Exam 1 — Part 2 — 2/15/2024

Instructions

This part is worth 40 points total. The exam (both parts) is worth 100 points total.

You have until the end of the class period to complete this part of the exam.

You may use your plebe-issue TI-36X Pro calculator.

You may refer to notes that *you have handwritten*, not to exceed *one side* of an 8.5" × 11" piece of paper.

You may *not* use any other materials.

No applications except for JupyterLab may be open on your laptop during the exam.

No collaboration allowed. All work must be your own.

Do not discuss the contents of this exam with any midshipmen until it is returned to you.

Type your answers **directly in this Jupyter notebook**, and submit this notebook (just the `ipynb` file) using the submission form on the course website.

Problem 1

a.

For $X \sim t(22)$, compute $P(-0.5 \leq X < 1.1)$.

Feedback. See Problem 4 from the Lesson 2 Exercises for a similar example.

In [ ]:

b.

For $Y \sim N(5, 13)$, compute the 0.42-quantile.

Feedback. See Lesson 2 (the section on *Random variables and distributions in R*) for a similar example.

In [ ]:

Problem 2

You have been recently been hired by Threads & Treasures, a local Simplexville clothing boutique. The data frame `Clothing` from the `Stat2Data` library contains several variables for 60 customers, including $Amount$, the amount spent during their latest visit (in dollars), and $Dollar12$, the amount spent over the last 12 months (in dollars).

Load the `Stat2Data` library and the `Clothing` data frame, and examine the first few rows by running the cell below.

In [1]:
```
library(Stat2Data)
data(Clothing)
head(Clothing)
```

A data.frame: 6 × 8

	ID	Amount	Recency	Freq12	Dollar12	Freq24	Dollar24	Card
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	1	0	22	0	0	3	400	0
2	2	0	30	0	0	0	0	0
3	3	0	24	0	0	1	250	0
4	4	30	6	3	140	4	225	0
5	5	33	12	1	50	1	50	0
6	6	35	48	0	0	0	0	0

a.

Compute the median amount spent by customers during their latest visit (in dollars).

Feedback. See Lesson 2 (the section on *An introduction to exploratory data analysis* for details on computing the median in R.

In [ ]:
```
```

b.

Create a histogram of the amount spent by customers over the last 12 months (in dollars). Your histogram should have 10 bins.

Feedback. See Lesson 2 (the section on *An introduction to exploratory data analysis* for details on creating a histogram in R.

In [ ]:
```
```

Problem 3

A person's systolic blood pressure can be a signal of serious issues in their cardiovascular system. Are there differences in average systolic blood pressure based on a person's size?

The data frame `Blood1` from the `Stat2Data` library has the systolic blood pressure ($SystolicBP$) and size ($Overwt$, where 0 = normal, 1 = overweight, 2 = obese) of 500 randomly chosen adults.

Load the `Blood1` data frame and examine the first few rows by running the cell below.

In [2]:
```
data(Blood1)
head(Blood1)
```

A data.frame: 6 × 3

	SystolicBP	Smoke	Overwt
	<int>	<int>	<int>
1	133	0	2
2	115	1	0
3	140	1	1
4	132	0	2
5	133	0	1
6	138	0	1

a.

Compute the one-way ANOVA table to test for differences in average systolic blood pressure between normal, overweight, and obese people.

Provide **only** the summary output for this part.

Hint. Remember to use `as.factor(Overwt)` to ensure that R treats $Overwt$ as a categorical variable.

Feedback. See Example 2 in Lesson 5, as well as STAT2 Exercises 5.54c and 5.73a assigned for homework, for details on how to compute the one-way ANOVA table in R. The homework exercises demonstrate the use of `as.factor()`.

In [ ]:

b.

According to the one-way ANOVA table you computed in part a, what is the p-value of the F-test that tests for differences in the average systolic blood pressure between normal, overweight, and obese people?

Using a significance level of 0.01, do you reject or fail to reject the null hypothesis? Briefly explain. Your explanation should be in complete sentences, with correct spelling and grammar.

Feedback. See Example 3 in Lesson 5, as well as STAT2 Exercises 5.54c and 5.73a assigned for homework, for similar examples.

Write your answer here. Double-click to edit.

c.

Based on your answer to part b, state your conclusion about the average systolic blood pressure of normal, overweight, and obese people.

Feedback. See Example 3 in Lesson 5, as well as STAT2 Exercises 5.54c and 5.73a assigned for homework, for similar examples.

Write your answer here. Double-click to edit.

d.

Using the rule of thumb covered in class, check whether **each population (group) has the same standard deviation**.

Use the code cell below to generate output that will help you check this condition.

In the Markdown cell below, comment on whether this condition is met, based on your output. Your explanation should be in complete sentences, with correct spelling and grammar.

Feedback. See Example 3 in Lesson 5, as well as STAT2 Exercises 5.54c and 5.73a assigned for homework, for similar examples.

In [ ]:

Write your answer here. Double-click to edit.

Problem 4

College students were asked to look at a photograph of an adult face and rate the person, on a scale of 1 (low) to 10 (high), for attractiveness. They were also asked to rate trustworthiness, faithfulness, and sexual dimorphism (i.e., how masculine or how feminine a face is). Overall, 68 students rated 170 faces.

The data frame `FaithfulFaces` from the `Stat2Data` library contains the data from this experiment. Each observation corresponds to one face. The variable $Trust$ contains the average trustworthiness rating, and the variable $Attract$ contains the average attractiveness rating.

For this problem, we want to understand the relationship between the average trustworthiness rating and average attractiveness rating of a face.

Load the `FaithfulFaces` data frame and examine the first few rows by running the cell below.

In [3]:
```
data(FaithfulFaces)
head(FaithfulFaces)
```

A data.frame: 6 × 7

	SexDimorph	Attract	Cheater	Trust	Faithful	FaceSex	RaterSex
	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<fct>	<fct>
1	4.900	2.000	1	4.118	6.294	F	M
2	2.583	2.083	1	4.000	6.588	F	M
3	4.300	2.667	1	3.941	5.824	F	M
4	4.083	2.917	0	4.118	5.412	F	M
5	4.500	3.500	0	3.235	4.000	F	M
6	2.000	1.833	0	4.647	6.647	F	M

a.

Fit a simple linear regression model that predicts the average trustworthiness rating of a face using the average attractiveness rating of a face.

Provide **only** the summary output for this part.

Feedback. See Example 2 in Lesson 7, as well as STAT2 Exercises 1.19 and 1.23 assigned for homework, for similar examples.

In [ ]:

b.

Is the **linearity** condition for simple linear regression met?

Using the code cell below, create a **single** diagnostic plot that will help you answer this question.

In the Markdown cell below, comment on whether the linearity condition is met, based on your diagnostic plot. Your explanation should be in complete sentences, with correct spelling and grammar.

Feedback. See Lesson 8 Part 1, especially the table on page 4, for details on how to check whether the linearity condition is met. See Lesson 8 Part 2 for a similar example.

In [ ]:

Write your answer here. Double-click to edit.

c.

Is the **equal variance** condition for simple linear regression met?

Using the code cell below, create a **single** diagnostic plot that will help you answer this question.

In the Markdown cell below, comment on whether the equal variance condition is met, based on your diagnostic plot. Your explanation should be in complete sentences, with correct spelling and grammar.

Feedback. See Lesson 8 Part 1, especially the table on page 4, for details on how to check whether the equal variance condition is met. See Lesson 8 Part 2 for a similar example.

In [ ]:

Write your answer here. Double-click to edit.

d.

Is the **normality** condition for simple linear regression met?

Using the code cell below, create a **single** diagnostic plot that will help you answer this question.

In the Markdown cell below, comment on whether the normality condition is met, based on your diagnostic plot. Your explanation should be in complete sentences, with correct spelling and grammar.

Feedback. See Lesson 8 Part 1, especially the table on page 4, for details on how to check whether the normality condition is met. See Lesson 8 Part 2 for a similar example.

In [ ]:

Write your answer here. Double-click to edit.

Grading rubric

Problem	Weight
1a	0.2
1b	0.2
2a	0.2
2b	0.2
3a	0.4
3b	0.4
3c	0.4
3d	0.4
4a	0.4
4b	0.4
4c	0.4
4d	0.4
Max Score	**40**