STAT 350 Final Exam SPRING 2025

V1

Name: ______________________________________

PUID ______________________________

Instructor (circle one): Heekyung Ahn Evidence Matangi Timothy Reese Halin Shin

Class Start Time: ○ 10:30 AM ○ 11:30 AM ○ 12:30 PM ○ 1:30 PM ○ 2:30 PM ○ 3:30 PM ○ Online

As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do.
Accountable together - we are Purdue.

Instructions:

1. **IMPORTANT** Please write your **name** and **PUID clearly** on every **odd page.**
2. **Write your work in the box. Do not run over into the next question space.**
3. The only materials that you are allowed during the exam are your **scientific calculator**, **writing utensils, erasers, your crib sheets (2x)**, and **your picture ID**. Colored scratch paper will be provided if you need more room for your answers. Please write your name at the top of that paper also.
4. Keep your bag closed and cellphone stored away securely at all times during the exam.
5. If you share your calculator without permission or have a cell phone at your desk, you will get a **zero** on the exam. Do not take out your cell phone until you are next in line to submit your exam.
6. The exam is **120 minutes** long. If you need a bathroom break, raise your hand to request approval. Breaks will only be permitted when an escort is available, and you must confirm that no cell phones or electronic devices are in your possession before leaving the room.
7. **For free response questions you must show ALL your work to obtain full credit.** An answer without showing any work may result in **zero** credit. If your work is not readable, it will be marked wrong. Remember that work has to be shown for all numbers that are not provided in the problem or no credit will be given for them. All explanations must be in complete English sentences to receive full credit.
8. All numeric answers should have **four decimal places** unless stated otherwise.
9. After you complete the exam, please turn in your exam as well as your table and any scrap paper that you used. Please be prepared to **show your Purdue picture ID**. You will need to **sign a sheet** indicating that you have turned in your exam.
10. You are expected to uphold the honor code of Purdue University. It is your responsibility to keep your work covered at all times. Anyone caught cheating on the exam will automatically fail the course and will be reported to the Office of the Dean of Students.
11. It is strictly prohibited to smuggle this exam outside. Your exam will be returned to you on Gradescope after it is graded.

Your exam is not valid without your signature below. This means that it won't be graded.

I attest here that I have read and followed the instructions above honestly while taking this exam and that the work submitted is my own, produced without assistance from books, other people (including other students in this class), notes other than my own crib sheet(s), or other aids. In addition, I agree that if I tell any other student in this class anything about the exam BEFORE they take it, I (and the student that I communicate the information to) will fail the course and be reported to the Office of the Dean of Students for Academic Dishonesty.

Signature of Student: ____

You may use this page as scratch paper.
The following is for your benefit only.

Question Number	Total Possible	Your points
Problem 1 (True/False) (2 points each)	12	
Problem 2 (Multiple Choice) (3 points each)	18	
Problem 3	21	
Problem 4	32	
Problem 5	40	
Problem 6	42	
Total	**150+15** (Extra Credit) = **165**	

1. **(12 points, 2 points each) True/False Questions.**

1.1. Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two random variables with a large covariance

$$\mathbf{COV}(\boldsymbol{X},\boldsymbol{Y}) = \mathbf{100{,}000},$$

and the individual standard deviation $\boldsymbol{\sigma_X}$ and $\boldsymbol{\sigma_Y}$ are unknown but finite.

Ⓣ or Ⓕ From this information, we can conclude that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are strongly correlated.

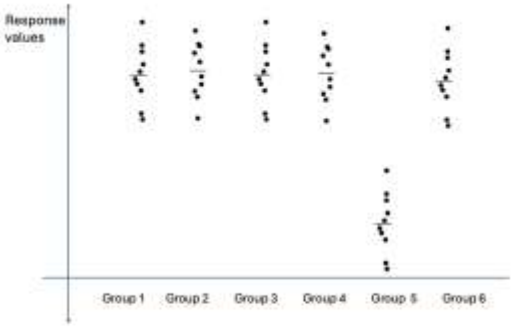
1.2. A Reddit user claims that when one encounters a Pokémon in the wild, there is a 2% chance that it is a legendary Pokémon. Assume that each encounter is independent, each has the same 2% chance, and each wild Pokémon encountered is either legendary or not legendary. Assuming the claim is true:

Ⓣ or Ⓕ the number of legendary Pokémon encountered is approximately normal if a sufficiently large number of wild Pokémon are observed.

1.3. Bayes' Theorem is often used for revising probabilities based on new evidence.

Ⓣ or Ⓕ Bayes' Theorem applies when the events of interest, $\boldsymbol{A_1}, \boldsymbol{A_2}, \dots, \boldsymbol{A_k}$ are mutually exclusive (disjoint), and the evidence event $\boldsymbol{B}$ has positive probability.

1.4. The plot below shows the response values of an experiment, organized by different treatment groups.



Ⓣ or Ⓕ According to the plot, an ANOVA F-test for this dataset will most likely result in a failure to reject $\boldsymbol{H_0}$ because most groups are shaped very similarly.

1.5. Let $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$and $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$ be the slope and intercept of the regression line computed from a dataset $(\boldsymbol{x_1},\boldsymbol{y_1}),(\boldsymbol{x_2},\boldsymbol{y_2}),\cdots,(\boldsymbol{x_n},\boldsymbol{y_n})$. Suppose a function $\boldsymbol{l}$ is defined as

$$\boldsymbol{l}(\boldsymbol{a},\boldsymbol{b}) = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \left(\boldsymbol{y_i} - (\boldsymbol{a} + \boldsymbol{b}\boldsymbol{x_i})\right)^{\mathbf{2}}.$$

Ⓣ or Ⓕ Then $\boldsymbol{l}$ attains the smallest value possible by plugging in $\widehat{\boldsymbol{\beta}}_{\mathbf{0}}$ for $\boldsymbol{a}$ and $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$ for $\boldsymbol{b}$.

1.6. Interpolation and extrapolation are terms used for making predictions with a regression model.

Ⓣ or Ⓕ Extrapolation typically yields safer, more reliable predictions than interpolation.

2. **(18 points, 3 points each) Multiple Choice Questions.** Indicate the correct answer by completely filling in the appropriate circle. If you indicate your answer by any other way, you may be marked incorrect. **For each question, there is only one correct option letter choice.**

2.1. Suppose $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{D}$ are non-empty events in the sampe sample space where $\boldsymbol{P}(\boldsymbol{C}) > \boldsymbol{P}(\boldsymbol{A} \cup \boldsymbol{B}) > \boldsymbol{P}(\boldsymbol{D}) > \mathbf{0.7}$. Which of the following statements is **TRUE**?

Ⓐ $\boldsymbol{D}$ must be a subset of $\boldsymbol{C}$.

Ⓑ $\boldsymbol{P}(\boldsymbol{A}' \cap \boldsymbol{B}') < \boldsymbol{P}(\boldsymbol{C}') < \mathbf{0.3}$ holds.

Ⓒ The two events $\boldsymbol{C}$ and $\boldsymbol{D}$ can be mutually exclusive.

Ⓓ If $\boldsymbol{A} \cap \boldsymbol{C} = \emptyset$, then $\boldsymbol{B} \cap \boldsymbol{C}$ must be a non-empty set.

Ⓔ The two events $\boldsymbol{A}$ and $\boldsymbol{B}$ are independent.

2.2. Which statement regarding the properties of common random variables is **FALSE**?

Ⓐ A **Poisson random variable** counts the number of events occurring in a fixed interval of time, area, or space.

Ⓑ A **Uniform random variable** assigns equal probability density across its entire support.

Ⓒ A **Binomial random variable** counts the number of independent trials required to achieve a specified number of successes.

Ⓓ An **Exponential random variable** measures the waiting time between consecutive independent events.

Ⓔ None of the statements listed above are false.

2.3. In an Amish community 🏡 🚜, an accountant wants to determine if the daily profit 💰 from selling groceries has increased in 2024 compared to 2023. However, due to a recent flood, the accountant was only able to obtain daily profit data for 55 days in 2023 and 82 days in 2024. The calculated test statistic follows the $\boldsymbol{t}$**-distribution** with $\mathbf{df} \approx \mathbf{117.25}$. Historically the distribution of daily profit prior to 2023 exhibited slight positive skewness, and current samples appear consistent with historical patterns. Which of the following statements in the accountant's report is **FALSE**?

Ⓐ Daily profit reports were mixed in a random order, and only a few reports at the top of the pile were readable after the flood. Thus, the SRS assumption is naively satisfied.

Ⓑ Since the daily profit is measured for the same Amish community in both years, a two-sample matched-pairs t-test should be used for analysis.

Ⓒ The combined sample size is sufficiently large for the t-test to be robust against slight positive skewness

Ⓓ A $\boldsymbol{Z}$-table is used when calculating the $\boldsymbol{p}$**-value** due to the limited access to electricity. The use of $\boldsymbol{Z}$-table is a reasonable approximation because $\boldsymbol{t}$**-distribution** becomes approximately **Normal** when $\mathbf{df}$ is sufficiently large.

2.4. In a simple linear regression, the predictor variable is recorded in pounds (lbs). What restriction, if any, does this put on the units of the response variable?

Ⓐ It must be measured in some other unit of mass or weight.

Ⓑ It cannot be expressed in pounds or any other weight units.

Ⓒ It may be measured in whatever units are appropriate for the outcome.

Ⓓ It must be measured in pounds exactly, matching the predictor units.

Ⓔ Each of the statements given above is simultaneously valid and applicable.

2.5. If the error variance is constant across all values of the predictor (homoscedasticity), what overall pattern should appear in a plot of residuals versus the predictor?

Ⓐ The residuals spread out progressively, forming a fan shape.

Ⓑ The residual spread narrows progressively, forming a funnel shape.

Ⓒ Residual spread first widens and then narrows across the range.

Ⓓ Residual spread stays about the same, forming a uniform band.

Ⓔ Residuals align along a single slanted line rising or falling.

2.6. A researcher performs ANOVA to analyze a dataset, and they mistakenly used the entire sample size $\boldsymbol{n}$ instead of $\boldsymbol{n_i}$ when calculating group variances $\boldsymbol{s_i^2}$. Assuming the formula below was used for calculating the $\mathbf{MS_E}$ value, which of the following statements is **TRUE**?

$$\mathbf{MS_E} = \frac{\sum_{i=1}^{k}(\boldsymbol{n_i} - \mathbf{1})\boldsymbol{s_i^2}}{\boldsymbol{n} - \boldsymbol{k}}$$

Ⓐ The $\mathbf{MS_E}$ value is overestimated, so it increases the chance of rejecting $\boldsymbol{H_0}$.

Ⓑ The $\mathbf{MS_E}$ value is overestimated, so it decreases the chance of rejecting $\boldsymbol{H_0}$.

Ⓒ The $\mathbf{MS_E}$ value is overestimated, but the ANOVA results remain the same.

Ⓓ The $\mathbf{MS_E}$ value is underestimated, so it increases the chance of rejecting $\boldsymbol{H_0}$.

Ⓔ The $\mathbf{MS_E}$ value is underestimated, so it decreases the chance of rejecting $\boldsymbol{H_0}$.

Ⓕ The $\mathbf{MS_E}$ value is underestimated, but the ANOVA results remain the same.

(Matching) Fill in the Blank.

3. **(21 points)** 🧬 A biological research firm hired seven new intern statisticians 👩‍🔬 👨‍🔬. As part of their training program, the firm requires each intern to run a mock experiment and submit a report answering a unique statistical question. Being new to the position, the interns made various mistakes ⚠️ during this procedure. Next to each mistake, write the letter corresponding to its most serious consequence. The letters may be used more than once or not at all.

□ Arthur created a confidence interval, while a lower confidence bound was necessary.

□ To expedite the data-collecting process, Bill asked all his extended family members to participate in his experiment.

□ Charlie used a two-sample independent approach, while it was correct to use the matched-pairs approach.

□ Ron was unaware that the population SD was known, so he instead used the sample SD and the t-procedure to create a lower confidence bound.

□ Molly correctly obtained a lower confidence bound of **-0.73** with **95%** confidence. She reported that the true parameter is above **-0.73** with the probability of **0.95**.

□ After rejecting the null hypothesis for an ANOVA F-test, Fred proceeded to construct the confidence intervals for pairwise differences of means using Bonferroni method. Given the family-wise Type I error of α, he used the critical value of $t_{\alpha/2,df}$ for each interval.

□ Ginny conducted the F-test for linear regression and failed to reject the null hypothesis. She reported that the explanatory variable and the response variable can be seen as independent of each other.

A. The margin of error will be larger than necessary.

B. The margin of error will be smaller than necessary.

C. The sample will be a biased representation of the population.

D. The result will be impacted by extraneous variation.

E. Inaccurate message about the result will be conveyed to the non-statisticians in their team.

F. This does not lead to any serious consequences.

Free Response Questions. Show all work, clearly label your answers, and use **four decimal places**.

4. **(32 points)** At a high-frequency trading venue, the waiting time (in milliseconds, ms) until the next buy order arrives follows an **Exponential distribution** $\boldsymbol{T_{\text{Buy}}} \sim \textbf{Exp}(\boldsymbol{\lambda_{\text{Buy}}})$, and the waiting time until the next sell order arrives independently follows an **Exponential distribution** $\boldsymbol{T_{\text{Sell}}} \sim \textbf{Exp}(\boldsymbol{\lambda_{\text{Sell}}})$.

   Define the **signed arrival-time difference** between the next buy and sell orders as:

$$\mathbf{R} = \boldsymbol{T_{\text{Buy}}} - \boldsymbol{T_{\text{Sell}}}$$

   a) **(4 points)** Assume $\boldsymbol{\lambda_{\text{Buy}}} = \mathbf{3}$ **per millisecond (**$\boldsymbol{ms^{-1}}$**)** and $\boldsymbol{\lambda_{\text{Sell}}} = \mathbf{1}$ **per millisecond (**$\boldsymbol{ms^{-1}}$**).** Determine the **expected value** $\boldsymbol{E}[\mathbf{R}]$ and **variance** $\mathbf{Var}(\mathbf{R})$ of the signed difference $\mathbf{R}$. *(Clearly show your calculation steps.)*

   b) **(12 points)** Under these assumptions, it can be shown that the random variable $\mathbf{R}$ follows a Laplace distribution (double-exponential) with parameters:

$$\boldsymbol{\mu} = \boldsymbol{E}[\mathbf{R}], \ \ \textbf{and} \quad \boldsymbol{b} = \frac{\mathbf{1}}{\boldsymbol{\lambda_{\text{Buy}}}} + \frac{\mathbf{1}}{\boldsymbol{\lambda_{\text{Sell}}}}.$$

   The probability density function (pdf) of a Laplace-distributed random variable $\mathbf{R}$ is:

$$\boldsymbol{f_{\mathrm{R}}(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{2}\boldsymbol{b}}\, \boldsymbol{e}^{\frac{\boldsymbol{x}-\boldsymbol{\mu}}{\boldsymbol{b}}} & \boldsymbol{x} < \boldsymbol{\mu} \\ \dfrac{\mathbf{1}}{\mathbf{2}\boldsymbol{b}}\, \boldsymbol{e}^{\frac{-(\boldsymbol{x}-\boldsymbol{\mu})}{\boldsymbol{b}}} & \boldsymbol{x} \geq \boldsymbol{\mu} \end{cases}$$

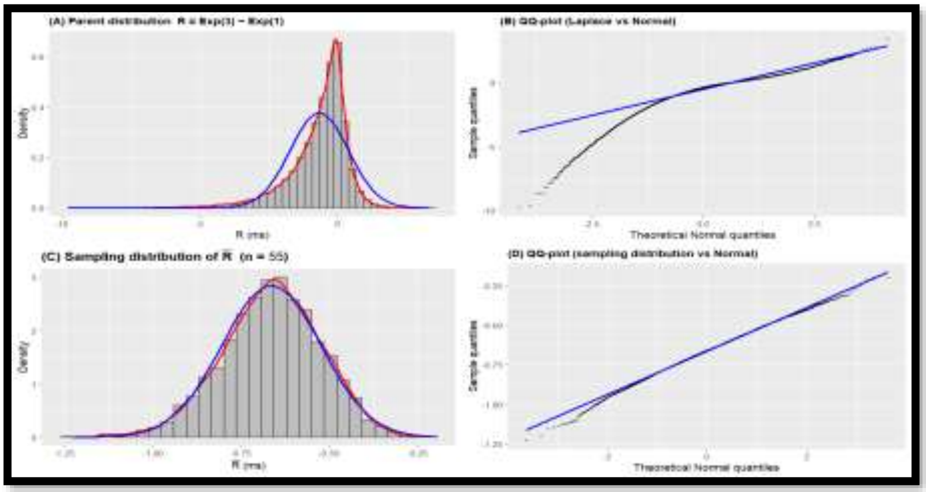
   Using your results from part (a), explicitly write down the parameters $\boldsymbol{\mu}$ and $\boldsymbol{b}$**,** and clearly compute the following probability: $\boldsymbol{P}(-\mathbf{1} \leq \boldsymbol{R} \leq +\mathbf{1})$.

c) (6 points) Now suppose that we record the next **55 independent observations** of signed arrival-time differences, each following the same Laplace distribution $\boldsymbol{R}$ defined above. Define the sample mean of these 55 differences as:

$$\overline{\mathbf{R}} = \frac{\mathbf{1}}{\mathbf{55}} \sum_{i=1}^{55} \mathbf{R_i}$$

Determine the exact mean $\boldsymbol{E}[\overline{\mathbf{R}}]$ and variance $\mathbf{Var}(\overline{\mathbf{R}})$.

d) (7 points) Based on the context of the problem and the provided plots, do you think it is reasonable to approximate $\overline{\mathbf{R}}$ with a Normal distribution using the Central Limit Theorem (CLT)? Answer **yes** or **no**, justify your answer by referring to the characteristics observed in the histograms and QQ-plots **(Figures (A)-(D))**.



e) **(3 points)** Even if you are uncertain about the Normal approximation, we will *use* it because the exact sampling distribution of $\bar{\mathbf{R}}$ is intractable.
Using the Normal approximation to $\bar{\mathbf{R}}$ with the correct mean $E[\bar{\mathbf{R}}]$ and standard deviation $\mathbf{SD}(\bar{\mathbf{R}})$ from part **d)**. Select the correct R code to compute the approximate probability that $P(-0.5 \leq \bar{\mathbf{R}} \leq 0)$.

Ⓐ pnorm(0, mean = $E[\bar{\mathbf{R}}]$, sd = $\mathbf{SD}(\bar{\mathbf{R}})$, lower.tail = TRUE)

− pnorm(-0.5, mean = $E[\bar{\mathbf{R}}]$, sd = $\mathbf{SD}(\bar{\mathbf{R}})$, lower.tail = FALSE)

Ⓑ pnorm(0, mean = $E[\bar{\mathbf{R}}]$, sd = $\mathbf{SD}(\bar{\mathbf{R}})$, lower.tail = FALSE)

− pnorm(-0.5, mean = $E[\bar{\mathbf{R}}]$, sd = $\mathbf{SD}(\bar{\mathbf{R}})$, lower.tail = TRUE)

Ⓒ pnorm(0, mean = $E[\bar{\mathbf{R}}]$, sd = $\mathbf{SD}(\bar{\mathbf{R}})$, lower.tail = TRUE)

− pnorm(-0.5, mean = $E[\bar{\mathbf{R}}]$, sd = $\mathbf{SD}(\bar{\mathbf{R}})$, lower.tail = TRUE)

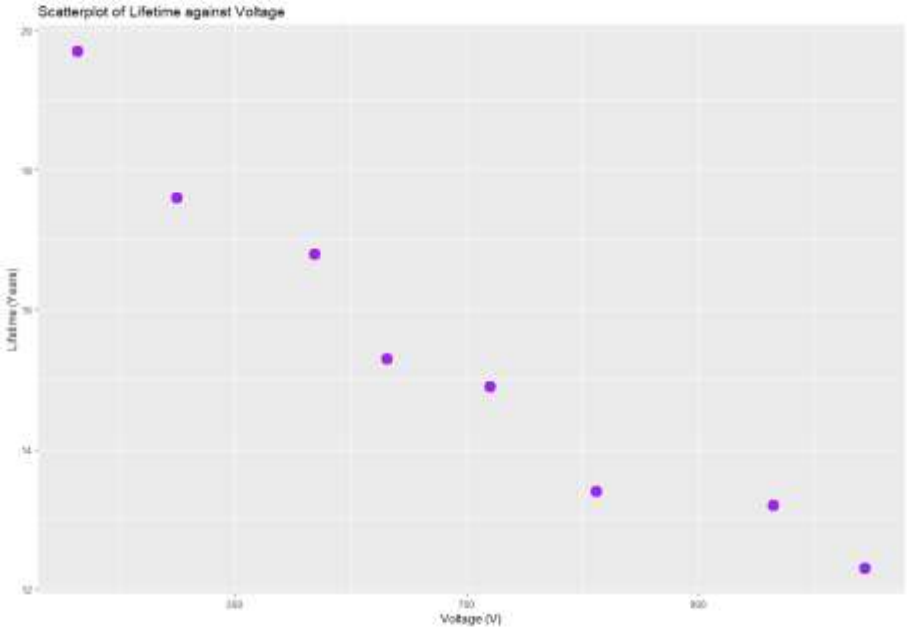
Ⓓ pnorm(0, mean = $E[\bar{\mathbf{R}}]$, sd = $\mathbf{SD}(\bar{\mathbf{R}})$, lower.tail = FALSE)

− pnorm(-0.5, mean = $E[\bar{\mathbf{R}}]$, sd = $\mathbf{SD}(\bar{\mathbf{R}})$, lower.tail = FALSE)

5. **(40 points)** Post silicon validation using machine learning tools helps engineers in semiconductor fabs to understand the behavior of chips under varying operating conditions. Purdue researchers want to establish the relationship between voltage (x) and the lifetime of semiconductor chips in electric vehicles (Y). They simulated the following data.

Voltage (V)	450	812	631	720	364	965	1044	569
Lifetime (years)	17.6	13.4	15.3	14.9	19.7	13.2	12.3	16.8

a) **(3 points)** Describe the relationship. strength, and direction of the relationship between lifetime of semiconductor chips and voltage in electric vehicles based on the scatter plot below.



(Answer on Next Page)

b) **(8 points)** The following output was realized from RStudio for the regression fit for these data.

```
Call:
lm(formula = Lifetime ~ Volt, data = jpr)
Residuals:
     Min       1Q   Median       3Q      Max 
-0.80339 -0.40072 -0.05738  0.48086  0.93905 
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 22.463962   0.782290  28.716 1.18e-07 ***
Volt        -0.010173   0.001073  -9.485 7.82e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.6771 on 6 degrees of freedom
Multiple R-squared:  0.9375,  Adjusted R-squared:  0.9271 
F-statistic: 89.96 on 1 and 6 DF,  p-value: 7.824e-05
```

(i) Write out the equation of the regression line.

(ii) Interpret the meaning of the estimate of the slope in (i) above.

(iii) Find the value of the coefficient of determination $\boldsymbol{R}^2$ from the output and interpret its meaning in the context of the problem

c) **(12 points)** Complete the ANOVA table. **Clearly, show your work in the box provided below.**

Source	Degrees of Freedom	Sum of Squares	Mean Square	F statistic	$Pr(> F)$
Regression			41.249	89.964	7.824e-05
Error					
Total		44			

d) **(7 points)** Based on the **ANOVA table** in **c)**, is there a significant linear association between the lifetime of the semiconductor chips and the voltage at a $\boldsymbol{\alpha = 0.001}$ level of **significance**? State the hypotheses and provide a **formal conclusion** in context.

e) **(10 points)** Given that $S_{xx} = 398569.8750$, construct a 99.9% confidence interval for β_1. Select an appropriate critical value for the calculation

```
> qt(0.001, 6, lower.tail = FALSE)      > qt(0.001/2, 6, lower.tail = FALSE)
[1] 5.207626                            [1] 5.958816
> qt(0.001, 7, lower.tail = FALSE)      > qt(0.001/2, 7, lower.tail = FALSE)
[1] 4.78529                             [1] 5.407883
> qt(0.001, 8, lower.tail = FALSE)      > qt(0.001/2, 8, lower.tail = FALSE)
[1] 4.500791                            [1] 5.041305
```

6. **(42 points)** In nature, it is uncommon to observe cats 😺 using vocalizations to communicate with other cats. However, when interacting with different species, cats easily copy and follow other species' preferred methods of communication. For example, domesticated cats 🐈 often become vocal 🎶 around humans, having learned that humans are not smart enough to understand cats' body language. On the other hand, rabbits 🐰 are known for being non-vocal 🫢, so cats tend to rely more on body language when interacting with them.

An ethologist 🧑‍🔬 plans to study **how vocal cats are when communicating** with **different species**. Five different species are selected; humans, dogs, parrots, rabbits, and hamsters. For each species, the **duration (in minutes)** of vocal communication per week for a group of **randomly selected 20 cats** is recorded. The table below provides the summary statistics for each group.

Group	👥Human	🐶Dog	🦜Parrot	🐰Rabbit	🐭Hamster
n_i	20	20	20	20	20
$\bar{x}_i$	50.1	37.4	44.3	15.6	24.3
s_i^2	12.3	9.10	10.4	9.35	9.40

a) **(3 points)** When analyzing the data through ANOVA, which of the following statements is **TRUE**?

Ⓐ The duration of vocal communication is a factor variable.

Ⓑ All groups must have the same sample size.

Ⓒ Measuring time in hours instead of minutes would change the ANOVA results.

Ⓓ Homogeneity of variance is automatically satisfied because all cats are sampled from the same population.

Ⓔ None of the above.

b) **(2 points)** Check the constant variance assumption using the summary statistics. Show your work and state clearly whether the assumption was satisfied or not.

c) **(12 points)** Fill each empty cell with the correct value to complete the ANOVA table. **Clearly, show your work in the box provided below.**

Source	Degrees of Freedom	Sum of Squares	Mean Square	F statistic
Factor				39.03792
Error		9842.808	103.6085	
Total	99			

d) **(4 points)** What is the estimated value of the assumed common variance among different species?

e) **(3 points)** Select the correct p-value associated with the ANOVA table in part c). Assume $\mathbf{F_{TS}}$, $\mathbf{df_A}$, and $\mathbf{df_E}$ are the correct test statistic, degrees of freedom for between groups, and degrees of freedom for within groups, respectively.

Ⓐ pf($\mathbf{F_{TS}}$, df1 = $\mathbf{df_A}$, df2 = $\mathbf{df_E}$, lower.tail = TRUE)

Ⓑ pf($\mathbf{F_{TS}}$, df1 = $\mathbf{df_A}$, df2 = $\mathbf{df_E}$, lower.tail = FALSE)

Ⓒ $\mathbf{2}$ ·pf($\mathbf{F_{TS}}$, df1 = $\mathbf{df_A}$, df2 = $\mathbf{df_E}$, lower.tail = FALSE)

Ⓓ The $\boldsymbol{p}$**-value** cannot be determined without specifying the significance level.

f) **(8 points)** The researchers obtain a $\boldsymbol{p}$**-value** less than **2e-16**. At **1% level of significance**, provide a formal decision and conclusion in the context of the problem.

g) **(10 points)** The R output below shows the Tukey's HSD results when using the **family-wise error rate** of **5%**. Use the R output and summary statistics, to draw a **graphical representation** of the Tukey's HSD results. Clearly indicate which species (or species, if multiple) has the shortest mean vocal communication duration in the population of species studied.

```
                      diff        lwr         upr     p adj
Hamster-Dog    -13.097560 -22.043802  -4.1513175 0.0009019
Human-Dog       12.701642   3.755400  21.6478844 0.0013940
Parrot-Dog       6.912555  -2.033688  15.8587969 0.2085189
Rabbit-Dog     -21.790044 -30.736286 -12.8438015 0.0000000
Human-Hamster   25.799202  16.852960  34.7454442 0.0000000
Parrot-Hamster  20.010114  11.063872  28.9563567 0.0000001
Rabbit-Hamster  -8.692484 -17.638726   0.2537583 0.0610890
Parrot-Human    -5.789088 -14.735330   3.1571548 0.3799648
Rabbit-Human   -34.491686 -43.437928 -25.5454436 0.0000000
Rabbit-Parrot  -28.702598 -37.648841 -19.7563561 0.0000000
```