Assignment 2

Gaja

Function definition

Define a function which generates a binary variable with a probit regression model for a given number of predictor variables $X \sim \mathcal{N}(0, 1)$, R^2 and *probability of having a value 1*, π_i.

```
genBin <- function(N, nvar, perc, R2){

  beta <- runif(nvar, 10, 20)
  X <- replicate(nvar, rnorm(N, 0, 1))

  s <- beta %*% cov(X) %*% beta
  resSigma <- (s / R2) - s
  sdLP <- sqrt(s+resSigma)[1]

  mult <- qnorm(1-perc)

  LP <- X %*% beta
  M <- mean(LP)
  R <- LP + rnorm(N, 0, sqrt(resSigma))

  return(list(out=ifelse(R > M+mult*sdLP, 1, 0), X=data.frame(X)))

}
```

Run

Run the code multiple times to obtain the same results by setting the same seed each time before running the function.

The function will be called two times, each time with the 1813544 seed for 1000 observations, 5 predictor variables, $R^2 = 0.1$ explained variance in the underlying latent variable and $\pi_i = 0.5$.

```
res <- lapply(1:2, function(x) {set.seed(1813544)
                                genBin(N=N, nvar=nv, R2=R2, perc=p)$out})
```

The percentage of observations with the same values in two calls is 100%.

Replicate many times with different parameters

Replicate the function 100 times by varying $R^2 \in [0.1, 0.5]$ and $\pi_i \in [0.1, 0.5]$ and setting the seed to 1813544.

```
set.seed(seed)
res <- lapply(1:nrow(conds),
              function(x) rowMeans(replicate(nrep,
                                             getRes(genBin(N,
                                                           nv,
                                                           conds[x,"p"],
                                                           conds[x,"R2"]),
                                                    "out",
                                                    "X")
                                             )
                                   )
              )
```

Results

The tables below show mean pseudo R^2 as estimated by McKelvey and Zavoina's estimator and mean percentage of 1s over 100 replications.

Mean percentage	Mean R^2	Pop. values
0.10023	0.1273988	0.1, 0.1
0.49889	0.0869309	0.5, 0.1
0.10152	0.5249056	0.1, 0.5
0.50036	0.4741064	0.5, 0.5

Or with latex:

Table 2: Varied factors.

Mean percentage	Mean R^2	Pop. values
0.10023	0.1273988	0.1, 0.1
0.49889	10.0869309	0.5, 0.1
bka bla	bla bla	bla bla

As we see, R^2_{MZ} is slightly deviates on average from the population parameter.

Session info

```
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Slovenian_Slovenia.utf8  LC_CTYPE=Slovenian_Slovenia.utf8
## [3] LC_MONETARY=Slovenian_Slovenia.utf8 LC_NUMERIC=C
```

```
## [5] LC_TIME=Slovenian_Slovenia.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] performance_0.10.1
##
## loaded via a namespace (and not attached):
##  [1] digest_0.6.31   lifecycle_1.0.3 magrittr_2.0.3  evaluate_0.19
##  [5] rlang_1.0.6     stringi_1.7.8   cli_3.4.1       rstudioapi_0.14
##  [9] vctrs_0.5.1     rmarkdown_2.19  tools_4.2.2     stringr_1.5.0
## [13] glue_1.6.2      xfun_0.35       yaml_2.3.6      fastmap_1.1.0
## [17] compiler_4.2.2  htmltools_0.5.4 insight_0.18.8  knitr_1.41
```

The end

A picture of a cute ferret to say goodbye, taken from this page.

Figure 1: A baby ferret.