# hw11

Jiayi

2024-12-07

## Q1 (a)

```r
# Data
soybean <- c(158, 171, 193, 199, 230, 243, 248, 248, 250, 267, 271, 316, 327, 329)
linseed <- c(141, 148, 169, 181, 203, 213, 229, 244, 257, 260, 271, 309)

z <- c(soybean, linseed)
group <- c(rep(1, length(soybean)), rep(2, length(linseed)))

# Kolmogorov-Smirnov statistic function
ks_statistic <- function(x, group) {
  Fn <- ecdf(x[group == 1]) # Empirical CDF for group 1
  Gm <- ecdf(x[group == 2]) # Empirical CDF for group 2
  max(abs(Fn(x) - Gm(x)))
}

# Observed statistic
obs_stat <- ks_statistic(z, group)

# Randomization distribution
set.seed(123)  # Ensure reproducibility
perm_stats <- replicate(10000, {
  perm_group <- sample(group)              # Randomize group labels
  ks_statistic(z, perm_group)              # Compute KS statistic for permuted data
})

# Plot the randomization distribution
hist(perm_stats, main = "Randomization Distribution of KS Statistic",
     xlab = "KS Statistic", col = "lightblue")
abline(v = obs_stat, col = "red", lwd = 2, lty = 2)
```
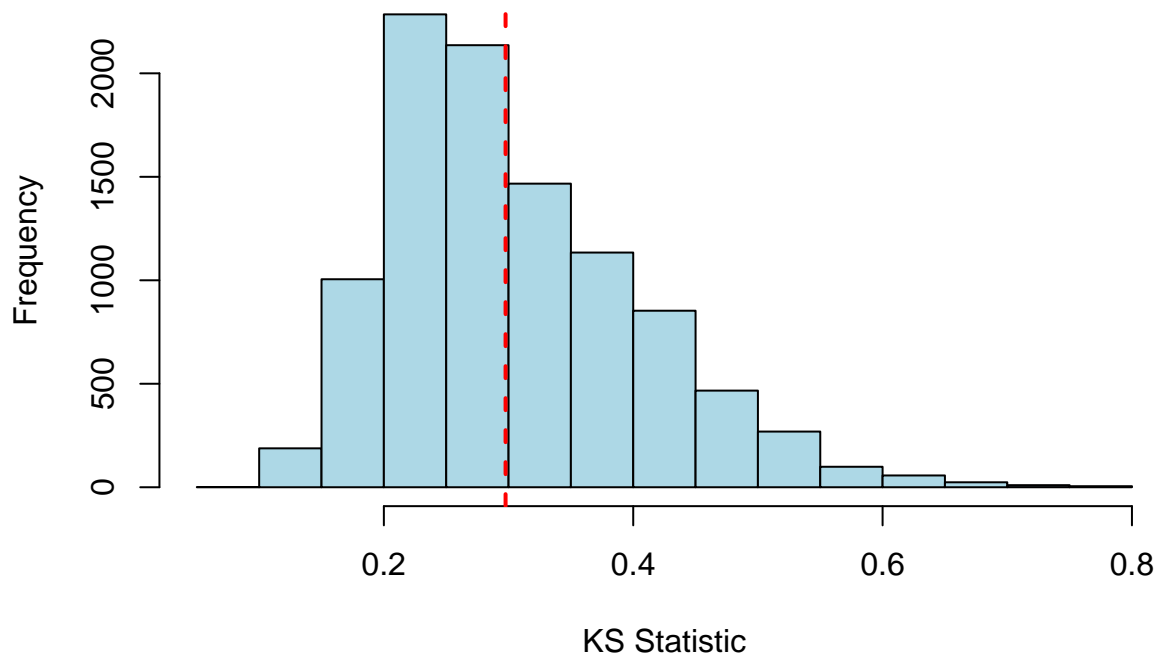
**Randomization Distribution of KS Statistic**



## Q1 (b)

```r
# Compute p-value
p_value <- mean(perm_stats >= obs_stat)
cat("The p-value is:", p_value)
```

```
## The p-value is: 0.4574
```

## Q2 (a)

```r
# Load the data
data(alfalfa, package = "faraway")

# Fit the model
model <- aov(yield ~ inoculum + shade + irrigation, data = alfalfa)

# Summary of the model
summary(model)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
```

```
## inoculum      4 155.89    38.97  12.709 0.000284 ***
## shade         4  87.40    21.85   7.125 0.003533 **
## irrigation    4  16.56     4.14   1.350 0.307872
## Residuals    12  36.80     3.07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table:

The treatment factor *inoculum* has a significant effect on yield (F = 12.709, p = 0.000284), as the p-value is *less than the significance level of 0.05.*

The blocking factor *shade* also has a significant effect on yield (F = 7.125, p = 0.003533), as its p-value is *below 0.05.*

The blocking factor *irrigation*, however, does not have a significant effect on yield (F = 1.350, p = 0.307872), as its *p-value is greater than 0.05.*

Thus, the treatment factor and the shade blocking factor are significant, while the irrigation blocking factor is not.

# Q2 (b)

```
# Tukey's Honest Significant Difference Test
tukey_results <- TukeyHSD(model, "inoculum", conf.level = 0.95)

# Display results
print(tukey_results)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = yield ~ inoculum + shade + irrigation, data = alfalfa)
##
## $inoculum
##      diff        lwr       upr      p adj
## B-A -0.72  -4.250202  2.810202 0.9633433
## C-A -0.08  -3.610202  3.450202 0.9999928
## D-A -0.86  -4.390202  2.670202 0.9326392
## E-A -6.60 -10.130202 -3.069798 0.0005166
## C-B  0.64  -2.890202  4.170202 0.9759059
## D-B -0.14  -3.670202  3.390202 0.9999332
## E-B -5.88  -9.410202 -2.349798 0.0014163
## D-C -0.78  -4.310202  2.750202 0.9515868
## E-C -6.52 -10.050202 -2.989798 0.0005764
## E-D -5.74  -9.270202 -2.209798 0.0017334
```

```
# Significant pairwise differences
significant <- tukey_results$inoculum[tukey_results$inoculum[, "p adj"] < 0.05, ]
significant
```

```
##      diff       lwr       upr      p adj
```

3

```
## E-A -6.60 -10.130202 -3.069798 0.0005166455
## E-B -5.88  -9.410202 -2.349798 0.0014163428
## E-C -6.52 -10.050202 -2.989798 0.0005764154
## E-D -5.74  -9.270202 -2.209798 0.0017334480
```

From the Tukey HSD test results:

Pairwise comparisons reveal that *inoculum E* is significantly different from several other levels *at an experiment-wise significance level of 0.05*:

- `  E-A: p = 0.0005166`

- `  E-B: p = 0.0014163`

- `  E-C: p = 0.0005764`

- `  E-D: p = 0.0017334`

No other pairwise comparisons show significance as their adjusted p-values exceed 0.05. In conclusion, inoculum level E stands out as being significantly different from levels A, B, C, and D, while the other pairwise differences are not statistically significant.