

## Paul Kogan-HW5

Paul Kogan

2021/10/10

```
pacman::p_load(ISwR, MASS)
rjct <- function(lvl) paste("<= alpha =", lvl,
  "so reject the null\n\t\t\t\t\tthypothesis that the")
conc <- \(test, hyp, lvl, cat = T) {
  lvl <- try(1 - attr(test$conf.int, "conf.level"))
  val <- try(test$p.value)
  lvl <- ifelse(length(lvl) == 0, .05, lvl)
  val <- ifelse(length(val) == 0, test[, 5], val)
  str <- paste("conclusion: p-value =", format(round(val, 4), nsmall = 4),
    ifelse(val <= lvl, rjct(lvl), rjct(lvl) %>%
      str_replace_all(c("<=" = ">", "so" = "so fail to"))), hyp)
  if (cat) cat(str)
  invisible(str)
}
```

**1**

```
data <- read.table("d_logret_6stocks.txt", header = T)
sset <- gather(data[, c("Pfizer", "Exxon", "Citigroup")], tic, ret, 1:2)
citi <- data$Citigroup
limo <- lm(Pfizer ~ Exxon, data)
cat("\na\t Intercept:\t", limo$coefficients[[1]],
    "\n \t Exxon:\t\t", limo$coefficients[[2]], "\nb\n\t")
(vaan <- anova(limo))
conc(vaan, "regression effects are not significant")
cat("\nc\n\t")
(grup <- anova(lm(ret ~ tic, sset)))
conc(grup, "means of the groups are equal")
cat("\nd\t")
(test <- binom.test(length(citi[citi > 0]), length(citi)))
conc(test, "proportion of positive returns is 0.5")
```

```
##
## a)      Intercept:  -0.005325057
##          Exxon:      0.354649
## b)
## Analysis of Variance Table
##
## Response: Pfizer
##           Df      Sum Sq   Mean Sq F value    Pr(>F)
## Exxon      1 0.0041609 0.0041609   8.774 0.004328 **
```

```
## Residuals 62 0.0294022 0.0004742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## conclusion: p-value = 0.0043 <= alpha = 0.05 so reject the null
##      hypothesis that the regression effects are not significant
## c)
## Analysis of Variance Table
##
## Response: ret
##      Df    Sum Sq   Mean Sq F value    Pr(>F)
## tic      1 0.001878 0.00187814   3.5509 0.06182 .
## Residuals 126 0.066645 0.00052893
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## conclusion: p-value = 0.0618 > alpha = 0.05 so fail to reject the null
##      hypothesis that the means of the groups are equal
## d)
## Exact binomial test
##
## data:  length(citi[citi > 0]) and length(citi)
## number of successes = 37, number of trials = 64, p-value = 0.2604
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4481586 0.7006192
## sample estimates:
## probability of success
##      0.578125
##
## conclusion: p-value = 0.2604 > alpha = 0.05 so fail to reject the null
##      hypothesis that the proportion of positive returns is 0.5
```

2

```
library(ISwR)
data <- na.omit(data.frame(igf1 = juul$igf1, tanner = juul$tanner))
cat("\na\nn\t")
(vaan <- anova(lm(igf1 ~ tanner, data)))
conc(vaan, "igf1 means of each tanner level are equal")
cat("\nb\nn")
print.data.frame(plyr::ddply(data, ~tanner, summarise, mean = mean(igf1)))
cat("\nc")
(test <- pairwise.t.test(data$igf1, data$tanner, p.adj = "bonf"))
pval <- test$p.value
cat("Tanner level pairs that appear to have a difference:\nn\t")
for (i in 1:nrow(pval)) {
  for (j in 1:ncol(pval)) {
    e <- pval[i, j]
    if (!is.na(e) && .5 > e) {
      pair <- paste(rownames(pval)[i], "and", colnames(pval)[j])
      cat(ifelse(!(i == nrow(pval) && j == ncol(pval)),
        paste0(pair, ", "), paste("and", pair)))
    }
  }
}
```

```
##
## a)
## Analysis of Variance Table
##
## Response: igf1
##          Df    Sum Sq Mean Sq F value    Pr(>F)
## tanner      1 10985605 10985605  686.07 < 2.2e-16 ***
## Residuals 790 12649728    16012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## conclusion: p-value = 0.0000 <= alpha = 0.05 so reject the null
##             hypothesis that the igf1 means of each tanner level are equal
## b)
##   tanner      mean
## 1      1 207.4727
## 2      2 352.6714
## 3      3 483.2222
## 4      4 513.0172
## 5      5 465.3344
##
## c)
## Pairwise comparisons using t tests with pooled SD
##
## data:  data$igf1 and data$tanner
##
##      1      2      3      4
## 2 < 2e-16 -      -      -
## 3 < 2e-16 9.9e-08 -      -
## 4 < 2e-16 5.5e-13 1.000 -
## 5 < 2e-16 1.3e-11 1.000 0.048
##
## P value adjustment method: bonferroni
## Tanner level pairs that appear to have a difference:
## 2 and 1, 3 and 1, 3 and 2, 4 and 1, 4 and 2, 5 and 1, 5 and 2, and 5 and 4
```

3

```
library(MASS)
(test <- fisher.test(survey$Smoke, survey$Exer))
conc(test, "students smoking habit is independent of their exercise level")
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  survey$Smoke and survey$Exer
## p-value = 0.4138
## alternative hypothesis: two.sided
##
## conclusion: p-value = 0.4138 > alpha = 0.05 so fail to reject the null
##             hypothesis that the students smoking habit is independent of their exercise level
```