

The Heritability of Facial Attractiveness

Alicia Purkey

Advisor: Dr. Matthew C. Keller

Department of Psychology and Neuroscience

University of Colorado at Boulder

Undergraduate Honors Thesis

Committee Members:

Dr. Matthew C. Keller: Department of Psychology and Neuroscience

Dr. Soo Rhee: Department of Psychology and Neuroscience

Dr. Christine MacDonald: Program for Writing and Rhetoric

Abstract

Sexual selection theories state that mate choice is dependent upon maximizing fitness. To understand sexual selection and fitness, we must understand the inheritance of sexually selected traits. The current study aims to estimate the heritability of facial attractiveness. 1599 twin subjects from the LTS and QIMR twin registries were rated on Facial Attractiveness by 8 raters. Facial Attractiveness appears to be predominately controlled by additive genetic variation and unique environmental effects. In the best fitting model, AE model, we found that A accounted for 65% of the phenotypic variation in Facial Attractiveness and E the remaining 35%. There was no sex-limitation found in Facial Attractiveness, showing that heritability for Facial Attractiveness is not more important for females or males.

## The Heritability of Facial Attractiveness

Attractiveness is a trait that is universally valued. It is the basis of our initial judgments about others around us and continuing appraisals about those we are familiar with (Langlois et al., 2000). Solomon Asch (1946) posited that our impression of people is quick and easy to develop. The sum of the traits we observe gives us an idea of a general impression of a person. Attractiveness is an overarching trait that informs our judgments about an individual according to Asch—he called it a central trait. People who are considered to be attractive are often thought to have other favorable qualities. Dion, Berscheid, and Walster (1972) explored how we unconsciously correlate beauty/attractiveness with positive traits. Subjects rated attractive faces to be more likely to have good jobs, marry earlier and have happy marriages, and better lives. This supports the attractiveness stereotype, or the ‘halo effect’, which assumes that more attractive individuals will have more desirable and advantageous lives. Inferences about traits from a person’s appearance can lead to advantages or disadvantages in social situations, such as in a job interview or a first date. J. C. Langlois et al. (2000) showed that individuals that are attractive actually do score higher on various measures (such as physical health, self-confidence and intelligence) than their less attractive counterparts. This is a source of dispute. Attractive individuals could score higher because they were treated better and had better opportunities or, on the other hand, attractive

individuals could be inherently better at certain things due to shared genetic contribution (the genes that control attractiveness could control other beneficial traits).

Evolutionary theories on attractiveness assume that attractiveness is not only an honest signal, but is also heritable. According to several evolutionary theories (Zahavi, 1975; Thornhill and Gangestad, 1993; Rowe and Houle, 1996; Scheib et al., 1999), physical appearance serves as a signal to potential mates to inform them about the genetic condition of an individual. The observation that an attractive individual inherits other advantageous traits can be explained by evolutionary theories. Fisher's 'sexy sons hypothesis' says that females inherit preferences for certain traits that are heritable in males (Cornwell and Perrett, 2008). The traits that are under sexual selection do not confer fitness under the sexy sons hypothesis, they are just traits that are desirable to the opposite sex. The 'good genes' selection theory states that individuals evolve preferences for mates who have traits that will increase reproductive success—passing good genes to offspring and providing a good environment for the offspring. The so-called 'good genes' that are passed to offspring could be those that increase fitness for parasite avoidance. The parasite-avoidance hypothesis and immunocompetence hypotheses assume that ancestors were faced with pathogens that may have caused developmental stability. Those who survived are said to have an advantage, passing immunocompetence to offspring. Attractiveness can be a signal of health (Kalick et al., 1998).

Attractiveness has been under a great deal of selection pressure (Scheib et al, 1999). The 'lek paradox' (Borgia, 1979; Kirkpatrick and Ryan, 1991; Andersson 1994) states that the persistent female choice for certain traits in males should erode the variation in males, yet the variation still exists. A possible solution to this paradox is that the expression of sexually selected traits depends on the correlation between the trait and the condition (Rowe and Houle, 1996). The condition refers to a number of genes or loci affecting physiology, such as nutrition or muscle mass, which are influenced by the environment (e.g. food availability or season). Sexually selected traits will maintain large genetic variation if there is a similarly large genetic variation in the condition. A further explanation for why sexually selected traits have high variance is that they are controlled by multiple loci, which increased the probability for mutation within the trait (Houle, 1998). Paired with selection for the most advantageous mutations, both mutation and selection provide means for a large genetic variation in fitness traits. These selected traits then become signals to the other sex that the possessor of the trait has beneficial genetic contributions.

Once Attractiveness was recognized as having a genetic component, the desire to investigate the genetic variance and heritability of the trait arose. The interest in quantifying the importance of the genetics involved in a number of traits led researchers to the so-called "twin method." Though early scientists, such as Sir Francis Galton, showed interest in studying twins to gain insight on genetic transmission, they didn't possess the tools necessary to glean heritability estimates from their research. Galton studied twins in 1875 in his paper "The History of Twins." He reasoned that twins could be used to study the effects of heredity and

environment. At that time, the knowledge or tools to assess the similarity of the genes between twins were not available. In his books *Hereditary Genius* and *English Men of Science*, Galton investigated anecdotal evidence from the lives of twins. He was famous for his use of questionnaires, which he used to try to parse apart the contribution of nature and nurture, or heredity and environment. Galton explored questions such as how similar twins were in simple traits such as preferences and physical characteristics and how prevalent twins were in families. While Galton didn't know about fraternal and identical twins, he made the observation that same-sex twins could be remarkably similar and also remarkably different. He concluded that nature, the genetic component, had a larger effect than nurture due to the similarity of twins along with their shared genes (Burbridge, 2001). Though Galton did contribute to statistics by providing correlation methods, he did not come up with the twin analysis method where MZ and DZ twins' phenotypic scores are correlated to find heritability.

Hermann W. Siemens and Curtis Merriman are credited with publishing the first twin studies and implementing heritability calculations that were undiscovered by the scientists before them. Siemens and Merriman both published twin studies in the same year, 1924. It was Siemens that proposed comparing correlations between monozygotic (MZ) and dizygotic (DZ) twins (Siemens, 1924). Merriman made the distinction between MZ and DZ twins, offering a calculation of the heritability coefficient:  $2(r_{MZ} - r_{DZ})$  (Merriman, 1924). The heritability coefficient gives the total possible variance in a trait due to genetic variance.

There are two types of heritability, broad-sense  $H^2$  and narrow-sense  $h^2$  (Feldman and Lewis, 1975). Both  $H^2$  and  $h^2$  contribute to the population phenotypic variance. Narrow-sense heritability is the additive genetic variance over the phenotypic variance ( $h^2 = \frac{V_A}{V_P}$ ). Broad-sense heritability is the all the genetic variation (additive, dominance, epistatic variation) over the phenotypic variance ( $H^2 = \frac{V_G}{V_P}$ ). Narrow -sense heritability contributes directly to the correlation or resemblance between siblings or parents while broad-sense heritability makes MZ twins more similar.

A pioneering effort in twin research was published by John Loehlin and R. C. Nichols established assumptions/conclusions about twins. An unstated conclusion in the Loehlin and Nichols research is that the correlation for MZ twins is always observed to be less than 1. This shows that there is no purely genetically controlled trait or behavior. MZ twins, however, tend to be more alike than DZ twin; though environment always has an influence, everything also has a genetic component (Loehlin and Nichols, 1976; Johnson et al, 2009). They also put forth what is now called the 'Loehlin and Nichols Hypothesis', which states that there is little difference in heritability between sexes and that most male and female heritabilities are moderate (Loehlin and Nichols, 1976). This finding has also been reproduced by Eric Turkheimer, one of Loehlin's graduate students (Johnson et al, 2009). A final and important finding was that the differences between MZ and DZ early

environments are not highly predictive of differences observed in the individual twins. This crucial finding is termed the 'equal environments assumption' (EEA).

Genetic analysis is traditionally separated into additive (A), non-additive (D), common environment (C), and unique environment (E) separate components. Additive genetic variation (A) arises from the sum of alleles at the loci that influence a trait. Non-additive genetic variation emerges from interactions between alleles at the same locus. Shared environment (C) includes the common environmental conditions shared by twins that make them more similar (ie. living in the same household, having the same parents, attending the same schools). Unique environment (E) accounts for the unshared experiences between twins that result in making them less similar (ie. going to a different school, having different friends). Because MZ twins share 100% of their genes and DZ twins share 50% of their genes, the magnitude of the difference between MZ and DZ twins estimates the total genetic variance ( $H^2 = A + D$ ). The estimate for E includes error and real environmental effects. The error in the unique environment variance comes from measurement error. It is difficult to discern whether the differences between MZ twins are due to simple measurement error or if it is the result of unique environmental influences.

The correlation between MZ twins is the narrow-sense heritability and shared environment ( $r_{MZ} = h^2 + c^2$ ). The correlation between DZ twins is the contributions of half of the narrow-heritability and shared environment ( $r_{DZ} = \frac{1}{2}h^2 + c^2$ ). For  $h^2$  to be a true estimate of heritability, the EEA must hold true.



When assuming  $c^2_{MZ} = c^2_{DZ}$ , the unique environment variance terms cancel out when the DZ correlation is subtracted from the MZ correlation to arrive at  $h^2 = 2(r_{MZ} - r_{DZ})$ , called Falconer's formula. If the shared environment is not equal between MZ and DZ twins, the heritability estimate will be overestimated because the twins appear more similar due to the increased contribution of shared environment on one set of twins. Hence, shared environment accounts for a greater similarity between DZ twins ( $r_{DZ} > 0.50 r_{MZ}$ ); the shared environment between the twins causes them to be more similar. A smaller similarity than expected between DZ twins ( $r_{DZ} < 0.50 r_{MZ}$ ) can be accounted for by D; non-additive genetic variance will cause DZ twins to be less alike than expected. Both C and D cannot be modeled at once because they rely on the same proportion of  $r_{DZ} : r_{MZ}$ ; the proportion of the correlations cannot simultaneously be less than and greater than 0.50. Therefore, the standard models we use to fit twin data are ACE and ADE.

Deviating from the standard models, data can be fit to models that estimate the genetic and environmental variances for males and females differently. The term used to describe different genetic contributions for each sex is 'sex-limitation'. There are two sex-limited models using separate sex differences that can be used to try and explain the data: qualitative and quantitative models. Qualitative sex differences are differences between males and females with a unique contribution from one sex that is not present in the other sex or a difference in the *kind* of genetic contribution between the opposite sex twins (measured here as a unique genetic contribution in males, R). Quantitative sex differences do not include R but measure

the difference in the *degree* to which the same genes contribute to a trait between the opposite sex twins.

*Understanding the genetics of facial attractiveness through twin research*

Because female animals have a higher investment in offspring, the 'sexy sons' theory states that females should look for an attractive male to pass good genes onto their offspring. In humans we expect that these difference should be reduced or even reversed because females and males are expected to have equal parental contribution. Mates will, therefore, be equally choosy and sexual selection will take its course in males and females. Understanding the sexual selection and mate-choice in humans can be achieved by observing the heritability of attractiveness, a component of mate-choice that is under sexual selection (Cornwell and Perrett, 2008). Cornwell and Perrett obtained pictures of 108 female undergraduates from the University of St. Andrews and pictures of the subjects' parents (95 pictures of biological fathers and 104 images of biological mothers). In a second study, 64 images of male undergraduates and pictures of their parents were collected (64 pictures of biological fathers and 61 pictures of biological mothers). The photos were rated on attractiveness by undergraduates on a 1-7 scale, 1 = not at all attractive and 7 = very attractive. Correlations between parents and offspring were determined to estimate heritability of attractiveness. They found that both a mother and father's attractiveness determine a daughter's attractiveness; attractiveness of a daughter can be predicted by either of her parents. However, a son's attractiveness was not clearly related to his parent's attractiveness. Attractive individuals were not necessarily found to have attractive sons. This discrepancy between the sexes can

be interpreted as an example of sex-limitation. It also appears that attractiveness has an additive genetic component for females, but a non-additive component for males. The family study design limits the scope of the study. When families are used to estimate heritability, there is no way to separate genetic contribution and shared environment. These uncertainties about the heritability of facial attractiveness require further investigation.

### Current Study

The current study aims to examine the heritability of Facial Attractiveness. Little is known about the heritability of Facial Attractiveness (Rhodes, 2006). The use of twins is required to gain a better understanding of the transmission of facial attractiveness from parent to offspring. Though we have an idea of how parental Attractiveness and offspring Attractiveness correlate, maximum likelihood-based heritability estimates are not available from the Cornwell and Perrett study. The current study uses the Facial Attractiveness ratings from two twin cohorts to model variance components of Facial Attractiveness. This study intends to answer the questions: is facial attractiveness a sex-limited trait and what quantifiable contribution does additive, dominance, shared environment and unique environment make to Facial Attractiveness? Ultimately, the current study aims to gain insight on the sexually selected trait and its genetic contribution to fitness to understand human mate choice in the context of current evolutionary theories.

### Predictions

In this study, Facial Attractiveness is predicted to be sex-limited based on the Cornwell and Perrett finding. The genetic model that is likely to fit best is a sex-limited ADE model, owing to the apparent non-additive genetics or inconsistencies with parental Facial Attractiveness when predicting male Facial Attractiveness (Cornwell and Perret, 2008). Female components of Facial Attractiveness will likely be mostly additive variance and little to no dominance variance.

### Method

The sample of twins was taken from two cohorts of twins. The first was the Longitudinal Twin Study (LTS) out of the Colorado Twin Registry. The second was a sample from the Queensland Institute of Medical Research (QIMR) twin database.

The sample consists of 1599 individuals, 1357 twins from QIMR sample (84.9%) and 242 individuals from LTS sample (Figure 1). Of the sample, 1474 were MZ and DZ twins. For MZ twins,  $n=566$ , including 327 female twins and 239 male twins. For DZ twins,  $n=908$ , including 296 female-female twins, 247 male-male twins, and 365 female-male twins (Figure 2). Ages of the twin subjects ranged from 15 to 23 years old. QIMR Twins' ages ranged from 15 to 22 years old (mean=16.2, SD=0.75). LTS Twins' ages are unavailable, but most were 21-23 years old. The sex composition of the sample was 875 females (54.7%) and 724 males (Figure 1). Non-twin siblings ( $n=125$ , not used in the current study) and DZOS twins ( $n=365$ ) came from the QIMR twin database.

-----  
Insert Figure 1,2 about here  
-----

*Photograph procedure*

LTS twins were taken into a photo room and asked to take their shoes off (in addition to glasses, jackets, etc) and asked to put an ID tag on which displays tester number and the date. The twins stood against a one-square-inch grid on the wall and were told they were going to have four photos taken, two full-body and two from the shoulders up. The first two photos of the full-body were taken from 3 meters away; the twins were asked to stand straight up with feet together and arms at sides. The last two photos were taken from 1 meter away; twins were asked to have as neutral of an expression on their faces as possible, as well as raising their arms slightly as demonstrated by the experimenter. Experimenters used either Nikon Coolpix S3100 or a Fujifilm Finepix A150 digital camera. If the twins blinked, smiled, turned away, or if any other mishap occurred the experimenter re-took the picture.

LTS photos were received in JPG format, 29.5 KB in size, and 300 X 400 pixels. Images were cropped to exclude ID tags and full-body pictures were cropped to match the close-up head and shoulders photos. The photo(s) took up approximately half of the computer screen. Individual photos for LTS twins were sized at 8.6 cm x 11.4 cm and took up approximately 60%-80% of the total photo

area. QIMR photos were received in JPG format, 18.5 KB in size, and 419 X 587 pixels. All photos were from the shoulder up. Photos for QIMR were sized at 17.2 cm x 22.8 cm with faces occupying 70%-80% of the total area. All photos were displayed against a black background. The images were pasted together, organized two-by-two, using Sprite Sheet Packer software.

### *Rating Procedure*

10 undergraduate research assistants from the University of Colorado Boulder (5 males, 5 females) rated photos of each subject. Eight raters rated all photos on Attractiveness (4 males, 4 females). Two raters (1 male, 1 female) rated all photos on Acne, Grooming, and Smiling, rating each trait at a time—all smiling ratings were made before grooming and acne. The first slide that raters saw was a blue screen, prompting: “In a moment, you are going to rate the following group of faces on Attractiveness. But first you will see a slideshow of all the faces. Use this time to get a sense of the range and variation among the faces in the trait of Attractiveness. Press any key to begin.” Photos were shown in groups of 50 twins each. Before actually rating the faces, raters saw a slideshow of the upcoming 50 twins to familiarize themselves with the sample. All photos in each group were members of the same sex. Each photo was displayed for 2 seconds. After the last photo of the slideshow was shown, a single individual’s photos (2-4 in the LTS twin sample and 1 in the QIMR sample) appeared on the screen. Once rating began, below the photo was a reminder of the trait being rated (i.e. “Please rate this face’s ATTRACTIVENESS”) and a scale from 1 to 7. The scale used for attractiveness was:

1=low attractiveness and 7=high attractive. The scale used for Acne was: 1=no acne and 7=heavy acne. The scale for grooming was 1=ungroomed and 7=well groomed. The scale for smiling was 1=no smile, 2=partial smile and 3=full smile. After all 50 faces in the group were rated in a randomized order, another slideshow would be displayed of 50 more randomly selected twin faces from the remaining sample. This continued until all faces were rated. Raters were permitted to discontinue rating after each 50 set was rated. Raters were instructed to make distribution of scores approximately uniform.

The mean for Facial Attractiveness ratings was 3.73 with a standard deviation of 1.71 (Figure 7). There were no differences between male and female subject facial attractiveness ratings ( $t = 1.3$ ,  $d.f. = 1575.97$ ,  $p = 0.19$ ). On average, female subjects were rated as slightly, but not significantly, more attractive than male subjects (female mean=3.767 and male mean=3.688, Figure 3).

-----

Insert Table 1 about here

-----

-----

Insert Figure 3 about here

-----

To assess reliability of facial attractiveness ratings, Cronbach's  $\alpha$  was used. It is defined as:

$$\alpha = \frac{k\bar{c}}{\bar{v} + (k-1)\bar{c}}$$

where  $\bar{c}$  is the average of the unique ratings covariances,  $\bar{v}$  is the average of the unique variances and  $k$  is the number of raters (from Table 1.)

Here reliability is found to be:

$$\alpha = \frac{(8)1.29}{2.88 + (7)1.29} = 0.867$$

The observed attractiveness ratings correlate strongly (0.867) with the “true” facial attractiveness of the subjects’ images, as indexed by the photographs.

### *Control Variables*

Control ratings for Acne, Grooming and Smiling were collected. All three ratings could possibly contribute to the subject’s rated Facial Attractiveness rating. Two raters rated the subjects on all three control ratings. One rater did not rate 42 participants (2.6%) on Smiling and 421 participants (26.3%) on Acne. Missing ratings were predicted by regressing the variable with missing values on the other five control variables. Both control raters Grooming and Acne ratings were moderately correlated (Grooming  $r = 0.54$ ,  $d.f. = 1597$ ,  $p < 0.0001$  [Figure 5] and Acne  $r = 0.62$ ,  $d.f. = 1597$ ,  $p < 0.0001$  [Figure 4]). Smiling ratings for both raters were strongly correlated ( $r = 0.82$ ,  $d.f. = 1597$ ,  $p < 0.0001$ [Figure 6]).



For all three control ratings, mean and variance differences existed between the two raters. After confirming the significance of variance differences, mean differences were tested using Welch's *t* test for unequal variances.

- Acne: Welch's  $t = -7.3191$ ,  $d.f. = 3080.583$ ,  $p < 0.0001$  (Figure 4).
- Grooming: Welch's  $t = -4.7323$ ,  $d.f. = 3183.352$ ,  $p = 0.0001$  (Figure 5).
- Smiling: Welch's  $t = 4.5365$ ,  $d.f. = 3187.134$ ,  $p = 0.0001$  (Figure 6).

-----  
Insert Figure 4, 5, 6 about here  
-----

Ratings were z-transformed to remove mean and variance differences between the raters and an average control rating each for Acne, Grooming and Smiling was calculated. Rater main effects and two-way interactions between Acne, Grooming, Smiling, subject sex and sample of origin were each significant. Raw Attractiveness ratings were regressed on Sex, Sample and control ratings (except for Grooming) separately for each of the ratings from the 8 raters, effectively controlling for sex of the rater. Regressions were calculated by modeling the raw data as a function of the 5 two-way interactions (Sample x Sex, Sample x Acne, Sample x Smiling, Sex x Acne, and Sex x Smiling) and well as the four simple effects (Sex, Sample, Acne, and Smiling). They are modeled to minimize the degree to which the predicted score for each individual's Facial Attractiveness rating deviates from their actual rating. The residual ratings are the degree to which the predicted rating

varies from the actual rating (residual= predicted rating – actual rating) so the only thing left is the portion of facial attractiveness that is not related to Sample, Sex, Acne, Smiling and 2-way interactions between those measures. Residuals from the 8 regressions were averaged to get a single attractiveness score for each subject (Figure 8).

-----  
Insert Figure 7,8 about here  
-----

#### Grooming

Grooming was not included in the regression of the data. Facial Attractiveness and Grooming are highly correlated ( $r=0.68$ , see Table 2). Preliminary analysis with Grooming included in the regression of the data revealed problems with the Grooming measure. The current measure of Grooming is most likely conflated with Facial Attractiveness itself. Since  $r=0.68$ ,  $r^2=0.46$  which means that 46% of the variance is lost in Facial Attractiveness when Grooming is controlled for. In the preliminary analysis, there was evidence of problems with the Grooming measure when looking at the variances of data controlled for grooming and data uncontrolled for grooming. There was much less variance in the controlled for Grooming data (0.74) than the uncontrolled for Grooming data (1.269, see Table 3).

-----  
Insert Table 2 about here  
-----

We can explain this variance difference and high correlation in a number of ways. The correlation between Facial Attractiveness and Grooming would resultantly be quite high because the two are actually measuring part of the same trait. This is a quick fix for future studies whereby Grooming is more clearly defined so that it doesn't overlap with how Facial Attractiveness is defined. Another possibility is that we perceive individuals that are more attractive as being better groomed. This confers with the halo effect. This would not be a proper situation to regress Grooming out because it would effectively take part of the measured Facial Attractiveness out. An even further explanation of this Grooming issue is that people who are attractive also groom more. The same genes and environmental components could control Grooming and Facial Attractiveness, supporting the good genes theory and mate-selection theory. A limitation to the current Grooming data is that a good argument cannot be made for regressing it out and using it as a true control. In future studies, Grooming should be a more clearly defined variable that is more reliable so true Facial Attractiveness could be measured that is not inextricably coupled with Grooming.

### *Twin Analysis*

To test the assumption that our sample is representative, a saturated model that allowed the means to vary and another saturated model that fixed the means were compared. The saturated model that fixed the means to be equal between the twins did not significantly reduce fit ( $\chi^2 = 0.948$ , d.f. =18, p=0.95). The estimated

variance/covariance matrix for the population is not significantly different than the sample variance/covariance matrix. By not significantly reducing fit by forcing the means to be equal, our sample is determined to be a representative group and further analysis can be conducted.

-----

Insert Table 3 about here

-----

The correlations between MZ twins is  $r=0.66$ . Female MZ twins correlate  $r=0.63$  and male MZ twins correlate  $r=0.70$ . The correlation between DZ twins is  $r=0.32$ . Female DZ twins correlate  $0.40$ , male DZ twins correlate  $r=0.26$ , and opposite sex DZ twins correlate  $r=0.29$  (Table 5).

-----

Insert Table 5 about here

-----

-----

Insert Figure 9-15 about here

-----

Quantitative and qualitative sex-limited ACE and ADE models were fit (Table 8). None of the sex-limited models fit better than the non-sex limited models (Table 7). The sex-limited model will tell us if there is different variance within a trait between males and females. However, none of the sex-limited models fit

significantly better than the standard models so further investigation into the standard models is statistically relevant.

### Results

In an ACE model, A accounts for 65% of the phenotypic variance in Facial Attractiveness, C accounts for 0%, and E accounts for 35% (-2LL=4283.899, d.f. =1451). Two sex-limited ACE models were fitted to the Facial Attractiveness ratings. In a qualitatively sex-limited model, A accounts for 37% of the phenotypic variance in Facial Attractiveness in females, C accounts for 25% and E accounts for 38% (-2LL=4278.979, d.f. =1448). For males, A accounts for 69% of the phenotypic variance in Facial Attractiveness, C accounts for 0.2%, E accounts for 31% and R (the unique variance component in males)=0. In a quantitatively sex-limited model, A accounts for 37% of the phenotypic variance in females, C accounts for 25%, and E accounts for 38% (-2LL=4278.979, d.f. = 1447). For males, A accounts for 69%, C accounts for 0.2% and E accounts for the remaining 31 % of phenotypic variance in Facial Attractiveness. In an ADE mode, the phenotypic variance in Facial Attractiveness can be accounted for 63% by A, 2% by D, and 35% by E (-2LL=4283.891, d.f. =1451). Two sex-limited ADE models were fitted. In a qualitatively sex-limited model, A accounts for 63% of the phenotypic variance in Facial Attractiveness in females, C accounts for 0.1% and E accounts for 37% (-2LL=4278.979, d.f. =1448). For males, A accounts for 44% of the phenotypic variance in Facial Attractiveness, C accounts for 26%, E accounts for 30% and R (the unique variance component in males)=0. In a quantitatively sex-limited model, A

accounts for 63% of the phenotypic variance in females, C accounts for 0.1%, and E accounts for 37%. For males, A accounts for 44%, C accounts for 26% and E accounts for the remaining 30 % of phenotypic variance in Facial Attractiveness. And finally, in an AE model, A accounts for 65% and E accounts for 35% of the phenotypic variance in Facial Attractiveness.

-----  
 Insert Table 7, 8, 9 about here  
 -----

Akaike Information Criterion, AIC, for each model was calculated and used to compare non-nested models. AIC is calculated as:

$$AIC = 2k - 2\ln(L)$$

where k is the number of parameters in the model and L is the maximized value of the likelihood of the model. The best fitting (most parsimonious) model is the one with the lowest AIC (see Table 6). For the ACE sex-limited models, the AICs for the quantitative sex-limited model (1382.979) and qualitative sex-limited model (1384.979) were substantially less parsimonious than the standard ACE model (1381.899). Similarly, for the ADE sex-limited models, the AICs for the quantitative sex-limited model (1384.719) and qualitative sex-limited model (1386.719) were considerably less parsimonious than the standard ADE model (1381.891).

Because the D estimate was so low in the ADE model, an AE model was fitted to the data (-2LL=4283.899, d.f. =1452, AIC=1379.899). The AE model is a reduced model in which the non-additive genetic variance component is removed. Therefore,

only the additive genetic and unique environmental variances contribute to the variance in Facial Attractiveness and the resemblance between co-twins can only result from additive genetic effects. To compare the two models, a Likelihood Ratio Test (LRT) was used, where the difference between the ADE and AE models is a  $\chi^2$  distribution. The degrees of freedom difference results from dropping the D parameter. The contribution of non-additive genetic variance is not necessary to explain the observed resemblance between twins in attractiveness ( $\Delta-2LL=3.172$ ,  $\Delta d.f. = 1$ ,  $p = 0.897$ ).

The best fitting model for Facial Attractiveness is the AE model. The AIC value of the AE model is 1379.899, the lowest of the models fitted. In this model 65% of the phenotypic variance is due to additive genetic variance and 35% is due to unique environment variance (Table 9). The AE model shows that Facial Attractiveness does not include dominance/non-additive genetic variation or shared environment variation. The lack of parsimony in the sex-limited models is evidence for the same genes controlling Facial Attractiveness in males and females, which challenges the sex-limited finding in the Cornwell and Perret experiment with parents and offspring. The lack of a dominance component doesn't seem to lend to the finding that Facial Attractiveness is unpredictable in sons because of its non-additivity.

## Discussion

Counter to the Cornwell and Perrett 2008 study, we found that the same genes control male and female Facial Attractiveness and in an additive manner. The

additivity of Facial Attractiveness is in agreement with the good genes theory and Fisher's sexy sons hypothesis. Facial Attractiveness acts as a different signal in each theory (signaling health in the good genes theory and signaling an evolved preference-based trait in Fisher's sexy sons hypothesis) and which is exactly true is not clear in this study. However, because Facial Attractiveness is heritable with a high additive genetic variance, we can say it serves as a meaningful signal that confers fitness. Facial Attractiveness wouldn't be as strong of a signal if it weren't predictably heritable. Additive inheritance ensures that offspring will have the mean Facial Attractiveness of the parental Facial Attractiveness. Since survival is important to both males and females, Facial Attractiveness should be important to both genders (Thornhill and Gangestad, 1993). This is consistent with our finding that there are no sex differences in Facial Attractiveness.

### Limitations and Future Studies

The current study collected Acne, Grooming and Smiling ratings to be used as controls. The Acne and Smiling ratings appeared to be reliably measuring what each measure was originally intended to measure. But, as Facial discussed above, Grooming proved to be less reliable and highly correlated with Facial Attractiveness ratings. In future studies, raters should be better instructed on how to rate Grooming and raters who rate Facial Attractiveness should be instructed to deliberately ignore Grooming when rating Facial Attractiveness.

Another limitation of the study was the way in which age was controlled for. The QIMR sample has ages of subjects available, but the LTS sample does not. There



are clear age differences between the samples; the QIMR sample has younger twins than the LTS sample. The closest this study came to controlling for age was to control for sample. Age is a big factor in Facial Attractiveness and having the ages of the LTS twins will aid in a better control of Age.

Picture quality is another complication with the data. The procedure for photographing the twins was different between the two samples. The LTS twins were taken more recently than the QIMR photos, which were taken starting in the 1980's. This presents a problem of photo quality. Also, there were more photos presented of the LTS twins; between 2-4 photos were displayed for each twin from the LTS twin sample, while only 1 photo was displayed for each twin from the QIMR sample.

In future studies, we plan to examine the shared heritability between Facial Attractiveness and other traits. By examining the heritability of each trait and the genetic variance they share with attractiveness, we can gain support for or evidence against the 'halo effect.' Also, the validity of the good genes theory and the mate-selection theory can be assessed. Does Facial Attractiveness share genetic variance with other beneficial traits as proposed by the good genes theory? Has Facial Attractiveness become related to other traits because of assortative mating? The traits we plan to examine along with Facial Attractiveness are sexual dimorphism, facial averageness, height, and IQ.

**Table 1.** Variance/Covariance matrix for attractiveness ratings.

	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7	Rater 8
Rater 1	3.28	1.47	1.49	1.28	1.40	0.95	0.91	1.69
Rater 2	1.47	3.26	1.45	1.40	1.50	1.19	0.98	1.77
Rater 3	1.49	1.45	3.02	1.32	1.44	1.13	1.02	1.72
Rater 4	1.28	1.40	1.32	3.01	1.37	1.03	0.83	1.63
Rater 5	1.40	1.50	1.44	1.37	2.97	1.05	0.96	1.89
Rater 6	.95	1.19	1.13	1.03	1.05	2.31	0.86	1.30
Rater 7	.91	0.98	1.02	0.83	0.96	0.86	2.03	1.13
Rater 8	1.69	1.77	1.72	1.63	1.89	1.30	1.13	3.15

**Table 2.** Correlations between facial ratings.

All Subjects	Attractiveness	Acne	Grooming
Attractiveness			
Acne	-0.34		
Grooming	0.68	-0.28	
Smiling	0.12	0.04	0.12

**Table 3.** Mean, Variance and Covariance estimates for twins from saturated model with fixed means.

Parameter	Estimate
Mean	-0.03
Variance	1.269
Covariance (MZF)	0.78
Covariance (MZM)	0.90
Covariance (DZF)	0.55
Covariance (DZM)	0.34
Covariance (DZOS)	0.36

**Table 4.** Original variance/covariance matrices for twins for attractiveness.

Variance		Twin 1	Twin 2
MZF	Twin 1	1.276	0.893
	Twin 2	0.893	1.525
MZM	Twin 1	1.257	0.826
	Twin 2	0.826	1.132
DZF	Twin 1	1.144	0.460
	Twin 2	0.460	1.139
DZM	Twin 1	1.162	0.311
	Twin 2	0.311	1.240
DZOS	Twin 1	1.383	0.396
	Twin 2	0.396	1.313

**Table 5.** Twin attractiveness ratings correlations.

	Correlations	d.f.	p
MZ	0.66	258	<2.2e-16
MZF	0.63	151	<2.2e-16
MZM	0.70	105	<2.2e-16
DZ	0.32	138	9.15e-7
DZF	0.40	174	9.095e-5
DZM	0.26	115	.00398
DZOS	0.29	174	9.095e-5

**Table 6.** AIC's of twin models.

Model	AIC
ACE	1381.899
ACE quantitative sex-limited	1382.979
ACE qualitative sex-limited	1384.979
AE	1379.899
ADE	1381.891
ADE quantitative sex-limited	1384.719
ADE qualitative sex-limited	1386.719

**Table 7.** Models of attractiveness with fit statistics.

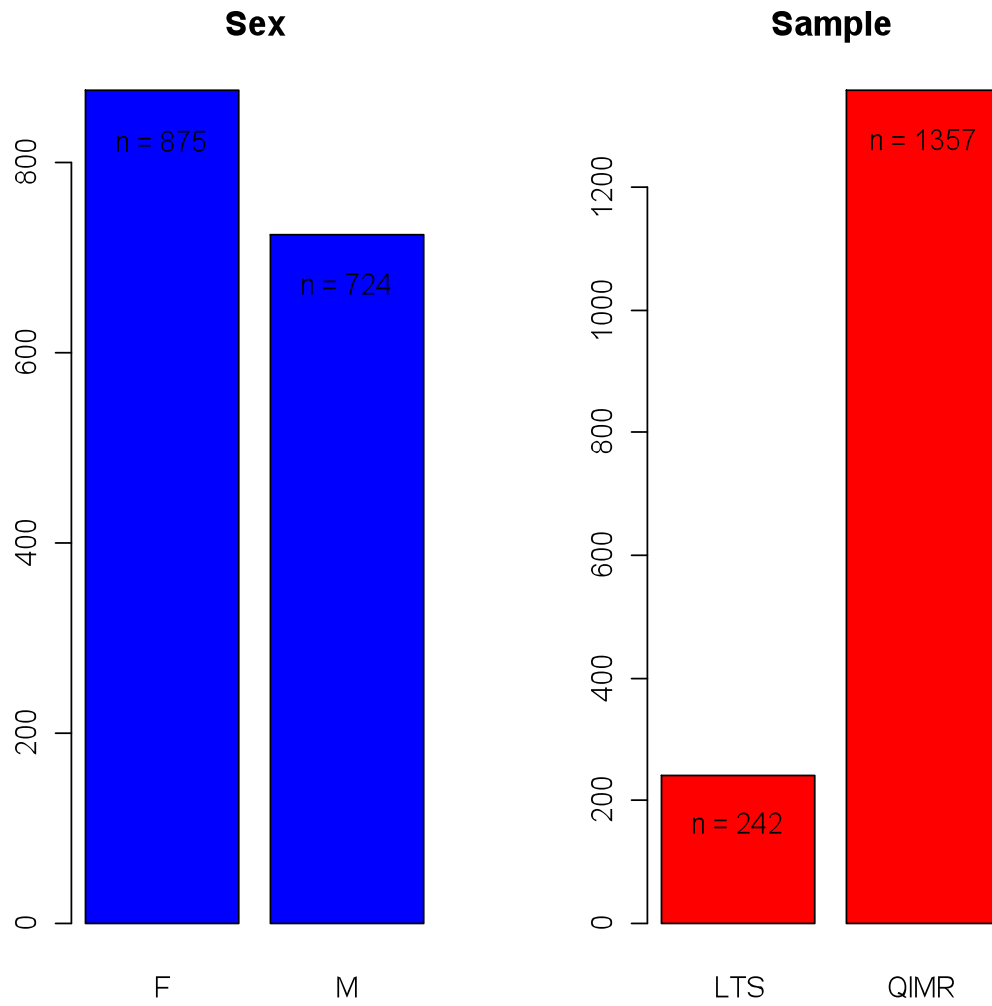
Model	-2LL	d.f.	$\Delta$ -2LL	$\Delta$ d.f.	p	AIC
ACE	4283.899	1451	4.92	-	-	1381.899
ACE-quant	4278.979	1448	-	3	0.178	1382.979
ACE-qual	4278.979	1447	-	4	0.296	1384.979
AE	4283.899	1452	4.92	1	1	1379.899
ADE	4283.891	1451	3.172	-	-	1381.891
ADE-quant	4280.719	1448	-	3	0.366	1384.719
ADE-qual	4280.719	1447	-	4	0.529	1386.719
AE	4283.899	1452	3.172	1	0.897	1379.899

**Table 8.** Sex-limited models of attractiveness.

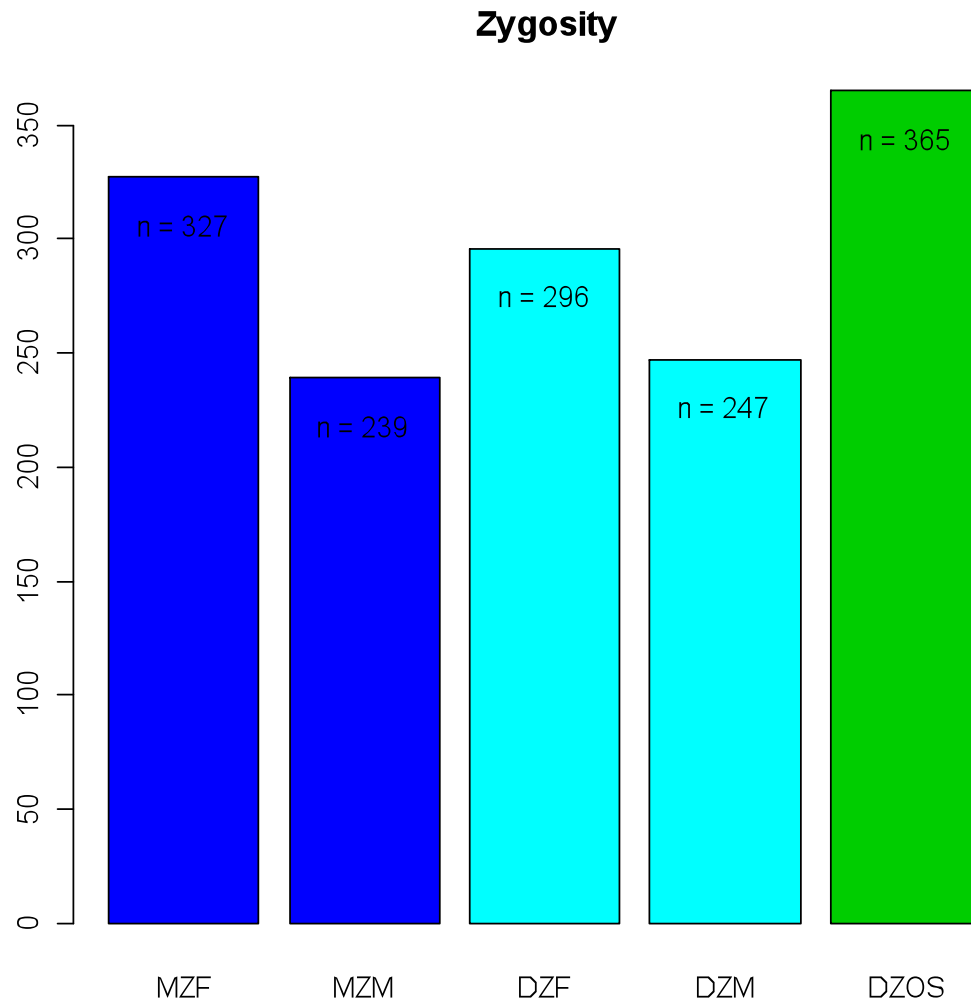
Model	Parameters						
	Af	Cf	Ef	Am	Cm	Em	Rg
ACE qual	0.37	0.25	0.38	0.69	0.002	0.31	0
ACE quant	0.37	0.25	0.38	0.69	0.002	0.31	-
ADE qual	0.63	0.001	0.37	0.44	0.26	0.30	0
ADE quant	0.63	0.001	0.37	0.44	0.26	0.30	-

**Table 9.** Non-sex limited models of attractiveness.

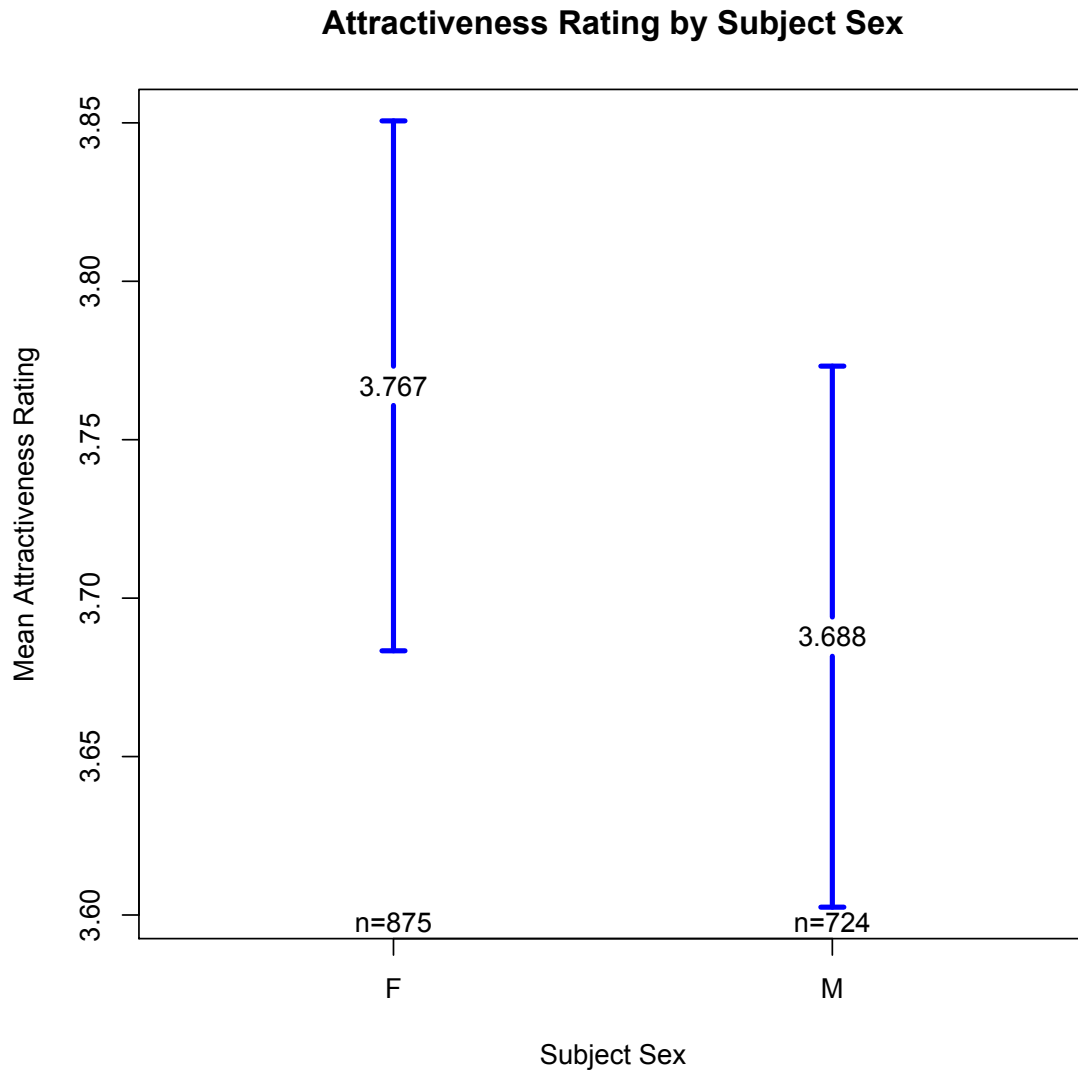
Model	Parameters		
	A	C	E
ACE	0.65	0	0.35
ADE	0.63	0.02	0.35
AE	0.65		0.35



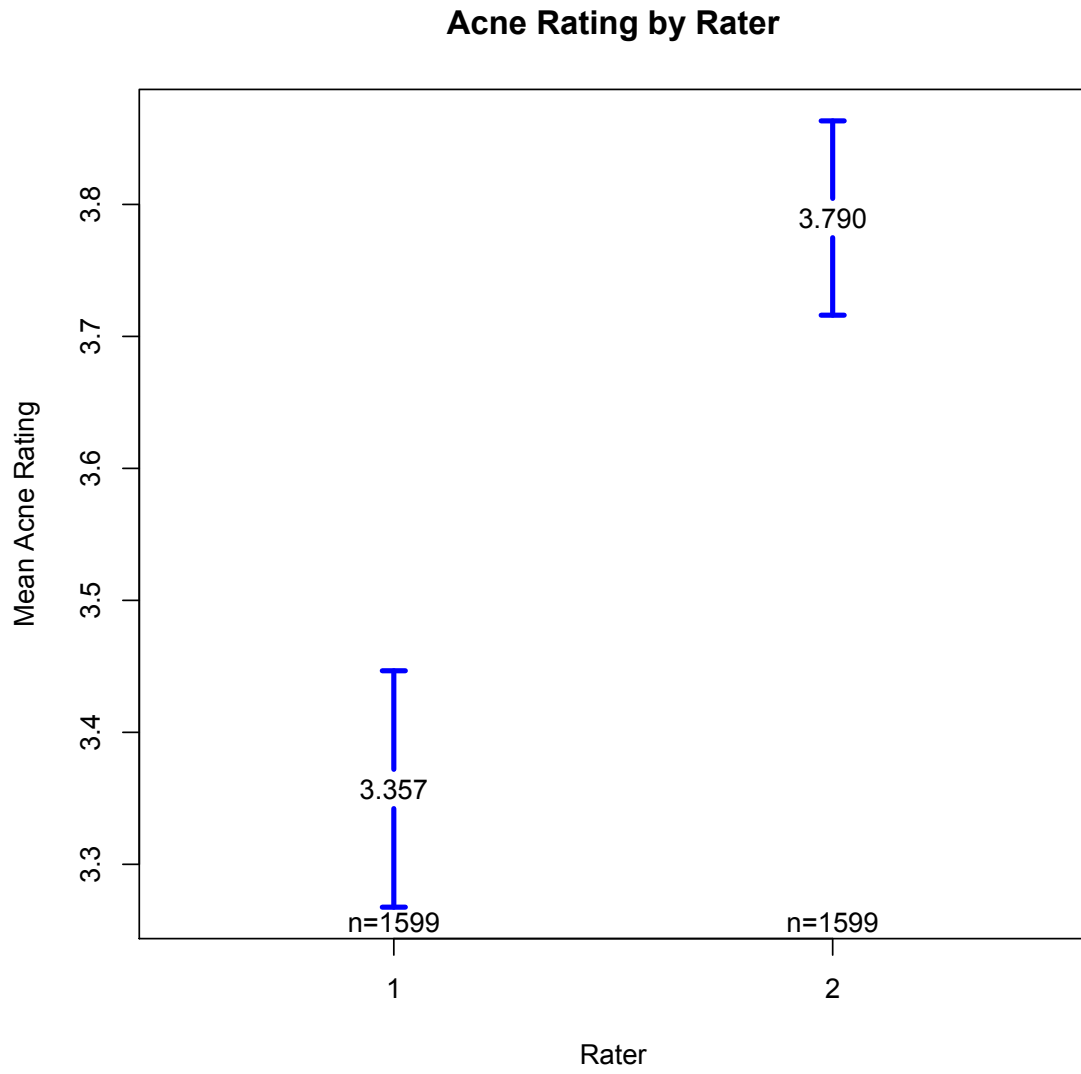
**Figure 1.** Sex and sample compositions.



**Figure 2.** Zygoty composition of the sample.

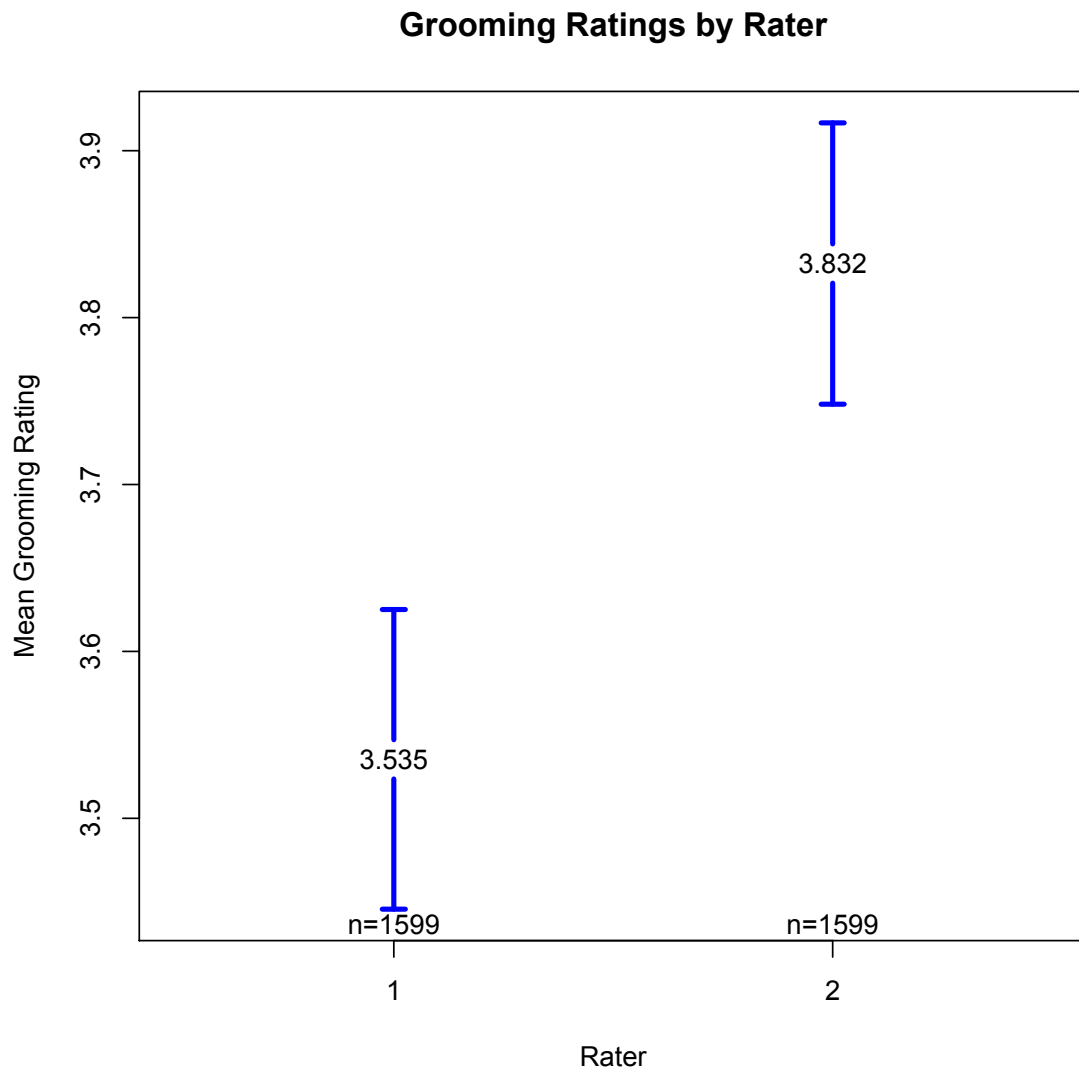


**Figure 3.** Attractiveness ratings plotted by the subject's sex.

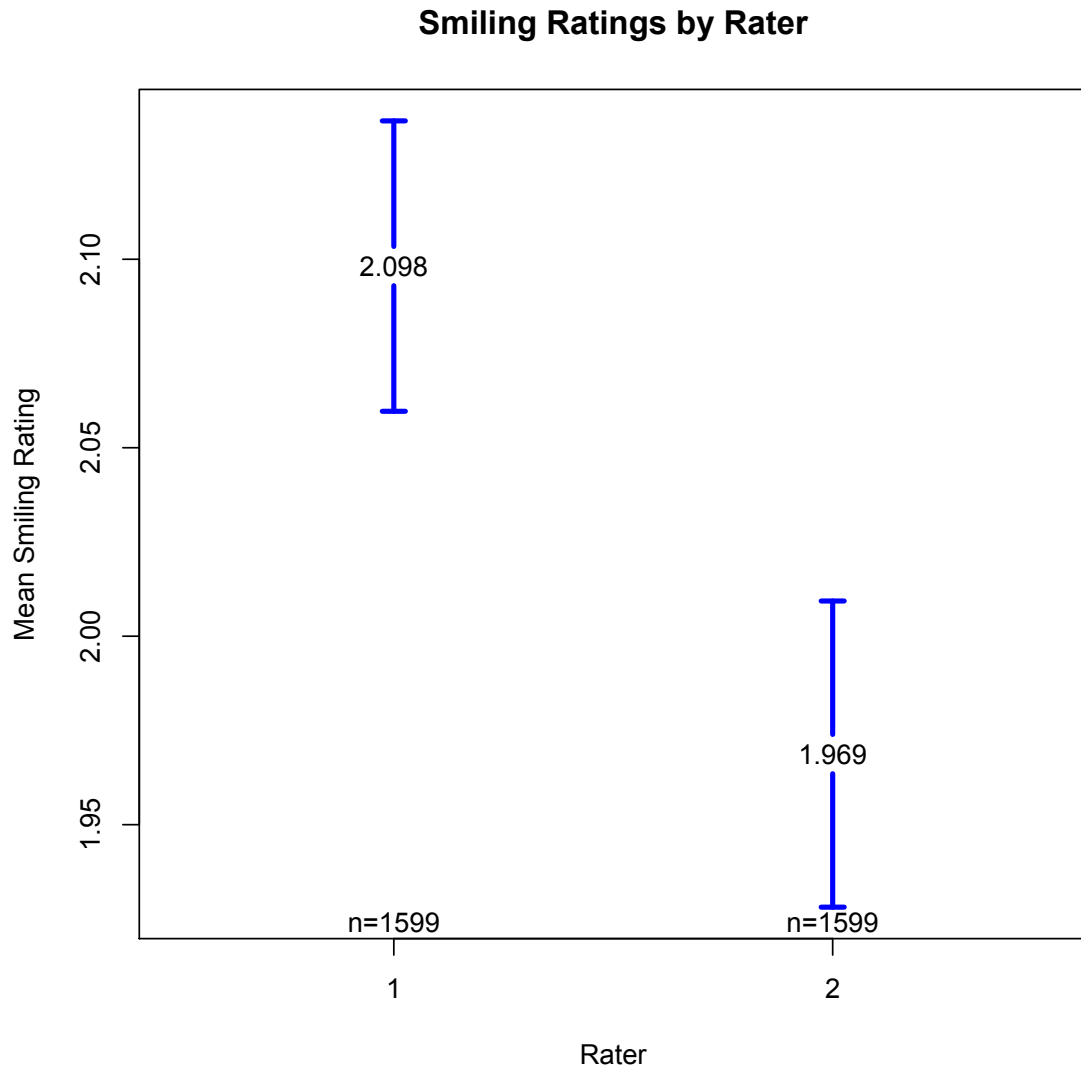


**Figure 4.** Control Variables: Acne ratings by rater.

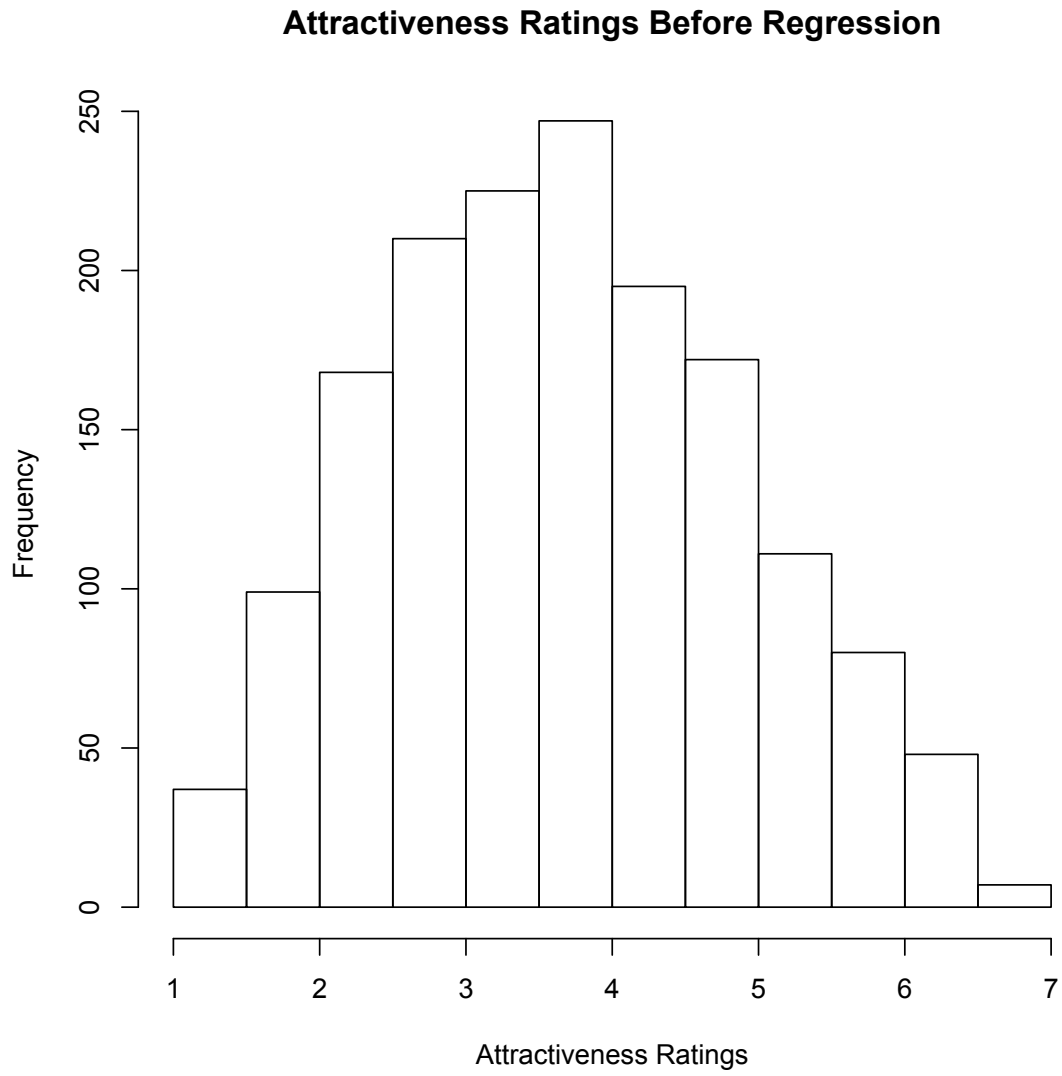




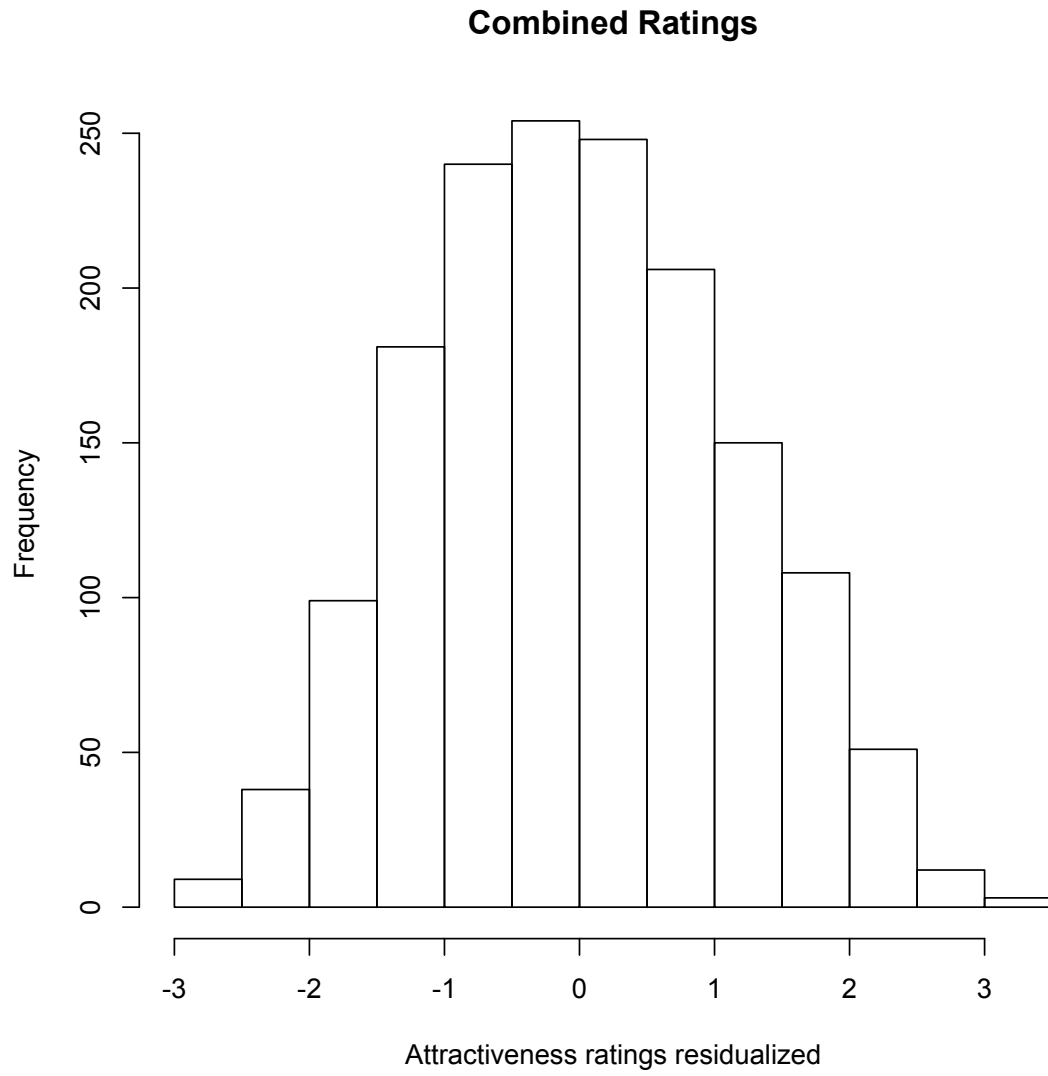
**Figure 5.** Control variables: Grooming ratings by rater.



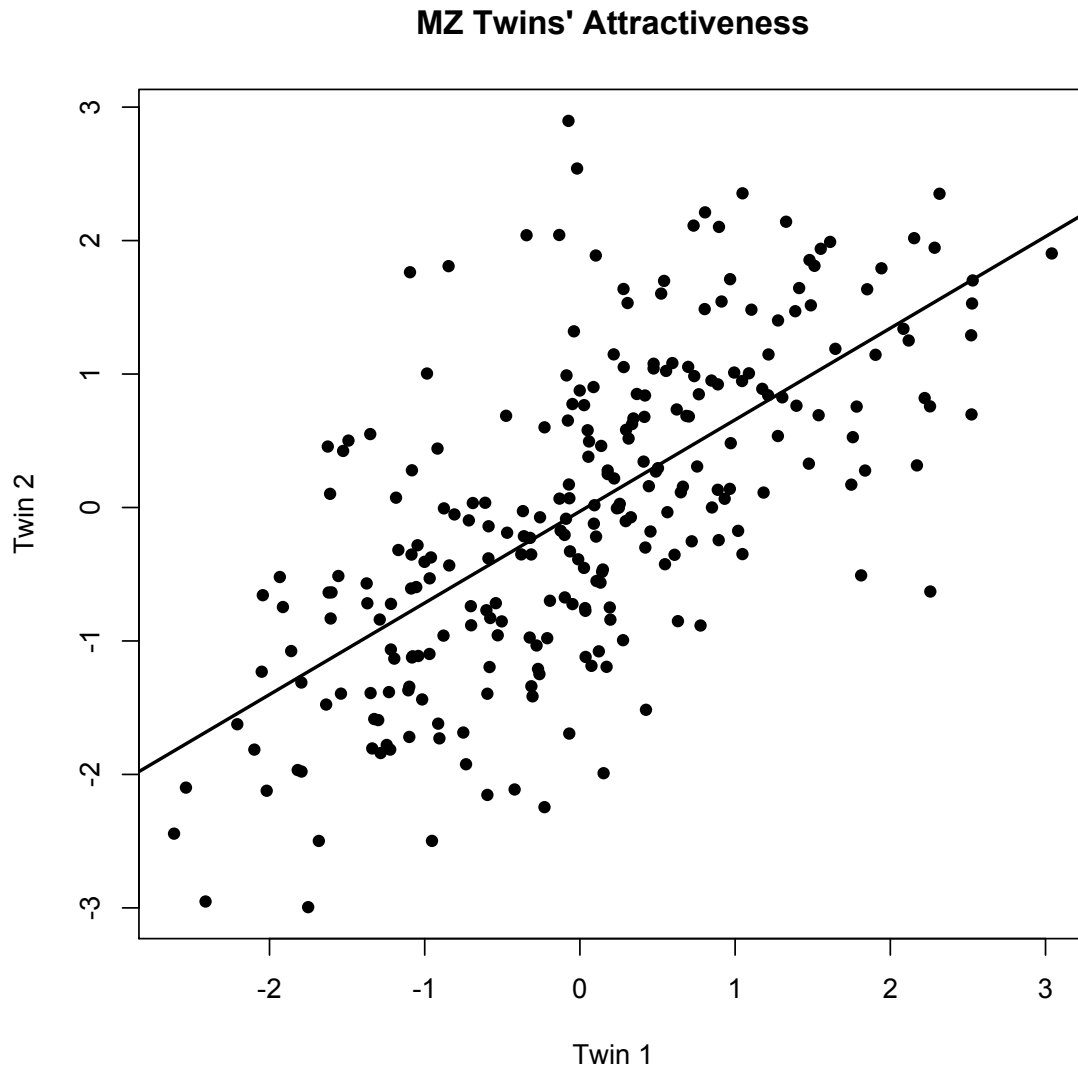
**Figure 6.** Control Ratings: Smiling ratings by rater.



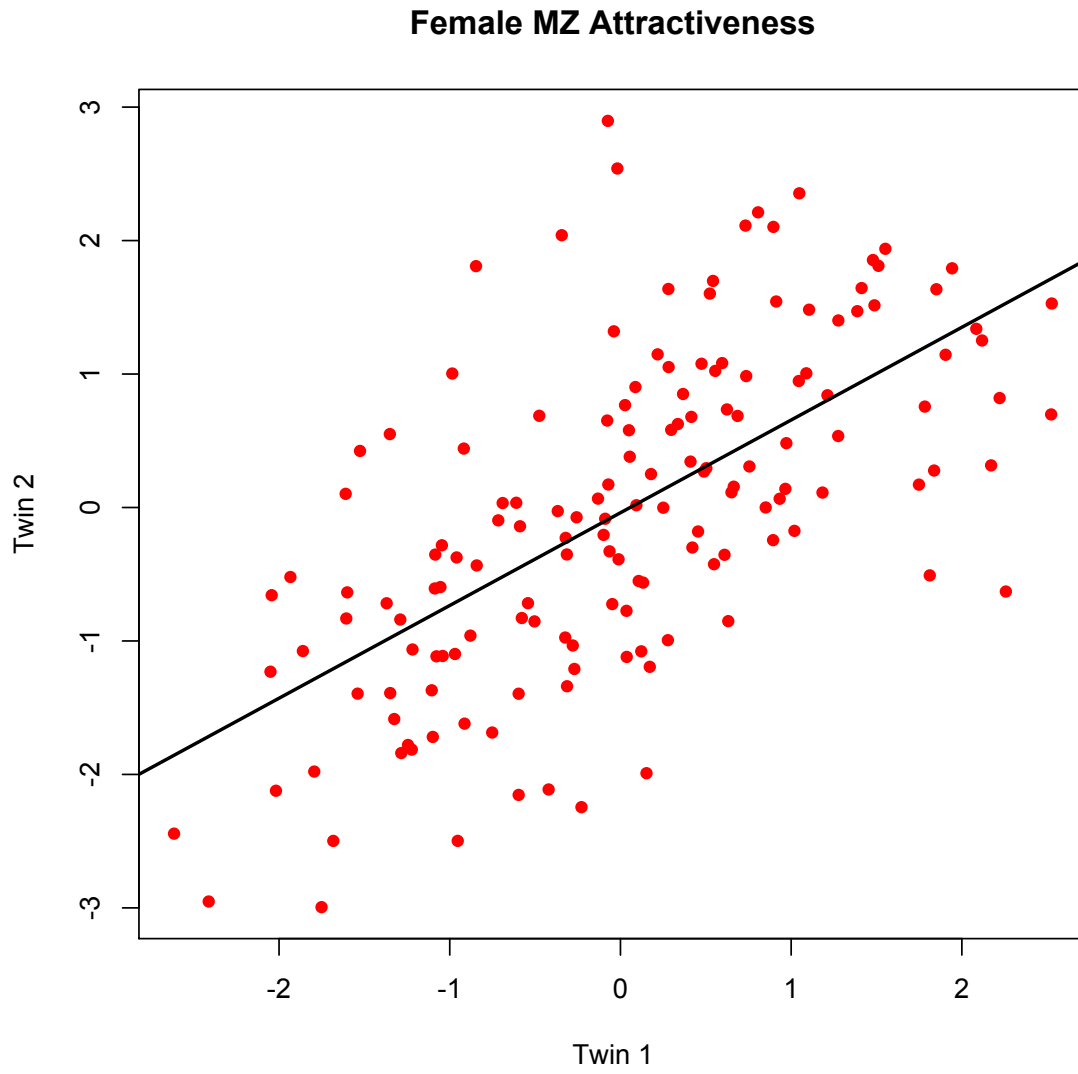
**Figure 7.** Attractiveness Ratings before regression.



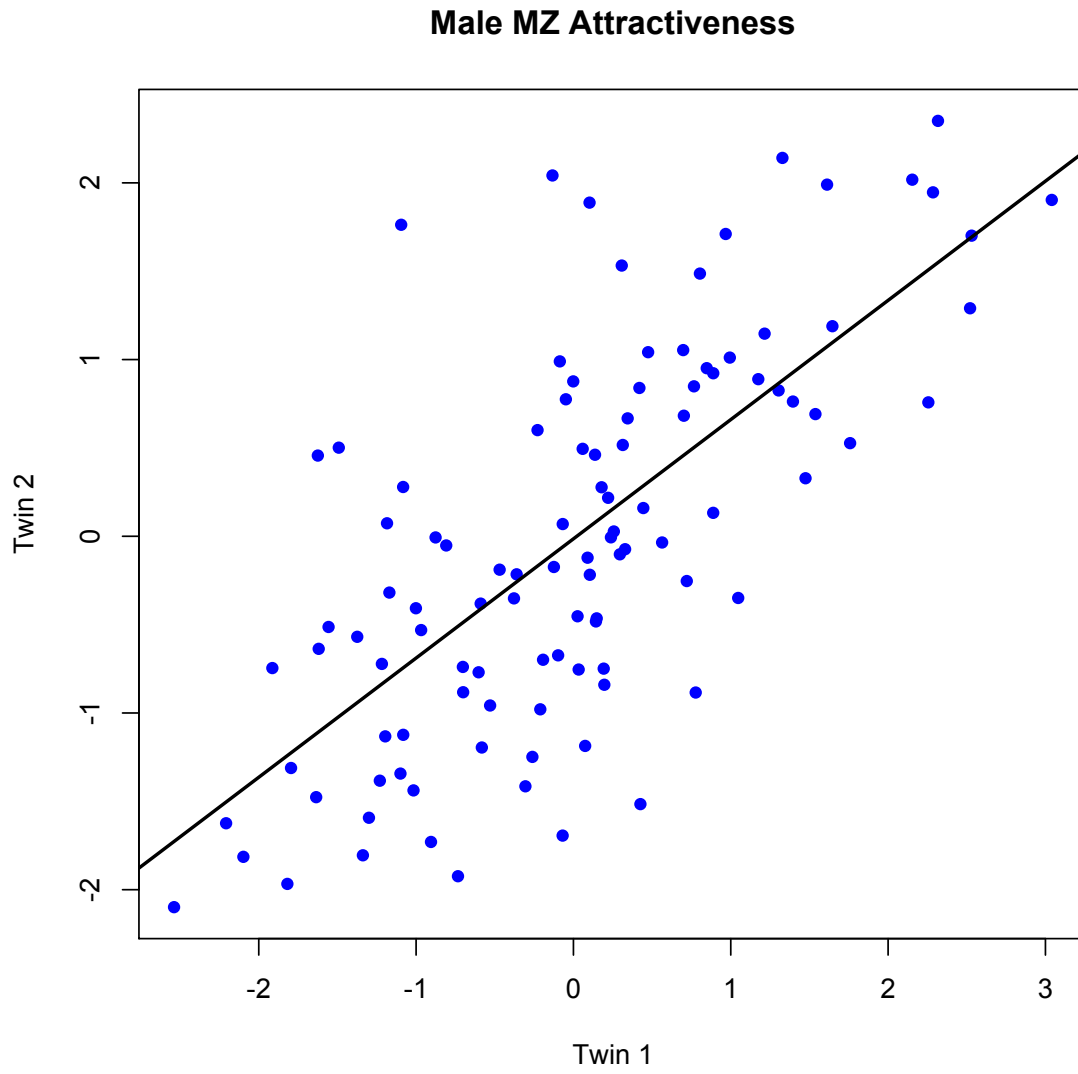
**Figure 8.** Bar graph of data after regression.



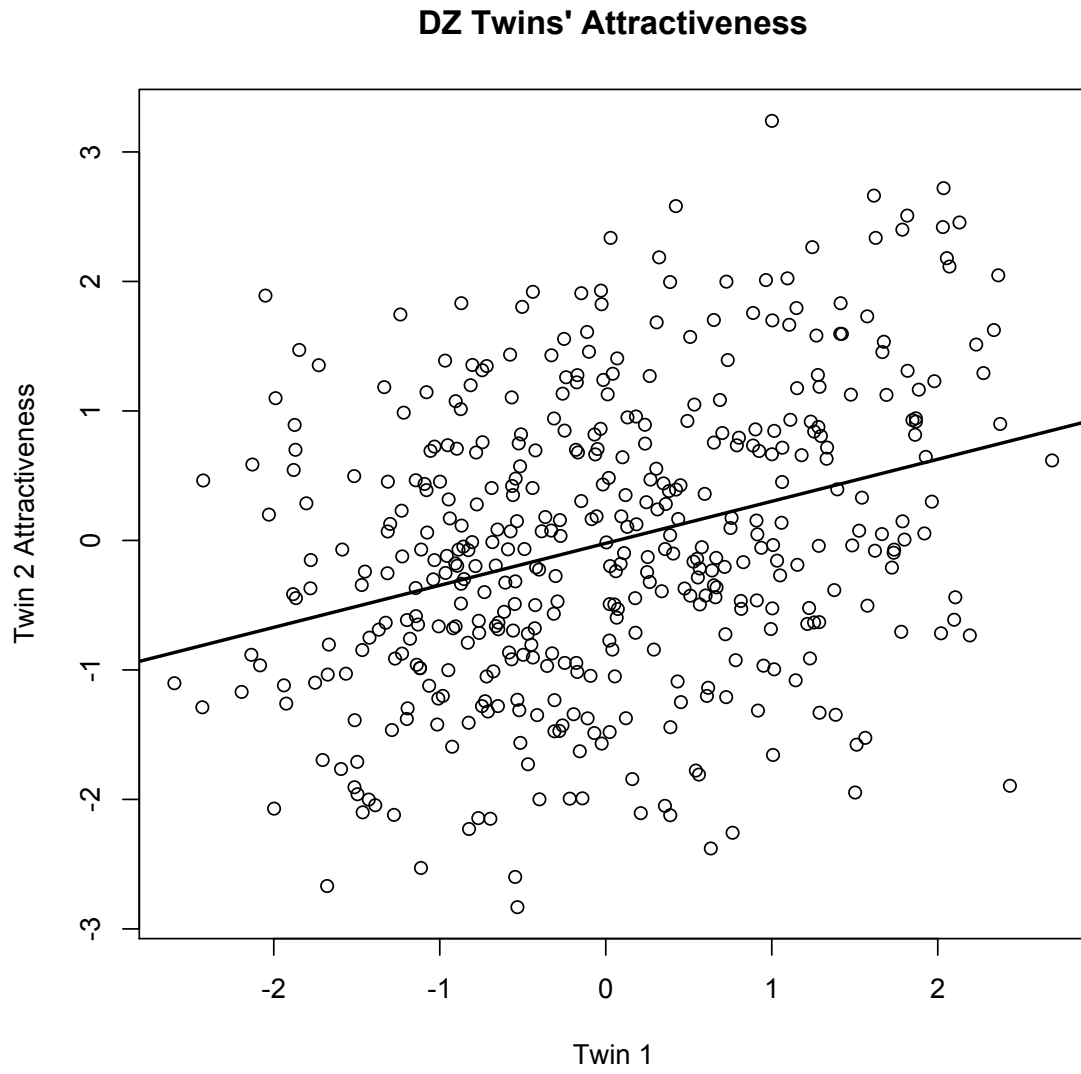
**Figure 9.** All MZ attractiveness ratings.



**Figure 10.** MZ female attractiveness ratings.

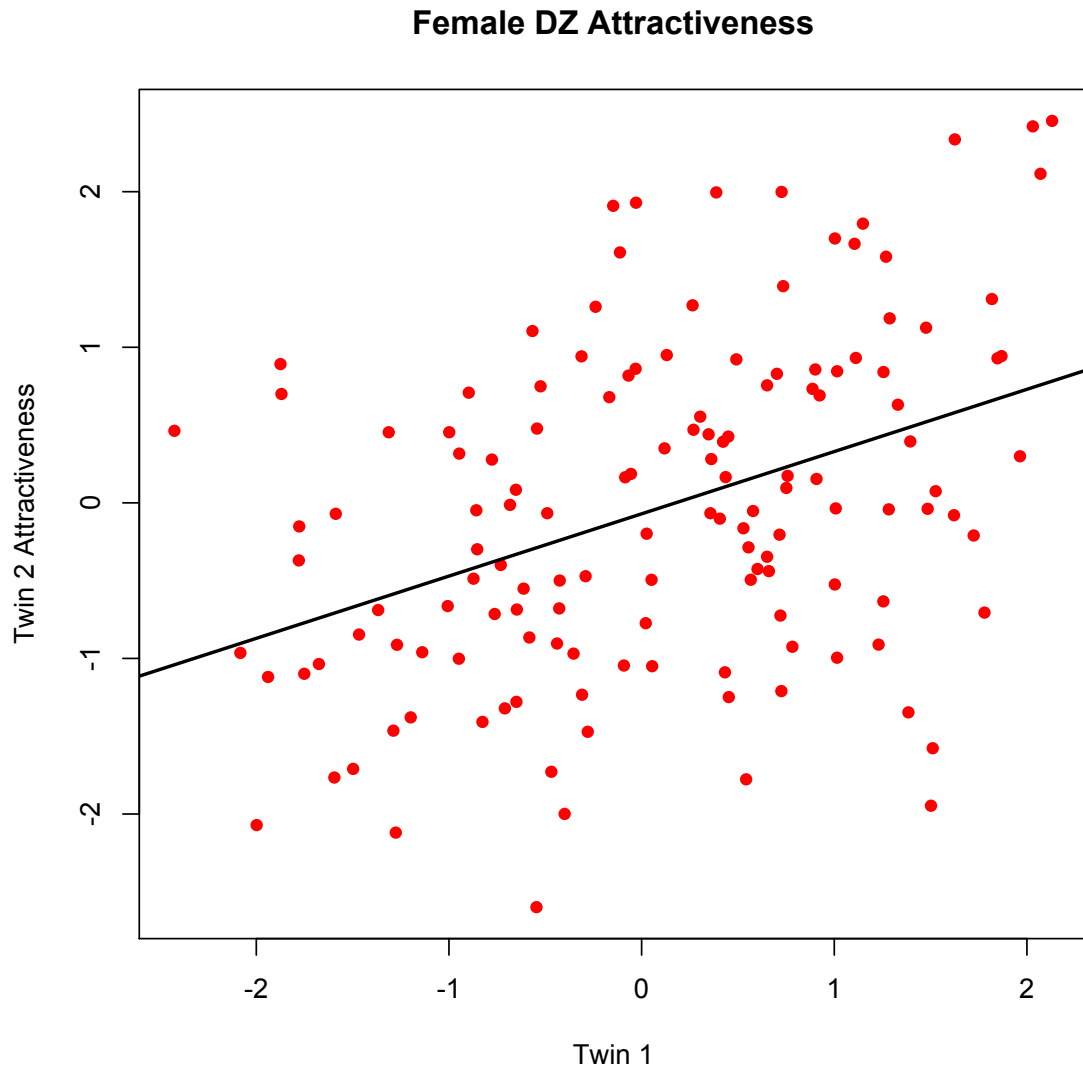


**Figure 11.** MZ male attractiveness ratings.

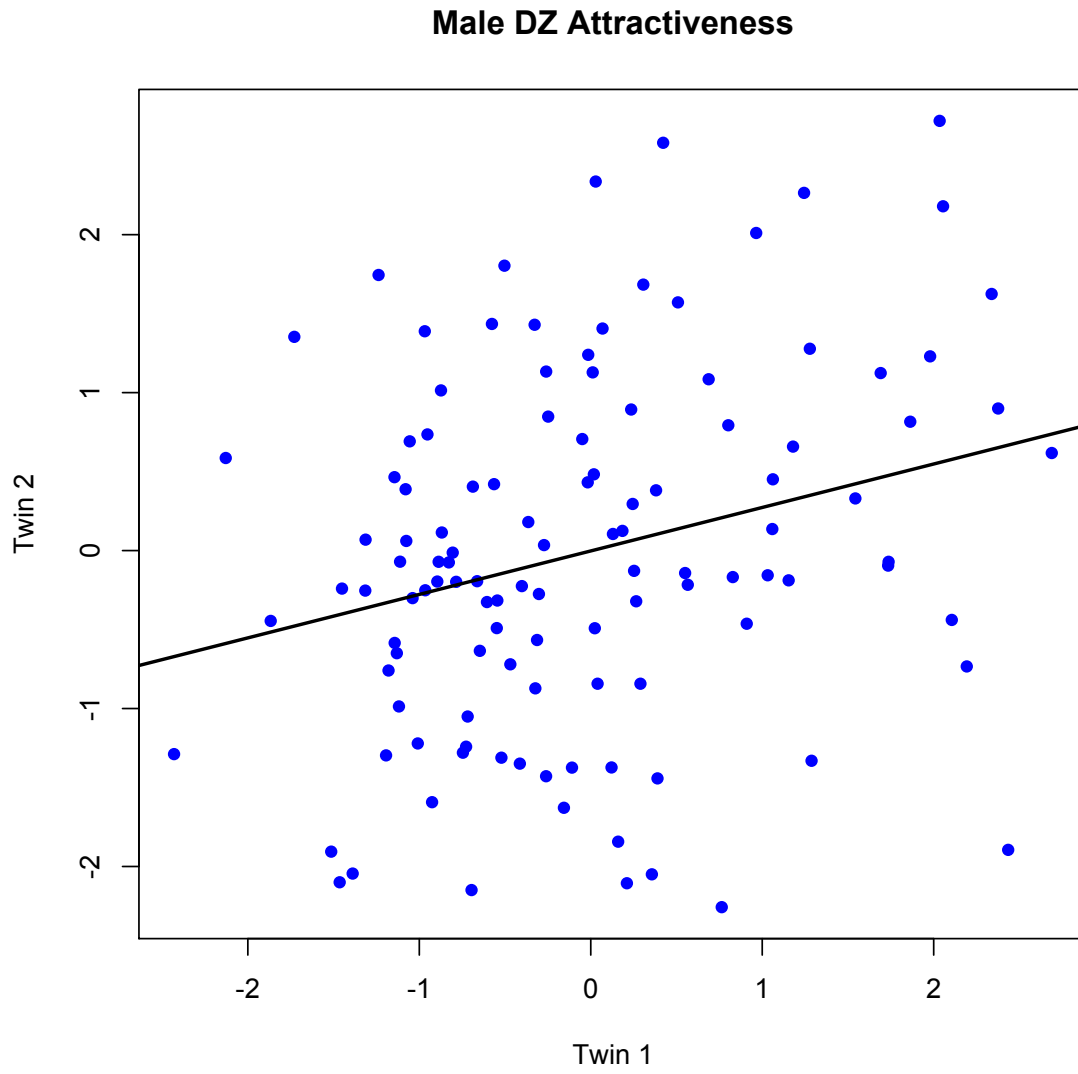


**Figure 12.** All DZ attractiveness ratings.

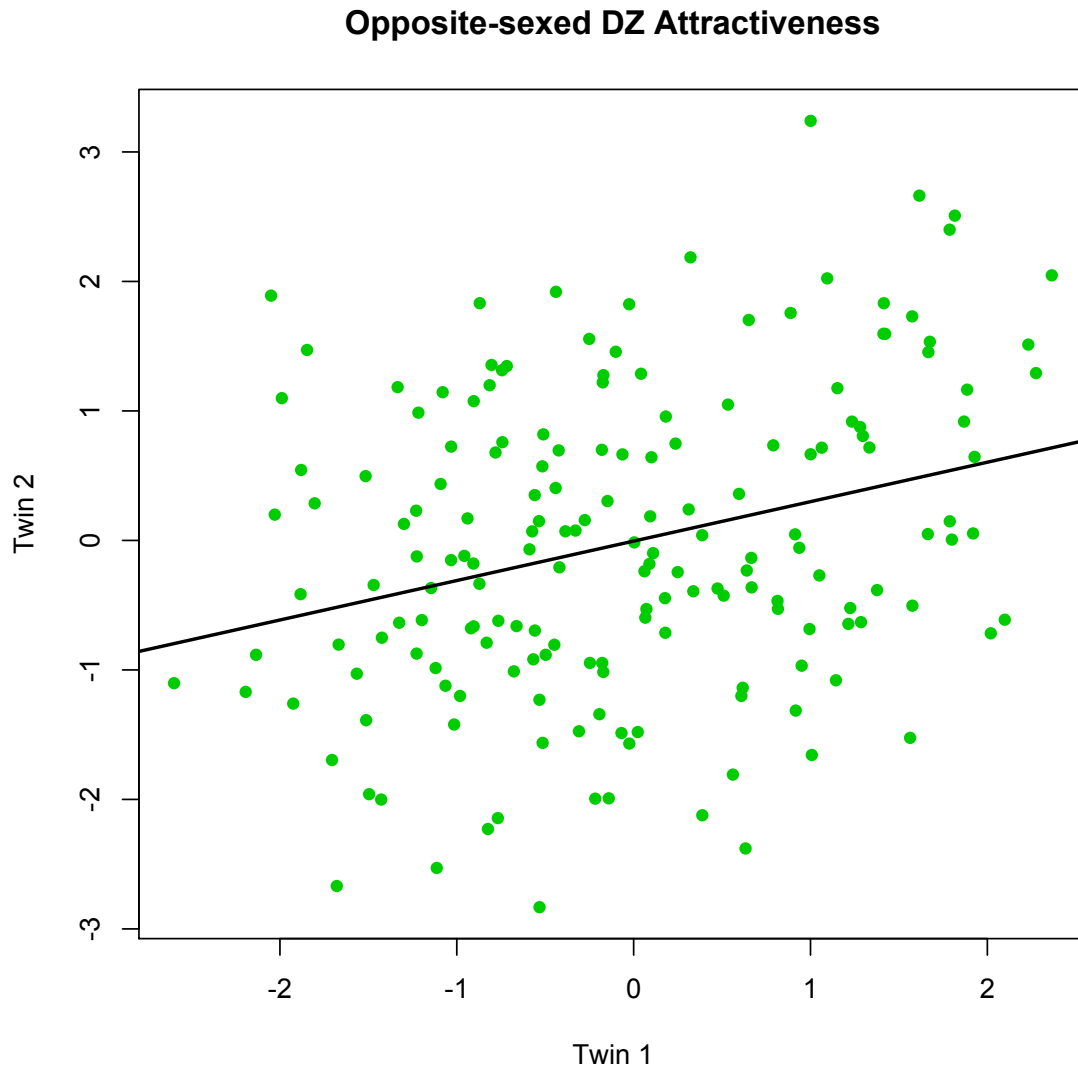




**Figure 13.** Female DZ attractiveness ratings.



**Figure 14.** Male DZ attractiveness ratings.



**Figure 15.** DZ opposite sex attractiveness ratings.

References

- Adamson, P. A., Doud Galli, S. K. (2004 Spring). Modern Concepts of Beauty. *Plastic Surgical Nursing*, 24(1), 32-36.
- Andersson, M. (1994). *Sexual Selection*. Princeton University Press.
- Asch, S. E. (1946) Forming Impressions of Personality. *The Journal of Abnormal and Social Psychology*, Vol 41(3), 258-290.
- Borgia, G. (1979). Sexual selection and the evolution of mating systems. In *Sexual selection and reproductive competition in insects* (ed. M.S. Blum & N. A. Blum), pp. 19-80. New York: Academic Press.
- Burbridge, D. (2001). Francis Galton on Twins, Heredity and Social Class. *The British Journal for the History of Science*. 34:3 (323-340).
- Cornwell, R. E., Perrett, D. I. (2008). Sexy sons and sexy daughters: the influence of parents' facial characteristics on offspring. *Animal Behavior*. 76, 1843-1853.
- Cowley, G., Springen, K. (1996, June). The biology of beauty. *Newsweek*, 127(23), 60.
- Dion, K., Berscheid, E., and Walster, E. (1972). What is Beautiful is Good. *Journal of Personality and Social Psychology*. Vol 23(3), 285-290.
- Feldman, M. W., Lewis, R. C. (1975). The Heritability Hang-up. *Science*. 190(4220), 1163-1168.
- Gangestad, S. W., Scheyd, G. J. (2005). The Evolution of Human Physical Attractiveness. *Annual Review of Anthropology*. Vol. 34, 523-548.

Gottschall, J. (2008, June). The “Beauty Myth” Is No Myth. *Human Nature*, 19(2), 174-188.

Houle, D. (1998). How should we explain variation in the genetic variance of traits? *Genetica*, 102, 241-253.

Johnson, W., Turkheimer, E., Gottesman, I. I., & Bouchard, T. J. (2009). Beyond heritability: Twin studies in behavioral research. *Current Directions in Psychological Science*, 18, 217-220.

Jones, D., Brace, C. L., Jankowiak, W., Laland, K. N., Musselman, L. E., Langlois, J. H., Roggman, L. A., Pérusse, D., Shweder, B., Symons, D. (1995). Sexual Selection, Physical Attractiveness, and Facial Neoteny. *Current Anthropology*. 36 (5), 723-748.

Kalick, S. M., Zebrowski, L. A., Langlois, J. H., Johnson, R. M. (1998) Does human facial attractiveness honestly advertise health? *Psychological Science*. 9 (1); 8-13.

Kirkpatrick, M., Ryan, M. J. (1991) The evolution of mating preferences and the paradox of the lek. *Nature, Lond.* 350, 33-38.

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, environment, and personality: A study of 850 sets of twins*. Austin, TX: University of Texas Press.

Langlois, J.H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims of myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390–423.

- Merriman, C. (1924). The intellectual resemblance of twins. *Psychological Monographs*. 33: 1-58.
- Rodes, G. (2006) The evolutionary psychology of facial beauty. *Annual Review of Psychology*. 57: 199-226.
- Rowe, D. C., Clapp, M., Wallis, J. (1986). Physical Attractiveness and the Personality Resemblance of Identical Twins. *Behavior Genetics*. Vol 17. (2), 191-201.
- Rowe, L. and Houle, D. (1996). The lek paradox and the capture of genetic variance by condition dependent traits. *Proc R Soc B*, **263**: 1415–1421.
- Scheib, J. E., Gangestad, S. W., Thornhill, R. (1999). Facial attractiveness and symmetry. *Proc. R. Soc. Lond. B*. 266, 1913-1917.
- Sheppard, L. D., Goffin, R. D., Lewis, R. J., Olson, J. (2011) The Effect of Target Attractiveness and Rating Method on the Accuracy of Trait Ratings. *Journal of Personnel Psychology*. 10(1): 24-33.
- Siemens, H. W. (1924) *Zwillingspathologie*. Ihre Bedeutung; ihre Methodik, ihre bisherigen Ergebnisse. Berlin. Springer Verlag.
- Thornhill, R., & Gangestad, S.W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature*, 4, 237–269.
- Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology*, 53, 205–214.