```
# A marketing research consultant evaluated the effects of the fee schedule, scope of work,
# and type of supervisory control on the quality of work performed under contract by
# independent marketing research agencies. The quality of work performed was measured
# by an index taking into account several characteristics of quality. Four agencies were
# chosen for each factor level combination and the quality of their work evaluated.



# a) Regress quality on agency, fee and an interaction between sup and scope.
# State the estimated regression equation and use drop1 to test which terms are significant.




mrcontract = expand.grid(agency=LETTERS[1:4], sup=c("local","travel"),scope=c("in-house", "subco
ntract"), fee=c("high","med","low"))

mrcontract$quality=c(124.3,120.6,120.7,122.6,112.7,110.2,113.5,108.6,115.1,119.9,115.4,117.3,88.
2,96,96.4,90.1,119.3,118.9,125.3,121.4,113.6,109.1,108.9,112.3,117.2,114.4,113.4,120,92.7,91.1,9
0.7,87.9,90.9,95.3,88.8,92,78.6,80.6,83.5,77.1,89.9,83,86.5,82.7,58.6,63.5,59.8,62.3)



# write.csv(my_data, "my_data.csv", row.names = FALSE)

View(mrcontract)
```

```
# Create the interaction term
mrcontract$sup_scope <- interaction(mrcontract$sup, mrcontract$scope)

# Fit the model
model <- lm(quality ~ agency + fee + sup_scope, data = mrcontract)

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

```
##
## Call:
## lm(formula = quality ~ agency + fee + sup_scope, data = mrcontract)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0208 -1.9292 -0.3406  1.8458  4.5167
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 122.4542     1.1360 107.796  < 2e-16 ***
## agencyB                       0.1250     1.0710   0.117    0.908
## agencyC                       0.1500     1.0710   0.140    0.889
## agencyD                      -0.5667     1.0710  -0.529    0.600
## feemed                       -0.9625     0.9275  -1.038    0.306
## feelow                      -31.1563     0.9275 -33.591  < 2e-16 ***
## sup_scopetravel.in-house    -10.9500     1.0710 -10.224 1.36e-12 ***
## sup_scopelocal.subcontract   -5.4417     1.0710  -5.081 9.70e-06 ***
## sup_scopetravel.subcontract -30.2333     1.0710 -28.229  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.623 on 39 degrees of freedom
## Multiple R-squared:  0.9838, Adjusted R-squared:  0.9805
## F-statistic: 295.9 on 8 and 39 DF,  p-value: < 2.2e-16
```

```
# Regression equation derived using OLS - >


# quality = 122.45 + 0.125 x agencyB + 0.15 x agencyC + (-0.56) x agencyD + (-0.96) x feemed +
(-31.15) x feelow
#          + (-10.95) x sup_scopetravel.in-house + (-5.44) x sup_scopelocal.subcontract + (-30.2
3) x sup_scopetravel.subcontract
```

```
drop1(model, test = "F")
```

```
## Single term deletions
##
## Model:
## quality ~ agency + fee + sup_scope
##           Df Sum of Sq     RSS     AIC  F value Pr(>F)
## <none>                   268.4 100.624
## agency     3       4.1   272.5  95.344   0.1964 0.8982
## fee        2   10044.3 10312.7 271.757 729.7061 <2e-16 ***
## sup_scope  3    6241.2  6509.6 247.672 302.2756 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Inferences that can be made from the drop1 test.

# 1. Dropping agency leads to an increase in RSS or Sum of squared Errors by 4.1.
# The p-value is much higher than 0.05, hence based on the results it can be inferred
# that agency does not significantly contribute to the model.

# 2. Dropping fee leads to an increase in RSS or Sum of squared Errors by 10044.3.
# The p-value is much lesser than 0.05, hence based on the results it can be inferred
# that fee does significantly contribute to the model.

# 3. Dropping interaction between sup and scope leads to an increase in RSS or Sum of squared Er
rors by 6241.2.
# The p-value is much lesser than 0.05, hence based on the results it can be inferred
# that the interacation term does significantly contribute to the model.
```

```
# b) Are there differences in quality between the agencies? To receive full credit state the nul
l and alternative hypotheses, find the P
# value, state you decision (reject or not), and summarize your conclusion.


# Load dplyr package
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Fit the linear model
model <- lm(mrcontract$quality ~ mrcontract$agency, data = mrcontract)

# Perform ANOVA
anova_result <- anova(model)

print(anova_result)
```

```
## Analysis of Variance Table
##
## Response: mrcontract$quality
##                   Df  Sum Sq Mean Sq F value Pr(>F)
## mrcontract$agency  3     4.1    1.35  0.0036 0.9997
## Residuals         44 16553.8  376.22
```

```
# Optional: Check if the ANOVA result is significant
if (anova_result$`Pr(>F)`[1] < 0.05) {
  cat("There are significant differences between the groups.\n")
} else {
  cat("There are no significant differences between the groups.\n")
}
```

```
## There are no significant differences between the groups.
```

```
# The p-value is much greater than 0.05 for the ANOVA test. Hence, we cannot reject the null hyp
othesis,
# i.e. there are no differences in quality between the agencies


# The null hypothesis in an ANOVA test is ->

# H0 - There are no differences among the group means.
# muAgency-A = muAgency-B = muAgency-C = muAgency-D

# H-alternative -> At least one group mean is different from the others.
```

```
# c) Are there differences in quality between the fee values? To receive full credit state the n
ull and alternative hypotheses, find the P
# value, state you decision (reject or not), and summarize your conclusion.


# Load dplyr package

library(dplyr)

# Fit the linear model
model <- lm(mrcontract$quality ~ mrcontract$fee, data = mrcontract)

# Perform ANOVA
anova_result <- anova(model)

# Print the results
print(anova_result)
```

```
## Analysis of Variance Table
##
## Response: mrcontract$quality
##                Df  Sum Sq Mean Sq F value    Pr(>F)
## mrcontract$fee  2 10044.3  5022.1  34.696 7.645e-10 ***
## Residuals      45  6513.6   144.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Optional: Check if the ANOVA result is significant
if (anova_result$`Pr(>F)`[1] < 0.05) {
  cat("There are significant differences between the groups.\n")
} else {
  cat("There are no significant differences between the groups.\n")
}
```

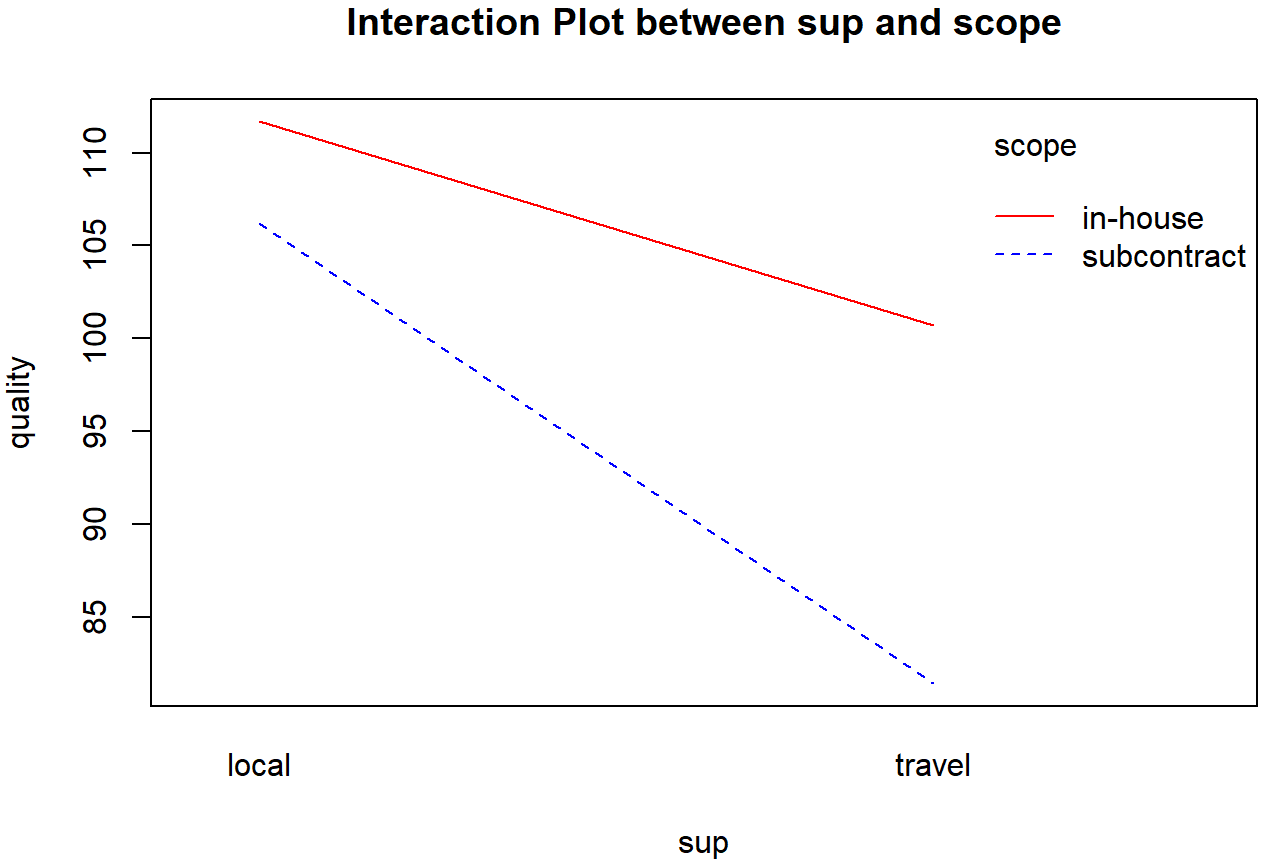
```
## There are significant differences between the groups.
```

```
# The p-value is much smaller than 0.05 for the ANOVA test. Hence, we can reject the null hypoth
esis,
# i.e. there are differences in quality between the fees levels.


# The null hypothesis in an ANOVA test is ->

# H0 - There are no differences among the group means for different fees levels.
# muAgency-A = muAgency-B = muAgency-C = muAgency-D

# H-alternative -> At least one group mean is different from the others.
```

```
# d) What does the coefficient for feemed tell you? Test whether it is different from 0 and disc
uss what the results of this tell you from a # managerial perspective


# feemed -> Estimated coeff metrics derived from the

#     Estimated coeff -> -0.9625
#     Std. Error -> 0.9275
#     t-value -> -1.038
#     Pr(>|t|) -> 0.306


# The co-efficient tells us that there is a negative impact of -0.9625 to the quality (relative
to high fees),
# when the fee is medium level.
# However, the p-value for this term is much higher than 0.05. The null-hypothesis is that this
term (fee-med) has no impact on quality,
# i.e. the co-efficient is 0. Hence, it can be said that it is not different from 0.

# From a managerial perspective, a plausible inference is that when the fees is medium level, th
e workers are neither too motivated
# to perform in a way that exceeds expectations, nor in a way that is below subpar.







# e) Is the interaction between sup and scope significant? To receive full credit state
# the null and alternative hypotheses, find the P value, and state you decision
# (reject or not).


# Reviewing the results from the drop-1 test for the interaction term.

#     sup_scope
#         Df->                3
#         Sum of Sq ->        6241.2
#         RSS ->              6509.6
#         AIC ->              247.672
#         F value ->          302.2756
#         Pr(>F) ->           <2e-16 ***


# We can clearly see that the sum of square errors increases by 6241.2 when sup_scope (the inter
action term) is dropped.
# And the p-value for the F-test is much less than 0.05.

# Hence, we can reject the null hypothesis that the interaction between sup and scope is insigni
ficant.
```

```
# The null hypothesis is:->

#     Predictor can be removed from the model without affecting model's effectiveness or R.S.S.
(residual sum of squares),
#     i.e. the coefficient is 0.

# The alternative hypothesis is:->
#     Removing the predictor from the model will affecting model's effectiveness or R.S.S. (resi
dual sum of squares) by
#     increasing it significantly, i.e. the coefficient is not 0.
```

```
# Create an interaction plot
interaction.plot(mrcontract$sup, mrcontract$scope, mrcontract$quality,
                 xlab = "sup", ylab = "quality",
                 trace.label = "scope",
                 main = "Interaction Plot between sup and scope",
                 col = c("red", "blue"), lty = 1:2)
```




```
# Answer: The presence of non-parallel lines in an interaction plot indicates that
```