Exam 3 – Practice

Question 1 (Monday Breakups Redux)

Facebook is a social networking website. One piece of data that members of Facebook often report is their relationship status: single, in a relationship, married, it's complicated, etc.

With the help of Lee Byron of Facebook, David McCandless examined changes in peoples' relationship status, in particular, breakups. A plot of the results showed that there were repeated peaks on Mondays. Based on this initial examination of data, McCandless speculated that breakups are reported at higher frequency on Mondays.

To test this research hypothesis, McCandless collected a random sample of 75 breakups reported on Facebook within the last year. Of these sampled breakups, 20 occurred on a Monday.

Research Question: Are people more likely to break up on Mondays than other days of the week?

Part A: Describe the parameter in context and provide the appropriate symbol.

The proportion of breakups reported on Monday = p

Part B: Is this an observational study or an experiment? How does this affect our conclusions? Explain.

Observational Study. We cannot make causal conclusions.

Part C: State the null and alternate hypothesis in words **and** using the notation we've seen in class.

```
# H0: p = 1/7
# HA: p > 1/7
```

Part D: Are the conditions met to conduct a hypothesis test for this scenario? State the conditions and whether they are met or not met (show work or no credit given)

```
# Yes. We have a random sample.
# n*p0 = 75 / 7 > 10 [met]
# n*p0 = 75 * (6 / 7) > 10 [met]
```

Part E: Find the value of the test-statistic for this hypothesis test.

```
Z = (20/75 - 1/7) / sqrt((1/7) * (6/7) / 75)
Z
```

```
## [1] 3.064129
```

Part F: The p-value for this test is 0.0010. Write up a conclusion to answer the research question using the strength of evidence approach.

With a p-value of 0.0010, there is overwhelming evidence that breakups are reported more frequently on Mondays.

Question 2 (Sleep Deprivation Redux)

A CDC report on sleep deprivation rates shows that the proportion of California residents who reported insufficient rest or sleep during each of the preceding 30 days is 8.0%, while this proportion is 8.8% for Oregon residents. These data are based on simple random samples of 11,545 California and 4,691 Oregon residents.

Goal: Conduct a hypothesis test to determine if these data provide strong evidence that the rate of sleep deprivation is different for the two states.

Part A: Describe the parameter in context and provide the appropriate symbols.

The parameter is the difference in proportion of residents who report insufficient sleep for California minus Oregon (or other way around) = $p_C - p_O$

Part B: State the null and alternate hypothesis in words **and** using the notation we've seen in class.

H0: the proportion of residents who report insufficient sleep is the same for California and Oregon, $p_C = p_O$

HA: the proportion of residents who report insufficient sleep is *not* the same for California and Oregon, $p_C \neq p_O$

Part C: Compute $\widehat{p}_{pool}$.

```
(11545*.8 + 4691*.88) / (11545 + 4691)
```

```
## [1] 0.8231141
```

Part D: Compute the test-statistic Z.

```
Z = (.088 - .08) / sqrt(.0823*(1-.0823)*(1/11545 + 1/4691))
Z
```

```
## [1] 1.681241
```

Part E: Suppose the p-value for this test is ≈ 0.00001. Write a conclusion to this question using 'strength of evidence'.

There is overwhelming evidence the rate of sleep deprivation is different for the two states (use the research question wording if able).

Question 3 (Lead Exposure)

Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 micro-grams per liter (micro-g/L) and a SD of 37.74 micro-g/L; a previous study of individuals from a nearby suburb, with no history of exposure, found an average blood level concentration of 35 micro-g/L. (Mortada et al. 2000)

Part A: Write down the hypotheses that would be appropriate for testing if the police officers appear to have been exposed to a different concentration of lead.

H0: the officers have been exposed to the same concentration of lead $\rightarrow$ the mean lead concentration of these officers and the nearby suburb are the same, $\mu_{officer} = 35$

HA: the mean lead concentration of these officers and the nearby suburb are the same, $\mu_{officer} = 35$.

Part B: Explicitly state and check all conditions necessary for inference on these data (show work or no credit given)

```
# Random sample (not met)
# Sample size 52 > 30 (normality condition met)
```

Part C: Compute the test-statistic for the hypothesis in Part A (show work for credit).

```
T = (124.32 - 35) / (37.74 / sqrt(52))
T
```

```
## [1] 17.06666
```

Part D: What is the degrees of freedom corresponding to this test-stat?

```
# df = n-1 = 52-1 = 51
```

Part D: The p-value for this test is $4.96 * 10^{-23}$. Write a conclusion to the researcher's question in context at a signifance level of $\alpha = .05$.

There is overwhelming evidence that the mean lead concentration in officers' blood is different than individuals from the suburb, which is to say there is overwhelming evidence the officers were exposed to a different lead concentration.

Question 4 (Fuel Economy)

Each year the US Environmental Protection Agency (EPA) releases fuel economy data on cars manufactured in that year. Below are summary statistics on fuel efficiency (in miles/gallon) from random samples of cars with manual and automatic transmissions manufactured in 2021.

Research Question: Do these data provide strong evidence of a difference between the average fuel efficiency of cars with manual and automatic transmissions in terms of their average highway mileage?

MPG	Mean	Std. Dev.	n
Automatic	28.7	3.90	25
Manual	30.1	5.13	25

Part A: What type of hypothesis test do we want to conduct for this scenario? Circle one of the following: (proportion / diff in proportions / mean / difference in means)

difference in means

Part B: Check the conditions to conduct a hypothesis test for this research question. (show work or no credit given)

The random sample condition is met. The sample size conditions are not met, as both are less than 30 and there is no mention of the populations being Normal.

Part C: Calculate the test-statistic and state the corresponding df.

```r
T = (30.9 - 28.7) / sqrt(3.9^2 / 25 + 5.13^2 / 25)
T
```

```
## [1] 1.706978
```

```r
df = 25-1
df
```

```
## [1] 24
```

Part D: The p-value corresponding to this test is 0.0504. Explain how your conclusion to this research question would differ if you used the 'strength of evidence' approach or by using the 'decision making' approach with a signifance level of $\alpha = .05$.

We would get borderline/weak evidence that the fuel economy of Automatic vehicles is different from Manual using strength of evidence.

If we use the decision making approach, then the p-value is not significant and so there is no statistical difference between the fuel economies of the two vehicle types.

Question 5 (Cocaine Addiction)

Cocaine addicts have been reported to have a significant depletion of stimulating neurotransmitters and thus continue to use cocaine to avoid feelings of depression and anxiety. A 3-year study with 72 chronic cocaine users compared an antidepressant drug called desipramine with lithium and a placebo (lithium is the standard treatment for cocaine addiction). One third of the subjects were randomly assigned to each treatment group with the following results.

Observed	Relapse	No Relapse
Desipramine	10	14
Lithium	18	6
Placebo	20	4

Expected	Relapse	No Relapse
Desipramine	16	8
Lithium	16	8
Placebo	16	8

Part A: Describe the null hypothesis of this study.

The null hypothesis is that the type of medication and whether someone relapses are independent (not influencing) each other. In other words, the medication has no effect on relapses.

Part B: What type of χ^2 test would we use to answer this, Goodness of fit or independence?

Independence (there are 2 variables)

Part C: Compute the χ^2 test statistic and the degrees of freedom. (show work or no credit given).

```r
# Chi2 test statistic
(10-16)^2/16 + (18-16)^2/16 + (20-16)^2/16 + (14-8)^2/8 + (6-8)^2/8 + (4-8)^2/8
```

```
## [1] 10.5
```

```r
# df
(3-1)*(2-1)
```

```
## [1] 2
```

Part D: Suppose we were testing at a signifance level of $\alpha = .01$ and we got a p-value of 0.023. Write a conclusion to this study. (1-2 sentences)

With a p-value of 0.023 and a signifance level of $\alpha = .01$, we do not reject the null hypothesis that the treatment and relapsing are independent. There is no significant association between them at the $\alpha = .01$ level.

Part E: Write a conclusion to this study using the strength-of-evidence approach. (1-2 sentences)

With a p-value of 0.023, there is moderate evidence of an association between the treatment and relapses. There is moderate evidence that the treatment affects whether someone relapses.