



## Research Article

# A Statistical Analysis on The Effects of a Nation's Individual Network Usage on A Business's Capacity for Innovation

Muhammad C. Alba<sup>1\*</sup>

<sup>1</sup>The Pennsylvania State University, University Park Campus, PA, United States of America

### Article Information

Received: 19 May 2020

Revised version received: 30 June 2020

Accepted: 05 July 2020

Published: 27 July 2020

#### Cite this article as:

M.C. Alba (2020) *Int. J. Soc. Sc. Manage.* 7(3): 113-130. DOI: [10.3126/ijssm.v7i3.29765](https://doi.org/10.3126/ijssm.v7i3.29765)

#### \*Corresponding author

Charles Alba,

The Pennsylvania State University, University Park Campus, PA, United States of America

Email: [cja5553@psu.edu](mailto:cja5553@psu.edu)

Peer reviewed under authority of IJSSM

© 2020 IJSSM, Permits unrestricted use under the CC-By-NC license.

OPEN ACCESS



This is an open access article & it is licensed under a [Creative Commons Attribution Non-Commercial 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/)

[International](https://creativecommons.org/licenses/by-nc/4.0/)

(<https://creativecommons.org/licenses/by-nc/4.0/>)

**Keywords:** ICT networks; Enterprise; Innovation Capacity; Network Readiness; Individual Network Usage; ICT policy; Public Policy

### Abstract

In this paper, we aimed to look at a Nations Individual Network usage and its effects on a business's capacity for innovation. We do so through the utilization of data from 2016 World Economic Forum (WEF) Networked Readiness Index report. In order to best determine the effects of a Nations Individual Network usage on a business's capacity for innovation, we create the most suitable supervised Machine Learning model through stepwise regression. We did so with the use of the numerous variables in the 6th pillar of the Network Readiness report, which reflect a nations individual network usage, against the Capacity for innovation Index found in the 7th pillar of the Network Readiness Index. Of the 6 initial prior variables in the 6th Pillar reflecting Individual Network Usage, stepwise regression outputs a best fit and an appropriate model that contains predictors "Use of virtual social networks", "Rate of Mobile broadband subscriptions" and "Rate of Mobile phone subscriptions" as significant factors in predicting a Business's Capacity for innovation. The results are utilized to make policy-proposals to both governments and businesses with the aim of improving a nations Business's Capacity for Innovation.

**Correction:** At the request of the author, on 17th October 2022, the name of the author was changed FROM Charles Alba TO Muhammad C. Alba.

### Introduction

As the world continues to evolve amidst globalization, the need for internet and technological exposures amongst individuals of nations remains more prevalent than ever. Such is displayed by the mere fact that as of 2011, the United Nations has considered the Internet access a fundamental human right (Ford, 143). This individualistic need has certainly fueled desires of technological innovation amongst corporations. Thus, in this paper, we aim to look at a Nations Individual Network usage and its

effects on a business's capacity for innovation. We do so through the utilization of data from 2016 World Economic Forum (WEF) Networked Readiness Index report. In order to best determine the effects of a Nations Individual Network usage on a business's capacity for innovation, we create the most suitable supervised Machine Learning model through stepwise regression. We proceed so with the use of the numerous variables in the 6<sup>th</sup> pillar of the Network Readiness report, which reflect a nations individual network

usage, against the Capacity for innovation Index found in the 7<sup>th</sup> pillar of the Network Readiness Index

#### **What is the WEF Network Readiness Index Report?**

The Network Readiness Index report, published annually by the World Economic Forum until 2016, is designed to measure “how well an economy is using information and communications technologies to boost competitiveness and well-being” (Keith, 2016). It aims to ultimately reflect Networked Readiness as a key indicator towards assessing how well a country performs in the digital world. The report contains a variety of indicators which could be divided into 3 overall categories: Overall Environment, Readiness and Usage. These 3 categories comprise of 10 composite pillars, each consisting of several variables. This structure could be best summarized below.

##### 1) Overall Environment sub-index

- 1st Pillar: Political and regulatory environment
- 2nd Pillar: Business and innovation environment
- 9th Pillar: Economic impact
- 10th Pillar: Social impact

##### 2) Readiness sub-index

- 3rd Pillar: Infrastructure
- 4th Pillar: Affordability
- 5th Pillar: Skills

##### 3) Usage sub-index

- 6th Pillar: Individual usage
- 7th Pillar: Business usage
- 8th Pillar: Government usage

(Baller *et al.*, 2016, p. 5; Malisuwan *et al.*, 2016, p. 405)

As clearly shown by the above structure, the Network Readiness Index report hopes to use a large variety of variables in determining “whether a country possesses the drivers necessary for digital technologies to meet their potential” (Breene, 2016) and most importantly assess if these technologies, in reality, translate to positive impacts towards the social, economic and business environment.

The paper was first published by the World Economic Forum in 2001. The last Network Readiness Index report published by the World Economic Forum was in 2016.

In this paper, we shall focus on the variables in “Individual Usage”, the 6<sup>th</sup> pillar, and its impact on a business’s “Capacity for Innovation” index, found under the 7<sup>th</sup> pillar the Network Readiness Index Report.

#### **Linkages between Individual Network Use and Capacity of Business Innovation**

With technology becoming more complex yet crucial over-time, a nation’s Individual Network Usage, which signifies exposure and activeness with regards to network technology, remains more critical than ever. The ability of an individual within any nation to gain network access in

different ways ensures exposure through this rapid and important phase of technological advancements.

Through this exposure, individuals would be able to familiarize themselves with imperative network tools. By gaining the proficiency of network technology, they would be able to also test out new forms of technology. Such is noted by Grant, where the “role of the individual as the primary actor in knowledge creation and the principle repository of knowledge is essential to piercing the veil of organizational knowledge and clarifying the role of organizations in the creation and application of knowledge” (Grant, 1996, p. 121).

This process gives one the ability to apply these individual exposures to their workplace. Thus, capacity towards Business Innovation is expected to be improved and enhanced. This is in line with Rothaermel and Hess’s literature, whom emphasizes that “individual-level factors (are crucial) in explaining firm-level heterogeneity in innovation” (Rothaermel and Hess, 2007, p. 195).

Individual Network Usage, while powerful, could be easily underestimated. As noted by Ibarra, “impact of individual attributes and formal position” impacts “administrative innovation roles to a greater extent than it mediated their impact on technical roles” (Ibarra, 2017).

## **Methods**

### **Data Collection**

As mentioned earlier, in this paper, we aim to find a nation’s Individual Network Usage, with variables from the 6<sup>th</sup> pillar, and its impact on a business’s “Capacity for Innovation” index, found under the 7<sup>th</sup> pillar the Network Readiness Index Report. Therefore, we would discuss two data collection interests in this group – Individual Network Usage and Capacity for Innovation.

#### **Individual Network Usage:**

Based on the Data made available by the World Economic Forum, we gain access to 6 variables that constitute Individual Network Usage. They are: Households w/ Internet access (%), Households w/ personal computer (%), Individuals using Internet (%), Mobile broadband subs/100 pop, Mobile phone subscriptions/100 pop, Use of virtual social networks (1-7). Their individual descriptions are as follows:

Households w/ Internet access (%): According to the WEF, the percentage of households with internet access is “calculated by dividing the number of in-scope households (where at least one household member is aged 15–74) with Internet access by the total number of in-scope households” (Baller *et al.*, 2016, p. 275).

Households w/ personal computer (%): According to the WEF, “the proportion of households with a computer is

calculated by dividing the number of households with a computer by the total number of households. A computer refers to a desktop or a laptop computer. It does not include equipment with some embedded computing abilities such as mobile cellular phones” (Baller, Dutta, Lanvin *et al.*, 2016, p. 275).

Individuals using Internet (%): According to the WEF, this represents the “proportion of individuals who used the Internet in the last 12 months” (Baller *et al.*, 2016, p. 274).

Mobile broadband subs/100 pop: According to the WEF, “Mobile broadband subscriptions refers to the sum of standard mobile broadband and dedicated mobile broadband subscriptions to the public Internet” (Baller *et al.*, 2016, p. 274).

Mobile phone subscriptions/100 pop: According to the WEF, “A mobile telephone subscription refers to a subscription to a public mobile telephone service that provides access to the Public Switched Telephone Network using cellular technology, including prepaid SIM cards active during the past three months. This includes both analog and digital cellular systems (IMT-2000, Third Generation, 3G) and 4G subscriptions, but excludes mobile broadband subscriptions via data cards or USB modems” (Baller *et al.*, 2016, p. 275).

Use of virtual social networks (1-7): The use of virtual social networks was determined through survey questioning, specifically asking “In your country, how widely are virtual social networks used (e.g., Facebook, Twitter, LinkedIn)? Where 1 = not at all used and 7 = used extensively” (Baller *et al.*, 2016, p. 275),

*Business Capacity for Innovation:*

According to the World Economic Forum, the Capacity for Innovation is calculated through the weighted average of surveyed questionnaire “In your country, to what extent do companies have the capacity to innovate? [1 = not at all; 7 = to a great extent]” (Baller *et al.*, 2016, p. 275).

**Data Analysis**

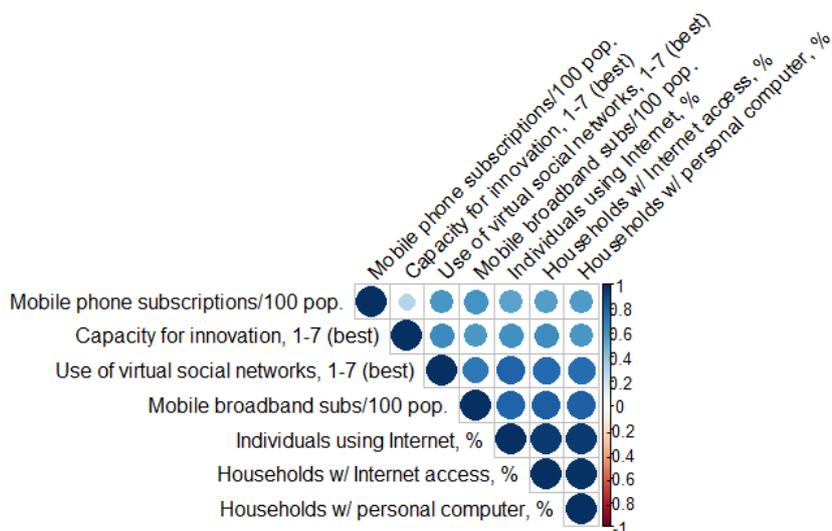
*Initial Model*

To reiterate, our main goal is to find a Nations Individual Network usage and its effects on a business’s capacity for innovation. In order to best determine the effects of a Nations Individual Network usage on a business’s capacity for innovation, we must create the most suitable supervised Machine Learning model through stepwise regression. Our initial model would produce the following:

$$\begin{aligned}
 & \text{(Capacity for innovation)} \\
 & = \beta_0 + \beta_1(\text{Use of virtual social networks}) \\
 & + \beta_2(\text{Households w/ personal computer, \%}) \\
 & + \beta_3(\text{Individuals using Internet, \%}) \\
 & + \beta_4(\text{Mobile broadband subs/100 pop.}) \\
 & + \beta_5(\text{Mobile phone subscriptions/100 pop.}) \\
 & + \beta_6(\text{Households w/ Internet access, \%})
 \end{aligned}$$

Such would give us a correlation plot which could be inferred from Fig. 1.

However, we could not help but notice that variable “Households with Internet access” is closely associated with many of the other variables such as “Individuals Using Internet” and “Households with personal computer”. This would severely affect the collinearity between variables and could jeopardize the validity of assumptions of Multiple Linear regression. Such is further confirmed by the VIF output of the above Multiple Linear Regression, evident from Fig. 2.



**Fig. 1:** Correlation Plot of initial model.

```
vif(innovation_model)
##           `Households w/ Internet access, %`
##                                     38.877634
##           `Households w/ personal computer, %`
##                                     41.329417
##           `Individuals using Internet, %`
##                                     15.285014
##           `Mobile broadband subs/100 pop.`
##                                     3.516182
##           `Mobile phone subscriptions/100 pop.`
##                                     1.790479
## `Use of virtual social networks, 1-7 (best)`
##                                     3.326839
```

Fig. 2: VIF output of initial model

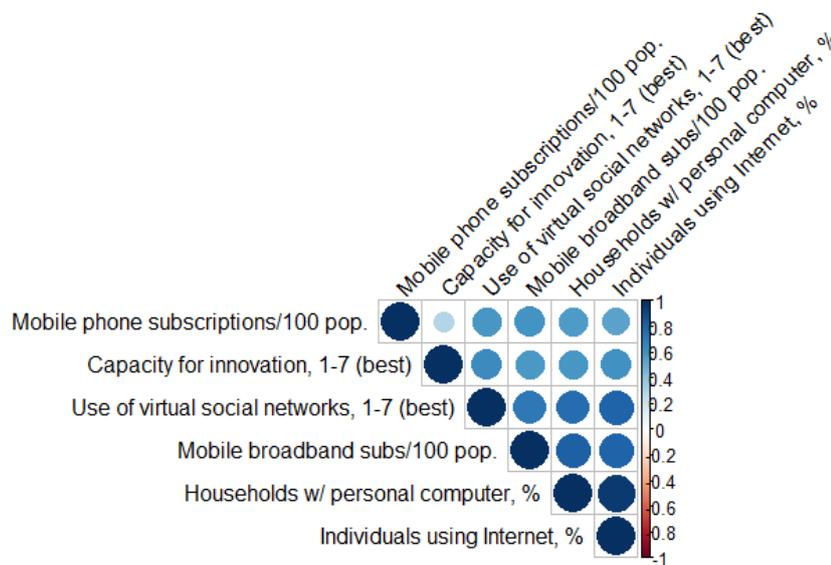


Fig. 3: Correlation plot of new suggested model.

It is at this note that we shall remove “Households w/ internet access” as a variable for our model, which would later undergo stepwise regression. Doing so also prevents the fallacy of data dredging. Therefore, our new suggested model would be as follows:

$$\begin{aligned}
 & \text{(Capacity for innovation)} \\
 & = \beta_0 + \beta_1(\text{Use of virtual social networks}) \\
 & + \beta_2(\text{Households w/ personal computer, \%}) \\
 & + \beta_3(\text{Individuals using Internet, \%}) \\
 & + \beta_4(\text{Mobile broadband subs/100 pop.}) \\
 & + \beta_5(\text{Mobile phone subscriptions/100 pop.})
 \end{aligned}$$

This model gives us a correlation plot which could be inferred from Fig. 3. It is evident from Fig. 3 that all the variables have a positive correlation with each other.

*Stepwise Model*

After undergoing bi-directional AIC Stepwise regression (results found in Results component), we obtained the following Stepwise Model:

$$\begin{aligned}
 & \text{(Capacity for innovation)} \\
 & = \beta_0 + \beta_1(\text{Use of virtual social networks}) \\
 & + \beta_4(\text{Mobile broadband subs/100 pop.}) \\
 & + \beta_5(\text{Mobile phone subscriptions/100 pop.})
 \end{aligned}$$

This model is based off bi-directional Stepwise regression from the earlier initial Models. This means that the variables (*Use of virtual social networks*), (*Mobile broadband subs/ 100 pop.*) and (*Mobile phone subscriptions/ 100 pop.*) are variables that allow us to obtain the best-fitting model.

Before presenting the summary of results of the Stepwise Regression Model, we shall test for assumptions of this model:

*Independence* – The assumption of independence would be met based off the fact that a country’s network usage does not and should not affect another’s country usage, justified through the collection and compilation of the defined data by the World Economic Forum

**Linearity** – The linearity assumption has been met as suggested by the residual plot (Fig. 4) which shows no fitted pattern.

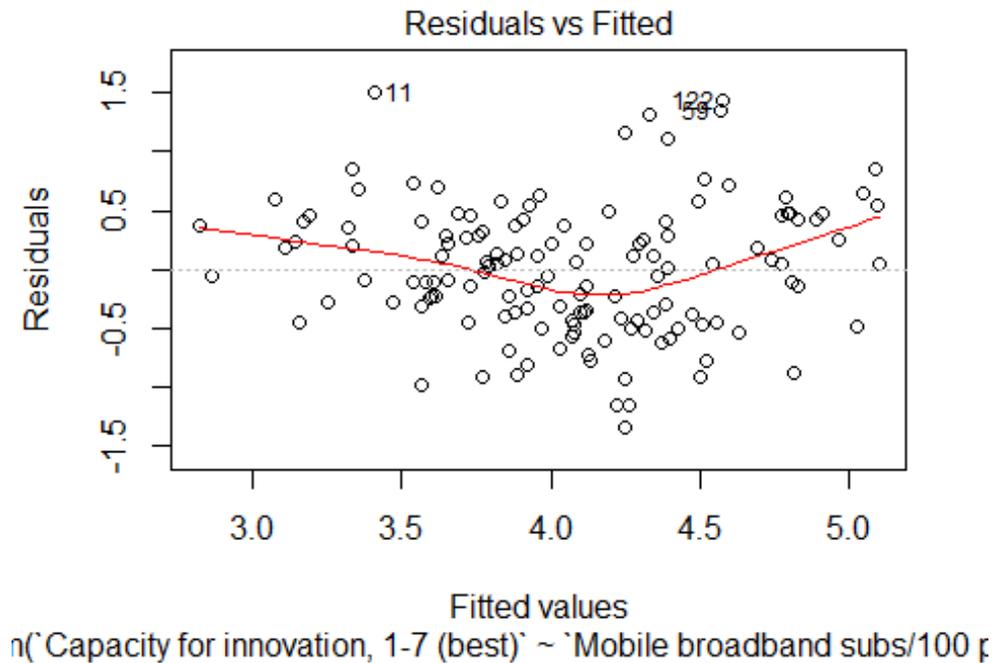
**Normality** – The assumption of normality has been met given the linear nature of the QQplot (Fig. 5) as well as the fact that the p-value of the Anderson-Darling normality test is more than 0.05 (Fig. 6).

**Homogeneity of Variance** – The assumption of equal variance is met given that residuals are spread equally along the ranges of predictors and we see a somewhat horizontal

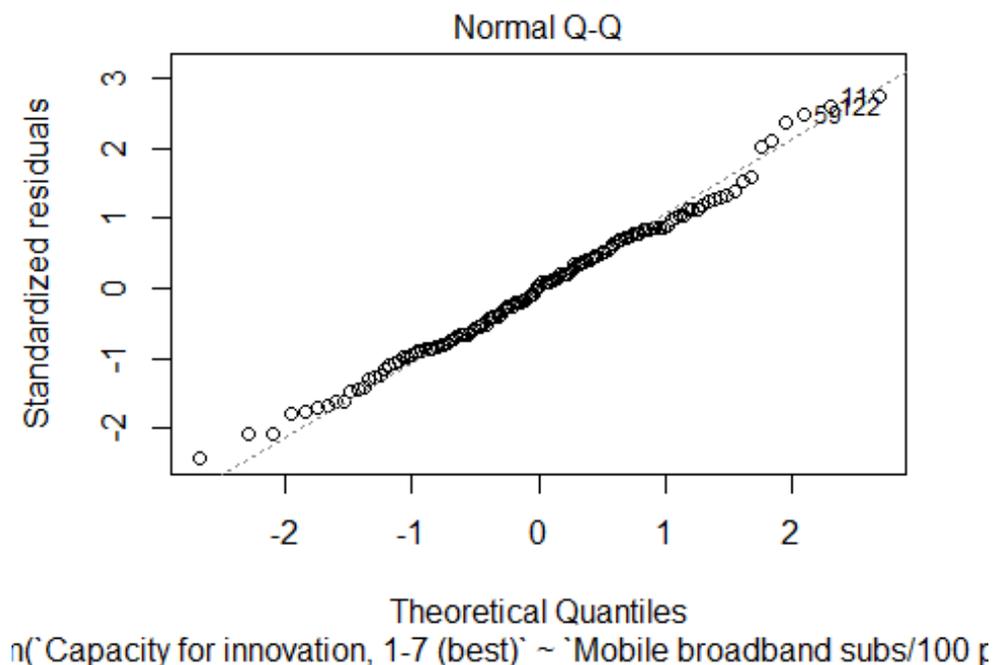
line with equally spread points in the scale-location plot (Fig. 7). This is further verified by the studentized Breusch-Pagan test which gives us a p-value greater than 0.05 (Fig. 8).

**Collinearity**– The assumptions of collinearity are met given that our VIF values reflected are all less than 3 (Fig. 9).

Now that the stepwise model has passed all the necessary assumptions for a multiple linear regression, we shall proceed with the results of our model.



**Fig. 4:** Residual vs Fitted Plot to test for linearity.



**Fig. 5:** QQplot to test for normality.

```
ad.test(studres(step.model2))  
  
##  
## Anderson-Darling normality test  
##  
## data: studres(step.model2)  
## A = 0.28312, p-value = 0.6293
```

Fig. 6: Anderson-Darling Test for normality.

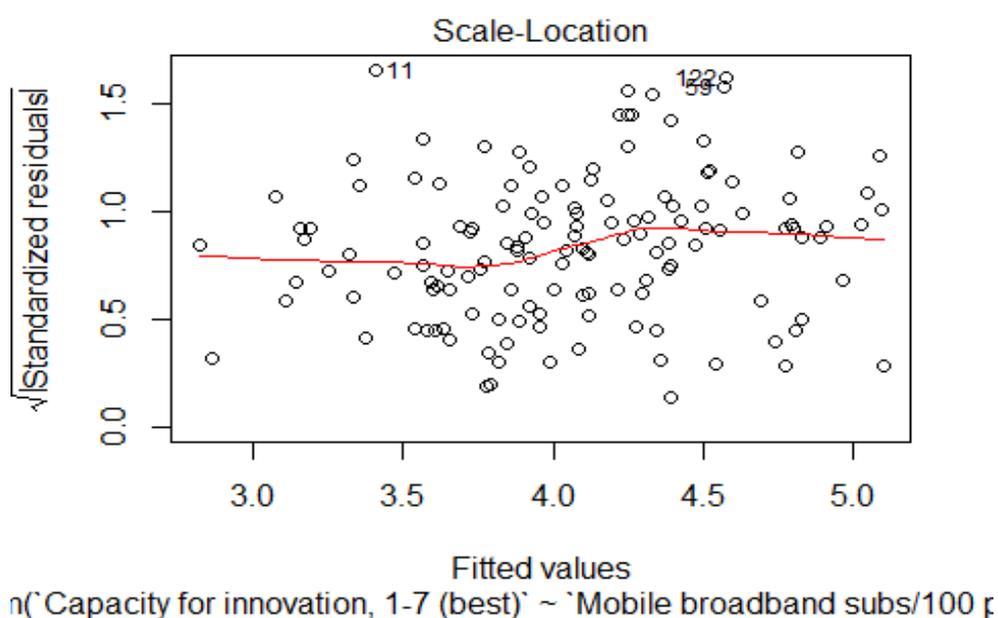


Fig. 7: Scale Location Plot.

```
bptest(step.model2)  
  
##  
## studentized Breusch-Pagan test  
##  
## data: step.model2  
## BP = 2.3881, df = 3, p-value = 0.4958
```

Fig. 8: Breusch-Pagan test results.

```
vif(step.model2)
##           `Mobile broadband subs/100 pop.`
##                               2.280004
##           `Mobile phone subscriptions/100 pop.`
##                               1.689539
## `Use of virtual social networks, 1-7 (best)`
##                               2.231945
```

Fig. 9: VIF output to test for Collinearity.

**Results**

For reiteration purposes, we aim to find how a nation’s Individual Network Usage, with variables from the 6<sup>th</sup> pillar, impact’s its business’s “Capacity for Innovation” index, found under the 7<sup>th</sup> pillar the Network Readiness Index Report.

As mentioned, our AIC bi-directional Stepwise Regressions obtains the following output, based off our initial model:

$$\begin{aligned}
 &(Capacity\ for\ innovation) \\
 &= \beta_0 + \beta_1(Use\ of\ virtual\ social\ networks) \\
 &+ \beta_4(Mobile\ broadband\ subs/100\ pop.) \\
 &+ \beta_5(Mobile\ phone\ subscriptions/100\ pop.)
 \end{aligned}$$

As such, our results are summarized as (also available in Table. 1).

Our results reflect that Use of virtual social networks, Rate of Mobile broadband subscriptions and Rate of Mobile phone subscriptions are significant factors within our best fit model in predicting a nations business’s capacity for

innovation. Our Adjusted R-squared value suggests that 44.46% of variation in a Nations Capacity for Innovation could be explained by its use of virtual social networks, its rate of Mobile broadband Subscriptions (per 100 population) and Mobile Phone subscriptions (per 100 population).

To interpret the intercept in context, based on our model, if a nation does not have any mobile broadband subscription and mobile phone subscriptions as well as no use of virtual social networks, the expected Capacity for Innovation is 1.27 on a scale of 1 to 7.

However, should a country have perfect use of virtual social networks, have every citizen subscribe to mobile phones and mobile broadband, the expected Capacity for innovation would be 5.35 on a scale of 1 to 7.

Such figures symbolize that our model is appropriate for it falls within the [1,7] possible range in terms of the “Capacity to Innovate” index.

Table 1: Summary of Results

Summary of results				
Coefficient	Estimate	t-value	Pr(> t )	Adj R-squared
$\beta_0$	1.272706	2.707	0.007682 **	0.4446
$\beta_1$	0.544855	5.422	2.66e-07 ***	
$\beta_4$	0.006968	3.453	0.000741 ***	
$\beta_5$	-0.004372	-2.512	0.013209 *	
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

### Discussions and limitations to the model

It is worth noting that Mobile related subscriptions do take significant precedence over computer related subscriptions, based on our determined model. Therefore, one could argue towards the vitality of mobile devices and its related network access at a national scale in warranting its Business innovation capacity.

It is through this finding where governments should be recommended to increase and promote their nations usage of mobile phone and mobile broadband subscriptions as well the utility of virtual social networks, should one wish to raise their business’s capacity for innovation. They could do so by making such services affordable and accessible in various ways. These include:

- the de-monopolization of related industries to increase competition and drive prices down, investing in mobile-related network infrastructure (Hassan, 2011, p. 8).
- ease restrictions on foreign investment to raise the accessibility of virtual social network usage (Hassan, 2011, p. 8).

Such could henceforth indirectly benefit their nations overall economy.

Businesses too could utilize such findings in playing a role towards raising an individual network usage, knowing that it could boost their very own capacity for innovation. They could introduce schemes such as:

- offering subsidized or corporate rates for mobile phone and mobile broadband subscriptions (Wolf,

2001, p. 6), particularly in areas with expensive mobile subscription rates.

- conduct courses in which its employees are taught how to set up and utilize virtual social networks (Lynch, Sandra, 1998, p. 64).

These policy suggestions are also summarized in Table. 2.

With regards to the limitation to the model, it is worth noting that the stepwise model’s adjusted R-squared of 44.46% is rather limited and not large. Therefore, should one conduct any future works on this model, it hopes to include other pillars of network usage such as government usage in the hopes of achieving a much larger R-square variation with similar outcomes.

### Conclusion

This study aimed to look at a Nations Individual Network usage and its effects on a business’s capacity for innovation. We utilized data from the 2016 World Economic Forum (WEF) Networked Readiness Index report and created the most suitable supervised Machine Learning model through stepwise regression. Our initial model included numerous variables in the 6<sup>th</sup> pillar of the Network Readiness report, which reflect a nations individual network usage, against the Capacity for innovation Index. Of the 6 initial prior variables in the 6<sup>th</sup> Pillar, stepwise regression outputs a best-fit and an appropriate model that contains predictors Use of virtual social networks, Rate of Mobile broadband subscriptions and Rate of Mobile phone subscriptions as significant factors in predicting a Business’s Capacity for innovation.

**Table 2:** Policy recommendations summary.

<u>Policy Recommendations</u>		
<i>Stakeholders</i>	<b>Government</b>	<b>Businesses</b>
<i>Recommended Policy</i>	the de-monopolization of related industries to increase competition and drive prices down, investing in mobile-related network infrastructure	offering subsidized or corporate rates for mobile phone and mobile broadband subscriptions – particularly in areas with expensive mobile subscription rates.
	ease restrictions on foreign investment to raise the accessibility of virtual social network usage	conduct courses in which its employees are taught how to set up and utilize virtual social networks

## Acknowledgement

I would like to acknowledge my former course instructor, Assistant Professor Dr. Helen Greatrex of Penn State's College of Earth & Mineral Sciences as well as Assistant Teaching Professor Dr. Marc Rigas of Penn State's College of Information Sciences & Technology for equipping me with the skills needed to complete this research.

## Conflict of Interest

Authors declare no conflict of interest with this study.

## References

- Baller S, Dutta S, Lanvin B, *et al.* (2016) The Global Information Technology Report 2016 *World Economic Forum* Retrieved from [http://www3.weforum.org/docs/GITR2016/WEF\\_GITR\\_Full\\_Report.pdf](http://www3.weforum.org/docs/GITR2016/WEF_GITR_Full_Report.pdf)
- Breene K (2016) What is 'networked readiness' and why does it matter? *World Economic Forum* Retrieved from <https://www.weforum.org/agenda/2016/07/what-is-networked-readiness-and-why-does-it-matter/>
- Ford P (2014) Freedom of Expression through Technological Networks: Accessing the Internet as a Fundamental Human Right. *Wis Int'l L J* **32**(142): 143–168. Retrieved from <https://heinonline.org/HOL/LandingPage?handle=hein.journals/wisint32&div=8&id=&page=>
- Grant RM (1996). *Toward a knowledge-based theory of the firm.* *Strategic management journal* **17**(S2): 109-122
- Hassan AO (2011). Telecommunications Reform and Effects of Competition on Availability, Quality and Cost of Services in Nigeria. *Public Policy and Administration Research* **1**(3): 8–20.
- Ibarra H (2017) Network Centrality, Power, and Innovation Involvement: Determinants of Technical and Administrative Roles. *Academy of Management Journal* **36**(3). DOI: <https://doi.org/10.5465/256589>
- Lynch LM & Black SE (1998). Beyond the Incidence of Employer-Provided Training. *Industrial and Labor Relations Review* **52**(1): 64–81. DOI: <https://doi.org/10.2307/2525243>
- Malisuwan S, Kaewphanuekrungsi W, Tiamnara, N & Suriyakrai N (2016) Thailand's Position in the Network Readiness Index (NRI): Analysis and Recommendations. *Journal of Economics, Business and Management*, **4**(5): 404–409. Retrieved from <http://www.ioebm.com/vol4/426-B040.pdf>
- Rothaermel FT & Hess AM (2007). Building Dynamic Capabilities: Innovation Driven by Individual-, Firm-, and Network-Level Effects. *Organization Science* **18**(6): 898–921. DOI: <https://doi.org/10.1287/orsc.1070.0291>
- Talukder, M. (2012). Factors affecting the adoption of technological innovation by individual employees: An Australian study. *Procedia - Social and Behavioral Sciences* **40**: 52–57. DOI: <https://doi.org/10.1016/j.sbspro.2012.03.160>
- Wolf G (2001) Mobile Internet Business Strategies, 1–36. Retrieved from [https://www.researchgate.net/profile/Gerrit\\_Wolf/publication/2394568\\_Mobile\\_Internet\\_Business\\_Strategies/links/00b49532f0cfd5b2a9000000/Mobile-Internet-Business-Strategies.pdf](https://www.researchgate.net/profile/Gerrit_Wolf/publication/2394568_Mobile_Internet_Business_Strategies/links/00b49532f0cfd5b2a9000000/Mobile-Internet-Business-Strategies.pdf)

## APPENDIX – R code

### downloading packages and reading data

```
# download libraries and packages
library(data.table)
library(leaps)

## Warning: package 'leaps' was built under R version 3.6.3

library(ggplot2)

## Registered S3 methods overwritten by 'ggplot2':
##   method      from
## [.quosures   rlang
## c.quosures   rlang
## print.quosures rlang

library(bit64)

## Loading required package: bit

## Attaching package bit

## package:bit (c) 2008-2012 Jens Oehlschlaegel (GPL-2)
## creators: bit bitwhich
## coercion: as.logical as.integer as.bit as.bitwhich which
## operator: ! & | xor != ==
## querying: print length any all min max range sum summary
## bit access: length<- [ [<- [[ [[<-
## for more help type ?bit

##
## Attaching package: 'bit'

## The following object is masked from 'package:data.table':
##
##   setattr

## The following object is masked from 'package:base':
##
##   xor

## Attaching package bit64

## package:bit64 (c) 2011-2012 Jens Oehlschlaegel
## creators: integer64 seq :
## coercion: as.integer64 as.vector as.logical as.integer as.double as.character as.bin
## logical operator: ! & | xor != == < <= >= >
## arithmetic operator: + - * / %% %%% ^
## math: sign abs sqrt log log2 log10
## math: floor ceiling trunc round
```

```
## querying: is.integer64 is.vector [is.atomic} [length] format print str
## values: is.na is.nan is.finite is.infinite
## aggregation: any all min max range sum prod
## cumulation: diff cummin cummax cumsum cumprod
## access: length<- [ [<- [[ [[<-
## combine: c rep cbind rbind as.data.frame
## WARNING don't use as subscripts
## WARNING semantics differ from integer
## for more help type ?bit64

##
## Attaching package: 'bit64'

## The following object is masked from 'package:bit':
##
##   still.identical

## The following objects are masked from 'package:base':
##
##   %in%, :, is.double, match, order, rank

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##   between, first, last

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(tidyverse)

## Registered S3 method overwritten by 'rvest':
##   method          from
##   read_xml.response xml2

## -- Attaching packages ----- tidyverse 1.2.1 --

## v tibble  2.1.1    v purrr   0.3.3
## v tidyr   0.8.3    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0

## Warning: package 'purrr' was built under R version 3.6.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::between() masks data.table::between()
## x dplyr::filter()  masks stats::filter()
## x dplyr::first()   masks data.table::first()
## x dplyr::lag()     masks stats::lag()
```

```
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()

library(BSDA)

## Warning: package 'BSDA' was built under R version 3.6.2
## Loading required package: lattice

##
## Attaching package: 'BSDA'

## The following object is masked from 'package:datasets':
##
##   Orange

library(rmarkdown)

## Warning: package 'rmarkdown' was built under R version 3.6.2

library(olsrr)

## Warning: package 'olsrr' was built under R version 3.6.1

##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##   rivers

library(lattice)
library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:olsrr':
##
##   cement

## The following object is masked from 'package:dplyr':
##
##   select

library(nortest)
library(lmtest)

## Warning: package 'lmtest' was built under R version 3.6.3
## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.6.1

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

library(MuMIn)

## Warning: package 'MuMIn' was built under R version 3.6.3
```

```
library(corrplot)

## Warning: package 'corrplot' was built under R version 3.6.3

## corrplot 0.84 loaded

library(StepReg)

## Warning: package 'StepReg' was built under R version 3.6.3

library(lmtest)
library(caret)

## Warning: package 'caret' was built under R version 3.6.1

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##   lift

library(car)

## Warning: package 'car' was built under R version 3.6.1

## Loading required package: carData

##
## Attaching package: 'carData'

## The following objects are masked from 'package:BSDA':
##
##   Vocab, Wool

##
## Attaching package: 'car'

## The following object is masked from 'package:purrr':
##
##   some

## The following object is masked from 'package:dplyr':
##
##   recode

# read the files
data<-fread("data.csv")
data<-data[complete.cases(data), ]
colnames(data) <- as.character(unlist(data[1,]))
data = data[-1, ]
data<-data[data$`Subindicator Type` != "Rank", ]
individualAndBusiness<-c("Mobile phone subscriptions/100 pop.", "Individuals using Internet, %", "Households w/ personal computer, %", "Households w/ Internet access, %", "Mobile broadband and subs/100 pop.", "Use of virtual social networks, 1-7 (best)", "Firm-level technology absorption, 1-7 (best)", "Capacity for innovation, 1-7 (best)")
data<-data[Indicator%in%individualAndBusiness]
data <- subset(data, select = -c(`Subindicator Type`))
data<-dcast(data=data, formula = `Country Name`~Indicator, value.var = "2016")
innovation_data<-subset(data, select = -c(`Firm-level technology absorption, 1-7 (best)`))
```

## initial model analysis

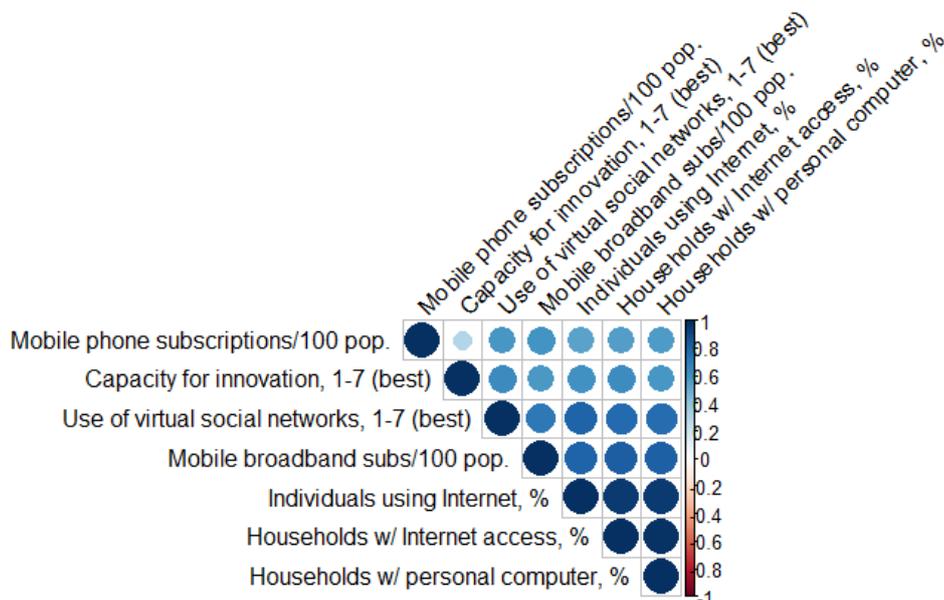
```

innovation_model<-lm(`Capacity for innovation, 1-7 (best)`~`Households w/ Internet access, %`
`+`Households w/ personal computer, %`+`Individuals using Internet, %`+`Mobile broadband subs/100 pop.`
+`Mobile phone subscriptions/100 pop.`+`Use of virtual social networks, 1-7 (best)`
), data=innovation_data)
vif(innovation_model)

##           `Households w/ Internet access, %`
##                               38.877634
##           `Households w/ personal computer, %`
##                               41.329417
##           `Individuals using Internet, %`
##                               15.285014
##           `Mobile broadband subs/100 pop.`
##                               3.516182
##           `Mobile phone subscriptions/100 pop.`
##                               1.790479
## `Use of virtual social networks, 1-7 (best)`
##                               3.326839

innovation_dataW0country<-subset(data, select = -c(`Country Name`, `Firm-level technology absorption, 1-7 (best)`))
res<-cor(innovation_dataW0country)
corrplot(res, type = "upper", order = "hclust",
          tl.col = "black", tl.srt = 45)

## Warning in corrplot(res, type = "upper", order = "hclust", tl.col =
## "black", : Not been able to calculate text margin, please try again with a
## clean new empty window using {plot.new(); dev.off()} or reduce tl.cex
    
```

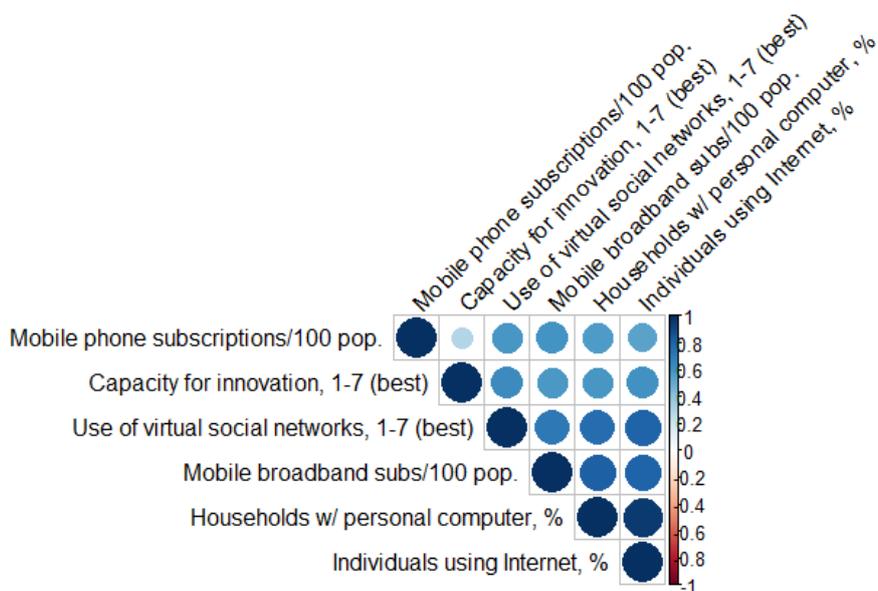


**remodelled corrplot w “Households w/ internet access” removed**

```

innovation_dataW0country<-subset(innovation_dataW0country, select = -c(`Households w/ Internet access, %`))
res<-cor(innovation_dataW0country)
corrplot(res, type = "upper", order = "hclust",
          tl.col = "black", tl.srt = 45)
    
```

```
## Warning in corrplot(res, type = "upper", order = "hclust", tl.col =
## "black", : Not been able to calculate text margin, please try again with a
## clean new empty window using {plot.new(); dev.off()} or reduce tl.cex
```

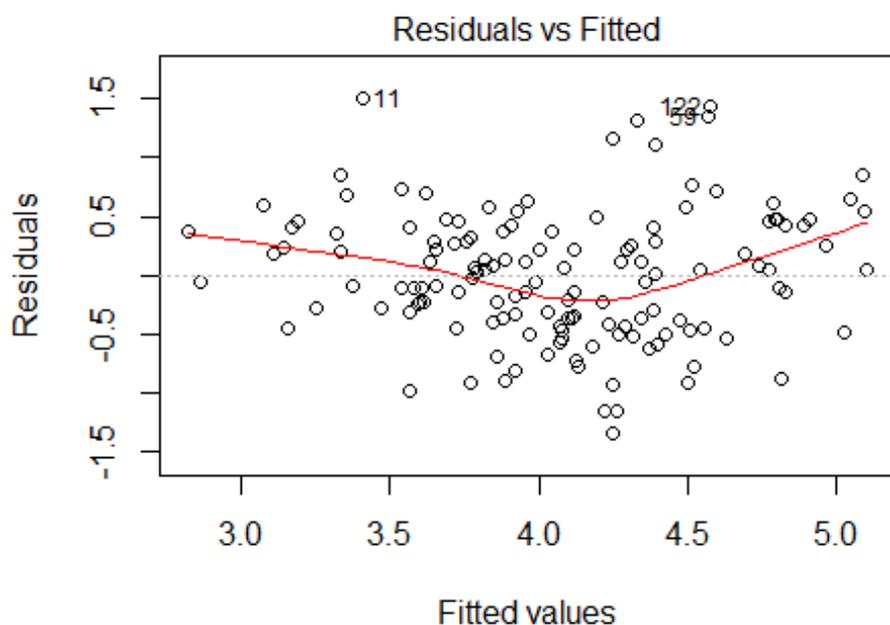


### re-modelled model and stepwise regression

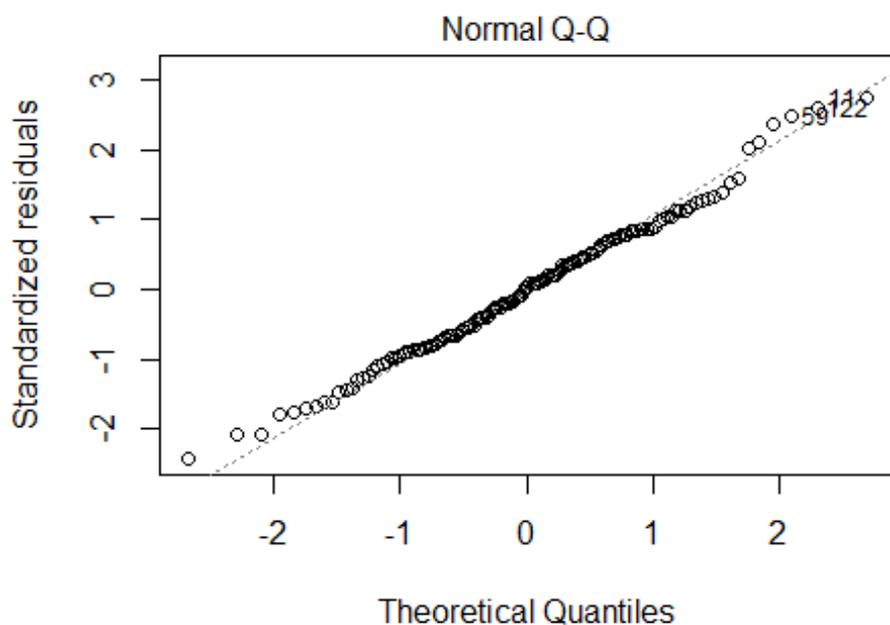
```
innovation_model2<-lm(`Capacity for innovation, 1-7 (best)`~`Households w/ personal computer
, %`+`Individuals using Internet, %`+`Mobile broadband subs/100 pop.`+`Mobile phone subscrip
tions/100 pop.`+`Use of virtual social networks, 1-7 (best)`, data=innovation_data)
step.model2 <- stepAIC(innovation_model2, direction = "both", trace = FALSE)
```

### testing for assumptions

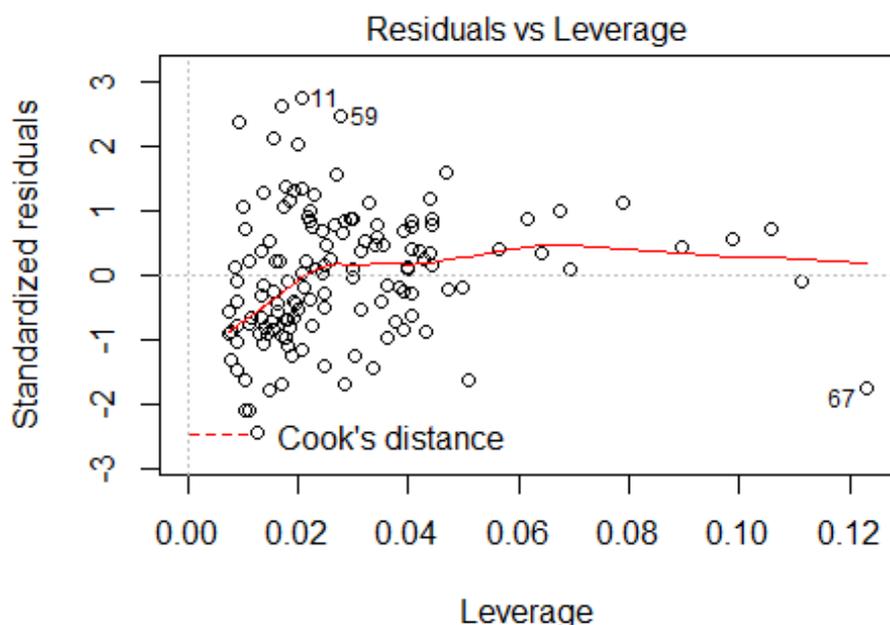
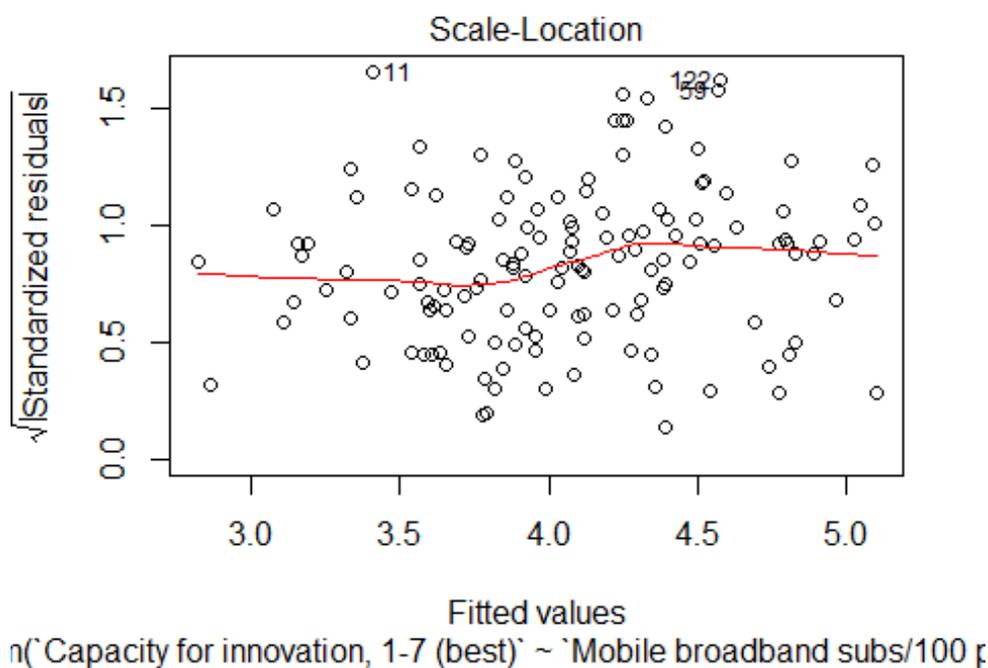
```
plot(step.model2)
```



n('Capacity for innovation, 1-7 (best)' ~ 'Mobile broadband subs/100 p



n('Capacity for innovation, 1-7 (best)' ~ 'Mobile broadband subs/100 p



```
ad.test(studres(step.model2))

##
## Anderson-Darling normality test
##
## data: studres(step.model2)
## A = 0.28312, p-value = 0.6293

bptest(step.model2)

##
## studentized Breusch-Pagan test
```

```
##
## data:  step.model2
## BP = 2.3881, df = 3, p-value = 0.4958

vif(step.model2)

##          `Mobile broadband subs/100 pop.`
##                2.280004
##          `Mobile phone subscriptions/100 pop.`
##                1.689539
## `Use of virtual social networks, 1-7 (best)`
##                2.231945
```

### summary of stepwise model results

```
summary(step.model2)

##
## Call:
## lm(formula = `Capacity for innovation, 1-7 (best)` ~ `Mobile broadband subs/100 pop.` +
##      `Mobile phone subscriptions/100 pop.` + `Use of virtual social networks, 1-7 (best)`,
##      data = innovation_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33225 -0.39372  0.01578  0.39409  1.49875
##
## Coefficients:
##
##              Estimate Std. Error t value
## (Intercept)      1.272706   0.470216   2.707
## `Mobile broadband subs/100 pop.`
##      0.006968   0.002018   3.453
## `Mobile phone subscriptions/100 pop.`
##     -0.004372   0.001741  -2.512
## `Use of virtual social networks, 1-7 (best)`
##      0.544855   0.100481   5.422
##
##              Pr(>|t|)
## (Intercept)      0.007682 **
## `Mobile broadband subs/100 pop.`
##      0.000741 ***
## `Mobile phone subscriptions/100 pop.`
##      0.013209 *
## `Use of virtual social networks, 1-7 (best)`
##      2.66e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5542 on 134 degrees of freedom
## Multiple R-squared:  0.4568, Adjusted R-squared:  0.4446
## F-statistic: 37.56 on 3 and 134 DF,  p-value: < 2.2e-16
```