Can smartphones cross the digital divide? The demographics of technology uptake. *

Laurence Warner†

7/29/18

1 Abstract

Technological change in communication has changed American society so much that some deem it the 'third industrial revolution'. Internet, high-speed broadband and cell phone technology all swept through American society at an incredibly rapid rate, yet there was a 'digital divide' in speed of uptake across socioeconomic groups in society, which still persists today to a certain extent. Internet-enabled smartphones represent a huge change in the ability of people to access the internet ubiquitously, and some detect a bright future for a 'mobile-only' internet. However, this paper shows that whereas smartphone ownership is indeed associated with a socioeconomic elite, smartphone dependence is found mainly among society's disadvantaged. The sociological theory of the usage gap suggests that this could forebode widening inequality in the future.

Keywords: Technology, Internet, Demographics, Sociology, Statistical Inference, Limited Dependent Variable Models

*Iteration

†lpwarner@uchicago.edu. My thanks to Dr. Benjamin Soltoff for his advice and guidance.

2 Introduction

2.1 Internet

The dawn of the internet age has been one of the most profoundly impactful technological changes on our economy. The Economist described it as causing the ’third industrial revolution’ [4]. However, as the theories of the historical materialists advised, changes to the economic base lead to changes in society. For the average American, daily life in 2018 hardly resembles 1998: communication, information and entertainment are all primarily routed through online channels.

2.2 Phones

Looking with a historical perspective, mobile phone technology has experienced a phenomenally rapid spread [3], even faster than that of the internet. And since Apple’s iPhone raised the bar of the technological possibilities of the internet-enabled smartphone in 2007, this variant of mobile technology has been almost ubiquitous in the West. As Steve Jobs famously announced, it combines the media capabilities of the iPod and the internet connectivity of a broadband-connected laptop into the design of a handheld cellular device.

2.3 Perceived universality

However, internet connection has not reached everybody in society. Quite the contrary, technological uptake, like any other economic process, occurs along existing lines of social stratification. This has led to the coinage of the phrase 'the digital divide' to describe the divide between internet haves and have-nots [8]. Whilst in the past,

2.4 Smartphone dependence

Some predict that the mobile web will make the desktop web experience redundant. Forbes counsels companies to switch from a mobile-friendly to a mobile-first web strategy, and Uber, recently named the largest player in the automobile industry, eschews a web experience [5].

This could lead some to conclude that the economic elites of the near future will not need broadband. This state of engaging with the internet without having a home broadband connection is sometimes called 'smartphone dependence'. Whilst this name could be slightly misleading - laptops can be taken to coffee shops, it implies an addictive element which may not be present - it is a useful heuristic to describe those who can generally access the internet at home only on a handheld screen.

This paper investigates whether the growing phenomenon of smartphone dependence conforms with other trends: being led by economic elites, or whether it actually represents a negative digital divide.

2.5 Why it matters

Understanding who is taking up new smartphone technology is vital for understanding social trends. As already implied, technology uptake is very closely related with economic activity.

In addition, there is growing concern around the public health impacts of technological uptake, particularly amongst the young who are the most reactive. There is a growing body of research around the concept of internet addiction and problematic internet use [1, 6, 10]. There have been suggestions that the smartphone is an exceptionally impactful technological device upon mental health [14, 9]. Most strikingly, a UT Austin experiment found that the mere presence of a smartphone reduces one's available cognitive capacity [13].

Evidently, how social groups uptake and use smartphone technology - productively and healthily; or otherwise - may determine their socioeconomic future, so it is vital for us to understand.

3 Theory & Literature

3.1 Digital Divide

A digital divide is a state in which there are structural factors influencing who in society possesses and uses technology. In their underrated study of the digital divide, communication theorists Tsetsi and Rains divide digital divide research into two streams [11].

Level 1 of the digital divide regards ownership. New technology is often highly expensive: only those in the economic elite can afford them, with early research focussing on poorer parts of society missing out on broadband internet [12].

3.2 Usage gap hypothesis

Level 2 of the digital divide theory concerns how different individuals use the same device differently.

Its heritage is from the theory of the knowledge gap: as information diffuses through media, higher socioeconomic status (SES) groups acquire information at a faster rate, thus exacerbating inequality further.

Scholars extended this idea to knowledge on how to use technologies like the Internet Le14. For example, higher SES groups will be educated to install tools like ad blockers to make their online time more productive.

3.3 Where do smartphones fit in?

Many would assume that smartphone ownership is merely a question of the Level 1 Digital Divide. Many observe that mobile phone ownership is correlated with socioeconomic status - indeed a recent study by University of Chicago scholars found iPhone ownership to be the single most precise predictor of being in the top income quartile of the country - and assume that eschewing broadband connection is a benign social trend [5].

3.4 Previous studies

However, Napoli and Obar questioned this conventional wisdom in their piece "Second Class Netizens: Race and the Emerging Mobile Internet Underclass" [7]. They question the substantive claim that mobile technology represents a productive innovation. They show that for many tasks - such as submitting job applications and deeper research - smartphones are inferior to desktops from a design viewpoint. Moreover, according to the usage gap hypothesis, they show that disadvantaged members of society gravitate to less ostensibly productive activities on the smartphone such as social media and are less savvy at effectively searching for enriching information. Moreover, I would posit a novel argument: that the knowledge gap's emphasis on speed of information acquisition is no longer relevant: the knowledge gap is now on knowing which (immediately available) information is *not* worth absorbing.

Tsetsi and Rains nuanced this perspective, making the "horses for courses" argument - the smartphone is superior for many tasks, but not others, so smartphone dependence will harm social mobility. In this sense they blend Level 1 and Level 2 of the digital divide theory, arguing that the range of devices you own affects how you use those devices.

From this perspective they went to the data with the following hypothesis: demographics like race, sex, age, income and education are associated with smartphone dependence. Using the Pew Research Center's survey data, considered the gold standard for surveys on digital life, they found that minorities, less educated and lower income groups were more likely to be smartphone dependent.

4 Data

4.1 Pew

We follow Tsetsi and Rains in using the Pew Research Center's survey data. Whereas they used a 2012 survey, we use a new comprehensive survey conducted in 2018: Pew Research Center?s Internet Project Core Trends Survey. It can be accessed freely online here: http://www.pewinternet.org/datasets/

Interviews with a nationally representative sample of 2,002 adults were conducted between January 3-10 2018. The target population for the study is non-institutionalized persons age 18 and over, living in the US. Most of the interviews were conducted using cellphones (n = 1,502) with the remainder conducted using landlines (n=500); both groups were included in the final sample. According to the Pew Research Center, the landline sample was collected using a proportional sample based on listed telephone households. The cellphone sample was selected systematically from dedicated wireless numbers. This dataset contains questions about social media use in 2018 and attitudes toward the internet and whether Americans think it's good or bad for society. Random digit dialing was used to collect survey responses and the final sample was weighted to represent the American adult population. The sample response rate was 11%.

Whilst this sample response rate does pose some issues of bias, Pew's statistical corrections assuage these fears. Besides we don't have much choice: in order to gain a full empirical picture of the relationship between demographics and device usage, old-fashioned data collection methods must be used. Whilst major ground has been broken regarding smartphone usage habits using the rich data collection that such devices enable, like the Menthal Project or Cambridge University's Device Analyzer, both research projects based upon Android apps, unfortunately such studies exclude the very people we are interested in understanding - those who eschew either smartphone or broadband technology [2].

4.2 Cleaning and pre-processing

The dataset was wrangled using the Python library Pandas. I chose to discard rows which contained answers including 'Don't Know' or 'Refused to answer'. This reduced the total size of the dataset to N = 1561, which fortunately is considered sufficient for more inferential and Machine Learning (ML) models.

4.2.1 Target variables

For both of the target variables, smartphone and broadband usage, I decided to construct ternary variables of the following form: full ownership (smartphone/broadband), part-way there ('dumb' internet-disabled phone/other internet), rejection (no phone/internet):

	home	outside home	no internet
smart	0.673286	0.119795	0.017297
dumb	0.082639	0.032671	0.045484
none	0.007687	0.001281	0.019859

Figure 1: Smartphone & Broadband Usage

A clearer picture emerges from a visualization.

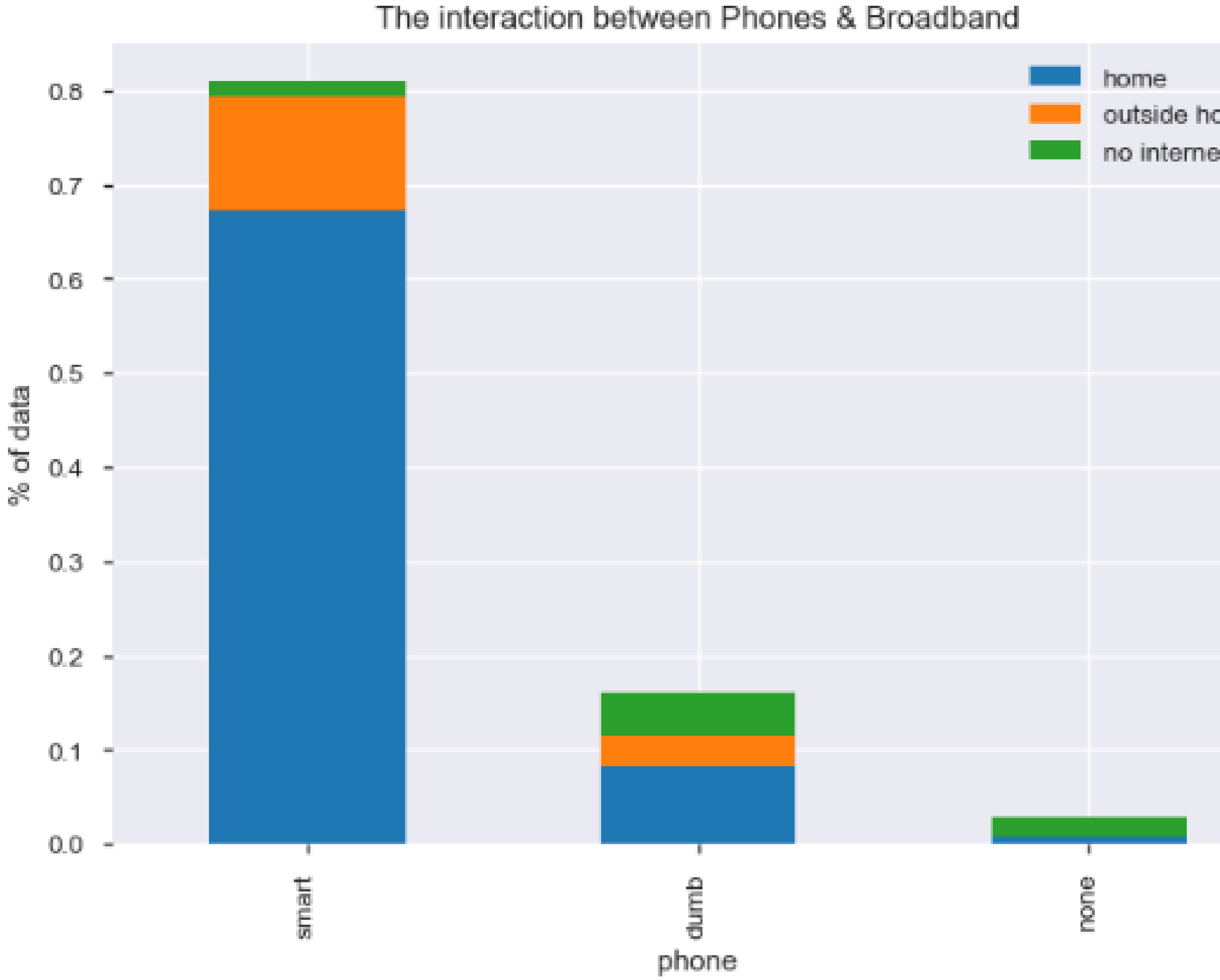

Figure 2:

As can be seen the vast majority of society today (80%) are smartphone users. However, the orange segment eschews traditional internet. We follow Tsetsi and Rains in defining these people as our smartphone dependent population (we also include the green segment, even though they supposedly don't use their smartphones for internet) and construct a new binary label 'dep'. I separately group those with dumb and no phones to create a binary smartphone label 'smart_bin'.

4.2.2 Predictor variables

We categorized age according to the following cutoffs:

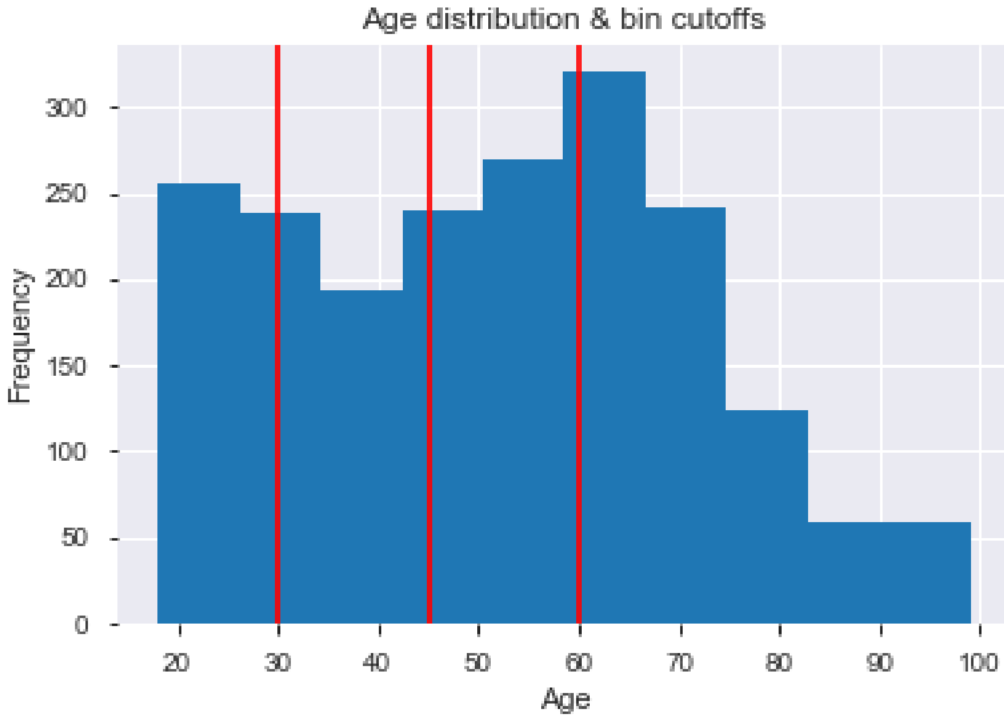

Figure 3: Age distribution & cutoffs

I left ordered categorical variables - age, education, income - as ascending integers. Even though this is not a completely accurate scale, it is is a sufficient approximation. I binarized race using one-hot encoding (having to discard the handful of respondents who answered 'Islander' to avoid multicollinearity).

5 Methods & Results

5.1 Logistic Regression: interpretation

Before discussing my major results, I want to explore the interpretability of logistic regression. I will use the outcome variable of smartphone ownership (=1 if smartphone, =0 if none) and a univariate model which includes a single predictor variable: *sex* (which =1 if female and =0 if male). The logit model chooses the beta coefficients in the following model to minimize the sum of the squared errors:

$$logit(p) = log(p/(1-p)) = \beta_0 + \beta_1 * sex \quad (1)$$

Thus the exponent of β_1 can be interpreted as the impact of being female upon the log-odds ratio. If the coefficient is negative, being female means you are less likely to have a smartphone and vice versa. Let's look at the model built from our data:

```
Optimization terminated successfully.
         Current function value: 0.483610
         Iterations 6
                           Logistic Regression
==============================================================================
Dep. Variable:              smart_bin   No. Observations:                 156
Model:                          Logit   Df Residuals:                     156
Method:                           MLE   Df Model:
Date:            Mon, 30 Jul 2018   Pseudo R-squ.:               0.00588
Time:                        03:02:08   Log-Likelihood:              -757.8
converged:                       True   LL-Null:                     -762.3
                                        LLR p-value:                 0.00274
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975
------------------------------------------------------------------------------
const          2.0228      0.206      9.815      0.000       1.619       2.42
sex           -0.3863      0.129     -2.991      0.003      -0.640      -0.13
==============================================================================
```

Figure 4: Logistic regression output

The negative coefficient (-0.39) shows that females are indeed less likely to own smartphones in our dataset. This is corroborated by a visual cross-tabulation of the data:

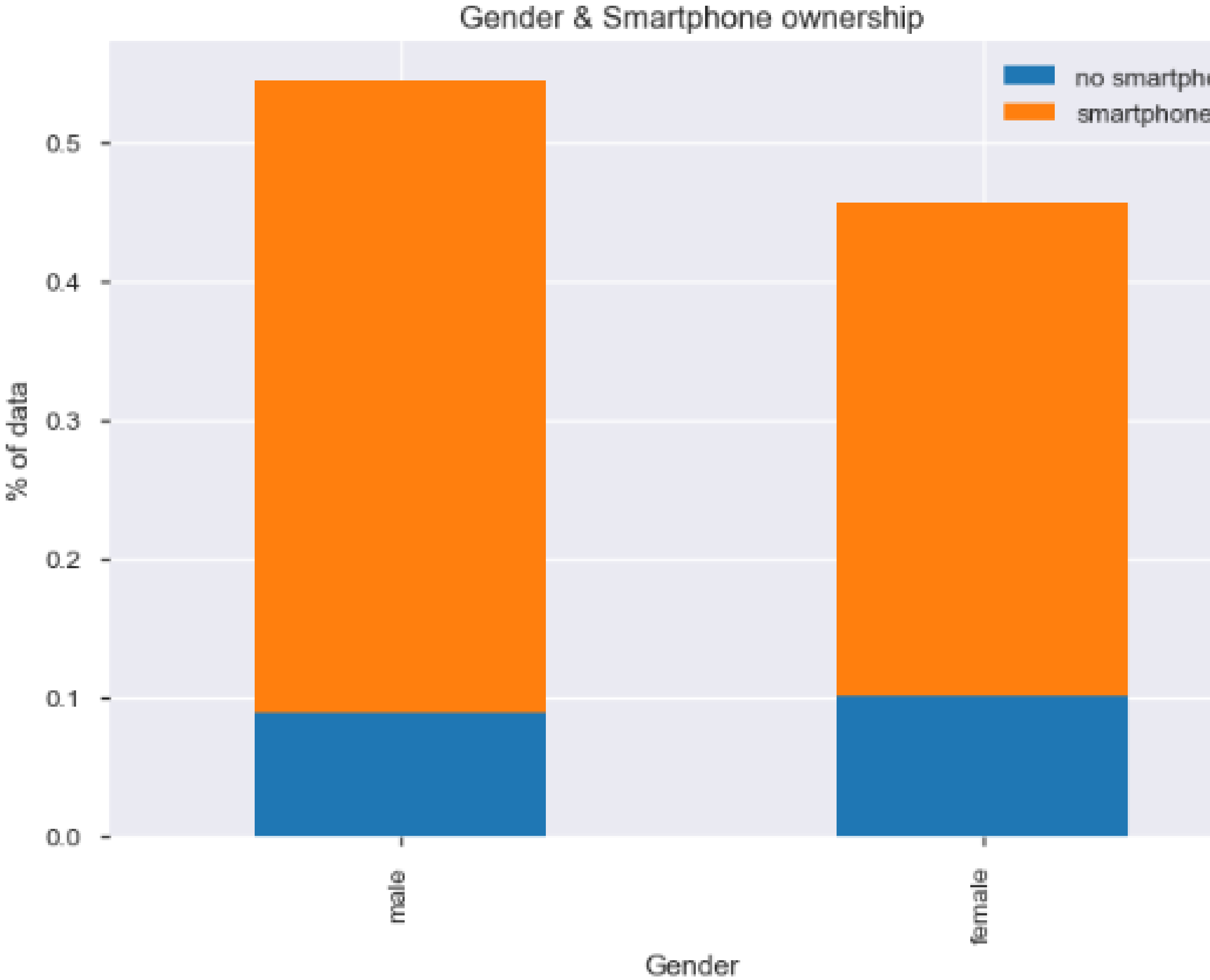

Figure 5:

Moreover the fact that the p-value was below 5% means that this finding is statistically significant: it is not just a random artefact of a small sample.

We can be even more precise about our inference. By taking the exponent, we can show that the log odds ratio decreases to 0.69 (taking the exponent of the coefficient) from males to females.

5.2 Models

We are now ready to analyze the multivariate logistic regressions for our two outcome variables. Firstly, smartphone ownership:

```
Optimization terminated successfully.
         Current function value: 0.377380
         Iterations 8
                           Logistic Regression
==============================================================================
Dep. Variable:              smart_bin   No. Observations:                  156
Model:                          Logit   Df Residuals:                      155
Method:                           MLE   Df Model:                            1
Date:                Mon, 30 Jul 2018   Pseudo R-squ.:                   0.224
Time:                        19:29:09   Log-Likelihood:                 -591.3
converged:                       True   LL-Null:                        -762.3
                                        LLR p-value:                  2.086e-6
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.8539      0.493      5.785      0.000       1.887       3.82
sex           -0.0680      0.149     -0.458      0.647      -0.359       0.22
age           -0.9467      0.084    -11.304      0.000      -1.111      -0.78
educ           0.2239      0.049      4.557      0.000       0.128       0.32
non_hisp      -0.4864      0.306     -1.591      0.112      -1.086       0.11
inc            0.3011      0.036      8.459      0.000       0.231       0.37
white         -0.0136      0.387     -0.035      0.972      -0.773       0.74
black          0.0528      0.428      0.124      0.902      -0.785       0.89
asian          1.0223      0.839      1.218      0.223      -0.623       2.66
other         -0.0074      0.937     -0.008      0.994      -1.843       1.82
native         0.0839      0.655      0.128      0.898      -1.200       1.36
==============================================================================
```

Figure 6: Logistic regression

The significant variables here are age, education and income (note that sex is no longer statistically significant, implying the effect was through interaction with those terms, probably income primarily). Let's consider the directions: higher SES individuals (higher education and income) are more likely to own smartphones. Also, older people are less likely to (this may have more to do with usage aptitude etc.).

Now let us consider the same analysis for a new outcome variable: whether an individual is smartphone dependent.

```
Optimization terminated successfully.
         Current function value: 0.337199
         Iterations 7
                           Logistic Regression
==============================================================================
Dep. Variable:                    dep   No. Observations:                 156
Model:                          Logit   Df Residuals:                     155
Method:                           MLE   Df Model:                           1
Date:                Mon, 30 Jul 2018   Pseudo R-squ.:                  0.156
Time:                        19:29:09   Log-Likelihood:                -526.3
converged:                       True   LL-Null:                       -623.8
                                        LLR p-value:                 1.807e-3
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975
------------------------------------------------------------------------------
const          2.4003      0.446      5.384      0.000       1.526       3.27
sex           -0.2434      0.163     -1.494      0.135      -0.563       0.07
age           -0.3054      0.072     -4.264      0.000      -0.446      -0.16
educ          -0.2246      0.054     -4.151      0.000      -0.331      -0.11
non_hisp      -0.8502      0.234     -3.636      0.000      -1.309      -0.39
inc           -0.2428      0.038     -6.466      0.000      -0.316      -0.16
white          0.4396      0.305      1.440      0.150      -0.159       1.03
black          0.7382      0.353      2.092      0.036       0.047       1.43
asian          0.0807      0.584      0.138      0.890      -1.063       1.22
other          1.7867      0.770      2.320      0.020       0.277       3.29
native         1.3406      0.563      2.382      0.017       0.237       2.44
==============================================================================
```

Figure 7: Logistic regression

Let's observe what has changed. The same variables are still significant: but the effect of education and income has reversed. In addition, some race variables (non-hisp, black, other, native) are now significant, in the direction that those ethnic minorities are more likely to be smartphone dependent.

6 Discussion & Conclusion

We find that smartphone ownership in general has followed the same Level 1 digital divide trajectory of previous technologies, with high SES individuals taking it up first.

However, we find an opposite effect when considering folks for whom the smartphone is their primary device to engage with the internet. This reinforces Tsetsi and Rains's findings from the similar 2012 dataset.

Considering the usage gap hypothesis, this does not bode well for the future of disadvantaged and ethnic minority groups in reaping the benefits of technological progress.

7 Avenues for further investigation

Data from other countries could be considered - for example developing countries.

Deeper outlier analysis could be done to identify whether there are indeed young, high-SES mobile-only users (think Evan Spiegel), and perhaps even young, high-SES users who eschew smartphone technology.

8 Iteration: note for Dr. Soltoff

I decided to leave out the machine learning analysis in this iteration, as I felt it was dubiously relevant to the theoretical focus. However, I can reintroduce it with some relevance for the final paper if you think it is not currently sufficiently computationally intensive.

Style note: I can remove the subsection headings if you think best. I can also try to include references names in parentheses. Happy to take other style pointers too.

References

[1] Emma Louise Anderson, Eloisa Steen, and Vasileios Stavropoulos. "Internet use and Problematic Internet Use: A systematic review of longitudinal research trends in adolescence and emergent adulthood". In: *International Journal of Adolescence and Youth* 22.4 (2017), pp. 430–454.

[2] Ionut Andone et al. "How age and gender affect smartphone usage". In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM. 2016, pp. 9–12.

[3] Manuel Castells et al. *Mobile communication and society: A global perspective*. Mit Press, 2009.

[4] The Economist. *The third industrial revolution*. 2012. URL: http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm (visited on 07/28/2018).

[5] Forbes. *The Shift From Mobile-Friendly To Mobile-First: What Your Brand Should Know*. 2017. URL: https://www.forbes.com/sites/gabrielshaoolian/2017/07/13/the-shift-from-mobile-friendly-to-mobile-first-what-your-brand-should-know/#60701f194626 (visited on 07/28/2018).

[6] Min Kwon et al. "Development and validation of a smartphone addiction scale (SAS)". In: *PloS one* 8.2 (2013), e56936.

[7] Philip M Napoli and Jonathan A Obar. “SECOND CLASS NETIZENS”. In: *Race and Gender in Electronic Media: Content, Context, Culture* (2016), p. 79.

[8] Pippa Norris et al. *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press, 2001.

[9] Dmitri Rozgonjuk and Jon Elhai. “Problematic smartphone usage, emotion regulation, and social and non-social smartphone use”. In: *Proceedings of the Technology, Mind, and Society*. ACM. 2018, p. 35.

[10] Maya Samaha and Nazir S Hawi. “Relationships among smartphone addiction, stress, academic performance, and satisfaction with life”. In: *Computers in Human Behavior* 57 (2016), pp. 321–325.

[11] Eric Tsetsi and Stephen A Rains. “Smartphone Internet access and use: Extending the digital divide and usage gap”. In: *Mobile Media & Communication* 5.3 (2017), pp. 239–255.

[12] Jan AGM Van Dijk. *The deepening divide: Inequality in the information society*. Sage Publications, 2005.

[13] Adrian F Ward et al. “Brain drain: the mere presence of one?s own smartphone reduces available cognitive capacity”. In: *Journal of the Association for Consumer Research* 2.2 (2017), pp. 140–154.

[14] Claire A Wolniewicz et al. “Problematic smartphone use and relations with negative affect, fear of missing out, and fear of negative and positive evaluation”. In: *Psychiatry research* 262 (2018), pp. 618–623.