**Final Project Statistical Interpretation**

In our research project, we have decided to conduct a Pearson's Chi-squared test with Yates' continuity correction to determine the statistical significance of the difference between two categorical variables which represent the use of two different languages in expressing irony.

Model 1: Finding statistical significance of language difference

The result:

Pearson's Chi-squared test with Yates' continuity correction

data: irony\_language

X-squared = 11.666, df = 1, p-value = 0.0006365

> print(chisq\_lang$expected)

kz cn

serious 86.59504 82.40496

ironic 161.40496 153.59504

Interpretation:

1. First of all, it is important to know the null thesis for the Chi-square test; That is, there is no relationship between the two categorical variables.
2. The p-value’s typical significance level is 0.05, if the p-value is larger than 0.05 then it suggests that there is strong evidence to conform the null hypothesis of no association. In contrast, if the p-value is smaller than 0.05 then the null hypothesis should be rejected.
3. P-value in this case: The p-value (0.0006365) is the probability of observing the data or more extreme data under the null hypothesis that there is no association between the variables. A low p-value indicates strong evidence against the null hypothesis.
4. Therefore, there is a statistically significant difference between the use of the two languages in expressing irony.

Model 2: Finding statistical significance of age difference

The result:

Pearson's Chi-squared test with Yates' continuity correction

data: irony\_age

X-squared = 0.63341, df = 1, p-value = 0.4261

> print(chisq\_age$expected)

below 30 above 30

serious 141.4153 27.58471

ironic 263.5847 51.41529

Interpretation:

Given the p-value of 0.4261, which is greater than the typical significance level of 0.05, we fail to reject the null hypothesis. Therefore, there is insufficient evidence to conclude that there is a significant association between age groups and the expression of irony. In other words, there is no statistically significant difference in the distribution of "serious" and "ironic" responses between age groups "below 30" and "above 30".

Model 3: Finding statistical significance between CN and KZ no emoji difference

The result:

Pearson's Chi-squared test with Yates' continuity correction

data: irony\_noemoji

X-squared = 0.6321, df = 1, p-value = 0.4266

Interpretation:

Given the p-value of 0.4266, which is greater than the typical significance level of 0.05, we fail to reject the null hypothesis. Therefore, there is insufficient evidence to conclude that there is a significant association between the use of "no emoji" expressions in Kazakh and Chinese languages.

Model 4: Finding statistical significance between CN and KZ emoji difference

The result:

Pearson's Chi-squared test with Yates' continuity correction

data: irony\_emoji\_kzcn

X-squared = 11.302, df = 1, p-value = 0.0007742

Interpretation:

Given the p-value of 0.0007742, which is lower than the critical value of 0.05, we can successfully reject the null thesis. Therefore, we can conclude that there is a statistically significant difference in the distribution of "serious" and "ironic" responses based on the use of emojis in Kazakh and Chinese languages.

In order to make sure that the Kazakh language and Chinese language indeed play a role in affecting the perception of irony, we have also run the logistic regression using a generalized linear mixed model (GLMM) to examine the relationship between the independent variables "kz" (Kazakh language) and "cn" (Chinese language).

Model 5: Finding statistical significance of language difference using logistic regression

The result:

Formula: Answer ~ kz + cn + (1 | response)

Data: final\_irony

AIC BIC logLik deviance df.resid

615.9 628.4 -304.9 609.9 481

Scaled residuals:

Min 1Q Median 3Q Max

-1.1531 -0.7033 -0.5370 1.1153 1.8622

Random effects:

Groups Name Variance Std.Dev.

response (Intercept) 0.2517 0.5017

Number of obs: 484, groups: response, 61

Fixed effects:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.0446 0.1791 -5.833 5.45e-09 \*\*\*

kz 0.7169 0.2390 2.999 0.00271 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr)

kz -0.741

fit warnings:

fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

> exp(0.7169)

[1] 2.048074

Interpretation:

1. The number "0.00271" followed by "\*\*" indicates the p-value associated with the coefficient estimate for the predictor variable "kz" in the logistic regression model.
2. As we know that if p is smaller than 0.05 then we will have reason to reject the null thesis.
3. Therefore, there is a statistically significant difference between the use of the two languages in expressing irony.