Exam 3 Review

Review 1

The holidays are upon us! You want to know if happiness (scale of 1-10) this time of year is impacted by which holiday folks celebrate (Christmas, Hanukkah, Kwanzaa) and ugly sweaters (ugly, not ugly). You're specifically interested in whether these interact. That is, you think that an ugly Christmas sweater might bring more joy than an ugly Hanukkah sweater or ugly Kwanzaa sweater. Here are the data:

	Ugly Sweater	Not Ugly Sweater
Christmas	7, 6, 9, 8	4, 5, 3, 5
Hanukkah	4, 8, 10, 7	4, 3, 4, 4
Kwanzaa	8, 6, 5, 5	2, 3, 2, 5

Calculate the ANOVA for main effects and interaction. Be able to interpret all *F* tests. Do this using a model comparison approach.

Fill out this table:

	SS	df	MS	F
Holiday	$SS_{Holiday}$	$df_{Holiday}$	$\frac{SS_{Holiday}}{df_{Holiday}}$	$\frac{MS_{Holiday}}{MS_W}$
Sweater	$SS_{Sweater}$	$df_{Sweater}$	$\frac{SS_{Sweater}}{df_{Sweater}}$	$\frac{MS_{Sweater}}{MS_W}$
Holiday x Sweater	SS_{HxS}	df_{HxS}	$\frac{SS_{HxS}}{df_{HxS}}$	$\frac{MS_{HxS}}{MS_W}$
Error/Residuals	SS_W	df_W	$\frac{SS_W}{df_W}$	

Review 1 Answers

Create DataFrame

```
holidays <- data.frame(
  Happiness = c(7, 6, 9, 8, 4, 5, 3, 5, 4, 8, 10, 7,
                4, 3, 4, 4, 8, 6, 5, 5, 2, 3, 2, 5),
  Holiday = rep(c("Christmas", "Hanukkah", "Kwanzaa"), each = 8),
  Sweater = rep(c(rep("Ugly", 4), rep("Not Ugly", 4)), 3)
) %>%
  mutate(
    Holiday = as.factor(Holiday),
    Sweater = as.factor(Sweater)
  )

holidays <- holidays %>%
  select(Holiday, Sweater, Happiness)
```

```
holidays %>% group_by(Holiday) %>%
  slice_head(n = 3) %>%
  kable("html") %>%
  kable_styling(font_size = 16)
```

Holiday	Sweater	Happiness
Christmas	Ugly	7
Christmas	Ugly	6
Christmas	Ugly	9
Hanukkah	Ugly	4
Hanukkah	Ugly	8
Hanukkah	Ugly	10
Kwanzaa	Ugly	8
Kwanzaa	Ugly	6
Kwanzaa	Ugly	5

Determine Degrees of Freedom

```
# three cells for holiday: christmas, hanukkah, kwanzaa
df_holiday <- 3 - 1
df_sweater <- 2 - 1
df_interaction <- (3-1) * (2-1)

#holiday has 3 levels, sweater has 2 levels
df_w <- 24 - (3 * 2)

c(df_holiday, df_sweater, df_interaction, df_w)
```

```
## [1]  2  1  2 18
```

Fill out DF Column:

	SS	df	MS	F
Holiday	$SS_{Holiday}$	2	$\frac{SS_{Holiday}}{df_{Holiday}}$	$\frac{MS_{Holiday}}{MS_W}$
Sweater	$SS_{Sweater}$	1	$\frac{SS_{Sweater}}{df_{Sweater}}$	$\frac{MS_{Sweater}}{MS_W}$
Holiday x Sweater	SS_{HxS}	2	$\frac{SS_{HxS}}{df_{HxS}}$	$\frac{MS_{HxS}}{MS_W}$
Error/Residuals	SS_W	18	$\frac{SS_W}{df_W}$	

Sum of Squares for Holiday (Restricted Model)

```
grandMean <- mean(holidays$Happiness)

#making a function to calc error
calculate_error <- function(data, predictions) {
  errors <- data$Happiness - predictions
  sum(errors^2)
}

#also known as the error of the restricted model
ss_restricted <- calculate_error(holidays, rep(grandMean, nrow(holidays)
ss_restricted
```

```
## [1] 110.9583
```

Sum of Squares for Holiday (Full Model)

```
holiday_means <- holidays %>%
  group_by(Holiday) %>%
  summarize(mean_happiness = mean(Happiness), .groups = "drop")

predictions_holiday <- holidays %>%
  left_join(holiday_means, by = "Holiday") %>%
  pull(mean_happiness)

#also known as the error of the full model
ss_full_holiday <- calculate_error(holidays, predictions_holiday)
ss_full_holiday
```

```
## [1] 102.875
```

Sum of Squares Holiday is...

```
SSHoliday <- ss_restricted - ss_full_holiday
dfh <- nlevels(holidays$Holiday) - 1
MSHoliday <- SSHoliday / dfh
SSHoliday
```

```
## [1] 8.083333
```

SS Sweater

```
sweater_means <- holidays %>%
  group_by(Sweater) %>%
  summarize(mean_happiness = mean(Happiness), .groups = "drop")

predictions_sweater <- holidays %>%
  left_join(sweater_means, by = "Sweater") %>%
  pull(mean_happiness)

#calcing the error of the full model for sweater
ss_full_sweater <- calculate_error(holidays, predictions_sweater)
SSSweater <- ss_restricted - ss_full_sweater
SSSweater
```

```
## [1] 63.375
```

SS Interaction

The goal is to create a table like this, and calculate the SS Interaction:

	Ugly Sweater	Not Ugly Sweater	Row Means
Christmas	7, 6, 9, 8	4, 5, 3, 5	**5.875**
Hanukkah	4, 8, 10, 7	4, 3, 4, 4	**5.5**
Kwanzaa	8, 6, 5, 5	2, 3, 2, 5	**4.5**
Column Means	**6.92**	**3.67**	**5.29**

SS Interaction

```
means = holidays %>%
  group_by(Holiday, Sweater) %>%
  summarize(m = mean(Happiness))

means
```

```
## # A tibble: 6 × 3
## # Groups:   Holiday [3]
##   Holiday   Sweater      m
##   <fct>     <fct>    <dbl>
## 1 Christmas Not Ugly  4.25
## 2 Christmas Ugly      7.5
## 3 Hanukkah  Not Ugly  3.75
## 4 Hanukkah  Ugly      7.25
## 5 Kwanzaa   Not Ugly  3
## 6 Kwanzaa   Ugly      6
```

SS Interaction

Holiday	Sweater	mean_happiness	row_mean	col_mean	interaction_effect
Christmas	Not Ugly	4.25	5.875	3.666667	0.000
Christmas	Ugly	7.50	5.875	6.916667	0.000
Hanukkah	Not Ugly	3.75	5.500	3.666667	-0.125
Hanukkah	Ugly	7.25	5.500	6.916667	0.125
Kwanzaa	Not Ugly	3.00	4.500	3.666667	0.125
Kwanzaa	Ugly	6.00	4.500	6.916667	-0.125

```
#calcing SSInteraction
SSInteraction <- sum((interaction_table$interaction_effect)^2) * 4
SSInteraction
```

```
## [1] 0.25
```

SS Within

```
predictions_within <- holidays %>%
  left_join(interaction_means, by = c("Holiday", "Sweater")) %>%
  pull(mean_happiness)

SSWithin <- calculate_error(holidays, predictions_within)
SSWithin
```

```
## [1] 39.25
```

Fill out the SS Column

	SS	df	MS	F
Holiday	8.08	2	$\frac{SS_{Holiday}}{df_{Holiday}}$	$\frac{MS_{Holiday}}{MS_W}$
Sweater	63.37	1	$\frac{SS_{Sweater}}{df_{Sweater}}$	$\frac{MS_{Sweater}}{MS_W}$
Holiday x Sweater	0.25	2	$\frac{SS_{HxS}}{df_{HxS}}$	$\frac{MS_{HxS}}{MS_W}$
Error/Residuals	39.25	18	$\frac{SS_W}{df_W}$	

Calculate the Mean Squares

```
MSHoliday <- SSHoliday / df_holiday
MSSweater <- SSSweater / df_sweater
MSInteraction <- SSInteraction / df_interaction
MSWithin <- SSWithin / df_w

c(MSHoliday, MSSweater, MSInteraction, MSWithin)
```

```
## [1]  4.041667 63.375000  0.125000  2.180556
```

Fill out the MS Column

	SS	df	MS	F
Holiday	8.08	2	4.04	$\frac{MS_{Holiday}}{MS_W}$
Sweater	63.37	1	63.37	$\frac{MS_{Sweater}}{MS_W}$
Holiday x Sweater	0.25	2	0.13	$\frac{MS_{HxS}}{MS_W}$
Error/Residuals	39.25	18	2.18	

Calculate the F Statistic

```
F_Holiday <- MSHoliday / MSWithin
F_Sweater <- MSSweater / MSWithin
F_Interaction <- MSInteraction / MSWithin

c(F_Holiday, F_Sweater, F_Interaction)
```

```
## [1]  1.85350318 29.06369427  0.05732484
```

Fill out the F Stat Column

	SS	df	MS	F
Holiday	8.08	2	4.04	1.854
Sweater	63.37	1	63.37	29.064
Holiday x Sweater	0.25	2	0.13	0.057
Error/Residuals	39.25	18	2.18	

Double Checking w/ aov() Output

```
summary(aov(Happiness ~ Holiday * Sweater, data = holidays))
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## Holiday          2   8.08    4.04   1.854    0.185
## Sweater          1  63.37   63.37  29.064 4.01e-05 ***
## Holiday:Sweater  2   0.25    0.12   0.057    0.944
## Residuals       18  39.25    2.18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

--

What do we conclude about the main effects?

Do the effects of Holiday and Sweater interact?

Review 2

Do you believe these findings? Which type of validity is threatened and why?

How about Statistical Conclusion Validity?

- low statistical power
- violations of assumptions
- fishing and the error rate problem
- unreliable measures
- restricted range
- unreliable group assignment

Review 2

Do you believe these findings? Which type of validity is threatened and why?

How about Internal Validity?

- ambiguous temporal precedence
- selection
- attrition
- history
- maturation
- regression
- testing
- instrumentation

Review 2

Do you believe these findings? Which type of validity is threatened and why?

How about Construct Validity?

- inadequate explication of constructs
- construct confounding
- reactive self-report changes
- reactivity to the experimental situation
- experimenter expectancy
- novelty and disruption effects

Review 2

Do you believe these findings? Which type of validity is threatened and why?

How about External Validity?

- sampling bias
- experimenter effects
- hawthorne effect
- testing effects
- situation effects

Review 3

What effect sizes have we learned? Eta-Squared!

Review 3

What effect sizes have we learned? Eta-Squared!

```
holiday.aov <- aov(Happiness ~ Holiday * Sweater, data = holidays)
eta_squared(holiday.aov, partial = FALSE)
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter       |     Eta2 |       95% CI
## ----------------------------------------
## Holiday         |     0.07 | [0.00, 1.00]
## Sweater         |     0.57 | [0.29, 1.00]
## Holiday:Sweater | 2.25e-03 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
c(SSHoliday/ss_restricted, SSSweater/ss_restricted)
```

```
## [1] 0.07285017 0.57116035
```

Review 3

Now Partial Eta-Squared!

```
holiday.aov <- aov(Happiness ~ Holiday * Sweater, data = holidays)
eta_squared(holiday.aov, partial = TRUE)
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter       | Eta2 (partial) |       95% CI
## ------------------------------------------------
## Holiday         |           0.17 | [0.00, 1.00]
## Sweater         |           0.62 | [0.35, 1.00]
## Holiday:Sweater |       6.33e-03 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
c(SSHoliday/(SSWithin + SSHoliday),
  SSSweater/(SSWithin + SSSweater),
  SSInteraction/(SSWithin + SSInteraction))
```

```
## [1] 0.170774648 0.617539586 0.006329114
```

Review 4

What tests did we learn about?

- Compare sample mean to population mean with unknown sigma
- Compare 2 means with within-groups design
- Compare 2 means with between-groups design
- Compare 3+ levels of IV
- 2+ IVs (main effects/interactions)

Review 4 Answers

Situation	Which Test?
Compare sample mean to population mean with unknown sigma	Single sample *t* test
Compare 2 means with within-groups design	Paired samples *t* test
Compare 2 means with between-groups design	Independent samples *t* test
Compare 3+ levels of IV	One-way ANOVA (between groups)
2+ IVs (main effects/interactions)	Two-way ANOVA

End of Review

- Good luck on Exam!