

Assignment1

Philip & Philipp

23 Feb 2016

```
# Load packages and create BibTeX file for R-packages
PackagesUsed <- c("ggplot2", "repmis", "doBy")

setwd("~/Dropbox/4_Spring_2016/2_Collaborative Data Analysis/PandP_Ass1")

# Load PackagesUsed and create .bib BibTeX file
repmis::LoadandCite(PackagesUsed, file = "Ass1Packages.bib", install = FALSE)

## Loading required package: survival
```

Analysis of the dataset ‘occupationalStatus’

The dataset consists of a contingency table between the occupational status measured on an 8-point scales for fathers and their sons.

```
dist_combined <- rbind(dist_fathers, dist_sons)
barplot(dist_combined,
        col=c("navyblue", "darkkhaki"),
        main="Distribution of occupational status among fathers and sons",
        xlab = "Occupational status categories",
        legend = c("Fathers", "Sons"),
        ylim = c(0, 1600),
        beside = TRUE
)
```

Distribution of occupational status among fathers and sons

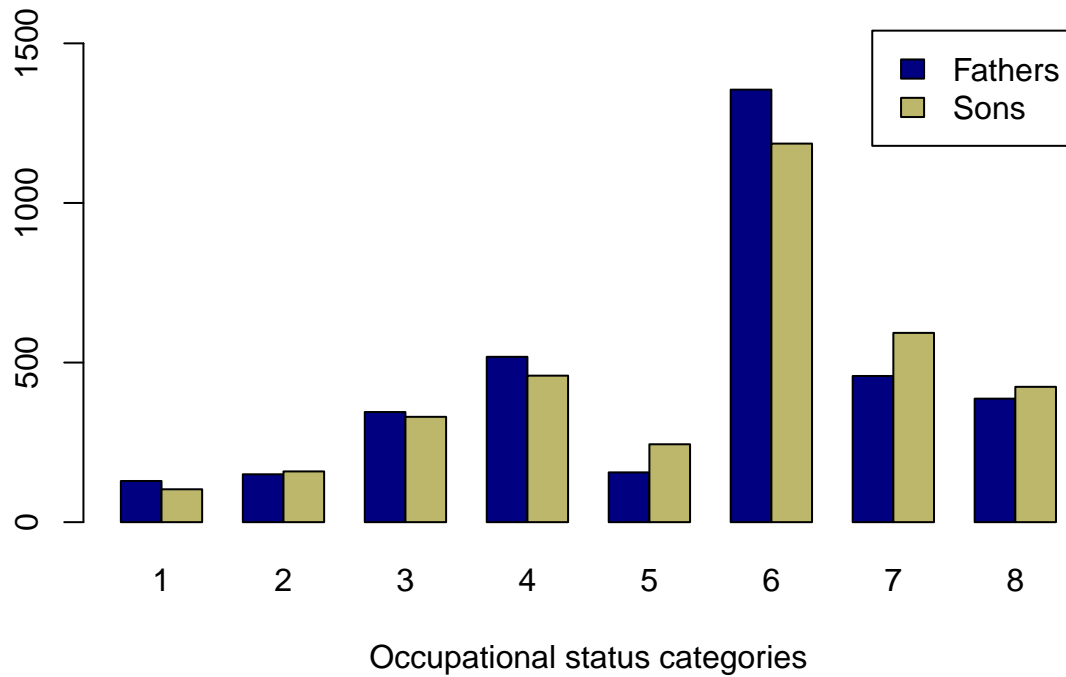


Figure 1 shows the distribution of occupational status for sons and fathers respectively. There does not appear to be any major generational shifts, though sons are slightly overrepresented in occupational status group 7 and 8.

```
knitr::kable(frequency_table, digits = 2)
```

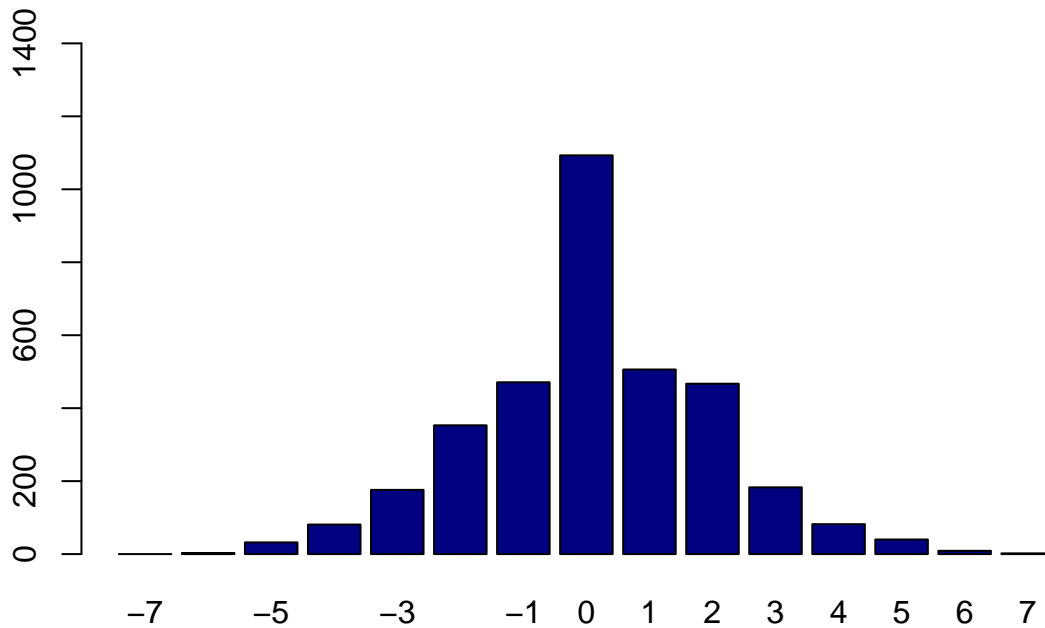
Category	Fathers	Sons
1	0.04	0.03
2	0.04	0.05
3	0.10	0.09
4	0.15	0.13
5	0.04	0.07
6	0.39	0.34
7	0.13	0.17
8	0.11	0.12

Table 1 is equivalent to figure 1 except that it shows the frequency distribution for each status category.

Generational mobility

Figure 1 and table 1 show the overall distribution of occupational status, but is not informative on the extent to which there are generational mobility. Figure 2 shows the distribution of generational mobility. -7 indicates observations where the father had social status 8 and the son 1. As such, a zero is when father and son had the same occupational status.

```
barplot(collapsed$Freq.sum,
        names = collapsed$difference,
        col = "Navyblue",
        ylim = c(0, 1400)
)
```



Analysis of the dataset ‘Life Cycle Savings’

Summary Statistics

```
knitr::kable(res_df, digits = 2)
```

	sr	pop15	pop75	dpi	ddpi
mean	9.67	35.09	2.29	1106.76	3.76
sd	4.48	9.15	1.29	990.87	2.87
median	10.51	32.58	2.17	695.66	3.00
minimum	0.60	21.44	0.56	88.94	0.22
maximum	21.10	47.64	4.70	4001.89	16.71
s.size	50.00	50.00	50.00	50.00	50.00

```
res <- lapply(LifeCycleSavings, function(x) rbind( mean = mean(x) ,
                                                  sd = sd(x) ,
                                                  median = median(x) ,
                                                  minimum = min(x) ,
                                                  maximum = max(x) ,
                                                  s.size = length(x) ) )
res_df <- as.data.frame(res)
```

Packages used (R Core Team 2015), (Højsgaard and Halekoh 2015), (Wickham and Chang 2015) and (Gandrud 2016).

References

Gandrud, Christopher. 2016. *Repmis: Miscellaneous Tools for Reproducible Research*. <https://CRAN.R-project.org/package=repmis>.

Højsgaard, Søren, and Ulrich Halekoh. 2015. *DoBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities*. <https://CRAN.R-project.org/package=doBy>.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Wickham, Hadley, and Winston Chang. 2015. *Ggplot2: An Implementation of the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.