

Julian Burgoff

11/21/2022

Analysis of Environmental Data Lab

Lab 10

1. `rm(list = ls())`

```
rope = read.csv(here("data", "rope.csv"))
rope$rope.type = factor(rope$rope.type)
```

```
n_obs = length(rope$rope.type)
n_groups = length(unique(rope$rope.type))
```

```
ss_tot = sum((rope$p.cut - mean(rope$p.cut))^2)
df_tot = n_obs - 1
```

```
agg_sum_sq_resids = aggregate(
  x = rope$p.cut,
  by = list(rope$rope.type),
  FUN = function(x) sum((x - mean(x))^2)
)
ss_within = sum(agg_sum_sq_resids$x)
df_within = n_obs - n_groups
```

```
ss_among = ss_tot - ss_within
df_among = n_groups - 1
```

```
ms_within = ss_within / (n_obs - n_groups)
ms_among = ss_among / (n_groups - 1)
```

```
f_ratio = ms_among / ms_within
f_pval = pf(f_ratio, df1 = df_among, df2 = df_within, lower.tail = FALSE)
```

2. I do not think there are equal variances among the groups, they have different widths of error bars.

3. `bartlett.test(p.cut ~ rope.type, data=rope)`

4. The p-value of the bartlett test is significant which means we can reject the null hypothesis that there is no difference in variances among the groups. This means that we will violate the assumption of ANOVA that the distributions have the same variance.

5. Rope.typeBLAZE
6. 0.36714
7. $0.36714 + (-0.10164) = 0.26550$

8. Shapiro-Wilk normality test

```
data: residuals(fit_rope_1)
W = 0.91144, p-value = 7.238e-07
```

9. The p-value is less than 0.05 so we reject the null hypothesis that the residuals are normally distributed.

10.

```
VEL=subset(rope, rope.type %in% c("VEL"))
resids_VEL= VEL$p.cut-mean(VEL$p.cut)
shapiro.test(resids_VEL)
```

Shapiro-Wilk normality test

```
data: resids_VEL
W = 0.90171, p-value = 0.0444
```

```
SB=subset(rope, rope.type %in% c("SB"))
resids_SB= SB$p.cut-mean(SB$p.cut)
shapiro.test(resids_SB)
```

Shapiro-Wilk normality test

```
data: resids_SB
W = 0.91358, p-value = 0.0746
```

```
BLAZE= subset(rope, rope.type %in% c("BLAZE"))
resids_BLAZE= BLAZE$p.cut-mean(BLAZE$p.cut)
shapiro.test(resids_BLAZE)
```

Shapiro-Wilk normality test

```
data: resids_BLAZE
W = 0.79481, p-value = 0.0005471
```

```
BS= subset(rope, rope.type %in% c("BS"))
resids_BS= BS$p.cut-mean(BS$p.cut)
shapiro.test(resids_BS)
```

Shapiro-Wilk normality test

```
data: resids_BS
W = 0.92292, p-value = 0.1128
```

```
PI= subset(rope, rope.type %in% c("PI"))
resids_PI= PI$p.cut-mean(PI$p.cut)
shapiro.test(resids_PI)
```

Shapiro-Wilk normality test

```
data: resids_PI
W = 0.92594, p-value = 0.129
```

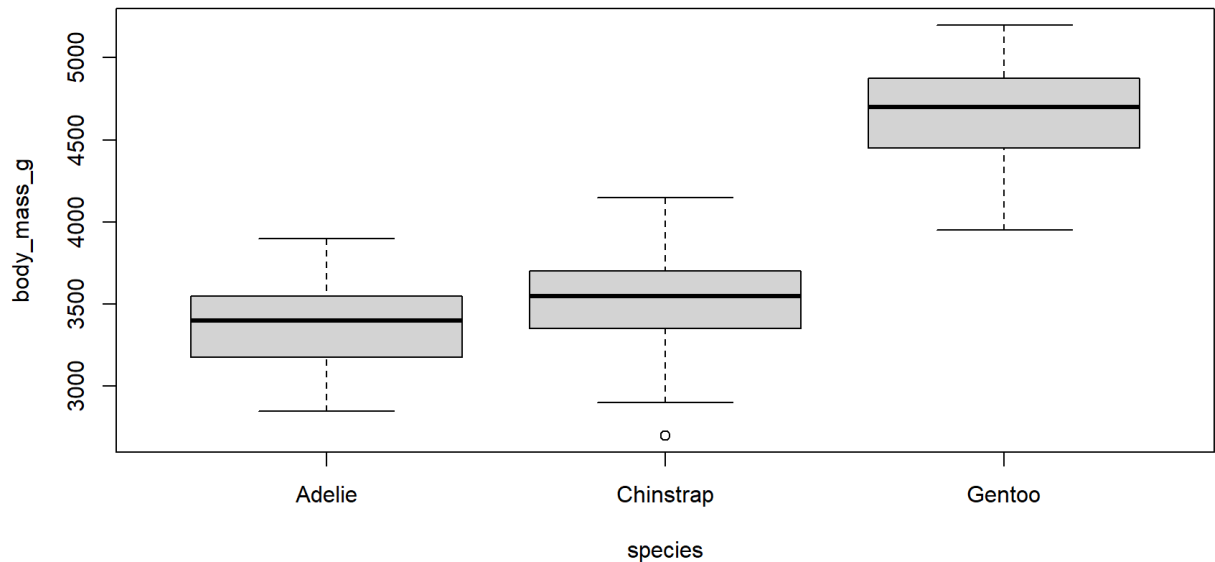
```
XTC= subset(rope, rope.type %in% c("XTC"))
resids_XTC= XTC$p.cut-mean(XTC$p.cut)
shapiro.test(resids_XTC)
```

Shapiro-Wilk normality test

```
data: resids_XTC
W = 0.88636, p-value = 0.02307
```

The residuals are normally distributed for PI, BS, and SB rope types.

11. Given the results of the normality tests, half of the groups do not meet the assumption of normality which might affect the integrity of the comparisons of means in the ANOVA.



12.

13. The variances look extremely even judging by the symmetry of the whiskers on each box.

14. p-value = 0.9056. The p-value is extremely high which indicates that the variances are very similar among groups.

15. Shapiro-Wilk normality test

data: lm_pen\$residuals

W = 0.99077, p-value = 0.3639

The p value is greater than 0.05 which means that the residuals are normally distributed and meet the normality assumption.

16. All the pairs of penguin species have significantly different body masses.

17. Chinstrap and Adelie look pretty similar visually, but are statistically significantly different. Gentoo looks different than the other two groups.