STP 531

Applied Analysis of Variance

Homework 2

Nathan A. Nguyen

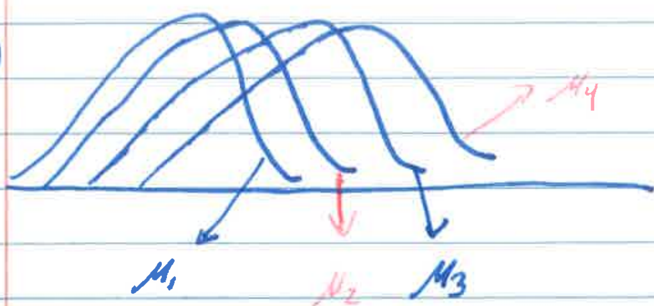
10 February 2021

#16.5

$\mu_1 = 5.1$, $\mu_2 = 6.3$, $\mu_3 = 7.9$, $\mu_4 = 9.5$, $\sigma = 2.8$

Assume ANOVA 16.2: $Y_{ij} = \mu_i + \varepsilon_{ij}$

a.)



b.) $n_i = n = 100$

$E\{MSE\} = \sigma^2 = (2.8)^2 = 7.84$

$E\{MSTR\} = \sigma^2 + \dfrac{\sum n_i(\mu_i - \mu_.)^2}{r-1}$

$\mu_. = \dfrac{\sum n_i \mu_i}{r} = \dfrac{n \sum \mu_i}{nr}$

$\Rightarrow E\{MSTR\} = \sigma^2 + \dfrac{n\sum(\mu_i - \mu_.)^2}{r-1}$

$\mu_. = \dfrac{5.1 + 6.3 + 7.9 + 9.5}{4} = 7.2$

$\Rightarrow E\{MSTR\} = 7.84 + 100\left(\dfrac{(5.1-7.2)^2 + (6.3-7.2)^2 + (7.9-7.2)^2 + (9.5-7.2)^2}{3-1}\right)$

$\approx 374.$

$E\{MSTR\} \approx 374 \gg 7.84 = E\{MSE\}$

$\Rightarrow$ factor level means (μ_i) are not equal.

$\mu_2 = 6.3$
$\mu_3 = 7.9.$

c.) Suppose then that $\mu_1 = 5.1$, $\mu_{2*} = 5.6$, $\mu_{3*} = 9.0$, $\mu_4 = 9.5$, $\sigma = 2.8$.

$$E\{MSTR\} = \sigma^2 + \frac{n\sum(\mu_i - \mu_.)^2}{r-1}$$

$\sigma^2 = 7.84$, $\mu_. = \frac{n\sum \mu_i}{nr}$, $\mu_. = \frac{5.1 + 5.6 + 9.0 + 9.5}{4} = \cancel{7.25}\ 7.3$

$$E\{MSTR\} = 7.84 + 100\left(\frac{(5.1-7.3)^2 + (5.6-7.3)^2 + (9.0-7.3)^2 + (9.5-7.3)^2}{4-1}\right)$$

$\approx 523.$

The $E\{MSTR\}$ is substantially larger because groups 2 and 3 are shifted further from the weighted mean $\mu_.$.

#16.6 $E\{MSE\} = \sigma^2$

$$E\{MSTR\} = \sigma^2 + \frac{\sum n_i(\mu_i - \mu_.)^2}{r-1}$$

$$\mu_. = \frac{\sum n_i \mu_i}{n_T}$$

if all μ_i are equal, then $\mu_i = \mu_i$ and hence

$$E\{MSTR\} = \sigma^2 + \frac{\sum n_i(\mu_i - \mu_.)^2}{r-1}$$

$$= \sigma^2 + 0 = \sigma^2,$$

Then $F^* = MSTR/MSE \doteq 1$.

$E\{MSTR\} > E\{MSE\}$ if there are differences in factor level means, and since we have a $(\text{quantity})^2$ term, it follows that $E\{MSTR\} > E\{MSE\} > 0$. This is why F^* is a one-tailed test.

ij

\# 16.8 b-f.　　　　　　　　　　　　　　　　　　　　\# 3

$r = 3$ (blue, green, orange).　　　　ANOVA 16.2: $Y_{ij} = \mu_i + \varepsilon_{ij}$

b.) fitted values. $Y_{ij} = \bar{Y}_{i.}$

~~$\hat{Y}_{ij} = Y_j$~~　　$\hat{Y}_{1j} = \bar{Y}_{1.} = \dfrac{28+26+31+27+35}{5} = 29.4$

$\hat{Y}_{2j} = \bar{Y}_{2.} = \dfrac{34+29+25+31+29}{5} = 29.6$

$\hat{Y}_{3j} = \bar{Y}_{3.} = \dfrac{31+25+27+29+28}{5}$ 28

c.) Residuals. $\underbrace{(28-29.4)}_{-1.4}$, $\underbrace{(26-29.4)}_{-3.4}$, $\underbrace{(31-29.4)}_{1.6}$, $\underbrace{(27-29.4)}_{-2.4}$

e_{ij}

i \ j	1	2	3	4	5
1	-1.4	-3.4	1.6	-2.4	5.6
2	4.4	-0.6	-4.6	1.4	-0.6
3	3	-3	-1	1	0

repeat that for each ith factor... Not going to write it all out.

$n_i = n = 5$

d.) ANOVA Table.

Source	SS	df	MS.
Between treatments: (columns)	7.6	~~3~~ 2	$7.6/2 = 3.8$
Error:	~~116~~ 116.4	~~15~~ 12	$116.4/(15-3) = 9.7$
total:		14	

$SSTR = \sum n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$　　$5\left((29.4-29)^2 + (29.6-29)^2 + (28-29)^2\right) = 7.6$

$\bar{Y}_{..} = \sum_i^r \sum_j^{n_i} Y_{ij} =$

$\bar{Y}_{..} = \dfrac{435}{15} = 29.$　　$n_i = 5.$

$SSE = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2 = \sum_i \sum_j e_{ij}^2 =$

53.2 i_1
+ 43.2 i_2
+ 20 i_3
116.4

#1 e.) $\alpha = 0.10$

$H_0: \mu_1 = \mu_2 = \mu_3$

H_a: Not all μ_i are equal for $i = 1, 2, 3$.

test stat: $F^* = \dfrac{MSTR}{MSE} = \dfrac{3.8}{9.7} \approx 0.392$

$1 - \alpha = 1 - 0.10 = 0.9$.

$F(0.90; 2; 12) = 2.81$

If $F^* \leq F_{crit}$, Conclude H_0.

if $F^* > F_{crit}$, Conclude H_a.

$F^* \approx 0.392 < 2.81$.

~~We~~ There is not enough evidence to reject the null hypothesis, Conclude H_0.

$P\{F(2, 12) > F^* = 0.39\}$. 2.806.

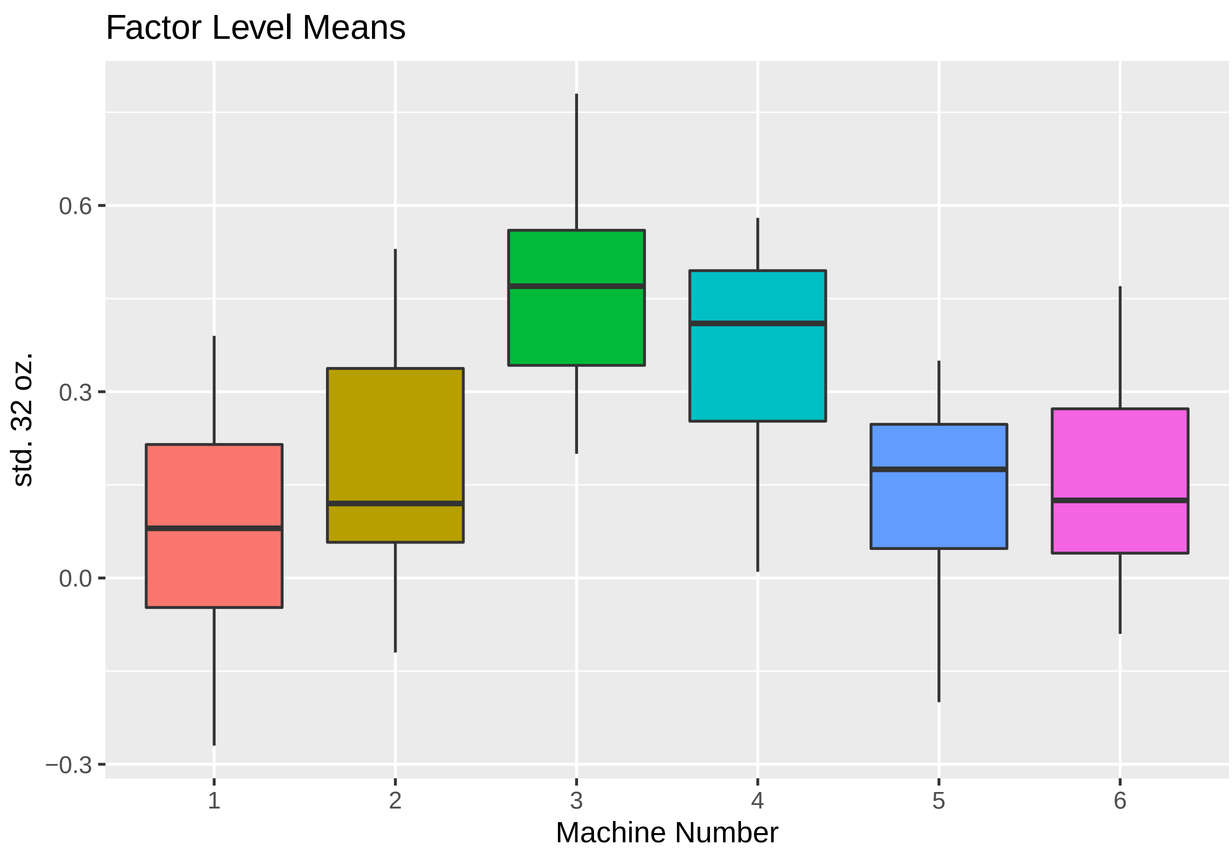
$P = 0.68$

f.) The analysis found that the factor level means were equal ($\mu_1 = \mu_2 = \mu_3$). In other words blue, green, or orange colour paper made no difference in the mean response rate. However, I recommend that a Control group (white paper) be included in a follow-up study so that we can make direct/rigour analysis before jumping to fast conclusions.

Question 16.11

$$
\begin{aligned}
Y_{ij} &= \mu_{ij} + \epsilon_{ij} \\
n_T &= 120 \\
n_i &= 20 \\
i &= 1, 2, \ldots, 6 \text{ and } j = 1, 2, \ldots, 20
\end{aligned}
$$

Part A



Part B

Table 1: Fitted Values

group	mean
1	0.0735
2	0.1905
3	0.4600
4	0.3655
5	0.1250
6	0.1515

Part C

Table 2: Deviations

Group_1	Group_2	Group_3	Group_4	Group_5	Group_6
-0.2135	0.2695	-0.25	0.1245	-0.315	-0.1015
0.1265	-0.0805	0.32	0.2145	0.145	-0.2015
-0.0035	-0.0705	-0.14	0.1545	-0.065	0.1285
0.1065	0.2795	-0.01	-0.0755	-0.015	0.3185
0.3065	0.0495	-0.24	-0.0955	0.105	-0.0315
0.0265	-0.1305	-0.11	0.1845	0.025	0.1185
-0.1135	-0.3105	0.08	0.0345	-0.115	-0.0715
-0.3435	0.1395	-0.22	-0.2255	0.095	0.0185
0.1965	-0.1305	0.01	0.1145	0.165	0.2785
-0.2835	-0.2205	0.16	-0.0255	0.015	-0.2215
0.3165	-0.1405	0.01	-0.3555	0.075	0.0485
-0.1435	0.3395	0.09	-0.0355	0.175	-0.1415
-0.0935	0.2295	0.13	-0.1855	-0.235	-0.0515
0.2065	0.0995	0.25	-0.2355	0.145	0.0085
0.0165	0.1695	-0.01	0.1145	-0.325	-0.2115
0.0565	-0.1505	0.02	0.1745	0.115	-0.0215
0.1865	-0.0205	-0.02	0.1445	0.075	0.2785
-0.0035	-0.1705	0.04	0.0545	0.015	0.1985
-0.0835	-0.0805	-0.26	0.0845	0.225	-0.2415
-0.2635	-0.0705	0.15	-0.1655	-0.305	-0.1015

The residuals sum to zero $-2.3592239 \times 10^{-16}$, which is essentially 0.

Part D

Table 3: ANOVA table

Source	SS	df	MS
Between Machines	2.29	5	0.46
Error	3.53	114	0.03
Total	5.82	119	

Part E

Let $\alpha = 0.05$

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$
$$H_a : \text{not all } \mu_i \text{ are equal}$$
$$F^* = \frac{MSTR}{MSE}$$

- if $F^* \leq F(1-\alpha; r-1; n_T - r)$, conclude H_0.
- if $F^* > F(1-\alpha; r-1; n_T - r)$, conclude H_a.

$$F^* = \frac{MSTR}{MSE} = \frac{0.46}{0.03} \approx 15.33$$
$$F(1-\alpha; r-1; n_T - r) = F(0.95, 5, 114) \approx 2.29$$

Since $F^* > F_{critical}$ there is sufficient evidence to suggest that the factor level means are all not equal. In other words, there exists a difference between the treatments.

Part F

The p-value is $3.6363746 \times 10^{-11}$, which is is consistent with part E – that there is sufficient evidence to reject the null hypothesis and proceed with proceed with the notion that not all of the factor level means are equal.

Part G

The variation variation between groups is indeed large compared the variation within groups. Based on the plots alone, we can see that there is nothing really standing out with regards to the width of each of the groups' box. In other words, the interval of each box are approximately the same.

As for the variation between groups. That is obviously seen. if we isolate group 2 and 3, their 1st-3rd quartiles don't even fall within the other 4 groups' distributions. Also, $E\{MSE\} < E\{MSTR\}$ where $E\{MSTR\}$ is larger than an order of magnitude – more evidence to suggest that the variation between groups is more important than variation within groups.

#16.19 Model: $Y_{ij} = \mu + \underbrace{\tau_1 X_{ij1} + \tau_2 X_{ij2} + \cdots + \tau_{r-1} X_{ij,r-1}}_{} + \varepsilon_{ij}$

$\mu_{\cdot} = \dfrac{\sum_i^r \mu_i}{r}$ factor effects unweighted means.

$\Rightarrow \sum \mu_i = r\mu_{\cdot}$ r = # factor levels.

$\tau_i = \mu_i - \mu_{\cdot} \Rightarrow \mu_i = \mu_{\cdot} + \tau_i$. $r = \cancel{4}\ 3$

$n_i = 5$

$\underset{\sim}{Y}, X, \underset{\sim}{\beta}$ $\tau_3 = -\tau_1 - \tau_2$

$$\underset{15\times 1}{\underset{\sim}{Y}} = \begin{pmatrix} 28 \\ 26 \\ 31 \\ 27 \\ 35 \\ 34 \\ 29 \\ 25 \\ 31 \\ 29 \\ 31 \\ 25 \\ 27 \\ 29 \\ 28 \end{pmatrix} \begin{matrix} \left.\right\} r_1 \\ \left.\right\} r_2 \\ \left.\right\} r_3 \end{matrix}, \quad \underset{\cancel{15\times(3+1)}\ 15\times 3}{X} = \begin{array}{c} \\ \left(\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{array}\right) \\ \tau_3 = -\tau_1 - \tau_2 \end{array} \begin{matrix} \tau_1 \\ \\ \tau_2 \end{matrix}, \quad \underset{3\times 1}{\beta} = \begin{pmatrix} \mu_{\cdot} \\ \tau_1 \\ \tau_2 \end{pmatrix}$$

b.) Compute $X\beta$. $\underset{(15\times 3)}{X}\ \underset{(3\times 1)}{\beta} = \underset{15\times 1}{[\ \ ]}$

$$\underset{15\times 3}{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix}} \underset{3\times 1}{\begin{pmatrix} \mu. \\ \tilde{\tau}_1 \\ \tilde{\tau}_2 \end{pmatrix}} = \underset{(15\times 1)}{\begin{pmatrix} \mu. + \tilde{\tau}_1 \\ \mu. + \tilde{\tau}_1 \\ \mu. + \tilde{\tau}_1 \\ \mu. + \tilde{\tau}_1 \\ \mu. + \tilde{\tau}_2 \\ \mu. + \tilde{\tau}_2 \\ \mu. + \tilde{\tau}_2 \\ \mu. + \tilde{\tau}_2 \\ \mu. + \tilde{\tau}_2 \\ \mu. - \tilde{\tau}_1 - \tilde{\tau}_2 \\ \mu. - \tilde{\tau}_1 - \tilde{\tau}_2 \\ \mu. - \tilde{\tau}_1 - \tilde{\tau}_2 \\ \mu. - \tilde{\tau}_1 - \tilde{\tau}_2 \\ \mu. - \tilde{\tau}_1 - \tilde{\tau}_2 \end{pmatrix}} = \underset{(15\times 1)}{\begin{pmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{pmatrix}}$$

μ_1, μ_2, μ_3

Recall: $\mu_i \equiv \mu. + \tilde{\tau}_i$

θ ?

#16.15 $(\tau_1, \tau_2, \tau_3 ?)$

ANOVA factor effects (16.62)

and $\mu_.$ (16.63)

$\cancel{Y_{ij} = 1} \; Y_{ij} = \mu_. + \tau_i + \varepsilon_{ij}$

$\tau_i = \mu_i - \mu_.$

$\mu_. = \dfrac{\sum_i^r \mu_i}{r}$ } (16.63) unweighted Mean.

$r = 3, \mu_1 = 65, \mu_2 = 80, \mu_3 = 95, \sigma = 3.$

$\cancel{\mu} \; \mu_. = \dfrac{\sum_{i=1}^{3} \mu_i}{3} = \dfrac{65+80+95}{3} = 80$

$\tau_1 = \mu_1 - \mu_. = 65 - 80 = -15$

$\tau_2 = \mu_2 - \mu_. = 80 - 80 = 0$

$\tau_3 = \mu_3 - \mu_. = 95 - 80 = 15$

$\mu_. = 80.$

#16.34 determine sample sizes ($n_i = n$).
Assume $\sigma = 0.15$.

a.) n_i if differences in mean amount is to be detected with probability $\underline{0.70}$, ~~and~~ when Range ($\Delta = 0.15$), controlling for type I error at $\alpha = 0.05$?

I.) $\alpha = 0.05$, $\quad r = 6$

II.) Ratio $= \frac{\Delta}{\sigma} = \frac{0.15}{0.15} = 1$

From B.12, $\Delta/\sigma = 1$, $r = 6$, $\alpha = 0.5$, the required sample size for each cell is $n = 22$.

b.) With $n_i = n = 20$, what is the power of the test if:
$\mu_1 = 0.09$, $\mu_2 = 0.18$, $\mu_3 = 0.30$, $\mu_4 = 0.20$, $\mu_5 = 0.10$, $\mu_6 = 0.20$?

$$\phi = \frac{1}{\sigma}\sqrt{\frac{n}{r}\left(\Sigma(\mu_i - \mu_.)^2\right)}, \quad \mu_. = \frac{\Sigma \mu_i}{r}.$$

$$\mu_. = \frac{0.09 + 0.18 + 0.30 + 0.20 + 0.10 + 0.20}{6} \approx 0.17833$$

let $\lambda = \Sigma(\mu_i - \mu_.)^2$

$$= (0.09 - 0.17833)^2 + (0.18 - 0.17833)^2 + (0.30 - 0.17833)^2 + (0.20 - 0.17833)^2 + (0.10 - 0.17833)^2 + \cancel{(0.20)}\ (0.20 - 0.17833)^2$$

$\lambda = 0.0297$

So $\phi = \frac{1}{\sigma}\sqrt{\frac{n}{r}\lambda}$

$$= \frac{1}{0.15}\sqrt{\frac{22}{6}(0.0297)} \approx 2.2$$

$\nu_1 = 5 \quad \nu_2 = 114 \quad$ B.11 $\rightarrow 1 - \beta > 0.97 =$ Power.

c.) Pro ≥ 0.95, Smallest mean fill. differs by mean ≥ 0.10.

What is required Sample sizes

$1-\alpha = 0.95$ or greater
let the difference be $\lambda = 0.10$

$\frac{\lambda\sqrt{n}}{\sigma}$ Table B.13, $r=6$.

we get $\frac{\lambda\sqrt{n}}{\sigma} = 3.1591$

$\sqrt{n} = \cancel{\sigma} \; \left(\frac{\sigma(3.1591)}{\lambda}\right)^2$
n

$n = \left(\frac{0.15(3.1591)}{0.10}\right)^2 = 22.45$, round up $[22.45]$

So, we need a Sample size of 23 or more for this minimum smallest mean detection

$n = 23$.

☆ ①

17.3 (i) $\mu_1 + 3\mu_2 - 4\mu_3$

(ii) $0.3\mu_1 + 0.5\mu_2 + 0.1\mu_3 + 0.1\mu_4$

(iii) $\dfrac{\mu_1+\mu_2+\mu_3}{3} - \mu_4$.

a.) (i) and (iii) are Contrasts

(i) $c_1 = 1,\ c_2 = 3,\ c_3 = -4$

(iii) $c_1 = \frac{1}{3},\ c_2 = 1/3,\ c_3 = 1/3,\ c_4 = -1.$

b.) (i) $(1)^2 + (3)^2 + (-4)^2 = 26,$

$$s^2(\hat{L}) = \frac{26\,MSE}{n}$$

(iii) $= (1/3)^2 + (1/3)^2 + (1/3)^2 + (1)^2 = \frac{4}{3}$

$$s^2(\hat{L}) = \frac{4}{3n}(MSE)$$

(ii) $(0.3)^2 + (0.5)^2 + (0.1)^2 + (0.1)^2 = 0.36.$

$$s^2(\hat{L}) = \frac{0.36\,MSE}{n}.$$

#17.5 $r = 5$, $n_i = 5$

a.) T, S, B if $g = 2, 5, 10$ 95% Confidence.

$q(0.95, 5, 25-5) = q(0.95, 5, 20)$ Table B.9

$= 4.23$

$F(0.95, 4, 20)$, B.4, $\frac{1}{\sqrt{2}}(4.23) \approx 2.99$

$= 2.87$

$\sqrt{4(2.87)} \approx 3.39$

- $T = \frac{1}{\sqrt{2}} q(1-\alpha; r; n_T - r)$
- $S^2 = (r-1) F(1-\alpha; r-1; n_T - r)$
- $B = t(1-\alpha/2g; n_T - r)$

g	T	S	B
2	2.99	3.39	2.42
5	2.99	3.39	2.845
10	2.99	2.79	3.15

$g=2$ $t\left(1-\frac{0.05}{4}, 20\right) = 2.42$

$g=5$ $t\left(1-\frac{0.05}{10}, 20\right) = 3.845$

$g=10$. $t\left(1-\frac{0.05}{2(10)}, 20\right) = 3.15$

III

6.) $n_i = 20.$

g	T	S	B
2	2.79	3.14	2.28
5	2.79	3.14	2.63
10	2.79	3.14	2.87.

$$q(0.95; 5; 95) = 3.94$$

$$F(0.95; 4; 95) = 2.46.$$

$$T = \frac{1}{\sqrt{2}}(3.94) \approx 2.79$$

$$S = \sqrt{4(2.46)} \simeq 3.14$$

$$g=2 \quad t\left(1-\frac{0.05}{4}, 95\right) \simeq 2.28$$

$$g=5 \quad t\left(1-\frac{0.05}{10}, 95\right) \simeq 2.63$$

$$g=10 \quad t\left(1-\frac{0.05}{20}, 95\right) \simeq 2.87$$

$\frac{x}{\sqrt{2}}$ $\frac{2\sqrt{2}}{2}$

(1)

#11 Gram-Schmidt

Let $A = \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix} \end{array}$, goal $A = \boxed{Q}R$

Q;

Check $\det(Q) = \pm 1$, if yes, all columns are orthonormal.

$Q = [e_1 \vdots e_2 \vdots e_3 \vdots e_4 \vdots]$.

$u_k = v_k - \sum_{j=1}^{k-1} \text{proj}_{u_j}(v_k)$, $\text{proj}_u(v) = \frac{u \cdot v}{u \cdot u} u$.

and e_k is the orthonormal vector.

$e_k = \frac{u_k}{\|u_k\|}$. let $u_1 = v_1 = [1\ -1\ 0\ 0]^T$

$e_1 = \frac{u_1}{\|u_1\|} = \frac{u_1}{\sqrt{1^2 + (-1)^2 + 0^2 + 0^2}} = \frac{u_1}{\sqrt{2}}$

$\Rightarrow e_1 = (\sqrt{2}/2, -\sqrt{2}/2, 0, 0)^T$ (1)

$u_2 = V_2 - \frac{u_1 \cdot v_2}{u_1 \cdot u_1} u_1$

$u_1 \cdot u_1 = 2$.

$u_1 \cdot v_2 = ((1 \cdot 0) + (-1(1)) + 0 + 0) = -1$

~~$\frac{-1}{2} \langle u_1 \rangle$, $u_2 =$~~

$V_2 - \frac{1}{2} u_1$

$\langle 0, 1, -1, 0, 0 \rangle - \frac{1}{2} \langle 1, -1, 0, 0 \rangle$

$u_2 = \langle -0.5, 1.5, -1, 0, 0 \rangle$

$e_2 = \frac{u_2}{\|u_2\|} =$

$\|u_2\| = 1.8708$.

Q

(2)

$e_2 = \langle -0.5/1.8708,\ 1.5/1.8708,\ -1/1.8708,\ 0, 0\rangle,$

$$u_3 = v_3 - \frac{u_1 \cdot v_3}{u_1 \cdot u_1} u_1 - \frac{u_2 \cdot v_3}{u_2 \cdot u_2} u_2$$

then $e_3 = \dfrac{u_3}{\|u_3\|}$

$$\langle 0, 0, 1, -1, 0\rangle - \frac{u_1 \cdot v_3}{0} - \frac{-1}{3.5}\langle -0.5, 1.5, -1, 0, 0\rangle.$$

$$u_3 = \langle -0.1429,\ 0.4286,\ 0.7143,\ -1,\ 0\rangle.$$

$$e_3 = \frac{u_3}{\|u_3\|},\quad \|u_3\| = 1.3093.$$

$$e_3 = \langle -0.1429/1.3093,\ \frac{-0.4286}{1.3093},\ \frac{0.7143}{1.3093},\ \frac{-1}{1.3093},\ 0\rangle.$$

$$u_4 = v_4 - \frac{u_1 \cdot v_4}{u_1 \cdot u_1} u_1 - \frac{u_2 \cdot v_4}{u_2 \cdot u_2} u_2 - \frac{u_3\, v_4}{u_3\, u_3} u_3$$

$$\langle 0, 0, 0, 0, 1, -1\rangle - \cancel{\frac{0}{2}}\ \cancel{\frac{0}{3.5}}$$

$$Q = \begin{array}{c} \begin{array}{cccc} e_1 & e_2 & e_3 & e_4 \end{array} \\ \begin{pmatrix} \frac{\sqrt{2}}{2} & -0.5/1.8708 & -0.1429/1.3093 & - \\ & 1.5/1.8708 & -0.4286/1.3093 & - \\ \frac{-\sqrt{2}}{2} & -1/1.8708 & 0.7143/1.3093 & - \\ 0 & 0 & & \\ 0 & 0 & -1/1.3093 & - \\ & & 0 & - \end{pmatrix} \end{array}$$

0 ③

didn't obtain the last vector $\underset{\sim}{e}_4$,
I messed up in my calculation somewhere
somewhere, and I did Not want to
go back and do Gram-Schmidt by
hand for a 5x4 Matrix again.

The point is Gram-Schmidt is

$$u_k = v_k - \sum_{i}^{k-1} \text{Proj}_{u_i}(v_k).$$

Appendix

Question 16.11

Part A

```
getdata <- function(...){
  e = new.env()
  name = data(..., envir = e)[1]
  e[[name]]
}

data <- getdata("FillingMachines")

data2 <- data
# drop redundant row column
data2 <- subset(data2, select = -row)
#force group to factor variables (indicators)
data2$group <- as.factor(data2$group)

p1 <- data2 %>% ggplot(aes(x = group, y = y)) +
  geom_boxplot(aes(fill = factor(group))) +
  theme(legend.position = "none") +
  labs(title = "Factor Level Means",
       x = "Machine Number",
       y = "std. 32 oz.")
p1
```

Part B

```
factor_level_means <- data2 %>%
  group_by(group) %>%
  summarise(mean = mean(y))

kable(factor_level_means[], caption = "Fitted Values",
      format = "markdown") %>%
  kable_styling(position = "center")
```

Part C

```
deviation <- data2 %>%
  group_by(group) %>%
  summarise(resid = y-mean(y))

corgi = as.data.frame(cbind(Group_1 = deviation$resid[deviation$group == 1],
                          Group_2 = deviation$resid[deviation$group == 2],
                      Group_3 = deviation$resid[deviation$group == 3],
                      Group_4 = deviation$resid[deviation$group == 4],
                      Group_5 = deviation$resid[deviation$group == 5],
                      Group_6 = deviation$resid[deviation$group == 6]))
zero = sum(corgi$Group_1) + sum(corgi$Group_2) + sum(corgi$Group_3)
+ sum(corgi$Group_4) + sum(corgi$Group_5) + sum(corgi$Group_6)
machine1_dev <- as.data.frame(subset(deviation, group == 1))
```

```r
kable(corgi[], caption = "Deviations", format = "markdown") %>%
  kable_styling(position = "center")
```

Part D

```r
n_T <- 120
n_i <- 20
Y.. <- sum(data2$y)
YBar.. <- Y.. / n_T
SSTR <- n_i*((0.0735-YBar..)^2+(0.1905-YBar..)^2+(0.46-YBar..)^2+
               (0.3655-YBar..)^2+(0.125-YBar..)^2+(0.1515-YBar..)^2)
SSE <- sum((deviation$resid)^2)
MSTR <- SSTR/5
MSE <- SSE/(n_T-6)
```

Part F

```r
p.value <- pf(F.star, df1 = 5, df2 = 114, lower.tail = FALSE)
```