Math 338 Midterm 1 Lab Portion

October 9/10, 2019

You have 50 minutes to complete this exam and an extra 10 minutes to Knit this file (click the button that says Knit) and upload the resulting pdf file to Titanium.

You may refer to your notes, your textbook, and any pre-existing online reference (eBook/Sapling, R/Rguroo help, anything on Titanium). You may search for help online, but you must cite any source found through the search. You may ask Dr. Wynne to clarify what a question is asking for. You may not ask other people for help or use any other resources. For full credit, show all work except for final numerical calculations (which can be done using a scientific/graphing calculator or R).

Problem 1

Maggie, a soccer player, can make 75% of her penalty kicks. At the end of practice, she attempts 10 penalty kicks. Assume that each penalty kick attempt is independent of the other attempts.

What is the probability that she makes exactly 7 kicks? [1.5 pts]

```r
X <- 7
n <- 10
p <- 0.75
dbinom(X, n, p)
```

```
## [1] 0.2502823
```

Given she will make **exactly** seven kicks, she will have around a 25% chance of doing so.

What is the probability that she makes more than 7 kicks? [2 pts]

```r
X <- 7
n <- 10
p <- 0.75
pbinom(X, n, p, lower.tail = FALSE) + dbinom(X, n, p)
```

```
## [1] 0.7758751
```

The likelihood of her **getting more** than seven kicks will be approximately 78%

Problem 2

Sagawa, et al. (2019) investigated the effect of a drug called DHL-HisZnNa on preventing alopecia (hair loss) due to chemotherapy in breast cancer patients. Read the following excerpt from the paper:

"…our null hypothesis was that the incidence of grade 2 alopecia [severe hair loss] in our patients would be 95%. We also hypothesized that the study drug would reduce the incidence of grade 2 alopecia by 10% [i.e., to 85%]." [text in brackets added by me for clarification]

Ultimately, Sagawa and colleagues enrolled 100 participants undergoing chemotherapy for breast cancer.

Using a standard Type I Error rate of 5%, what is the power of a binomial test to detect their specific alternative hypothesis that the drug would reduce the incidence to 85%? [4 pts]

```
alpha <- 0.05
n <- 100
p0 <- 0.05
p1 <- 0.15

critical_region <- qbinom(alpha, size = n, prob = p0, lower.tail = FALSE)

power <- pbinom(critical_region, size = n, prob = p1, lower.tail = FALSE) + dbinom(critical_region, siz
# power assignment is cut off so you need to view the code in the source file
print(power)
```

```
## [1] 0.9725243
```

The power of dropping the hair loss rate to 85% is around 97%.

Do you believe 100 subjects would be a large enough sample size for the researchers to detect the hypothesized decrease in the incidence of grade 2 alopecia? Why or why not? [1.5 pts]

Given the population size and power computed above, we can make the claim that a *bigger* population is needed. Our power is not even in the ball park of the 80% minimum threshold.

If the researchers only managed to enroll 50 participants, would their study have more power, less power, or the same amount of power? Justify your answer. (Hint: you can either re-compute power with the new sample size, use mathematical reasoning, or do a combination of the two) [2 pts]

```
alpha <- 0.05
n <- 50
p0 <- 0.05
p1 <- 0.15

critical_region <- qbinom(alpha, size = n, prob = p0, lower.tail = FALSE)

power <- pbinom(critical_region, size = n, prob = p1, lower.tail = FALSE) + dbinom(critical_region, size
# power assignment is cut off so you need to view the code in the source file
print(power)
```

```
## [1] 0.8878948
```

As we stated above, the sample needs to bigger, so by decreasing the sample size, we will subsequently drop the power. This makes the results **less** reliable.

Problem 3

Maimaran and Salant (2019) investigated whether limiting the amount of carrots a child could eat affects how the child feels about the taste of the carrots. The carrots.csv file on Titanium contains sample data from 49 children who were given a plate of carrots to eat. There are two variables:

- Condition: whether the researchers told the child that the carrots on the plate were the only carrots left. Limit means that the researchers told the child; Control means they didn't.
- Rating: whether the child rated the carrots on the plate as Yummy, OK, or Yucky.

```
data <- read.csv("~/Downloads/carrots.csv")
```

Produce a bar graph to visualize the sample data. Full credit will be given for a properly labeled graph that correctly displays the information. [2 pts]

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
data <- read.csv("~/Downloads/carrots.csv")
summary(apply(data, 2, as.factor))
```

```
##    Condition    Rating
##  Control:24   OK   : 8
##  Limit  :25   Yucky: 1
##               Yummy:40
```

```r
nrow(data)
```

```
## [1] 49
```

```r
dim(data)[1]
```

```
## [1] 49
```

```r
carrot_plot <- qplot(data$Condition, geom= "bar")
carrot_plot_labeled <- carrot_plot + labs(title= "", x = "Condition", y = "Rating")
# please see attached RPlots.pdf file, knitting process complained
```

Perform an appropriate statistical hypothesis test to determine whether simply telling children they can't have more carrots affects how they feel about the taste of carrots. Use either Fisher's significance testing approach or the null hypothesis significance test approach (as appropriate) and assume a significance level of 0.05. [4 pts]

Full credit will be given for:

- Choosing an appropriate test
- Writing an appropriate null hypothesis for that test in the real-world context and, if appropriate, the corresponding alternative hypothesis
- Checking any necessary assumptions
- Performing the test correctly in software and including the output
- Correctly identifying the value of the test statistic and the p-value from the output
- Making an appropriate conclusion in the real-world context

Hypothesis Listings

- Null: children are equally likely to not be affected when told they cannot have a carrot and their opinion on carrot taste.
- Alternative: children are **not** equally likely to be affected when told they cannot have a carrot and their opinion on carrot taste.

Assumption Checking

Assumptions made clear in the null and alternate hypothesis.

```r
X <- 41
n <- 49
p_null <- 0.5
alpha <- 0.05
binom.test(X, n, p, alternative = "g")
```

```
##
##  Exact binomial test
##
## data:  X and n
## number of successes = 41, number of trials = 49, p-value = 0.1046
## alternative hypothesis: true probability of success is greater than 0.75
## 95 percent confidence interval:
##  0.7246484 1.0000000
## sample estimates:
## probability of success
##              0.8367347
```

Replace this text with the values of your test statistic and p-value.

Replace this text with an appropriate decision and a conclusion in the real-world context.