao longo das 6 observações;
- O caso 3. revela-se contrário aos anteriores, uma vez que o *p-value* associado ao valor do χ^2 é de 0,5209, superior ao intervalo de confiança de 0,05, logo não significativo, pelo que a hipótese nula não é rejeitada, mostrando a independência das variáveis em causa. Ou seja, a melhoria de classificação no

ranking das *apps* não está associada ao aumento do número de *downloads* da mesma;

- Por fim, o caso 4. assemelha-se ao caso anterior, o valor do χ^2 é de 1.1823e-27 e o *p-value* é 1, superior ao valor crítico definido, o que nos permite, também, aceitar a hipótese nula e afirmar a inexistência de associação entre a melhoria de classificação no ranking da *app* e a redução do número de *downloads* que a mesma sofre ao longo das 6 observações;

Segue-se a construção efetiva do segundo modelo de regressão logística, cujo modo de apresentação segue uma lógica sequencial, tendo em conta os passos dados no programa de suporte à análise dos resultados (RStudio). Assim, primeiramente, são apresentadas as variáveis sobre as quais a regressão incide, bem como a fórmula utilizada para o modelo *glm*. Seguidamente, é apresentado o output relativo ao modelo obtido (*summary*), que nos fornece os valores estimados dos coeficientes, valores esses que nos elucidam das variáveis que são estatisticamente significativas para o modelo em questão.

A seleção das variáveis respeita a mesma lógica da Regressão 1, tendo agora em conta o detalhe da variação ocorrida em cada uma delas. Ou seja, no caso da variável *Rankchanges*, temos agora o sinal (positivo ou negativo) da sua alteração ao longo das 6 observações (*Rankups* e *Rankdowns*). Analogamente, o mesmo foi feito para a variável *Dwnlchanges*.

(Os passos em RStudio são iguais em todas as regressões, pelo que a consulta remete sempre para o Apêndice 4)

Depois de definidas todas as variáveis, a fase seguinte englobou a obtenção do modelo de regressão logística múltipla (a partir de agora denominado de *mylogit2*), através da função *glm*:

```
mylogit2 <- glm(formula = Rankimp ~ Rankups + Rankdowns + Dwnlups +  
Dwnldowns, family = "binomial", data = Logit2)
```

O output relativo ao modelo obtido através do comando *summary(mylogit2)* encontra-se exposto na tabela seguinte:

Tabela 8. Output relativo ao modelo de regressão logística múltipla

```

> mylogit2 <- glm(formula = Rankimp ~ Rankups + Rankdowns + Dwnlups + Dwnldowns,
family = "binomial", data = Logit2)
> summary(mylogit2)

Call:
glm(formula = Rankimp ~ Rankups + Rankdowns + Dwnlups + Dwnldowns, family =
"binomial", data = Logit2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8478 -0.3684  0.3028  0.7403  1.6478

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.41914    0.58740   5.821  5.86e-09 ***
Rankups     -1.59620    0.15628  -10.214 < 2e-16 ***
Rankdowns   0.30917    0.17447   1.772  0.0764 .
Dwnlups     0.08508    0.33139   0.257  0.7974
Dwnldowns   6.70590   441.37170  0.015  0.9879
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 626.31 on 539 degrees of freedom
Residual deviance: 418.58 on 535 degrees of freedom
AIC: 428.58
Number of Fisher Scoring iterations: 13

```

Seguidamente, passamos à análise dos valores estimados dos coeficientes presentes no output. Para tal, recorreu-se a um teste de significância estatística (ou teste de hipóteses), o qual corresponde a uma regra decisória que nos permite rejeitar ou não rejeitar uma hipótese estatística com base nos resultados de uma amostra e para um dado nível de significância.

Temos, então, como hipóteses:

H0: as variáveis são independentes vs H1: as variáveis são dependentes,

para um nível de significância de 0,05 (e, conseqüentemente, intervalo de confiança de 95%).

A partir do output da **Tabela 8** pode-se concluir que apenas as variáveis *Rankups* e *Rankdowns* são estatisticamente significativas para o modelo em questão. O *p-value* associado a *Rankups* é inferior a um nível de significância de 0,05, o que permite descartar a hipótese nula e aceitar a influência da variável *Rankups* na melhoria de classificação no ranking das *apps*.

A interpretação dos coeficientes baseia-se no seguinte raciocínio (Institute for Digital Research and Education):

- Coeficiente > 1 , aumenta a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão;
- Coeficiente < 1 , diminui a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão.

Estes coeficientes são ainda alvos de um cálculo ($e^{\text{coeficiente}}$), cujo resultado nos indica a maior ou menor probabilidade de o evento ocorrer, dada a variável independente que está a causar a mudança.

Assim, e *ceteris paribus*:

- Quando a variável *Rankups* regista uma variação positiva de uma unidade, as chances de a *app* melhorar a sua classificação diminuem em $e^{(-1,596)} = 0.20$ (16,7%¹⁵).

¹⁵ Chance = $(p/(1-p)) \rightarrow 0,20 = (p/(1-p)) \rightarrow p = 16,7\%$

Como a probabilidade é inferior a 50%, significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que a variável *Rankups* piore a classificação das apps do que a sua melhoria.

4.3. Modelo de Regressão Logística 3

Primeiramente, recorreu-se ao Teste de Independência Qui-Quadrado (χ^2) para testar a associação da variável dependente *Rankimp_maintain* com cada uma das variáveis (Tabela 9) independentes selecionadas (descritas na Tabela 10 **Error! Reference source not found.****Error! Reference source not found.**).

Tabela 9. Descrição das variáveis para o modelo de regressão logística múltipla (Logit3)

<i>Abreviatura</i>	<i>Variável</i>	<i>Unidades de medida</i>
Rankimp_maintain	<i>Variável dependente relativa ao resultado da diferença entre a 1ª e 6ª observação: melhorou, piorou ou ficou na mesma posição do ranking (situação em que o zero está incluído na melhoria ou estagnação de posição)</i>	<i>0 (piorou) e 1 (melhorou/igual)</i>
Rankchanges	<i>Variável independente discreta relativa ao número de alterações na posição do ranking entre observações</i>	<i>[0; 5]</i>
PVchanges	<i>Variável independente discreta relativa ao número de alterações ocorridas na versão atual</i>	<i>[0; 5]</i>

Dwnlchanges	<i>Variável independente</i>	<i>[0; 2]</i>
	<i>discreta relativa ao número de alterações ocorridas no intervalo de downloads</i>	

Tabela 10. Teste de Independência Qui-Quadrado (3)

1. Rankimp_maintain + Rankchanges

```
> Q1 <- df [c ("Rankimp_maintain", "Rankchanges", "PVchanges", "Dwnlchanges")]
> table (Q1$Rankimp_maintain, Q1$Rankchanges)
```

	Rankchanges					
	0	1	2	3	4	5
Rankimp_maintain 0	0	0	2	5	26	99
Rankimp_maintain 1	1	89	27	48	73	170

```
> chisq.test(table (Q1$Rankimp_maintain, Q1$Rankchanges))
Pearson's Chi-squared test
data: table(Q1$Rankimp_maintain, Q1$Rankchanges)
X-squared = 62.841,
df = 5,
p-value = 3.14e-12
```

2. Rankimp_maintain + PVchanges

```
> Q2 <- df [c ("Rankimp_maintain", "Rankchanges", "PVchanges", "Dwnlchanges")]
> table (Q2$Rankimp_maintain, Q2$PVchanges)
```

	PVchanges					
	0	1	2	3	4	5
Rankimp_maintain 0	61	21	12	16	12	10
Rankimp_maintain 1	240	73	41	31	17	6

```
> chisq.test(table (Q2$Rankimp_maintain, Q2$PVchanges))

Pearson's Chi-squared test

data: table(Q2$Rankimp_maintain, Q2$PVchanges)
X-squared = 22.558,
df = 5,
p-value
= 0.00041
```

3. Rankimp_maintain + Dwnlchanges

```
> Q3 <- df [c ("Rankimp_maintain", "Rankchanges", "PVchanges", "Dwnlchanges")]  
> table (Q3$Rankimp_maintain, Q3$Dwnlchanges)
```

	Dwnlchanges		
	0	1	2
Rankimp_maintain 0	113	17	2
Rankimp_maintain 1	363	42	3

```
> chisq.test(table (Q3$Rankimp_maintain, Q3$Dwnlchanges))
```

Pearson's Chi-squared test

```
data: table(Q3$Rankimp_maintain, Q3$Dwnlchanges)  
X-squared = 1.393,  
df = 2,  
p-value = 0.4983
```

Tendo em consideração as hipóteses *H0: as variáveis são independentes vs H1: as variáveis são dependentes*, e a partir do output supra apresentado podemos retirar as seguintes conclusões:

- Para o caso 1., o valor do χ^2 é de 62.841 e o *p-value* associado é de 3.14e-12, significativo para um intervalo de confiança de 0,05, pelo que podemos rejeitar a hipótese nula e concluir que as variáveis não são independentes ou, por outras palavras, que estão associadas uma à outra (existência de associação entre a melhoria (estagnação) de classificação no ranking da *app* e o número de alterações no ranking que a mesma sofre entre a 1.^a e a 6.^a observação);
- Para o caso 2., o valor do χ^2 é de 22.558 e o *p-value* é 0.00041, inferior ao valor crítico definido, o que nos permite, também, rejeitar a hipótese nula e afirmar a existência de associação entre a melhoria (estagnação) de classificação no ranking da *app* e o número de alterações na versão que a mesma sofre ao longo das 6 observações;
- Por fim, o caso 3. revela-se contrário aos anteriores, uma vez que o *p-value* associado ao valor do χ^2 é de 0,4983, superior ao intervalo de confiança de 0,05, logo não significativo, pelo que a hipótese nula não é rejeitada, mostrando a independência das variáveis em causa. Ou seja, a melhoria (estagnação) de classificação no ranking das *apps* não é influenciada pelo aumento ou diminuição do número de *downloads* da mesma.

Segue-se a construção efetiva do terceiro modelo de regressão logística, cujo modo de apresentação segue uma lógica sequencial, tendo em conta os passos dados no programa de suporte à análise dos resultados (RStudio). Assim, primeiramente, são apresentadas as variáveis sobre as quais a regressão incide, bem como a fórmula utilizada para o modelo *glm*. Seguidamente, é apresentado o output relativo ao modelo obtido (*summary*), que nos fornece os valores estimados dos coeficientes, valores esses que nos elucidam das variáveis que são estatisticamente significativas para o modelo em questão. A definição de tais variáveis pode ser vista de forma análoga à Regressão 1, com a única diferença ao nível da variável dependente. Como já foi referido anteriormente, dado que existem *apps* que mantêm a sua classificação no ranking ao longo do período observável, optou-se por criar a variável *Rankimp_maintain*, com o intuito de detalhar ainda mais o estudo em causa.

Depois de definidas todas as variáveis, a fase seguinte englobou a obtenção do modelo de regressão logística múltipla (a partir de agora denominado de *mylogit3*), através da função *glm*:

```
mylogit3 <- glm(formula = Rankimp_maintain ~ Rankchanges + PVchanges +  
                Dwnlchanges, family = "binomial", data = Logit3)
```

O output relativo ao modelo obtido através do comando *summary* (*mylogit3*) encontra-se exposto na tabela seguinte:

Tabela 11. Output relativo ao modelo de regressão logística múltipla

```
> mylogit3 <- glm(formula = Rankimp_maintain ~ Rankchanges + PVchanges +  
Dwnlchanges, family = "binomial", data = Logit3)  
> summary(mylogit3)  
  
Call:  
glm(formula = Rankimp_maintain ~ Rankchanges + PVchanges + Dwnlchanges, family =  
"binomial", data = Logit3)
```

Deviance Residuals:				
Min	1Q	Median	3Q	Max
-2.5656	0.1819	0.4142	0.8861	1.3442
Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.93358	0.61749	7.990	1.35e-15 ***
Rankchanges	-0.84023	0.13195	-6.368	1.92e-10 ***
PVchanges	-0.22329	0.07521	-2.969	0.00299 **
Dwnlchanges	0.10931	0.29313	0.373	0.70922

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 600.64 on 539 degrees of freedom				
Residual deviance: 512.23 on 536 degrees of freedom				
AIC: 520.23				
Number of Fisher Scoring iterations: 6				

Seguidamente, passamos à análise dos valores estimados dos coeficientes presentes no output. Para tal, recorreu-se a um teste de significância estatística (ou teste de hipóteses), o qual corresponde a uma regra decisória que nos permite rejeitar ou não rejeitar uma hipótese estatística com base nos resultados de uma amostra e para um dado nível de significância.

Temos, então, como hipóteses:

H0: as variáveis são independentes vs H1: as variáveis são dependentes,

para um nível de significância de 0,05 (e, conseqüentemente, intervalo de confiança de 95%).

A partir do output da Tabela 11 pode-se concluir que apenas as variáveis *Rankchanges* e *PVchanges* são estatisticamente significativas para o modelo em questão. O *p-value* associado a estas variáveis é inferior a um nível de significância de 0,05, o que permite descartar a hipótese nula e aceitar a associação entre as variáveis *Rankchanges* e *PVchanges* e a melhoria de classificação no ranking das *apps*.

A interpretação dos coeficientes baseia-se no seguinte raciocínio (Institute for Digital Research and Education):

- Coeficiente > 1 , aumenta a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão;
- Coeficiente < 1 , diminui a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão.

Estes coeficientes são ainda alvos de um cálculo ($e^{\text{coeficiente}}$), cujo resultado nos indica a maior ou menor probabilidade de o evento ocorrer, dada a variável independente que está a causar a mudança.

Assim, e *ceteris paribus*:

- Quando a variável *Rankchanges* regista uma variação positiva de uma unidade, as chances de a *app* melhorar a sua classificação no ranking diminuem em $e^{(-0.840)} = 0,43$ (30,1%¹⁶).
- Quando a variável *PVchanges* regista uma variação positiva de uma unidade, as chances de a *app* melhorar a sua classificação no ranking diminuem em $e^{(-0.223)} = 0,80$ (44,4%¹⁷).

Como ambas as probabilidades são inferiores a 50%, significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que as variáveis *Rankchanges* e *PVchanges* piorem a classificação no ranking das *apps* do que a sua melhoria.

4.4. Modelo de Regressão Logística 4

Primeiramente, recorreu-se ao Teste de Independência Qui-Quadrado (χ^2) para testar a associação da variável dependente *Rankimp_maintain* com cada uma das variáveis (**Error! Reference source not found.**) independentes seleccionadas (Tabela 4):

¹⁶ Chance = $(p/(1-p)) \rightarrow 0,43 = (p/(1-p)) \rightarrow p = 30,1\%$

¹⁷ Chance = $(p/(1-p)) \rightarrow 0,80 = (p/(1-p)) \rightarrow p = 44,4\%$

Tabela 12. Descrição das variáveis para o modelo de regressão logística múltipla (Logit4)

<i>Abreviatura</i>	<i>Variável</i>	<i>Unidades de medida</i>
Rankimp_maintain	<i>Variável dependente relativa ao resultado da diferença entre a 1ª e 6ª observação: melhorou, piorou ou ficou na mesma posição do ranking (situação em que o zero está incluído na melhoria ou estagnação de posição)</i>	<i>0 (piorou) e 1 (melhorou/igual)</i>
Rankups	<i>Variável independente discreta relativa ao número de subidas no ranking entre observações</i>	<i>[0; 5]</i>
Rankdowns	<i>Variável independente discreta relativa ao número de descidas no ranking entre observações</i>	<i>[0; 5]</i>
Dwnlups	<i>Variável independente discreta relativa ao número de subidas ocorridas no intervalo de downloads</i>	<i>[0; 2]</i>
Dwnldowns	<i>Variável independente discreta relativa ao número de descidas ocorridas no intervalo de downloads</i>	<i>0 e 2</i>

Tabela 13. Teste de Independência Qui-Quadrado (4)

1. Rankimp_maintain + Rankups

```
> Q1 <- df [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns)]
> table (Q1$Rankimp_maintain, Q1$Rankups)
```

		Rankups					
		0	1	2	3	4	5
Rankimp_maintain	0	1	8	40	68	14	1
Rankimp_maintain	1	116	125	123	44	0	0

```
> chisq.test(table (Q1$Rankimp_maintain, Q1$Rankups))
```

Pearson's Chi-squared test

```
data: table(Q1$Rankimp_maintain, Q1$Rankups)
X-squared = 185.85,
df = 5,
p-value < 2.2e-16
```

2. Rankimp_maintain + Rankdowns

```
> Q2 <- df [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns
")]
> table (Q2$Rankimp_maintain, Q2$Rankdowns)
```

		Rankdowns					
		0	1	2	3	4	5
Rankimp_maintain	0	4	24	75	27	2	0
Rankimp_maintain	1	2	107	127	129	37	6

```
> chisq.test(table (Q2$Rankimp_maintain, Q2$Rankdowns))
```

Pearson's Chi-squared test

```
data: table(Q2$Rankimp_maintain, Q2$Rankdowns)
X-squared = 40.17,
df = 5,
p-value = 1.38e-07
```

3. Rankimp_maintain + Dwnlups

```
> Q3 <- df [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns
")]
> table (Q3$Rankimp_maintain, Q3$Dwnlups)
```

		Dwnlups		
		0	1	2
Rankimp_maintain	0	113	17	2
Rankimp_maintain	1	364	42	2

```
> chisq.test(table (Q3$Rankimp_maintain, Q3$Dwnlups))
```

Pearson's Chi-squared test

```
data: table(Q3$Rankimp_maintain, Q3$Dwnlups)
X-squared = 2.1714,
df = 2,
p-value = 0.3377
```

4. Rankimp_maintain + Dwnldowns

```
> Q4 <- df [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns")]  
> table (Q4$Rankimp_maintain, Q4$Dwnldowns)
```

	Dwnldowns	
	0	2
Rankimp_maintain 0	132	0
Rankimp_maintain 1	407	1

```
> chisq.test(table (Q4$Rankimp_maintain, Q4$Dwnldowns))
```

Pearson's Chi-squared test

```
data: table(Q4$Rankimp_maintain, Q4$Dwnldowns)  
X-squared = 1.0776e-28,  
df = 1,  
p-value = 1
```

Tendo em consideração as hipóteses *H0: as variáveis são independentes vs H1: as variáveis são dependentes*, e a partir do output supra apresentado podemos retirar as seguintes conclusões:

- Para o caso 1., o valor do χ^2 é de 185.85 e o *p-value* associado é $< 2.2e-16$, significativo para um intervalo de confiança de 0,05, pelo que podemos rejeitar a hipótese nula e concluir que as variáveis não são independentes ou, por outras palavras, que estão associadas uma à outra (existência de associação entre a melhoria (estagnação) de classificação no ranking da *app* e o número de subidas no ranking que a mesma sofre entre a 1.^a e a 6.^a observação);
- Para o caso 2., o valor do χ^2 é de 40.17 e o *p-value* é $1.38e-07$, inferior ao valor crítico definido, o que nos permite, também, rejeitar a hipótese nula e afirmar a existência de associação entre a melhoria (estagnação) de classificação no ranking da *app* e o número de descidas no ranking que a mesma sofre ao longo das 6 observações;
- O caso 3. revela-se contrário aos anteriores, uma vez que o *p-value* associado ao valor do χ^2 é de 0,3377, superior ao intervalo de confiança de 0,05, logo não significativo, pelo que a hipótese nula não é rejeitada, mostrando a independência das variáveis em causa. Ou seja, a melhoria (estagnação) de

classificação no ranking das *apps* não é influenciada pelo aumento do número de *downloads* da mesma;

- Por fim, o caso 4. assemelha-se ao caso anterior, o valor do χ^2 é de 1.0776e-28 e o *p-value* é 1, superior ao valor crítico definido, o que nos permite, também, aceitar a hipótese nula e afirmar a inexistência de associação entre a melhoria (estagnação) de classificação no ranking da *app* e a redução do número de *downloads* que a mesma sofre ao longo das 6 observações.

Segue-se a construção efetiva do quarto modelo de regressão logística, cujo modo de apresentação segue uma lógica sequencial, tendo em conta os passos dados no programa de suporte à análise dos resultados (RStudio). Assim, primeiramente, são apresentadas as variáveis sobre as quais a regressão incide, bem como a fórmula utilizada para o modelo *glm*. Seguidamente, é apresentado o output relativo ao modelo obtido (*summary*), que nos fornece os valores estimados dos coeficientes, valores esses que nos elucidam das variáveis que são estatisticamente significativas para o modelo em questão.

A lógica de seleção destas variáveis foi conduzida pelo descrito na Regressão 2, tendo sempre em atenção a substituição da variável dependente Rankimp por Rankimp_maintain.

Depois de definidas todas as variáveis, a fase seguinte englobou a obtenção do modelo de regressão logística múltipla (a partir de agora denominado de *mylogit4*), através da função *glm*:

```
mylogit4 <- glm(formula = Rankimp_maintain ~ Rankups + Rankdowns + Dwnlups  
                + Dwnldowns, family = "binomial", data = Logit4)
```

O output relativo ao modelo obtido através do comando *summary* (*mylogit4*) encontra-se exposto na tabela seguinte:

Tabela 14. Output relativo ao modelo de regressão logística múltipla

```
> mylogit4 <- glm(formula = Rankimp_maintain ~ Rankups + Rankdowns + Dwnlups +  
Dwnldowns, family = "binomial", data = Logit4)
```

```
> summary(mylogit4)
```

Call:

```
glm(formula = Rankimp_maintain ~ Rankups + Rankdowns + Dwnlups + Dwnldowns,  
family = "binomial", data = Logit4)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9370	0.1192	0.2941	0.6960	1.5054

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.87042	0.65401	5.918	3.26e-09 ***
Rankups	-1.60980	0.16711	-9.633	< 2e-16 ***
Rankdowns	0.21457	0.18387	1.167	0.243
Dwnlups	-0.09109	0.33123	-0.275	0.783
Dwnldowns	6.63577	441.37170	0.015	0.988

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 600.64 on 539 degrees of freedom

Residual deviance: 405.11 on 535 degrees of freedom

AIC: 415.11

Number of Fisher Scoring iterations: 13

Seguidamente, passamos à análise dos valores estimados dos coeficientes presentes no output. Para tal, recorreu-se a um teste de significância estatística (ou teste de hipóteses), o qual corresponde a uma regra decisória que nos permite rejeitar ou não

rejeitar uma hipótese estatística com base nos resultados de uma amostra e para um dado nível de significância.

Temos, então, como hipóteses:

H0: as variáveis são independentes vs H1: as variáveis são dependentes,

para um nível de significância de 0,05 (e, conseqüentemente, intervalo de confiança de 95%).

A partir do output da Tabela 14 pode-se concluir que apenas a variável *Rankups* é estatisticamente significativa para o modelo em questão. O *p-value* associado a esta variável é inferior a um nível de significância de 0,05, o que permite descartar a hipótese nula e aceitar a associação entre a variável *Rankups* e a melhoria (estagnação) de classificação no ranking das *apps*.

A interpretação dos coeficientes baseia-se no seguinte raciocínio (Institute for Digital Research and Education):

- Coeficiente > 1, aumenta a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão;
- Coeficiente < 1, diminui a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão.

Estes coeficientes são ainda alvos de um cálculo ($e^{\text{coeficiente}}$), cujo resultado nos indica a maior ou menor probabilidade de o evento ocorrer, dada a variável independente que está a causar a mudança.

Assim, e *ceteris paribus*:

- Quando a variável *Rankups* regista uma variação positiva de uma unidade, as chances de a *app* melhorar a sua classificação no ranking diminuem em $e^{(-1.6098)} = 0.20$ (16,7%¹⁸).

Como a probabilidade associada à variável *Rankups* é inferior a 50%, as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso

¹⁸ Chance = $(p/(1-p)) \rightarrow 0,20 = (p/(1-p)) \rightarrow p = 16,7\%$

de não mudar nada. Ou seja, é mais provável que a variável *Rankups* piore a classificação no ranking das *apps* do que a sua melhoria (estagnação).

A partir dos quatro modelos de regressão múltipla construídos, podemos concluir que as variáveis dependentes *Rankimp* e *Rankimp-maintain* apresentam resultados quase idênticos para as variáveis independentes utilizadas, pelo que podemos afirmar que a melhoria (estagnação) na classificação do ranking de uma *app* é influenciada pelas variáveis *PVchanges*, *Rankchanges* e *Rankups*.

Contudo, e para uma análise mais detalhada do estudo, procedeu-se ainda à criação de modelos de regressão logística múltipla utilizando apenas as *apps* da amostra que constassem em todas as 6 observações, ou seja, as *apps* que nunca desaparecem do ranking desde o início até ao final das observações que alimentaram a base de dados. Esta subamostra engloba um total de 249 *apps*, das quais 97 relativas a Cuidados Médicos (34 pagas e 63 gratuitas) e 152 relativas a Saúde & Fitness (57 pagas e 95 gratuitas).

No entanto, como neste caso muitas *apps* mantêm a sua posição ao longo do ranking, optamos por utilizar apenas a variável dependente *Rankimp_maintain* para corroborar o anteriormente descrito.

```
subset_6obs <- subset(df, Tot_Obs > 5)
```

Depois de criado o *subset* com as 6 observações, criamos os quatro modelos de regressão anteriores, de modo a poder comparar resultados.

4.5. Modelo de Regressão Logística 5

Primeiramente, recorreu-se ao Teste de Independência Qui-Quadrado (χ^2) para testar a associação da variável dependente *Rankimp* com cada uma das variáveis independentes selecionadas (Tabela 15):

Tabela 15. Teste de Independência Qui-Quadrado (5)

1. Rankimp_maintain + Rankchanges

```
> Q1 <- subset_6obs [c ("Rankimp_maintain", "Rankchanges", "PVchanges",
"Dwnlchanges")]
> table (Q1$Rankimp_maintain, Q1$Rankchanges)
```

	Rankchanges				
	0	2	3	4	5
Rankimp_maintain 0	0	0	3	16	79
Rankimp_maintain 1	1	1	5	26	118

```
> chisq.test(table (Q1$Rankimp, Q1$Rankchanges))
```

Pearson's Chi-squared test

data: table(Q1\$Rankimp, Q1\$Rankchanges)

X-squared = 1.3833,

df = 4,

p-value = 0.8471

2. Rankimp_maintain + PVchanges

```
> Q2 <- subset_6obs [c ("Rankimp_maintain", "Rankchanges", "PVchanges",
"Dwnlchanges")]
> table (Q2$Rankimp_maintain, Q2$PVchanges)
```

	PVchanges					
	0	1	2	3	4	5
Rankimp_maintain 0	39	17	11	11	12	8
Rankimp_maintain 1	91	22	12	12	9	5

```
> chisq.test(table (L$Rankimp, L$PVchanges))
```

Pearson's Chi-squared test

data: table(Q2\$Rankimp, Q2\$PVchanges)

X-squared = 11.907,

df = 5,

p-value = 0.0368

3. Rankimp_maintain + Dwnlchanges

```
> Q3 <- subset_6obs [c ("Rankimp_maintain", "Rankchanges", "PVchanges",
"Dwnlchanges")]
> table (Q3$Rankimp_maintain, Q3$Dwnlchanges)
```

	Dwnlchanges		
	0	1	2
Rankimp_maintain 0	81	15	2
Rankimp_maintain 1	135	14	2

```
> chisq.test(table (Q3$Rankimp, Q3$Dwnlchanges))
```

Pearson's Chi-squared test

```
data: table(Q3$Rankimp, Q3$Dwnlchanges)
```

```
X-squared = 2.3603,
```

```
df = 2,
```

```
p-value = 0.3072
```

Tendo em consideração as hipóteses *H0: as variáveis são independentes vs H1: as variáveis são dependentes*, e a partir do output supra apresentado podemos retirar as seguintes conclusões:

- Para o caso 1., o valor do χ^2 é de 1.3833 e o *p-value* associado é de 0.8471, superior ao intervalo de confiança de 0,05, logo não significativo, pelo que a hipótese nula não é rejeitada, mostrando a independência das variáveis em causa. Ou seja, a melhoria de classificação no ranking das *apps* não é influenciada pelo número de alterações no ranking que a mesma sofre ao longo das 6 observações;
- Para o caso 2., o valor do χ^2 é de 11.907 e o *p-value* é 0.0368, inferior ao valor crítico definido, logo significativo, o que nos permite rejeitar a hipótese nula e afirmar a existência de associação entre a melhoria de classificação no ranking da *app* e o número de alterações na versão que a mesma sofre ao longo das 6 observações;
- Por fim, o caso 3. revela-se contrário aos anteriores, uma vez que o *p-value* associado ao valor do χ^2 é de 0,3072, superior ao intervalo de confiança de 0,05, logo não significativo, pelo que a hipótese nula não é rejeitada, mostrando a independência das variáveis em causa. Ou seja, a melhoria de classificação no ranking das *apps* não está relacionada com o aumento ou diminuição do número de *downloads* da mesma.

Segue-se a construção efetiva do quinto modelo de regressão logística, cujo modo de apresentação segue uma lógica sequencial, tendo em conta os passos dados no programa de suporte à análise dos resultados (RStudio). Assim, primeiramente, são apresentadas

as variáveis sobre as quais a regressão incide, bem como a fórmula utilizada para o modelo *glm*. Seguidamente, é apresentado o output relativo ao modelo obtido (*summary*), que nos fornece os valores estimados dos coeficientes, valores esses que nos elucidam das variáveis que são estatisticamente significativas para o modelo em questão.

Assim, o quinto modelo (a partir de agora denominado de *mylogit5*) foi obtido através da função *glm*:

```
mylogit5 <- glm(formula = Rankimp_maintain ~ Rankchanges + PVchanges +
                Dwnlchanges, family = "binomial", data = Logit5)
```

O output relativo ao modelo obtido através do comando *summary* (*mylogit5*) encontra-se exposto na tabela seguinte:

Tabela 16. Output relativo ao modelo de regressão logística múltipla (*Logit5*)

```
> mylogit5 <- glm(formula = Rankimp_maintain ~ Rankchanges + PVchanges +
Dwnlchanges, family = "binomial", data = Logit5)
> summary(mylogit6)
```

Call:
glm(formula = Rankimp_maintain ~ Rankchanges + PVchanges + Dwnlchanges, family = "binomial", data = Logit5)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7301	-1.2555	0.8895	0.8895	1.4976

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.02351	1.12177	1.804	0.07125 .
Rankchanges	-0.26012	0.23087	-1.127	0.25989
PVchanges	-0.27064	0.08874	-3.050	0.00229 **
Dwnlchanges	-0.09695	0.34570	-0.280	0.77913

```
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 333.82 on 248 degrees of freedom
Residual deviance: 321.46 on 245 degrees of freedom
AIC: 329.46
Number of Fisher Scoring iterations: 4
```

Seguidamente, passamos à análise dos valores estimados dos coeficientes presentes no output. Para tal, recorreu-se a um teste de significância estatística (ou teste de hipóteses), o qual corresponde a uma regra decisória que nos permite rejeitar ou não rejeitar uma hipótese estatística com base nos resultados de uma amostra e para um dado nível de significância.

Temos, então, como hipóteses:

H0: as variáveis são independentes vs H1: as variáveis são dependentes,

para um nível de significância de 0,05 (e, conseqüentemente, intervalo de confiança de 95%).

A partir do output da Tabela 16. pode-se concluir que apenas a variável *PVchanges* é estatisticamente significativa para o modelo em questão. O *p-value* associado a esta variável é inferior a um nível de significância de 0,05, o que permite descartar a hipótese nula e aceitar a associação entre a variável *PVchanges* e a melhoria (estagnação) de classificação no ranking das *apps*.

A interpretação dos coeficientes baseia-se no seguinte raciocínio (Institute for Digital Research and Education):

- Coeficiente > 1, aumenta a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão;
- Coeficiente < 1, diminui a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão.

Estes coeficientes são ainda alvos de um cálculo ($e^{\text{coeficiente}}$), cujo resultado nos indica a maior ou menor probabilidade de o evento ocorrer, dada a variável independente que está a causar a mudança.

Assim, e *ceteris paribus*:

- Quando a variável *PVchanges* regista uma variação positiva de uma unidade, as chances de a *app* melhorar a sua classificação no ranking diminuem em $e^{(-0.27064)} = 0.76$ (43,2%¹⁹).

A probabilidade associada à variável *PVchanges* é inferior a 50%, o que significa que que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada.

4.6. Modelo de Regressão Logística 6

Primeiramente, recorreu-se ao Teste de Independência Qui-Quadrado (χ^2) para testar a associação da variável dependente *Rankimp* com cada uma das variáveis independentes seleccionadas (Tabela 17):

Tabela 17. Teste de Independência Qui-Quadrado (6)

```

1. Rankimp_maintain + Rankups

> Q1 <- subset_6obs [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",
"Dwnldowns")]
> table (Q1$Rankimp_maintain, Q1$Rankups)

              Rankups
Rankimp_maintain 0  1  3  27  53  13  1
Rankimp_maintain 1  6  41  72  32  0  0

> chisq.test(table (Q1$Rankimp, Q1$Rankups))
Pearson's Chi-squared test
data: table(Q1$Rankimp, Q1$Rankups)
X-squared = 67.824,
df = 5,

```

¹⁹ Chance = $(p/(1-p)) \rightarrow 0,76 = (p/(1-p)) \rightarrow p = 43,2\%$

p-value = 2.905e-13

2. Rankimp_maintain + Rankdowns

```
> Q2 <- subset_6obs [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
> table (Q2$Rankimp_maintain, Q2$Rankdowns)
```

	Rankdowns					
	0	1	2	3	4	5
Rankimp_maintain 0	4	19	53	20	2	0
Rankimp_maintain 1	1	2	49	70	25	4

```
> chisq.test(table (Q2$Rankimp, Q2$Rankdowns))
```

Pearson's Chi-squared test

data: table(Q2\$Rankimp, Q2\$Rankdowns)

X-squared = 58.456,

df = 5,

p-value = 2.532e-11

3. Rankimp_maintain + Dwnlups

```
> Q3 <- subset_6obs [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
> table (Q3$Rankimp_maintain, Q3$Dwnlups)
```

	Dwnlups		
	0	1	2
Rankimp_maintain 0	81	15	2
Rankimp_maintain 1	135	14	2

```
> chisq.test(table (Q3$Rankimp, Q3$Dwnlups))
```

Pearson's Chi-squared test

data: table(Q3\$Rankimp, Q3\$Dwnlups)

X-squared = 2.3603,

df = 2,

p-value = 0.3072

4. Rankimp_maintain + Dwnldowns

```
> Q4 <- subset_6obs [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
> table (Q4$Rankimp_maintain, Q4$Dwnldowns)
```

	Dwnldowns	
	0	1
Rankimp_maintain 0	110	

```
Rankimp_maintain 1 139
> chisq.test(table(Q4$Rankimp, Q4$Dwnldowns))
Pearson's Chi-squared test
data: table(Q4$Rankimp, Q4$Dwnldowns)
X-squared = 3.3775,
df = 1,
p-value = 0.06609
```

Tendo em consideração as hipóteses *H0: as variáveis são independentes vs H1: as variáveis são dependentes*, e a partir do output supra apresentado podemos retirar as seguintes conclusões:

- Para o caso 1., o valor do χ^2 é de 67.824 e o *p-value* associado é de 2.905e-13, inferior ao intervalo de confiança de 0,05, pelo que a hipótese nula é rejeitada, mostrando a dependência das variáveis em causa. Ou seja, a melhoria de classificação no ranking das *apps* é influenciada pelo número de subidas de posição no ranking que a mesma sofre ao longo das 6 observações;
- Para o caso 2., o valor do χ^2 é de 58.456 e o *p-value* é 2.532e-11, inferior ao valor crítico definido, logo significativo, o que nos permite rejeitar a hipótese nula e afirmar a existência de associação entre a melhoria de classificação no ranking da *app* e o número de descidas de posição no ranking que a mesma sofre ao longo das 6 observações;
- Os casos 3. e 4. revelam-se contrários aos anteriores, uma vez que os *p-value* associados ao valor do χ^2 são de 0,3072 e 0,06609, respetivamente, superiores ao intervalo de confiança de 0,05, logo não significativos, pelo que as hipóteses nulas não são rejeitadas, mostrando a independência das variáveis em causa. Ou seja, a melhoria de classificação no ranking das *apps* não é influenciada pelo aumento ou diminuição do número de *downloads* da mesma.

Segue-se a construção efetiva do sexto modelo de regressão logística, cujo modo de apresentação segue uma lógica sequencial, tendo em conta os passos dados no programa de suporte à análise dos resultados (RStudio). Assim, primeiramente, são apresentadas as variáveis sobre as quais a regressão incide, bem como a fórmula utilizada para o

modelo *glm*. Seguidamente, é apresentado o output relativo ao modelo obtido (*summary*), que nos fornece os valores estimados dos coeficientes, valores esses que nos elucidam das variáveis que são estatisticamente significativas para o modelo em questão.

Assim, o sexto modelo (a partir de agora denominado de *mylogit6*) foi obtido através da função *glm*:

```
mylogit6 <- glm(formula = Rankimp_maintain ~ Rankupss + Rankdowns + Dwnlups
                + Dwnldowns, family = "binomial", data = Logit6)
```

O output relativo ao modelo obtido através do comando *summary* (*mylogit6*) encontra-se exposto na tabela seguinte:

Tabela 18. Output relativo ao modelo de regressão logística múltipla (*Logit6*)

```
> mylogit7 <- glm(formula = Rankimp_maintain ~ Rankups + Rankdowns + Dwnlups +
Dwnldowns, family = "binomial", data = Logit6)
> summary(mylogit6)
```

Call:

```
glm(formula = Rankimp_maintain ~ Rankchanges + Rankups + Rankdowns + Dwnlups +
Dwnldowns, family = "binomial", data = Logit6)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.4833	-0.9827	0.3062	0.7005	1.7296

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.0877	1.1625	0.936	0.349462
Rankups	-1.0156	0.2726	-3.725	0.000195 ***
Rankdowns	0.7411	0.2760	2.685	0.007257 **

Dwnlups	-0.3825	0.3895	-0.982	0.326132
Dwnldowns	NA	NA	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 333.82 on 248 degrees of freedom				
Residual deviance: 254.62 on 245 degrees of freedom				
AIC: 262.62				
Number of Fisher Scoring iterations: 5				

Seguidamente, passamos à análise dos valores estimados dos coeficientes presentes no output. Para tal, recorreu-se a um teste de significância estatística (ou teste de hipóteses), o qual corresponde a uma regra decisória que nos permite rejeitar ou não rejeitar uma hipótese estatística com base nos resultados de uma amostra e para um dado nível de significância.

Temos, então, como hipóteses:

H0: as variáveis são independentes vs H1: as variáveis são dependentes,

para um nível de significância de 0,05 (e, conseqüentemente, intervalo de confiança de 95%).

A partir do output da Tabela 18. pode-se concluir que apenas as variáveis *Rankups* e *Rankdowns* são estatisticamente significativas para o modelo em questão. O *p-value* associado a estas variáveis é inferior a um nível de significância de 0,05, o que permite descartar a hipótese nula e aceitar a associação entre as variáveis *Rankups* e *Rankdowns* e a melhoria (estagnação) de classificação no ranking das *apps*.

A interpretação dos coeficientes baseia-se no seguinte raciocínio (Institute for Digital Research and Education):

- Coeficiente > 1, aumenta a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão;
- Coeficiente < 1, diminui a probabilidade de o evento ocorrer quando há uma mudança positiva na variável independente em questão.

Estes coeficientes são ainda alvos de um cálculo ($e^{\text{coeficiente}}$), cujo resultado nos indica a maior ou menor probabilidade de o evento ocorrer, dada a variável independente que está a causar a mudança.

Assim, e *ceteris paribus*:

- Quando a variável *Rankups* regista uma variação positiva de uma unidade, as chances de a *app* melhorar (igualar) a sua classificação no ranking diminuem em $e^{(-1.60980)} = 0.20$ (16,7%²⁰).
- Quando a variável *Rankdowns* regista uma variação positiva de uma unidade, as chances de a *app* melhorar (igualar) a sua classificação no ranking diminuem em $e^{(0.7411)} = 2.10$ (67,7%²¹).

Neste caso, a probabilidade associada à variável *Rankups* é inferior a 50%, o que significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que a variável *Rankups* piore a classificação no ranking das *apps* do que a sua melhoria (estagnação).

Por outro lado, a probabilidade associada à variável *Rankdowns* é superior a 50%, o que significa que quanto maior o número de descidas da classificação no ranking das *apps*, maior a probabilidade de ocorre uma melhoria (estagnação) da sua classificação no ranking.

²⁰ Chance = $(p/(1-p)) \rightarrow 0,20 = (p/(1-p)) \rightarrow p = 16,7\%$

²¹ Chance = $(p/(1-p)) \rightarrow 2,1 = (p/(1-p)) \rightarrow p = 66,7\%$

5. Discussão

Tabela 19. P-value e respetivo código de significância (*) para cada um dos Modelos de Regressão Logística (MRL)

<i>MRL</i>	<i>Rankchanges</i>	<i>PVchanges</i>	<i>Dwnlchanges</i>	
1	3.74e-11 ***	0.00838 **	0.49300	
3	1.92e-10 ***	0.00299 **	0.70922	
5	0.25989	0.00229 **	0.77913	

<i>MRL</i>	<i>Rankups</i>	<i>Rankdowns</i>	<i>Dwnups</i>	<i>Dwnldowns</i>
2	< 2e-16 ***	0.0764	0.7974	0.9879
4	< 2e-16 ***	0.243	0.783	0.988
6	0.000195 ***	0.007257 **	0.326132	NA

Tabela 20. Teste de Independência resultante de cada um dos Modelos de Regressão Logística (MRL)

<i>MRL</i>	<i>Rankchanges</i>	<i>PVchanges</i>	<i>Dwnlchanges</i>	
1	Associadas	Associadas	Não associadas	
3	Associadas	Associadas	Não associadas	
5	Não associadas	Associadas	Não associadas	

<i>MRL</i>	<i>Rankups</i>	<i>Rankdowns</i>	<i>Dwnups</i>	<i>Dwnldowns</i>
2	Associadas	Associadas	Não associadas	Não associadas
4	Associadas	Associadas	Não associadas	Não associadas
6	Associadas	Associadas	Não associadas	Não associadas

Tabela 21. Resultados de um dos Modelos de Regressão Logística

<i>MRL</i>	<i>Resultado</i>
------------	------------------

1 As probabilidades associadas às variáveis *Rankchanges* e *PVchanges* são inferiores a 50%, o que significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que as variáveis *Rankchanges* e *PVchanges* piorem a classificação no ranking das *apps* do que a sua melhoria.

2 A probabilidade associada à variável *Rankups* é inferior a 50%, o que significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que a variável *Rankups* piore a classificação das *apps* do que a sua melhoria.

3 As probabilidades associadas às variáveis *Rankchanges* e *PVchanges* são inferiores a 50%, o que significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que as variáveis *Rankchanges* e *PVchanges* piorem a classificação no ranking das *apps* do que a sua melhoria.

4 A probabilidade associada à variável *Rankups* é inferior a 50%, o que significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que a variável *Rankups* piore a classificação no ranking das *apps* do que a sua melhoria (estagnação).

5 A probabilidade associada à variável *PVchanges* é inferior a 50%, o que significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada.

A probabilidade associada à variável *Rankups* é inferior a 50%, o que significa que as chances em causa são inferiores quando se compara o caso de aumentar uma unidade com o caso de não mudar nada. Ou seja, é mais provável que a variável *Rankups* piore a classificação no ranking das *apps* do que a sua melhoria (estagnação).

- 6 Por outro lado, a probabilidade associada à variável *Rankdowns* é superior a 50%, o que significa que quanto maior o número de descidas da classificação no ranking das *apps*, maior a probabilidade de ocorre uma melhoria (estagnação) da sua classificação no ranking.

6. Conclusão

Esta investigação dedicou-se à análise e à compreensão do fenómeno de Mobile Health (m-Health), o qual consiste na provisão por via eletrónica de informação e de serviços de saúde através de dispositivos de comunicação móveis, dada a sua capacidade de criar, armazenar, recuperar e transmitir dados em tempo real entre utilizadores finais, com o objetivo de melhorar quer a segurança dos mesmos, quer a qualidade dos serviços de saúde (Akter et al., 2013).

O *m-health* envolve a utilização de tecnologias *wireless* que permitem a transmissão de vários conteúdos de dados e serviços, os quais são de fácil acesso para trabalhadores da área da saúde, através de dispositivos móveis, como *smartphones*, PDA's, computadores portáteis e tablets. Assim, é considerado um serviço de saúde personalizado e interativo com o objetivo de providenciar um acesso ubíquo e universal à informação e aconselhamento médico por parte de qualquer utilizador (Akter et al., 2013).

No sentido de compreender o processo de adoção e difusão das aplicações móveis de saúde foi realizado um estudo quantitativo, o qual teve como suporte a construção de uma base de dados com toda a informação disponível na fonte selecionada (loja online da plataforma *Android* – *Google Play/Play Store*). O ponto de partida deste estudo foi a classificação no ranking atribuída a cada uma das *apps* constantes na amostra selecionada (ranking esse que foi atualizado mensalmente durante 5 meses), com o intuito de identificar os fatores que influenciam o comportamento das *apps* ao nível da sua classificação.

Essa análise foi complementada com estatística descritiva e modelos de regressão logística, o que, em conjunto, nos permitiu concluir que são as *apps* que mais descidas registam no ranking que representam o conjunto mais significativo da amostra (73%) e que, ao contrário do que seria expectável, a variável *Rankdowns* (número de descidas no ranking de cada *app*) é o único fator que aumenta a probabilidade de uma *app* melhorar (ou manter) a sua posição no ranking.

Ora, a perspetiva inicial apontava para que fossem as variáveis *PVchanges* e *Dwnlchanges* a impactarem o ranking das *apps*, mas a verdade é que no que respeita ao número de *downloads* a que uma *app* é sujeita, a *Google Play/Play Store* disponibiliza apenas classes de *downloads*, as quais compreendem intervalos demasiado alargados. Deste modo, não nos é possível ter noção do número exato de *downloads*, pelo que aponto esta realidade como uma das principais limitações do estudo.

Ainda no âmbito das limitações, a falta de um repositório geral de aplicativos móveis (de saúde e não só), implicou um esforço significativo na construção de base de dados, cuja informação foi introduzida manualmente.

O período temporal de observação também pode ser visto como uma limitação deste estudo, pois 5 meses não nos permite obter uma imagem totalmente fidedigna da evolução de cada variável selecionada.

Por fim, relativamente a perspetivas de investigações e trabalhos futuros, sugiro, primeiramente, a definição de um espetro temporal maior de observação das variáveis (mínimo de 9 meses), bem como fazer um levantamento dos principais problemas ao nível de regulação, já que a informação disponível acerca desta temática não é suficiente para deixar os utilizadores totalmente confortáveis com a utilização de certos aplicativos móveis.

Referências Bibliográficas

- Akter, S., Ray, P., & D'Ambra, J. (2013). Continuance of *mHealth* services at the bottom of the pyramid: the roles of service quality and trust. *Electronic Markets*, 23(1), 29-47. doi: 10.1007/s12525-012-0091-5
- Boulos, M. N. K., Wheeler, S., Tavares, C., & Jones, R. (2011). How *smartphones* are changing the face of *mobile* and participatory *healthcare*: an overview, with example from eCAALYX. *BioMedical Engineering Online*, 10(1), 24. doi: 10.1186/1475-925x-10-24
- Burdette, S. D., Herchline, T. E., & Oehler, R. (2008). Practicing Medicine in a Technological Age: Using *Smartphones* in Clinical Practice. *Clinical Infectious Diseases*, 47(1), 117-122. doi: 10.1086/588788
- Cohn, A. M., Hunter-Reel, D., Hagman, B. T., & Mitchell, J. (2011). Promoting behavior change from alcohol use through *mobile* technology: the future of ecological momentary assessment. *Alcohol. Clin. Exp. Res.*, 2209–2215.
- Dennison, L., Morrison, L., Conway, G., & Yardley, L. (2013). Opportunities and challenges for *smartphone* applications in supporting *health* behavior change: qualitative study. *Journal of Medical Internet Research*, 112–133.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, ACM*, 9-15.
- Dogson, M., Gann, D., & Salter, A. (2008). Networks and Communities. *Oxford University Press Inc*, 86-93.
- Dosi, G. (1982). Technological paradigms and technological trajectories. *Research Policy*, 11(3), 147-162. doi: [http://dx.doi.org/10.1016/0048-7333\(82\)90016-6](http://dx.doi.org/10.1016/0048-7333(82)90016-6)
- Easterby-Smith, M., Thorpe, R., & Jackson, P. R. (2012). Management research. *Sage*.
- Elias, B. L., Fogger, S. A., McGuinness, T. M., & D'Alessandro, K. R. (2014). *Mobile apps* for psychiatric nurses. 42-47.
- Fagerberg, J. (2003). Innovation: A Guide to the Literature.

- Goldbach, H., AChang, Y., Kyer, A., Ketshogileng, D., Taylor, L., & Chandra, A. (2014). Evaluation of generic medical information accessed via *mobile* phones at the point of care in resource-limited settings. *J. Am. Med. Inform. Assoc.* , 37-42.
- Helf, C., & Hlavacs, H. (2016). *Apps* for life change: Critical review and solution directions. *Entertainment Computing*, 14, 17-22. doi: <http://dx.doi.org/10.1016/j.entcom.2015.07.001>
- Hildebrand, D., Lyman Ott, R., & Brian Gray, J. (2004). *Basic Statistical Ideas for Managers* (2 ed.): Thomson South-Western.
- Ho, J. C., & Lee, C. S. (2015). A typology of technological change: Technological paradigm theory with validation and generalization from case studies. *Technological Forecasting and Social Change*, 97, 128-139. doi: 10.1016/j.techfore.2014.05.015
- Hovenga, E., Kidd, M. R., Garde, S., & Hullin, L. C. C. (2009). *Health informatics – an introduction. Stud. Health Technol. Inform.*, 9-15.
- Hussain, M., Al-Haiqi, A., Zaidan, A. A., Zaidan, B. B., Kiah, M. L. M., Anuar, N. B., & Abdulnabi, M. (2015). The landscape of research on *smartphone* medical *apps*: Coherent taxonomy, motivations, open challenges and recommendations. *Computer Methods and Programs in Biomedicine*, 122(3), 393-408. doi: <http://dx.doi.org/10.1016/j.cmpb.2015.08.015>
- Lippman, H. (2013). How *apps* are changing family medicine. *J. Fam. Pract.*, 362-367.
- Oh, H., Rizo, C., Enkin, M., & Jadad, A. (2005). What is *eHealth* (3): a systematic review of published definitions. *J. Med. Internet Res.* , 23-37.
- Oliveira, E. R., & Ferreira, P. (2014). Métodos de Investigação Da Interrogação à Descoberta Científica. *Grupo Editorial Vida Económica*.
- Rogers, E. M. (2003). Diffusion of Innovations. *Free Press*.
- Schumpeter, J. A. (1962). Capitalism, Socialism and Democracy
Chapter VII: The Process of Creative Destruction. *Harper Torchbooks*, 80-86.
- Silow-Carroll, S., & Smith, B. (2013). *Clinical Management Apps: Creating Partnerships Between Providers and Patients*. 1-10.
- Silva, B. M. C., Rodrigues, J. J. P. C., de la Torre Díez, I., López-Coronado, M., & Saleem, K. (2015). *Mobile-health: A review of current state in 2015. Journal of Biomedical Informatics*, 56, 265-272. doi: <http://dx.doi.org/10.1016/j.jbi.2015.06.003>
- Szumilas, M. (2010). Explaining Odds Ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3), 227-229.

World Economic Forum (2017), “*Digital Transformation Initiative: Unlocking \$100 Trillion for Business and Society from Digital Transformation*”, pp. 50-51

PricewaterhouseCoopers (2015), “*Desafios da saúde: responder hoje aos desafios de amanhã*”, pp. 3.

Portal da Gestão, <https://www.portal-gestao.com/gurus-da-gest%C3%A3o/7610-clayton-christensen,-autor-da-inova%C3%A7%C3%A3o-disruptiva.html>, acessado a 25 de janeiro de 2017.

Público, <https://www.publico.pt/2013/04/06/tecnologia/noticia/destruicao-criativa-1590360>, acessado a 25 de janeiro de 2017.

Apêndice 1

Tabela 22. Conjunto das derivadas das observações recolhidas

<i>Variável</i>	<i>Explicação da variável</i>	<i>Observações</i>
Tot_Obs	Número de observações de cada <i>app</i>	Discreta
Rankmean	Média das posições no ranking de cada <i>app</i> , ao longo das 6 observações	[1; 862,5]
Rankmed	Mediana das posições no ranking por <i>app</i>	[1; 999]
Rankmax	Posição máxima no ranking de cada <i>app</i>	[1; 999]
Rankmin	Posição mínima no ranking de cada <i>app</i>	[1; 180]
Ranksd	Desvio-padrão das posições no ranking por <i>app</i>	[0; 537,4]
Rankpct5	Percentil 5 das posições no ranking de cada <i>app</i>	[1; 384,8]
Rankpct95	Percentil 95 das posições no ranking de cada <i>app</i>	[1; 999]
Ranksum	Somatório das posições no ranking de cada <i>app</i>	[6; 5175]
Rankchanges	Número de alterações na posição do ranking entre observações	[0; 5]
Rankdowns	Número de descidas no ranking	[0; 5]
Rankups	Número de subidas no ranking	[0; 5]
Rankdif16	Diferença de posições entre a 1ª e 6ª observação	[-136; 998]
Rankimp	Resultado da diferença entre a 1ª e 6ª observação: melhorou, piorou ou ficou na mesma posição do ranking (situação em que o zero está excluído)	Melhou=1 Pior/Igual=0
Rankimp/maintain	Resultado da diferença entre a 1ª e 6ª observação: melhorou, piorou ou ficou na mesma posição do ranking (situação em que o zero está incluído)	Melhorou/Igual=1 Pior=0

Pmean	Preço médio	[0; 35,99]
Pmed	Preço mediano	[0; 35,99]
Pmax	Preço máximo	35,99
Pmin	Preço mínimo	0
Psd	Desvio-padrão do preço	[0; 3,2]
Ppct5	Percentil 5 dos preços entre a 1 ^a e a 6 ^a observação	[0; 35,99]
Ppct95	Percentil 95 dos preços entre a 1 ^a e a 6 ^a observação	[0; 35,99]
Pchanges	Número de alterações nos preços	[0; 3]
Pdowns	Número de descidas nos preços	[0; 2]
Pups	Número de subidas nos preços	[0; 2]
Pdif16	Diferença de preços entre a 1 ^a e a 6 ^a observação	[-2,08; 7,86]
Pdin	Resultado da diferença de preços entre a 1 ^a e a 6 ^a observação: melhorou (1), piorou (-1) ou não sofreu alterações (0)	Melhorou=1 Piorou=-1 Igual=0
PImean	Média do intervalo dos valores dos produtos integrados em aplicações	[0 – 9,09]
PImax	Máximo do intervalo dos valores dos produtos integrados em aplicações	[0 – 99,99]
PImin	Mínimo do intervalo dos valores dos produtos integrados em aplicações	[0 – 9,24]
PVchanges	Número de alterações ocorridas na versão atual	[0; 5]
Datechanges	Número de alterações ocorridas na data de atualização	[0; 5]
Dwnlchanges	Número de alterações ocorridas na classe de <i>downloads</i>	[0; 2]
Dwnldowns	Número de descidas no número de <i>downloads</i>	0 e 2
Dwnlups	Número de subidas no número de <i>downloads</i>	[0; 2]
Sumclasses	Somatório das classes de <i>downloads</i> entre a 1 ^a e a 6 ^a observação	[D12; D96]
Dwnldif16	Diferença entre as classes de <i>downloads</i> entre a 1 ^a e a 6 ^a observação	[-2; 11]

Apêndice 2

Tabela 23. Conjunto das variáveis presentes na Base de Dados

<i>Variável</i>	<i>Explicação da variável</i>	<i>Observações</i>
Name	Nome da <i>app</i>	
Supplier	Nome do fornecedor	
Category	Cada segmento foi dividido em 4 grandes categorias	- Apoio/Suporte - Monitorização de saúde - Saúde feminina - Outros
Subcategory	Cada uma das 4 categorias foi subdividida em várias subcategorias	Apoio/Suporte - Didático - Recursos médicos - Organização -Outros Monitorização de saúde - Medicação - Treino físico - Alimentação Saúde feminina - Gravidez - Outros Outros - Outros
SF/CM	A amostra escolhida engloba 2 segmentos: Saúde & Fitness (SF) e Cuidados Médicos (CM)	
Free/Paid	De um total de 540 <i>apps</i> (270 para cada categoria), definiu-se uma proporção de 2/3 para as <i>apps</i> pagas e de 1/3 para as <i>apps</i> gratuitas. Ou seja, 180 são <i>apps</i> pagas (90 CM + 90 SF) e 360 são gratuitas (180 CM + 180 SF)	
O1	Observação 1 (data da recolha dos dados)	14,17, 18 dezembro 2016
O2	Observação 2 (data da recolha dos dados)	14, 16, 18 janeiro 2017
O3	Observação 3 (data da recolha dos dados)	14, 17, 18 fevereiro 2017
O4	Observação 4 (data da recolha dos dados)	14, 17, 18 março 2017
O5	Observação 5 (data da recolha dos dados)	14, 17, 18 abril 2017
O6	Observação 6 (data da recolha dos dados)	14, 17, 18 maio 2017

Tot_Obs	Número de observações de cada <i>app</i>	Discreta
Rankmean	Média das posições no ranking de cada <i>app</i> , ao longo das 6 observações	[1; 862,5]
Rankmed	Mediana das posições no ranking por <i>app</i>	[1; 999]
Rankmax	Posição máxima no ranking de cada <i>app</i>	[1; 999]
Rankmin	Posição mínima no ranking de cada <i>app</i>	[1; 180]
Ranksd	Desvio-padrão das posições no ranking por <i>app</i>	[0; 537,4]
Rankpct5	Percentil 5 das posições no ranking de cada <i>app</i>	[1; 384,8]
Rankpct95	Percentil 95 das posições no ranking de cada <i>app</i>	[1; 999]
Ranksum	Somatório das posições no ranking de cada <i>app</i>	[6; 5175]
Rankchanges	Número de alterações na posição do ranking entre observações	[0; 5]
Rankdowns	Número de descidas no ranking	[0; 5]
Rankups	Número de subidas no ranking	[0; 5]
Rankdif16	Diferença de posições entre a 1ª e 6ª observação	[-136; 998]
Rankimp	Resultado da diferença entre a 1ª e 6ª observação: melhorou, piorou ou ficou na mesma posição do ranking (situação em que o zero está excluído)	Melhou=1 Pior/Igual=0
Rankimp/maintain	Resultado da diferença entre a 1ª e 6ª observação: melhorou, piorou ou ficou na mesma posição do ranking (situação em que o zero está incluído)	Melhorou/Igual=1 Pior=0
P1	Preço 1 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 1)	[0; 35,99]
P2	Preço 2 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 2)	[0; 35,99]
P3	Preço 3 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 3)	[0; 35,99]
P4	Preço 4 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 4)	[0; 35,99]

P5	Preço 5 ((Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 5)	[0; 35,99]
P6	Preço 6 (Preço indicado no <i>Google Play</i> para aquisição da <i>app</i> , na observação 6)	[0; 35,99]
Pmean	Preço médio	[0; 35,99]
Pmed	Preço mediano	[0; 35,99]
Pmax	Preço máximo	35,99
Pmin	Preço mínimo	0
Psd	Desvio-padrão do preço	[0; 3,2]
Ppct5	Percentil 5 dos preços entre a 1ª e a 6ª observação	[0; 35,99]
Ppct95	Percentil 95 dos preços entre a 1ª e a 6ª observação	[0; 35,99]
Pchanges	Número de alterações nos preços	[0; 3]
Pdowns	Número de descidas nos preços	[0; 2]
Pups	Número de subidas nos preços	[0; 2]
Pdif16	Diferença de preços entre a 1ª e a 6ª observação	[-2,08; 7,86]
Pdin	Resultado da diferença de preços entre a 1ª e a 6ª observação: melhorou (1), piorou (-1) ou não sofreu alterações (0)	Melhorou=1 Piorou=-1 Igual=0
PI1	Produtos integrados em aplicações 1 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 1)	Refere-se apenas às <i>apps</i> gratuitas
PI2	Produtos integrados em aplicações 2 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 2)	Refere-se apenas às <i>apps</i> gratuitas
PI3	Produtos integrados em aplicações 3 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 3)	Refere-se apenas às <i>apps</i> gratuitas
PI4	Produtos integrados em aplicações 4 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 4)	Refere-se apenas às <i>apps</i> gratuitas
PI5	Produtos integrados em aplicações 5 (Preço dos produtos integrados da <i>app</i> indicado no <i>Google Play</i> , na observação 5)	Refere-se apenas às <i>apps</i> gratuitas
PI6	Produtos integrados em aplicações 6 (Preço dos produtos integrados	Refere-se apenas às <i>apps</i> gratuitas

	da <i>app</i> indicado no <i>Google Play</i> , na observação 6)	
PV1	Versão atual 1 (no momento da observação 1)	
PV2	Versão atual 2 (no momento da observação 2)	
PV3	Versão atual 3 (no momento da observação 3)	
PV4	Versão atual 4 (no momento da observação 4)	
PV5	Versão atual 5 (no momento da observação 5)	
PV6	Versão atual 6 (no momento da observação 6)	
PVchanges	Número de alterações ocorridas na versão atual	[0; 5]
Date 1	Data de atualização 1 (data da última atualização da <i>app</i> indicada no momento da observação 1)	[23 de Maio de 2012; 27 de Janeiro de 2016]
Date 2	Data de atualização 2 (data da última atualização da <i>app</i> indicada no momento da observação 2)	[23 de Maio de 2012; 18 de Janeiro de 2017]
Date 3	Data de atualização 3 (data da última atualização da <i>app</i> indicada no momento da observação 3)	[26 de Julho de 2013; 28 de Fevereiro de 2017]
Date 4	Data de atualização 4 (data da última atualização da <i>app</i> indicada no momento da observação 4)	[26 de Julho de 2013; 19 de Março de 2017]
Date 5	Data de atualização 5 (data da última atualização da <i>app</i> indicada no momento da observação 5)	[26 de Julho de 2013; 18 de Abril de 2017]
Date 6	Data de atualização 6 (data da última atualização da <i>app</i> indicada no momento da observação 6)	[26 de Julho de 2013; 18 de Maio de 2017]
Datechanges	Número de alterações ocorridas na data de atualização	[0; 5]
Dwnl 1	Classe de <i>downloads</i> 1 (no momento da observação 1)	D1 = 5 - 10 D2 = 10 - 50 D3 = 50 - 100 D4 = 100 - 500 D5 = 500 - 1.000 D6 = 1.000 - 5.000 D7 = 5.000 - 10.000 D8 = 10.000 - 50.000 D9 = 50.000 - 100.000 D10 = 100.000 - 500.000 D11 = 500.000 - 1.000.000

		D12 = 1.000.000 - 5.000.000
		D13 = 5.000.000 - 10.000.000
		D14 = 10.000.000 - 50.000.000
		D15 = 50.000.000 - 100.000.000
		D16 = 100.000.000 - 500.000.000
Dwnl 2	Classe de <i>downloads</i> 2 (no momento da observação 2)	<i>Idem</i>
Dwnl 3	Classe de <i>downloads</i> 3 (no momento da observação 3)	<i>Idem</i>
Dwnl 4	Classe de <i>downloads</i> 4 (no momento da observação 4)	<i>Idem</i>
Dwnl 5	Classe de <i>downloads</i> 5 (no momento da observação 5)	<i>Idem</i>
Dwnl 6	Classe de <i>downloads</i> 6 (no momento da observação 6)	<i>Idem</i>
Dwnlchanges	Número de alterações ocorridas na classe de <i>downloads</i>	[0; 2]
Dwnldowns	Número de descidas no número de <i>downloads</i>	0 e 2
Dwnlups	Número de subidas no número de <i>downloads</i>	[0; 2]
Android Version	Versão <i>Android</i> requerida para cada <i>app</i>	[1.5; 4.4]

Apêndice 3

Explicação do código glm associado às fórmulas *mylogit*:

mylogit corresponde apenas ao nome definido para apresentar o modelo, o qual diz respeito a um modelo linear generalizado (“*glm*”) que fornece uma descrição simbólica do preditor linear e da distribuição de erro, ou seja, permite apurar a correlação entre variáveis; a “*formula*” é indicativa das variáveis em análise: a primeira respeita à variável que se pretende explicar (dependente) e as seguintes são as variáveis explicativas (independentes); o elemento “*family*” diz respeito ao tipo de variável que resulta do modelo glm, o que, por defeito em associação a uma função logit, é binomial (isto é, uma variável com resposta positiva ou negativa); o elemento “*data*” remete para o comando através do qual o RStudio absorve a informação necessária para o cálculo da fórmula.

Apêndice 4

Passos do RStudio

```
# Convert to numeric
df$Rankmean <- as.numeric(df$Rankmean)
df$Rankmed <- as.numeric(df$Rankmed)
df$Ranksd <- as.numeric(df$Ranksd)
df$Rankpct5 <- as.numeric(df$Rankpct5)
df$Rankpct95 <- as.numeric(df$Rankpct95)
df$P1 <- as.numeric(df$P1)
df$P2 <- as.numeric(df$P2)
df$P3 <- as.numeric(df$P3)
df$P4 <- as.numeric(df$P4)
df$P5 <- as.numeric(df$P5)
df$P6 <- as.numeric(df$P6)
df$Pmean <- as.numeric(df$Pmean)
df$Pmed <- as.numeric(df$Pmed)
df$Pmax <- as.numeric(df$Pmax)
df$Pmin <- as.numeric(df$Pmin)
df$Psd <- as.numeric(df$Psd)
df$Ppct5 <- as.numeric(df$Ppct5)
df$Ppct95 <- as.numeric(df$Ppct95)
df$PI1 <- as.numeric(df$PI1)
df$PI2 <- as.numeric(df$PI2)
df$PI3 <- as.numeric(df$PI3)
df$PI4 <- as.numeric(df$PI4)
df$PI5 <- as.numeric(df$PI5)
df$PI6 <- as.numeric(df$PI6)
df$PImean <- as.numeric(df$PImean)
df$PImax <- as.numeric(df$PImax)
df$PImin <- as.numeric(df$PImin)
df$PV1 <- as.numeric(df$PV1)
df$PV2 <- as.numeric(df$PV2)
df$PV3 <- as.numeric(df$PV3)
df$PV4 <- as.numeric(df$PV4)
df$PV5 <- as.numeric(df$PV5)
```

```

df$PV6 <- as.numeric(df$PV6)
df$Version <- as.numeric(df$version)

# Convert unordered factor to ordered factor
df$Dwnl1 <- ordered(df$Dwnl1, levels = c('D1', 'D2', 'D3', 'D4', 'D5', 'D6', 'D7', 'D8', 'D9',
'D10', 'D11', 'D12', 'D13', 'D14', 'D15', 'D16'))
df$Dwnl2 <- ordered(df$Dwnl2, levels = c('D1', 'D2', 'D3', 'D4', 'D5', 'D6', 'D7', 'D8', 'D9',
'D10', 'D11', 'D12', 'D13', 'D14', 'D15', 'D16'))
df$Dwnl3 <- ordered(df$Dwnl3, levels = c('D1', 'D2', 'D3', 'D4', 'D5', 'D6', 'D7', 'D8', 'D9',
'D10', 'D11', 'D12', 'D13', 'D14', 'D15', 'D16'))
df$Dwnl4 <- ordered(df$Dwnl4, levels = c('D1', 'D2', 'D3', 'D4', 'D5', 'D6', 'D7', 'D8', 'D9',
'D10', 'D11', 'D12', 'D13', 'D14', 'D15', 'D16'))
df$Dwnl5 <- ordered(df$Dwnl5, levels = c('D2', 'D3', 'D4', 'D5', 'D6', 'D7', 'D8', 'D9', 'D10',
'D11', 'D12', 'D13', 'D14', 'D15', 'D16'))
df$Dwnl6 <- ordered(df$Dwnl6, levels = c('D2', 'D3', 'D4', 'D5', 'D6', 'D7', 'D8', 'D9', 'D10',
'D11', 'D12', 'D13', 'D14', 'D15', 'D16'))

```

De seguida, recorreu-se à função *is.na* (*Logit1*) para despistar os valores nulos (*NA*). Uma vez que todas as células estão preenchidas com o termo *FALSE*, então não temos nenhum valor em falta.

```

is.na(Logit)
      Rankimp Rankchanges PVchanges Dwnlchanges
[1,] FALSE      FALSE      FALSE      FALSE
[2,] FALSE      FALSE      FALSE      FALSE
[3,] FALSE      FALSE      FALSE      FALSE
[4,] FALSE      FALSE      FALSE      FALSE
[5,] FALSE      FALSE      FALSE      FALSE
[... ] ...          ...          ...          ...

```

Teste de independência Qui-Quadrado

#logit1

```
Q1 <- df [c ("Rankimp", "Rankchanges", "PVchanges", "Dwnlchanges")]
```

```
table (Q1$Rankimp, Q1$Rankchanges)
```

```
chisq.test(table (Q1$Rankimp, Q1$Rankchanges))
```

```
Q2 <- df [c ("Rankimp", "Rankchanges", "PVchanges", "Dwnlchanges")]
```

```
table (Q2$Rankimp, Q2$PVchanges)
```

```
chisq.test(table (Q2$Rankimp, Q2$PVchanges))
```

```
Q3 <- df [c ("Rankimp", "Rankchanges", "PVchanges", "Dwnlchanges")]
```

```
table (Q3$Rankimp, Q3$Dwnlchanges)
```

```
chisq.test(table (Q3$Rankimp, Q3$Dwnlchanges))
```

#logit2

```
Q1 <- df [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns")]
```

```
table (Q1$Rankimp, Q1$Rankups)
```

```
chisq.test(table (Q1$Rankimp, Q1$Rankups))
```

```
Q2 <- df [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns")]
```

```
table (Q2$Rankimp, Q2$Rankdowns)
```

```
chisq.test(table (Q2$Rankimp, Q2$Rankdowns))
```

```
Q3 <- df [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns")]
```

```
table (Q3$Rankimp, Q3$Dwnlups)
```

```
chisq.test(table (Q3$Rankimp, Q3$Dwnlups))
```

```
Q4 <- df [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups", "Dwnldowns")]
```

```
table (Q4$Rankimp, Q4$Dwnldowns)
```

```
chisq.test(table (Q4$Rankimp, Q4$Dwnldowns))
```

#logit3

```
Q1 <- df [c ("Rankimp_maintain", "Rankchanges", "PVchanges", "Dwnlchanges")]
```

```
table (Q1$Rankimp_maintain, Q1$Rankchanges)
```

```
chisq.test(table (Q1$Rankimp_maintain, Q1$Rankchanges))
```

```
Q2 <- df [c ("Rankimp_maintain", "Rankchanges", "PVchanges", "Dwnlchanges")]
```

```
table (Q2$Rankimp_maintain, Q2$PVchanges)
```

```
chisq.test(table (Q2$Rankimp_maintain, Q2$PVchanges))
```

```
Q3 <- df [c ("Rankimp_maintain", "Rankchanges", "PVchanges", "Dwnlchanges")]
```

```
table (Q3$Rankimp_maintain, Q3$Dwnlchanges)
```

```
chisq.test(table (Q3$Rankimp_maintain, Q3$Dwnlchanges))
```

#logit4

```
Q1 <- df [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
table (Q1$Rankimp_maintain, Q1$Rankups)
```

```
chisq.test(table (Q1$Rankimp_maintain, Q1$Rankups))
```

```
Q2 <- df [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
table (Q2$Rankimp_maintain, Q2$Rankdowns)
```

```
chisq.test(table (Q2$Rankimp_maintain, Q2$Rankdowns))
```

```
Q3 <- df [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
table (Q3$Rankimp_maintain, Q3$Dwnlups)
```

```
chisq.test(table (Q3$Rankimp_maintain, Q3$Dwnlups))
```

```
Q4 <- df [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
table (Q4$Rankimp_maintain, Q4$Dwnldowns)
```

```
chisq.test(table (Q4$Rankimp_maintain, Q4$Dwnldowns))
```

#logit5

```
Q1 <- subset_6obs [c ("Rankimp_maintain", "Rankchanges", "PVchanges",  
"Dwnlchanges")]
```

```
table (Q1$Rankimp_maintain, Q1$Rankchanges)
```

```
chisq.test(table (Q1$Rankimp_maintain, Q1$Rankchanges))
```

```
Q2 <- subset_6obs [c ("Rankimp_maintain", "Rankchanges", "PVchanges",  
"Dwnlchanges")]
```

```
table (Q2$Rankimp_maintain, Q2$PVchanges)
```

```
chisq.test(table (Q2$Rankimp_maintain, Q2$PVchanges))
```

```
Q3 <- subset_6obs [c ("Rankimp_maintain", "Rankchanges", "PVchanges",  
"Dwnlchanges")]
```

```
table (Q3$Rankimp_maintain, Q3$Dwnlchanges)
```

```
chisq.test(table (Q3$Rankimp_maintain, Q3$Dwnlchanges))
```

#logit6

```
Q1 <- subset_6obs [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
table (Q1$Rankimp_maintain, Q1$Rankups)
```

```
chisq.test(table (Q1$Rankimp_maintain, Q1$Rankups))
```

```
Q2 <- subset_6obs [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
table (Q2$Rankimp_maintain, Q2$Rankdowns)
```

```
chisq.test(table (Q2$Rankimp_maintain, Q2$Rankdowns))
```

```
Q3 <- subset_6obs [c ("Rankimp_maintain", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
table (Q3$Rankimp_maintain, Q3$Dwnlups)
```

```
chisq.test(table (Q3$Rankimp_maintain, Q3$Dwnlups))
```

```
Q4 <- subset_6obs [c ("Rankimp", "Rankups", "Rankdowns", "Dwnlups",  
"Dwnldowns")]
```

```
table (Q4$Rankimp, Q4$Dwnldowns)
```

```
chisq.test(table (Q4$Rankimp, Q4$Dwnldowns))
```