Face Identification Game

An exploration of the (best) linguistic predictors of game success

By Lotte Verheijen

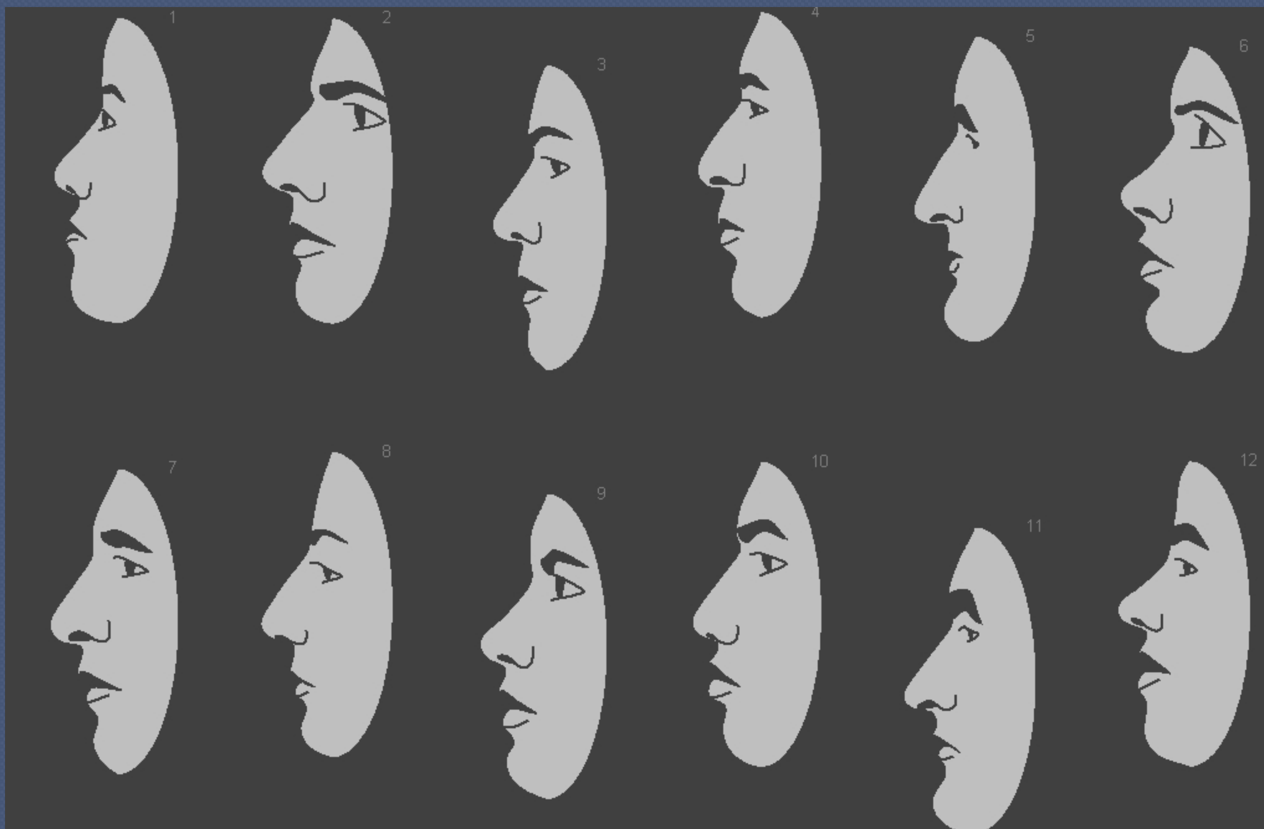
Content

- Introduction
  - Research design
  - Variables/Data
  - Research questions
  - Expectations (based on CMC coordination research)
- Method
  - Multiple logistic regression
    - assumptions
  - Model selection
    - Unclean data
    - Clean data + assumptions
    - Exclude an outlier + assumptions
- Tentative conclusions: best model
- Questions?

Research Design

- Face identification: CMC coordination game
- Akin to the tangram task
  - Director: sees 1 face
  - Matcher: sees 12 faces
  - Alternated
  - They have to match (via communicating) the one face to the director is seeing to one out of the 12 the matcher is seeing (Bangerter & Clark, 2003)

Research Design



Matcher

Director

Research Design

Director

Matcher

Research Design

mouth, eyes, and brows are all drooped down	, 24ee
brows are very drooped down	, 24ee
looks sad	, 24ee
he has a fat chin right?	, liao
like blockish	, liao
? or is it rounder	, liao
yea, the bottom of his chin is a little bit slanted	, 24ee
doesnt bulge out too much	, 24ee
nose is the thing that stands out	, 24ee
eye brows arent bushy?	, liao
no	, 24ee
kinda thin	, 24ee
mouth and chin are the second thing that stands out the most	, 24ee
his are looking forward?	, liao
then the forehead	, 24ee
yea straight ahead	, 24ee
his forehead to the nose is making sorta like a paperclip s?	, liao
err that was a bad question	, liao
noo....?	, 24ee
i think i know	, liao
Correct. Your score is 10	, tgliaoliao
Correct. Your score is 10	, nkyung24ee
/10((NEWLINE))	
Game success: 1	

Variables

- Quantified conversation analysis
- Outcome: Success of Trial (binomial)
- Predictors:

  - 1. Deletes per trial
  - 2. Typing time per trial
  - 3. Nr. of text turns per trial
  - 4. Total trials attempted
  - 5. Characters typed per trial
  - 6. Nr. of question marks per trial

Research Questions

- What is the best model for prediciting game success in the face identification game and which variables are included in this model?

- What is the influence of cleaning up the data on such a model? (regarding AIC levels, assumptions)

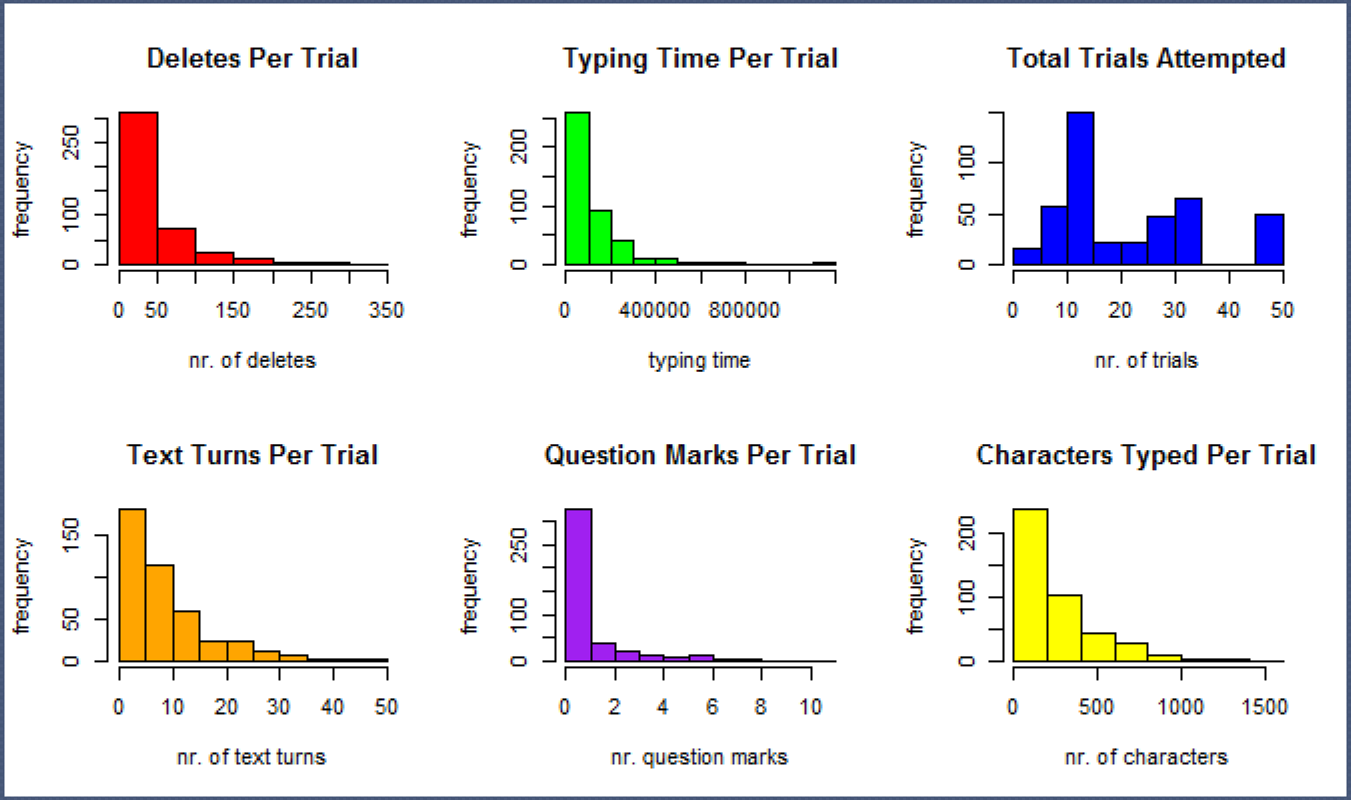
Expectations

- Exploration of (best) predictors… unknown
- Type of predictor
  - 1. Deletes per trial = negative
  - 2. Typing time per trial = negative
  - 3. Nr. of text turns per trial = negative
  - 4. Total trials attempted = positive/negative*
  - 5. Characters typed per trial = negative
  - 6. Nr. of question marks per trial = negative
- *cutoff point instead of practice effect

CMC Coordination

- Uncertainty or repair hints at decreased coordination (deletes, typing time, question marks, also related to nr. of text turns and characters typed) (Hayashi, Raymond & Sidnell, 2013).
- High turn rate: lower coherence/coordination level (Smith, Cadiz & Burkhalter, 2000; Herring, 1999)
- Decreased coordination is assumed to lead to a lower success rate (Garrod & Anderson, 1987)

Unclean Data: First Look



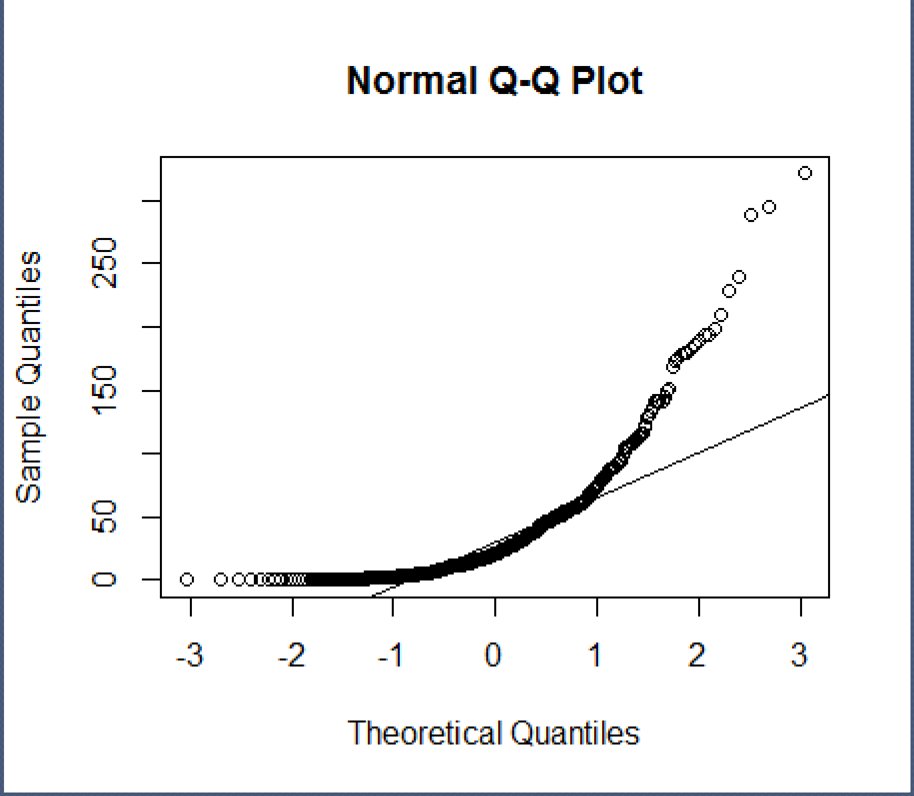
Unclean Data: First Look

- Non-normal distribution
- QQplots and

Shapiro-Wilk tests Confirm this.

See, for example, 'Deletes'

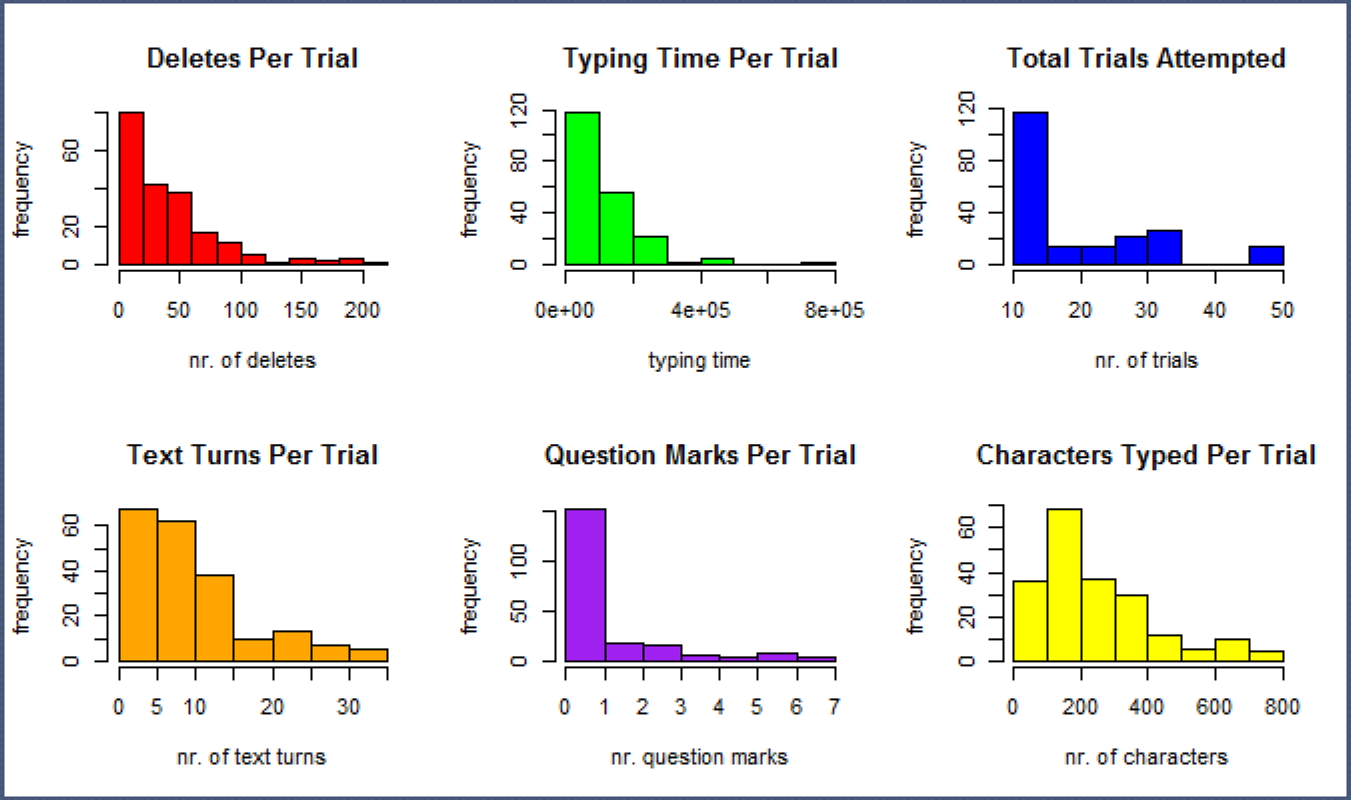
W = 0.74059,
p-value < 2.2e-16



Clean data

- Completed more than 12 trials
- Only first 12 faces
  - see the actual strategies
  - nice subset for a pilot study

Clean Data: First look



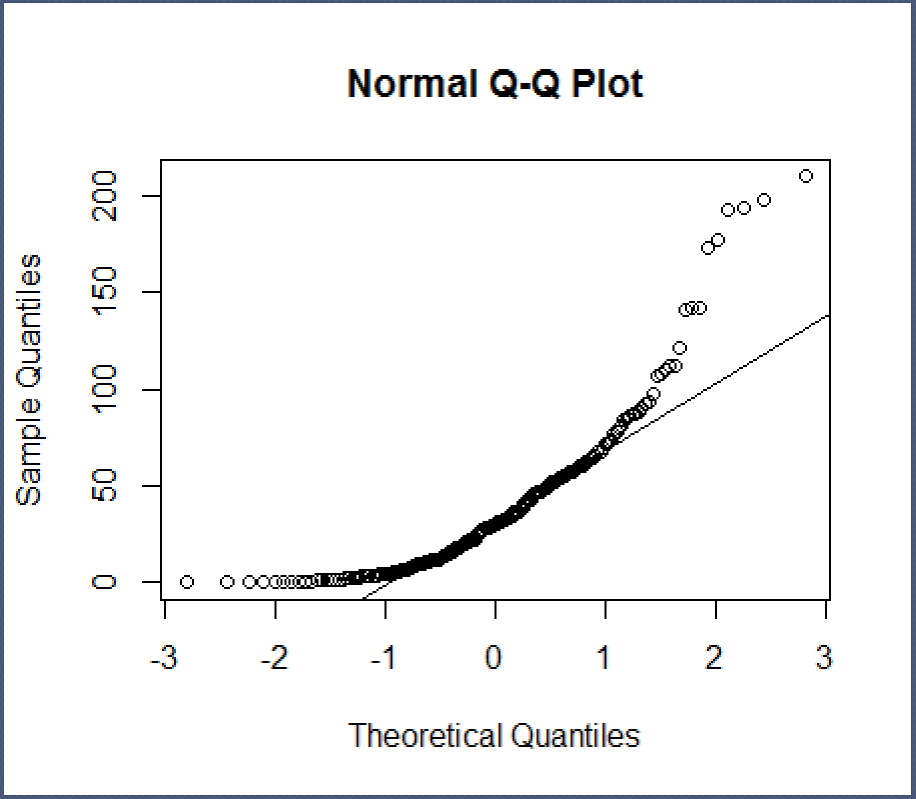
Clean Data: First Look

- Non-normal distribution
- QQplots and

Shapiro-Wilk tests
Confirm this.

See, for example,
'Deletes'

W = 0.82373,
p-value = 2.239e-14



Multiple logistic regression

- Outcome variable: categorical
- Predictor variables: continuous

- Non-normal distribution

- Binomial

Multiple Logistic Regression

- Best model: no over/underfitting
- AIC instead of R-squared
- Deviances: related to model fit
  - "the null deviance measures the discrepancy between the intercept only model and the data and the [residual] deviance measures the discrepancy between the fitted model and the data. In other words, the smaller the deviance is, the better the model fits the data" (Speelman, 2014, p. 24).
  - the lower the better

Assumptions

- Multicolinearity:
  - predictors should not be too highly correlated. VIF.
- Dispersion:
  - there should be no over/underdispersion. Overdispersion: large difference between residual deviance and df → coefficents will seem more confident than they are. Smaller SE's (Nerbonne).
- Linearity of the logit:
  "The linearity assumption in logistic regression is (…) that there is a linear relationship between any continuous predictors and the logit of the outcome variable (…) Look at whether the interaction terms between the predictor and its log transformation is significant" (Field, 2012, p. 321). It should not be significant.

Step(): Model Selection

- Full model → best model
- Forward/backward
- Critique: "a joke, not based on a theory (…) unconstrained coefficients" (Gelman, 2014).
- Backward better, according Field (2012, p. 265)
  - supressor effects: when a predictor is significant only when another is held constant. Forward elimination is more likely to exclude factors involved in suppressor effects and consequently more likely to make a Type II error (saying something is insignificant when it is significant).

- Gives you the model with the lowest AIC . Is it also the best model?

Unclean Data: Full Model

```
Call:
glm(formula = SuccessOfTrial ~ DeletesPerTrial + TypingTimePerTrial +
    CharactersTypedPerTrial + NumberOfTextTurnsPerTrial +
TotalTrialsAttemptedByDyad +
    NoOfQuestionMarksPerTrial, family = "binomial", data = fd)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1658  -1.2026   0.7323   0.7881   1.8802

Coefficients:
                            Estimate Std. Error z value
(Intercept)                1.294e+00  3.662e-01   3.533
DeletesPerTrial            1.284e-02  4.120e-03   3.115
TypingTimePerTrial        -3.084e-06  1.125e-06  -2.742
CharactersTypedPerTrial   -1.071e-04  1.077e-03  -0.099
NumberOfTextTurnsPerTrial -3.565e-02  3.162e-02  -1.127
TotalTrialsAttemptedByDyad -2.700e-03  1.067e-02  -0.253
NoOfQuestionMarksPerTrial  2.477e-02  1.201e-01   0.206
                           Pr(>|z|)
(Intercept)                0.000411 ***
DeletesPerTrial            0.001838 **
TypingTimePerTrial         0.006115 **
CharactersTypedPerTrial    0.920821
NumberOfTextTurnsPerTrial  0.259591
TotalTrialsAttemptedByDyad 0.800234
NoOfQuestionMarksPerTrial  0.836588
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 494.27  on 424  degrees of freedom
Residual deviance: 479.33  on 418  degrees of freedom
AIC: 493.33

Number of Fisher Scoring iterations: 4
```

Unclean Data: Best Model

```
Call:
glm(formula = SuccessOfTrial ~ DeletesPerTrial + TypingTimePerTrial,
    family = "binomial", data = fd)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1340  -1.2202   0.7556   0.7917   1.6524

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)         1.033e+00  1.431e-01   7.219 5.22e-13 ***
DeletesPerTrial     1.024e-02  3.636e-03   2.816  0.00486 **
TypingTimePerTrial -3.369e-06  1.048e-06  -3.213  0.00131 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 494.27  on 424  degrees of freedom
Residual deviance: 482.48  on 422  degrees of freedom
AIC: 488.48

Number of Fisher Scoring iterations: 4
```

- (Best model according to step() also included text turns’. That model had a lower AIC (487.4), but was not significantly different and in order to avoid overfitting, this model was chosen)

ANOVA

- AIC:

  - 493.33 vs. 488.84

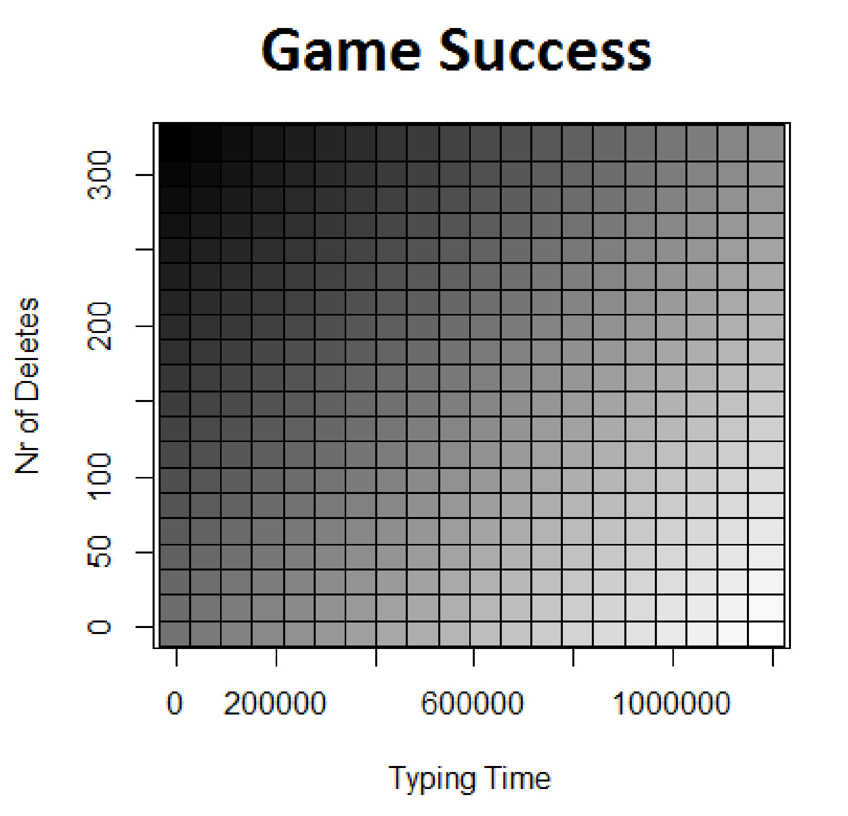
- ANOVA comparison:

  - p: 0.5339

  - no significant difference. However, this model is simpler, and has an AIC of >2 points lower, and is therefore better (no overfitting)

Predictors

- Best predictors: Deletes & Typing Time
- Typing Time: negative
- Deletes: unexpected positive predictor



Clean Data: Full Model

```
Call:
glm(formula = SuccessOfTrial ~ DeletesPerTrial + TypingTimePerTrial +
    CharactersTypedPerTrial + NumberOfTextTurnsPerTrial +
TotalTrialsAttemptedByDyad +
    NoOfQuestionMarksPerTrial, family = "binomial", data = fdc)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1411  -0.9368   0.5916   0.8111   2.5874

Coefficients:
                            Estimate Std. Error z value
(Intercept)                2.890e+00  6.349e-01   4.553
DeletesPerTrial            1.374e-02  6.685e-03   2.055
TypingTimePerTrial        -5.610e-06  2.291e-06  -2.449
CharactersTypedPerTrial    2.911e-03  2.296e-03   1.268
NumberOfTextTurnsPerTrial -1.431e-01  5.693e-02  -2.514
TotalTrialsAttemptedByDyad -5.699e-02 1.821e-02  -3.130
NoOfQuestionMarksPerTrial  1.098e-01  1.719e-01   0.638
                          Pr(>|z|)
(Intercept)               5.29e-06 ***
DeletesPerTrial            0.03985 *
TypingTimePerTrial         0.01433 *
CharactersTypedPerTrial    0.20486
NumberOfTextTurnsPerTrial  0.01194 *
TotalTrialsAttemptedByDyad 0.00175 **
NoOfQuestionMarksPerTrial  0.52320
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 240.38  on 201  degrees of freedom
Residual deviance: 214.70  on 195  degrees of freedom
AIC: 228.7

Number of Fisher Scoring iterations: 4
```

Clean Data: Best Model

```
Call:
glm(formula = SuccessOfTrial ~ DeletesPerTrial + TypingTimePerTrial +
    NumberOfTextTurnsPerTrial + TotalTrialsAttemptedByDyad, family =
"binomial",
    data = fdc)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1215  -1.0081   0.6063   0.7954   2.7030

Coefficients:
                             Estimate Std. Error z value
(Intercept)                 2.940e+00  6.114e-01   4.809
DeletesPerTrial             1.591e-02  6.509e-03   2.445
TypingTimePerTrial         -5.044e-06  2.183e-06  -2.311
NumberOfTextTurnsPerTrial  -7.814e-02  3.249e-02  -2.405
TotalTrialsAttemptedByDyad -5.837e-02  1.793e-02  -3.256
                           Pr(>|z|)
(Intercept)                1.52e-06 ***
DeletesPerTrial             0.01450 *        Way more significant than in
full model
TypingTimePerTrial          0.02085 *        Less significant than in full
model
NumberOfTextTurnsPerTrial   0.01618 *        Less significant than in full
model
TotalTrialsAttemptedByDyad  0.00113 **       More significant than in full
model
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 240.38  on 201  degrees of freedom
Residual deviance: 216.89  on 197  degrees of freedom
AIC: 226.89

Number of Fisher Scoring iterations: 4
```

ANOVA

- AIC:
  - 228.7 vs. 226.89
- ANOVA comparison:
  - p: 0.3358
  - no significant difference. However, this model is simpler and therefore better (no overfitting)
  - smaller model does differ significantly from other smaller models

Assumptions

- Multicolinearity:
  - could not use VIF, but deletes/typing time/tekst turns are highly corrlated ($r > .5$) → assumption violated.

- Dispersion: Binomial response variables rarely lead to over/underdispersion
  - Residual Deviance = 216.89
  - Df = 197
  - No over or underdispersion (also checked this with 'family=quasibinomial'

- Linearity of the logit

Linearity of the logit

```
Call:
glm(formula = SuccessOfTrial ~ DeletesPerTrial + TypingTimePerTrial +
    NumberOfTextTurnsPerTrial + TotalTrialsAttemptedByDyad +
    deleteslogint + timelogint + turnslogint + trialslogint,
    family = binomial, data = fdc)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3739  -0.9522   0.5468   0.7957   1.4307

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                4.547e+00  2.525e+00   1.801   0.0717 .
DeletesPerTrial            1.288e-01  5.681e-02   2.268   0.0233 *
TypingTimePerTrial        -1.456e-04  6.779e-05  -2.148   0.0317 *
NumberOfTextTurnsPerTrial -1.307e-01  2.643e-01  -0.494   0.6211
TotalTrialsAttemptedByDyad -1.795e-01  3.650e-01  -0.492   0.6228
deleteslogint             -2.106e-02  1.052e-02  -2.002   0.0453 *
timelogint                 1.057e-05  5.067e-06   2.086   0.0370 *
turnslogint                1.613e-02  7.182e-02   0.225   0.8223
trialslogint               2.521e-02  8.460e-02   0.298   0.7657
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 226.24  on 191  degrees of freedom
Residual deviance: 199.06  on 183  degrees of freedom
  (10 observations deleted due to missingness)
AIC: 217.06

Number of Fisher Scoring iterations: 4
```

Linearity of the logit

- Deletes & Typing Time interaction terms are in fact significant → problematic
- Linearity of the logit **cannot** be assumed

Model Fit

- AIC: 226.89
- Deviances:

  - Null: 240.38

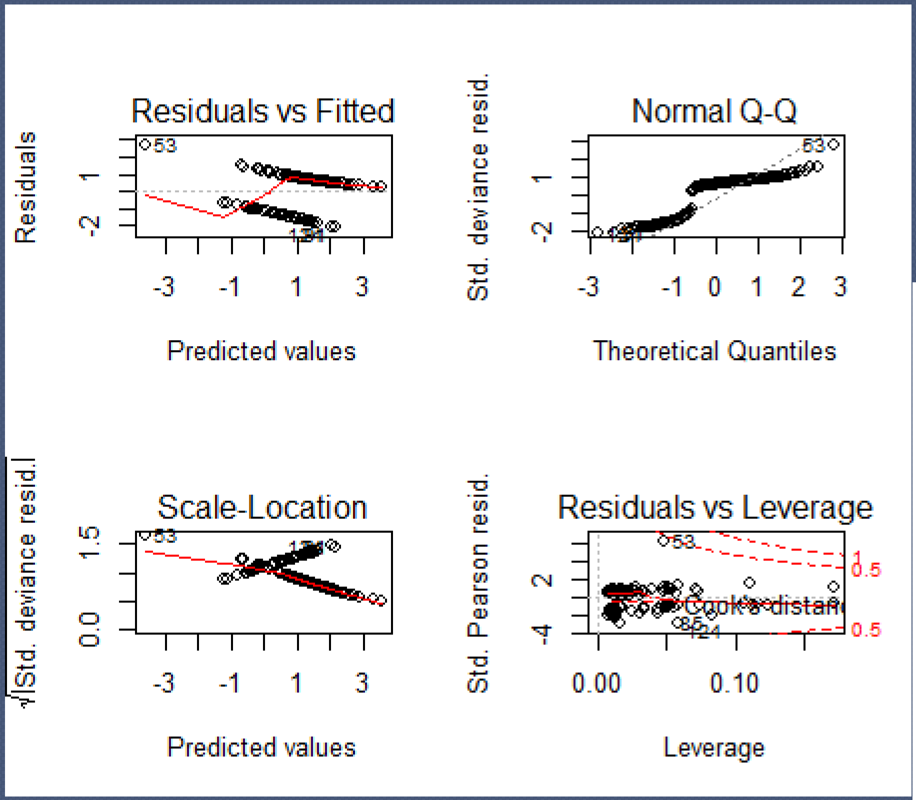
  - Residual: 216.89
- Goodness of fit:

  1 - pchisq(23.49, 4)

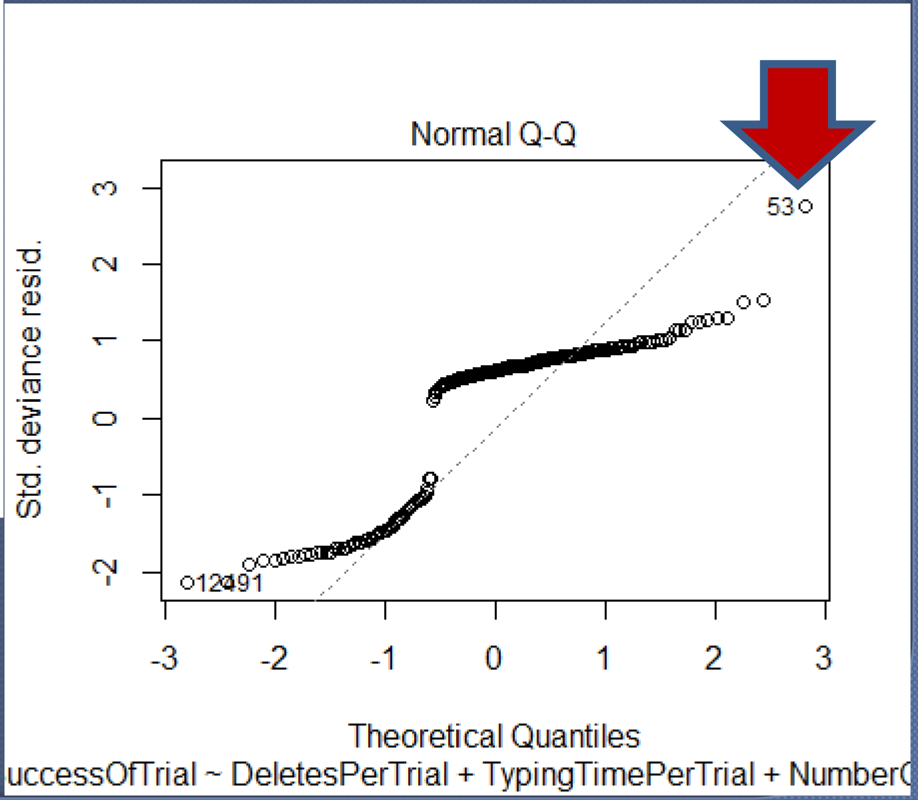
  (difference in deviance, difference in df)

  p = 0.0001010535, better than chance ($p < .05$)

Residuals



- Regression models for binomial response variables are not meant to have normally distributed residuals



Outlier Excluded: Best Model

```
> summary(fdc1.out <- glm(SuccessOfTrial ~ DeletesPerTrial +
TypingTimePerTrial + NumberOfTextTurnsPerTrial +
TotalTrialsAttemptedByDyad, family = "binomial", data = fdc, subset=-53))

Call:
glm(formula = SuccessOfTrial ~ DeletesPerTrial + TypingTimePerTrial +
    NumberOfTextTurnsPerTrial + TotalTrialsAttemptedByDyad, family =
"binomial",
    data = fdc, subset = -53)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2507  -0.7901   0.5630   0.7925   1.6650

Coefficients:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                 3.334e+00  6.429e-01   5.186 2.15e-07 ***
DeletesPerTrial             2.237e-02  7.157e-03   3.126 0.001772 **
TypingTimePerTrial         -8.266e-06  2.656e-06  -3.113 0.001852 **
NumberOfTextTurnsPerTrial  -9.263e-02  3.456e-02  -2.681 0.007350 **
TotalTrialsAttemptedByDyad -6.526e-02  1.847e-02  -3.534 0.000409 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 239.71  on 200  degrees of freedom
Residual deviance: 207.49  on 196  degrees of freedom
AIC: 217.49          LOWEST DEVIANCE AND AIC

Number of Fisher Scoring iterations: 4
```

Assumptions

- Multicolinearity:
  - could not use VIF, but deletes/typing time/tekst turns are highly corrlated (r > .5) → assumption violated.
- Dispersion
  - Residual Deviance = 207.49
  - Df = 196
  - No over or underdispersion
- Linearity of the Logit

Linearity of the Logit

```
Call:
glm(formula = SuccessOfTrial ~ DeletesPerTrial + TypingTimePerTrial +
    NumberOfTextTurnsPerTrial + TotalTrialsAttemptedByDyad +
    deleteslogint + timelogint + turnslogint + trialslogint,
    family = binomial, data = fdc, subset = -53)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1636  -0.7918   0.5525   0.7932   1.7194

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                3.687e+00  2.589e+00   1.424    0.154
DeletesPerTrial            8.853e-02  6.184e-02   1.432    0.152
TypingTimePerTrial        -9.621e-06  9.506e-05  -0.101    0.919
NumberOfTextTurnsPerTrial -1.788e-01  2.756e-01  -0.648    0.517
TotalTrialsAttemptedByDyad -1.691e-01 3.701e-01  -0.457    0.648
deleteslogint             -1.308e-02  1.163e-02  -1.125    0.261
timelogint                 9.655e-08  7.223e-06   0.013    0.989
turnslogint                2.372e-02  7.465e-02   0.318    0.751
trialslogint               2.475e-02  8.576e-02   0.289    0.773

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.60  on 190  degrees of freedom
Residual deviance: 194.09  on 182  degrees of freedom
  (10 observations deleted due to missingness)
AIC: 212.09

Number of Fisher Scoring iterations: 4
```

Linearity of the logit

- None of the interaction terms are significant
- Linearity of the logit **can** be assumed

Model Fit

- AIC = 217.49
- Deviances
  - Null: 239.71
  - Residual: 207.49
- Goodness of fit:
  - 1 - pchisq(32.22, 4)= 1.724908e-06
  - $p < .05$
  - It can be assumed that the model performs better than chance

Report: Best Model

fdc1.out	B (SE)	Lower	95% CI for odds ratio Odds Ratio	Upper
Included				
Constant	3.334e+00 (6.429e-01)	8.3322635	28.0542325	104.8929683
Deletes	2.237e-02 (7.157e-03)	1.0091658	1.0226228	1.0380148
Typing Time	-8.266e-06 (2.656e-06)	0.9999859	0.9999917	0.9999965
Text Turns	-9.263e-02 (3.456e-02)	0.8487203	0.9115323	0.9729605
Trials Attempted	-6.526e-02 (1.847e-02)	0.9024395	0.9368262	0.9707146

Odds ratio: "indicator of the change in odds resulting from a unit change in the predictor" (Field, 2012, p. 319)

Lack of graphs

- “Graphs aren't very useful for showing the results of multiple logistic regression; instead, people usually just show a table of the independent variables, with their *P* values and perhaps the regression coefficients”. (McDonald, 2009)
- Visualisation of a 5-dimensional model (4 predictors, 1 outcome) is often too complicated

Tentative conclusions

- 1. Best Model/predictors:
  - Best AIC: 217.49
  - Includes: Deletes, typing time, nr. of text turns and total trials attempted
- 2. Cleaning up the data:
  - Cleaning the data leads to slightly different and significantly better models than having unclean data
  - Excluding a very influential outlier leads to a significantly lower AIC, lower deviances, and linearity of the logit
- 3. Deletes is a positive predictor in all models: unexpected
- Best predictors:
  - in and of itself, total trials attempted is the best predictor

Questions?

- Shoud I exclude more outliers?
- Exclude total trials attempted? Interpretation of this variable is dodgy…
- Shall I add graphs of the multiple logistic regression model?

References

- Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue.*Cognitive Science*, *27*(2), 195-225 Coltekin, C. (2014). Retrieved from http://coltekin.net/cagri/R/R-exercises.html
- Field, A. (2012). *Discovering statistics using R.* Sage.Chicago.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218.
- Gelman, A. (2014). "Why we hate stepwise regression". Retrieved from http://andrewgelman.com/2014/06/02/hate-stepwise- regression/
- Hayashi, M., Raymond, G., & Sidnell, J. (2013). Conversational repair and human understanding: an introduction. In Hayashi, M., Raymond, G., & Sidnell, J. (Eds.) *Conversational repair and human understanding* (Vol. 30) (pp. 1-40). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511757464.001
- Herring, S. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication*, *4*(4), 0-0.
- McDonald, J. H. (2009). *Handbook of biological statistics* (Vol. 2, pp. 173-181). Baltimore, MD: Sparky House Publishing.Chicago
- Smith, M., Cadiz, J. J., & Burkhalter, B. (2000, December). Conversation trees and threaded chats. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (pp.97-105). ACM.
- Speelman, D. (2014). Logistic regression. *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, *43*, 487. ISO 690