Replication of Hsieh et al. (2023): The Role of Chromatic Cues in Facial Age Perception

Shemsi Dkhissy, Hrvoje Žlepalo, Berk Ceylan, Youssef Barrima, and Khaled Miraoui

École polytechnique fédérale de Lausanne (EPFL)

HUM-403: Experimental Cognitive Psychology

Dr. Domicele Jonauskaite

December 21, 2025

Abstract

This study replicates Experiments 1 and 2 of **Hsieh et al. (2023)** to evaluate the role of chromatic cues in facial age estimation. While structural features like wrinkles are known predictors of age, the contribution of holistic skin colour remains less understood. Using a within-subjects design in a controlled laboratory setting ($N = 40$), we presented 200 facial stimuli from the "**Chicago Face Database**" (Ma et al., 2015) manipulated in the **CIE Lab*** colour space. The study aims to verify two key mechanisms: (1) whether the presence of colour information reduces **inter-observer variability** in age estimates compared to greyscale images, and (2) whether manipulating the red-green axis systematically **biases** age perception, with increased redness signaling youth. By controlling for environmental settings and making the experiment in an offline setting and monitor calibration—limitations present in the original online study—we test the robustness of these chromatic effects. We discuss how these findings suggest that the brain uses skin colour as a direct cue for health and youth, rather than relying on facial shape alone.

Replication of Hsieh et al. (2023): The Role of Chromatic Cues in Facial Age Perception

Determining the age of another person is a rapid, automatic perceptual process that underpins much of human social interaction (**Hsieh et al., 2023**; Mayes et al., 2010; Nikitin et al., 2024; Porcheron et al., 2013). Perceived age dictates not only social hierarchy and etiquette but also legal responsibilities and perceived reproductive fitness (Christensen et al., 2009; Lazzari et al., 2025; Sanghi, 2022; Voelkle et al., 2012). Given this ecological importance, vision science has long sought to identify the specific visual cues that drive age estimation (George and Hole, 1995). Traditionally, research has focused on structural and morphological cues (Burt and Perrett, 1995). We know, for instance, that macroscopic changes such as the deepening of wrinkles, the sagging of soft tissue, and the changing shape of cartilaginous features like the nose and ears are potent indicators of aging (Baldasso et al., 2019; Mark et al., 1980; Porcheron et al., 2013). While these "topographic changes are undeniably critical" (Porcheron et al., 2013), they do not act in isolation. "The surface properties of the skin—specifically its pigmentation and colour distribution—offer a parallel stream of information" (Fink et al., 2006) that the visual system utilizes to gauge age, yet this channel remains less understood compared to structural markers, , likely due to the methodological challenge of isolating subtle chromatic cues from texture in naturalistic stimuli.

The role of colour in face perception is well-established in domains such as health and attractiveness (Fink and Matts, 2008). Evolutionary psychology suggests that skin coloration serves as a reliable signal of physiological status; for example, increased skin redness is frequently associated with cardiovascular health and oxygenation (Re et al., 2011; Stephen et al., 2009), while yellowness often signals a robust immune system and a diet rich in carotenoids (Stephen et al., 2011). However, the specific contribution of these chromatic cues to age perception remains a subject of ongoing investigation. Previous work has often conflated skin texture (smoothness) with skin colour (Fink et al., 2006) or

restricted analysis to specific features, such as the luminance contrast between the lips and the surrounding skin (Porcheron et al., 2013). It remains an open question whether the holistic colour information of a face—independent of its structural features—serves as a reliable cue for age (Russell et al., 2019), and whether specific colour channels (such as the red-green axis) actively bias an observer's judgment (**Hsieh et al., 2023**; Porcheron et al., 2017).

Understanding the role of colour is important because it changes how we think about age perception. If colour alone can make a face look younger, it means our brains do not just look at wrinkles and face shape to decide how old someone is. Instead, the visual system likely combines shape information with colour cues—specifically the idea that redder skin looks healthier and therefore younger. This study aims to test this specific effect. By removing shape and texture differences from our images, we can test if skin colour is just a minor detail or a major signal that the brain uses to judge age.

The Replicated Study

The present study serves as a direct replication of Experiments 1 and 2 by **Hsieh et al. (2023)**, titled "Colour information biases facial age estimation and reduces inter-observer variability." In their original work, the authors investigated the influence of holistic chromatic information on age estimation by systematically manipulating facial images within the CIE Lab* colour space in an online setting. This approach allowed them to dissociate colour information from luminance to test its independent effects. In Experiment 1, they presented participants with facial stimuli in two conditions: full colour and greyscale (with chromatic channels a^* and b^* removed). This was designed to test whether the presence of colour information aids observers in converging on a consistent age estimate. In Experiment 2, they specifically manipulated the red-green (a^*) channel, creating versions of faces with increased or decreased redness to test for directional bias in age perception.

Their findings revealed distinct roles for colour in face processing. First, in

Experiment 1, they demonstrated that the mere presence of colour information significantly reduced inter-observer variability compared to greyscale images. Specifically, the standard deviation of age ratings was lower for colour images, suggesting that colour provides a "stabilizing" cue that helps different observers agree on a target's age. However, they found no significant bias in the median age rating between colour and greyscale conditions, indicating that the mere presence of colour does not automatically make a face look younger or older—it simply makes the judgment more precise. Second, in Experiment 2, they found a systematic bias: faces with increased red-green contrast (appearing "redder") received significantly lower median age estimates than those with decreased contrast. These results imply that facial redness is likely processed as a specific cue for youthfulness, potentially due to its association with vascular health, while general colour availability primarily serves to reduce ambiguity in the estimation process.

The Current Study

This replication adopts the core methodology of the main paper (**Hsieh et al., 2023**), but transitions the experimental environment from an online platform to a controlled laboratory setting. This modification represents a shift from "replication" to "adaptation" in terms of delivery, though the experimental logic remains identical. The shift to a laboratory setting addresses a key limitation often found in online colour perception studies: the lack of control over monitor calibration and ambient lighting. By strictly controlling the viewing environment, we aim to reduce noise and ensure that the subtle chromatic manipulations, particularly the red-green contrast changes, are perceived accurately by all participants. Given this controlled setting and the expected reduction in outlier data compared to the original online sample, we will utilise the mean rather than the median for our analysis of central tendency. Our primary research question remains consistent with the original authors: How does colour information affect age estimation?

Our procedure involves a within-subjects design where participants view a sequence of 200 facial stimuli derived from the Chicago Face Database (Ma et al., 2015). We have

maintained the original study's stimulus processing pipeline, converting images to CIE Lab* space using a Python script (see Appendix A) to create distinct variations per face as needed: Original Colour, Greyscale, Increased Red-Green Contrast, and Decreased Red-Green Contrast. Participants will view each face for 2000ms and provide an estimated age. We have maintained the sample size and counterbalancing strategies to ensure statistical power comparable to the original work.

Based on the significant findings of **Hsieh et al. (2023)**, we propose the following directional hypotheses:

- **H1 (Experiment 1 - Variability):** We predict that the standard deviation of age estimates will be significantly lower for faces shown in Colour compared to Greyscale.
- **H2 (Experiment 2 - Bias):** We predict that faces with increased Red-Green contrast will have a significantly lower mean estimated age than faces with decreased Red-Green contrast.
- **H3 (Experiment 1 - Mean Age):** We expect no significant difference in the mean estimated age between the Colour and Greyscale conditions.
- **H4 (Experiment 2 - Variability):** We expect no significant difference in the standard deviation of age estimates between the increased and decreased Red-Green contrast conditions.

By rigorously testing these hypotheses, this study aims to cement the understanding of colour as a primary, rather than secondary, input in the complex cognitive calculation of age.

Acknowledgements. We acknowledge the use of Google Gemini (Model 2.5 Flash) to assist with the refinement of the grammatical structure and paraphrasing some parts to improve clarity and flow of this report. All narrative content, scientific argumentation, and original writing were performed solely by the authors.

Method

Participants

We plan to recruit a total of 40 participants ($N = 40$) from the EPFL and UNIL community, as well as from our personal acquaintances. We aim to achieve an even gender split to ensure a balanced sample. The recruitment will be conducted in person for an offline laboratory setting. Given that the experiment is estimated to take approximately 20–30 minutes per person, we anticipate that recruitment may be challenging due to the time commitment required. To incentivize participation, we plan to offer a small food reward (e.g., chocolate) upon completion. All participants will be naive to the specific hypotheses of the study regarding colour and age perception.

Materials

The stimuli for this study were selected from the Chicago Face Database (Ma et al., 2015). We selected a stratified random subset of 200 high-resolution images of neutral-expression faces to serve as the base stimuli. To ensure demographic diversity consistent with the original study, the stimulus set was composed of 30% Asian, 30% White, 15% Latino, and 15% Black faces, balanced for gender (50% female, 50% male). These images were then processed using a custom Python script (see Appendix A) to generate the specific chromatic conditions required for our replication (see Figure 1). The script converted the original images from RGB to the CIE Lab* colour space, allowing us to manipulate colour channels independently of luminance.

For Experiment 1, we created two versions of each face: a "Full Colour" version and a "Greyscale" version (where the a^* and b^* channels were set to zero). For Experiment 2, we created two modified versions based on the red-green axis: an "Increased Contrast" version (where the a^* channel values were multiplied by 1.5) and a "Decreased Contrast" version (where the a^* channel values were divided by 1.5).

Figure 1

Example of Experimental Stimuli Conditions

(a) *Full Colour* (b) *GreyScale* (c) *RG+* (d) *RG-*

Note. Stimuli adapted from the Chicago Face Database (Ma et al., 2015). Digital manipulations follow the protocol of **Hsieh et al. (2023)**.

The experiment was programmed using the jsPsych library De Leeuw, 2015, a JavaScript framework for creating behavioral experiments in a web browser. The study will be run on a standard laptop computer with a colour-calibrated screen to ensure accurate rendering of the chromatic manipulations.

Procedure

The experiment follows a within-subjects design, meaning each participant will complete both Experiment 1 and Experiment 2 in a single session. The session will take place in a quiet room with controlled lighting to minimize glare and colour distortion on the screen. Upon arrival, participants will be briefed on the general nature of the experiment, which is to estimate the age of faces and type a "2-digit number," and will provide informed consent. They will then be seated in front of the laptop at a viewing distance of approximately 60 cm. The experiment is divided into two main parts, the order of which will be counterbalanced across participants.

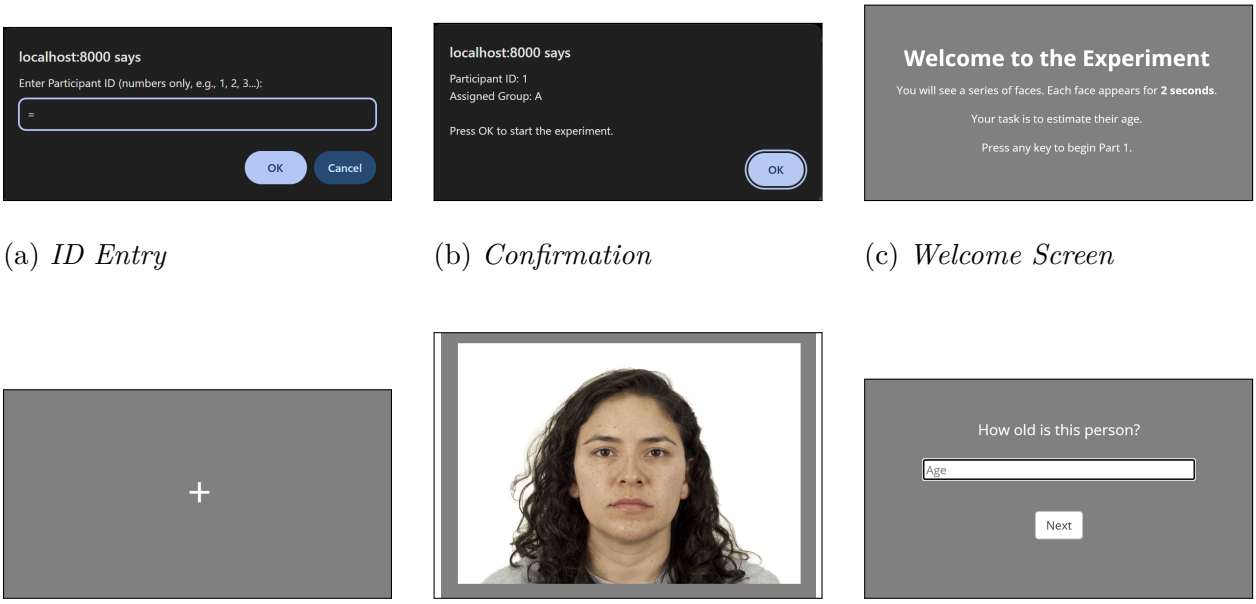
Experiment 1 (Colour vs. Greyscale): Participants will view a series of 100 trials. In each trial, a fixation cross will appear for 500 ms, followed by a target face presented for 2000 ms. After the face disappears, a text box will appear on the screen, prompting the participant to type their estimate of the person's age in years. The stimuli

in this experiment will be a mix of Full Colour and Greyscale images, presented in a randomized order.**Experiment 2 (Red-Green Contrast):** The procedure for the second block is identical to the first, but the stimuli will consist of 100 trials featuring faces with either Increased or Decreased Red-Green contrast.

We implemented a counterbalancing system based on participant ID numbers (odd/even), assigning participants to either Group "**A**" or "**B**" to determine the specific subset of faces shown in each condition, ensuring that no participant sees the exact same face in two different conditions. The total session duration is expected to be approximately 20 minutes. Figure 2 illustrates the complete experimental workflow, including the initial setup screens and the repeated trial sequence.

Figure 2

Experimental Procedure and Trial Sequence Screens



(a) *ID Entry* (b) *Confirmation* (c) *Welcome Screen*

(d) *Fixation Cross* (e) *Stimulus (2000ms)* (f) *Response Window*

Note. The top row displays the one-time setup screens per participant. The bottom row shows the sequence of a single trial, which repeats 100 times per experiment and each participant complete experiment "1" and "2".

Planned Statistical Analysis

The independent variables of the experiments are selected images from the Chicago Face Database, which differ across experiments, together with their condition (depending on the experiment, that is either **Chromatic Condition** with two levels: Full Colour and Greyscale. for the first one, or **Red-Green Contrast** using CIE Lab* with two levels: Increased (RG+) and Decreased (RG-). for the second one). Next, the dependent variables are then, of course, the perceived ages the participants label the given images. The results of these measurements will be treated as in the original paper, with the exception of taking averages of participants' guesses per image, instead of the medians. The reason for this substitution is that it is more plausible to not have exclusively integers as data points for performing the statistical tests, as the tests assume data is coming from a normal distribution, with some possibly unknown parameters. This decision is also motivated by the transition to a controlled laboratory environment, which is expected to minimize the extreme outliers and noise characteristic of online data collection. So, the mean will serve as a more sensitive metric for detecting subtle shifts in central tendency.

For the purposes of testing the bias of image condition on the perception of age in both experiments, we employ paired sample **t**-tests (see Appendix B), with the paired data being average perceived ages per image, with different conditions of the same CFD image forming a pair. As for the inter-observer variability, the paired sample t-test will prove helpful once again, with the difference being that our data points are now sample standard deviations of our list of guesses for each image in each condition. The significance level of all the statistical tests done in our replication will be equal to 95%.

Transparency, Openness, and Reproducibility

The design and analysis plan for this replication were pre-registered on OSF and can be accessed at this link. All experimental stimuli, Chicago Face Database data set, the presentation code (jsPsych), and the Python script used for image generation are available at this link.

References

Baldasso, R. P., Damascena, N. P., Deitos, A. R., Machado, C. E. P., Franco, A., & de Oliveira, R. N. (2019). Morphologic alterations ear, nose and lip detected with aging through facial photoanthropometric analysis. *The Journal of forensic odonto-stomatology*, *37*(2), 25.

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *259*(1355), 137–143.

Christensen, K., Thinggaard, M., McGue, M., Rexbye, H., Aviv, A., Gunn, D., van der Ouderaa, F., Vaupel, J. W., et al. (2009). Perceived age as clinically useful biomarker of ageing: Cohort study. *Bmj*, *339*.

De Leeuw, J. R. (2015). Jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*(1), 1–12.

Fink, B., Grammer, K., & Matts, P. J. (2006). Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces. *Evolution and Human Behavior*, *27*(6), 433–442.

Fink, B., & Matts, P. J. (2008). The effects of skin colour distribution and topography cues on the perception of female facial age and health. *Journal of the European Academy of Dermatology and Venereology*, *22*(4), 493–498.

George, P. A., & Hole, G. J. (1995). Factors influencing the accuracy of age estimates of unfamiliar faces. *Perception*, *24*(9), 1059–1073.

Hsieh, J. Y., Boyce, W. P., Goddard, E., & Clifford, C. W. (2023). Colour information biases facial age estimation and reduces inter-observer variability. *Scientific Reports*, *13*(1), 13224.

Lazzari, E., Compans, M.-C., & Beaujouan, E. (2025). Change in the perceived reproductive age window and delayed fertility in europe [Publisher: Routledge

_eprint: https://doi.org/10.1080/00324728.2023.2298678]. *Population Studies*, *79*(1), 81–101. https://doi.org/10.1080/00324728.2023.2298678

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*, 1122–1135.

Mark, L. S., Pittenger, J. B., Hines, H., Carello, C., Shaw, R. E., & Todd, J. T. (1980). Wrinkling and head shape as coordinated sources of age-level information. *Perception & Psychophysics*, *27*(2), 117–124. https://doi.org/10.3758/BF03204298

Mayes, A. E., Murray, P. G., Gunn, D. A., Tomlin, C. C., Catt, S. D., Wen, Y. B., Zhou, L. P., Wang, H. Q., Catt, M., & Granger, S. P. (2010). Environmental and lifestyle factors associated with perceived facial age in chinese women [Publisher: Public Library of Science]. *PLOS ONE*, *5*(12), e15270. https://doi.org/10.1371/journal.pone.0015270

Nikitin, J., Luchetti, M., Alimanovic, N., Rupprecht, F. S., & Sutin, A. R. (2024). Are social interactions perceived as more meaningful in older adulthood? *Psychology and Aging*, *39*(6), 599.

Porcheron, A., Mauger, E., & Russell, R. (2013). Aspects of facial contrast decrease with age and are cues for age perception [Publisher: Public Library of Science]. *PLOS ONE*, *8*(3), e57985. https://doi.org/10.1371/journal.pone.0057985

Porcheron, A., Mauger, E., Soppelsa, F., Liu, Y., Ge, L., Pascalis, O., Russell, R., & Morizot, F. (2017). Facial contrast is a cross-cultural cue for perceiving age. *Frontiers in Psychology*, *8*, 1208.

Re, D. E., Whitehead, R. D., Xiao, D., & Perrett, D. I. (2011). Oxygenated-blood colour change thresholds for perceived facial redness, health, and attractiveness. *PloS one*, *6*(3), e17859.

Russell, R., Batres, C., Courrèges, S., Kaminski, G., Soppelsa, F., Morizot, F., & Porcheron, A. (2019). Differential effects of makeup on perceived age [_eprint:

https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/bjop.12337]. *British Journal of Psychology*, *110*(1), 87–100. https://doi.org/10.1111/bjop.12337

Sanghi, S. (2022). Age as a site of law's meaning-making practices. *Jindal Global Law Review*, *13*(2), 221–229. https://doi.org/10.1007/s41020-022-00185-6

Stephen, I. D., Coetzee, V., Law Smith, M., & Perrett, D. I. (2009). Skin blood perfusion and oxygenation colour affect perceived human health. *PLoS ONE*, *4*(4), e5083. https://doi.org/10.1371/journal.pone.0005083

Stephen, I. D., Coetzee, V., & Perrett, D. I. (2011). Carotenoid and melanin pigment coloration affect perceived human health. *Evolution and Human Behavior*, *32*(3), 216–227. https://doi.org/10.1016/j.evolhumbehav.2010.09.003

Voelkle, M. C., Ebner, N. C., Lindenberger, U., & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychology and aging*, *27*(2), 265.

Appendix A

Image Conversion Script

Below is the Python code used to convert the Chicago Face Database images into the four experimental conditions (Color, Greyscale, RG+, RG-).

```
import os
import cv2
import numpy as np
from skimage import color, io, img_as_ubyte

# --- CONFIGURATION ---
INPUT_FOLDER = 'raw_images'
OUTPUT_FOLDER = 'processed_faces'

# Ensure output folder exists
if not os.path.exists(OUTPUT_FOLDER):
    os.makedirs(OUTPUT_FOLDER)

# Get list of all images
files = [f for f in os.listdir(INPUT_FOLDER) if f.endswith(('.jpg', '.png'
    , '.jpeg'))]
print(f"Found {len(files)} images. Processing...")

for filename in files:
    # 1. Load Image (Standard RGB)
    img_path = os.path.join(INPUT_FOLDER, filename)
    img_rgb = io.imread(img_path)

    # Get the base filename without extension (e.g., "CFD-WM-001")
    base_name = os.path.splitext(filename)[0]

    # --- VERSION 1: ORIGINAL COLOR (Just copy/save) ---
    # We save it as standard RGB to ensure format consistency
```

```
    save_path = os.path.join(OUTPUT_FOLDER, f"{base_name}_color.jpg")
    io.imsave(save_path, img_rgb)

    # 2. Convert to CIE LAB for manipulation
    # skimage converts images to float64 range [0, 100] for L, [-128, 127] for a/b
    img_lab = color.rgb2lab(img_rgb)

    # --- VERSION 2: GREYSCALE (Exp 1) ---
    # Keep L* (channel 0), Set a* (channel 1) and b* (channel 2) to 0
    lab_grey = img_lab.copy()
    lab_grey[:, :, 1] = 0 # a* = 0
    lab_grey[:, :, 2] = 0 # b* = 0

    # Convert back to RGB
    img_grey = color.lab2rgb(lab_grey)
    # Save (img_as_ubyte converts 0-1 float back to 0-255 image)
    io.imsave(os.path.join(OUTPUT_FOLDER, f"{base_name}_grey.jpg"), img_as_ubyte(img_grey))


    # --- VERSION 3: RED-GREEN INCREASED (Exp 2 RG+) ---
    # Multiply a* channel by 1.5
    lab_rg_plus = img_lab.copy()
    lab_rg_plus[:, :, 1] = lab_rg_plus[:, :, 1] * 1.5

    # Convert back and Save
    # Note: conversion handles clipping automatically if colors go out of bounds
    img_rg_plus = color.lab2rgb(lab_rg_plus)
    io.imsave(os.path.join(OUTPUT_FOLDER, f"{base_name}_rg_plus.jpg"), img_as_ubyte(img_rg_plus))

```

```
57
58      # --- VERSION 4: RED-GREEN DECREASED (Exp 2 RG-) ---
59      # Divide a* channel by 1.5
60      lab_rg_minus = img_lab.copy()
61      lab_rg_minus[:, :, 1] = lab_rg_minus[:, :, 1] / 1.5
62
63      # Convert back and Save
64      img_rg_minus = color.lab2rgb(lab_rg_minus)
65      io.imsave(os.path.join(OUTPUT_FOLDER, f"{base_name}_rg_minus.jpg"),
        img_as_ubyte(img_rg_minus))
66
67  print("Done! Check the 'processed_faces' folder.")
```

Appendix B

Statistical Analysis code

Below is the R code that will be used in the statistical analysis and handling of the data provided by the participants.

```
means_unmodified<-sample(18:67,100,replace=TRUE)
means_modified<-sample(18:67,100,replace=TRUE)
variances_unmodified<-sample(9:36,100,replace=TRUE)
variances_modified<-sample(9:36,100,replace=TRUE)

mA<-matrix(nrow=20,ncol=100)
mB<-matrix(nrow=20,ncol=100)
dfA<-data.frame(mA)
dfB<-data.frame(mB)
s="IM"

rnAB<-c(1:100)
for (i in 1:100){
  rnAB[i]=paste(c(s,as.character(i)),collapse="")
}
rownames(dfA)=c(1:20)
rownames(dfB)=c(1:20)
colnames(dfA)=rnAB
colnames(dfB)=rnAB

for (i in 1:100){
  dfA[,i]=trunc(rnorm(n=20,mean=means_unmodified[i],sd=sqrt(variances_
   unmodified[i])))
  dfB[,i]=trunc(rnorm(n=20,mean=means_modified[i],sd=sqrt(variances_
   modified[i])))
}
write.csv(dfA,file="TrialA",row.names=FALSE)
dfA=read.csv(file="TrialA",header=TRUE)
```

```
write.csv(dfB,file="TrialB",row.names=FALSE)
dfB=read.csv(file="TrialB",header=TRUE)

dfA[is.na(dfA)]=0
dfB[is.na(dfB)]=0
avg_unmodified=c(1:100)
avg_modified=c(1:100)
for (i in 1:100){
  avg_unmodified[i]=sum(dfA[,i])/(20-sum(dfA[,i]==0))
  avg_modified[i]=sum(dfB[,i])/(20-sum(dfB[,i]==0))
}
avg_differences=avg_unmodified-avg_modified
t.test(avg_differences,alternative="two.sided",conf.level=0.95)

sd_unmodified=c(1:100)
sd_modified=c(1:100)
for (i in 1:100){
  sd_unmodified[i]=0
  sd_modified[i]=0
  j=1
  while(dfA[j,i]!=0 && j<=20){
    sd_unmodified[i]=sd_unmodified[i]+(dfA[j,i]-avg_unmodified[i])*(dfA[j,
   i]-avg_unmodified[i])
    j=j+1
  }
  j=1
  while(dfB[j,i]!=0 && j<=20){
    sd_modified[i]=sd_modified[i]+(dfB[j,i]-avg_modified[i])*(dfB[j,i]-avg
   _modified[i])
    j=j+1
  }
  sd_unmodified[i]=sqrt((1/(20-sum(dfA[,i]==0)-1))*sd_unmodified[i])
  sd_modified[i]=sqrt((1/(20-sum(dfB[,i]==0)-1))*sd_modified[i])
```

```
}
sd_differences=sd_unmodified-sd_modified
t.test(sd_differences,alternative="two.sided",conf.level=0.95)
```