# Epistemic Test

20/4/2022

## Research Questions

1. What is the relations between the similarity of the job in low status job, the similarity of the job in high status job, status of the job in similar job, and status of the job in different job, according to the perception of the claims made by the workers?
2. Does the more similar the job is, the more respondents can correctly testify the claims made by the job holder?
3. Does the perception/testification of the claims made by the workers in various conditions affected by gender and age?
4. Does the perception/testification of the claims made by the workers in various conditions affected by the status, similarity, gender, and age?

## Identifying Relations according to the Perception of the claims made by workers

```
head (Epistemic_Test,6)
```

```
## # A tibble: 6 x 7
##   RaterID AVG_1 AVG_2 AVG_3 AVG_4   Age Gender
##     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>
## 1       1   5.5  2     6.75  4.25    54      1
## 2       2   7.5  6.75  5.5   7.75    51      2
## 3       3   7    6     7.25  5.75    31      2
## 4       4   7    6     8.5   7       31      2
## 5       5   7    5     8     5.75    52      1
## 6       6   7    6.5   7.75  7.75    65      2
```

```
cor(Epistemic_Test,method=c("pearson"))
```

```
##              RaterID        AVG_1       AVG_2        AVG_3       AVG_4
## RaterID  1.000000000 -0.041369753 -0.02571448  0.008489672 -0.02375506
## AVG_1   -0.041369753  1.000000000  0.38369848  0.557077136  0.29322401
## AVG_2   -0.025714484  0.383698477  1.00000000  0.239447999  0.58669174
## AVG_3    0.008489672  0.557077136  0.23944800  1.000000000  0.24222250
## AVG_4   -0.023755056  0.293224011  0.58669174  0.242222495  1.00000000
## Age     -0.121480254 -0.003374841  0.10182132 -0.078716069  0.21632362
## Gender   0.074381681 -0.037119204  0.03983584  0.045282889  0.17033962
##                  Age       Gender
## RaterID -0.121480254  0.074381681
## AVG_1   -0.003374841 -0.037119204
## AVG_2    0.101821318  0.039835839
## AVG_3   -0.078716069  0.045282889
## AVG_4    0.216323622  0.170339622
## Age      1.000000000 -0.003861318
## Gender  -0.003861318  1.000000000
```

Between the similarity of the job in low status job (AVG_1 and AVG_2), there is positive relation. It means

that the more respondents answering "surely true" to the claims made by the highly similar job in low status job, the higher is the perception of the claims made by workers in low similar job in low status job.

Between the similarity of the job in high status job (AVG_3 and AVG_4), there is positive relation. It means that the more respondents answering "surely true" to the claims made by the highly similar job in high status job, the higher is the perception of the claims made by workers in low similar job in high status job.

Between the status of the job in similar job (AVG_1 and AVG_3), there is positive relation. It means that the more respondents answering "surely true" to the claims made by high status job in similar job, the higher is the perception of the claims made by workers in low status job in similar job.

Between the status of the job in different job (AVG_2 and AVG_4), there is positive relation. It means that the more respondents answering "surely true" to the claims made by high status job in different job, the higher is the perception of the claims made by workers in low status job in different job.

## Statistical Significance of the testification of the claims made by the job holder

Assumption: It is assumed that the claims are true

```
Epistemic_Test.long <- Epistemic_Test %>%
  gather(key = "group", value = "score", AVG_1, AVG_2)
res <- t.test(score ~ group, data = Epistemic_Test.long, paired = TRUE, alternative="less")
res
```

```
##
##  Paired t-test
##
## data:  score by group
## t = 24.851, df = 600, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 1.716978
## sample estimates:
## mean of the differences
##                1.610233
```

```
Epistemic_Test.long <- Epistemic_Test %>%
  gather(key = "group", value = "score", AVG_3, AVG_4)
res <- t.test(score ~ group, data = Epistemic_Test.long, paired = TRUE, alternative="less")
res
```

```
##
##  Paired t-test
##
## data:  score by group
## t = 11.987, df = 600, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 1.005894
## sample estimates:
## mean of the differences
##               0.8843594
```

It is hypothesized that the more similar the job is, the respondents can correctly the claims made by the job holder. With p-value more than 0.05, the H0 is failed to be rejected, which means that the respondents indeed can correctly testify the claim made by the job holder the more similar the job is.

### Gender and age factor on the testification made by the respondents of the claims

From the P-value below, it is known that the testification score on job claims for low status and highly similar job of the respondents are not affected by gender and age.

```
m <- polr(as.factor(CODE_1) ~ Gender + Age, data = Epistemic_Test_Code, Hess=TRUE)
ctable <- coef(summary(m))
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p))
```

```
##               Value  Std. Error    t value      p value
## Gender -0.131367693 0.145781627 -0.9011265 3.675210e-01
## Age    -0.001527222 0.005713576 -0.2672971 7.892404e-01
## 1|2    -3.173546065 0.368537356 -8.6111924 7.230455e-18
## 2|3    -0.376073235 0.326289665 -1.1525748 2.490850e-01
```

From the P-value below, it is known that the testification score on job claims for low status and low similarity job of the respondents are affected by age. For older respondents, the odds of being more likely to testify the claim as "surely true" is 0.00098 times [i.e., 1/1.014] that of younger respondents, holding constant all other variables.

```
sa <- polr(as.factor(CODE_2) ~ Gender + Age, data = Epistemic_Test_Code, Hess=TRUE)
ctable <- coef(summary(sa))
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p))
```

```
##             Value  Std. Error    t value      p value
## Gender 0.17157440 0.144536140 1.18706916 2.352003e-01
## Age    0.01353188 0.005719622 2.36586986 1.798777e-02
## 1|2    0.02123628 0.320622804 0.06623447 9.471912e-01
## 2|3    2.69849642 0.344372233 7.83598723 4.651732e-15
```

```
#exp(cbind(OR = coef(sa), ci))
```

From the P-value below, it is known that the testification score on job claims for high status and highly similar job of the respondents are affected by age. For older respondents, the odds of being more likely to testify the claim as "surely true" is 1.015 times [i.e., 1/0.985] that of younger respondents, holding constant all other variables.

```
m <- polr(as.factor(CODE_3) ~ Gender + Age, data = Epistemic_Test_Code, Hess=TRUE)
ctable <- coef(summary(m))
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p))
```

```
##              Value Std. Error    t value      p value
## Gender  0.01615963 0.14612370  0.1105887 9.119425e-01
## Age    -0.01482510 0.00572171 -2.5910254 9.569042e-03
## 1|2    -3.14365238 0.35947765 -8.7450566 2.229040e-18
## 2|3    -0.67922300 0.32612411 -2.0827132 3.727737e-02
```

```
#exp(cbind(OR = coef(m), ci))
```

From the P-value below, it is known that the testification score on job claims for high status and lowsimilar job of the respondents are affected by age and gender. For older respondents, the odds of being more likely to testify the claim as "surely true" is 0.00096 times [i.e., 1/1.031] that of younger respondents, holding constant all other variables. For specific gender respondents, the odds of being more likely to testify the claim as "surely true" is 0.621 times [i.e., 1/1.610] that of another gender respondents, holding constant all other variables.

```
m <- polr(as.factor(CODE_4) ~ Gender + Age, data = Epistemic_Test_Code, Hess=TRUE)
ctable <- coef(summary(m))
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p))
```

```
##             Value  Std. Error   t value      p value
## Gender 0.47637392 0.146290446 3.2563570 1.128517e-03
## Age    0.03046043 0.005810331 5.2424596 1.584501e-07
## 1|2    0.24524175 0.324333063 0.7561417 4.495642e-01
## 2|3    2.89608545 0.349050155 8.2970467 1.067314e-16
```

```
#exp(cbind(OR = coef(m), ci))
```

## Gender, age, job similarity, and job status factors on the testification made by the respondents of the claims

From the P-value below, it is known that the testification score on job claims for low status and low similarity job of the respondents are affected by age, gender, dummy similar job, and dummy high status job. This logistic regression is an addition to answer number 2 and number 3.

```
four <- polr(as.factor(CODE) ~ HIGH_STATUS + SIMILAR + Gender + Age, data = Epistemic_Test_4, Hess=TRUE)
ctable <- coef(summary(four))
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
(ctable <- cbind(ctable, "p value" = p))
```

```
##                    Value  Std. Error   t value      p value
## HIGH_STATUS  0.433229549 0.079942029  5.419296 5.983405e-08
## SIMILAR      1.506648208 0.084779738 17.771324 1.178753e-70
## Gender       0.127035744 0.072603724  1.749714 8.016769e-02
## Age          0.006886214 0.002857732  2.409678 1.596659e-02
## 1|2         -0.495736544 0.172275609 -2.877578 4.007405e-03
## 2|3          2.059669372 0.178753462 11.522403 1.017339e-30
```

## Limitations of this analysis strategy

1. To answer those four research questions, it needs different data. For the question number 1 and 2, I used average testification score for each condition.
2. While to answer question number 3, I used the coded average score which categorize 9 scores into 3 categories, which are: surely false (1-3), not sure (4-6), and surely true (7-9).
3. Furthermore, to answer question number 4, I used the coded average and put 2 more variables, which make the logistic regression become more detailed than the answer number 3.
4. I am not sure whether the answer of respondents (means same age and gender variables) could be replicated as such to accomodate the different conditions in job status and job similarity in the dataset.
5. Gender is not specifically revealed, making it different to state in the interpretation.
6. Decision Tree could be tried.

## Visualization

I plot all of the predicted probabilities for the different conditions; connected with a line, colored by level of the outcome, whether the respondents will say "surely true" on the claims, and facetted by level of job similarity and status. A custom label function also be used, to add clearer labels showing what each column and row of the plot represent.

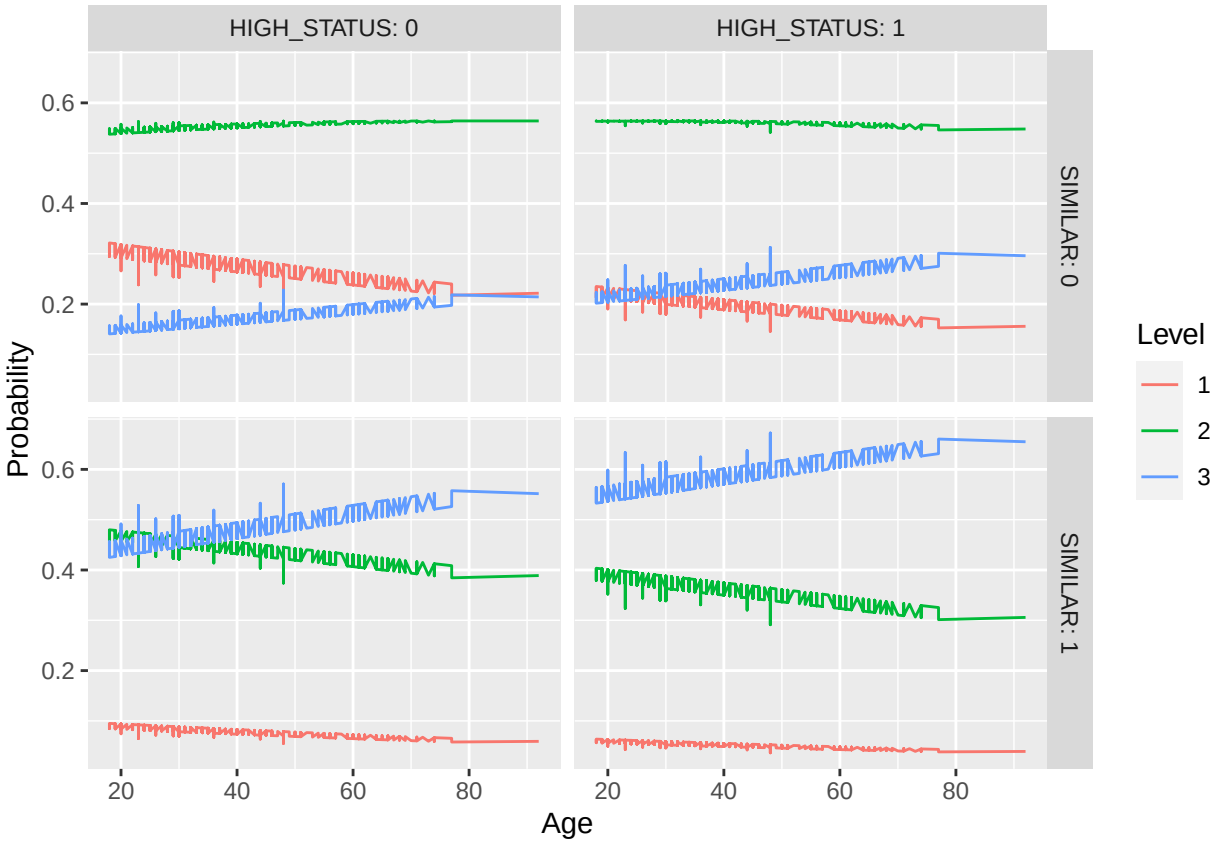
```
newdat <- cbind(Epistemic_Test_4, predict(four, Epistemic_Test_4, type = "probs"))
head(newdat)
```

```
##   RaterID SIMILAR HIGH_STATUS CODE Age Gender          1         2         3
## 1       1       1           0    2  54      1 0.07576877 0.4377604 0.4864708
## 2       2       1           0    3  51      2 0.06864750 0.4182937 0.5130588
## 3       3       1           0    3  31      2 0.07799335 0.4433629 0.4786437
## 4       4       1           0    3  31      2 0.07799335 0.4433629 0.4786437
## 5       5       1           0    3  52      1 0.07673888 0.4402302 0.4830310
## 6       6       1           0    3  65      2 0.06273420 0.4001705 0.5370953
```

```
lnewdat <- melt(newdat, id.vars = c("RaterID", "SIMILAR", "HIGH_STATUS", "CODE", "Age", "Gender"),  vari
## view first few rows
head(lnewdat)
```

```
##   RaterID SIMILAR HIGH_STATUS CODE Age Gender Level Probability
## 1       1       1           0    2  54      1     1  0.07576877
## 2       2       1           0    3  51      2     1  0.06864750
## 3       3       1           0    3  31      2     1  0.07799335
## 4       4       1           0    3  31      2     1  0.07799335
## 5       5       1           0    3  52      1     1  0.07673888
## 6       6       1           0    3  65      2     1  0.06273420
```

```
ggplot(lnewdat, aes(x = Age, y = Probability, colour = Level)) +
  geom_line() + facet_grid(SIMILAR ~ HIGH_STATUS, labeller="label_both")
```




Notice that in similar jobs, older people tend to answer "surely true" of the claims made, which probably because elders tend to have more experiences in the similar job.