**1. Reading data**

> jhk<-read.csv("work result.csv")

**2. Confirmation of the number of rows and columns**

> dim(jhk)

[1] 800 11

**3. Confirmation of variables**

> colnames(jhk)

[1] "ID" "Gender" "Department" "Career"

[5] "Collaboration" "Assertiveness" "Skill" "Knowledge"

[9] "Stress" "Total" "Previous.year"

**4. Confirmation of first 4 rows**

> head(jhk, 4)

ID Gender Department Career Collaboration Assertiveness Skill Knowledge

1 1 M A　 Middle 70 45 65 71

2 2 F B　 Expert 45 62 51 72

3 3 M A　 Middle 54 70 55 70

4 4 M A　 Expert 51 63 53 65

Stress Total Previous.year

1 53 251 248

2 64 227 211

3 61 249 242

4 60 232 240

**5. Drawing a histogram**

> library(lattice) # Reading a statistical package

> histogram(~Stress, data=jhk, breaks=10, type="count")# type should be “percent”, “count”, or “density”

グラフ, ヒストグラム

自動的に生成された説明

**6. Finding central tendencies**

> mean(jhk$Stress)

[1] 55.0525

> mean(jhk$Stress)

**[1] 55.0525**

> median(jhk$Stress)

**[1] 55**

> sort(jhk$Stress)

[1] 36 37 37 39 40 40 41 41 41 41 42 42 42 42 42 42 43 43 43 43 43 43 43 43

[25] 43 43 43 43 43 44 44 44 44 44 44 45 45 45 45 45 45 45 45 45 45 45 45 45

[49] 45 45 45 46 46 46 46 46 46 46 46 46 46 46 46 46 46 46 46 47 47 47 47 47

[73] 47 47 47 47 47 48 48 48 48 48 48 48 48 48 48 48 48 48 48 48 48 48 48 48

[97] 48 48 48 48 48 48 48 48 48 48 48 48 48 49 49 49 49 49 49 49 49 49 49 49

[121] 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 50

[145] 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 51 51 51 51 51 51

[169] 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51

[193] 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51 51

[217] 51 51 51 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52

[241] 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52 52

[265] 52 52 52 52 52 52 52 53 53 53 53 53 53 53 53 53 53 53 53 53 53 53 53 53

[289] 53 53 53 53 53 53 53 53 53 53 53 53 53 54 54 54 54 54 54 54 54 54 54 54

[313] 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54

[337] 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54

[361] 54 54 54 54 54 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55

[385] 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55

[409] 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55 55

[433] 55 55 55 55 55 55 55 55 55 55 55 56 56 56 56 56 56 56 56 56 56 56 56 56

[457] 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 57 57

[481] 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57

[505] 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57 57

[529] 57 57 57 57 57 57 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58

[553] 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58

[577] 58 58 58 58 59 59 59 59 59 59 59 59 59 59 59 59 59 59 59 59 59 59 59 59

[601] 59 59 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60

[625] 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60 60

[649] 60 60 60 60 60 61 61 61 61 61 61 61 61 61 61 61 61 61 61 61 61 61 61 61

[673] 61 61 61 61 61 61 61 61 61 61 61 61 61 61 62 62 62 62 62 62 62 62 62 62

[697] 62 62 62 62 62 62 62 62 62 63 63 63 63 63 63 63 63 63 63 63 63 63 63 63

[721] 63 63 63 63 63 63 63 63 63 64 64 64 64 64 64 64 64 64 64 64 64 64 64 64

[745] 64 64 64 64 64 64 64 64 64 64 64 64 64 64 64 65 65 65 65 65 65 65 65 65

[769] 66 66 66 66 66 66 66 66 66 66 66 66 66 66 67 67 67 67 67 67 67 68 68 68

[793] 68 69 69 70 71 73 73 73

> sort(table(jhk$Stress))

36 39 70 71 37 40 69 73 41 68 42 44 67 65 47 43 66 45 46 50 62 59 63 53 64 48

1 1 1 1 2 2 2 3 4 4 6 6 7 9 10 13 14 16 16 19 19 22 24 30 30 32

61 49 56 58 60 52 57 51 54 55

33 34 35 46 51 52 56 57 64 78

>sort(table(jhk$Career))

Expert Middle Young

177 308 315

**7. Finding dispersion**

> sd(jhk$Stress)

[1] 6.02288

> var(jhk$Stress)

[1] 36.27509

**Comparison of univariate data among groups**

**8. Drawing histogram by groups**

> histogram(~Collaboration|Career+Gender, data=jhk, breaks=10)

グラフ, ヒストグラム

自動的に生成された説明

**9. Comparison of dispersion by groups**

> tapply(jhk$Collaboration,jhk$Gender, mean) # tapply: (quantitative variable, variable of a group, name of function)

F M

52.92114 50.05590

> tapply(jhk$Collaboration, jhk$Gender, sd)

F M

9.845809 10.116078

**10. Drawing boxplot**

> boxplot(jhk$Stress, horizontal=TRUE) # drawing vertically: horizontal=FALSE

グラフ, 箱ひげ図

自動的に生成された説明

> boxplot(Collaboration~Gender,data=jhk,horizontal=TRUE)

# Box plot of collaboration by gender

グラフ, 箱ひげ図

自動的に生成された説明

> boxplot(Collaboration~Gender,data=jhk,horizontal=TRUE)

**11. Summary of statistics**

> summary(jhk$Skill)

Min. 1st Qu. Median Mean 3rd Qu. Max.

22.00 44.00 50.00 50.01 56.00 82.00

**12. F-test for homogeneity of variance**

> var.test(Collaboration~Gender, data=jhk)

F test to compare two variances

data: Collaboration by Gender

F = 0.94728, num df = 316, denom df = 482, p-value = 0.6026

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.776489 1.160565

sample estimates:

ratio of variances

0.9472802

#df(degree of freedom)

**13. Independent-samples t-test (homogeneity of variance)**

> t.test(Collaboration~Gender, data=jhk, var.equal=TRUE)

Two Sample t-test

data: Collaboration by Gender

t = 3.9599, df = 798, p-value = 8.167e-05

alternative hypothesis: true difference in means between group F and group M is not equal to 0

95 percent confidence interval:

1.444935 4.285535

sample estimates:

mean in group F mean in group M

52.92114 50.05590

**13. Welch T-test (NOT homogeneity of variance)**

> t.test(Collaboration~Gender, data=jhk,var.equal=FALSE)

Welch Two Sample t-test

data: Collaboration by Gender

t = 3.9823, df = 688.8, p-value = 7.552e-05

alternative hypothesis: true difference in means between group F and group M is not equal to 0

95 percent confidence interval:

1.452562 4.277908

sample estimates:

mean in group F mean in group M

52.92114 50.05590

**14. Paired T-test**

> score<-c(jhk$Total,jhk$Previous.year)

> year<-c(rep("Total", 800),rep("Previous.year", 800))

> t.test(score~year, paired=TRUE)

Paired t-test

data: score by year

t = -0.85118, df = 799, p-value = 0.3949

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2.045664 0.808164

sample estimates:

mean of the differences

-0.61875

**Basic analysis of multivariate data**

**15. Summarizing multivariate data by column**

> varname<-c("Collaboration", "Assertiveness", "Skill", "Knowledge")

> jhk2<-jhk[,varname] # Extracting 4 variables from data frame

> apply(jhk2, 2, mean) 　　 # Finding mean by each variable. (2nd argument: 1 is row, 2 is column)

Collaboration 　　Assertiveness Skill Knowledge

51.19125 58.00500 50.00625 63.00375

> apply(jhk2,2,sd) # Finding SD by each variable

Collaboration Assertiveness Skill Knowledge

10.101471 11.992905 10.006567 7.995539

**16. Summarizing multivariate data by row**

> apply(jhk2,1,sum) # Finding sum of 4 variables by person

[1] 251 230 249 232 220 236 259 237 294 254 238 308 207 233 232 235 231 169

[19] 137 297 221 157 224 244 266 213 187 153 255 218 254 249 230 197 219 195

> apply(jhk2,1,sd) # Finding SD of 4 variables by person

[1] 12.1209186 11.9582607 8.9582364 7.0237692 10.0000000 11.6332856

[7] 5.7373048 12.8160056 8.3864971 13.9642400 5.2599113 7.7028133

**17. Comparing multivariate distributions between groups**

> by(jhk2,jhk$Gender,apply,2,mean) #Finding mean of 4 variables by gender

jhk$Gender: F

Collaboration Assertiveness Skill Knowledge

52.92114 57.41640 50.37855 63.05678

----------------------------------------------------------

jhk$Gender: M

Collaboration Assertiveness Skill Knowledge

50.05590 58.39130 49.76190 62.96894

> by(jhk2,jhk$Gender,apply,2,sd) #Finding SD of 4 variables by gender

jhk$Gender: F

Collaboration Assertiveness Skill Knowledge

9.845809 11.233291 10.153034 7.662786

----------------------------------------------------------

jhk$Gender: M

Collaboration Assertiveness Skill Knowledge

10.116078 12.462645 9.912231 8.214264

**18. Standardizing data (Standardization)**

> zscore<-scale(jhk2) # finding the z score ( adjusting to mean=0 & SD=1)

> head(zscore,2)

Collaboration Assertiveness Skill Knowledge

[1,] 1.8619812 -1.0843912 1.49839107 1.000089

[2,] -0.6129058 0.3331136 0.09930979 1.125159

> tscore<-zscore\*10+50 # finding the t score ( adjusting to mean=50 & SD=10)

> head(tscore, 2)

Collaboration Assertiveness Skill Knowledge

[1,] 68.61981 39.15609 64.98391 60.00089

[2,] 43.87094 53.33114 50.99310 61.25159

**19. Drawing Scatter plot**

> Skill<-jhk$Skill

> Knowledge<-jhk$Knowledge

> plot(Skill, Knowledge, xlab="Skill", ylab="Knowledge")

グラフ, 散布図

自動的に生成された説明

**20. Drawing Scatter plot matrix**

> cas<-c("Collaboration", "Assertiveness", "Stress")

> plot(jhk[,cas])

ダイアグラム

自動的に生成された説明

**21. Drawing Scatter plot by group**

> xyplot(Knowledge~Skill|Career+Department, data=jhk) # read ‘library(lattice)’ beforehand

カレンダー

自動的に生成された説明

**22. Finding correlation of 2 variables**

> cor(jhk$Collaboration,jhk$Stress)

[1] -0.507292

**23. Correlations matrix**

> cor(jhk[,cas])

Collaboration Assertiveness Skill

Collaboration 1.0000000 0.3486028 0.4915577

Assertiveness 0.3486028 1.0000000 0.3800334

Skill 0.4915577 0.3800334 1.0000000

**24. Covariance matrix**

> cov(jhk[,cas])

Collaboration Assertiveness Skill

Collaboration 102.03972 42.23183 49.68716

Assertiveness 42.23183 143.82976 45.60698

Skill 49.68716 45.60698 100.13138

**25. Hypothesis testing on the correlation coefficient**

> library(psych) #reading a package ‘psych’

> corresult<-corr.test(jhk[,cas])

> corresult$t # finding t-value

Collaboration Assertiveness Skill

Collaboration Inf 10.50673 15.94539

Assertiveness 10.50673 Inf 11.60631

Skill 15.94539 11.60631 Inf

> corresult$p # finding p-value

Collaboration Assertiveness Skill

Collaboration 0.000000e+00 2.837099e-24 2.029922e-49

Assertiveness 2.837099e-24 0.000000e+00 1.379569e-28

Skill 6.766406e-50 6.897844e-29 0.000000e+00

**26. Multiple regression analysis**

> jhk3<-lm(Total~Collaboration,data=jhk) # simple regression analysis

> summary(jhk3)

Call:

lm(formula = Total ~ Collaboration, data = jhk)

Residuals:

Min 1Q Median 3Q Max

-53.039 -13.448 -0.455 13.007 76.545

Coefficients:

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

(Intercept) 107.61384 3.57059 30.14 <2e-16 \*\*\*

Collaboration 2.21529 0.06843 32.37 <2e-16 \*\*\*

---

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

Residual standard error: 19.54 on 798 degrees of freedom

Multiple R-squared: 0.5677, Adjusted R-squared: 0.5672

F-statistic: 1048 on 1 and 798 DF, p-value: < 2.2e-16

-----------------------------------------------------------------------------------------------

> jhk4<-lm(Total~Collaboration+Assertiveness,data=jhk) # multiple regression analysis

> summary(jhk4)

Call:

lm(formula = Total ~ Collaboration + Assertiveness, data = jhk)

Residuals:

Min 1Q Median 3Q Max

-35.389 -8.356 -0.039 8.405 34.349

Coefficients:

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

(Intercept) 58.71561 2.72719 21.53 <2e-16 \*\*\*

Collaboration 1.66562 0.04710 35.37 <2e-16 \*\*\*

Assertiveness 1.32810 0.03967 33.48 <2e-16 \*\*\*

---

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

Residual standard error: 12.6 on 797 degrees of freedom

Multiple R-squared: 0.8204, Adjusted R-squared: 0.8199

F-statistic: 1820 on 2 and 797 DF, p-value: < 2.2e-16

-----------------------------------------------------------------------------------------------------------

# multiple regression analysis

> jhk5<-lm(Total~Collaboration+Assertiveness+Skill+Knowledge+Stress, data=jhk)

> summary(jhk5)

Call:

lm(formula = Total ~ Collaboration + Assertiveness + Skill +

Knowledge + Stress, data = jhk)

Residuals:

Min 1Q Median 3Q Max

-2.7759 -1.5995 0.7824 1.1951 2.0965

Coefficients:

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

(Intercept) 2.596789 0.778753 3.335 0.000894 \*\*\*

Collaboration 0.940841 0.008237 114.216 < 2e-16 \*\*\*

Assertiveness 1.018141 0.004843 210.228 < 2e-16 \*\*\*

Skill 1.015079 0.006689 151.763 < 2e-16 \*\*\*

Knowledge 1.014757 0.007455 136.115 < 2e-16 \*\*\*

Stress -0.063451 0.011566 -5.486 5.53e-08 \*\*\*

---

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

Residual standard error: 1.424 on 794 degrees of freedom

Multiple R-squared: 0.9977, Adjusted R-squared: 0.9977

F-statistic: 6.937e+04 on 5 and 794 DF, p-value: < 2.2e-16

Reference

川端一光、岩間徳兼、鈴木雅之(2018). 「Rによる多変量解析入門 データ分析の実践と理論」オーム社