STAT565_Lab

Shen Qu

Mar 11, 2019

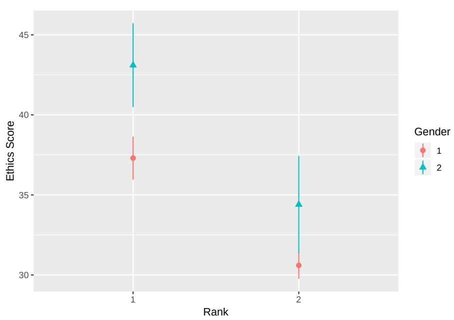
Problem 1 Unbalanced 2-Factor Factorial (Non-significant interaction)

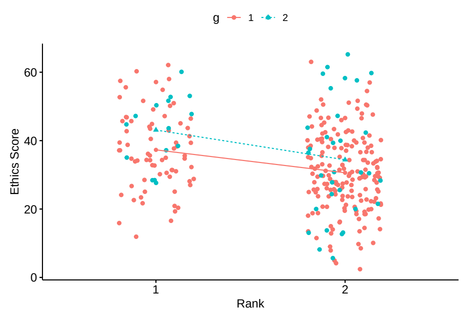
Source: R.D. White, Jr. (1999). "Are Women More Ethical? Recent Findings on the Effects of Gender Under Moral Development," Journal of Public Administration Research and Theory, Vol. 9, 3, pp.459-471. Description: Ethics scores for a sample of members of the U.S. Coast Guard by Gender (1=Male, 2=Female) and Rank (1=Officer, 2=Enlisted). Data simulated to match cell means and SDs.

Variables/Columns:Gender, 8; Rank, 16; Ethics Score,18-24.

(a) Plot the data and report the plot here (A plot with data and means of treatment combinations). Do not report code here. Describe the observed relationship between two factors.

```
## 'data.frame':    299 obs. of  5 variables:
##  $ Gender: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Rank  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Score : num  35 25.1 40.5 51 50.2 ...
##  $ g     : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ r     : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
```





(b) Obtain the numerical summary for each treatment combination and factor levels separately. Report them here in a tabular form

factorial	min	Q1	median	Q3	max	mean	sd	n	missing
Gender 1	2.42	25.04	31.92	40.11	63.03	32.51	11.6	252	0
Gender 2	5.65	27.73	37.57	49.06	65.22	37.18	15.72	47	0
Rank 1	11.91	30.38	37.19	46.62	62.1	38.3	11.34	87	0
Rank 2	2.42	23.44	30.25	39.37	65.22	31.17	12.29	212	0
GR 1.1	11.91	30	35.9	45.22	62.1	37.3	11.38	72	0
GR 2.1	27.65	36.12	44.74	51	60.11	43.1	10.14	15	0
GR 1.2	2.42	23.65	29.78	38.39	63.03	30.6	11.16	180	0
GR 2.2	5.65	21.14	30.75	44.64	65.22	34.4	17.19	32	0

(c) Fit the two-factor factorial model and report the complete ANOVA table here. Do not report code here. The complete ANOVA table should have a row for each of the following: main effects of each treatment, two-factor interaction effects, error and total.Report the adjusted (Type III) SS for each effect because of unbalanced design. Note that, adjusted SS and SSE do not add up to SSTotal

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
g	1	860.9	860.9	6.034	0.01461
r	1	3048	3048	21.36	5.693e-06
g:r	1	34.1	34.1	0.239	0.6253
Residuals	295	42087	142.7	NA	NA
Total	298	46030.1	154.46	NA	NA

Type III	Sum Sq	Df	F value	Pr(>F)
(Intercept)	180136	1	1263	1.327e-108
g	785.2	1	5.503	0.01964
r	2021	1	14.17	0.000202
g:r	34.1	1	0.239	0.6253
Residuals	42087	295	NA	NA
Total	46030.1	298	NA	NA

(d) Based on the ANOVA table write your conclusion appropriately. Perform all the necessary tests and report the conclusion along with the p-value. (no need to do pairwise comparison to save the time)

The line plot shows that two lines are approximately parallel. Difference in score between ranks is same for different genders. There is not obvious interaction effect.

According to ANOVA table, there is not a significant interaction effect from ranks and genders on the score at 5% significance level (P-value=0.6253). That means, effect of ranks and effect of genders on score is independent. Therefore, examine the main effects.

According to type III ANOVA table, there is a significant effect from genders and ranks on the score at 5% significance level (P-value=0.01964, 0.000202, respectively).

(e) Report the code here without output.

```
table_ethics <- read.table("Ethics.txt",header = T)
table_ethics$g <-  as.factor(table_ethics$Gender)
table_ethics$r <-  as.factor(table_ethics$Rank)
model_ethics <- aov(Score ~ g*r, data=table_ethics)
str(table_ethics)
ggplot(data = table_ethics, aes(x = r, y = Score, colour=g,shape = g)) +
   stat_summary() + labs(y= "Ethics Score", x="Rank", color="Gender", shape="Gender")
favstats(Score ~ Gender, data=table_ethics)
favstats(Score ~ Rank, data=table_ethics)
favstats(Score ~ Gender+Rank, data=table_ethics)
pander(summary(model_ethics))
sum((table_ethics$Score- mean(table_ethics$Score))^2)
# To obtain Adjusted SS for factorial design in R when there is an unbalanced design #
# First, it is necessary to set the contrasts option in R.#
options(contrasts = c("contr.sum","contr.poly"))
# Because the multi-way ANOVA model is over-parameterised, #
# it is necessary to choose a contrasts setting that sums to zero.#
# otherwise the ANOVA analysis will give incorrect results with respect #
# to the expected hypothesis. (The default contrasts type does not satisfy this requirement.)

# Use with drop1 function or Anova function with type = "III" as below #
drop1(model_ethics, .~., test="F")
#Type III sum of squares give the adjusted (for all other effects) SS and p values of each effect #
Anova(model_ethics, type="III")
```

Problem 2: 3-Factor Factorial (Non-significant interaction)

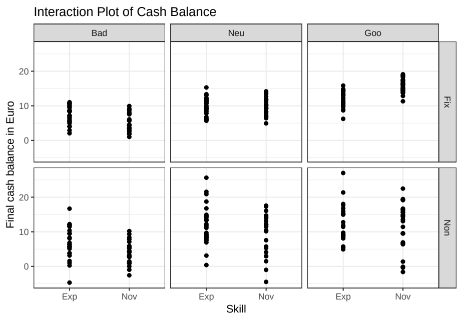
Results from 3-Factor ANOVA to investigate effects of poker skill (average/expert), poker hand (bad/neutral/good), and bet limit (fixed/none) on winnings. There were 25 individuals in each of 12 combinations (each individual in only one treatment).Use a=0.05

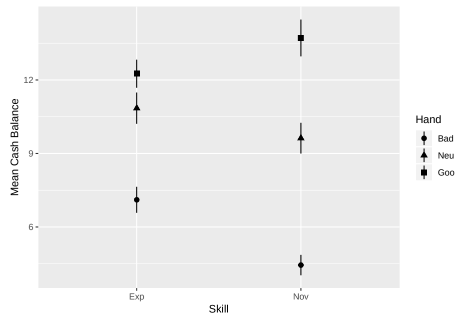
Variables/Columns: Skill: 1=Expert, 2=Average; Hand:1=Bad, 2=Neutral, 3=Good; Limit: 1=Fixed, 2=None; Final Cash Balance (euros)

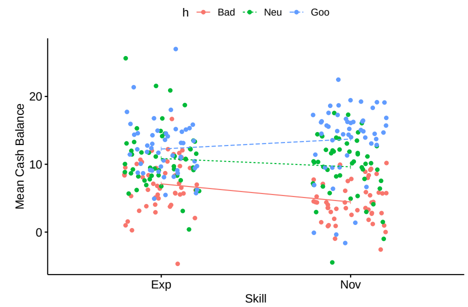
(a) Plot the data and report the plot here (A plot with data and means of treatment combinations). Do not report code here.

```
## 'data.frame':    300 obs. of  7 variables:
##  $ skill: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ hand : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ limit: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ cash : num  4 5.55 9.45 7.19 5.71 5.32 8.52 4.06 4.97 6.18 ...
##  $ s    : Factor w/ 2 levels "Exp","Nov": 1 1 1 1 1 1 1 1 1 1 ...
##  $ h    : Factor w/ 3 levels "Bad","Neu","Goo": 1 1 1 1 1 1 1 1 1 1 ...
```

```
##  $ l    : Factor w/ 2 levels "Fix","Non": 1 1 1 1 1 1 1 1 1 1 ...
```







(b) Obtain the numerical summary for each treatment combination and factor levels separately. Report them here in a tabular form

s	min	Q1	median	Q3	max	mean	sd	n	missing
Exp	-4.68	7.145	9.59	12.35	26.99	10.07	4.64	150	0
Nov	-4.47	4.4	9.415	14.16	22.48	9.262	5.755	150	0

h	min	Q1	median	Q3	max	mean	sd	n	missing
Bad	-4.68	3.4	5.7	8.41	16.68	5.777	3.618	100	0
Neu	-4.47	7.827	10.16	12.39	25.63	10.24	4.501	100	0
Goo	-1.61	9.713	13.96	15.98	26.99	12.98	4.762	100	0

s.h.l	min	Q1	median	Q3	max	mean	sd	n	missing
Exp.Bad.Fix	2.07	5.55	7.19	9.45	11.04	7.329	2.56	25	0
Nov.Bad.Fix	1.02	3.43	4.39	7.53	9.94	5.05	2.55	25	0
Exp.Neu.Fix	5.68	8.33	9.51	11.56	15.3	9.85	2.52	25	0
Nov.Neu.Fix	4.93	8.21	9.62	11.63	14.18	9.8	2.42	25	0
Exp.Goo.Fix	6.24	10.56	12.37	14.12	15.84	12.12	2.32	25	0
Nov.Goo.Fix	11.31	14.49	16.09	17.27	19.11	15.8	1.94	25	0
Exp.Bad.Non	-4.68	3.82	6.42	10.34	16.68	6.889	4.71	25	0
Nov.Bad.Non	-2.57	1.19	4.02	5.82	10.19	3.84	3.24	25	0
Exp.Neu.Non	0.39	8.31	11.14	14.53	25.63	11.85	5.771	25	0
Nov.Neu.Non	-4.47	5.32	11.48	13.93	17.56	9.45	5.861	25	0
Exp.Goo.Non	4.93	8.76	11.44	15.33	26.99	12.39	5.31	25	0
Nov.Goo.Non	-1.61	6.94	13.55	16.31	22.48	11.63	6.7	25	0

(c) Fit the three-factor factorial model and report the complete ANOVA table here. Do not report code here. The complete ANOVA table should have a row for each of the following: main effects of each treatment, two-factor interaction effects, error and total.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
s	1	49.17	49.17	2.839	0.09308
h	2	2647	1323	76.41	2.389e-27
l	1	31.68	31.68	1.829	0.1773
s:h	2	219	109.5	6.324	0.002052
s:l	1	119.1	119.1	6.878	0.009191
h:l	2	97.29	48.64	2.809	0.06192
s:h:l	2	42.38	21.19	1.224	0.2957
Residuals	288	4987	17.32	NA	NA
Total	299	8192.67	27.40	NA	NA

(d) Based on the ANOVA table write your conclusion appropriately. Perform all the necessary tests and report the conclusion along with the p-value.

The line plot shows that not all lines are parallel. Difference in cash between skills is not same for different hands. There could be an interaction effect.

According to ANOVA table, there is not a significant interaction effect from skills, hands and limits on the cash at 5% significance level (P-value=0.2957). there is a significant interaction effect from skills versus hands and skills versus limits on the cash at 5% significance level (P-value=0.002052, 0.009191, respectively). That means, effect of skills versus hands and skills versus limits on the cash are not independent. Therefore, examine the simple effects.

(e) Suppose that we want to compare the average final cash balance between two types of Skill (Expert versus Average) when the type of Hand is 3 (Good) and Limit is 1 (Fixed). This is the simple effect contrast we learned in class. Test this contrast and report conclusion with p value.

The $H0$ for contrast is $\mu_{131} - \mu_{231} = 0$. The $H1$ is they don't equal zero.

The P-value of this contrast test is 0.0019, which is small enough. We can reject the $H0$ and conclude that the average final cash balance between two types of Skill (Expert versus Average) when the type of Hand is 3 (Good) and Limit is 1 (Fixed) is different at 5% significance level.

(f) Report the code here without output.

```
#If the 3-factor interaction is significant, then test 2-factor interaction for each level of 3rd factor#
#To do all pairwise comparisons at each level of skill(s) factor  #
contrast(lsmeans(model_poker, ~h*l|s), interaction = "pairwise") # You may use and adjustment method as shown in t
#To do all pairwise comparisons at each level of hand (h) factor  #
contrast(lsmeans(model_poker, ~s*l|h), interaction = "pairwise", adjust="bonferroni")
#To write contrasts in terms of cell means #
#First, create a 3-factor interaction term of factors s, h and l #
#Then fit the model with that factor #
table_poker$shl<-interaction(table_poker$s,table_poker$h,table_poker$l)
# Save the LSmeans for interaction terms of factors s, h and l #
# Before writing a contrast, check the order of terms in previous lsmf output #
# Then create a vector of contrast in terms of LSmeans of treatment combinations #
# Below is the contrast to test mean for two levels of skill factor (Expert versus Average) #
# when the type of Hand is 3 (Good) and Limit is 1 (Fixed) #
summary(contrast(lsmeans(lm(cash~shl, table_poker), 'shl'), list(c1 = c(rep(0,4), 1, -1, rep(0,6)))))
```