STAT 523 HW 5

Md Muhtasim Billah

2/20/2020

Problem 1.

An experiment investigated the effects of different bleaching chemicals on pulp brightness. Three chemicals were selected at random from a large population of potential bleaching chemicals. Summary statistics on pulp brightness are given below:

Bleaching Chemical	Sample Size	Sample Mean	Sample SD
1	6	72.843	8.607
2	6	83.386	5.146
3	5	77.257	5.460

An ANOVA table for analyzing the data is given below:

Source	DF	SS	MS	F	p-value
Chemical	2	336.0	168.02	3.781	0.0486
Error	14	622.1	44.44		
Total	16	958.1			

Refer to formula set 10K for all the problems below.

a. Briefly explain why a random-effects model is appropriate here.
b. Carry out the hypothesis test based on the ANOVA table at the $\alpha = 0.10$ level of significance. Show all 4 steps in hypothesis testing that we follow in class. Use the p-value provided above.
c. Estimate σ^2 the variability due to bleaching chemicals. Show some calculations.
d. What is an estimate of total variance in a single pulp brightness value?
e. What proportion of total variation in a single pulp brightness value is attributed to differences among bleaching chemicals?

Answer

a.

For this experiment, the three chemicals were not fixed by the researcher, rather, they were selected at random from a large population of potential bleaching chemicals. If we use a one way fixed model, the model will be "overparameterized". To avoid this, a one way random-effects model will be more appropriate for this experiment.

b.

The 4-step hypothesis testing is given below based on the provided ANOVA table.

Step1:

Null hypothesis, $H_0 : \sigma_A^2 = 0$.
Alternative hypothesis, $H_a : \sigma_A^2 > 0$.
Level of significane, $\alpha = 0.10$

Step2:

From the ANOVA table, F-statistic,

$$F = 3.781$$

Step3:

From the ANOVA table,

$$p - value = 0.0486$$

Step4:

It is evident that,

$$0.0486 < 0.10$$

$$p - value < \alpha$$

So, we can reject the null hypothesis. This indicates that there is difference in the pulp brightness across different bleaching chemicals.

c.

The variability due to the bleaching chemicals,

$$\hat{\sigma}_A^2 = \frac{MSTr - MSE}{r}$$

Where,
$MSTr = 168.02$ $MSE = 44.44$ And, for unequal sample size,

$$r = \frac{n - \sum J_i^2/n}{I - 1}$$

$$r = \frac{17 - (6^2 + 6^2 + 5^2)/17}{3 - 1}$$

$$r = 5.647$$

Finally, pluggin in all these values, we get,

$$\hat{\sigma}_A^2 = \frac{168.02 - 44.44}{5.647}$$

So, the estimate of the variability due to the bleaching chemicals is,

$$\hat{\sigma}_A^2 = 21.88$$

d.

An estimate of total variance in a single pulp brightness value,

$$Var(X_{ij}) = \hat{\sigma}^2 + \hat{\sigma}_A^2 = MSE + \hat{\sigma}_A^2 = 44.44 + 21.88 = 66.32$$

e.

Now,

$$\frac{\hat{\sigma}_A^2}{Var(X_{ij})} = \frac{21.88}{66.32} = 0.32989$$

The proportion of total variation in a single pulp brightness value that is attributed to differences among bleaching chemicals is 32.99%.

Problem 2.

(From Jay L. Devore, Chapter 11: Section 11.1, Exercise 4.) In an experiment to see whether the amount of coverage of light-blue interior latex paint depends either on the brand of paint or on the brand of roller used, one gallon of each of four brands of paint was applied using each of three brands of roller.

The two-way additive ANOVA table resulting from this data (number of square feet covered) is provided as below.

An ANOVA table for analyzing the data is given below:

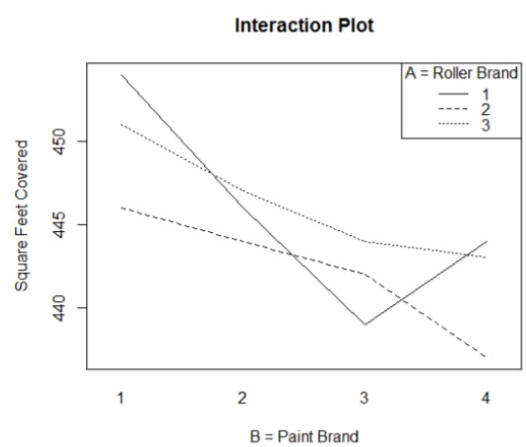
Source	DF	SS	MS	F-value	p-value
A=Roller Brand	2	38.0	19.00	2.803	0.1381
B=Paint Brand	3	159.58	53.19	7.848	0.0169
Residuals	6	40.67	6.78		
Total	11	238.25			

a. An interaction plot of the data is given on the right. For the two-way additive model and ANOVA analysis above to be valid, the plot should show (roughly) parallel curves. Discuss what the plot tells you.
b. Show by hand calculations that SSA=SS(Roller Brand)=38.00.
c. State and test hypotheses appropriate for deciding whether paint brand has any effect on coverage. Complete the 4-step procedure we follow in class. Use the given p-value in the ANOVA table.
d. Repeat part (c) for brand of roller. Use the given P-value in the ANOVA table..
e. Comment on each plot of residuals below.

Answer

a.

The given interaction plot,



From the above plot, we see that the the curve for roller brand 1 is not “roughly” parallel with roller brand 2 and 3 which indicates that there is interaction between brand 1 with brand 2 and 3. But since the brand 2 and 3 are “roughly” parallel, it can be inferred that they don’t have any interaction.

b.

Hand calculations for SSA=SS(Roller Brand) are given below.

$$SSA = J\sum_{i=1}^{I}(\bar{x}_{i.} - \bar{x}_{..})^2$$

$$SSA = 4 \times [(445.75 - 444.75)^2 + (442.25 - 444.75)^2 + (446.25 - 444.75)^2]$$

$$SSA = 4 \times [1 + 6.25 + 2.25]$$

$$SSA = 4 \times 9.5 = 38.00$$

c.

The 4-step hypotheses test appropriate for deciding whether paint brand (factor B) has any effect on coverage is shown below.

Step1:

Null hypothesis, $H_0 : \beta_1 = \beta_2 = 0$.
Alternative hypothesis, $H_a : H_0$ is false.
Level of significane, $\alpha = 0.05$

Step2:

From the ANOVA table, F-statistic,

$$F = \frac{MSB}{MSE} = 7.848$$

Step3:

From the ANOVA table,

$$p - value = 0.0169$$

Step4:

It is evident that,

$$0.0169 < 0.05$$

$$p - value < \alpha$$

So, we reject the null hypothesis. This indicates that the paint brands make a significant difference on coverage.

d.

The 4-step hypotheses test appropriate for deciding whether roller brand (factor A) has any effect on coverage is shown below.

Step1:

Null hypothesis, $H_0 : \alpha_1 = \alpha_2 = 0$.
Alternative hypothesis, $H_a : H_0$ is false.
Level of significane, $\alpha = 0.05$

Step2:

From the ANOVA table, F-statistic,

$$F = \frac{MSA}{MSE} = 2.803$$

Step3:

From the ANOVA table,

$$p - value = 0.1381$$

Step4:

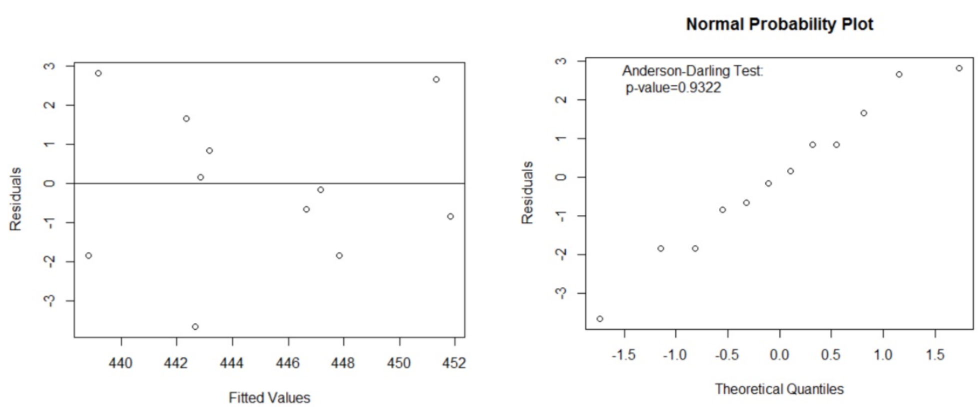
It is evident that,

$$0.1381 > 0.05$$

$$p - value > \alpha$$

So, we retain the null hypothesis. This indicates that the roller brands doesn't make a significant difference on coverage.

e.



From the residual vs fitted plot on the left, we can see that for the coverage values in the range of 446 to 448, the vertical spread is much smaller than the rest of the plot. So, it indicates that there is a possibiity that the assumption of constant variance is violated.

From the normal probability plot on the right, it seems that the residuals follow a very linear trend which indicates that the residuals belong to a normal distribution. This can also be reassured by the p-value from the Anderson-Darling test given on the plot. Since, the p-value is greater than $\alpha = 0.05$, the null is retained which means that the distribution of the residuals is normal.

Problem 3.

(From Jay L. Devore, Chapter 11: Section 11.1, Exercise 6.) A particular county employs three assessors who are responsible for determining the value of residential property in the county. To see whether these assessors differ systematically in their assessments, 5 houses are selected, and each assessor is asked to determine the market value of each house. With factor A denoting assessors ($I = 3$) and factor B denoting houses ($J = 5$), suppose $SSA = 11.7$, $SSB = 113.5$, and $SSE = 25.6$.

a. Test $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ at level 0.05. (H_0 states that there are no systematic differences among assessors.)

Answer

Here,
Mean square error for factor A, $MSA = SSA/dfA = 11.7/(3-1) = 5.85$.
Mean square error for factor B, $MSB = SSB/dfB = 113.5/(5-1) = 28.375$.
Mean square error for residuals, $MSE = SSE/(dfA * dfB) = 25.6/(2*4) = 3.2$.

The 4-step hypotheses test is shown below.

Step1:

Null hypothesis, $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$.
Alternative hypothesis, $H_a : H_0$ is false.
Level of significane, $\alpha = 0.05$.

Step2:

The F-statistic,

$$F = \frac{MSA}{MSE} = \frac{5.85}{3.2} = 1.828$$

Step3:

```r
1-pf(1.828,2,8)
```

```
## [1] 0.2219023
```

From R we get,

$$p - value = 0.222$$

Step4:

It is evident that,

$$0.222 > 0.05$$

$$p - value > \alpha$$

So, we retain the null hypothesis. This indicates that there are no systematic differences among assessors.