4QZ3 Midterm - 400138679

Jeff Suitor

10/22/2021

Setup

Q1

Transdermal - Water soluble so not easily absorbed, but less than 500 Daltons so some absorbed. Only choice with a very low concentration and assuming that it takes a long time to absorbed. Therefore A is the only appropriate choice. The reason is that A has the lowest concentration.

Ingested - Has a very long absorption time and because it is non metabolized so you can model as a 1 compartment system. Therefore, the decrease in concentration should be linear. Therefore F is the only appropriate choice for long time of action and linear decrease.

Inhaled - Has a fast time of action but not as fast as intravenous. Also, water soluble drugs are not absorbed as quickly by the lungs as lipid soluble meaning that the choice of the fastest time of action is definitely not appropriate. Therefore, C which has a fast but not the fastest time of action is.

Intravenous - Fastest time of action and 100% bioavailable. Therefore B is the only appropriate choice.

Q2

Yes, absorbed faster because lipid soluble.
Yes, they are ingestible.
Yes, they are absorbed fast through the capillaries.
No, Lipid soluble drugs are not filtered through the kidney.

Q3

No, lipid soluble so absorbed fast.
No, less than 500 Daltons.
No, nonmetabolized.
No, water soluble are trapped in circulatory not lipid soluble.
Yes, expired through the lungs.
No, lipid soluble so no kidney.

Q4

```
6.65 > 5.26
```

```
## [1] TRUE
```

```
12.36 > 13.76
```

```
## [1] FALSE
```

There is an effect of treatment. But there is no difference between patients.

Q5

```
k = 36
ci = 0.33 * 1000 # mmol
cfinal = 0.008
t = round((-k * log(cfinal/ci))/log(2), 2)
t
```

```
## [1] 551.96
```

$$
\begin{aligned}
C(t) &= C_i * 2^{\frac{-t}{k}} \\
0.008 &= 330 * 2^{\frac{-t}{36}} \\
\frac{0.008}{330} &= 2^{\frac{-t}{36}} \\
ln(\frac{0.008}{330}) &= ln(2^{\frac{-t}{36}}) \\
ln(\frac{0.008}{330}) &= \frac{-t}{36} * ln(2) \\
\frac{ln(\frac{0.008}{330})}{ln(2)} &= \frac{-t}{36} \\
t &= \frac{-36 * ln(\frac{0.008}{330})}{ln(2)} \\
t &= 551.96
\end{aligned}
$$

Q6

```
k = 44
ci = 1.6 * 1000 # mmol
cfinal = 0.51
t = (-k * log(cfinal/ci))/log(2)
dc = ci * 2^(-t/k) * log(2) * -1/k
sigma_t = round((cfinal^2 / dc^2)^0.5, 4)
t
```

```
## [1] 511.0726
```

```
dc
```

```
## [1] -0.008034206
```

```
sigma_t
```

```
## [1] 63.4786
```

$$
\begin{aligned}
C(t) &= C_i * 2^{-t/k} \\
t &= \frac{-k * ln(\frac{C(t)}{C_i})}{ln(2)} \\
t &= \frac{-44 * ln(\frac{0.51}{1600})}{ln(2)} \\
t &= 511.0726
\end{aligned}
$$

$$
\begin{aligned}
C(t) &= C_i * 2^{-t/k} \\
\frac{dC}{dt} &= C_i * 2^{-t/k} * ln(2) * -1/k \\
\frac{dC}{dt} &= 1600 * 2^{-t/44} * ln(2) * -1/44 \\
t &= -0.008034206
\end{aligned}
$$

$$
\begin{aligned}
(\sigma_{C(t)})^2 &= (\sigma_t)^2 * (\frac{dC}{dt})^2 \\
(\sigma_t)^2 &= \frac{(\sigma_{C(t)})^2}{(\frac{dC}{dt})^2} \\
(\sigma_t) &= \sqrt{\frac{(\sigma_{C(t)})^2}{(\frac{dC}{dt})^2}} \\
t &= 63.4786
\end{aligned}
$$

Q7

```
mean1 = 8.4
mean2 = 6.46
sigma = 8.09
mode = 16.0
median = 13.0
phi = 3

n = (phi * sigma/ (mean1 - mean2))^2
ceiling(n)
```

```
## [1] 157
```

$$
\begin{aligned}
n &= (\frac{\phi * \sigma}{mean_1 - mean_2})^2 \\
&= (\frac{3 * 8.09}{8.4 - 6.46})^2 \\
&= ceiling(156.5) \\
&= 157
\end{aligned}
$$

Q8

Yes, Lumped parameter model. No
No
No

Q9

No
No
Yes, this is the description of poor model fitting
No

Q10

```
countries = c("Italy", "Spain", "Portugal", "Netherlands")
num_countries = 4
num_players_per_country = 4
num_players = 16
num_replicates = 3
data = {}
data$values = as.numeric(unlist(read.table('Q10.txt')))
data$country = gl(4, num_players_per_country*num_replicates, labels=factor(countries))
data$unique = gl(length(data$values), 1)
data$players = gl(num_players, num_replicates)
summary(aov(values ~ players, data = data)) # Players
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## players     15  74.77   4.984   0.782  0.687
## Residuals   32 203.98   6.374
```

```
summary(aov(values ~ country + players, data=data)) # Error
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## country      3   1.80   0.598   0.094  0.963
## players     12  72.97   6.081   0.954  0.509
## Residuals   32 203.98   6.374
```

Not effective because there is no significance in the Anova therefore you accept the null hypothesis that there is no difference.

Q11

```
data = {}
data$values = as.numeric(unlist(read.table('Q11.txt')))
data$drug = gl(7, 86, labels=c("D1", "D2", "D3", "D4", "D5", "D6", "D7"))
data$unique = gl(length(data$values), 1)
model = aov(values ~ drug, data=data)
model_sum = summary(model)
model_sum
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## drug          6 121140   20190    80.3 <2e-16 ***
## Residuals   595 149600     251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P is less than 0.05 so there is a difference between the different drugs. Need to perform post hoc testing to determine which drugs are different.

```
LSD.test(model, "drug", console=TRUE)
```

```
##
## Study: model ~ "drug"
##
## LSD t Test for values
##
## Mean Square Error:  251.4284
##
## drug,  means and individual ( 95 %) CI
##
##      values       std  r      LCL      UCL   Min    Max
## D1 35.23919  4.335714 86 31.88111 38.59726 18.90  46.49
## D2 16.47709  5.364546 86 13.11902 19.83517  8.28  31.19
## D3 21.21802 13.044613 86 17.85995 24.57610  3.78  83.60
## D4 28.73756 19.013714 86 25.37948 32.09563  3.30  99.60
## D5 14.83674  6.027741 86 11.47867 18.19482  4.96  33.44
## D6 58.13791 27.040672 86 54.77983 61.49598 16.20 148.99
## D7 40.05291 20.327497 86 36.69483 43.41098  5.82  88.83
##
## Alpha: 0.05 ; DF Error: 595
## Critical Value of t: 1.963959
##
## least Significant Difference: 4.749035
##
## Treatments with the same letter are not significantly different.
##
##      values groups
## D6 58.13791      a
## D7 40.05291      b
## D1 35.23919      c
## D4 28.73756      d
## D3 21.21802      e
## D2 16.47709     ef
## D5 14.83674      f
```

Therefore, D5 and D2 are not significantly different. D2 and D3 are not significantly different. All other combinations are significantly different. To improve the design I would have used blocking to reduce the experimental error while being aware of potential blocking error. Another option could be to use subsampling to increase the number of observations per experimental unit allowing for a more detailed statistical analysis.

Q12

```
F = 550 / 60 / 1000
V = 150/1000
R = 0.1
t = 20
V
```

```
## [1] 0.15
```

```
F
```

```
## [1] 0.009166667
```

```
C20 = R/F*(1-exp(-F/V*t))
C20
```

```
## [1] 7.695547
```

From 0 to 20 s (Model 4):
R = rate of injection = 0.1 $\frac{mols}{s}$
F = flow rate = 550 $\frac{ml}{min}$ = $\frac{550}{60*1000}\frac{L}{s}$ = 33 L/s
V = volume = 150 ml = 0.15 L
C(t) = $\frac{mols}{L}$

$$C(t) = \frac{R}{F}(1 - e^{-(F/V)t})$$
$$C(t) = \frac{0.1}{0.0092}(1 - e^{-(0.0092/0.15)t})$$

From 20s onward (Model 1):

$$C(20) = \frac{0.1}{0.0092}(1 - e^{-(0.0092/0.15)t})$$
$$C(20) = 7.7\frac{mols}{L}$$
$$C(t) = C_i e^{-(F/V)t}$$
$$C(t) = C(20)e^{-(0.0092/0.15)(t-20)})$$

Q13

Due to the machine only outputting the 12 mean and standard deviation value there is a large amount of data missing. Because the graduate student doesn't have the individual data for each of the wells he is unable to validate his data and ensure the accuracy of his assumptions. Due to only having the mean and standard deviations the graduate student is unable to check the data for kurtosis to view how outlier prone the data is. Additionally, he is unable to check his data for skewness to check the normality of his data or whether there is asymmetry about the sample mean.

Ideally the graduate student would want to re run the experiment this time collecting the data for the 96 samples as well the information for each person. The student can however,still compare kurtosis and skewness between the different people to view if his population selection is normally distributed and not outlier prone. This can be done by taking the mean of and standard deviation of all the people and using their means to perform the kurtosis and skewness calculations.

Q14

Multiple linear regression attempts to fit a model involving 2 or more explanatory variables to a single output variable using a linear regression.

In the case of this example the model would be:
$y = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 d_3 + e$

Where d_1 d_2 and d_3 represent the different drugs while β_0 β_1, β_2, and β_3 are the partial regression coefficients for the intercept and the various drugs, and e is the associated with the model. It is important that when gathering data that it is gathered in an ethical manner. Participants must give their informed consent and should be made aware of the risks and benefits of participation. Furthermore, it is important to make sure that participant data is anonymized and kept confidential. During the experimental design risk assessment and minimization should also occur ensuring participants are as safe as possible. It is also important during the design to ensure the testing is being administered by trained professionals and that the research is conducted in a transparent and properly accountable way. Finally, any findings should be for the benefit of all and should be made publically available.

Each patient will first have their tumour volume measured before being assigned a dosage for the drug cocktail. Based on an appropriate frequency, the tumour volume for each patient will be measured. Once the experiment has been completed a least squares analysis will be performed to determine the partial regression coefficients. Afterwards, a multiple regression Anova and determine if the null hypothesis that none of these drugs impact tumour volume can be accepted or rejected. Should there be significance in the Anova and the null hypothesis cannot be rejected post hoc testing such as an LSD or Duncan's test can be performed to asses which drugs are significant. There are also several assumptions that are made at this point such as the error coming from the normal distribution and that the errors are uncorelated. It is important to test these assumption either via overfitting the model or examining the residuals for assumption validity. Outlier tests such as generalized ESD test can be performed to analyze the data for outlier values.

The model will have to undergo reduction testing to check if model parameters can be removed to better optimize the design.

Q15

Constants:
R_S
P_{SA}
C_{SA}
T_S
Q_{max}
T_{max}

$G = \frac{1}{C_{SA}}(Q_{AO} - \frac{P_{SA}}{R_S})$

$$Q_{AO} = \begin{matrix} \frac{Q_{max}t}{T_{max}} & 0 \le t \le T_{max} \\ \frac{Q_{max}(T_S - t)}{T_S - T_{max}} & T_{max} \le t \le T_s \\ 0 & T_s \le t \le T \end{matrix}$$

$0 \le t \le T_{max}$:

$$G = \frac{1}{C_{SA}}(\frac{Q_{max}t}{T_{max}} - \frac{P_{SA}}{R_S})$$
$$\frac{dG}{dt} = \frac{Q_{max}}{C_{SA}T_{max}}$$

$T_{max} \le t \le T_s$:

$$G = \frac{1}{C_{SA}}(\frac{Q_{max}(T_S - t)}{T_S - T_{max}} - \frac{P_{SA}}{R_S})$$
$$\frac{dG}{dt} = \frac{-Q_{max}}{C_{SA} * (T_S - T_{max})}$$

$T_s \le t \le T$:

$$G = \frac{1}{C_{SA}}(0 - \frac{P_{SA}}{R_S})$$
$$\frac{dG}{dt} = 0$$

$(\sigma_G)^2 = (\sigma_t)^2 * (\frac{dG}{dt})^2$

$$\sigma_G = \begin{matrix} \sqrt{(0.05)^2 * (\frac{Q_{max}}{C_{SA}T_{max}})^2} & 0 \le t \le T_{max} \\ \sqrt{(0.05)^2 * (\frac{-Q_{max}}{C_{SA} * (T_S - T_{max})})^2} & T_{max} \le t \le T_s \\ 0 & T_s \le t \le T \end{matrix}$$