State 461 HW5 Jiaqi Li

1. (a)

$$Y_{it} = \mu + \tau_i + \varepsilon_{it}, \quad i = A, B, C, \quad t = 1, 2.$$

note : C stands for "control".
A stands for "drug A" & B stands for "drug B".

(c) For $\{Y_{it}\}$, $i = A, B, C$, $t = 1, 2$

$\bar{Y}_{A\cdot} = \frac{-14-4}{2} = -9$

$\bar{Y}_{B\cdot} = \frac{5-1}{2} = 2$

$\bar{Y}_{C\cdot} = \frac{-2+6}{2} = 2.$

$\bar{Y}_{\cdot\cdot} = \frac{-14-4+5-1-2+6}{6} = -\frac{5}{3}$

$$SST = 2\times(-9+\tfrac{5}{3})^2 + 2\times(2+\tfrac{5}{3})^2 + 2\times(2+\tfrac{5}{3})^2 = \frac{484}{3}$$

$$SSTOT = (-14+\tfrac{5}{3})^2 + (-4+\tfrac{5}{3})^2 + (5+\tfrac{5}{3})^2 + (-1+\tfrac{5}{3})^2 + (-2+\tfrac{5}{3})^2 + (6+\tfrac{5}{3})^2 = \frac{784}{3}$$

$$SSE = SSTOT - SSE = 100$$

(b) $E(\Delta BP \text{ of } B) = \mu + \tau_B = \bar{Y}_{B\cdot} = 2$

$\Delta BP$ of B $\sim N(2, \sigma^2/2)$ since $\bar{Y}_{B\cdot} \sim N(2, \sigma^2/2)$

(d) $\hat{\sigma}^2 = \frac{SSE}{n-v}$ where $n=6$, $v=3$

$= \frac{100}{6-3} = \frac{100}{3}.$

(e) $E(\Delta BP \text{ of } A - \Delta BP \text{ of } C) = (\mu+\tau_A) - (\mu+\tau_C) = \bar{Y}_{A\cdot} - \bar{Y}_{C\cdot} = -11$

$\Delta BP$ of A $- \Delta BP$ of C $\sim N(-11, \sigma^2)$

since $\bar{Y}_{A\cdot} - \bar{Y}_{C\cdot} \sim N(-9-2, \frac{\sigma^2}{2} + \frac{\sigma^2}{2}) = N(-11, \sigma^2)$

(f) When $\tau_A = \tau_C$, $(\mu+\tau_A) - (\mu+\tau_C) = 0$.
Then,
$$\hat{Y}_{A\cdot} - \hat{Y}_{C\cdot} \sim N(0, \sigma^2).$$

(g) Let $K(\sigma) = \sigma$
Then,
$$\frac{\hat{Y}_{A\cdot} - \hat{Y}_{C\cdot}}{K(\sigma)} \sim N(0/\sigma^2, \sigma^2/\sigma^2) = N(0,1).$$

(h)
$$\frac{SSE}{\sigma^2} = \frac{\sum_i \sum_t (Y_{it} - \bar{Y}_{i\cdot})^2}{\sigma^2}$$
$$= \left[\frac{\sum_i \sum_t (Y_{it} - \bar{Y}_{i\cdot})}{\sigma}\right]^2 = \sum_i \sum_t \left(\frac{Y_{it} - \bar{Y}_{i\cdot}}{\sigma}\right)^2$$

Since $\frac{Y_{it} - (\mu+\tau_i)}{\sigma} \sim N(0,1)$ where $Y_{it} \overset{iid}{\sim} N(\mu+\tau_i, \sigma^2)$,
$$\left(\frac{Y_{it} - (\mu+\tau_i)}{\sigma}\right)^2 \sim \chi^2_1$$

Then, $\sum_i \sum_t \left(\frac{Y_{it} - (\mu+\tau_i)}{\sigma}\right)^2 \sim \chi^2_6$
Now, replace $\mu+\tau_i$ with $\bar{Y}_{i\cdot}$, we have
$$\frac{SSE}{\sigma^2} = \sum_i \sum_t \left(\frac{Y_{it} - \bar{Y}_{i\cdot}}{\sigma}\right)^2 \sim \chi^2_{6-3} = \chi^2_3.$$

(i) Since $\frac{SSE}{\sigma^2}$ is chi-square distribution with degree of freedom 3 and $\frac{\hat{Y}_{A\cdot} - \hat{Y}_{C\cdot}}{K(\sigma)}$ is standard normal distribution,
$$\frac{(\hat{Y}_{A\cdot} - \hat{Y}_{C\cdot})/K(\sigma)}{\sqrt{SSE/[(n-v)\sigma^2]}} \sim t_3$$

(j) ANOVA (one-way):

| Source | Deg. of Freedom | Sum of Square | Mean Square | Ratio |
|---|---|---|---|---|
| Treatment | $3-1=2$ | $484/3$ | $(484/3)/2 = 242/3$ | $(242/3)/(100/3) = 2.42$ |
| Error | $6-3=3$ | $100$ | $100/3$ | |
| Total | $6-1=5$ | $784/3$ | | |

$H_0$: there are no differences in mean blood pressure change between women getting any treatment.

Under null hypothesis, test statistic $T^* = 2.42$
Test statistic is F distribution (ie. $F_{2,3}$)

(k)

```
> library(knitr)
> model1=aov(effect~drug,data=data)
> kable(anova(model1),format="markdown")


|           | Df|    Sum Sq|   Mean Sq| F value|     Pr(>F)|
|:----------|--:|---------:|---------:|-------:|----------:|
|drug       |  2|  161.3333|  80.66667|    2.42|  0.2367052|
|Residuals  |  3|  100.0000|  33.33333|      NA|         NA|
```

**$H_0$: Drugs do not significantly affect the mean blood pressure of women who take those drugs.**

**$H_1$: Drugs have significant impacts on the mean blood pressure of women who take those drugs.**

By observing the table, we have:

Test Statistic is F distribution with $F_{2,3} = 2.42$

p-value = 0.2367052 > 0.05 = α (which is the significant level)

Thus, we accept null hypothesis because the test statistic greater than 2.42 is more than 5% of the time.

Thus, drugs do not significantly affect the mean blood pressure of women who take those drugs.

(i)

```
> lsm.drug=lsmeans(model1, ~drug)
> kable(summary(contrast(lsm.drug,method="pairwise",adjust="tukey"),infer=c(T,T),level=0.5,side="
two-sided"))


|contrast         | estimate|       SE| df|   lower.CL|  upper.CL|   t.ratio|   p.value|
|:----------------|--------:|--------:|--:|----------:|---------:|---------:|---------:|
|Control - Drug A |       11| 5.773503|  3|   3.688629| 18.311371|  1.905256| 0.2816743|
|Control - Drug B |        0| 5.773503|  3|  -7.311371|  7.311371|  0.000000| 1.0000000|
|Drug A - Drug B  |      -11| 5.773503|  3| -18.311371| -3.688629| -1.905256| 0.2816743|
```

**$H_0$: $\tau_C - \tau_A = 0$**

**$H_1$: $\tau_C - \tau_A \neq 0$**

Test Statistic $T^*$ is t distribution with degree of freedom = 3 and t = 1.905256.

p-value = 0.2816743 > 0.05 = α

Since p-value is larger than α, which is the significant level, we accept null hypothesis.

Thus, the mean blood pressure of women getting the Control group is not significantly different from the mean blood pressure of women getting Drug A.

## 2.

```
> model2=aov(weightloss~type,data=experiment)
> kable(anova(model2),format="markdown")


|          | Df|    Sum Sq|  Mean Sq|  F value| Pr(>F)|
|:---------|--:|---------:|--------:|--------:|------:|
|type      |  2| 16.122050| 8.061025| 104.4512|  6e-07|
|Residuals |  9|  0.694575| 0.077175|       NA|     NA|
```

**$H_0$: Soap type does not have significant effects on the mean weight loss of a soap over 24 hours.**

**$H_1$: Soap type does have significant effects on the mean weight loss of a soap over 24 hours.**

Test Statistic $T^*$ is F distribution with $F_{2,9} = 104.4512$.

p-value = $6 \times 10^{-7} < 0.05 = \alpha$

Since p-value is smaller than α, which is the significant level, we reject null hypothesis.

Thus, soap type does have significant effects on the mean weight loss of a soap over 24 hours.

## 3.

By using one-way ANOVA model:

```
> model3=aov(time~push,data=light)
> kable(anova(model3),format="markdown")


|          | Df|    Sum Sq|   Mean Sq|   F value|    Pr(>F)|
|:---------|--:|---------:|---------:|---------:|---------:|
|push      |  3| 0.0080471| 0.0026824| 0.2454844| 0.8638292|
|Residuals | 28| 0.3059529| 0.0109269|        NA|        NA|
```

Assumption:

All variable residuals are normally distributed (ie. $\epsilon \sim N(0,\sigma)$).

All σ for each variable are the same.

All $Y_{it}$ are iid normal random variables.

**$H_0$: There are no differences in mean waiting time for different number of pushing the button.**

**$H_1$: Pushing the button has significant impacts on waiting time.**

Test Statistic $T^*$ is F distribution with $F_{3,28} = 0.2454844$.

p-value = $0.8638292 > 0.05 = \alpha$

Since p-value is larger than α, which is the significant level, we accept null hypothesis.

Thus, the number of push does not significantly affect the mean waiting time for light change.

```
> lsm.light=lsmeans(model3, ~push)
> kable(summary(contrast(lsm.light,method="pairwise",adjust="tukey"),infer=c(T,T),leve
l=0.5,side="two-sided"))


|contrast |   estimate|        SE| df|   lower.CL|  upper.CL|    t.ratio|   p.value|
|:--------|----------:|---------:|--:|----------:|---------:|----------:|---------:|
|0 - 1    |  0.0361429| 0.0515138| 28| -0.0368215| 0.1091072|  0.7016149| 0.8955744|
|0 - 2    |  0.0131429| 0.0515138| 28| -0.0598215| 0.0861072|  0.2551327| 0.9940396|
|0 - 3    | -0.0048571| 0.0612075| 28| -0.0915517| 0.0818375| -0.0793553| 0.9998162|
|1 - 2    | -0.0230000| 0.0467480| 28| -0.0892141| 0.0432141| -0.4919994| 0.9602303|
|1 - 3    | -0.0410000| 0.0572544| 28| -0.1220954| 0.0400954| -0.7161022| 0.8898585|
|2 - 3    | -0.0180000| 0.0572544| 28| -0.0990954| 0.0630954| -0.3143863| 0.9890012|
```

**$H_0$: $\tau_0 = \tau_1 = \tau_2 = \tau_3 = 0$**

**$H_1$: At least one τ is different than others.**

Test Statistic $T^*$ is t distribution with degree of freedom = 28 and t = 0.7016149, 0.2551327, -0.0793553, -0.4919994, -0.7161022, -0.3143863 for contrast pair “0-1”, “0-2”, “0-3”, “1-2”, “1-3”, “2-3”.

By observing the chart, we have that p-value of each contrast is greater than 5%, which is the significant level. Thus, we accept the null hypothesis.

Thus, each mean waiting time for pushing the button 0, 1, 2, 3 times is not significantly different from each other.

## R code:

```
#1-----------------------------------------------------------------------------------------------------------

drug=c(rep("Drug A",2),rep("Drug B",2),rep("Control",2))

effect=c(-14,-4,5,-1,-2,6)

data=data.frame(drug=as.factor(drug),effect=effect)

data


install.packages("knitr")

library(knitr)

model1=aov(effect~drug,data=data)

kable(anova(model1),format="markdown")


install.packages("lsmeans")

library(lsmeans)

lsm.drug=lsmeans(model1, ~drug)

kable(summary(contrast(lsm.drug,method="pairwise",adjust="tukey"),infer=c(T,T),level=0.5,side="two-sided"))


#2-----------------------------------------------------------------------------------------------------------

type=c(rep("Regular",4),rep("Deordorant",4),rep("Moisturizing",4))

weightloss=c(-0.30,-0.10,-0.14,0.40,2.63,2.61,2.41,3.15,1.86,2.03,2.26,1.82)

experiment=data.frame(type,weightloss)

experiment


model2=aov(weightloss~type,data=experiment)

kable(anova(model2),format="markdown")


#3-----------------------------------------------------------------------------------------------------------

push=c(rep("0",7),rep("1",10),rep("2",10),rep("3",5))

time=c(38.14,38.20,38.31,38.14,38.29,38.17,38.20,38.28,38.17,38.08,38.25,38.18,38.03,37.95,38.26,38.30,38.21,3
8.17,38.13,38.16,38.30,38.34,38.34,38.17,38.18,38.09,38.06,38.14,38.30,38.21,38.04,38.37)

light=data.frame(push,time)
```

```
light


model3=aov(time~push,data=light)

kable(anova(model3),format="markdown")


lsm.light=lsmeans(model3, ~push)

kable(summary(contrast(lsm.light,method="pairwise",adjust="tukey"),infer=c(T,T),level=0.5,side="two-sided"))
```