Practice Exam #2 (Sta-209, F25)

Directions

- Answer each question using *no more than specified number of sentences* and not attempt to avoid these guidelines by using run-on sentences. Answers that are unnecessarily verbose may result in point loss.
- Do not include superfluous information in your answers, you may be penalized if you make an inaccurate statement even if you go on to provide a correct answer. Your answers should be clear, concise, and include only what is needed to answer the question that was asked.

Formula Sheet

Standard Errors:

Statistic	Standard Error	Conditions
$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$	$np \geq 10$ and $n(1-p) \geq 10$
$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$	normal population or $n \geq 30$
$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$n_i p_i \geq 10$ and $n_i(1-p_i) \geq 10$ for $i \in \{1,2\}$
$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	normal populations or $n_1 \geq 30$ and $n_2 \geq 30$

Definitions:

- Odds: frequency of an event divided by the frequency of the complimentary event (ie: A/B)
- Risk: proportion of the time that an event occurs (ie: A/(A+B))
- Type I error: Rejecting H_0 when it is true
- Type II error: Not rejecting H_0 when it is false

Confidence Interval Calibration Values:

Distribution	c for 95%	c for 99%
Normal	1.96	2.58
$t(df = 5)$	2.57	4.03
$t(df = 25)$	2.06	2.79

Section 1 - True/False

Directions: Clearly indicate whether each of the following statements is true or false. You do not need to explain your reasoning and you will not receive a better score for doing so.

1. The main difference between an *experimental study* and an *observational study* is whether the researchers manipulated the explanatory variable.
2. In a retrospective, case-control study investigating oral cancer the researchers obtain their data by recruiting a sample of participants with oral cancer and an entirely separate sample of participants without oral cancer.
3. For categorical outcomes, odds ratios and relative risks are generally considered a more useful measure of effect size than differences in proportions when the sample size is large.
4. Consider a random sample $n = 10$ games played by an NFL team where the sample odds of the team winning are 4. These sample odds suggest that the team lost exactly 2 games in the sample.
5. The null hypothesis in a Chi-squared Goodness Fit test is always set up to suggest each category as being equally likely within the population.
6. When finding the p-value in a Chi-squared test we only consider the area of the null distribution corresponding to values greater than the observed test statistic, even when we are interested in differences that could go in either direction.
7. In one-way ANOVA, rejecting the null hypothesis suggests that the observed data do not follow a Normal distribution.
8. In one-way ANOVA, the fit of the null and alternative models are measured using the sum of their squared residuals.
9. Lowering the threshold for statistical significance from $\alpha = 0.05$ to $\alpha = 0.01$ will increase the likelihood of a hypothesis test producing a Type II error.
10. When performing many hypothesis tests as part of a single experiment, family-wise Type I error rate control procedures, such as the Bonferroni adjustment, typically result in more Type II errors than false discovery rate control procedures.
11. Suppose a scientifically rigorous study reports a 95% confidence interval estimate for the mean cholesterol level of US adults as (202.4, 225.6). This interval suggests 95% of the US adult population has a cholesterol level between 202.4 and 225.6.
12. Bootstrapping is a way to increase the sample size of a study by resampling the observed cases with replacement.
13. Consider a 95% confidence interval estimate for ρ, the correlation between two variables in a population. If a sample of $n = 50$ cases produces an interval estimate of (0.01, 0.26) then we'd expect a hypothesis test of $H_0 : \rho = 0$ to produce a p-value less than 0.05.

Section 2 - Conceptual Questions

Directions: Answer each question in about 3-sentences. Do not include unnecessary details, as you will be penalized for any inaccurate statements, regardless of whether they are relevant or not. Aim to clearly answer the question that was posed, not to demonstrate your knowledge of related topics.

1. In your own words, explain the concept of the "confidence level" in confidence interval estimation. That is, what does the confidence level tell us about an interval? What inherent problem in interval estimation does the "confidence level" address?

2. Suppose you and one of your friends are trying to estimate the proportion of Grinnell students who are double majors. You each take a representative sample of students and construct a 95% confidence using a statistically valid procedure. However, you sampled $n = 60$ students but your friend only sampled $n = 30$ students. Whose interval is more likely to contain the true proportion? Explain your reasoning.

3. Consider a Chi-squared test of independence where we are interested in determining whether there is an association between three different exposure groups and three different health outcomes. In your own words, explain how you'd go about finding the expected counts within each group under the null hypothesis of this test.

4. Log-transformations are often used on the outcome variable prior to performing one-way ANOVA. In your own words, explain the purpose of a log-transformation in this context. That is, what are the reasons why someone might log-transform their outcome variable prior to performing an ANOVA test?

Section 3 - Application #1

Researchers in Florida collected data from a random sample of $n = 26$ lakes across the state, recording the median mercury level of large mouth bass (ppm) in each lake. The average mercury level in the sampled lakes was 0.55 ppm, and the standard deviation was 0.53 ppm.

Part A: Consider the calculation of a 99% confidence interval estimate for the average mercury level in all lakes in Florida using this sample. Provide both components of the margin of error for this interval.

Part B: Provide the 99% confidence interval estimate for the average mercury level in all lakes in Florida using the margin of error you calculated in Part A.

Part C: The EPA has a maximum acceptable level of 1.0 ppm for the mercury concentration in predatory and bottom-dwelling species of fish, with levels lower than this being deemed "safe". Does the confidence interval you calculated in Part B suggest that it is plausible that lakes in Florida, on average, have unsafe levels of mercury according to the EPA?

Part D: Assuming everything else remains unchanged, which of the following will *decrease* the width of the interval you calculated in Part B.

- i) An additional 20 lakes are sampled, bringing the total sample size to $n = 46$
- ii) You change the confidence level to 95%
- iii) Lakes are sampled until the interval's margin of error reaches 0.1

Part E: It is possible that *sampling bias* causes the confidence interval you calculated in Part B to fail to contain the true mercury level of all lakes in Florida? Briefly explain your reasoning.

Section 4 - Application #2

We've previously analyzed the "flicker" data set, where researchers measured the critical flicker frequency (the variable `Flicker`), which is the highest frequency where a person can differentiate between a solid and flickering light source, for subjects with three different eye colors: blue, brown, and green (the variable `Color`). The researchers hypothesized that eye color was associated with critical flicker frequency.

Below are some descriptive statistics from this study that you should use throughout this question:

```
## Overall summary
flicker %>%
  summarize(mean_flicker = mean(Flicker),
            median_flicker = median(Flicker),
            sd_flicker = sd(Flicker),
            n = n())
```

```
##   mean_flicker median_flicker sd_flicker  n
## 1     26.75263           26.8   1.845526 19
```

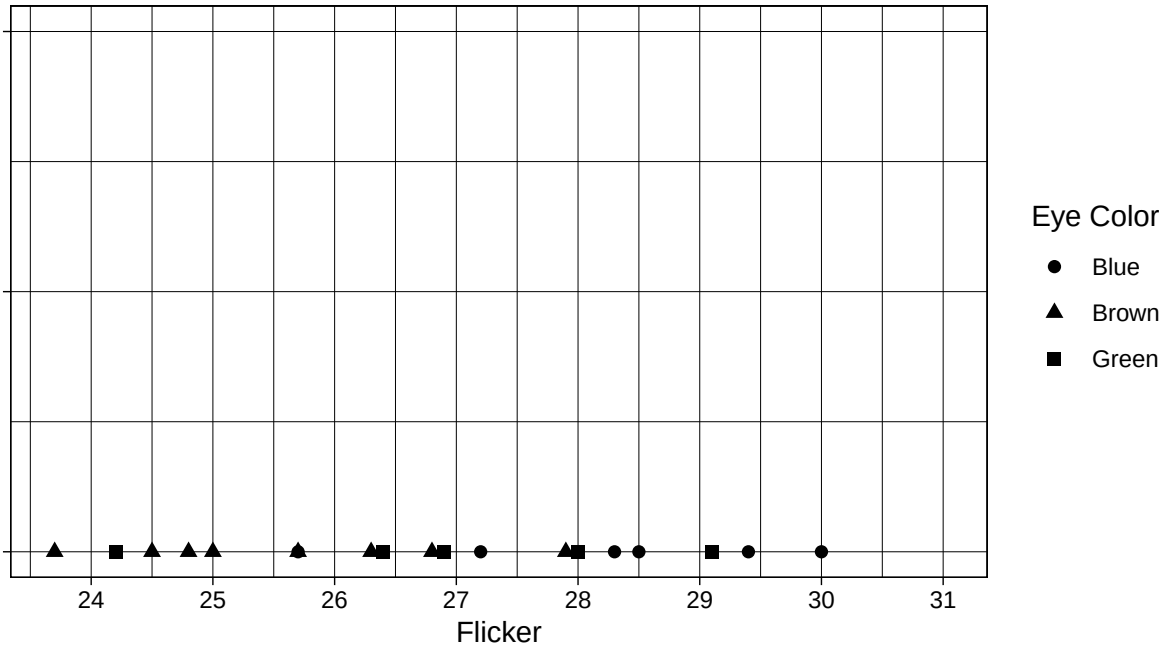
```
## Summary by eye color
flicker %>%
  group_by(Color) %>%
```

```
  summarize(mean_flicker = mean(Flicker),
            median_flicker = median(Flicker),
            sd_flicker = sd(Flicker),
            n = n())
```

```
## # A tibble: 3 x 5
##   Color mean_flicker median_flicker sd_flicker     n
##   <chr>        <dbl>          <dbl>      <dbl> <int>
## 1 Blue          28.2           28.4       1.53     6
## 2 Brown         25.6           25.4       1.37     8
## 3 Green         26.9           26.9       1.84     5
```

Part A: One-way ANOVA can be expressed as a comparison of statistical models. With this in mind, briefly describe the *null model* using either words or statistical symbols (or both).

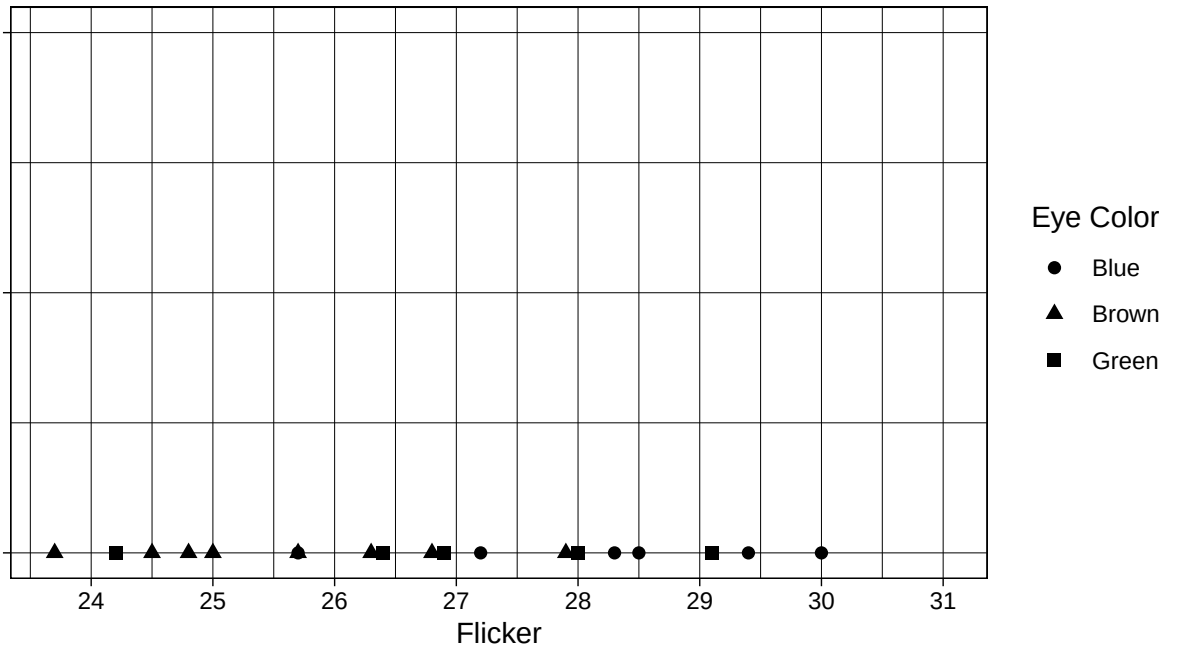
Part B: The *null model* in one-way ANOVA involves using a certain probability distribution to model the data. Sketch this distribution on the graph below. Be careful where you place the center of the distribution.



Part C: One-way ANOVA compares models using sums of squares, which involve the squared deviations between observed and predicted values (ie: squared residuals). More specifically, $SS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. With this is in mind, what is the *residual* for the observation with the highest observed critical flicker frequency *under the null model*. Use the data shown on the x-axis of the graph given above.

Part D: Now briefly describe the *alternative model* involved in one-way ANOVA using either words or statistical symbols (or both).

Part E: Similar to Part B, sketch the *alternative* model involved in one-way ANOVA on the graph given below:



Part F: Similar to Part C, what is the *residual* for the observation with the highest observed critical flicker frequency *under the alternative model*.

Part G: The sum of squared residuals for the null model, SST, in this application is 61.31. Do you expect the sum of squared residuals for the alternative model to be larger, smaller, or approximately equal to this value? State "larger", "smaller", or "approximately equal" and briefly explain your reasoning.

Part H: The F-statistic and p-value for the one-way ANOVA that analyzes the relationship between `Color` and `Flicker` are $F = 4.8$ and $p = 0.023$ respectively. Based upon these results, provide a brief conclusion in regard to the researcher's hypothesis described in the introduction of Question 2. You may assume the conditions of the test have been met.

Part I: Shown below are the results of post-hoc pairwise testing for the ANOVA model described throughout this application. Briefly explain how these results allow you to make a more precise conclusion than the one you made in Part H.

```
model = aov(Flicker ~ Color, data = flicker)
TukeyHSD(model)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Flicker ~ Color, data = flicker)
##
## $Color
##                  diff        lwr        upr     p adj
## Brown-Blue  -2.595833 -4.7621317 -0.4295349 0.0181490
## Green-Blue  -1.263333 -3.6922387  1.1655720 0.3934460
```

```
## Green-Brown  1.332500 -0.9542388  3.6192388 0.3155339
```