

HW2-ADA

Code ▼

Jing Qian (jq2282)

1.a. Compute the analysis of variance table for the data.

Hide

```
anova(lm(data[,2]~factor(data[,1])))
```

Analysis of Variance Table

Response: data[, 2]

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(data[, 1])	3	0.52652	0.175506	4.4891	0.01185 *
Residuals	25	0.97740	0.039096		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

From the ANOVA table, we could find that p-value = 0.01185, which is smaller than $\gamma = 0.05$. So we could reject H_0 , which means at least two diet means are significantly different.

1.b. Compute a 95% confidence interval for $\mu_1 - \mu_2$.

Hide

```
t.test(data[,2][data[,1]==1], data[,2][data[,1]==2])
```

Welch Two Sample t-test

data: data[, 2][data[, 1] == 1] and data[, 2][data[, 1] == 2]

t = 1.1625, df = 9.124, p-value = 0.2745

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.1325412 0.4139698

sample estimates:

mean of x mean of y

3.745714 3.605000

The 95% confidence interval for $\mu_1 - \mu_2$ could be estimated by running t-test of sample 1 and sample 2. From above, we could find the 95% confidence interval for $\mu_1 - \mu_2$ is [-0.1325412, 0.4139698].

1.c. Compute a 95% confidence interval for $L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$.

Hide

```
Lhat = (mean(data[,2][data[,1]==1])+mean(data[,2][data[,1]==2])-mean(data[,2][data[,1]==3])-mean(data[,2][data[,1]==4]))*0.5
n1 = sum(data[,1]==1)
n2 = sum(data[,1]==2)
n3 = sum(data[,1]==3)
n4 = sum(data[,1]==4)
n = n1+n2+n3+n4
Lse = 0.5*sqrt(0.039096*(1/n1+1/n2+1/n3+1/n4))
tL = qt(1-0.05/2, n-4)
c(Lhat - tL*Lse, Lhat + tL*Lse)
```

```
[1] -0.2373650  0.0672459
```

For $L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$, $c_1 = c_2 = 0.5$, $c_3 = c_4 = -0.5$. Then $\hat{L} = \sum_{i=1}^4 c_i \bar{Y}_i = -0.08505952$,

$SE(\hat{L}) = \sqrt{\sigma^2 \sum_{i=1}^4 \frac{c_i^2}{n_i}} = 0.5 \sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_4} \right)} = 0.3982396$. So a 95% confidence interval for L is $\hat{L} \pm t_{\frac{\gamma}{2}}(d) SE(\hat{L}) = [-0.237365, 0.0672459]$

1.d. Test $H_0 : L = 0$ against $H_a : L \neq 0$ using $\gamma = 0.05$.

Hide

```
abs(Lhat) > tL*Lse
```

```
[1] FALSE
```

From above, we could see that $|SE(\hat{L})| < t_{\frac{\gamma}{2}}(d) SE(\hat{L})$, so we could not reject $H_0 : L = 0$.

2.a. Obtain estimates of μ , α_1 , α_2 , α_3 and α_4 .

We could get the estimation by doing regression using

$$D_i = \begin{cases} 1, & \text{if diet is } i \\ -1, & \text{if diet is 4} \\ 0, & \text{otherwise} \end{cases}$$

where $i = 1, 2, 3$ with $Y_{ij} = \mu^* + \alpha_i D_i + \epsilon_{ij}$.

Hide

```
a = factor(data[,1])
model2 = lm(data[,2]~a, contrasts = list(a="contr.sum"))
summary(model2)
```

```

Call:
lm(formula = data[, 2] ~ a, contrasts = list(a = "contr.sum"))

Residuals:
    Min       1Q   Median       3Q      Max
-0.3857 -0.0950 -0.0525  0.1250  0.4443

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.71789    0.03698  100.549  <2e-16 ***
a1             0.02783    0.06450   0.431   0.6698
a2            -0.11289    0.06173  -1.829   0.0794 .
a3            -0.11955    0.06801  -1.758   0.0910 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1977 on 25 degrees of freedom
Multiple R-squared:  0.3501,    Adjusted R-squared:  0.2721
F-statistic: 4.489 on 3 and 25 DF,  p-value: 0.01185

```

So $\hat{\mu} = 3.71789$, $\hat{\alpha}_1 = 0.02783$, $\hat{\alpha}_2 = -0.11289$, $\hat{\alpha}_3 = -0.11955$, $\hat{\alpha}_4 = -\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\alpha}_3 = 0.20461$.

2.b. Test $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ against H_a : Not H_0 at $\gamma = 0.05$. Compare your answer with 1.a.

Following we show the ANOVA table for regression model in 2.a. We could see that the p-value is 0.01185, which is less than 0.05. So $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ is rejected.

[Hide](#)

```
anova(model2)
```

Analysis of Variance Table

```

Response: data[, 2]
      Df Sum Sq Mean Sq F value Pr(>F)
a       3  0.52652  0.175506   4.4891 0.01185 *
Residuals 25  0.97740  0.039096
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Comparing the ANOVA table with the one in Problem 1, we could see that the two tables are the same, which shows that the two null hypotheses are equivalent: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ and $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$.

3.a. Complete the table.

Source	df	SS	MS	F
feed	2	23.43	11.72	17.50

Source	df	SS	MS	F
error	7	4.67	0.67	
total	9	28.10		

3.b. Test the null hypothesis is that all the chicken feeds have the same effect on the length of the major axis against the alternative is that the feed has some causal effect. Use $\alpha = 0.05$.

$F_{\gamma}(k-1, n-k) = F_{0.05}(2, 7) = 4.74$. From the table above, $F > F_{\gamma}(k-1, n-k)$. So H_0 hypothesis could be rejected. The chicken feeds do have some causal effect on the length of the major axis.

[Hide](#)

```
qf(.95, df1=2, df2=7)
```

```
[1] 4.737414
```