Tukey HSD tests

2024-10-04

```
df <- read.csv("scaffold_split_metrics.csv")

metric_name <- "prec"
df[[metric_name]] <- as.numeric(df[[metric_name]])
df$method <- factor(df$method)

formula <- as.formula(paste(metric_name, "~ method + (1 | cv_cycle)"))
m <- lmer(formula, data = df)
m_sum <- summary(m)$coefficients
df_residual <- as.integer(m_sum[nrow(m_sum), "df"])

test_out <- multcomp::glht(m, linfct = mcp(method = "Tukey"), df = df_residual)
summary(test_out)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmer(formula = formula, data = df)
##
## Linear Hypotheses:
##                                 Estimate Std. Error t value Pr(>|t|)
## chemprop_st - chemprop_mt == 0 -0.016123   0.006302  -2.558    0.036 *
## lgbm_morgan - chemprop_mt == 0 -0.007599   0.006302  -1.206    0.456
## lgbm_morgan - chemprop_st == 0  0.008524   0.006302   1.353    0.374
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
confint(test_out)
```

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmer(formula = formula, data = df)
##
## Quantile = 2.4187
## 95% family-wise confidence level
```

```
##
##
## Linear Hypotheses:
##                                 Estimate  lwr       upr
## chemprop_st - chemprop_mt == 0 -0.016123 -0.031366 -0.000881
## lgbm_morgan - chemprop_mt == 0 -0.007599 -0.022842  0.007643
## lgbm_morgan - chemprop_st == 0  0.008524 -0.006718  0.023767
```

```
old.par <- par(mai=c(1.5,3,1,1))
plot(test_out)
```