

hw12

Jiayi

2024-12-14

Q1.(a)

```
# Load the necessary library
library(faraway)

# Load the dataset
data(denim)

fixed_model <- lm(waste ~ supplier, data = denim)
summary(fixed_model)

##
## Call:
## lm(formula = waste ~ supplier, data = denim)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.432  -4.377  -1.323   2.639  61.368
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.5227     2.1021   2.152  0.0341 *
## supplier2      4.3091     2.9728   1.450  0.1507
## supplier3      0.3089     3.0879   0.100  0.9206
## supplier4      2.9667     3.0879   0.961  0.3392
## supplier5      5.8542     3.4491   1.697  0.0931 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.86 on 90 degrees of freedom
## Multiple R-squared:  0.04901,    Adjusted R-squared:  0.006747
## F-statistic:  1.16 on 4 and 90 DF,  p-value: 0.334

anova(fixed_model)

## Analysis of Variance Table
##
## Response: waste
##           Df Sum Sq Mean Sq F value Pr(>F)
## supplier   4  450.9  112.730   1.1596  0.334
## Residuals 90 8749.1   97.212
```

The p-value is 0.334, which is much larger than 0.05. Therefore, the supplier is not significant at 0.05.

Q1.(b)

```
library(lme4)

## Loading required package: Matrix

# Fit a mixed model with supplier as a random effect
random_model <- lmer(waste ~ 1 + (1 | supplier), data = denim)
summary(random_model)

## Linear mixed model fit by REML ['lmerMod']
## Formula: waste ~ 1 + (1 | supplier)
## Data: denim
##
## REML criterion at convergence: 702.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.9095 -0.4363 -0.1669  0.3142  6.3817
##
## Random effects:
## Groups Name Variance Std.Dev.
## supplier (Intercept) 0.6711 0.8192
## Residual 97.3350 9.8658
## Number of obs: 95, groups: supplier, 5
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 6.997 1.078 6.49

# Estimated variance components
VarCorr(random_model)

## Groups Name Std.Dev.
## supplier (Intercept) 0.81918
## Residual 9.86585
```

The estimated variance components for the random effects model are as follows:

- Variance for supplier: 0.6711
- Residual variance: 97.3350

This indicates that the variability in waste is primarily due to residual variance, with a small contribution from variability between suppliers.

Q1.(c)

```
# Compute ICC
variance_components <- as.data.frame(VarCorr(random_model))
icc <- variance_components$vcov[1] / sum(variance_components$vcov)
icc
```

```
## [1] 0.006847128
```

The intraclass correlation is 0.006847128.

Q1.(d)

```
library(boot)
```

```
##
## Attaching package: 'boot'

## The following objects are masked from 'package:faraway':
##
##      logit, melanoma
```

```
# Parametric bootstrap for Supplier variance
set.seed(123)
boot_res <- bootMer(random_model, FUN = function(x) {
  as.numeric(VarCorr(x)$supplier[1, 1])
}, nsim = 1000, use.u = TRUE)

# Bootstrap p-value
observed_variance <- VarCorr(random_model)$supplier[1, 1]
p_value <- mean(boot_res$t >= observed_variance)
p_value
```

```
## [1] 0.345
```

The p-value is 0.345.

Q2.(a)

```
# Load the dataset
data(eggprod)

# Fit the mixed model
egg_model <- lmer(eggs ~ treat + (1 | block), data = eggprod)
summary(egg_model)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: eggs ~ treat + (1 | block)
## Data: eggprod
##
## REML criterion at convergence: 85.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.71233 -0.47453 -0.02845  0.64196  1.42942
##
## Random effects:
## Groups Name Variance Std.Dev.
## block (Intercept) 129.9 11.40
## Residual 386.9 19.67
## Number of obs: 12, groups: block, 4
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 349.00 11.37 30.702
## treatF -6.25 13.91 -0.449
## treatO -42.50 13.91 -3.056
##
## Correlation of Fixed Effects:
## (Intr) treatF
## treatF -0.612
## treatO -0.612 0.500
```

```
library(emmeans)
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```
best_treatment <- emmeans(egg_model, "treat")
print(best_treatment)
```

```
## treat emmean SE df lower.CL upper.CL
## E 349 11.4 7.99 323 375
## F 343 11.4 7.99 317 369
## O 306 11.4 7.99 280 333
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

Treatment E is the best treatment, with significantly higher egg production compared to *Treatment O* ($t=-3.056$ gives a p-value 0.015 smaller than 0.05). Also, The confidence intervals for Treatments E and O do not overlap, which confirms the difference between them is statistically significant.

However, Treatment E is *not significantly better than Treatment F* ($t=-0.449$ gives a p-value 0.667 smaller than 0.05). Also, the confidence intervals for Treatments E and F overlap, which means the difference between them is not statistically significant.

Q2.(b)

```
# Test the significance of block effects using likelihood ratio test
null_model <- lm(eggs ~ treat, data = eggprod)
anova(egg_model, null_model)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: eggprod
```

```
## Models:
```

```
## null_model: eggs ~ treat
```

```
## egg_model: eggs ~ treat + (1 | block)
```

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
## null_model	4	113.58	115.52	-52.788	105.58			
## egg_model	5	114.89	117.31	-52.444	104.89	0.6871	1	0.4071

The significant level is 0.4071, which is much larger than 0.05. This means the random effects for blocks are not significant.