Homework 6

Brian Cervantes Alvarez
May 22, 2024
ST 553 Statistical Methods

Question 1

1.1

i.

Same as HW5

ii.

Same as HW5

iii.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Manufacturer	2	0.87451214	0.43725607	5.93	0.0076
StressLevel	2	10.59276333	5.29638167	71.84	<.0001
Manufactu*StressLeve	4	1.04101976	0.26025494	3.53	0.0198

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Manufacturer	2	0.86386576	0.43193288	5.86	0.0079
StressLevel	2	10.53691921	5.26845961	71.46	<.0001
Manufactu*StressLeve	4	1.04101976	0.26025494	3.53	0.0198

iv.

```r
ALT <- read.csv("ALT2.csv")

# Log-transform the response variable
ALT$log_life <- log(ALT$life)

# Convert variables to factors and set contrasts
ALT$manuf <- as.factor(ALT$manuf)
ALT$stress <- as.factor(ALT$stress)
contrasts(ALT$manuf) <- contr.sum
contrasts(ALT$stress) <- contr.sum

# Fit the model with interaction
model <- lm(log_life ~ manuf * stress, data = ALT)

# Extract coefficient estimates
coefficients <- coef(model)

# Calculate the variance of the residuals (sigma squared)
sigma_squared <- sum(resid(model)^2) / model$df.residual

# Display the results
print(coefficients)
```

```
   (Intercept)         manuf1         manuf2        stress1        stress2 
    6.49816039    -0.13842078    -0.07893290    -0.73582306     0.55410913 
manuf1:stress1 manuf2:stress1 manuf1:stress2 manuf2:stress2 
   -0.28922294     0.02960576     0.09557111     0.09591959 
```

```r
print(paste("Sigma squared:", sigma_squared))
```

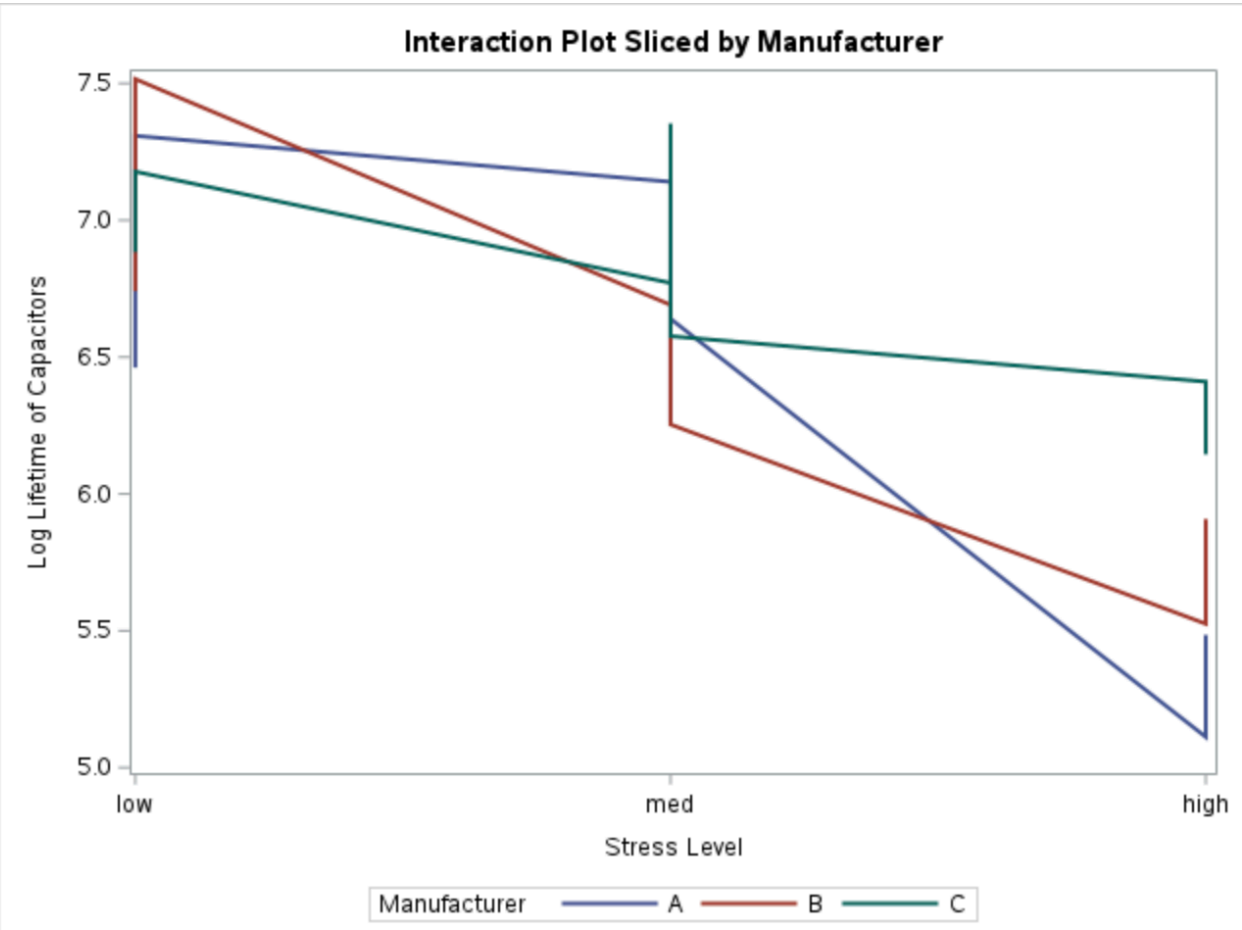
```
[1] "Sigma squared: 0.073729446912098"
```

v.

Same as HW5

vi.

```
proc sgplot data=alt2;
    series x=StressLevel y=logLifetime / group=Manufacturer lineattrs=(thickness=0.5);
    xaxis label='Stress Level';
    yaxis label='Log Lifetime of Capacitors';
    title 'Interaction Plot Sliced by Manufacturer';
run;
```



This chart helps us understand how well capacitors from three different manufacturers hold up under increasing levels of stress. Capacitors from manufacturer A generally last less long as stress increases. On the other hand, manufacturer B's capacitors show a peculiar behavior where they seem to do worse at a medium stress level but better at a higher one. In contrast, manufacturer C's capacitors do slightly better at a medium level but significantly worse at high stress levels. This plot is useful for choosing the right capacitors depending on the expected stress conditions in their usage environment.

vii.

```
proc glm data=alt2;
   class Manufacturer StressLevel;
   model logLifetime = Manufacturer StressLevel Manufacturer*StressLevel;
   where StressLevel = 'high';
   estimate 'Diff C-B High Stress' Manufacturer -1 0 1;
   estimate 'Diff C-A High Stress' Manufacturer -1 1 0;
   lsmeans Manufacturer / pdiff=all cl;
run;
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Diff C-B High Stress	0.90461458	0.11103535	8.15	<.0001
Diff C-A High Stress	0.37831659	0.11103535	3.41	0.0078

Least Squares Means for Effect Manufacturer				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
2	1	0.378317	0.088075	0.668558
3	1	0.904615	0.614373	1.194856

Diff C-B High Stress:

- Estimate: 0.9046
- Standard Error: 0.1110
- t Value: 8.15
- p-value: < 0.0001
- 95% Confidence Interval: [0.6144, 1.1949]

Diff C-A High Stress:

- Estimate: 0.3783
- Standard Error: 0.1110
- t Value: 3.41
- p-value: 0.0078
- 95% Confidence Interval: [0.0881, 0.6686]

1.2

When data from a specific group is missing, like capacitors from manufacturer B under medium stress, we can still analyze the remaining data by adjusting our approach. One way is to modify the statistical model to exclude the missing group, focusing on the data we have. Mixed-effects models and ANCOVA are helpful here. Mixed-effects models handle unbalanced data by considering the variability within groups, while ANCOVA combines ANOVA and regression to control for other factors, helping adjust for the missing data. It's important to focus on interaction terms because the main effects might be misleading without all the data. When reporting results, we need to mention the missing group and discuss how it might affect our findings. Additionally, using other statistical methods or doing sensitivity analyses can help check the strength of our conclusions from the incomplete data set.

Question 2

```
proc glmpower data=pilot;
   class A B;
   model response = A|B;
   contrast 'A1 vs A2 at B3' A 1 -1 0 A*B 0 0 1 0 0 -1 0 0 0;
   power
      stddev = 3.87298334620
      alpha = 0.05
      power = 0.80
      ntotal = .;
run;
```

Computed N Total						
Index	**Type**	**Source**	**Test DF**	**Error DF**	**Actual Power**	**N Total**
1	Effect	A	2	9	0.849	18
2	Effect	B	2	387	0.807	396
3	Effect	A*B	4	279	0.803	288
4	Contrast	A1 vs A2 at B3	1	27	0.905	36

The total number of observations required are given the table above (N Total).

Question 3

Part 3.1

Given the model $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, where $\alpha_i \sim$ iid $N(0, \sigma_\alpha^2)$ and $\epsilon_{ij} \sim$ iid $N(0, \sigma^2)$, and both are independent, the distribution of y_{ij} is,

$$y_{ij} \sim N(\mu, \sigma_\alpha^2 + \sigma^2)$$

by properties of normal distributions.

Part 3.2

Now, $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ and $y_{ij'} = \mu + \alpha_i + \epsilon_{ij'}$, so the covariance can be expressed as,

$$\mathrm{Cov}(y_{ij}, y_{ij'}) = \mathrm{Cov}(\mu + \alpha_i + \epsilon_{ij}, \mu + \alpha_i + \epsilon_{ij'}) = \mathrm{Var}(\alpha_i) = \sigma_\alpha^2$$

Thus, the variances of y_{ij} are,

$$\mathrm{Var}(y_{ij}) = \sigma_\alpha^2 + \sigma^2$$

Therefore, the correlation is given,

$$\rho(y_{ij}, y_{ij'}) = \frac{\mathrm{Cov}(y_{ij}, y_{ij'})}{\sqrt{\mathrm{Var}(y_{ij})\mathrm{Var}(y_{ij'})}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma^2}$$

Part 3.3

Since $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ and $y_{i'j} = \mu + \alpha_{i'} + \epsilon_{i'j}$, and α_i and $\alpha_{i'}$ are independent, as well as ϵ_{ij} and $\epsilon_{i'j}$, the covariance is,

$$\text{Cov}(y_{ij}, y_{i'j}) = 0$$

Hence, the correlation is,

$$\rho(y_{ij}, y_{i'j}) = \frac{0}{\sqrt{(\sigma_\alpha^2 + \sigma^2)(\sigma_\alpha^2 + \sigma^2)}} = 0$$

Part 3.4

Using $\mathbf{Y} = [y_{11}, y_{12}, y_{21}, y_{22}, y_{31}, y_{32}, y_{41}, y_{42}]'$, we can write the $Var(Y)$ as,

$$\text{Var}(\mathbf{Y}) = \sigma_\alpha^2 \mathbf{J} + \sigma^2 \mathbf{I}$$

where $\mathbf{J}$ is an 8×8 matrix of ones and $\mathbf{I}$ is the 8×8 identity matrix.

As a result, the variance-covariance matrix Σ can be fully expanded to,

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 + \sigma^2 & \sigma_\alpha^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\alpha^2 + \sigma^2 & \sigma_\alpha^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\alpha^2 + \sigma^2 & \sigma_\alpha^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\alpha^2 + \sigma^2 & \sigma_\alpha^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma^2 \end{pmatrix}$$