ST332/ST409 Group Project

A report on characteristics associated with simian lines

Jamie Wreathall 1513854, Samuel Forbes 1629094, Thomas Newport 1514065, Priyal Soni 1511082, Claire Zheng 1506395

8 May 2018

Contents

1 Introduction

The study of palmar creases goes back to ancient Chinese and Indian civilisations as early as 3000 BC[5]. In palmistry, the two major transverse lines in the upper half of the palm are called the heart line and head line. When the heart line and head line are fused together, this is known as a simian line. This is presented as one major line across the hand from edge to edge[1] (see Fig 1). It is often referenced as a 'Proximal Transverse Palmar Crease' or 'Single Palmar Crease.' Since the work of Paul Broca (1877), the simian line has been used for studying variations among populations of different ethnicities[2].

Figure 1: Palmar crease terminology
Source: *https://jefffinley.org/simian-line-palmistry*

There are two other horizontal lines that transverse the full palm (Fig 2). The Sydney crease is known as the extended proximal palmar crease. The Suwon crease has been described as a "long heart line that crosses the entire palm appearing to join a head line at the radial edge [with] a second head line present"[3]. Park et al.[4] introduced the concept of the Suwon crease (named after *Suwon, Korea*, the location of the study) in 2010.

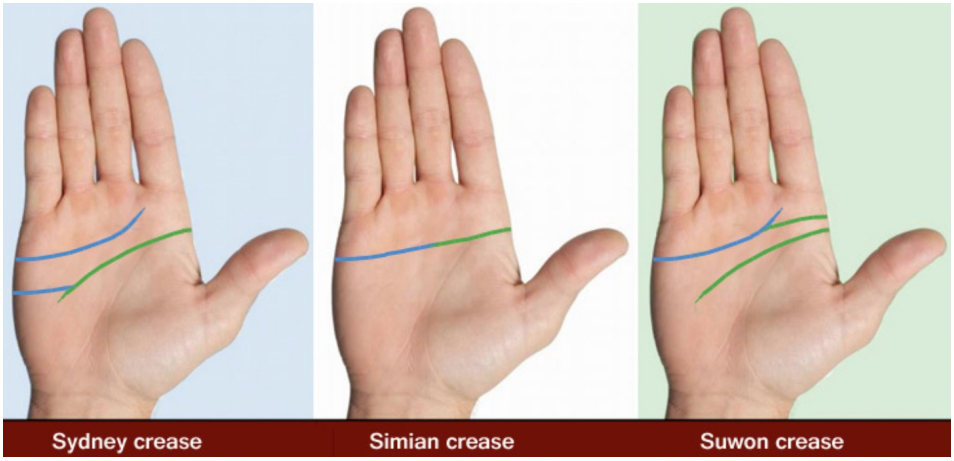


Figure 2: Other transverse palmar creases
Source: *simianline.handresearch.com/simianline/simian-line-how-to-recognize.htm*

In recent years, more medical attention has been drawn to simian lines due to the correlation between palmar creases and numerous chromosomal anomalies and conditions. Recent research has revealed that males are twice as likely as females to have simian lines, and it also tends to run in families. In its non-symptomatic form, it is more common among Asians and Native Americans than other populations[5]. Dar et al.[6] found that white newborns in the USA were significantly more likely to have simian lines than black newborns.

It must also be noted that palm creases do not always signify abnormalities[7,8]. However, they still are helpful in illustrating certain characteristics of populations of different ethnic origins. They are also easily analysable

in a practical sense since they can be observed without physical pain or monetary costs.

The aim of this project was to collect and analyse data with the interest of finding any links between simian lines and medical conditions, personality traits, and beliefs. This report will include a literature review, the data collection methodology, analysis of the data collected, limitations of the study and final conclusions.

2 Literature Review

Sharma and Sharma[5] provide an important paper which motivates the present report. The following reasons exemplify why this paper is important to our work:

1. Unlike many online resources related to simian lines, this report is not influenced by any religious or spiritual beliefs. In other words, there is no personal bias in the analysis of the findings of the research.
2. This report is based on studies carried out in populations of Central India. The results may be relevant to our survey as many of our participants were of Indian heritage.
3. This report not only look at simian lines, but also includes results related to Sydney and Suwon lines, which accounted for 36.6% combined of our survey population.

Sharma and Sharma found the proportion of people in Central India with simian, Sydney and Suwon creases to be 14.4%, 3.6% and 2.4% respectively. This simian line prevalence rate was compared to that of other ethnicities, and it was found that the figure for Central Indians was similar to that of Chinese, Koreans and Gypsies, but was much higher than that of Pygmies, Swiss and Germans (amongst others). Simian and Suwon creases were more common in males and Sydney creases more common in females.

The respondent base used by Sharma and Sharma was 1000 randomly selected Indian subjects aged between 5 and 70 and observations were then analysed by Fisher's exact test, with a significance level of 0.05. However, this paper only provided limited insight into the association between simian lines and medical conditions, as the main findings were geographically restricted. Therefore, it is necessary to look at parallel investigations into other population groups.

Hidayah, Tjong and Roesma's paper[9] on dermatoglyphics with respect to simian lines in Indonesia found increased frequency of whorl-type fingerprints and other dermatoglyphic characteristics in the simian crease group. The study comprised 30 cases (simian crease group) and 30 controls (normal group). Purposive sampling was used for the simian crease group and random sampling was used for the control group. Apart from the difference in fingerprint patterns, the simian crease group also showed a significantly higher IQ score than the control group; these results are shown in the table below. Although this study revealed some interesting results and could potentially lead to further medical research on the relationship between fingerprint patterns and generic characteristics, the results of this study cannot necessarily be generalised to populations of different ethnicities.

Group	Total	Ratio	Average
Normal	30	63.5-141.5	106.78
Palmar Crease	30	93.5-150.5	113.72

Furthermore, Down's syndrome is perhaps the most well-known medical condition linked to the presence of a simian line. This link dates back to R.L. Down in 1906. For example, 61.1% of patients with Down's syndrome in south India have a single palmar crease[10]. We also investigated this relationship in our study, but none of the survey respondents reported having Down's syndrome. There have also been some suggested links between some psychiatric conditions such as schizophrenia and simian lines[11], although there isn't substantial evidence provided.

3 Data Collection & Methodology

3.1 How data was collected

We decided to collect data via an online survey, together with several other groups within the module. This was for several reasons:

1. Due to the low prevalence of simian lines, either a large group of people needed to be reached or specific populations needed to be contacted.
2. If a specific population (e.g a simian lines Facebook group) was contacted by multiple different groups, response rates were likely to be split across the different survey. Therefore, it seemed logical to approach such groups as a collective to obtain the maximum amount of data.
3. We were being assessed on our analysis and discussion, so it seemed sensible to work with other groups on the data collection part of the assignment.

We had many meetings to discuss the types of questions each group wanted to include in the survey and split up the work to ensure each group was researching and collating information. Consequently, we decided on a final survey and sent this out to other students, family etc. We also sent a copy of the survey to a simian lines Facebook group, the results of which were stored separately to those of the main survey.

It was hard to coordinate with multiple groups and decide on a final survey in a timely manner. For this reason, we decided to create a Facebook group for all members of the larger data collection group as we found it to be an easier way to discuss the survey, as opposed to trying to decide a time to meet or emailing.

Given the need to collect a large quantity of data and the necessity to ensure anonymity, we created an anonymous online survey using Google Docs. Following discussion between the groups, the survey was split into 5 sections, focusing on the presence of a simian line, the individual's background, personality, faith and beliefs, and medical information.

When deciding on the questionnaire, one of the key issues was finding an agreeable method to assess the respondent's personality. Our initial thought was to employ the Myers-Briggs questionnaire, consisting of 44 questions. However, given that one of the priorities was the time spent answering the survey, we decided on employing the Ten-Item Personality Inventory (TIPI). This questionnaire includes 10 questions - 2 questions relating to each one of the 'big five' personality traits. The big five comprises 'Extraversion, Agreeableness, Conscientiousness, Emotional stability and Openness to Experience.' Although the TIPI is less comprehensive than the Big Five Personality Test upon which it is based, it has been shown that it is adequate if a quick overview of personality is needed[12]. Each of the 10 questions consists of two descriptors, representing each of the domain personality traits. Each statement starts with the phrase "I see myself as" and was ranked on a 7-point scale from 1 (disagree strongly) to 7 (agree strongly). Consequently, a single score for each of the 'big five' personality traits can be created by combining the two scales — we will discuss this further in the Formal Testing section.

The questionnaire was advertised to a large number of people through social media and word of mouth, encouraging participants to refer others where possible. The individual participants were not specifically chosen, which allowed the questionnaire to have anonymous data. The number of required participants was also not specified and instead, we set a deadline of the end of week 10 (Term 2) for surveys to be completed.

Unfortunately, during data collection, a significant amount of time was wasted due to the indecisiveness of groups when creating the survey. Most group members did not voice their opinions on question options to put in the survey which led to some delays in finalising the survey. Although a few individuals took the initiative to arrange the meetings/contact other group members, the majority of group members failed to take more responsibilities which created some inefficiencies in the process. These inefficiencies and delays could have been avoided if each group had an assigned 'group leader' responsible for communicating with other groups and supervising group members to be up to speed with assigned work on the project. Ideally, we could have spent more time on reaching out to more individuals and collecting more data rather than creating the survey itself.

Advantages of Google Docs

There are several advantages to using Google Docs for this project.

1. Anonymity of responses.
2. Easily able to distribute the link to the survey on any social media platform, around the world.
3. The data can be exported as a .csv file, without the need to manually transfer data to a spreadsheet, thus eliminating the possibility of errors.
4. There can be multiple people allowed to edit and access the results of the survey at any time.
5. Possibility of using a wide range of question types. For example, drop down lists or text boxes.

Disadvantages of Google Docs

However, we must also consider that there are some weaknesses of using a Google Docs form.

1. We are not asking for responses in person so respondents could make up data or lie.
2. The form automatically categorises responses, which causes repeated categories. For example, if people start the word with a capital letter, ('Maths & Stats' or 'maths and stats'), they get put into a different categories. This prolongs the data cleaning time.
3. The way in which questions are worded and consequently interpreted has a large impact on the corresponding answers.
4. The Google Doc does not have an advanced output analysis package.
5. The website only allowed us to put images, and not videos, in the survey. This may have restricted people's understanding of simian line concepts.

3.2 Which data was collected

It was decided that we would collect data on a number of individual attributes that are listed below:

- Presence of simian line
- Dominant hand
- Age
- Gender
- Sexual orientation
- Ethnicity
- Occupation
- Education level
- Degree course area
- Faith
- Medical conditions
- Siblings
- Personality (Ten item personality measure)

A full copy of the survey and brief explanatory notes can be found in the appendix.

3.3 Possible sources of bias

Non-respondents' bias: Our data is based on the participants who completed the questionnaire only. We do not possess any data for those who either didn't submit the survey or those who never attempted it, either because of the length of time to complete it or because they weren't interested in doing so. This can lead to bias as those not sampled may portray different characteristics to those sampled.

Diagnostic suspicion bias: Given that the survey was titled about simian lines, potential respondents may have been more likely to complete it if they had a simian line.

Fake data: Given that this was an anonymous online questionnaire, respondents may make up data. This can lead to inaccurate data in the sample. In order to try and solve this, anomalous results are investigated during the 'data cleaning' part of the analysis.

Misclassification bias: Inaccurate diagnoses tend to reduce the apparent strengths of relationships.

Socially non-acceptable bias: If the respondent didn't want to label themselves in a way that might be seen as negative (e.g. too extraverted), they could have chosen a lower score, while increasing the score for traits with positive associations.

Unacceptable disease bias: Socially unacceptable diseases are often under-reported. As this report asked personal questions about a variety of illnesses some respondents may not have responded accurately.

Cognitive bias (Motivational bias): Some participants may hold certain (positive) beliefs about simian lines which makes them more likely to self diagnose their hands with simian lines when they do not have them.

4 Preprocessing and Demographics of Data

The Google Docs survey was completed by 418 people at the time of starting our analysis (we could have waited for longer but the response frequency was slowing and we felt we had a good sample size) and a .csv file was automatically generated with the results. A row of the file represented a particular person's questionnaire response and a column represented the responses of all 418 people for a particular variable asked in the questionnaire. Responses to the question of degree subject were categorised into Science, Arts, Humanities, Languages, Miscellaneous and none. The same was done with the responses for ethnicity which were categorised into Asian, White, Mixed, Black and Hispanic. Any obvious erroneous data was removed for example a response of 42 older siblings was removed. This subject's other responses were kept however, as they seemed genuine. Whether to keep this subject's other responses or not was a cause for debate within our group but fortunately data of this nature was small enough to be insignificant.

For each of the five personality variables (extraversion, agreeableness, conscientiousness, emotional stability, openness to new experiences), two relevant questions were asked in order to measure that personality trait. For example, questions 1 ("I see myself as extraverted, enthusiastic") and 6 ("I see myself as reserved, quiet") were asked on the topic of extraversion. The response to question 6 was subtracted from the response to question 1 to give an overall score for extraversion. The same process was done for the other personality variables to give a score ranging from -6 (does not display the personality trait) to 6 (strongly displays with the personality trait).

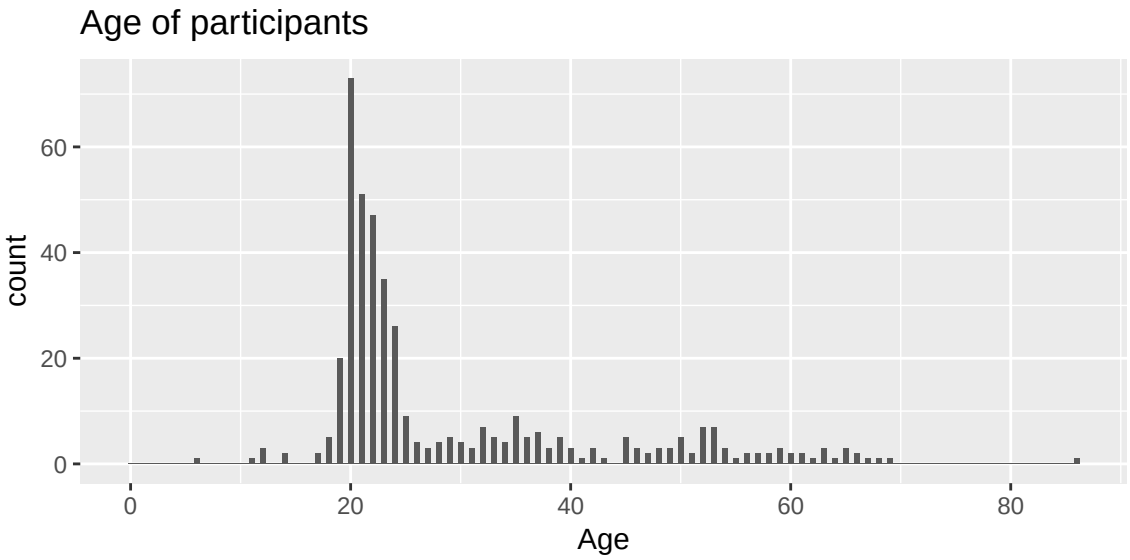
More than half of the participants (around 56%) were female, with one participant identifying as nonbinary and the rest as male. A mean and median age of 28.9 and 23 respectively were calculated with the youngest age being 6 and the oldest age 86. A cluster of people between the ages of 18 and 24 were observed, see histogram below, comprising roughly 55% of the sample; this is unsurprising, as many of the respondents were students. Other than this main cluster, the ages were fairly evenly spread. In terms of ethnicity, most respondents were White (67%) or Asian (29%). Around 90% of respondents had 2 older siblings or less, though one respondent had 8 older siblings.

In general, the majority of people identified themselves as more agreeable, conscientious, open to new experiences and emotionally stable than not. However, ratings for extraversion tended to be fairly neutral. Roughly 44% of participants were spiritual or believed in God, but a smaller proportion of respondents (around 22%) said they often engaged in religious practice.

34 medical conditions were reported with the most common conditions being anxiety (reported by 17%), allergies (16%), vision problems (12%), depression (11%), insomnia (6%), and anaemia and respiratory problems (both 5%).

12% of subjects had a simian line on their left hand and 14% had a simian line on their right, with 15% having a simian line on at least one hand. 18% of males had a simian line, compared to 13% of females. 18%

of Asian respondents had at least one simian line compared to 13% of Whites. There was a higher incidence of simian lines than expected, as 1% of Whites and up to 14% of Asians are expected to have simian lines[3].



5 Initial Analysis

Throughout this analysis, our main variable of interest was the presence of a simian line on either hand (i.e. an indicator that was 1 if a participant selected option D on either the right or left hand question, and 0 otherwise). We began by considering the contingency tables between the simian line indicator and certain demographic variables. We wanted to identify existing relationships between any demographic variables and simian lines in order to control for these associations in formal analysis later. For example, previous studies suggest that there is a relationship between ethnicity and simian line prevalence[5]. It is also the case that people of different ethnicities vary in religious beliefs and practice. So when testing for the effect of having a simian line on religious beliefs, it is sensible to control for ethnicity. This initial analysis allowed us to identify other variables that could be related to having a simian line and hence have a confounding effect on this study's results.

A Chi-squared test for statistical independence was used on these contingency tables to determine if there was a relationship between simian lines and the other variable. The Chi-squared statistic for a contingency table with r rows and c columns is defined as

$$\mathcal{X}^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

where O_{ij} is the observed number in cell (i,j), $E_{ij} = Np_ip_j$ is the expected number in cell $(i,\ j)$, $p_i = \dfrac{O_{i\cdot}}{N} = \sum_{j=1}^{c}\dfrac{O_{ij}}{N}$ is the proportion in row i, $p_j = \dfrac{O_{\cdot j}}{N} = \sum_{i=1}^{r}\dfrac{O_{ij}}{N}$ is the proportion in column j and N is the total number in the survey (the sum of all the cells). Under the assumption of independence of the variables $\mathcal{X}^2 \sim \mathcal{X}^2_{n-p}$, where n is the number of cells and $p = r + c - 1$.

Two demographic variables were found to have a significant relationship with having a simian line on either hand:

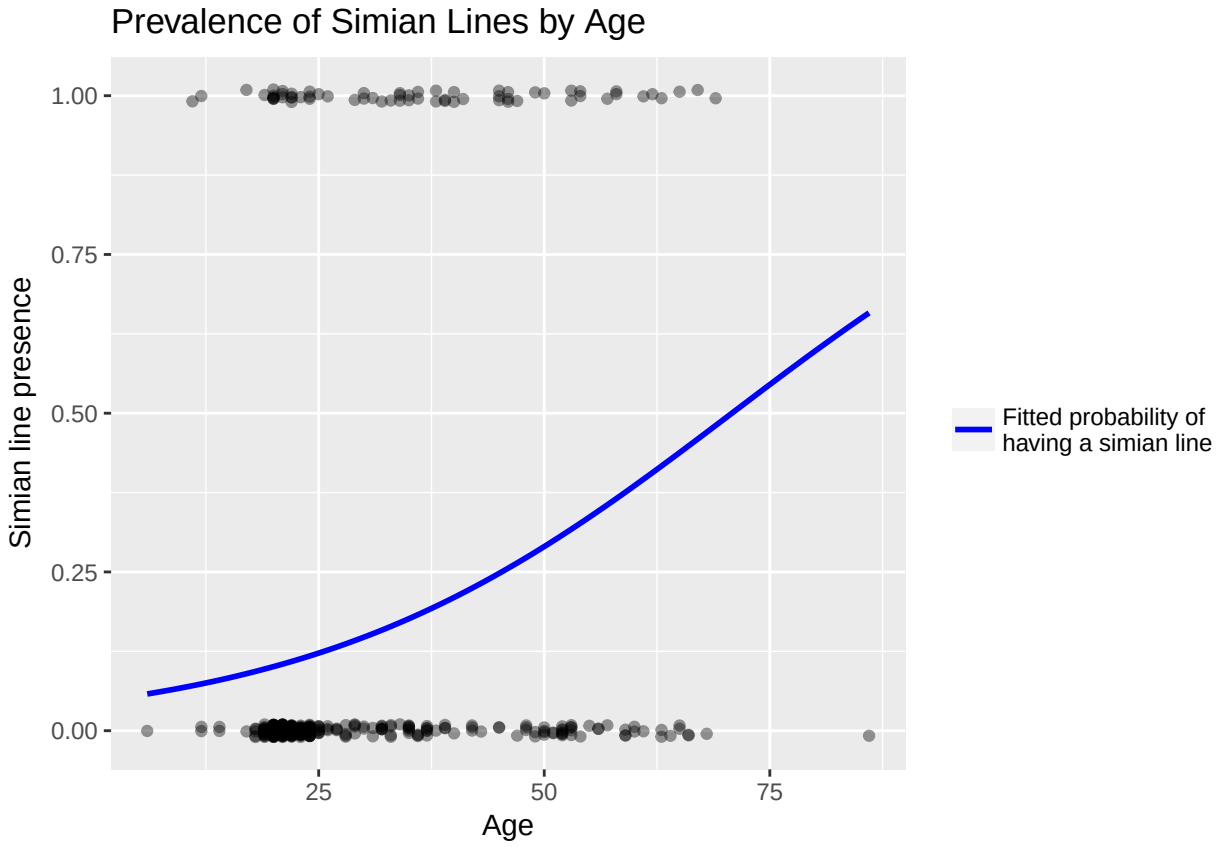
- Number of older siblings (p = 0.004953). Participants with more older siblings reported having a simian line more frequently. The contingency table for these two variables is below, along with the R output for the chi-squared test.

Table 1: Number of older siblings versus simian line on either hand

	0	1	2	3	4	5	6	8
No simian line	170	112	45	17	6	2	0	1
Simian line	22	20	9	5	5	2	1	0

```
##
##  Pearson's Chi-squared test
##
## data:  Simian_vs_older_siblings
## X-squared = 20.302, df = 7, p-value = 0.004953
```

- Age ($p = 0.00000001656$). In general, older participants were more likely to report having a simian line. The contingency table for this variable has been omitted due to the high number of different ages in the sample. Instead, the data is plotted below, with a logistic regression line overlayed for the sake of illustration. Individual points are jittered slightly to avoid overlap.



It should be noted that chi-squared testing may be unreliable when column totals are small (for example, in a contingency table including age, there is only one person aged 6, one aged 11, etc.). However, these results were still considered to be important enough that we wanted to control for these variables in the next part of the study.

6 Formal Testing & Results

To test the effect of simian lines on predicting the belief/personality/medical variables, we fitted a logistic regression model for each variable. In each model we used the simian line indicator as a predictor, along

with the demographic variables that we wanted to control for: ethnicity, gender, age, and number of older siblings. The latter two were included in the models due to the results of our initial analysis, while ethnicity and gender have been found in previous studies to have a relationship with simian lines. For each model, a Wald test was automatically performed (with the R glm function) on the hypothesis that the coefficient was 0 for each coefficient of the predictors. Under the hypothesis $H_0 : \beta_i = 0$ where β_i is the coefficient of predictor i, the Wald statistic is

$$W = \frac{\hat{\beta}_i}{\text{se}(\hat{\beta}_i)} \sim \mathcal{N}(0, 1)$$

where $\hat{\beta}_i$ is the maximum likelihood estimator for β_i and $\text{se}(\hat{\beta}_i)$ is the standard error for $\hat{\beta}_i$. W is used to determine the p-value and significance of β for a two sided test. Using this test we could see if a simian line was a significant predictor for the variable of interest.

In total, we fitted 15 models. The variables of interest were the seven most common medical conditions (i.e. those reported by at least 15 participants), the three belief variables, and the five personality variables. To counteract data dredging bias, we applied the Bonferroni correction to our results: a result was only considered to be significant if it had a p-value of less than $\frac{0.05}{15} = 0.00\dot{3}$. Taking the standard significance level (5%) and dividing it by the number of tests being done (in our case, 15) ensures that, under the various null hypotheses, the expected overall false positive rate of the study remains at 5%. The downside of this method is that it reduces the power of the tests and hence increases the false negative rate.

6.1 Medical Conditions

In the model fitted to predict the presence of insomnia, for example, this gives a logistic regression model along the lines of:

$$\mathbb{P}(\text{Insomnia}) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 \text{Age} + \beta_2 \text{Ethnicity} + \beta_3 \text{Gender} + \beta_4 \text{Older.Siblings} + \beta_5 \text{Simian.Either} + \varepsilon))}$$

where:

- `Age` is the individual's age in years
- `Ethnicity` is an indicator for the the individual being of a given ethnicity.

  Note: In our model, there are actually several ethnicity terms — each being an indicator for a different ethnicity, and each one having a different coefficient.
- `Gender` is an indicator for the individual's gender

  Again, there was more than one term for gender in our model, accounting for participants who identified as nonbinary genders.
- `Older.Siblings` is the number of older siblings the individual has
- `Simian.Either` is an indicator that is 1 if the individual has a simian line on either hand, and 0 otherwise.
- ε is the random error distributed by the standard logistic distribution (as per the assumptions of logistic regression)

```
##
## Call:
## glm(formula = Insomnia ~ Age + Ethnicity + Gender + Older.Siblings +
##     Simian.Either, family = binomial(link = "logit"), data = palms)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -0.8588  -0.3280  -0.2911  -0.2113   2.7813
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -2.956e+00  6.363e-01  -4.646 3.38e-06 ***
## Age                  -7.997e-03  1.823e-02  -0.439 0.660888
## EthnicityAsian;White -1.526e+01  1.518e+03  -0.010 0.991983
## EthnicityBlack        5.690e-01  1.226e+00   0.464 0.642482
## EthnicityHispanic    -1.524e+01  3.956e+03  -0.004 0.996927
## EthnicityWhite        1.148e-01  5.008e-01   0.229 0.818701
## EthnicityWhite;Black -1.464e+01  2.797e+03  -0.005 0.995823
## GenderM              -7.893e-01  4.852e-01  -1.627 0.103760
## GenderNB              7.957e-01  5.595e+03   0.000 0.999887
## Older.Siblings        1.377e-01  1.598e-01   0.862 0.388831
## Simian.Either         1.673e+00  4.966e-01   3.369 0.000755 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 177.77  on 414  degrees of freedom
## Residual deviance: 161.09  on 404  degrees of freedom
##   (3 observations deleted due to missingness)
## AIC: 183.09
##
## Number of Fisher Scoring iterations: 16
```

So in this model, the presence of a simian line was a significant predictor for insomnia (at the aforementioned $0.\dot{3}\%$ level). We performed the same test for the other common medical conditions (anxiety, allergies, vision problems, depression, anaemia, and respiratory problems), but the presence of a simian line was not significant to the required $0.\dot{3}\%$ level in any of these models. It did have a p-value of less than 0.05 (but higher than 0.003) in both the anxiety and the depression models, but this is not considered to be a significant result after applying the Bonferroni correction to the significance criterion.

6.2 Belief Variables

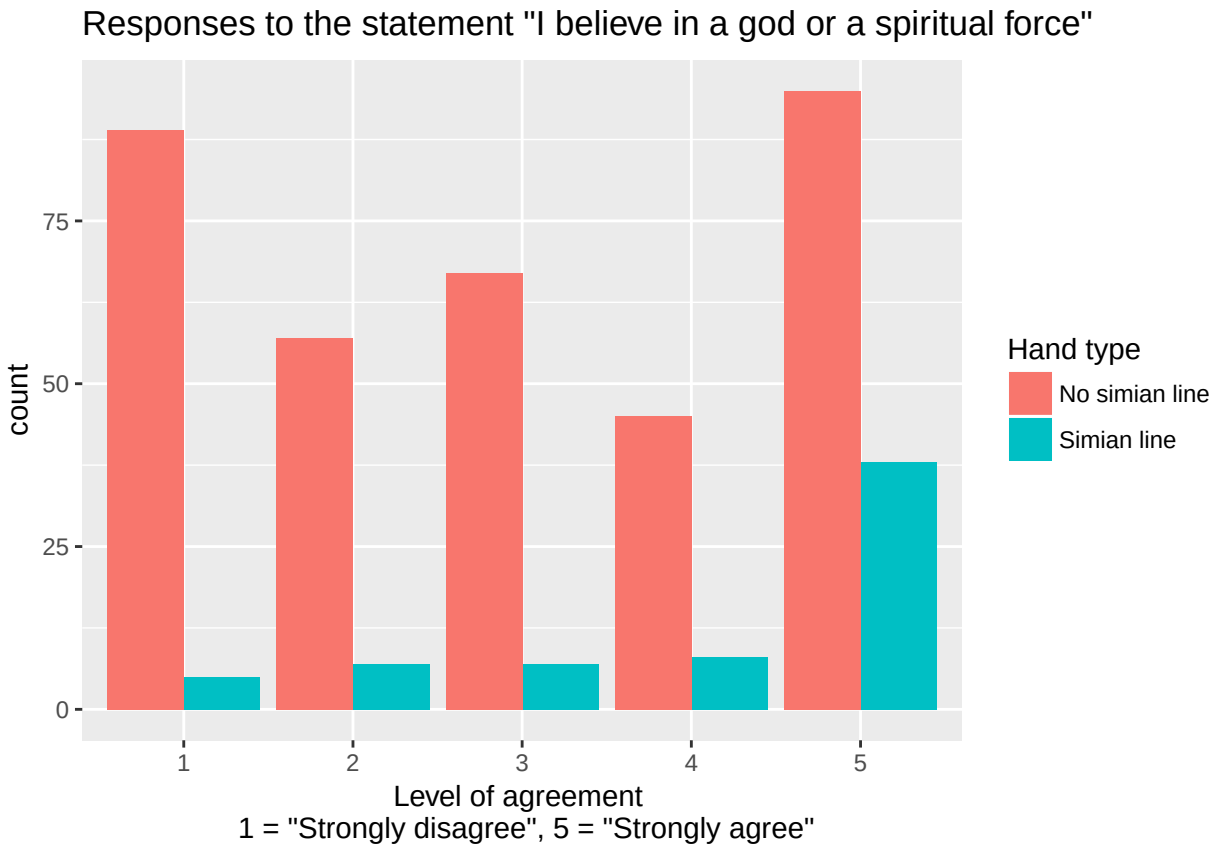
The belief variables were all collected as 5-point Likert scales, and so ordinal regression would likely be the best choice for fitting an accurate model. However, it is unclear how to test for the significance of a specific variable within an ordinal regression model, so we instead coerced the data into a binary form and used logistic regression. To do this, we took responses of 4 or 5 as "Agree" and 1–3 as "Do not agree".

Note: We also considered using multinomial regression but this gives a different p-value for each different category of the response, whereas we wanted to have a single coefficient to interpret in order to test the effect of having a simian line on the response variables.

For the question "I believe there is a God or a spiritual force", for example, we took "Slightly agree" and "Strongly agree" to be a single "Belief" indicator. In this model, the presence of a simian line was significant. Below are the R output for this logistic regression model and a bar chart showing the high level of religious belief amongst participants with simian lines.

```
##
## Call:
## glm(formula = Belief.BIN ~ Age + Ethnicity + Gender + Older.Siblings +
##     Simian.Either, family = binomial(link = "logit"), data = palms)
##
```

```
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9243  -0.8296  -0.7258   1.0108   1.7244
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            9.615e-02  3.114e-01   0.309  0.75749
## Age                    1.034e-02  8.779e-03   1.178  0.23879
## EthnicityAsian;White   8.362e-01  1.140e+00   0.734  0.46304
## EthnicityBlack         9.402e-01  1.121e+00   0.839  0.40161
## EthnicityHispanic     -1.710e+01  1.455e+03  -0.012  0.99063
## EthnicityWhite        -1.513e+00  2.459e-01  -6.152 7.64e-10 ***
## EthnicityWhite;Black -1.625e+01  1.026e+03  -0.016  0.98736
## GenderM                2.887e-02  2.231e-01   0.129  0.89704
## GenderNB               1.221e+00  2.058e+03   0.001  0.99953
## Older.Siblings         2.391e-01  9.912e-02   2.413  0.01584 *
## Simian.Either          1.079e+00  3.295e-01   3.273  0.00106 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 570.42  on 414  degrees of freedom
## Residual deviance: 487.49  on 404  degrees of freedom
##   (3 observations deleted due to missingness)
## AIC: 509.49
##
## Number of Fisher Scoring iterations: 14
```



It is unsurprising that in this model and both of the other belief models (i.e. the models for responses to the statements "My belief is important to me" and "I often engage in religious practice"), ethnicity was the most significant predictor. However, in the latter two belief models, the presence of a simian line was not a significant predictor.

6.3 Personality Variables

As the personality data was also ordinal, we again converted the data to binary. For example, a score of greater than zero on the Extraversion scale was taken to mean "Extraverted" and less than or equal to zero was "Not extraverted". However, the presence of a simian line was not a significant predictor in this model, nor was it significant in any of the other personality models (i.e. Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experiences). The simian line predictor did have a p-value of 0.034 in the Conscientiousness model, but as mentioned before, this result is not considered to be significant after applying the Bonferroni correction.

In general, age and gender were the most commonly significant predictors in the personality variables' models.

7 Study Limitations

7.1 Sampling method

We worked with several other groups to send the survey to a large number of people. Some sampling biases were discussed in the "Sources of Bias" section above, with non-respondents' bias and diagnostic suspicion bias in particular being weaknesses of handing a survey out to many people that may or may not complete it for a variety of reasons. This could possibly affect the results of the study.

Another weakness of this sampling method (and the fact that the survey was anonymous) is that we don't know who the survey was sent to; if our group had handed the survey out, we would have a better idea of who the participants were, and whether or not they were likely to fill out the survey honestly.

7.2 Sample

A large proportion of the sample were students or the family and friends of students. Also, the vast majority of participants were Asian or White. Due to the sample being unrepresentative of the population of the country (or world) as a whole, the results of the study cannot be generalised to these larger populations.

Another limitation of the study with regards to the sample was that the number of participants with simian lines, while higher than expected, was still too low in some cases for hypothesis tests to be significant. This was also the case with many of the medical conditions; when only a handful of people had the condition, it was hard to make any inferences about the relationship between that condition and simian lines. In order to properly investigate the relationship between rarer medical conditions and simian lines, participants would probably have to be found via hospitals or support groups in order to get a large enough sample size.

7.3 Survey questions

One limitation of the survey questions was that the ethnicity question was far too broad with its categories. In particular, "Asian" encompasses a multitude of different ethnicities. In the literature review we found that the prevalence of simian lines varies a lot even within India alone, so it is clear that grouping all Asian participants together into one ethnicity does not capture the vast differences between participants of Asian heritage. This could have affected the study as, again using the "Asian" ethnicity example, different cultures in Asia have different religions and levels of religiosity. So if participants of certain ethnicities are more likely to both have a simian line and be very religious, then this could explain why it seems like there is a strong link between simian lines and religious belief, even after we have attempted to control for ethnicity.

The personality measure used (i.e. TIPI) is a way of briefly capturing some aspects of participants' personalities. This brief measure was chosen so that participants would not become bored while filling out the survey and hence not participate. However, it may be the case that personality cannot be accurately assessed with so few questions, which would reduce the applicability of this study's research into the links between personality and simian lines.

7.4 Survey format

Self-reported data can often be very unreliable. In particular, participants in this study may have been bad at identifying the lines on their palm (or simply misunderstood the question). The observed prevalence of simian lines in this study is far higher than we expected, which could be due to misclassification by participants. This could have been reduced by interviewing participants in person or having them submit a picture of their hand, so that a researcher could classify them more accurately. Another aspect of self-reporting that may have been inaccurate is the personality data. Participants were asked to rate how they see themselves, but a person's perception of themself can be vastly different to how others perceive them.

Also, while the survey being anonymous was necessary for ethical reasons, and may have reduced unacceptable disease bias (as participants' data wasn't traceable back to them), it could have allowed participants to lie without fear of consequences.

7.5 Hypothesis testing

In the hypothesis tests for belief and personality variables, ordinal data was treated as binary for ease of analysis. However, this does not fully capture the data, and more sophisticated methods may have

been more appropriate (such as multinomial or ordinal regression, or different tests altogether). Another potential weakness of the logistic regression models is that the interaction between different variables was not included; it is likely that factors such as age, gender and ethnicity often combine to have complicated effects on health or personality. Also, it is possible that other variables that were missed by our initial analysis (or weren't collected in this study at all) could have had a confounding effect on both simian lines and medical/personality/belief data.

While the Bonferroni correction allowed us to apply multiple hypothesis tests without increasing the false positive rate, the power of each individual test was consequently very low. This means that our tests may have produced false negatives by having a very low significance criterion (alpha).

8 Conclusions

A survey with the aim of investigating the relationship between simian lines and personality, religion, and medical conditions was filled out by 418 participants. Our initial testing and literature review found a number of demographic variables that are linked to the presence of simian lines. These variables were controlled for when fitting logistic regression models. In these models, the presence of a simian line was found to be a significant predictor of both insomnia and belief in a god or spiritual force. Results were only taken to be significant if they had a p-value of less than $0.00\dot{3}$ in order to overcome the multiple comparisons problem. It is also noted that the coefficient for the simian line predictor had a p-value of less than 0.05 (but higher than $0.00\dot{3}$) in the models for anxiety, depression, conscientiousness and whether or not their degree subject was in the Arts. This could warrant further study into these relationships, but is not taken to be a significant result in this study.

9 References

[1] Martijn van Mensvoort. "SIMIAN LINE - How to recognize" http://simianline.handresearch.com/simianline/simian-line-how-to-recognize.htm. Accessed on 10/04/18.

[2] Bali, R. S. (1994). Anthropology of crease morphogenesis: a scientific analysis. New Dehli: Concept Publishing Company, pp. 20 - 21.

[3] Paul, J. N., Osunwoke, E. A. and Paul, C. W. (2016). Prevalence of simian, Suwon, and Sydney creases in acquired idiopathic blindness in some selected schools for the blind in Nigeria. Annals of Bioanthropology, 4 (2), pp. 75 - 78.

[4] Park, J. S., Shin, D. S., Jung, W. and Chung, M. S. (2010). Improved analysis of palm creases. Anatomy & cell biology, 43 (2), pp. 169 - 177.

[5] Sharma, D. K. and Sharma, V. (2011). Prevalences of Simian, Sydney and Suwon creases and their association with each other, body sides, handedness, sex and anomalies/diseases/syndromes in a population of Central india. International Journal of Morphology, 29 (3), pp. 1069 - 1075.

[6] Dar, H., Schmidt, R. and Nitowsky, H.M., 1977. Palmar crease variants and their clinical significance: a study of newborns at risk. Pediatric research, 11 (2), pp. 103 - 108.

[7] Hernández, M. (1985). Palmar creases in Spaniards. Anthropologischer Anzeiger, 43 (2), pp.187-190.

[8] De Lestrange, M.T. (1969). The transverse crease in Europe: Index and comparative study of different samples cited in the literature. American journal of physical anthropology, 30 (2), pp.173 - 182.

[9] Hidayah, M., Tjong, D. H. and Roesma, D. I. (2016). The Study of Dermatoglyphic in Simian Crease Group (The Human Masukake-Gata) at Minangkabau Ethnic, West Sumatra, Indonesia. Int. J. Pure App. Biosci, 4 (1), pp. 9 - 14.

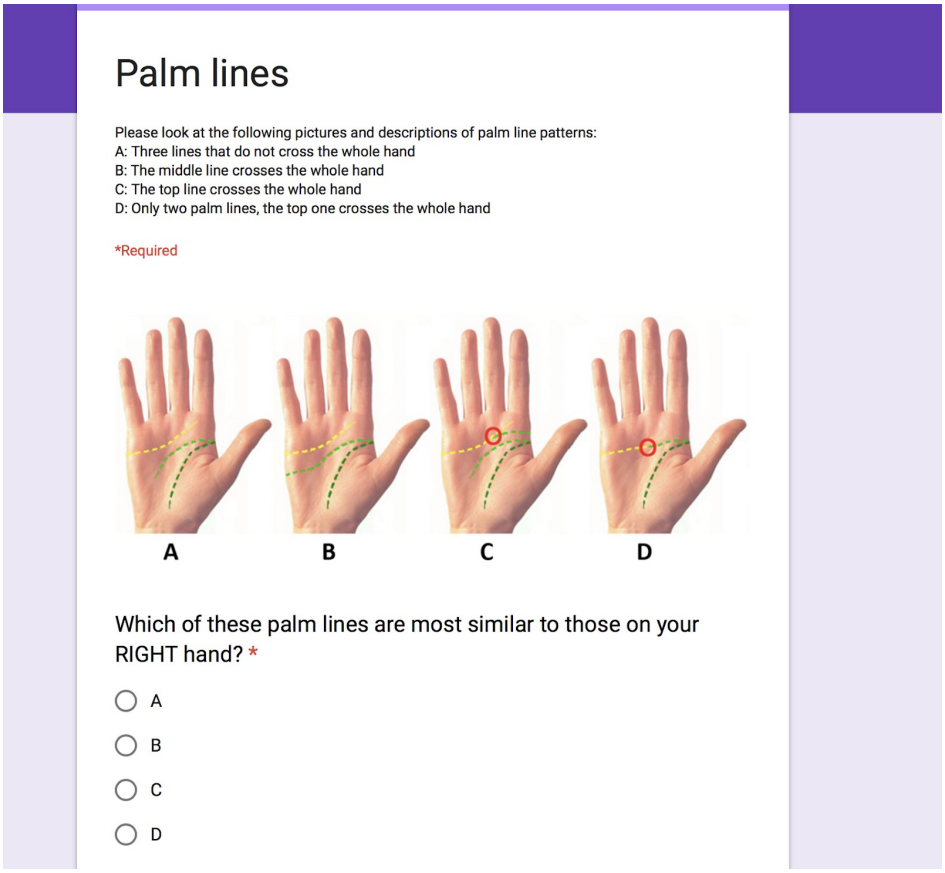
[10] Sureshbabu, R., Kumari, R., Ranugha, S., Sathyamoorthy, R., Udayashankar, C. and Oudeacoumar, P., 2011. Phenotypic and dermatological manifestations in Down Syndrome. Dermatology online journal, 17 (2).

[11] Huang, C. J., Chiu, H. J., Lan, T. H., Wang, H. F., Kuo, S. W., Chen, S. F., Yu, Y. H., Liu, Y. R., Hu, T. M. and Loh, E. W. (2010). Significance of morphological features in schizophrenia of a Chinese population. Journal of psychiatric research, 44 (2), pp. 63 - 68.

[12] Gosling, S.D., Rentfrow, P.J. and Swann Jr, W.B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in personality, 37 (6), pp. 504 - 528.

10 Appendix

10.1 Google Docs survey screenshot



10.2 Survey questions

SECTION 1

Which of these palm lines are most similar to those on your right hand?
The image above was included in the questionnaire to illustrate the patterns and this was a compulsory question. The question was implemented using a multiple choice ratio field with the 4 letters being the options.

Which of these palm lines are most similar to those on your left hand?
The image above was included in the questionnaire to illustrate the patterns and this was a compulsory question. The question was implemented using a multiple choice ratio field with the 4 letters being the options.

Which is your dominant hand?
This was a compulsory question and was implemented using a multiple choice ratio field with the options being:

- Left hand
- Right hand
- Other

'Other' was implemented due to for example, ambidexterity.

SECTION 2

What is your age?
This question was implemented using a short-answer text box and was a compulsory question.

What is your gender?
This was implemented using a multiple choice ratio field with the options being "Male", "Female", "Prefer not to say". As well as an additional option of "Other" to input your own test. This was also a compulsory question.

Which of the following terms best describes your sexual orientation?
This question was implemented using a multiple choice ratio field with the options:

- Heterosexual
- Homosexual
- Bisexual
- Prefer not to say
- In order to remain inclusive, we also included an "Other" field.

What is your ethnicity? If you have mixed ethnicity, tick all that apply.
This question was integrated using a list with the options being:

- Black or African American
- Asian/ Asian British
- White
- Indian/ Indian British
- Arab/ Arab British In order to account for other possibilities, an "Other" field was also included. This was a compulsory question.

What is your occupation?
This question was implemented using a short-answer text box and was a compulsory question.

What is your education level?
This question was integrated using a multiple choice ratio field with the options being as follows:

- Degree or equivalent
- Higher education
- A level or equivalent
- GCSEs grades A-C or equivalent
- Other qualifications
- No qualification
- Don't know

If you went to university, what subject did/do you study? If you did not go to university, please leave this blank.
This question was implemented using a short-answer text field, with the option to leave it blank if applicable.

SECTION 3

In this section, all of the statements below were implemented. Each statement had to individually be ranked from 1 to 5, with 1 being "Strongly disagree" and 5 being "Strongly agree." This was not a compulsory question since it is not viable to assume that all potential respondents would like to disclose their beliefs about their faith.

- I believe there is a God or a spiritual force
- My belief (including atheism) is important to me
- I often engage in religious practice

SECTION 4

Do you have any of these conditions? Tick all that apply
This question was integrated using a list with the options being

- Diabetes (Type I)
- Diabetes (Type II)
- Cardiovascular disease
- Anaemia
- Respiratory problems
- Vision problems
- Hearing problems
- Allergies
- Autism
- Anxiety
- Depression
- Insomnia
- Substance abuse
- Down's syndrome
- Fetal alcohol syndrome
- Prefer not to say
- None

In order to account for other possibilities, "Other" field was also included.

How many older siblings do you have?
This question was implemented using a short-answer text box. Additionally, this was a compulsory question.

How many older siblings of the same sex do you have?
This question was implemented using a short-answer text box and was also a compulsory question. This question was included at the request of another group, who believed that the number of siblings a participant had that were of the same sex as them might be of interest. However, due to potential ambiguity in the wording of the question, we chose not to analyse this data.

SECTION 5

This section consists of the TIPI framework. Each statement needed to be ranked on a 7-point scale ranging from "Disagree strongly" being 1, to "Agree strongly" being 7.

- I see myself as extroverted, enthusiastic.
- I see myself as critical, quarrelsome.
- I see myself as dependable, self-disciplined.
- I see myself as anxious, easily upset.
- I see myself as open to new experiences, complex.
- I see myself as reserved, quiet.
- I see myself as sympathetic, warm.
- I see myself as disorganised, careless.
- I see myself as calm, emotionally stable.
- I see myself as conventional, uncreative.

10.3 Group meetings information

Date	Topics of Discussion/Achievements	Members Attended
5/02/2018	Methods of data collection, population of study, job allocation within group	All
12/02/2018	Collaboration on data collection with 5 other groups, what to include in the survey and allocation of survey sections	Group members of all 6 groups
16/02/2018	Finalising the 'personality' section of the survey	Priyal, Claire, Tom, Sam
21/02/2018	Finalising the survey and setting deadline for the data collection	Representatives from each of the 6 groups
05/03/2018	Started sending out survey	
09/03/2018	Discussion on the data collected and job allocation	All
23/03/2018	Updates on the progress of each section of the report	Priyal, Jamie, Sam, Claire
30/03/2018	Finalising the report and discussion of data analysis	Priyal, Jamie, Sam, Claire

10.4 Technical Appendix

The R code that was used to produce the graphs and statistical analysis in this report can be found in a technical appendix in the online submission of this report. The R packages used were `dplyr`, `ggplot2` and `knitr`. Further information on these packages can be found on their respective CRAN pages:

https://cran.r-project.org/web/packages/dplyr/index.html
https://cran.r-project.org/web/packages/ggplot2/index.html
https://cran.r-project.org/web/packages/knitr/index.html