# Exam 2

Aurora Travers

2022-09-08

I. Introduction

We want to find which food and amount of food is most effective in gaining weight. We are interested in finding this out because underweight people want to gain weight in a healthy manner, and overweight people want to avoid food that would make them gain weight. To solve this question, I would use two factor ANOVA.
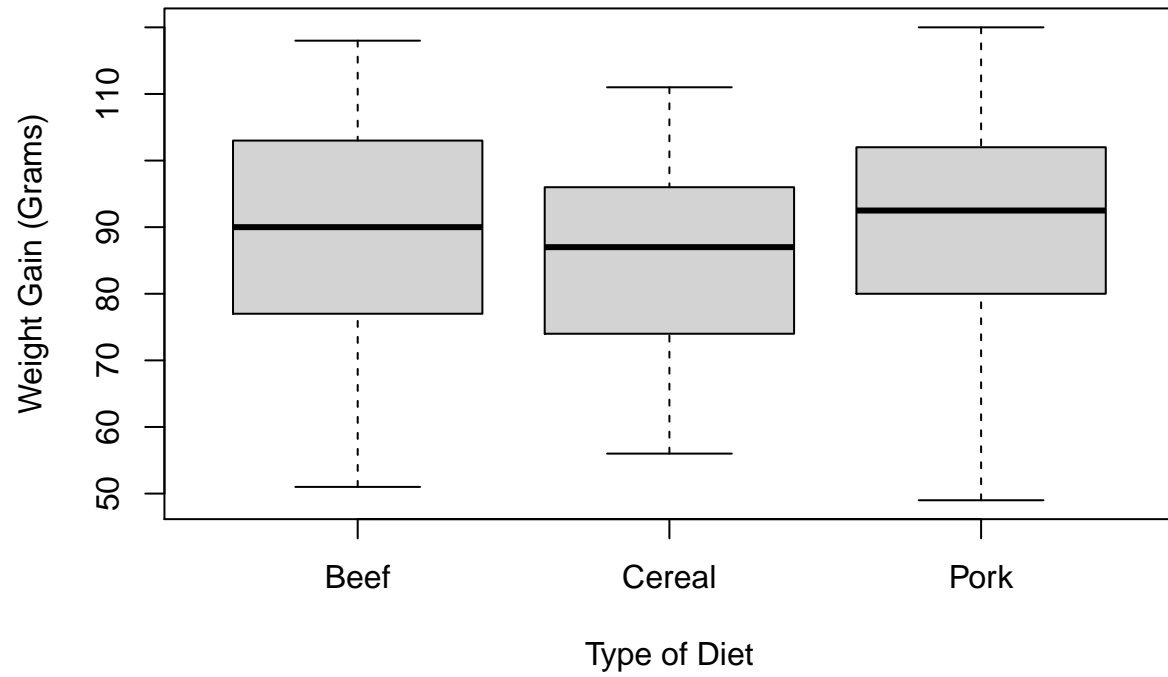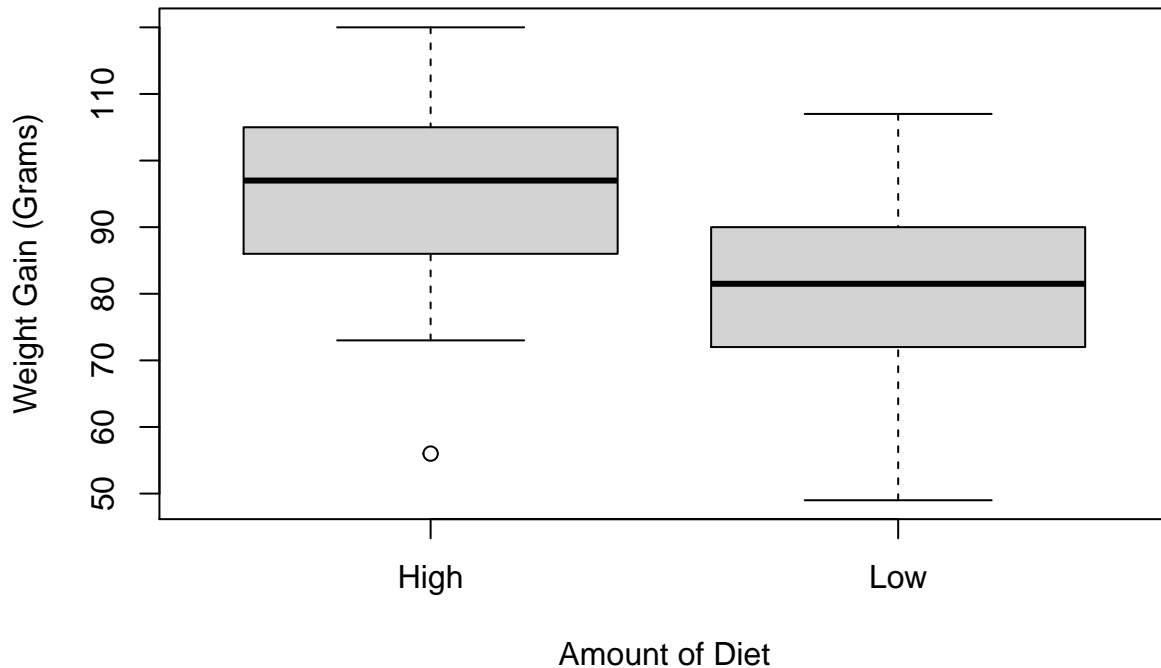
II. Summary
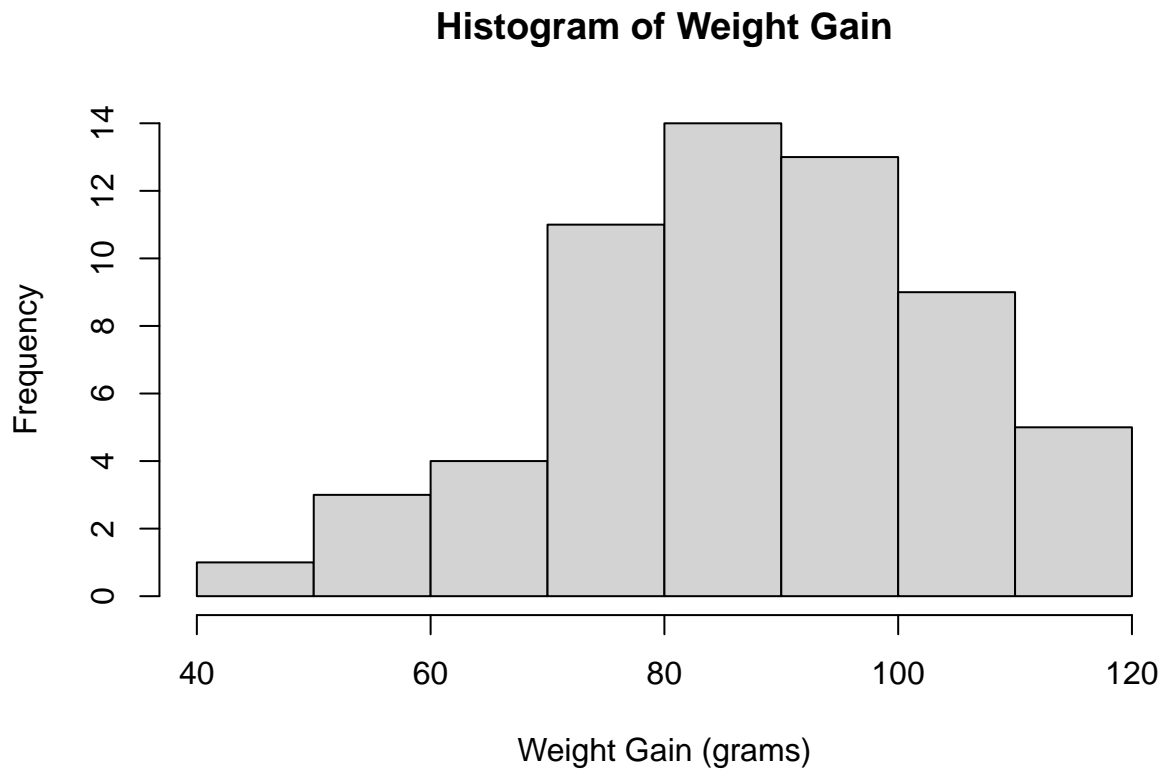
Histogram of Weight Gain

Boxplot of Weight Gain depending on diet
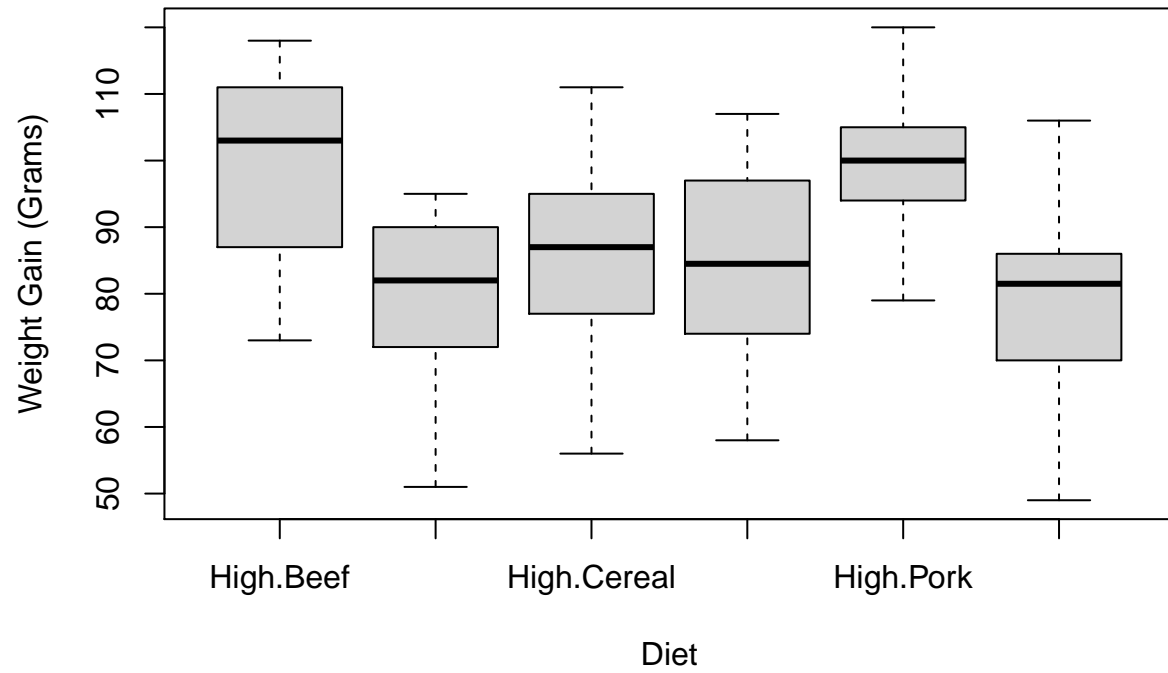
Boxplot of Weight Gain depending on diet

Boxplot of Weight Gain depending on diet

The histogram suggests the data is approximatly normal since there is no skewness. Also, the sample size is larger than 30 so by the Central Limit Theorem the distribution is normal. The most common weight gain is between 80 and 90 grams.

The boxplots suggests high beef is most effective in weight gain, and low beef and low pork is least effective in weight gain.

```
##    Amount   Type Weight.mean Weight.std Weight.samp_size
## 1   High   Beef   100.00000   15.13642         10.00000
## 2    Low   Beef    79.20000   13.88684         10.00000
## 3   High Cereal    85.90000   15.02184         10.00000
## 4    Low Cereal    83.90000   15.70881         10.00000
## 5   High   Pork    99.50000   10.91635         10.00000
## 6    Low   Pork    78.70000   16.54657         10.00000
```

The summary values are suggested above. High beef has the highest sample mean which is 100. Low pork has the lowest sample mean which is 78.7. Low pork has the highest standard deviation which is 16.54657. High pork has the lowest standard deviation which is 10.91635.

III. Diagnostics

Since we are going to perform a two factor ANOVA, the assumptions are as follows: 1. All subjects are randomly selected (independent) 2. Factor A levels are independent 3. Factor B levels are independent 4. $\epsilon_{ijk} \sim$ Normal $(0, \sigma_\epsilon^2)$

I believe the assumptions are met because the data is normal, random, and independent.

To view the data as a whole, there isn't any extreme values that we need to exclude.

IV. Analysis

Interaction Effect:

```
## [1] 60
```

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

```
## [1] 3
```

```
## [1] 16198.93
```

```
## [1] 13030.67
```

```
## [1] 15932.4
```

```
## [1] 12764.13
```

```
## [1] 11586
```

```
## [1] 0.0923003
```

Interpretation: the proportion in reduction in error when adding the interaction between the amount of diet and type of diet to the model with the main effect amount of diet and type of diet is 9.23%.

No, we don't need to include the interaction effect.

```
## [1] 2.74552
```

```
## [1] 0.07318788
```

```
## Analysis of Variance Table
##
## Response: Weight
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## Amount       1  3168.3  3168.3 14.7666 0.0003224 ***
## Type         2   266.5   133.3  0.6211 0.5411319
## Amount:Type  2  1178.1   589.1  2.7455 0.0731879 .
## Residuals   54 11586.0   214.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test statistic is 2.74552 and p-value is 0.07318788.

Amount effects and Type effects:

```
## [1] 0.1955849
```

```
## [1] 0.01645376
```

```
## Analysis of Variance Table
##
## Response: Weight
##           Df  Sum Sq Mean Sq F value     Pr(>F)
## Amount     1  3168.3  3168.3  14.102 0.0004033 ***
## Residuals 58 13030.7   224.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
## Response: Weight
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Type       2   266.5  133.27  0.4768 0.6232
## Residuals 57 15932.4  279.52
```

Test statistics is 14.102 and p-value is 0.0004033 for Amount. Test statistics is 0.4768 and p-value is 0.6232 for Type.

```
## [1] 0.02045431

## [1] 0.1988568
```

R2{A + B|A} is 0.0205, or 2.05 %, and R2{A + B|B} is 0.199, or 19.9 %.

V. Interpretation

Conclusion for interactions: Since p-value is $0.07318788 > \alpha = 0.05$, we fail to reject the null hypothesis. At the 5% significance level, we conclude the model with no interaction effects between amount of diet and type of diet is statistically equivalent to a model with interaction effects.

Conclusion for Amount effects: Since p-value is $0.0004033 < \alpha = 0.05$, we reject the null hypothesis. At the 5% significance level, we conclude the model with Amount of diet is statistically better fit than a model without Amount of diet.

Conclusion for Type effects: Since p-value is $0.6232 > \alpha = 0.05$, we fail to reject the null hypothesis. At the 5% significance level, we conclude the model with no Type effects is statistically equivalent to a model with Type effects.

The proportion in reduction of error when adding information regarding amount with information including type is 2.05 %.

The proportion in reduction of error when adding information regarding type with information including type is 19.9 %.

I would suggest using a model without interaction, Amount effect, and Type effect. We already concluded that the model with no Type effects is statistically equivalent to a model with Type effects. Moreover, the conditional R squared for both Amount effect and Type effect is lower than 20%, which means they have no to minimum effect to the model. So we can exclude all three.

$Y_i = \beta_0 + \beta_1 X_{a_2;i} + \beta_2 X_{b_2;i} + \beta_3 X_{b_3;i} + \epsilon_i$

$X_{a_2} = 1$ if $A = a_2$ $X_{a_2} = 0$ if otherwise

$X_{b_2} = 1$ if $B = b_2$ $X_{b_2} = 0$ if otherwise

$X_{b_3} = 1$ if $B = b_3$ $X_{b_3} = 0$ if otherwise

ANOVA Estimate

$E\{Y_i\} = \beta_0 = \bar{Y}_{11.}$ $E\{Y_i\} = \beta_0 + \beta_2 = \bar{Y}_{12.}$ $E\{Y_i\} = \beta_0 + \beta_3 = \bar{Y}_{13.}$ $E\{Y_i\} = \beta_0 + \beta_1 = \bar{Y}_{21.}$ $E\{Y_i\} = \beta_0 + \beta_1 + \beta_2 = \bar{Y}_{22.}$ $E\{Y_i\} = \beta_0 + \beta_1 + \beta_3 = \bar{Y}_{13.}$

VI. Conclusion

Our model doesn't include Amount effect and Type effect. Our model also doesn't have interaction effect. High Beef is most effective in gaining weight, and low pork is least effective in gaining weight.

```r
knitr::opts_chunk$set(echo = TRUE)
rat <- read.csv("rat.csv")
hist(rat$Weight, xlab = "Weight Gain (grams)", main = "Histogram of Weight Gain")
boxplot(Weight~Amount+Type, data = rat, xlab = "Diet", ylab = "Weight Gain (Grams)", main = "Boxplot of
boxplot(Weight~Type, data = rat, xlab = "Type of Diet", ylab = "Weight Gain (Grams)", main = "Boxplot o
boxplot(Weight~Amount, data = rat, xlab = "Amount of Diet", ylab = "Weight Gain (Grams)", main = "Boxpl
aggregate(Weight~Amount+Type, data = rat, FUN = function(x) {c(mean = mean(x), std = sd(x), samp_size =
Q <- quantile(rat$Weight, probs=c(.25, .75), na.rm = FALSE)
iqr <- IQR(rat$Weight)
up <-  Q[2]+1.5*iqr
low<- Q[1]-1.5*iqr
eliminated<- subset(rat, rat$Weight > (Q[1] - 1.5*iqr) & rat$Weight < (Q[2]+1.5*iqr))
n_T <- length(rat$Weight)
n_T
a <- length(unique(rat$Amount))
a
b <- length(unique(rat$Type))
b

model_ept <- lm(Weight~1, data = rat)
model_A <- lm(Weight~Amount, data = rat)
model_B <- lm(Weight~Type, data = rat)
model_APB <- lm(Weight~Amount + Type, data = rat)
model_AB <- lm(Weight~Amount * Type, data = rat)

sse_ept <- tail(anova(model_ept)$'Sum Sq', n = 1)
sse_A <- tail(anova(model_A)$'Sum Sq', n = 1)
sse_B <- tail(anova(model_B)$'Sum Sq', n = 1)
sse_APB <- tail(anova(model_APB)$'Sum Sq', n = 1)
sse_AB <- tail(anova(model_AB)$'Sum Sq', n = 1)

sse_ept
sse_A
sse_B
sse_APB
sse_AB
R_ab_apb <- (sse_APB-sse_AB)/sse_APB
R_ab_apb
test_stats <- ((sse_APB - sse_AB)/(n_T - a - b + 1 - n_T + a*b))/(sse_AB/(n_T-a*b))
test_stats
p_value <- 1-pf(test_stats, (n_T - a - b + 1 - n_T + a*b), (n_T - a*b))
p_value

anova(model_AB)
R_A_EPT <- (sse_ept-sse_A)/sse_ept
R_A_EPT

R_B_EPT <- (sse_ept-sse_B)/sse_ept
R_B_EPT
```

```
anova(model_A)
anova(model_B)
R_A_APB <- (sse_A-sse_APB)/sse_A
R_A_APB

R_B_APB <- (sse_B-sse_APB)/sse_B
R_B_APB
```