U6614: Assignment 2

your name (your uni)

2025-09-16

Please submit your knitted .pdf file along with the corresponding R markdown (.rmd) via Courseworks by 11:59pm on the due date.

Before knitting your rmd file as a pdf, you will need to install TinyTex for Latex distribution by running the following code:

```
tinytex::install_tinytex()
```

Please visit this link for more information on TinyTex installation.

1 Load and inspect ACCESS 2018 data:

1a) Inspect the data frame and state the data types for each column.

- make sure to inspect the age, gender, caste, income source, state & educ columns

```
load("access2018.RData")
str(access2018)
```

```
## tibble [9,008 x 10] (S3: tbl_df/tbl/data.frame)
##  $ HHID        : chr [1:9008] "236582/10" "222500/10" "231781/11" "257562/4" ...
##   ..- attr(*, "label")= chr "Household ID"
##   ..- attr(*, "format.stata")= chr "%10s"
##  $ date        : chr [1:9008] "15-May-18" "27-May-18" "Mar 18, 2018" "Apr 23, 2018" ...
##   ..- attr(*, "label")= chr "Date of interview"
##   ..- attr(*, "format.stata")= chr "%12s"
##  $ state       : chr [1:9008] "BIHAR" "BIHAR" "BIHAR" "BIHAR" ...
##   ..- attr(*, "label")= chr "State"
##   ..- attr(*, "format.stata")= chr "%14s"
##  $ relationship: 'haven_labelled' num [1:9008] 2 5 1 1 1 4 4 1 1 1 ...
##   ..- attr(*, "label")= chr "Relationship to the household head"
##   ..- attr(*, "format.stata")= chr "%13.0g"
##   ..- attr(*, "labels")= Named num [1:5] 1 2 3 4 5
##   .. ..- attr(*, "names")= chr [1:5] "Self" "Spouse" "Mother/Father" "Son/Daughter" ...
##  $ age         : num [1:9008] 35 43 65 70 57 27 32 60 27 58 ...
##   ..- attr(*, "label")= chr "Age (years)"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ gender      : chr [1:9008] "Female" "Male" "Male" "Male" ...
##  $ education   : chr [1:9008] "Grade 5" "Grade 5" "Grade 5" "Grade 12" ...
##  $ caste       : chr [1:9008] "Other Backward Class" "Other Backward Class" "General" "General" ...
##  $ incsource   : chr [1:9008] "Daily labourer" "Agriculture (own land)" "Agriculture (own land)" "Ag
##  $ month_exp   : num [1:9008] 7000 9000 6000 5000 16000 10000 8000 10000 5000 15000 ...
##   ..- attr(*, "label")= chr "Monthly expenditure (rupee/month)"
```

```
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  - attr(*, "na.action")= 'omit' Named int [1:64] 89 1382 1663 2417 2424 2442 2598 2686 2765 2794 ...
##   ..- attr(*, "names")= chr [1:64] "89" "1382" "1663" "2417" ...
```

```
summary(access2018$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   32.00   42.00   43.29   55.00   98.00
```

```
summary(access2018$gender) #summary is not very useful with character variables
```

```
##    Length     Class      Mode
##      9008 character character
```

```
summary(access2018$caste) #summary is not very useful with character variables
```

```
##    Length     Class      Mode
##      9008 character character
```

```
summary(access2018$education) #summary is not very useful with character variables
```

```
##    Length     Class      Mode
##      9008 character character
```

```
summary(access2018$state) #summary is not very useful with character variables
```

```
##    Length     Class      Mode
##      9008 character character
```

Age is a numeric variable, whereas gender, caste, income source, state and education are all stored as character strings.

1b) Use the mutate function to create new column for gender:

- gender.fac = as.factor(gender),
- check if it worked by calling the str() function

```
mutate(access2018, gender.fac = as.factor(gender))
```

```
#Note: the output is an object, we're just not assigning it to the environment
#but we can put the entire operation within str() to inspect the output
str(mutate(access2018, gender.fac = as.factor(gender)),
    give.attr = FALSE)
```

```
## tibble [9,008 x 11] (S3: tbl_df/tbl/data.frame)
##  $ HHID        : chr [1:9008] "236582/10" "222500/10" "231781/11" "257562/4" ...
##  $ date        : chr [1:9008] "15-May-18" "27-May-18" "Mar 18, 2018" "Apr 23, 2018" ...
##  $ state       : chr [1:9008] "BIHAR" "BIHAR" "BIHAR" "BIHAR" ...
##  $ relationship: 'haven_labelled' num [1:9008] 2 5 1 1 1 4 4 1 1 1 ...
##  $ age         : num [1:9008] 35 43 65 70 57 27 32 60 27 58 ...
##  $ gender      : chr [1:9008] "Female" "Male" "Male" "Male" ...
##  $ education   : chr [1:9008] "Grade 5" "Grade 5" "Grade 5" "Grade 12" ...
##  $ caste       : chr [1:9008] "Other Backward Class" "Other Backward Class" "General" "General" ...
##  $ incsource   : chr [1:9008] "Daily labourer" "Agriculture (own land)" "Agriculture (own land)" "Ag
##  $ month_exp   : num [1:9008] 7000 9000 6000 5000 16000 10000 8000 10000 5000 15000 ...
##  $ gender.fac  : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 2 2 1 2 2 ...
```

1c) In a single pipe, include gender.fac in a new data frame called access2018.temp1.

- also create factors for caste, income source, state and education

- exclude the columns for HHID and date after creating access2018.temp1
- print the first 5 observations

```
access.temp1 <- access2018 %>%
  mutate(gender.fac = as.factor(gender),
         caste.fac = as.factor(caste),
         education.fac = as.factor(education),
         state.fac = as.factor(state),
         incsource.fac = as.factor(incsource)) %>%
    select(-HHID, -date)
```

```
head(access.temp1, n = 5)
```

```
## # A tibble: 5 x 13
##   state relationship   age gender education caste incsource month_exp gender.fac
##   <chr> <hvn_lbll>   <dbl> <chr>  <chr>     <chr> <chr>         <dbl> <fct>
## 1 BIHAR 2              35 Female Grade 5   Othe~ Daily la~      7000 Female
## 2 BIHAR 5              43 Male   Grade 5   Othe~ Agricult~      9000 Male
## 3 BIHAR 1              65 Male   Grade 5   Gene~ Agricult~      6000 Male
## 4 BIHAR 1              70 Male   Grade 12  Gene~ Agricult~      5000 Male
## 5 BIHAR 1              57 Male   Grade 10  Gene~ Agricult~     16000 Male
## # i 4 more variables: caste.fac <fct>, education.fac <fct>, state.fac <fct>,
## #   incsource.fac <fct>
```

1.0.1 1d) Inspect caste.fac, gender.fac, incsource.fac, education.fac and state.fac

using the levels() function. What package is the levels() function located in? {-}

```
levels(access.temp1$gender.fac)
```

```
## [1] "Female" "Male"
```

```
levels(access.temp1$caste.fac)
```

```
## [1] "General"              "Other Backward Class" "Others"
## [4] "Scheduled Caste"      "Scheduled Tribe"
```

```
levels(access.temp1$education.fac)
```

```
## [1] "Grade 10"          "Grade 12"          "Grade 5"
## [4] "Graduate or above" "No School "
```

```
levels(access.temp1$state.fac)
```

```
## [1] "BIHAR"          "JHARKHAND"      "MADHYA PRADESH" "ODISHA"
## [5] "UTTAR PRADESH"  "WEST BENGAL"
```

```
levels(access.temp1$incsource.fac)
```

```
## [1] "Agriculture (own land)" "Casual agri labour"     "Cattle rearing"
## [4] "Cultivation (leased)"   "Daily labourer"         "Other"
## [7] "Own business"           "Salaried job"
```

```
#?levels #'?' before a function name will tell you what the function does and what package it is a part
```

Using ?levels, we see that levels is a base R function. This means that it generally can't be called within a pipe, as it is not part of the tidyverse.

1e) Use filter() to only include rows only for the state "Uttar Pradesh."

- store as a new object access_UP
- print the first 5 observations
- confirm your data only includes observations for Uttar Pradesh

```
access_UP <- access.temp1 %>%
  filter(state == "UTTAR PRADESH")

head(access_UP, n = 5)
```

```
## # A tibble: 5 x 13
##   state relationship   age gender education caste incsource month_exp gender.fac
##   <chr> <hvn_lbll>   <dbl> <chr>  <chr>     <chr> <chr>         <dbl> <fct>
## 1 UTTA~ 1               70 Female "No Scho~ Othe~ Daily la~      4000 Female
## 2 UTTA~ 4               30 Male   "No Scho~ Sche~ Daily la~      3000 Male
## 3 UTTA~ 1               35 Male   "No Scho~ Sche~ Daily la~      2500 Male
## 4 UTTA~ 1               72 Male   "Grade 1~ Othe~ Agricult~      3000 Male
## 5 UTTA~ 4               21 Male   "Graduat~ Gene~ Cultivat~      8000 Male
## # i 4 more variables: caste.fac <fct>, education.fac <fct>, state.fac <fct>,
## #   incsource.fac <fct>
```

```
#validate
summary(access_UP$state.fac)
```

```
##          BIHAR      JHARKHAND MADHYA PRADESH         ODISHA  UTTAR PRADESH
##              0              0              0              0           3002
##    WEST BENGAL
##              0
```

1f) What is the female share of respondents in Uttar Pradesh? Compare it to the female share of respondents in Bihar?

```
#Share of female respondents in UP
summary(access_UP$gender.fac)
```

```
## Female   Male
##   1011   1991
```

```
prop.table(table(access_UP$gender.fac))
```

```
##
##    Female      Male
## 0.3367755 0.6632245
```

```
t1_UP <- prop.table(table(access_UP$gender.fac))

#Share of female respondents in Bihar
access_bihar <- access.temp1 %>%
  filter(state == "BIHAR")
summary(access_bihar$gender.fac)
```

```
## Female   Male
##    352   1158
```

```
prop.table(table(access_bihar$gender.fac))
```

```
##
```

```
##    Female      Male
## 0.2331126 0.7668874
```

```
t1_bihar <- prop.table(table(access_bihar$gender.fac))
```

The share of female respondents in UP is 0.337, compared to 0.233 in Bihar.

1g) Remove the access2018.temp1 object from memory using the rm() function.

```
rm(access.temp1)
rm(access_bihar)
```

2 Describe the access_UP data frame

2a) What is the unit of observation (or unit of analysis)?

```
str(access_UP)
```

```
## tibble [3,002 x 13] (S3: tbl_df/tbl/data.frame)
##  $ state        : chr [1:3002] "UTTAR PRADESH" "UTTAR PRADESH" "UTTAR PRADESH" "UTTAR PRADESH" ...
##   ..- attr(*, "label")= chr "State"
##   ..- attr(*, "format.stata")= chr "%14s"
##  $ relationship : 'haven_labelled' num [1:3002] 1 4 1 1 4 1 5 1 1 1 ...
##   ..- attr(*, "label")= chr "Relationship to the household head"
##   ..- attr(*, "format.stata")= chr "%13.0g"
##   ..- attr(*, "labels")= Named num [1:5] 1 2 3 4 5
##   .. ..- attr(*, "names")= chr [1:5] "Self" "Spouse" "Mother/Father" "Son/Daughter" ...
##  $ age          : num [1:3002] 70 30 35 72 21 55 40 51 66 55 ...
##   ..- attr(*, "label")= chr "Age (years)"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ gender       : chr [1:3002] "Female" "Male" "Male" "Male" ...
##  $ education    : chr [1:3002] "No School " "No School " "No School " "Grade 10" ...
##  $ caste        : chr [1:3002] "Other Backward Class" "Scheduled Caste" "Scheduled Caste" "Other Bac
##  $ incsource    : chr [1:3002] "Daily labourer" "Daily labourer" "Daily labourer" "Agriculture (own
##  $ month_exp    : num [1:3002] 4000 3000 2500 3000 8000 4000 2000 15000 8000 15000 ...
##   ..- attr(*, "label")= chr "Monthly expenditure (rupee/month)"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ gender.fac   : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 2 2 2 1 2 ...
##  $ caste.fac    : Factor w/ 5 levels "General","Other Backward Class",..: 2 4 4 2 1 1 4 1 4 2 ...
##  $ education.fac: Factor w/ 5 levels "Grade 10","Grade 12",..: 5 5 5 1 4 1 5 4 5 3 ...
##  $ state.fac    : Factor w/ 6 levels "BIHAR","JHARKHAND",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ incsource.fac: Factor w/ 8 levels "Agriculture (own land)",..: 5 5 5 1 4 1 5 7 5 5 ...
##  - attr(*, "na.action")= 'omit' Named int [1:64] 89 1382 1663 2417 2424 2442 2598 2686 2765 2794 ...
##   ..- attr(*, "names")= chr [1:64] "89" "1382" "1663" "2417" ...
```

Here each row represents the responses of an individual on behalf of the households. So you could really say the unit of observation is the household or the individual, depending on the nature of the analysis and variables you are using.

2b) How many individuals are observed?

- how many of these are from 'Reserved' castes (i.e: Scheduled Caste, Scheduled Tribe, Other Backward Class")?

- how many are "General" caste members? [**Note: for this question, use the 'table' and 'prop.table' functions. 'table' will give the number of observations belonging to each caste category.'prop.table will give the proportion**]

```
#Number of total observations in access_UP
nrow(access_UP)
```

```
## [1] 3002
```

```
#Ans: 3002

#Number of 'Reserved' and 'General' individuals
table(access_UP$reserved)
```

```
##
##  General Reserved
##      708     2294
```

```
#Ans: 708 (G), 2294 (R)

#Proportion of 'Reserved' and 'General' individuals
prop.table(table(access_UP$reserved))
```

```
##
##   General  Reserved
## 0.2358428 0.7641572
```

```
#Ans: 23.5% (G), 76.4% (R)
```

The total number of observations in the access_UP data frame is 3002. Within this, there are 708 General category observations and 2294 Reserved category observations. Therefore, the proportion of General category observations is 0.236, while Reserved category individuals make up 0.764 of the total observations.

2c) Do you think the sample is representative of all households in India?

- how would you describe the population represented by this sample?
- is there more information you would like to see to assess the representativeness of the sample?

The ACCESS 2018 dataset provides valuable insights into rural households in six states (Bihar, Jharkhand, Madhya Pradesh, Odisha, Uttar Pradesh, and West Bengal), but it is not fully representative of all Indian households. These less economically diversified states rely more on agriculture and informal labor, while more diversified states like Maharashtra, Gujarat, and Tamil Nadu have significant service and manufacturing sectors, leading to different spending behaviors. The six states also have a higher proportion of Scheduled Castes (SC), Scheduled Tribes (ST), Other Backward Classes (OBC), and Muslim populations compared to more diversified states, which may have a greater presence of General castes and different religious groups. These demographic and economic differences shape spending behaviors, resulting in varied expenditure patterns across regions.

To determine the representativeness of this sample, additional information is needed. This includes the sampling methodology (e.g., random or stratified sampling), whether the data is weighted to account for unequal selection probabilities, and a comparison of key demographic characteristics (caste, education, income, occupation, religion) with national datasets like the Census of India. We would also need to assess the urban-rural balance in the sample, as the dataset focuses on rural areas, and determine how representative these six states are compared to the rest of India. With this information, we could better evaluate how closely the sample mirrors the broader Indian population.

2d) What is the average age of individuals in the sample? youngest and oldest person?

```
summary(access_UP$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   30.00   42.00   43.44   55.00   98.00
```

```
#Avg age: 43.41 years, youngest: 18, oldest: 98.
```

The average age of individuals in this sample is 43.44, with the youngest individual being 18 years old, and the oldest being 98 years old.

3 Earnings per week for different groups in Uttar Pradesh

3a) Find the observation for the top monthly expenditure using the summarize() function.

- assign this to a new object called max_exp_obs1

```
 max_exp_obs1 <- access_UP %>%
    summarize(max_exp = max(month_exp))
  max_exp_obs1
```

```
## # A tibble: 1 x 1
##   max_exp
##     <dbl>
## 1   56000
```

3b) Find max monthly expenditures using the arrange function instead of summarize.

```
access_UP %>%
    arrange(desc(month_exp)) %>%
    select(month_exp) %>%
    head(n = 1)
```

```
## # A tibble: 1 x 1
##   month_exp
##       <dbl>
## 1     56000
```

3c) Use the filter function to subset for the observation with max monthly expenditure.

- (don't hardcode the max expenditure to filter on, refer to the max_exp_obs1 object from a)
- store in new data frame max_exp_obs2
- confirm it worked

```
max_exp_obs2 <- access_UP %>%
                filter(month_exp == max_exp_obs1$max_exp)

#validate
summary(max_exp_obs2$month_exp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   56000   56000   56000   56000   56000   56000
```

3d) What is the age, gender and caste of the top monthly spender in the sample?

```
max_exp_obs2 %>%
  select(age, gender, caste) %>%
  head(n = 1)
```

```
## # A tibble: 1 x 3
##     age gender caste
##   <dbl> <chr>  <chr>
## 1    43 Male   Other Backward Class
```

3e) List the age, gender and caste of the top 10 monthly spenders in the sample.

```
access_UP %>%
    arrange(desc(month_exp)) %>%
    select(age, gender, caste, month_exp) %>%
    head(n = 10)
```

```
## # A tibble: 10 x 4
##      age gender caste                month_exp
##    <dbl> <chr>  <chr>                    <dbl>
##  1    43 Male   Other Backward Class     56000
##  2    58 Male   Other Backward Class     50000
##  3    66 Male   Other Backward Class     50000
##  4    24 Male   Other Backward Class     50000
##  5    28 Male   Other Backward Class     50000
##  6    18 Male   General                  40000
##  7    65 Male   General                  40000
##  8    23 Female Other Backward Class     40000
##  9    42 Male   Other Backward Class     40000
## 10    20 Male   Other Backward Class     32000
```

3f) How many individuals spend more than the mean monthly expenditure amount of the sample?

```
 access_UP %>%
    filter(month_exp > mean(month_exp)) %>%
    nrow()
```

```
## [1] 1115
```

Caste-based monthly expenditure gaps in Uttar Pradesh.

4a) Use the filter function and 'reserved' dummy variable to subset observations belonging to the 'General' caste.

- assign to new data frame, access_UPgen
- sort in descending order of monthly expenditure
- check if it worked

```
access_UPgen <- access_UP %>%
  filter(reserved == "General") %>%
  arrange(desc(month_exp))
```

```
#validate
  summary(access_UPgen$reserved)
```

```
##  General Reserved
##      708        0
```

4b) Repeat part a, but this time, use the filter function to subset observations belonging to 'Reserved" castes, and assign them to a new data frame called 'access_UPres'.

```
access_UPres <- access_UP %>%
  filter(reserved == "Reserved")%>%
  arrange(desc(month_exp))

#validate
  summary(access_UPres$reserved)
```

```
##  General Reserved
##        0     2294
```

4c) Use summarise to find mean, min & max monthly expenditure for the General category and Reserved category of castes, separately.

- name each statistic appropriately (i.e. name each column in the 1-row table of stats)
- what is the gap in mean monthly expenditure between the two groups?

```
access_UPgen %>%
  summarise(avg_monthexp_gen = mean(month_exp),
            min_monthexp_gen = min(month_exp),
            max_monthexp_gen = max(month_exp),
            n_monthexp_gen = n())
```

```
## # A tibble: 1 x 4
##   avg_monthexp_gen min_monthexp_gen max_monthexp_gen n_monthexp_gen
##              <dbl>            <dbl>            <dbl>          <int>
## 1            7861.              300            40000            708
```

```
access_UPres %>%
  summarise(avg_monthexp_res = mean(month_exp),
            min_monthexp_res = min(month_exp),
            max_monthexp_res = max(month_exp),
            n_monthexp_res = n())
```

```
## # A tibble: 1 x 4
##   avg_monthexp_res min_monthexp_res max_monthexp_res n_monthexp_res
##              <dbl>            <dbl>            <dbl>          <int>
## 1            6475.                0            56000           2294
```

```
mean(access_UPgen$month_exp) - mean(access_UPres$month_exp)
```

```
## [1] 1385.483
```

```
#Ans: 1385.48
```

The gap in mean monthly expenditures between General and Reserved caste group households is 1385 Rupee.

4d) Research suggests that people belonging to the 'General' caste category own 65% of agricultural land in India (India Human Development Survey, 2020). They are also more likely to own land in a proportion that is much higher than their share of the population. We will now use this data to compare 'General' versus 'Reserved' households using the 'reserved' dummy variable to understand whether this is true.

- use the 'table' and 'prop.table' functions to explore the total number of landowners by their caste.
- here, 'landowners' are people whose primary source of income is agriculture on their own land. So, you can use the 'incsource.fac' variable to filter the 'access_UP' data frame for 'Agriculture (own land)'
- does the number of General and Reserved caste landowners seem proportional to the overall number of 'General' and 'Reserved' caste individuals in Uttar Pradesh's population (as calculated in Q2 b)?

```
table(access_UP$incsource.fac, access_UP$reserved)
```

```
##
##                           General Reserved
##   Agriculture (own land)     433      877
##   Casual agri labour           6       86
##   Cattle rearing               4       16
##   Cultivation (leased)        17       86
##   Daily labourer              94      874
##   Other                       18       24
##   Own business                55      182
##   Salaried job                81      149
```

```
prop.table(table(access_UP$incsource.fac,
                 access_UP$reserved),
           margin = 1) %>%
  round(3)
```

```
##
##                           General Reserved
##   Agriculture (own land)   0.331    0.669
##   Casual agri labour       0.065    0.935
##   Cattle rearing           0.200    0.800
##   Cultivation (leased)     0.165    0.835
##   Daily labourer           0.097    0.903
##   Other                    0.429    0.571
##   Own business             0.232    0.768
##   Salaried job             0.352    0.648
```

```
pt_4d <- prop.table(table(access_UP$incsource.fac,
                          access_UP$reserved),
                    margin = 1) %>% round(3)
```

Out of 708 General caste individuals, 433 are landowners, meaning 33.1 percent own land, while they make up 23.5% of the total population. For Reserved caste individuals, 877 out of 2294 are landowners, equating to 66.9 percent, even though they account for 76.4% of the population. This shows that General caste individuals are disproportionately represented among landowners compared to their share of the overall population.

The total number of observations in the access_UP data frame is 3002. Within this, there are 708 General category observations and 2294 Reserved category observations. Therefore, the proportion of General category observations is 0.236, while Reserved category individuals make up 0.764 of the total observations.

4e) Do differences in landowning explain the expenditure gap between General and Reserved castes? In other words, does the General-Reserved expenditure gap persist among landowning families? What is the gap in mean monthly expenditures of General vs Reserved category landowners?

```
access_UPgen_land <- access_UPgen %>%
  filter(incsource.fac == "Agriculture (own land)")

access_UPres_land <- access_UPres %>%
  filter(incsource.fac == "Agriculture (own land)")

gap_land <- mean(access_UPgen_land$month_exp) - mean(access_UPres_land$month_exp)
gap_land
```

```
## [1] 864.8861
```

```
#Ans: 864.89
```

The General-Reserved expenditure gap persists among landowners only but is lower in magnitude, falling to 865 Rupee for landowning households, compared to 1385 for all households.

4f) Is there a gap between the mean monthly expenditures of General caste male and Reserved caste male landowners? What about the same, but for the female landowners?

```
access_UPgen_male <- access_UPgen_land %>%
  filter(gender == "Male")

access_UPgen_fem <- access_UPgen_land %>%
  filter(gender == "Female")

access_UPres_male <- access_UPres_land %>%
  filter(gender == "Male")

access_UPres_fem <- access_UPres_land %>%
  filter(gender == "Female")

gap_male <- mean(access_UPgen_male$month_exp) - mean(access_UPres_male$month_exp)
gap_male
```

```
## [1] 838.7428
```

```
#Ans: 838.74

gap_fem <- mean(access_UPgen_fem$month_exp) - mean(access_UPres_fem$month_exp)
gap_fem
```

```
## [1] 915.3296
```

```
#Ans: 915.32
```

The General-Reserved expenditure gap for landowning households with male respondents is 839 Rupee, compared to 915 Rupee for female respondents. If we consider the respondent to be the head of household, this implies that being a women-led household doesn't explain the expenditure gap between General and Reserved landowning families.

4g) Does educational attainment explain part of the expenditure gap between General and Reserved caste landowners? What is gap between landowning General vs. Reserved household with a HS degree or more?

```
access_UP <- access_UP %>%
  mutate(degree = factor(if_else(education.fac %in% c("Grade 12",
                                                      "Graduate or above"),
                                 "HS degree or above",
                                 "Less than HS")))

#General v Reserved landowners who have degrees
access_UPgen_hs <- access_UP %>%
  filter(reserved == "General",
         incsource.fac == "Agriculture (own land)",
         degree == "HS degree or above")

access_UPres_hs <- access_UP %>%
  filter(reserved == "Reserved",
         incsource.fac == "Agriculture (own land)",
         degree == "HS degree or above")

gap_hs_land <- mean(access_UPgen_hs$month_exp) - mean(access_UPres_hs$month_exp)
gap_hs_land
```

```
## [1] 588.6288
```

```
#Ans: 588.62
```

The expenditure gap between General and Reserved caste landowners who have at least a high school degree or higher is 589. This suggests that even among more educated landowners, there is still a notable expenditure gap between the two groups, indicating that landowning and educational attainment do not fully explain the economic disparities between caste groups. Other structural factors likely contribute to the persistent expenditure gap.