```
#####################
# load libraries
# set wd
# clear global .envir
#####################

# remove objects
rm(list=ls())
# detach all libraries
detachAllPackages <- function() {
  basic.packages <- c("package:stats", "package:graphics",
"package:grDevices", "package:utils", "package:datasets",
"package:methods", "package:base")
  package.list <- search()[ifelse(unlist(gregexpr("package:",
search()))==1, TRUE, FALSE)]
  package.list <- setdiff(package.list, basic.packages)
  if (length(package.list)>0)  for (package in package.list)
detach(package,  character.only=TRUE)
}
detachAllPackages()

# load libraries
pkgTest <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[,  "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg,  dependencies = TRUE)
  sapply(pkg,  require,  character.only = TRUE)
}

# here is where you load any necessary packages
# ex: stringr
# lapply(c("stringr"),  pkgTest)

lapply(c("stringr","stats","ggplot2","rms"),  pkgTest)


# set wd for current folder
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
print(getwd())
#####################
# Problem 1
#####################

# load data
load(url("https://github.com/ASDS-TCD/StatsII_Spring2022/blob/main/
datasets/climateSupport.RData?raw=true"))

summary(climateSupport)
str(climateSupport)

# clean missing value if there are
climateSupport[climateSupport == "?"] <- NA
limateSupport <- na.omit(climateSupport)
```

```
str(climateSupport)

#convert factor "choice" to 0 and 1 according to its levels of value( it
means "Not supported" = 0, "Supported" = 1)
levels(climateSupport$choice) <- c("0", "1")

head(climateSupport)

# have a look the relationship between the variable countries and choice,
sanctions and choice.
ggplot(data = climateSupport, mapping = aes(x = choice, y = countries))+
  geom_jitter() +
  coord_flip()
ggplot(data = climateSupport, mapping = aes(x = choice, y = sanctions))+
  geom_jitter() +
  coord_flip()


#1 Run the logit regression
reg <- glm(choice ~ countries + sanctions,
           data = climateSupport,
           family = binomial(link = "logit"))

# summary output:
summary(reg)

dummy.coef(reg)
# Dummy variables are created: we can take the level in country number [20
of 192] and the level in sanctions[None] which are left as the respective
reference categories.
# so all other variable levels (dummy variables) could be interpered in
reference to them.


# Create the global null hypotheis with Likelihood ratio test

# Ho: none of the explanatory variables influences the choice on supporting
the policy (all slope =0)
# Ha: at least one slope not equal to 0

reg_null <- glm(choice ~ 1, data = climateSupport, family = "binomial") # 1
= fit an intercept only (i.e. sort of a "mean")
anova(reg_null, reg, test = "Chisq")

# P-value conducted from above <0.01, so we can reject global null
hypothesis, which means at least one predictor (in countires and sanctions)
is reliable in model.
# all the explanatory variables mentioned here are significant, because all
their P-value less than 0.05.
```

#2(a) We want to compare the odd change from 5% to 15% in sanction, for making it easier, we set reference to catagory sanction 5% by relevelling sanction scale.

```
levels(climateSupport$sanctions)
climateSupport$sanctions <-factor(climateSupport$sanctions, ordered =
FALSE)
climateSupport$sanctions <-relevel(climateSupport$sanctions, "5%")
levels(climateSupport$sanctions)

# set the log regression model in this condition:

reg2 <- glm(choice ~ countries + sanctions,
            data = climateSupport,
            family = binomial(link = "logit"))
summary(reg2)
```

I try to consider sanctions as a confounding Z, taking values 0 and 1 for sanction 5% and sanction15%, and regard the variable countries as X,
so the model that adjusts for confounding is log(E(Y|X,Z)/(1-E(Y|X,Z)))=log(P/(1-P)) = Betao+Beta1X+Beta2Z.
so Beta1 is the log(odds) between when X=1 and X=0 for each Z. exp(Beta1) is the odds ratio.

As countries [20 of 192] is reference, so we can regards countries [80 of 192] as X=0 and countries 160 of 192 as X=1, which is exactly what we want to have for 2(a).

So in this case, sanctions15% is significant (P<0.01) and has a Beta1 estimate of -0.32510. the two groups are sanctions15% and the sanctions5% (reference group).
Therefore, exp(-0.32510) =0.7225. It means that the odds of supporing the policy in the group of sanctions15% is 0.7225 times that of
supporint the policy in the group of sanctions5% (or equivalently 27.75% [= (0.7225-1)*100] lower) when controlling for all other variables.

2(b).
As we are not taking interaction term in this model, so there shouldn't be odds difference change in sanctions from 5% to 15% with that happend in countries [20 of 192] and countries [160 of 192].
It means that for the policy in which very few countries participate [20 of 192], the odds of supporing the policy in the group
of sanctions15% is exp(-0.32510) =0.7225 times that of supporint the policy in the group of sanctions5% (or equivalently 27.75%[= (0.7225-1)*100] lower) when controlling for all other variables as above.

2(c).
We can use coeffecients in reg2 here, and calculate the predicted probability of supporting a policy (Yi=1) if there are 80 of 192 countries participating with no sanctionsn.

```
summary(reg2)
```

```
# P(Yi=1|xi) =exp(Beta0 + Beta1Xi)/[1+exp(Beta0 +Beta1Xi)] = 1/[1+exp-
(-0.19186+0.33636*1)]=0.536

# So the predicted probability of supporting a policy if there are 80 of
192 countries participating with no sanctionsn is 53.6%


# (d) we can set the hypothesis
# Ho: Beta (increasing sanctions from 5% to 15% )|countries participate[20
of 192] = Beta (increasing sanctions from 5% to 15% )|countries
participate[160 of 192] ;
# Ha: Effect of increasing sanctions from 5% to 15% is different by
countries participate


#  we can use unique() with model.matrix() to create a matrix of different
combinations of factor levels to use with predict().

model.matrix( ~ unique(sanctions), data = climateSupport)
model.matrix( ~ as.factor(unique(sanctions)), data = climateSupport)


# create data with predicted values

reg3 <- glm(choice ~ countries * sanctions, data = climateSupport, family =
"binomial")
anova(reg3, reg, test = "Chisq")

# As P-value=0.3912 >0.05,
# so we cannot reject null hypothesis, there is not strong evident shows
that an iteractive effect of
# countries participants change and sanctions in different levels is a
significant predictor for odds of supporting the policy.
```