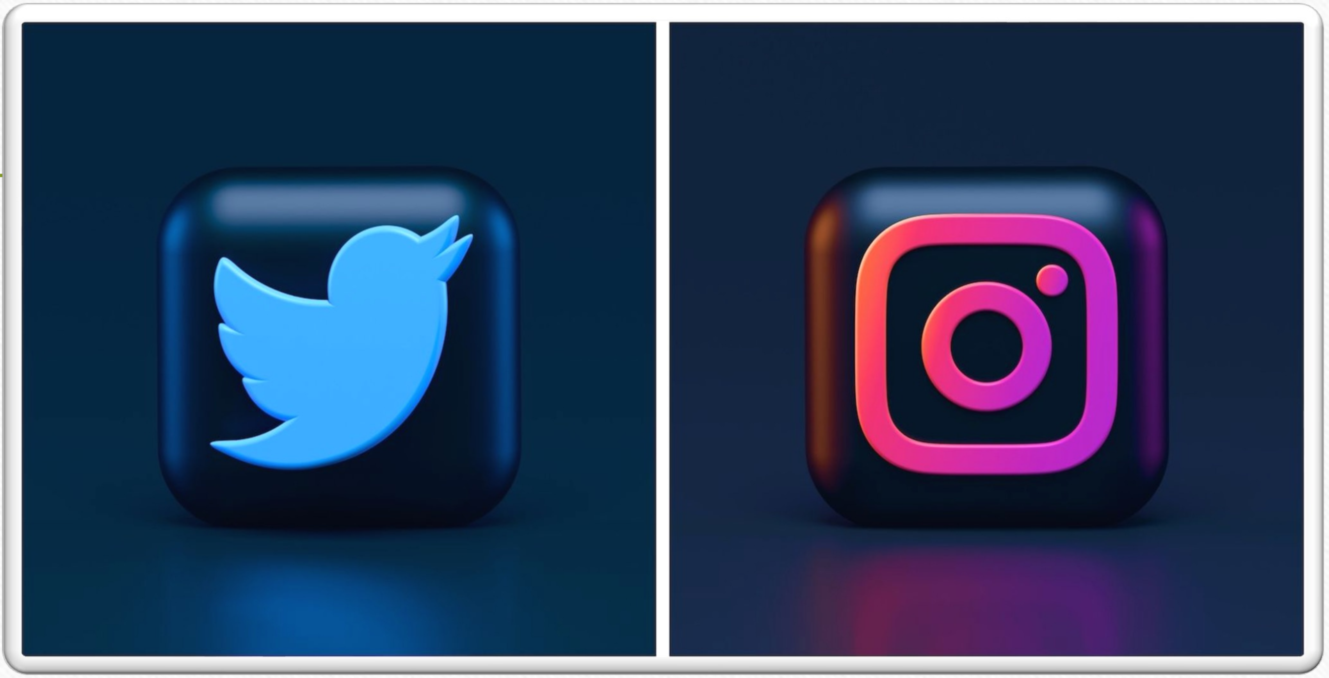
Applied Statistical Analysis II: Replication Study

Caitlín Cooney

Explaining Privacy Control on Instagram and Twitter: The Roles of Narcissism and Self-Esteem

- Nardis, Y., & Panek, E. (2019). Explaining privacy control on Instagram and Twitter: The roles of narcissism and self-esteem. Communication Research Reports, 36(1), 24-34.
- https://doi.org/10.1080/08824096.2018.1555522



The Original Study

- Abstract: A growing body of research examines the relationships between psychological traits and privacy behaviors on Social Networking Sites (SNSs) to understand why users control information about themselves. This study investigates how narcissism and self-esteem can explain tendencies to control privacy on two widely used platforms: Instagram and Twitter.
- Data from an online survey (n = 510) are analyzed using linear mixed models (glmm).
- Exhibitionism, superiority, self-esteem and privacy settings (i.e. public access to profiles) are examined in the context of 2 different types of SNS – Twitter and Instagram.

Testing 4 hypotheses

- *H1: Exhibitionism will be positively related to less privacy control on SNSs.*
- *H2: Superiority will be positively related to less privacy control on SNSs.*
- *H3: Self-esteem will be positively related to more privacy control on SNSs.*
- H4: *The positive relationship between self-esteem and more privacy control will be stronger for Instagram than Twitter.*

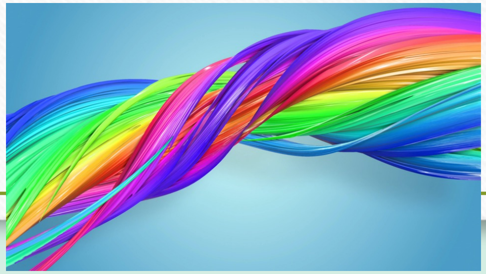
Original Method and Twist Method

GLMM

- Extension of GLM that allows for modelling of correlated data, such as repeated measurements or clustered data, by including random effects in the model.
- Subset the data to exclude anyone who did not use either Twitter or Instagram (which meant they didn't have a baseline of 'no use', but this is acceptable because people who self-select into using social media are likely going to have different levels of narcissism than those who opt out of SNS use).
- The remaining data consisted of people who use *either* Instagram or Twitter, or both.
- Clustering standard errors by ID means that each person is considered as one cluster, with the data points being whether they used one type of social media, or both.
- Because there are some participants who use both forms of SNS, and this violates the assumption of independence, as these observations are coming from the same person, this method of GLMM is one option for examining this relationship.

GLM with SE clustered by ID

- Another method to model binary or categorical outcomes.
- A simpler approach that does not explicitly model the random effects.
- Clustering standard errors by ID essentially groups observations with the same ID or cluster label together and estimates the variance of the parameter estimates within each group. This means that we are accounting for the correlation among observations within the same group or cluster (in this case, use of one or both SNS), which can help to improve the accuracy of our estimates.
- For my twist, each person is considered a single cluster, with one data point for Twitter use and one data point for Instagram use (so each point will only have 1 or 2 data points).



Code and results

Replicated code:

```r
####Mixed models predicting privacy setting####

#Model without interaction terms
mixed_priv <- glmer(privacy ~ Exhibitionism + Superiority + Self_Est
                      SNS + Female + (1|id),
                    data = analysis_reshaped,
                    family = "binomial",
                    glmerControl(optimizer = "bobyqa", optCtrl = lis

summary(mixed_priv)

#Calculating odds ratios
se <- sqrt(diag(vcov(mixed_priv)))
(tab <- cbind(Est = fixef(mixed_priv),
              LL = fixef(mixed_priv) - 1.96 * se,
              UL = fixef(mixed_priv) + 1.96 *se))
exp(tab)
```

Twist code:

```r
# My twist including p-values in code

# Model without interaction terms
glm_priv <- glm(privacy ~ Exhibitionism + Superiority + Self_Esteem + Age + SNS + Female,
                data = analysis_reshaped, family = "binomial")


# Compute coefficient estimates, standard errors, p-values, and odds ratios
coeffs_cl <- coeftest(glm_priv, vcov = vcovCL, cluster = analysis_reshaped$id)
odds_ratios <- exp(coeffs_cl[, "Estimate"])
p_values <- coeffs_cl[, "Pr(>|z|)"]

# Combine coefficients, odds ratios, and p-values into final table
coeffs_cl <- cbind(coeffs_cl[, c("Estimate", "Std. Error", "z value")],
                   Odds_Ratio = odds_ratios,
                   p_value = p_values)

coeffs_cl
```

Replicated results:

> results

	Coefficients	Standard Error	Z-value	Odds Ratio	p-values
(Intercept)	4.14033159	2.9730166	1.3926365	62.8236498	1.637297e-01
Exhibitionism	0.26782394	0.1154458	2.3199098	1.3071170	2.034576e-02
Superiority	-0.34552064	0.1244398	-2.7766078	0.7078517	5.492942e-03
Self_Esteem	-0.55806808	0.2533935	-2.2023769	0.5723137	2.763869e-02
Age	-0.04071846	0.1477484	-0.2755932	0.9600994	7.828605e-01
SNSInstagram	-1.27291390	0.2085753	-6.1028994	0.2800145	1.041614e-09
FemaleFemale	-1.29814863	0.3032550	-4.2807165	0.2730368	1.862925e-05

Twist results:

> coeffs_cl

	Estimate	Std. Error	z value	Odds_Ratio	p_value
(Intercept)	2.87377763	2.23646682	1.2849632	17.7037703	1.637297e-01
Exhibitionism	0.18869366	0.08241361	2.2895935	1.2076709	2.034576e-02
Superiority	-0.24073737	0.09247883	-2.6031620	0.7860480	5.492942e-03
Self_Esteem	-0.39793228	0.18425185	-2.1597193	0.6717075	2.763869e-02
Age	-0.02464955	0.11310892	-0.2179276	0.9756518	7.828605e-01
SNSInstagram	-0.91416326	0.12638941	-7.2329101	0.4008519	1.041614e-09
FemaleFemale	-0.91794590	0.21710033	-4.2282106	0.3993385	1.862925e-05

Interpretation

- Comparing the results of the two methods, we can see that they are quite similar. The direction and magnitude of the coefficients are generally consistent between the two models, with some small variation.
- The p-values are generally similar between the two models. Both models find statistically significant effects of Exhibitionism, Superiority, Self-Esteem, use of Instagram, and being female, while Age is not significant in either model.
- As Exhibitionism increases by one unit, the likelihood of having public settings increases by 1.3071170 in the GLMM model and 1.2076709 in the GLM model, providing support for H1.
- For a one-unit increase in "Superiority", the odds of having a public setting (compared to a private setting) decrease by a factor of 0.7078517 in the GLMM model and 0.77860480 in the GLM model. Thus, H2 is not supported.
- H3 expected self-esteem to be positively related to more privacy control on SNSs. As self-esteem increases, the likelihood of having public settings decreases by 0.5723137 in the GLMM model and 0.6717075 in the GLM model, providing support for H3.

Original

> results

	Coefficients	Standard Error	Z-value	Odds Ratio	p-values
(Intercept)	4.14033159	2.9730166	1.3926365	62.8236498	1.637297e-01
Exhibitionism	0.26782394	0.1154458	2.3199098	1.3071170	2.034576e-02
Superiority	-0.34552064	0.1244398	-2.7766078	0.7078517	5.492942e-03
Self_Esteem	-0.55806808	0.2533935	-2.2023769	0.5723137	2.763869e-02
Age	-0.04071846	0.1477484	-0.2755932	0.9600994	7.828605e-01
SNSInstagram	-1.27291390	0.2085753	-6.1028994	0.2800145	1.041614e-09
FemaleFemale	-1.29814863	0.3032550	-4.2807165	0.2730368	1.862925e-05

Twist

> coeffs_cl

	Estimate	Std. Error	z value	Odds_Ratio	p_value
(Intercept)	2.87377763	2.23646682	1.2849632	17.7037703	1.637297e-01
Exhibitionism	0.18869366	0.08241361	2.2895935	1.2076709	2.034576e-02
Superiority	-0.24073737	0.09247883	-2.6031620	0.7860480	5.492942e-03
Self_Esteem	-0.39793228	0.18425185	-2.1597193	0.6717075	2.763869e-02
Age	-0.02464955	0.11310892	-0.2179276	0.9756518	7.828605e-01
SNSInstagram	-0.91416326	0.12638941	-7.2329101	0.4008519	1.041614e-09
FemaleFemale	-0.91794590	0.21710033	-4.2282106	0.3993385	1.862925e-05

Interactions Replication and Twist

- In the original study, separate models were run with interaction effects between each component of narcissism and SNS.
- In keeping with this, I ran a glm with standard errors clustered by ID for each interaction.
- The results remained nonsignificant, supporting the study's findings that SNSs do not moderate the relationship between Exhibitionism and privacy setting, nor between Superiority and privacy setting.
- The interaction between self-esteem and SNS does indicate that the association between self-esteem and privacy control differed significantly across Instagram and Twitter, continuing to provide support for H4 when using a GLM.
- Twitter and Instagram did not differ in the likelihood of having public settings at the lowest level of self-esteem. However, as self-esteem increases, the likelihood of having public settings on Instagram is significantly lower than on Twitter.

Original:

```r

#Plotting Self_Esteem*SNS interaction
pred_mixed_priv_inter3 <- predict(mixed_priv_inter3, type="response")
summary(pred_mixed_priv_inter3)


# Please note that this code did not run when I was replicating the data
# Below I have included some edited code that allowed me to plot this interaction
self_esteem_df <- data.frame(Self_Esteem = analysis_reshaped$Self_Esteem,
                             pred_mixed_priv_inter3 = pred_mixed_priv_inter3, SNS = analysis_reshaped$SNS)

ggplot(data = self_esteem_df, aes(x = Self_Esteem, y = pred_mixed_priv_inter3, linetype = SNS)) +
  geom_smooth(method = "lm", col="black") +
  geom_blank() +
  labs(x = "Self Esteem", y = "Predicted Probability (Public Setting)") +
  xlim(1,4) +
  ylim(0,1)
```

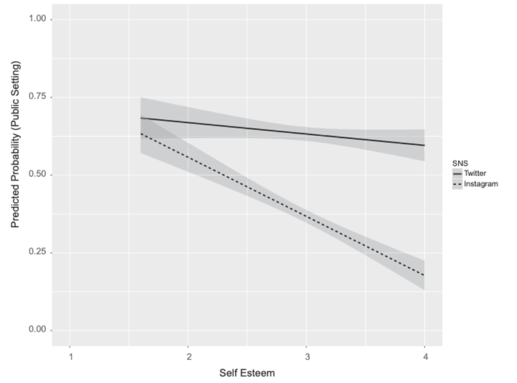
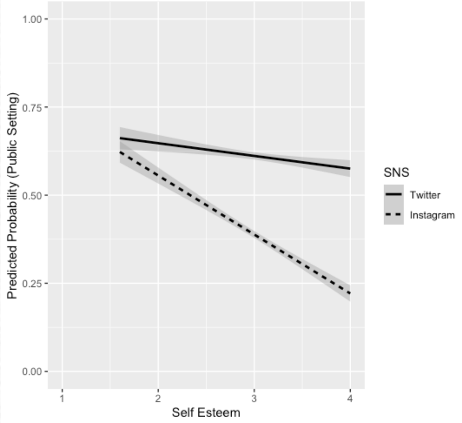


Figure 1 SNS as the moderator of the relationship between self-esteem and public setting.

Twist:

```r

#Model with Self_Esteem*SNS interaction
glm_priv_inter3 <- glm(privacy ~ Exhibitionism + Superiority + Age + Female + Self_Esteem*SNS, data = analy
summary(glm_priv_inter3)

# Compute coefficient estimates, standard errors, p-values, and odds ratios
coeffs_inter3 <- coeftest(glm_priv_inter3, vcov = vcovCL, cluster = analysis_reshaped$id)
odds_ratios_inter3 <- exp(coeffs_inter3[, "Estimate"])
p_values_inter3 <- coeffs_inter3[, "Pr(>|z|)"]

# Combine coefficients, odds ratios, and p-values into final table
coeffs_inter3 <- cbind(coeffs_inter3[, c("Estimate", "Std. Error", "z value")],
                       Odds_Ratio = odds_ratios_inter3,
                       p_value = p_values_inter3)

coeffs_inter3


#Plotting Self_Esteem*SNS interaction
pred_glm_priv_inter3 <- predict(glm_priv_inter3, type="response")
summary(pred_glm_priv_inter3)


# Please note that this code did not run when I was replicating the data
# Below I have included some edited code that allowed me to plot this interaction
self_esteem_df_glm <- data.frame(Self_Esteem = analysis_reshaped$Self_Esteem,
                                 pred_glm_priv_inter3 = pred_glm_priv_inter3, SNS = analysis_reshaped$SNS)

ggplot(data = self_esteem_df_glm, aes(x = Self_Esteem, y = pred_glm_priv_inter3, linetype = SNS)) +
  geom_smooth(method = "lm", col="black") +
  geom_blank() +
  labs(x = "Self Esteem", y = "Predicted Probability (Public Setting)") +
  xlim(1,4) +
  ylim(0,1)
```



Conclusion

- In conclusion, when dealing with binary data that violates the assumption of independence, both GLM and GLMM models can be used to account for the correlation among observations within clusters.
- My replication contribution shows that using a GLM with standard errors clustered by ID is an effective and simpler approach for modelling such data, as it can achieve similar results to a GLMM while requiring fewer assumptions and computational resources.
- Therefore, the use of a clustered GLM can be a valuable tool for researchers and practitioners working with binary data that exhibit clustering or dependence among observations.