STAT 260 R Assignment 3

Rodriguez Castro Raul V01030827

PART 1

```
fps_str = "97 135 114 111 119 118 114 114 97 135
96 113 128 115 117 115 103 106 131 123
129 99 119 113 93 116 115 104"
fps_vector = as.numeric(scan(text=fps_str, what=" "))
fps_vector
```

```
##  [1]  97 135 114 111 119 118 114 114  97 135  96 113 128 115 117 115 103 106 131
## [20] 123 129  99 119 113  93 116 115 104
```

a)

- Using the correct notation, define the parameter(s) of interest in this study. $\mu =$ The average fps at which the "flagship game" is able to run on the Ybox.

- State the null and alternative hypotheses in terms of the parameter(s). I went with a left side tailed test instead two-tailed test because, if the game runs at equal to or greater than 120 fps it's good for the consumer and supports the developer's claim. So we only want to test if it runs at lower fps since that would impact the consumer and developer's claim. $H_0 \geq 120$. $H_1 < 120$. • Copy and paste the appropriate R command and output for the hypothesis test into your document.

```
mu = mean(fps_vector)
t.test(fps_vector, mu = 120, alternative = 'less')
```

```
##
##  One Sample t-test
##
## data:  fps_vector
## t = -2.7914, df = 27, p-value = 0.004759
## alternative hypothesis: true mean is less than 120
## 95 percent confidence interval:
##      -Inf 117.6194
## sample estimates:
## mean of x
##  113.8929
```

With our calculations so far, it appears there is evidence to suggest the mean fps of the game on the YBox is less than 120.

b)

Value of test statistic: $t_{obs} = -2.7914$ Degrees of Freedom: 27 P-value: 0.004759 Strength of evidence: Since the p-value is really small $p - value \leq 0.01$, there is really strong evidence against the H_0 hypothesis. Which means we reject H_0 in favour of H_1.

c)

In plain English, there is sufficient evidence to suggest the YBox runs the flagship game at an average fps which is less than 120.

d)

```r
t.test(fps_vector, conf.level =  0.99)
```

```
##
##  One Sample t-test
##
## data:  fps_vector
## t = 52.057, df = 27, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  107.8311 119.9546
## sample estimates:
## mean of x
##  113.8929
```

e)

It is reasonable to assume the average fps is 110 as it is within our condifence interval. Any value $x := 107.8311 \leq x \leq 119.9546$

PART 2

```r
springfest = 78/116
autumfest = 110/182
```

a)

- Using the correct notation, define the parameter(s) of interest in this study. p_1: The local population proportion in favour of Springfest.
- State the null and alternative hypotheses in terms of the parameter(s). $H_0 : p_1 \leq 0.60$ $H_1 : p_1 > 0.60$ • Copy and paste the appropriate R command and output for the hypothesis test into your document.

```
binom.test(x = 78, n = 116, p = 0.60, alternative = 'greater')
```

```
##
##  Exact binomial test
##
## data:  78 and 116
## number of successes = 78, number of trials = 116, p-value = 0.06593
## alternative hypothesis: true probability of success is greater than 0.6
## 95 percent confidence interval:
##  0.5936533 1.0000000
## sample estimates:
## probability of success
##              0.6724138
```

There seems to be evidence to suggest that the proportion of voters in favour of Springfest exceeds 60%.

b)

Test Statistic: $\hat{p} = \frac{78}{116}$ P-value: 0.06593 Conclusion: given $\alpha = 0.05$ and the p-value $> \alpha$ we fail to reject the null hypothesis.

c)

There are two cases. case 1: Given that the p-value is less than 0.10, there is moderate evidence to suggest more than 60% of the local population is in favour of Springfest.

case 2: If we take into account the previous value given: $\alpha = 0.05$, then the evidence suggests at most 60% of the local population is in favour Springfest

d)

p_2: The local population proprotion in favour of Autumnfest.

```
  binom.test(x = 110,n = 182, conf.level = 0.90)
```

```
##
##  Exact binomial test
##
## data:  110 and 182
## number of successes = 110, number of trials = 182, p-value = 0.005943
## alternative hypothesis: true probability of success is not equal to 0.5
## 90 percent confidence interval:
##  0.5410686 0.6651463
## sample estimates:
## probability of success
##              0.6043956
```

90% Confidence interval = [0.5410686, 0.6651463]

e)

I am going to assume "Expansion B" is referring to the latter Autumnfest event. In this case saying 50% of voters are in favour of it is reasonable as it is within the range previously stated in part d) [0.5410686, 0.6651463].

PART 3

a)

```
 scottish_str = "5 5 10 12 20 10 2 15 15 20 20 15 15 15 20 16 18 20 15 25 60 5 10 15 10 10 10 20 3"

english_str = "4 24 19 2 15 5 16 10 24 26 21 28 7 22 20 27 20 27 11 1 29 27 19 10 17 23 13 39 6 24 11 

 daily_cigs_scottish = as.numeric(scan(text=scottish_str, what=" "))
 daily_cigs_english = as.numeric(scan(text=english_str, what=" "))

sd_scottish = sd(daily_cigs_scottish)
sd_english = sd(daily_cigs_english)

sd_scottish
```

```
## [1] 10.4488
```

```
sd_english
```

```
## [1] 9.113547
```

b)

```
sd_scottish / sd_english
```

```
## [1] 1.146513
```

Since $\frac{\sigma_{big}}{\sigma_{small}} \leq 1.4$ we can assume $\sigma_1^2 = \sigma_2^2$.

c)

Since the $n < 30$ for the number of samples from the Scottish samples and $\sigma_1^2 = \sigma_2^2$. We can perform a hypothesis test using a pooled T-distribution.

- Using the correct notation, define the parameter(s) of interest in this study. μ_1: The daily cigarettes consumed by Scottish smokers. μ_2: The daily cigarettes consumed by English smokers.

- State the null and alternative hypotheses in terms of the parameter(s). $H_0 : \mu_1 - \mu_2 = 0$ $H_1 : \mu_1 - \mu_2 \neq 0$
- Copy and paste the appropriate R command and output for the hypothesis test into your document.

```r
t.test(daily_cigs_scottish, daily_cigs_english, mu = 0, alternative='two.sided', var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  daily_cigs_scottish and daily_cigs_english
## t = -0.90425, df = 75, p-value = 0.3688
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.562534  2.464832
## sample estimates:
## mean of x mean of y
##  15.03448  17.08333
```

d)

Given the computed p-value $= 0.3688$, at a significance level of $\alpha = 0.10$, we fail to reject H_0.

e)

This means that there is not enough evidence to suggest the daily cigarettes consumed by English and Scottish smokers is different.

f)

```r
t.test(daily_cigs_scottish, daily_cigs_english, conf.level = 0.80, var.equal = TRUE )
```

```
##
##  Two Sample t-test
##
## data:  daily_cigs_scottish and daily_cigs_english
## t = -0.90425, df = 75, p-value = 0.3688
## alternative hypothesis: true difference in means is not equal to 0
## 80 percent confidence interval:
##  -4.9783831  0.8806819
## sample estimates:
## mean of x mean of y
##  15.03448  17.08333
```

80% Confidence interval: [-4.9783831 0.8806819]

PART 4

a)

```r
mean_lifespan = 6.2
simHeadset = rexp(500, rate = 1 / mean_lifespan)

hist(simHeadset,
     main = "Histogram of Simulated Headset Lifespans",
     xlab = "Lifespan (years)",
     ylab = "Frequency",
     col = "lightblue",
     border = "black")

mean_simHeadset = mean(simHeadset)
print("simHeadset mean: ")
mean_simHeadset
```

b)

```r
  simMean = numeric(300)
  mean_lifespan = 6.2

  for (i in 1:300) {
  simMean[i] = mean(rexp(500, rate = 1/mean_lifespan))
  }

  hist(simMean,
     main = "Histogram of Simulated Sample Mean Lifespans",
     xlab = "Sample Mean Lifespan (years)",
     ylab = "Frequency",
     col = "lightgreen",
     border = "black")

  mean_simMean = mean(simMean)
  print("simMean ")
  mean_simMean
```

c)

The histogram in (a) is right-skewed (asymmetric), which is typical for an exponential distribution. The histogram in (b), however, is approximately symmetric and resembles a normal distribution. This difference is explained by the Central Limit Theorem (CLT), which states that the sampling distribution of the sample mean will tend to be normal, regardless of the original distribution of the data, provided the sample size is sufficiently large.

d)

The mean of the simulated data in (a) should be close to the theoretical mean of the exponential distribution, which is 6.2 years. The mean of the sample means in (b) should also be close to 6.2 years. Once again, referring back to the Central Limit Theorem, as the number of samples increases, the sample approaches a normal distribution centered around the population mean. Therefore, we expect the means (a) and (b) to be similar. As both are derived from the same underlying distribution.