Laboratory 10

Table of contents

1 Reply on the Comments from Lab 09

1.1 The line plot and `ggplot2`

Although conducting ANOVA or ANCOVA in the world of `Python` or `R` is not as straightforward as training a even more complex ML model. Plotting could be easier. So challenge accepted. Recall back our second question:

> Draw a line chart using the Plots option. Let Horizontal Axis and Separate Lines be alcohol consumption and facial attractiveness respectively, and include error bars showing the 95% confidence intervals of the means.

It's quite simple to just draw two line-plot with error bar as shown in Figure 1:

```r
library(ggplot2)
library(haven)

goggles <- read_sav("goggles.sav")

goggles$facetype <- as.factor(goggles$facetype)
goggles$alcohol <- as.factor(goggles$alcohol)

ggplot(goggles, aes(x = alcohol, y = attractiveness,
                    group = facetype, color = facetype)) +
  stat_summary(fun = mean, geom = "point") +
  stat_summary(fun = mean, geom = "line") +
```

```
  stat_summary(fun.data = mean_cl_normal, geom = "errorbar", width = 0.1) +
  scale_x_discrete(labels = c("Placebo", "Low dose", "High dose")) +
  theme_light()
```

```
Warning: Computation failed in `stat_summary()`
Caused by error in `fun.data()`:
! The package "Hmisc" is required.
```

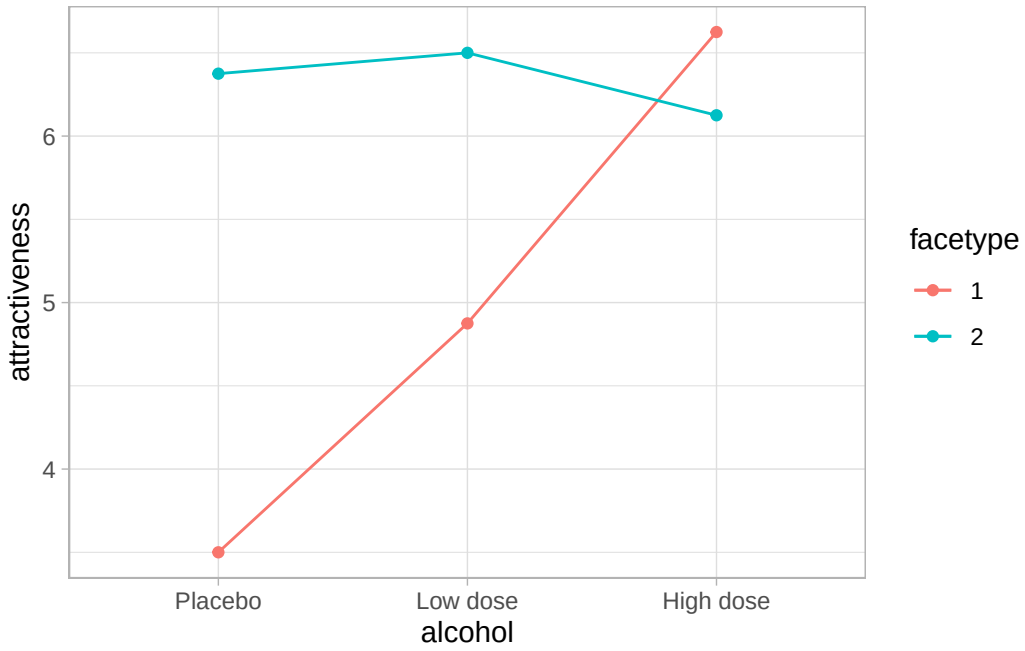

Figure 1: The interaction of type of face and alcohol consumption on the ratings

`ggplot2` is always the most elegant way to visualize the data and idea.

2 Context

A marketing manager tested the benefit of soft drinks for curing hangovers. He took 15 people and got them drunk. Then he measured how drunk the person got on a scale of `0 = straight edge` to `10 = flapping about like a haddock out of water`. The next morning as they awoke and dehydrated, he gave five of them water to drink, five of them Lucozade (a very nice glucose-based UK drink [1]) and the remaining five a leading brand of cola [2]. These people were randomly assigned to the three treatment groups. He measured how well they felt (on a scale from 0 = I feel like death to 10 = I feel really full of beans and healthy) two hours later. The dataset is saved in `hangover.sav`. The goal is to study whether different drinks (treatment) affect how the drunk people feel (outcome).

3 Objectives

1. Draw a scatterplot to check the relationship between how drunk the person was before treatment and how well they felt after treatment.

[1] And it is now owned by Japanese (SUNTORY bought it from GSK in 2013).
[2] The leading brand? That should be the reeeeeeeeeeeeed one.

2. Is there any evidence of unlucky randomization, i.e., people in the three groups differ in how drunk they were before the treatment?
3. Consider applying ANCOVA with how drunk they felt before treatment as a covariate. Assess whether there is an interaction between treatment and the covariate.
4. Conduct an ANCOVA and report the adjusted means of the three groups.
5. To study the causal relationship between the treatment and the outcome, is ANOVA a valid choice? How about ANCOVA? Which one do you prefer? Why?
6. Briefly summarize your findings.

4 Solutions

4.1 Q1-Q4, and Q6: result of ANCOVA and its findings

In this experiment, given:

- Y: Measures of how well a hangover person feel after the soft drink treatment.
- X_c: Measures of how drunk a person got last night.
- A: Types of drink. Group A_1 for water, A_2 for Lucozade and A_3 for Cola.

4.1.1 Data screening: scatterplot

The scatterplot between how drunk the person was before the treatment (X_c) and how well they felt after treatment (Y) is shown as Figure 2, X_c and Y showed a linear relation with no bivariate outliers.

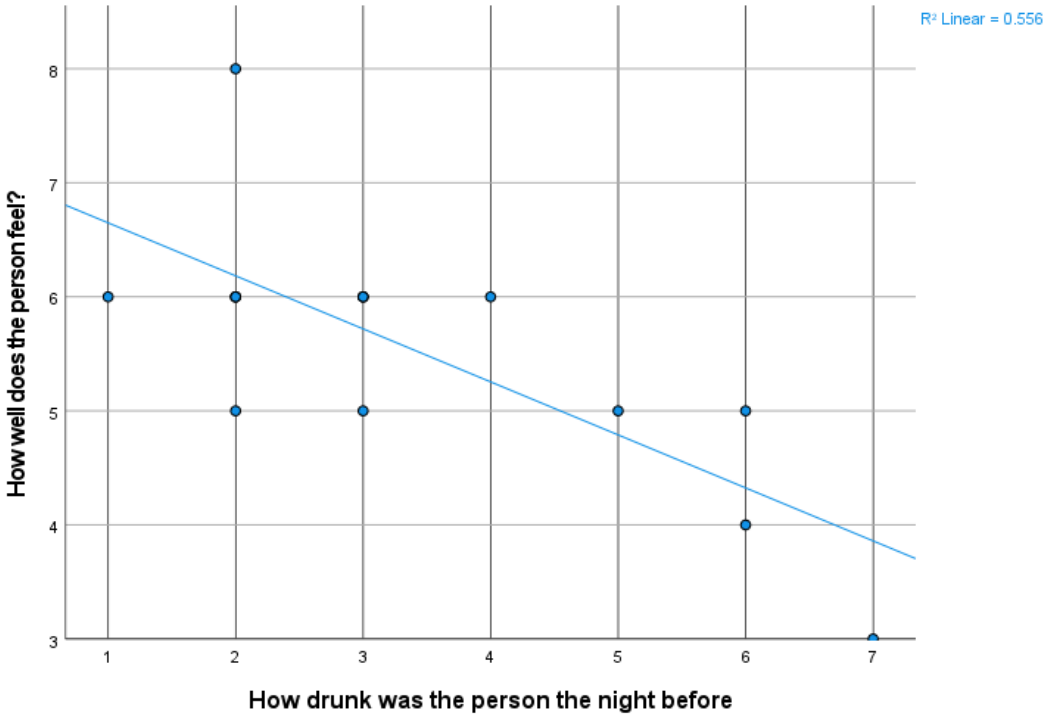

Figure 2: The outcome Y against the covariate X_c.

4.1.2 Sanity check: Independence of X_c and A

To reveal whether people in the three groups differ in how drunk they were before the treatment, a one-way ANOVA was conducted and the results were listed as Table 1.

Table 1: Result of one-way ANOVA for A on X_c.

Source	SS	df	MS	F	p
drink	8.400	2	4.200	1.355	.295
Error	37.200	12	3.100		

The result showed that the main effect of drink is not significant, with $F(2, 12) = 1.355$, $p = 0.295$, that indicated that statically there were no differences on the average drunk level between all three groups. That means the X_c and A are independent, one of the assumptions of ANCOVA appeared to be satisfied.

4.1.3 Sanity check: Homogeneity of regression slopes

Table 2: Result of ANCOVA for the interaction between A and X_c.

Source	SS	df	MS	F	p
drink * drunk	11.610	3	3.870	6.951	0.007
Error	6.124	11	0.557		

To assess whether there was an interaction between A and X_c, a preliminary ANCOVA was run using the SPSS GLM procedure with a custom model that included an $A \times X_c$ interaction term and the result was shown in Table 2. This interaction was statistically significant with an $F(3, 11) = 6.951$, $p = 0.007$. This indicates that the effect of the treatment on how well the person feels depends on how drunk they felt before the treatment.

In conclusion, there is a significant interaction between the treatment and the covariate, meaning the initial feeling of drunkenness influences the treatment effect on how well the person feels.[3] The assumption of homogeneity of regression slopes may be violated.

4.1.4 ANCOVA result and the adjusted means of Group A_1, A_2 and A_3.

Table 3: Result of ANCOVA.

Source	SS	df	MS	F	p	η_p^2
drunk	9.856	1	9.856	24.568	<0.001	0.691
drink	3.464	2	1.732	4.318	0.041	0.440
Error	4.413	11	0.401			

An ANCOVA was conducted to examine the effect of drink type on recovery while controlling for initial drunkenness and the result was shown in Table 3. The effect of how drunk the person felt the night before was significant, $F(1, 11) = 24.568$, $p < 0.001$, $\eta_p^2 = 0.691$, indicated that pre-existing drunkenness significantly affects how well the person feels. The treatment had a significant effect on how well the person felt, $F(2, 11) = 4.318$, $p = 0.041$, $\eta_p^2 = 0.440$.

The adjusted means were given in Table 4 [4], after the adjustment for the Y of drunkenness, participants who took Lucozade ($M_2 = 6.239$) felt better than participants who drunk water or cola ($M_1 = 5.110$, $M_3 = 5.252$).

[3] I tested on myself and this is true.

[4] The rank ordering of Y means across levels of groups in A changed slightly. I believe the adjusted mean is more close to the fact, see the reason in next footnote.

Table 4: Adjusted means (evaluated at initial drunkenness = 3.40).

		Mean of Y, adjusted for X_c	Unadjusted Mean of Y
Group A_1	Water	5.110	5.00
Group A_2	Lucozade	6.239	5.80
Group A_3	Cola	5.252	5.80

4.2 Q5: ANOVA or ANCOVA

Given all assumptions of ANCOVA or ANOVA were satisfied (although this may only happen in the statistician's Disneyland). In general, ANOVA is used to compare the means of different groups and determine if there are statistically significant differences between them. It assumes that all groups are treated equally and does not adjust for covariates that might affect the dependent variable. In contrast, ANCOVA extends ANOVA by including covariates that can influence the dependent variable. In this case, ANCOVA may reduce error variance and increases statistical power to detect treatment effects, especially with the small sample size (like this dataset with $n = 15$). Therefore, ANCOVA may provides a clearer understanding of the treatment effects. In this study, however, because of the significant interaction between the treatment and the covariate, the causal relationship between drink type and recovery from hangover tended to remain unclear [5][6].

5 Final thoughts

I felt lonely when I started writing this part. Initially I had many things to say but later I decided wrap them all into two sentences with bullet points.

- Thank you for all the tolerance for my foolish this semester.
- May all the cats at UM, as well as the force, be with you.

Happy thanksgiving and メリークリスマス.

[5]But it has been proven that consuming glucose-rich foods is an effective way to alleviate hangover symptoms because when people drink alcohol, their blood glucose levels tend to decrease, leading to hypoglycemia that makes people *feel like death*. The power of glucose was also confirmed by an experienced drinker who wrote this footnote.

[6]If I get another chance, I'd like to say that I prefer neither option and would choose causal inference instead. I'm determined to **almost** master this method after the Christmas and Lunar New Year breaks.

Also Greetings from our lovely cats, the Prince (in Figure 3) and A-bao (阿煲, see Figure 4). A-bao is the cat—and the real boss—at a coffee shop I love. If you ever pass through Guangzhou, I'll let the shop owner know and treat you to a hot Americano.

Figure 3: The Prince

Figure 4: A-bao