Homework 03

Joseph Blubaugh
jblubau1@tamu.edu

STAT 642-720

1)

a) i) Designed Experiment: create a random permutation of the experimental units and assign it to the treatments so that each treatment has the same number of experimental units

ii) Observational Experiment: you cannot randomly assign treatments to experimental units in this case so either treatment groups or charateristic group have to be formed where each unit as an equal chance of being selected. The randomization will hold true only within in individual groups.

b) i) μ_i is the mean response for the ith treatment

ii) β is the Best Linear Unbiased Estimator

c) e_ij are iid random variables with a normal distribution

d) You can do an F-Test to test of the Reduced Model vs. the Full Model

2)

a) The power of the test is 0.946295

```
mu_hat = 1/21 * (90*4 + 100*5 + 120*3 + 110*5 + 80*4)
lambda = (4*(90 - mu_hat)^2 +
            5*(100 - mu_hat)^2 +
            3*(120 - mu_hat)^2 +
            5*(110 - mu_hat)^2 +
            4*(80 - mu_hat)^2) / 150
F_stat = qf(.95, 4, 16)
1 - pf(F_stat, 4, 16, lambda)

[1] 0.946295
```

b) We would need at least 17 chickens per treatment to detect the difference

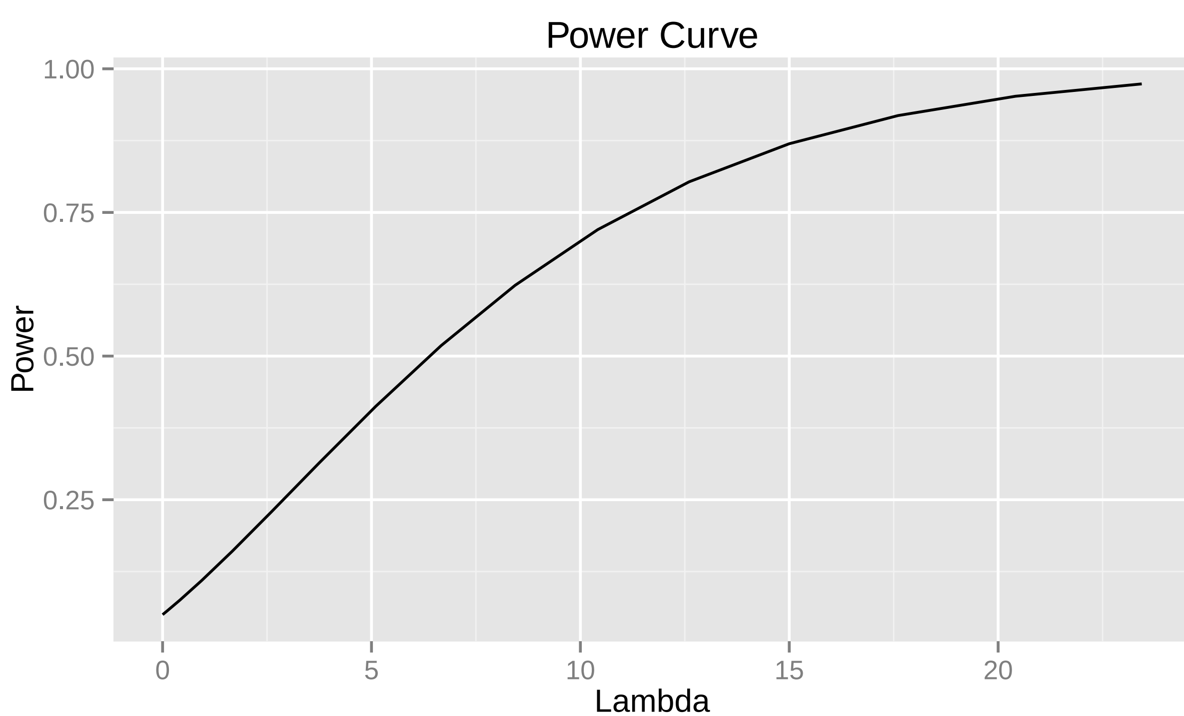
```
r = 2:20; t = 5; u1 = t - 1; u2 = t * (r - 1); S = 150; D = 20
L = (r * D^2) / (2 * S); phi = sqrt(L/t)

data.frame(
  Treaments = t,
  Reps = r,
  DF.1 = u1,
  DF.2 = u2,
  Lambda = round(L, 3),
  Phi = round(phi, 3) ,
  Power = round(1 - pf(qf(.99, u1, u2), u1, u2, L), 3)
)
```

	Treaments	Reps	DF.1	DF.2	Lambda	Phi	Power
1	5	2	4	5	2.667	0.730	0.028
2	5	3	4	10	4.000	0.894	0.064
3	5	4	4	15	5.333	1.033	0.116
4	5	5	4	20	6.667	1.155	0.182
5	5	6	4	25	8.000	1.265	0.256
6	5	7	4	30	9.333	1.366	0.336
7	5	8	4	35	10.667	1.461	0.418
8	5	9	4	40	12.000	1.549	0.497
9	5	10	4	45	13.333	1.633	0.572
10	5	11	4	50	14.667	1.713	0.641
11	5	12	4	55	16.000	1.789	0.703
12	5	13	4	60	17.333	1.862	0.757
13	5	14	4	65	18.667	1.932	0.803
14	5	15	4	70	20.000	2.000	0.843
15	5	16	4	75	21.333	2.066	0.875
16	5	17	4	80	22.667	2.129	0.902
17	5	18	4	85	24.000	2.191	0.924
18	5	19	4	90	25.333	2.251	0.941
19	5	20	4	95	26.667	2.309	0.955

3)

a)



b) The engineer would need 26 intersections

```r
reps = 10:27
treatments = 3
alpha = .05
variance = 12
u1 = 16
u2 = 17
u3 = 19
mean_u = (u1 + u2 + u3) / treatments
Lambda = reps * ( (u1 - mean_u)^2 + (u2 - mean_u)^2 + (u3 - mean_u)^2 ) / variance
df.1 = treatments - 1
df.2 = treatments * (reps - 1)
F.stat = qf(1 - alpha, df.1, df.2)
Power = 1 - pf(F.stat, df.1, df.2, Lambda)

data.frame(
  reps = reps,
  df.1 = df.1,
  df.2 = df.2,
  F.stat = F.stat,
  Lambda = Lambda,
  Power = Power
)
```

```
   reps df.1 df.2   F.stat    Lambda     Power
1    10    2   27 3.354131  3.888889 0.3671176
2    11    2   30 3.315830  4.277778 0.4035094
3    12    2   33 3.284918  4.666667 0.4390621
4    13    2   36 3.259446  5.055556 0.4736186
5    14    2   39 3.238096  5.444444 0.5070513
6    15    2   42 3.219942  5.833333 0.5392589
7    16    2   45 3.204317  6.222222 0.5701648
8    17    2   48 3.190727  6.611111 0.5997140
9    18    2   51 3.178799  7.000000 0.6278708
10   19    2   54 3.168246  7.388889 0.6546165
11   20    2   57 3.158843  7.777778 0.6799474
12   21    2   60 3.150411  8.166667 0.7038721
13   22    2   63 3.142809  8.555556 0.7264101
14   23    2   66 3.135918  8.944444 0.7475901
15   24    2   69 3.129644  9.333333 0.7674480
16   25    2   72 3.123907  9.722222 0.7860255
17   26    2   75 3.118642 10.111111 0.8033693
18   27    2   78 3.113792 10.500000 0.8195295
```

4)

a) Cell Means Model: $Y_{ij} = u_i + e_{ij}; i = 1, 2, 3; j = 1, ..., n; n = \sum_{i=1}^{t} n_i$

$$Y = \begin{bmatrix} y_{1,1} \\ y_{1,2} \\ \cdots \\ y_{1,12} \\ y_{2,1} \\ y_{2,2} \\ \cdots \\ y_{2,14} \\ y_{3,1} \\ y_{3,2} \\ \cdots \\ y_{3,11} \end{bmatrix} X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ . & . & . \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ . & . & . \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ . & . & . \\ 0 & 0 & 1 \end{bmatrix} \beta = \begin{bmatrix} 25.2 \\ 32.6 \\ 28.1 \end{bmatrix} e = \begin{bmatrix} e_{1,1} \\ e_{1,2} \\ \cdots \\ e_{1,12} \\ e_{2,1} \\ e_{2,2} \\ \cdots \\ e_{2,14} \\ e_{3,1} \\ e_{3,2} \\ \cdots \\ e_{3,11} \end{bmatrix}$$

b) We can conclude that there is not a significant difference in the means

```
u1 = 25.2; u2 = 32.6; u3 = 28.1
n1 = 12; n2 = 14; n3 = 11
s1 = 3.6; s2 = 4.8; s3 = 5.3

mu = ( 1 / (n1 + n2 + n3)) * (n1*u1 + n2*u2 + n3*u3)
se = ( 1 / (n1 + n2 + n3)) * (n1*s1^2 + n2*s2^2 + n3*s3^2)

lambda = ( n1*(u1-mu)^2 + n2*(u2-mu)^2 + n3*(u3-mu)^2 ) / se

1 - pf(qf(.99, 2, 34), 2, 34, lambda)
```

```
[1] 0.8330068
```

5)

a) $\tau_4 = -\frac{1}{3}\sum_{i=1}^{t-1} n_i\tau_i = -\frac{1}{3}(6(-2.3) + 3(-1.7) + 5(1.8)) = 3.3$

b) Randomizing the mice to the treatments will minimize any bias that might occur from differences in the mice

c) There is only one sample of paint from each manufacturer and each manufacturer might have provided only their best paint so the experiment can only measure the effectiveness of the single batch instead of the effectiveness of several paint samples

d) Answer: (c) τ_1 is the difference between the model mean and the mean of P5 because in SAS, by default the last treatment has a coefficient of 0, meaning the model μ is its coefficient, so Pl is 2.3 less than P5.

e) At least 6 samples per treatment are required

```
r = 2:6; t = 5; u1 = t - 1; u2 = t * (r - 1); S = 2; D = 5.1
L = (r * D^2) / (2 * S); phi = sqrt(L/t)

data.frame(
Treaments = t,
Reps = r,
DF.1 = u1,
DF.2 = u2,
Lambda = round(L, 3),
Phi = round(phi, 3) ,
Power = round(1 - pf(qf(.99, u1, u2), u1, u2, L), 3)
)
```

```
  Treaments Reps DF.1 DF.2 Lambda   Phi Power
1         5    2    4    5 13.005 1.613 0.145
2         5    3    4   10 19.508 1.975 0.499
3         5    4    4   15 26.010 2.281 0.796
4         5    5    4   20 32.512 2.550 0.936
5         5    6    4   25 39.015 2.793 0.984
```

```
No id variables; using all as measure variables
```

```
         .L          .Q          .C          ^4
 1.93531393 -0.03741657 -0.06008328  0.47689622
```

```
                .L         .Q            .C         ^4
[1,] 1 -0.6324555  0.5345225 -3.162278e-01  0.1195229
[2,] 1 -0.3162278 -0.2672612  6.324555e-01 -0.4780914
[3,] 1  0.0000000 -0.5345225 -3.096264e-16  0.7171372
[4,] 1  0.3162278 -0.2672612 -6.324555e-01 -0.4780914
[5,] 1  0.6324555  0.5345225  3.162278e-01  0.1195229
```

6)

a) C1 is not mutually orthogonal with the other constrasts. At most there can be t-1 mutually othogonal contrasts, but C1 is specific and we have to weight T30 against the other Temps so we really have t=4 and the 3 available mutually othgonal contrasts can be applied to C2, C3, and C4.

b)

Parameter	Estimate	Standard Error	t Value	Pval
T30 vs Other	6.21	1.034	6.00	<.0001
Linear	6.12	0.731	8.37	<.0001
Quadratic	-0.14	0.865	-0.16	0.8722
Cubic	-0.19	0.731	-0.26	0.7962

c) We have evidence to support all of the contrasts are not equal to 0

```
library(knitr)

Con = data.frame(
  C1 = c(-1, -1, -1, -1, 4),
  C2 = c(-2, -1, 0, 1, 2),
  C3 = c(2, -1, -2, -1, 2),
  C4 = c(-1, 2, 0, -2, 1)
)

n = 10; t = 5
Dh = data.frame(Dh = colSums(Con^2 / n))
Sh = (1.141542*sqrt((t - 1)* qf(.95, 4, 45))) * Dh
colnames(Sh) = "Sh"
Ch = data.frame( Ch = abs(colSums(c(6.21, 6.12, -.14, -.19) * Con)))
Con.t = t(Con)
colnames(Con.t) = c(1,2,3,4,5)

Scheffe = cbind(Con.t, Dh, Sh, Ch)

kable(Scheffe)
```

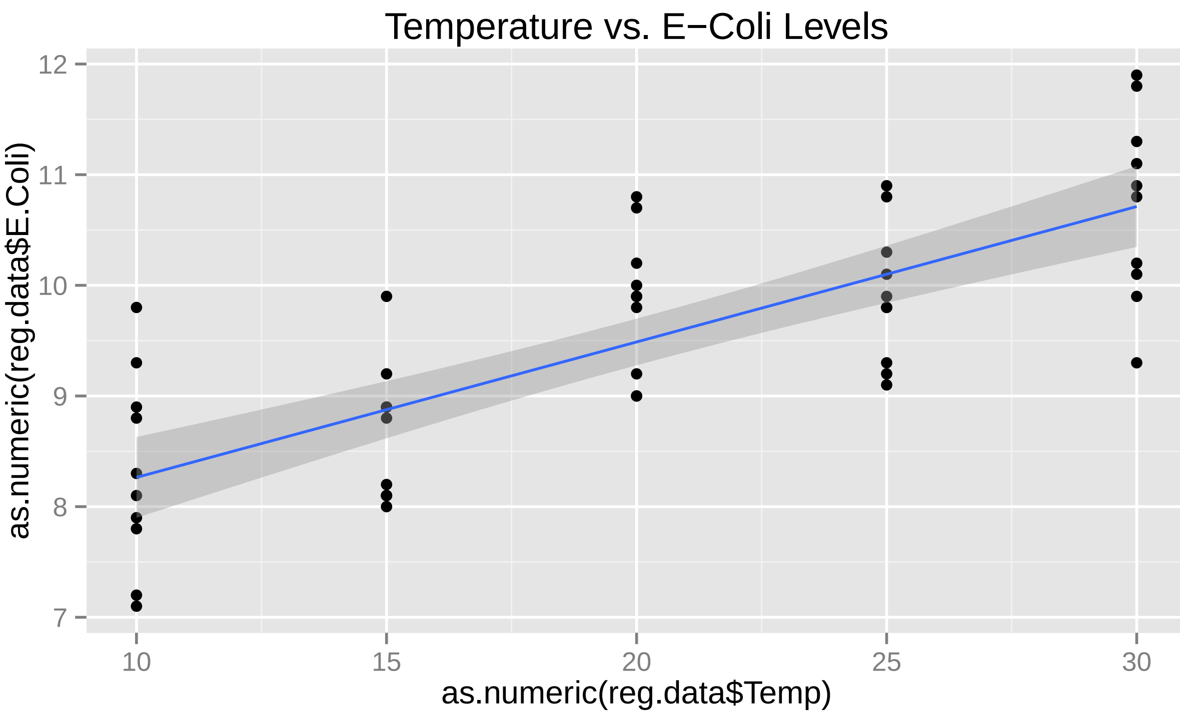
	1	2	3	4	5	Dh	Sh	Ch
C1	-1	-1	-1	-1	4	2.0	7.332560	12.84
C2	-2	-1	0	1	2	1.0	3.666280	6.35
C3	2	-1	-2	-1	2	1.4	5.132792	18.91
C4	-1	2	0	-2	1	1.0	3.666280	12.70

d) For Bonferroni we have to divide the significance level by M = 4 so the actual alpha we are looking for is $\alpha = .0125$. From the table in 6a, both the T30 vs Other and Linear Contrasts are still significant.

e) There is not significant evidence that the 3 contrasts are not equal to 0

```
         [,1]
[1,] 0.807593
```

f) There is a clear linear trend between Temperature and the detected E-Coli levels. A measure of correlation returns .76 which is a strong trend. The trend is easily detected with a simple plot.



7)

a) Temperate at 10 degrees has the lowest Ecoli Concentration

Temp	Est	LSMEAN 95% Confidence Limits
T10	8.32	7.85, 8.78
T15	8.62	8.15, 9.08
T20	9.85	9.38, 10.31
T25	9.92	9.45, 10.38
T30	10.73	10.26, 11.19

b) All of the temperature levels have a mean E-Coli concentration below the mean of the E-coli concentration at 30 degrees

c) Different Pairs using the Tukey method: (1 & 3), (1 & 4), (1 & 5) (2 & 3), (2 & 4), (2 & 5)

Least Squares Means for effect Temp Pr > |t| for H0: LSMean(i)=LSMean(j)

	1	2	3	4	5
1		0.8888	0.0002	0.0001	<.0001
2	0.8888		0.0042	0.0022	<.0001
3	0.0002	0.0042		0.9995	0.0713
4	0.0001	0.0022	0.9995		0.1143
5	<.0001	<.0001	0.0713	0.1143	