What Factors Into Animal Cuteness?
Tameez Latib

Introduction and goal:

The question of which animal companion is better than the other has been asked for ages. However, why don't we quantify this? While the answer clearly depends on the individual, perhaps we can extrapolate some useful information for an average person. Additionally, what makes an animal cute? Is it an inherent quality, or are there certain factors which make one better than another? In this project I try to answer some of these questions, by creating an experiment, gathering data, and finally using statistical methods to analyse and report results.

To this end, I propose some hypotheses to test. (1) Dogs are better than cats, (2) Sleeping animals are cuter, (3) Younger animals are cuter, and (4), animals with more of their body in the image are cuter. Lastly, I test (5) whether run order has effect on score

Methods (experimental design):

To test hypotheses (1)-(4), I propose a full factorial design experiment. There are 4 factors to test (adult, dog, sleep, paw), summarised in table 1 below. Note to test hypothesis 4, the factor here is only whether the animal's paw is visible or not. This is not a perfect test, but I believe it adequately captures how much of the animal's body is in the picture- if the paw is in the picture, the whole animal is in the image. Otherwise, it is only partly in the image (perhaps the tail or legs are cropped out)

I collected freely available stock images online, and asked users to rate them based on their cuteness on a scale of 0 to 10. Here a 0 would correspond to "ugliest" and 10 would correspond to "cutest". While collecting images, I noticed that there were many other variables that I could not account for, and so to increase the quality of results I obtained 2 images from each of the 4 factors. In doing so, I can test whether the image set significantly affects the score, since each one should be a replicate. In other words, we want the imageset factorial effect to be low so that we know there isn't much variance in the selected images- otherwise it is possible that a single image scores highly but this does not correlate to the factorial effects. Therefore, I end up with a 2^5 full factorial design (factors listed in table 1)

Factor	A	B	C	D	E
Low (-1)	Animal is younger	Animal is cat	Animal is awake	Paw is not visible	From set 1
High (+1)	Animal is older	Animal is dog	Animal is sleeping	Paw is visible	From set 2

Table 1: List of all 5 factors, resulting in 32 images total

Data Collection:

12 participants are shown all images in a random order (order from random.org), and asked to rate each image before seeing the next image. Each participant is a block here, since it is expected that different individuals have drastically different tastes and subjective opinions on “cuteness”. However, on the same image (same factors), the ratings each individual gives should be close. Similarly, participants cannot ‘go back’ and change their rating on a certain image. It is possible that run order affects the experiment (5). The reason is that after seeing many images, participants become fatigued and start rating more harshly. By randomizing the order and saving all run orders, we can also test for this.

The collected data can be viewed in the appendix, including run order.
Sample images from the dataset are shown below.


Results:

First, I try fit a main effects model to the 5 factors in table 1. I fit these factors to the average (over the 12 participants) score that each picture was given. The result is shown in image “Main Effects 2” below. We see that factors A and C are significant (at 5% level), and that E (the set effect) is insignificant (good!). Furthermore, C has factorial effect of about 1 score point (we double since it is factorial effect), and A has factorial effect of about -0.7 score points. From this, it is clear we should fit a model using only A and C- shown in “Main Effects 2,” with an R-squared of 0.40. The residuals are also plot below in “Main Effects 3”.

We can also look at the data itself, in table 2 summarizing the mean/variance. Since the means/variances are very different, it may be a good idea to transform the data to have mean 0 and variance 1. Then we can see what makes an animal relatively cuter per participant. We re-do the analysis on the normalized data. Still, A and C are the only important effects- outputs in “Normalised Main Effects”

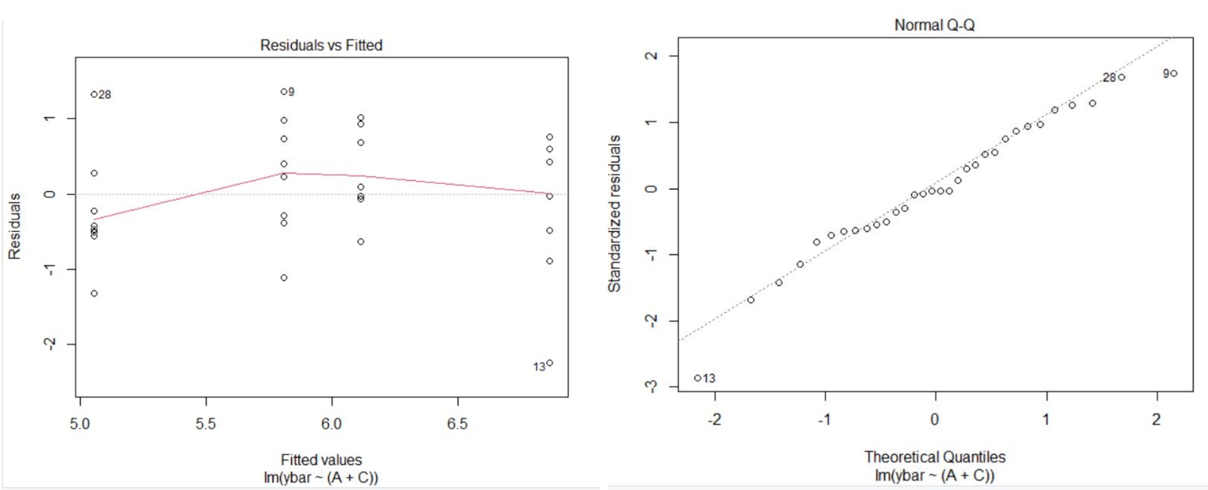
What factors into Animal Cuteness?
Tameez Latib

```
Call:
lm(formula = ybar ~ (A + B + C + D + E))

Residuals:
    Min      1Q  Median      3Q     Max
-2.1682 -0.3995 -0.0276  0.3870  1.4828

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.96042    0.14178  42.039  < 2e-16 ***
A           -0.37708    0.14178  -2.660 0.013218 *
B            0.19844    0.14178   1.400 0.173457
C            0.52917    0.14178   3.732 0.000936 ***
D            0.21667    0.14178   1.528 0.138550
E            0.05521    0.14178   0.389 0.700162
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.802 on 26 degrees of freedom
Multiple R-squared:  0.4946,	Adjusted R-squared:  0.3975
F-statistic:  5.09 on 5 and 26 DF,  p-value: 0.002196
```

Main Effects 1: Showing fit to all main factors

```
Call:
lm(formula = ybar ~ (A + C))

Residuals:
     Min       1Q   Median       3Q      Max
-2.24167 -0.47604 -0.03333  0.61354  1.35833

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.9604     0.1453  41.029  < 2e-16 ***
A            -0.3771     0.1453  -2.596  0.01466 *
C             0.5292     0.1453   3.643  0.00105 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8218 on 29 degrees of freedom
Multiple R-squared:  0.4082,	Adjusted R-squared:  0.3674
F-statistic:    10 on 2 and 29 DF,  p-value: 0.0004968
```

Main Effects 2: Showing fit to only A and C



Main Effects 3: Residuals and Normal Q-Q for model using A and C

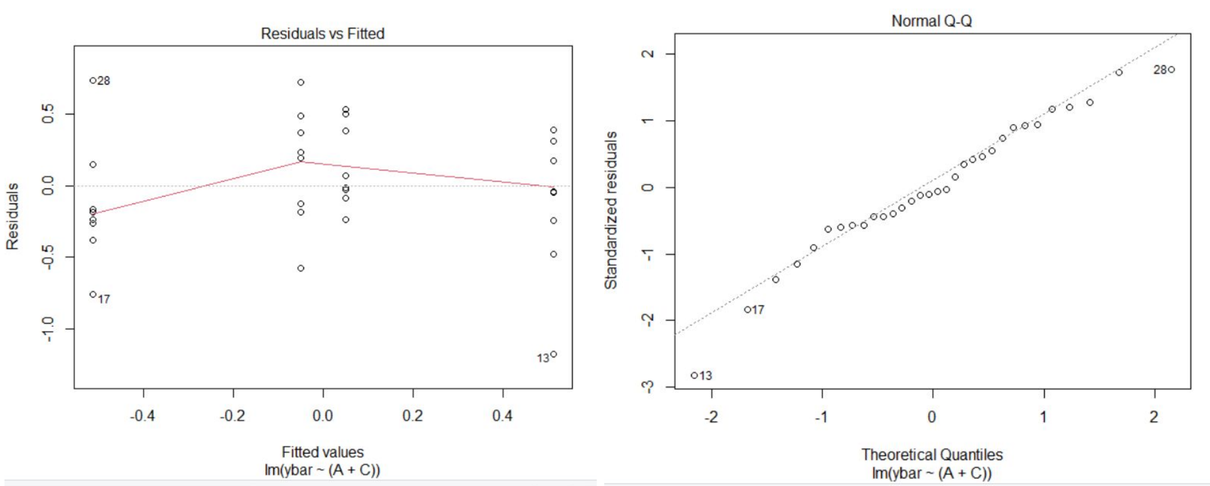
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
mean	3.78	6.77	7.12	5.31	4.21	4.32	2.10	8.27	7.65	6.40	7.10	8.47
var	4.56	3.37	2.89	3.54	6.31	3.46	4.81	2.26	2.35	5.25	1.30	2.40

Table 2: Summarizing mean/var for each participant's scores

```
Call:
lm(formula = ybar ~ (A + B + C + D + E))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.11121 -0.17193 -0.00193  0.22473  0.78963 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.850e-17  7.502e-02   0.000 1.000000    
A           -2.319e-01  7.502e-02  -3.090 0.004719 ** 
B            9.833e-02  7.502e-02   1.311 0.201447    
C            2.815e-01  7.502e-02   3.753 0.000888 ***
D            1.241e-01  7.502e-02   1.654 0.110214    
E            3.635e-02  7.502e-02   0.485 0.632074    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4244 on 26 degrees of freedom
Multiple R-squared:  0.5214,	Adjusted R-squared:  0.4293 
F-statistic: 5.664 on 5 and 26 DF,  p-value: 0.00116
```

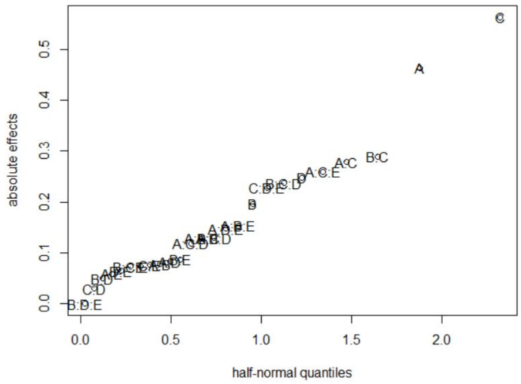
```
Call:
lm(formula = ybar ~ (A + C))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.17330 -0.23593 -0.03431  0.32397  0.73219 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept)  7.850e-17  7.717e-02   0.000  1.00000   
A           -2.319e-01  7.717e-02  -3.004  0.00544 **
C            2.815e-01  7.717e-02   3.648  0.00103 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4366 on 29 degrees of freedom
Multiple R-squared:  0.4351,	Adjusted R-squared:  0.3961 
F-statistic: 11.17 on 2 and 29 DF,  p-value: 0.0002534
```

Normalised Main Effects 1 and 2: Fit to all factors, and fit to A and C



Normalised Main Effects 3: Residuals and Normal Q-Q for model using A and C

Noticing that the normalised model gives slightly better results, (R-squared now 0.43 from 0.40 in our model) we check 2 and 3 factor interactions using the normalised data (since we have no aliasing with full factorial design). First we obtain the half-normal plot for location (mean) labeled "half normal plot", and also the factorial effects in "factorial effects". From the full model, we see only A, C, A:C and B:C are significant at 10% level so we use these (and also B) for our two-factor interaction model. Lastly plot residuals. Note our two-factor model has R-squared 0.60

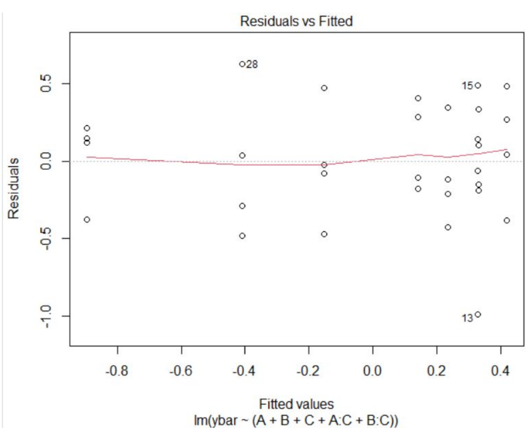


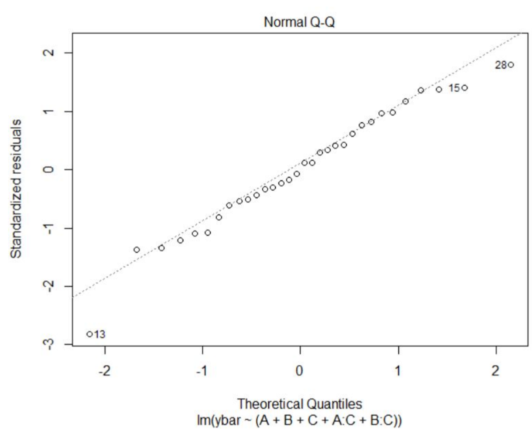
Half normal plot for location: Note the descending order C, A, B:C, A:C, A:C:E, ...

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.850e-17  6.898e-02   0.000  1.00000
A           -2.319e-01  6.898e-02  -3.361  0.01521 *
B            9.833e-02  6.898e-02   1.425  0.20393
C            2.815e-01  6.898e-02   4.081  0.00649 **
D            1.241e-01  6.898e-02   1.799  0.12220
E            3.635e-02  6.898e-02   0.527  0.61713
A:B          3.875e-02  6.898e-02   0.562  0.59462
A:C          1.391e-01  6.898e-02   2.017  0.09033 .
A:D         -4.151e-02  6.898e-02  -0.602  0.56935
A:E          2.941e-02  6.898e-02   0.426  0.68478
B:C         -1.446e-01  6.898e-02  -2.097  0.08085 .
B:D          2.481e-02  6.898e-02   0.360  0.73145
B:E          4.356e-02  6.898e-02   0.631  0.55101
C:D          1.488e-02  6.898e-02   0.216  0.83632
C:E         -3.814e-02  6.898e-02  -0.553  0.60028
D:E         -3.220e-02  6.898e-02  -0.467  0.65717
A:B:C        6.405e-02  6.898e-02   0.928  0.38898
A:B:D        6.448e-02  6.898e-02   0.935  0.38598
A:B:E        7.639e-02  6.898e-02   1.107  0.31053
A:C:D       -5.918e-02  6.898e-02  -0.858  0.42388
A:C:E       -1.297e-01  6.898e-02  -1.880  0.10908
A:D:E        7.361e-02  6.898e-02   1.067  0.32695
B:C:D        1.187e-01  6.898e-02   1.720  0.13617
B:C:E        3.621e-02  6.898e-02   0.525  0.61851
B:D:E       -9.853e-06  6.898e-02   0.000  0.99989
C:D:E       -1.139e-01  6.898e-02  -1.651  0.14989
---
```

```
Call:
lm(formula = ybar ~ (A + B + C + A:C + B:C))

Residuals:
     Min       1Q   Median       3Q      Max
-0.98788 -0.19490  0.00794  0.27401  0.62833

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.850e-17  6.872e-02   0.000 1.000000
A           -2.319e-01  6.872e-02  -3.374 0.002333 **
B            9.833e-02  6.872e-02   1.431 0.164391
C            2.815e-01  6.872e-02   4.097 0.000363 ***
A:C          1.391e-01  6.872e-02   2.024 0.053336 .
B:C         -1.446e-01  6.872e-02  -2.105 0.045137 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3887 on 26 degrees of freedom
Multiple R-squared:  0.5984,    Adjusted R-squared:  0.5212
F-statistic: 7.748 on 5 and 26 DF,  p-value: 0.0001421
```

Factorial Effects 1 and 2: Note the only significant factors are A, C, A:C, B:C at 10% level (left). Then we find R-squared of 0.59 for this model (right)





Factorial effects 3: Residuals and Normal Q-Q for model using A, B, C, A:C and B:C on normalised data

We can also check run order effect on rating. Simply looking at the correlation between score and run order, we see that per participant there are some correlations (Table 3), but overall there is no correlation, and so we do not need to include this in our model.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	All
Cor	.05	-.38	.09	.24	.23	-.08	.09	.13	.32	-.11	.00	.35	.05

Table 3: Correlations of run order vs score per participant

Discussion and conclusion:

If we look at the Main Effects model, we only have A and C as important factors. That is, younger animals and sleeping animals are cuter, they are both significant at 5% level, confirming hypothesis (2) and (3). Furthermore, dogs and cats are roughly equal, and the presence of the animal body is not significant either, so we cannot confirm hypotheses (1) and (4). We also see that the imageset is not significant, and only has a small effect. This means that even though there are many factors unaccounted for, we are measuring the effect of each factor instead of whether some chosen images are inherently better than others. However, as pointed out in Table 2, user scores varied dramatically. We normalize the data (to mean 0, variance 1 per block/participant) so that we can find out the relative impact of a certain factor. Then, when we perform the analysis again, we actually find that A and C are significant at the 1% level, and we have a better R-squared than previously. We can also look at the residuals and normal Q-Q plots, and see there are no large abnormalities. The biggest outlier is image 13, which is the only chihuahua. While we cannot say for certain, it seems that people disliked the chihuahua far more than other images. Otherwise, the assumptions of normality and constant variance are somewhat reasonable. The residual plot deviates slightly, but this is expected when only using main effects. Continuing with the normalised data, we add two and three factor interactions. We still find that the imageset (and all its interactions) are insignificant, but we now have that A:C and B:C are significant at 10% level. Old + sleeping is a cute combination, but unfortunately dog + sleeping is an ugly combination. We can also look at the residual and normal Q-Q plot here, and see our assumptions are reasonable. Also, using the two factor interactions we also increased R-squared from ~0.4 to ~0.6, which explains the better fit. Analysing the run order (table 3) we see no correlation overall, and thus we disprove hypothesis (5), that run order affects score. Lastly, to acknowledge some limitations: we do not factor in breed, color, expression, etc. I believe a larger scale study may yield more interesting results, but it would be wise to use fractional factorial in these cases, since 3-factor and higher effects can be ignored usually. Additionally, participants are similar in age so it is possible that older/younger people may have different opinion, and so a larger and more diverse sample would be necessary to avoid this issue.

APPENDIX:

"Animal.csv" has the data formatted as a full factorial design, with all scores as different columns. "Animal_order.csv" has the actual run order for each participant. P1 saw images first, with the order given by the "Order" column (from 1 as first image seen to 32 as last image seen). Next P2 saw images, again with "order" column listing the order of images seen.

Since there is a lot of data, I put it on google sheets here:

https://docs.google.com/spreadsheets/d/178HO68E3IXAjeXKu8ryTSXVbsLWp1_Qh4igALcwCaXI/edit?usp=sharing