

## STAT512 – Exam 2

### Spring 2018

**Honor Pledge:** I have not given, received, or used any unauthorized assistance on this exam.

**Signature:** \_\_\_\_\_

**Printed Name:** \_\_\_\_\_

**Instructions:**

- **Open book, open notes, calculator required. No computers or cell phones.**
- **Time limit is 1 hour 50 minutes - strictly enforced!**
- If an answer is in the computer output, use it; don't calculate it by hand.
- Show your work where appropriate. Put your final answer in the box (if provided).
- Make explanations brief and legible.
- All questions are worth 4 points except where noted. Maximum score is 100.
- Computer input/output is provided at the end of the exam.
- The exam contains a total of 13 pages (including computer input/output).
- If you run out of space, you may use the blank area on page 6.

**Questions 1 through 2:** An experiment was done to investigate the material hardness (**Y**) of 4 **trts** (A,B,C,D). The experiment took a total of 4 **days** to complete, because a group of 4 units are sprayed and left to dry in a ventilated hood overnight and tested the following day. It was thought that **position** (1,2,3,4) in the hood may affect the resulting hardness (**Y**) of the material. A design was implemented such that each treatment was represented exactly once on each day (1,2,3,4) and each position (1,2,3,4). There are a total of 4 trts \* 4 reps =16 observations.

	Day1	Day2	Day3	Day4
Pos1	A	D	C	B
Pos2	B	A	D	C
Pos3	C	B	A	D
Pos4	D	C	B	A

1. What is the name for this design? Circle one answer.

CRD

RCB

Latin Square

Split-Plot

2. Provide R code to fit an appropriate model. (I am just looking for a single line of R code.)

**Questions 3 through 13 (Feed Trial):** A study was done to investigate a new vitamin supplement for cattle. A successful supplement would increase **ADG** (average daily gain) in cattle. Three **Feed** formulations (A, B, C) and two levels of **Suppl** (0 or 1) were considered. Feeds are expected to be different, so the research question focused on the effect of **Suppl** for each **Feed**. A total of 24 pens of cattle were randomly assigned such that there were 4 pens for each of the 6 treatment combinations (2 **Suppl** x 3 **Feed**). **ADG** was recorded at the end of the study. The R input and output are labeled **Feed Trial**. Use  $\alpha = 0.05$ .

**Questions 3 through 7:** The following questions use **just the Modell1** results.

3. Considering the **emmeans1** **\$emmeans** output, a colleague notices that that SE is the same for each of the treatment combinations ( $SE = 0.072$ ). Briefly explain why the SE is the same.
  
  
  
  
  
  
  
  
  
  
4. Considering the **emmeans1** and **emmeans2** **\$contrasts** output, both are comparing levels of **Suppl**. Briefly explain the difference between the two sets of comparisons (for someone with little knowledge of statistics).
  
  
  
  
  
  
  
  
  
  
5. Calculate the **F test statistic** corresponding to the main effect of **Suppl**.

6. What is the **p-value** corresponding to the test of main effect of **Suppl**? Hint: This information appears elsewhere in the output.

7. A colleague notices that the **Feed:Suppl** interaction is not significant ( $F = 2.301$ ,  $p = 0.1288$ ). He says that since the interaction is not significant, based on the “Factorial Principle” that it is wrong to present the **emmeans1** comparisons and only the **emmeans2** comparisons should be considered. Do you agree? Briefly discuss.

**Questions 8 through 10 (Feed Trial continued):** The following questions compare **Models 1 and 2**.

8. Briefly explain the difference between Models 1 and 2.
9. Is there any advantage to using Model2 (as compared to Model1)? Discuss. Hint: Think in terms of power, but there may be other approaches.
10. Is there any disadvantage to using Model2 (as compared to Model1)? Discuss. Hint: Consider the emmeans3 and emmeans4 results.

**Questions 11 through 13 (Feed Trial continued):** Now **return to Model 1** and consider the following scenarios.

11. Suppose (just for this question), that the data had not been balanced. For example, 2 (out of 6) treatment combinations had 3 pens (instead of 4) for a total of 22 pens for the study. What modifications to the R code would be required?
12. Suppose (just for this question), that the investigators had recorded the initial weight (Weight0) of each pen and wanted to include Weight0 as a covariate in the model. Provide R code to fit an appropriate model. (I am just looking for a single line of R code.)
13. Each pen included 5 cows. In our analysis, the ADG value is really the average over the 5 individual cows. It is known that ADG for cows within the same pen will be correlated. Suppose (just for this question), that we had the ADG values for each of the 120 cows (6 Trts \* 4 Pens/Trt \* 5 cows/Pen) and the data contained a column giving PenID (1-24). Provide R code to fit an appropriate model. (I am just looking for a single line of R code).

**Questions 14 through 19 (Germination):** A study was done to investigate the effect of temperature on germination rate of seeds for a certain variety of wheat. Temperature (**Temp**) ranged between 5 and 30 degrees (Celsius). At each level of temperature 40 seeds were tested (**Ntest** = 40) and the number that germinated (**Ngerm**) were recorded. The goal of the analysis is to model the probability of germination as a function of temperature. The R input and output are labeled **Germination**.

14. Briefly explain the difference between **Models 1 and 2**. Note: I am looking for more than the R function!
15. For **Model1**, identify and interpret the slope. Be specific.
16. For **Model2**, identify and interpret the “slope”. Be specific.
17. Comparing the plots of **both models**, does one model seem to fit better? Or is there any evidence of “lack of fit” for either model?
18. Looking at the predicted values for **both models**, we get similar predicted values at Temp = 18 but noticeably different predicted values at Temp = 60. Explain why the predicted values are so different at Temp = 60. Be specific. (I am looking for more than that we are using two different models).

Temp	PredMod1	PredMod2
18	0.527	0.529
60	1.085	0.918

19. Do you feel “confident” in the prediction at Temp = 60 for Model1? What about Model2? Briefly discuss.

**Questions 20 through 25:** An RCB analysis was done using R. There are a total of 3 Trts \* 5 Blocks =15 observations. See ANOVA table below.

Anova Table (Type III tests)					
	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	333.55	1	733.732	3.716e-09	***
Block	5.75	4	3.161	0.0778	.
Trt	53.27	2	58.587	1.668e-05	***
Residuals	3.64	8			

20. Calculate the SE for pairwise **differences of Trt** means (ex: Trt1 vs Trt2). Show your work to receive full credit.

21. Would it be reasonable to include Block\*Trt interaction in this model? Discuss. Hint: Think about df.

22. Suppose (just for this question) that Block had been dropped from the model, would dfResid be higher or lower? (2 pts)

Higher      Lower

23. Suppose (just for this question) that Block had been dropped from this model, would SSResid be higher or lower? (2 pts)

Higher      Lower

24. All other things held equal, will a model with higher dfResid have higher or lower power? (2 pts)

Higher      Lower

25. All other things held equal, will a model with higher MSResid have higher or lower power? (2 pts)

Higher      Lower

Questions continue on the next page....

**Questions 26 through 27:** An experiment was done to compare Yield (Y) for 5 **Varieties** \* 3 **Herbicides** = 15 treatment combinations. Four **Blocks** of fifteen plots were laid out in a field. Within each block, five **Varieties** (A,B,C,D,E) were randomly assigned to North-South columns. Within each column, three **Herbicide** treatments (1,2,3) were randomly assigned. There are a total of 15 trts \* 4 blocks = 60 observations. The layout for block 1 is shown below (but remaining blocks are not shown!).

D2	A2	E1	B3	C1
D3	A1	E2	B1	C3
D1	A3	E3	B2	C2

26. What is the name for this design? Circle one answer.

CRD

RCB

Latin Square

Split-Plot

27. Provide R code for analysis. (I am just looking for a single line of R code.)

## Feed Trial (Questions 3 through 13)

```
library(dplyr)
library(car)
library(emmeans)
FeedData$Suppl <- as.factor(FeedData$Suppl)
str(FeedData)
SumStats <- summarize(group_by(FeedData, Feed, Suppl),
                      n = n(),
                      mean = mean(ADG),
                      sd   = sd(ADG),
                      SE    = sd/sqrt(n))

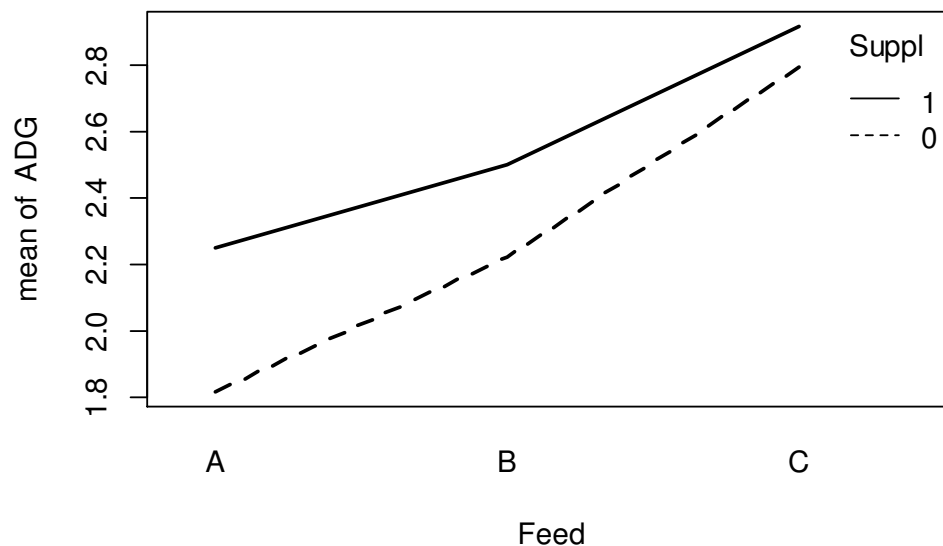
SumStats
with(interaction.plot(Feed, Suppl, ADG, lwd = 2), data = FeedData)
#MODEL1
options(contrasts = c("contr.sum", "contr.poly"))
Modell1 <- lm(ADG ~ Feed*Suppl, data = FeedData)
Anova(Modell1, type = 3)
#EMMEANS1
emmeans(Modell1, pairwise ~ Suppl|Feed)
#EMMEANS2
emmeans(Modell1, pairwise ~ Suppl)
#MODEL2
Model2 <- lm(ADG ~ Feed + Suppl, data = FeedData)
Anova(Model2, type = 3)
#EMMEANS3
emmeans(Model2, pairwise ~ Suppl|Feed)
#EMMEANS4
emmeans(Model2, pairwise ~ Suppl)
```

---

```
> library(dplyr)
> library(car)
> library(emmeans)
> FeedData$Suppl <- as.factor(FeedData$Suppl)
> str(FeedData)
Classes 'tbl_df', 'tbl' and 'data.frame':   24 obs. of  4 variables:
 $ Feed : chr  "A" "A" "A" "A" ...
 $ Suppl: Factor w/ 2 levels "0","1": 1 1 1 1 2 2 2 2 1 1 ...
 $ ADG  : num  1.55 1.81 1.99 1.93 2.2 ...
> SumStats <- summarize(group_by(FeedData, Feed, Suppl),
+                          n = n(),
+                          mean = mean(ADG),
+                          sd   = sd(ADG),
+                          SE    = sd/sqrt(n))
```

## Feed Trial continued (Questions 3 through 13)

```
> SumStats
# A tibble: 6 x 6
# Groups:   Feed [?]
  Feed Suppl      n mean      sd      SE
  <chr> <fct> <int> <dbl> <dbl> <dbl>
1 A     0         4  1.82  0.195  0.0976
2 A     1         4  2.25  0.045  0.0224
3 B     0         4  2.22  0.086  0.0430
4 B     1         4  2.50  0.167  0.0833
5 C     0         4  2.79  0.147  0.0735
6 C     1         4  2.92  0.165  0.0826
> with(interaction.plot(Feed, Suppl, ADG, lwd = 2), data = FeedData)
```



```
> #MODEL1
> options(contrasts = c("contr.sum", "contr.poly"))
> Model11 <- lm(ADG ~ Feed*Suppl, data = FeedData)
> Anova(Model11, type = 3)
Anova Table (Type III tests)
Response: ADG
```

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	140.052	1	6764.833	< 2.2e-16	***
Feed	2.752	2	66.458	4.885e-09	***
Suppl	0.459	1			
Feed:Suppl	0.095	2	2.301	0.1288688	
Residuals	0.373	18			

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



## Feed Trial continued (Questions 3 through 13)

```
> #EMMEANS1
> emmeans(Model1, pairwise ~ Suppl|Feed)
$emmeans
Feed = A:
  Suppl    emmean      SE df lower.CL upper.CL
0      1.816961 0.07194272 18 1.665815 1.968107
1      2.247085 0.07194272 18 2.095939 2.398231

Feed = B:
  Suppl    emmean      SE df lower.CL upper.CL
0      2.220061 0.07194272 18 2.068915 2.371207
1      2.498621 0.07194272 18 2.347475 2.649768

Feed = C:
  Suppl    emmean      SE df lower.CL upper.CL
0      2.794944 0.07194272 18 2.643798 2.946090
1      2.916420 0.07194272 18 2.765274 3.067566

$constrasts
Feed = A:
  contrast    estimate      SE df t.ratio p.value
0 - 1      -0.4301239 0.1017424 18  -4.228  0.0005

Feed = B:
  contrast    estimate      SE df t.ratio p.value
0 - 1      -0.2785603 0.1017424 18  -2.738  0.0135

Feed = C:
  contrast    estimate      SE df t.ratio p.value
0 - 1      -0.1214767 0.1017424 18  -1.194  0.2480

> #EMMEANS2
> emmeans(Model1, pairwise ~ Suppl)
NOTE: Results may be misleading due to involvement in interactions
$emmeans
  Suppl    emmean      SE df lower.CL upper.CL
0      2.277322 0.04153615 18 2.190058 2.364586
1      2.554042 0.04153615 18 2.466778 2.641306

$constrasts
  contrast    estimate      SE df t.ratio p.value
0 - 1      -0.2767203 0.05874098 18  -4.711  0.0002

Results are averaged over the levels of: Feed
```

## Feed Trial continued (Questions 3 through 13)

```
> #MODEL2
> Model12 <- lm(ADG ~ Feed + Suppl, data = FeedData)
> Anova(Model12, type = 3)
Anova Table (Type III tests)
Response: ADG
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	140.052	1	5986.075	< 2.2e-16 ***
Feed	2.752	2	58.807	4.204e-09 ***
Suppl	0.459	1	19.637	0.0002567 ***
Residuals	0.468	20		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #EMMEANS3
> emmeans(Model12, pairwise ~ Suppl|Feed)
$emmeans
Feed = A:
```

Suppl	emmean	SE	df	lower.CL	upper.CL
0	1.893662	0.06244515	20	1.763404	2.023921
1	2.170383	0.06244515	20	2.040124	2.300641

```
Feed = B:
```

Suppl	emmean	SE	df	lower.CL	upper.CL
0	2.220981	0.06244515	20	2.090723	2.351239
1	2.497701	0.06244515	20	2.367443	2.627960

```
Feed = C:
```

Suppl	emmean	SE	df	lower.CL	upper.CL
0	2.717322	0.06244515	20	2.587064	2.847580
1	2.994042	0.06244515	20	2.863784	3.124300

```
$contrasts
Feed = A:
```

contrast	estimate	SE	df	t.ratio	p.value
0 - 1	-0.2767203	0.06244515	20	-4.431	0.0003

```
Feed = B:
```

contrast	estimate	SE	df	t.ratio	p.value
0 - 1	-0.2767203	0.06244515	20	-4.431	0.0003

```
Feed = C:
```

contrast	estimate	SE	df	t.ratio	p.value
0 - 1	-0.2767203	0.06244515	20	-4.431	0.0003

See next page...

## Feed Trial continued (Questions 3 through 13)

```
> #EMMEANS4
> emmeans(Model12, pairwise ~ Suppl)
$emmeans
  Suppl    emmean      SE df lower.CL upper.CL
0      2.277322 0.04415539 20 2.185215 2.369428
1      2.554042 0.04415539 20 2.461936 2.646149

$contrasts
  contrast    estimate      SE df t.ratio p.value
0 - 1     -0.2767203 0.06244515 20  -4.431  0.0003
```

Results are averaged over the levels of: Feed

## Germination (Questions 14 through 19)

```
GermData$PropGerm <- GermData$Ngerm/GermData$Ntest
GermData
#Model1
Model1 <- lm(PropGerm ~ Temp, data = GermData)
summary(Model1)
#Model2
Model2 <- glm(cbind(Ngerm, Ntest-Ngerm) ~ Temp, family = binomial(link = "
logit"), data = GermData)
summary(Model2)
#Predictions (both models)
NewData <- data.frame(Temp = c(18, 60))
Predictions <- data.frame(NewData,
                           PredMod1 = predict(Model1, NewData),
                           PredMod2 = predict(Model2, NewData,
                                                type = "response"))

Predictions
```

---

```
> GermData$PropGerm <- GermData$Ngerm/GermData$Ntest
> GermData
  Temp Ntest Ngerm PropGerm
1     5    40    13   0.325
2    10    40    17   0.425
3    15    40    21   0.525
4    20    40    23   0.575
5    25    40    24   0.600
6    30    40    27   0.675

> #Model1
> Model1 <- lm(PropGerm ~ Temp, data = GermData)
> summary(Model1)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.288333   0.026730  10.787 0.000419 ***
Temp         0.013286   0.001373   9.678 0.000638 ***

Residual standard error: 0.02871 on 4 degrees of freedom
Multiple R-squared:  0.959,    Adjusted R-squared:  0.9488
F-statistic: 93.67 on 1 and 4 DF,  p-value: 0.0006377
```

## Germination continued (Questions 14 through 19)

```
> #Model2
> Model2 <- glm(cbind(Ngerm, Ntest-Ngerm) ~ Temp, family = binomial(link =
"logit"), data = GermData)
> summary(Model2)
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.87292     0.30515  -2.861 0.004228 **
Temp         0.05491     0.01585   3.464 0.000533 ***

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 13.11149  on 5  degrees of freedom
Residual deviance:  0.54339  on 4  degrees of freedom
AIC: 29.117

> #Predictions (both models)
> NewData <- data.frame(Temp = c(18, 60))
> Predictions <- data.frame(NewData,
+                           PredMod1 = predict(Model1, NewData),
+                           PredMod2 = predict(Model2, NewData, type = "response"))
> Predictions
  Temp PredMod1 PredMod2
1   18 0.5274762 0.5288117
2   60 1.0854762 0.9184448

> #Plots (both models)
> #Model1
> plot(PropGerm ~ Temp, data = GermData, main = "Model1")
> abline(coef(Model1))
> #Model2
> TempRange <- seq(0, 35, 0.5)
> phat <- predict(Model2, list(Temp = TempRange), type = "response")
> plot(PropGerm ~ Temp, data = GermData, main = "Model2")
> lines(phat ~ TempRange)
```

