Question 1

1)

Using the formula given in lecture 4 to calculate the confidence interval:

$$\left(\hat{\mu}_{\mathrm{ML}} - t_{\alpha/2,n-1}\frac{\hat{\sigma}_u}{\sqrt{n}},\ \hat{\mu}_{\mathrm{ML}} + t_{\alpha/2,n-1}\frac{\hat{\sigma}_u}{\sqrt{n}}\right)$$

The estimated mean of the data is:
10.367

There are 20 data points, so the percentile for a 95% confidence interval with 19 degrees of freedom is:
2.093

Using the formula given in the lecture to estimate variance:

$$\hat{\sigma}_u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{\mu}_{\mathrm{ML}})^2$$

The estimated variance of the data is:
7.478

Thus, our values for a 95% confidence interval are:

$$(10.367\ -\ 2.093 \times \frac{7.478}{\sqrt{20}},\ 10.367\ +\ 2.093 \times \frac{7.478}{\sqrt{20}})$$

Simplified:
(6.866, 13.868)

This means that the estimated average fuel efficiency of all wheel drive vehicles is 10.367 km/L, and that we have a 95% confidence that the true mean is between 6.866 km/l and 13.868 km/L

2)

An approximate z value for a 95% confidence interval is:
1.96

For **all wheel drive** cars, we can use the values from the previous question.

Estimated mean:
10.367

Estimated variance:
7.478

For **part-time 4 wheel drive** cars, we can calculate the values:

Estimated mean:
8.750

Estimated variance:
11.857

We can calculate an estimated difference of means by subtracting one estimated mean from the other:

10.367 - 8.750 = 1.617

The observed difference in fuel efficiency between AWD and par-time 4WD cars is 1.617km/L

Using the formula to calculate the 95% confidence interview from the lecture:

$$\left(\hat{\mu}_A - \hat{\mu}_B - z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}},\ \hat{\mu}_A - \hat{\mu}_B + z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}\right)$$

We can calculate the confidence interval, which is:

$$(10.367 - 8.750 - 1.96\sqrt{\frac{7.478}{20} + \frac{11.857}{25}},\ 10.367 - 8.750 + 1.96\sqrt{\frac{7.478}{20} + \frac{11.857}{25}})$$

Simplified:
(-0.188, 3.422)

These results mean that we have observed that AWD cars, on average, will travel 1.617km further per litre of fuel, and that we are 95% confident that the actual difference is between -0.188km/L and 3.422km/L.

3)

Using the formula from lecture 5, and our previous calculated values:

$$z_{(\hat{\mu}_x - \hat{\mu}_y)} = \frac{\hat{\mu}_x - \hat{\mu}_y}{\sqrt{\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}}}$$

We can calculate a test statistic for our hypothesis test:

$$\frac{10.367 - 8.750}{\sqrt{\frac{7.478}{20} + \frac{11.857}{25}}}$$

Simplified:

$$z(\hat{u}x - \hat{u}y) = 1.244$$

Now we perform a null hypothesis test. We are testing the null hypothesis that all wheel drives are, on average, less efficient than part-time 4 wheel drive cars. We can make the following statements:

H0 = AWD km travelled per litre of fuel >= P4WD km travelled per litre of fuel

HA = AWD km travelled per litre of fuel < P4WD km travelled per litre of fuel

Thus, we want to perform a left tailed test, as shown in the 3rd formula found in the lecture:

$$p \approx \begin{cases} 2\,\mathbb{P}(Z < -|z_{(\hat{\mu}_x - \hat{\mu}_y)}|) & \text{if } H_0 : \mu_x = \mu_y \text{ vs } H_A : \mu_x \neq \mu_y \\ 1 - \mathbb{P}(Z < z_{(\hat{\mu}_x - \hat{\mu}_y)}) & \text{if } H_0 : \mu_x \leq \mu_y \text{ vs } H_A : \mu_x > \mu_y \\ \mathbb{P}(Z < z_{(\hat{\mu}_x - \hat{\mu}_y)}) & \text{if } H_0 : \mu_x \geq \mu_y \text{ vs } H_A : \mu_x < \mu_y \end{cases}$$

When computing the value in r, we are left with

$$p \approx 0.893$$

This means that, when assuming that the null hypothesis is true, there is an approximately 89.3% chance that we observe results that are just as, or more extreme as our actual observations. With such a large p value, we do not have enough evidence from our tests to reject the null hypothesis.

4)

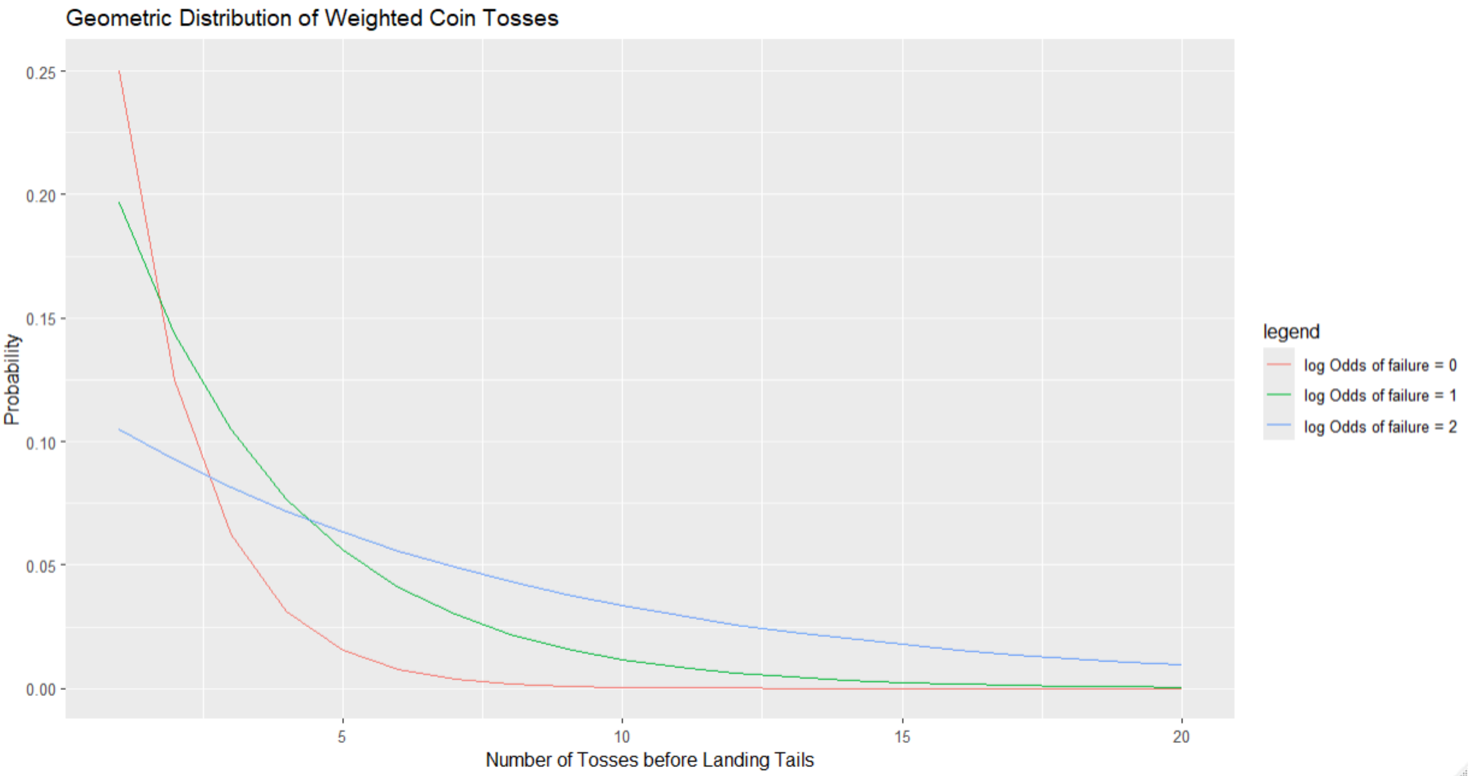
Some potential problems with my conclusions may be due to a wide variety of factors. There is only a sample size of 20 for AWD cars, and 25 for part time 4WD cars. This sample size may be too small to properly represent the total population. Additionally the complete data may not be normally distributed, which invalidates the results from our tests, which assume a normal distribution.

Code:

```
data = read.csv("fuel.ass2.2024.csv")
data.AWD <- subset(data, Type == "A")
data.PWD <- subset(data, Type == "P")

#Question 1.1

sampleMeanAWD <- mean(data.AWD$FA)
sampleMeanAWD

qt(p = 1 - 0.05/2, df = 19)

data.AWD$varianceCalc <- (data.AWD$FA - sampleMeanAWD) ^ 2

sampleVarianceAWD <- (1 / 19) * sum(data.AWD$varianceCalc)
sampleVarianceAWD

#Question 1.2

sampleMeanPWD <- mean(data.PWD$FA)
sampleMeanPWD

data.PWD$varianceCalc <- (data.PWD$FA - sampleMeanPWD) ^ 2

sampleVariancePWD <- (1 / 19) * sum(data.PWD$varianceCalc)
sampleVariancePWD

#Question 1.3
pnorm(1.244)
```

Question 2

1)



2)

2) $L(y|\phi) = \prod_{i=1}^{n} p(y^i|\phi)$

$\to \prod_{i=1}^{n}\left((e^{\phi}+1)^{-y^i-1} e^{y^i\phi}\right)$

$= (e^{\phi}+1)^{-\sum_{i=1}^{n} y^i - \sum_{i=1}^{n} 1} e^{\phi\sum_{i=1}^{n} y^i}$

$= (e^{\phi}+1)^{-\sum_{i=1}^{n}(y^i+1)} e^{\phi\sum_{i=1}^{n} y^i}$

3)

3) $-\log(L(y|\phi)) = -\log\left((e^{\phi}+1)^{-\sum_{i=1}^{n}(y^i+1)} e^{\phi\sum_{i=1}^{n} y^i}\right)$

$= \left(-\left(\sum_{i=1}^{n} y_i+1\right)\log(e^{\phi}+1) + \left(\sum_{i=1}^{n} y_i\right)\log e^{\phi}\right.$

$\hookrightarrow$ let $\sum_{i=1}^{n} y_i = S$

$\to -\left(-(S+n)\log(e^{\phi}+1) + S\,\overset{\phi}{\cancel{\log \phi}}\right)$

$\to (S+n)(\log(e^{\phi}+1)) - S\phi$

4)

4) $\frac{d}{d\phi} -\log L(y|\phi) = 0$, solve for ϕ

$$\frac{d}{d\phi} -\log L(y|\phi) = (s+n)\frac{e^{\phi}}{1+e^{\phi}} - s = 0$$

$$(s+n)\frac{e^{\phi}}{1+e^{\phi}} = s$$

$$\frac{e^{\phi}}{1+e^{\phi}} = \frac{s}{s+n}$$

$$e^{\phi} = (1+e^{\phi})\frac{s}{s+n}$$

$$e^{\phi} = \frac{s}{s+n} + \frac{se^{\phi}}{s+n}$$

$$e^{\phi} - \frac{se^{\phi}}{s+n} = \frac{s}{s+n}$$

$$e^{\phi}\left(1 - \frac{s}{s+n}\right) = \frac{s}{s+n}$$

$$e^{\phi} = \frac{s}{(s+n)\left(1-\frac{s}{s+n}\right)}$$

$$e^{\phi} = \frac{s}{s+n-s}$$

$$e^{\phi} = \frac{s}{n}$$

$\hat{\phi} = \log_e\left(\frac{s}{n}\right)$, where $s = \sum_{i=1}^{n} y_i$

$\therefore$ $\frac{s}{n}$ = sample mean

$\therefore$ $\cancel{\phi = \hat{4}y}$ $\hat{\phi} = \hat{\mu}_y$

5) I GOT NO CLUE HOW TO DO THIS, SHOULDVE GONE TO THE CONSULTATION

Question 3
1 and 2)

1) $\hat{p} = \frac{168}{240} = 0.7$

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.7 \times 0.3}{240}} \approx 0.0296$$

For 95% CI, $Z = 1.96$

$\therefore$ CI $= (\hat{p} - Z \times SE,\ \hat{p} + Z \times SE)$

$\rightarrow$ CI $= (0.7 - 0.29 \times 1.96,\ 0.7 + 0.29 \times 1.96)$

95% CI $= (0.643,\ 0.757)$ where right = success
left = failure

2) $H_0 \rightarrow \theta = 0.5$

$H_A \rightarrow \theta \neq 0.5$

$$Z_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{\sqrt{\theta_0(1-\theta_0)/n}}$$

$$Z_{\hat{\theta}} = \cancel{\frac{0.7 - 0.5}{}} \quad \frac{0.7 - 0.5}{\sqrt{\frac{0.25}{240}}} \approx \cancel{6.25}\ 6.197$$

$p \approx 0.213$

This p-val suggests a rejection of the null hypothesis ~~that~~ and that there is a preference in head-turning direction

3)
Using the following R code:

```
dataTilt <- c(rep(1, 168), rep(0, 72))
t.test(dataTilt)
```

We are left with a very small p-value of `2.2e-16,` which leads us to reject the null hypothesis, and that there is in fact a preference to which side people tilt their head when kissing.

4)

4) $\hat{\theta}_p = \frac{168 + 213}{240 + 240} \approx 0.794$

$\hat{\theta}_x = \frac{168}{240} \approx 0.7$

$\hat{\theta}_y = \frac{213}{240} \approx 0.888$

$$z(\hat{\theta}_x - \hat{\theta}_y) = \frac{0.7 - 0.888}{\sqrt{0.794(1-0.794)\left(\frac{1}{240} + \frac{1}{240}\right)}}$$

$$\approx -5.092$$

~~in RQ~~ $H_0 : \theta_x = \theta_y$
$H_A : \theta_x \neq \theta_y$

we are testing the hypothesis that the rate of right handedness = rate of tilting head to the right when kissing, i.e. $\theta_x = \theta_y$

using pnorm to do a 2-tailed test,

$p \approx 3.543 \times 10^{-7}$

this p-val suggests that we should reject the null hypothesis, and that the rates are most likely not the same for handedness and head tilting

Complete Code:

```
data = read.csv("fuel.ass2.2024.csv")
data.AWD <- subset(data, Type == "A")
data.PWD <- subset(data, Type == "P")

#Question 1.1

sampleMeanAWD <- mean(data.AWD$FA)
sampleMeanAWD

qt(p = 1 - 0.05/2, df = 19)

data.AWD$varianceCalc <- (data.AWD$FA - sampleMeanAWD) ^ 2

sampleVarianceAWD <- (1 / 19) * sum(data.AWD$varianceCalc)
sampleVarianceAWD

#Question 1.2

sampleMeanPWD <- mean(data.PWD$FA)
sampleMeanPWD

data.PWD$varianceCalc <- (data.PWD$FA - sampleMeanPWD) ^ 2

sampleVariancePWD <- (1 / 19) * sum(data.PWD$varianceCalc)
sampleVariancePWD

#Question 1.3
pnorm(1.244)

#Question 2.1
library(ggplot2)
e <- exp(1)

coinToss <- matrix(rep(1:20), nrow=20)

coinToss <- as.data.frame(coinToss)

colnames(coinToss) = "Y"

coinToss$phi0 <- (((e ^ 0 + 1)^(-(coinToss$Y) - 1)) * ((e) ^ (coinToss$Y * 0)))
coinToss$phi1 <- (((e ^ 1 + 1)^(-(coinToss$Y) - 1)) * ((e) ^ (coinToss$Y * 1)))
coinToss$phi2 <- (((e ^ 2 + 1)^(-(coinToss$Y) - 1)) * ((e) ^ (coinToss$Y * 2)))

ggplot(coinToss) +
  geom_line(aes(x = coinToss$Y, y = coinToss$phi0, color = "log Odds of failure
= 0")) +
  geom_line(aes(x = coinToss$Y, y = coinToss$phi1, color = "log Odds of failure
= 1")) +
  geom_line(aes(x = coinToss$Y, y = coinToss$phi2, color = "log Odds of failure
= 2")) +
  labs(title = "Geometric Distribution of Weighted Coin Tosses",
       x = "Number of Tosses before Landing Tails",
       y = "Probability",
```

```
       color = "legend")

#Question 3.2
2 * pnorm(-abs(1.244))

#Question 3.3
dataTilt <- c(rep(1, 168), rep(0, 72))
t.test(dataTilt)

#Question 3.4
2 * pnorm(-abs(-5.092))
```