Application of a Novel Quantile-Based Statistical Test to Reaction Times Data

Nick Chandler, Caleb Fuemmeler, Fiona Cleary, Samuel Burnett, Fernando Marmolejo-Ramos, Kimihiro Noguchi

Western Washington University and University of South Australia

PiMUC 2024

April 13, 2024

Outline

Dataset

What is our dataset?

Description of data:

- Sourced from the paper "Your Face and Moves Seem Happier When I Smile" [1].
- Concerned with the influence of facial action on perception.
- Participants are asked to evaluate a series of images, determining whether they appear happy or sad.
- Participants are either given a pen to hold in their teeth, forcing a smile or are administered the test with no control of facial expression.

Structure of analyses:

- Reaction times of participants.
- Stratified by nationality (Japanese & Swedish) and biological gender into 4 groups.
- Compared under 6 conditions (with/without pen) x (happy, neutral, sad faces).
- We apply a multi-sample test with all-pairwise comparisons to detect differences between these groups in any of the conditions.

Dataset

Happy, neutral, and sad faces

Here is how we defined happy, neutral, and sad faces:
Happy (7–10), neutral (4–6), and sad (0–3).

Figure: Happy, neutral, and sad faces

- Each participant's task was to indicate, as quickly and accurately as possible, whether the facial expression portrayed a "sad" or a "happy" emotional state while holding a pen in their teeth or no pen at all.
- The 11 faces were presented twice in random order, across 7 blocks containing all 11 stimuli. Each stimulus was shown 77 times in each of the 2 conditions (with or without pen in the participant's teeth), for a total of $22 \times 7 = 154$ trials.

Dataset

Relevant Statistic

- The statistic we use in our analysis, which we call "log range", is the natural log of the range of reaction times for each participant, given by $\ln\left(\max\left(RT\right) - \min\left(RT\right)\right)$.
- This statistic uses extreme values and addresses individual differences in reaction times.
- This use of the range of participants is novel to cognitive psychology research. Typically, the mean reaction time is used when multiple reaction times are available from the same participant.
- We conduct the test on the 10th, 25th, 50th, 75th, and 90th percentiles of the distribution of this statistic.

The Distribution We Assume for Log Range

The Gumbel Distribution: The Gumbel distribution, denoted by Gumbel(μ, β), $\mu \in \mathbb{R}$, $\beta > 0$, can be used to approximate the distribution of the minimum or maximum of a random sample.
Its PDF is:

$$f(x; \mu, \beta) = \frac{1}{\beta} e^{-(z+e^{-z})} \text{ , where, } z = \frac{x - \mu}{\beta} \text{ and, } x \in \mathbb{R}$$

.
Its CDF is:

$$F(x; \mu, \beta) = e^{-e^{-(x-\mu)/\beta}} \text{ and, } x \in \mathbb{R}$$

.
Median: $\mu - \beta \ln(\ln(2))$.
Mean: $E[X] = \mu + \beta\gamma$, where γ is the Euler-Mascheroni Constant.
Variance: $\sigma^2 = \frac{\pi^2}{6}\beta^2$.

Finding Parameters of the Gumbel Distribution

Maximum Likelihood Estimation:

$$L(\mu, \beta; x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{e^{-(z_i + e^{-z_i})}}{\beta}$$

The log-likelihood is given by:

$$-n \ln(\beta) - \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\beta}\right) + e^{-\left(\frac{x_i - \mu}{\beta}\right)}$$

Thus, after arg-maxing, we obtain:

$$\hat{\mu} = -\hat{\beta} \ln \left(\frac{\sum_{i=1}^{n} e^{-x_i/\hat{\beta}}}{n} \right)$$

$$\hat{\beta} = \bar{X} - \frac{\sum_{i=1}^{n} x_i e^{x_i/\hat{\beta}}}{\sum_{i=1}^{n} e^{-x_i/\hat{\beta}}}$$

Method-of-Moments Estimation:

$$\frac{\sum_{i=1}^{n} x_i}{n} = \mu_1(\hat{\theta}) = \hat{\mu} + \hat{\beta}\gamma,$$

$$\frac{\sum_{i=1}^{n} x_i^2}{n} = \mu_2(\hat{\theta}) = \hat{\sigma}^2 + \hat{E[X]}^2$$

$$= \frac{\pi^2}{6}\hat{\beta}^2 + \left(\hat{\mu} + \hat{\beta}\gamma\right)^2 = \frac{\pi^2}{6}\hat{\beta}^2 + \bar{x}^2$$

Therefore,

$$\hat{\mu} = \bar{x} - \hat{\beta}\gamma,$$

$$\hat{\beta} = \sqrt{\frac{6}{\pi^2}\left(\frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2\right)}$$

See [2] for details.

Suitability of the Data

Does the data satisfy the required assumptions?

To determine whether our dataset follows a Gumbel Distribution, we use Q-Q plots and Kolmogorov-Smirnov tests.

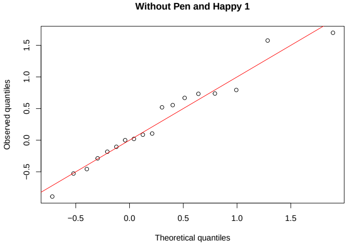


Figure: Q-Q plot of Group 1.

Kolmogorov-Smirnov Test

- H_0 : The dataset follows a Gumbel Distribution.
- H_A : The dataset does not follow a Gumbel Distribution.

Test Statistic:

$$D_n = \sup_x |F_n(x) - F(x)|$$

Results (Without Pen and Happy):

1. 0.8123 – Fail to Reject H_0
2. 0.5359 – Fail to Reject H_0
3. 0.5014 – Fail to Reject H_0
4. 0.7464 – Fail to Reject H_0

The Null and Alternative Hypotheses

Let $F^{-1}(p)$ represent the $100p$-th percentile of Gumbel(0, 1). Then, the $100p$-th percentile of Gumbel(μ_i, β_i) is given by $\mu_i + \beta_i F^{-1}(p)$, $i = 1, \ldots, 4$.

The null and alternative hypotheses are given by:

$$H_0^{(i_1,i_2)}: \mu_{i_1} + \beta_{i_1} F^{-1}(p) = \mu_{i_2} + \beta_{i_2} F^{-1}(p),\ i_1 \neq i_2,$$
$$H_A^{(i_1,i_2)}: \mu_{i_1} + \beta_{i_1} F^{-1}(p) \neq \mu_{i_2} + \beta_{i_2} F^{-1}(p),\ i_1 \neq i_2.$$

$$H_0 = \bigcap H_0^{(i_1,i_2)}: \mu_1 + \beta_1 F^{-1}(p) = \cdots = \mu_4 + \beta_4 F^{-1}(p),$$
$$H_A = \bigcup H_A^{(i_1,i_2)}: \mu_{i_1} + \beta_{i_1} F^{-1}(p) \neq \mu_{i_2} + \beta_{i_2} F^{-1}(p) \text{ for some } (i_1, i_2).$$

The Test Statistic

The test statistic for comparing the i_1-th and i_2-th sample is given by

$$T^{(i_1,i_2)} = \frac{[\hat{\mu}_{i_1} + \hat{\beta}_{i_1} F^{-1}(p)] - [\hat{\mu}_{i_2} + \hat{\beta}_{i_2} F^{-1}(p)]}{\sqrt{\sum_{j=1}^{2} \left[\widehat{\mathrm{Var}}(\hat{\mu}_{i_j}) + [F^{-1}(p)]^2 \widehat{\mathrm{Var}}(\hat{\beta}_{i_j}) + 2F^{-1}(p)\widehat{\mathrm{Cov}}(\hat{\mu}_{i_j}, \hat{\beta}_{i_j})\right]}},$$

where $(\hat{\mu}_{i_j}, \hat{\beta}_{i_j})$, $j = 1, 2$, are method-of-moment or maximum likelihood estimator or (μ_{i_j}, β_{i_j}).

Note that $\boldsymbol{T} = [T^{(2,1)}, T^{(3,1)}, T^{(4,1)}, T^{(3,2)}, T^{(4,2)}, T^{(4,3)}]'$ is approximately multivariate-t under H_0, i.e., $\boldsymbol{T} \sim \mathcal{T}_{6,\nu}(\mathbf{0}, \boldsymbol{\Sigma})$ approximately for some covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν.

The Degrees of Freedom Calculation

The degrees of freedom ν for the multivariate t-distribution is computed by applying an extension of Welch's degrees of freedom. Specifically, $\nu = \max\{1, \min\{\nu_1, \nu_2, \ldots, \nu_6\}\}$, where

$$\nu_j = \frac{\left(\sum_{i=1}^{4} k_{j,i}\hat{\beta}_i^2\right)^2}{\sum_{i=1}^{4} \frac{(k_{j,i}\hat{\beta}_i^2)^2}{n_i - 1}},$$

$$k_{j,i} = \frac{c_{j,i}^2}{n_i}[1.168 + 1.100\{F^{-1}(p)\}^2 + 0.096F^{-1}(p)],$$

and that $c_{j,i}^2 = 1$ if the j-th pair contains the i-th group and 0 otherwise. Moreover, n_i represents the i-th sample size. The constants in the equation above can be obtained from [2].

The Adjusted p-Value Computation

The (approximate) adjusted p-values, $p_{\text{adj},z}^{(i_1,i_2)}$, are computed by

$$p_{\text{adj},t}^{(i_1,i_2)} = 1 - \Pr\left(\bigcap\left\{|V^{(i_1,i_2)}| \leq t^{(i_1,i_2)}\right\}\right),$$

where $[V^{(2,1)}, V^{(3,1)}, V^{(4,1)}, V^{(3,2)}, V^{(4,2)}, V^{(4,3)}]' \sim \mathcal{T}_{6,\nu}(\mathbf{0}, \hat{\mathbf{\Sigma}})$. Here, $\hat{\mathbf{\Sigma}}$ represents the estimated covariance matrix of $\mathbf{\Sigma}$.

Similarly, 95% simultaneous confidence intervals are given by

$$T_{\text{num}}^{(i_1,i_2)} \pm t_{0.95,2,\nu,\hat{\mathbf{\Sigma}}} T_{\text{den}}^{(i_1,i_2)},$$

where $T_{\text{num}}^{(i_1,i_2)}$ and $T_{\text{den}}^{(i_1,i_2)}$ represent the numerator and denominator of $T^{(i_1,i_2)}$, respectively. Moreover, $t_{0.95,2,\nu,\hat{\mathbf{\Sigma}}}$ represents the two-sided 95-th percentile of $\mathcal{T}_{6,\nu}(\mathbf{0}, \hat{\mathbf{\Sigma}})$.

Power Curve Plots

- All-pairwise comparisons of the 25^{th} percentile of each group.
- Samples sizes are all equal to 10.
- $\binom{4}{2} = 6$ comparisons.
- H_0 is true when $x = 0$.
- Note that the multivariate t-approximation with the method-of-moment estimates (Power.t.MoM in light green) is the most robust method. The other estimation methods tend to be more liberal.

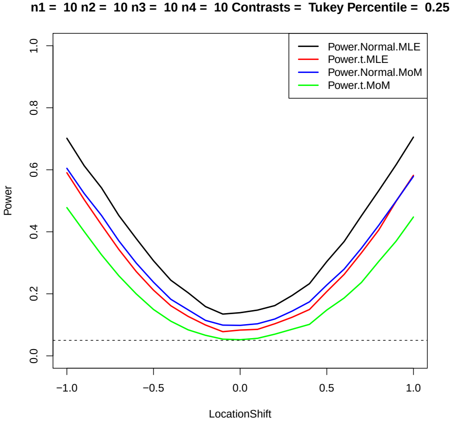
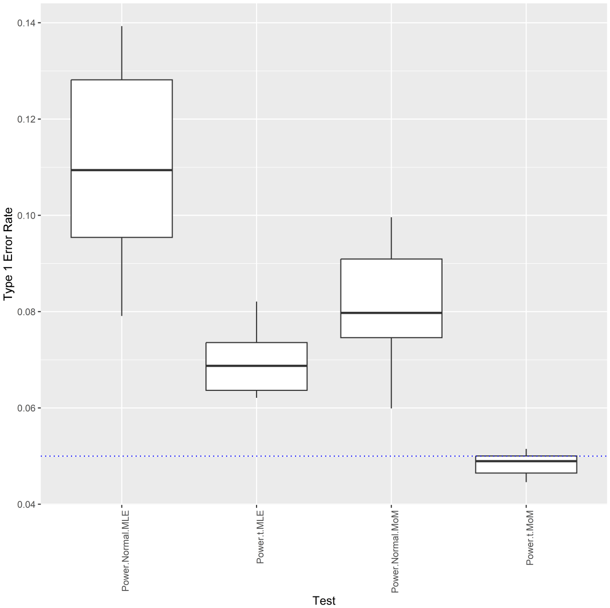


Figure: Power curves of four methods.

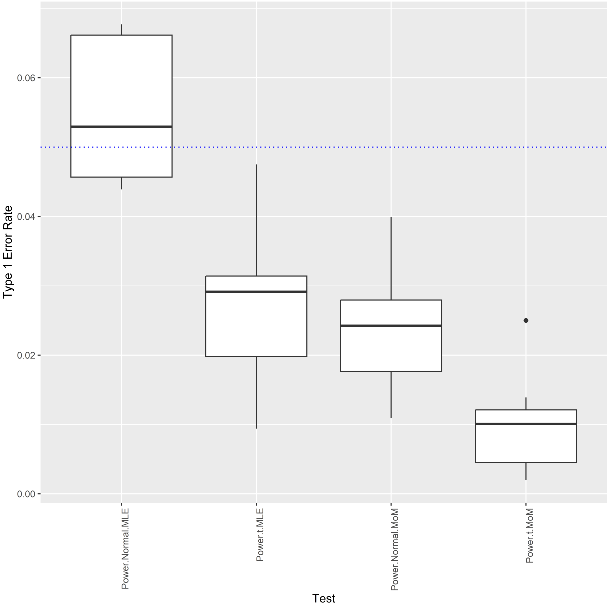
Simulation Study Results

- Case 2, 25^{th} percentile type 1 error rate box plots
- At the 25^{th} percentile, Power t MoM performs the best.



Simulation Study Results

- Case 2, 90^{th} percentile type 1 error rate box plots
- At higher percentiles Power t MLE performs better



Results

We examine the results for the condition: Without pen and happy.

- Group 1: Japanese Female ($n = 18$)
- Group 2: Japanese Male ($n = 20$)
- Group 3: Swedish Female ($n = 22$)
- Group 4: Swedish Male ($n = 14$)

10th Percentile

Comparison	2 vs 1	3 vs 1	4 vs 1	3 vs 2	4 vs 2	4 vs 3
p-Value	0.6204	0.7752	0.1956	0.1847	**0.0354**	0.4846
Test Statistic	-1.2098	0.9476	2.0420	2.0737	**2.8937**	1.4332

25th Percentile

Comparison	2 vs 1	3 vs 1	4 vs 1	3 vs 2	4 vs 2	4 vs 3
p-Value	0.7911	0.8778	0.1362	0.3960	**0.0474**	0.2555
Test Statistic	-0.9183	0.7386	2.2402	1.5890	**2.7597**	1.8823

Visualization of Differences

- Empirical CDFs (top row)
- Estimated CDFs assuming Gumbel distribution (middle row)
- Quantile Plots assuming Gumbel distribution (bottom row)
- The 10th and 25th percentiles show significant differences (bottom right).

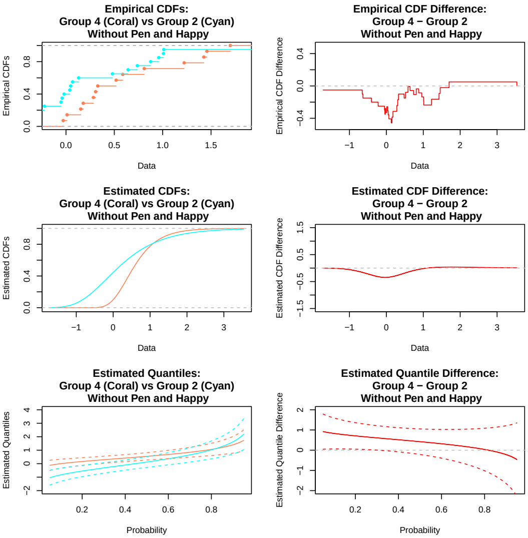


Figure: Plots of Group 4 vs. Group 2

Test Conclusions

From the results in the previous sections, we can make the following conclusions on the without pen and happy condition:

1. We reject H_0 in the comparison of Groups 2 and 4 in the 10^{th} percentile.
2. We reject H_0 in the comparison of Groups 2 and 4 in the 25^{th} percentile.

That is, in each of these, we see a statistically significant difference in the specified percentiles between Groups 2 and 4 at $\alpha = 0.05$. No other pairwise comparisons are statistically significant.

Interpretation

We can interpret each of the statistically significant results as follows:

1. There are statistically significant differences between Swedish male (Group 4) and Japanese male (Group 2) when comparing individuals with highly consistent reaction times (i.e., those with smaller reaction time ranges, which correspond to the lower percentiles).
2. In particular, the differences are found for the recognition of happy faces while not holding a pen in their lips.

Conclusion & Future Work

Conclusions:

- We have developed a new hypothesis test for comparing quantiles of multiple, independent samples assuming Gumbel distribution.
- The multivariate T approximation with the method-of-moments estimation seems to work well.
- We apply a new type of statistic, called the log range, to this test to compare different percentiles.

Future Work:

- Investigate further to find out why the power curves tend to be more liberal when the maximum likelihood estimation is applied.
- Identify statistics other than the log range that follow the Gumbel distribution.
- Extend our hypothesis test to the factorial designs.

References

1. Marmolejo-Ramos F., Murata A., Sasaki K., Yamada Y., Ikeda A., Hinojosa J.A., Watanabe K., Parzuchowski M., Tirado C., Ospina R. (2020). Your face and moves seem happier when I smile: Facial action influences the perception of emotional faces and biological motion stimuli. *Experimental Psychology* 67(1), 14–22.

2. Phien, H.N. (1987). A review of methods of parameter estimation for the extreme value type-1 distribution. *Journal of Hydrology* 90, 251–268.