---

test2 - Confidential

Course Name: -

Notices:

This exam has no notices

Question #: 1

This question is to be answered using *Python*.
Consider the stud_perf dataset once again. There were three measures of performance of the students:

Feature	Description	Details
G1	First period grade	from 0 to 20
G2	Second period grade	from 0 to 20
G3	Final grade	from 0 to 20

In this question, we shall investigate the *agreement* between G1 and G3 scores. If there is a high agreement, it means that teachers can drop one of the scores and still get a good estimate of final grade of a student.

1. First, read in the dataset, and divide the G1 and G3 scores into 10 bins:

- [0,2], (2,4], (4,6], (6,8], (8,10], (10,12], (12,14], (14,16], (16,18], (18,20]

- You should now have two new columns G1_bin and G3_bin in the dataset.

2. Next, create a contingency table with G1_bin in the rows and G3_bin in the columns.

3. Convert the cell counts into proportions. The sum of entries in all cells should now equal to 1. Create a visualisation of this table of proportions that is relevant to the goal of measuring agreement.

4. If we let pij be the proportion in row i and column j, then the strictest version of agreement is $\eta_0 = \sum_{i=1}^{10} p_{ii}$ What is the range of values for η_0? Which values correspond to higher agreement?

5. A less stringent measure of agreement is given by $\eta_1 = \sum_{|i-j| \leq 1} p_{ij}$ Compute η_1 separately for the five groups defined by Medu and summarise what you observe.

Item Weight: 8.0

Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.41	-	-	-	-	-

Question #: 2

In 1935, Sir R A Fisher described an experiment involving a British woman. The woman claimed that if she was presented with a cup of milk tea, she would be able to distinguish whether milk or tea was added to the cup first. To test, she was given 8 cups of tea, in four of which milk was added first.
The data collected was as follows:

	Actual Milk	Actual Tea
Guessed Milk	3	1
Guessed Tea	1	3

Suppose that Fisher's Exact Test was applied to assess if there was any association between her guesses and the truth. Under the null hypothesis of Fisher's Exact Test, what is the probability of observing the table above?

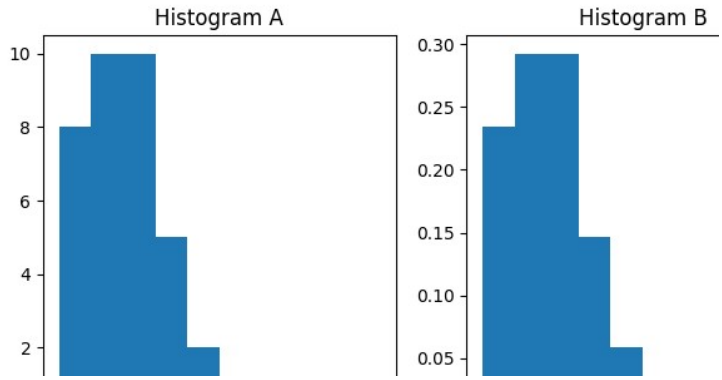
A. 0.5
B. 0.25
C. 0.5625
✓D. 0.228

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.23	-	-	-	-	0.22

Question #: 3

Consider the following two histograms (created with Python) from the liverpool dataset used in Tutorial 2:



What is the argument needed to convert the Histogram A into Histogram B?
liverpool.GF.hist(___1___)

1. Choice of: freq=False | density=True | type="percent" | type="density" - Correct Answer:freq=False

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.97	-	-	-	-	0.17

Question #: 4

The data in phones.csv contains information on phone calls made in Belgium from 1950 until 1973. Let Y be the number of calls, and X be the year variable. Read the data into Python as a pandas dataframe and answer the following questions:

1. Write a function that takes in three arguments: beta0, beta1, and the phones dataframe. It should compute the following L1-norm and return it: $\sum_{i=1}^{n}|Y_i-\beta_0-\beta_1 X_i|$

2. Iterate over a range of beta0 and beta1 values and find the pair that minimizes the L1-norm. Return this pair of beta0 and beta1 values.

3. Create a plot of this line, along with the OLS estimate, along with the datapoints.

4. Discuss the benefits of the L1-fit over the OLS fit.

Item Weight: 8.0

Item Psychometrics:

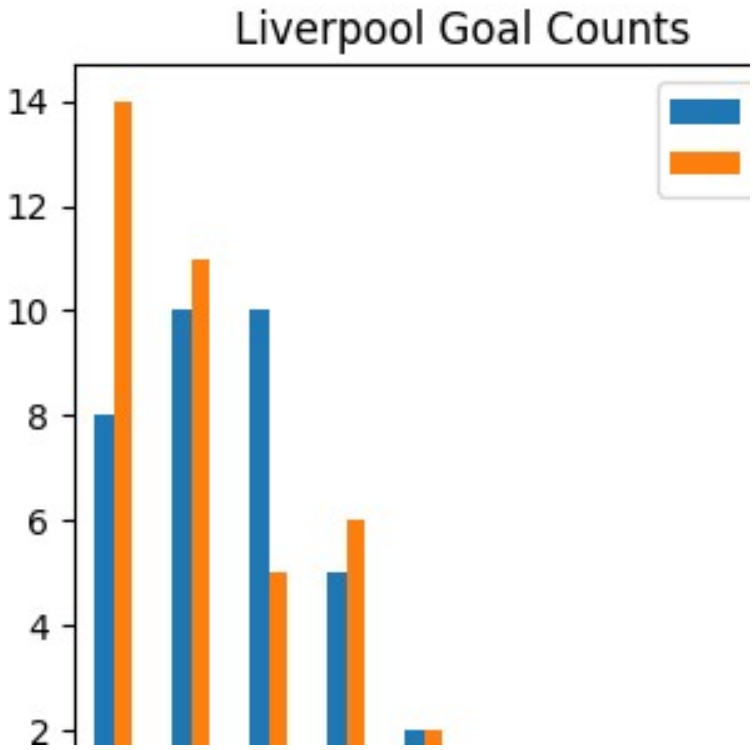
Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.44	-	-	-	-	-

Question #: 5

Suppose that the data in liverpool_2223_season.csv has been read into Python as liverpool. The following code tabulates the goal counts *for* and stores them in a column in goal_counts.

```
goal_counts =pd.DataFrame(np.zeros((10, 2), dtype='int'),
columns=['GF', 'GA'])
tmp2 =liverpool.GF.value_counts()
goal_counts.loc[tmp2.index, 'GF'] =tmp2
```

Continue the code to fill up the second column, which tabulates the goal counts *against* into the second column of goal_counts. Then create the following bar chart, which compares the GF and GA:



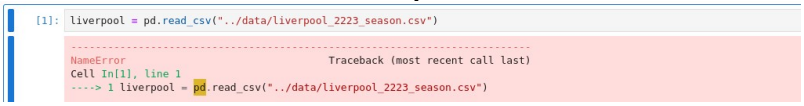
Item Weight: 4.0

Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.71	-	-	-	-	-

Question #: 6

What is the most likely solution to the error below?



A. The function to read the file is read.csv, not read_csv.
✓B. The pandas package has to be imported.
C. The pandas package has to be installed.
D. The separator has to be specified as sep=';'.

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.99	-	-	-	-	0.14

Question #: 7

The heifers dataset was introduced on in the topic on ANOVA. An analyst used SAS to run the ANOVA procedure and estimate the contrast comparing the Control group to the rest of the five groups. This is the code that the analyst used:

```
proc glm data=ST2137.HEIFERS;
class type;
model org=type / clparm;
means type / hovtest=levene welch plots=none;
lsmeans type / adjust=tukey pdiff alpha=.05;
estimate 'control vs. rest' type -1 5 -1 -1 -1 -1 / divisor=5;
run;
```

The essential output for this particular contrast can be seen in the following figures:

Class Level Information		
Class	Levels	Values
type	6	Alfacyp Control Enroflox Fenbenda Ivermect

Number of Observations Read	34
Number of Observations Used	34

Level of type	N	org	
		Mean	Std
Alfacyp	6	2.89500000	0.1167
Control	6	2.60333333	0.1187
Enroflox	6	2.71000000	0.1619
Fenbenda	6	2.83333333	0.1235
Ivermect	6	3.00166667	0.1094

Dependent Variable: org

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confi
control vs. rest	-0.25566667	0.05489709	-4.66	<.0001	-0.36811826

Use Python to recreate

1. the point estimate of the contrast, and

2. the confidence interval for the contrast.

Item Weight: 6.0

Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.42	-	-	-	-	-

Question #: 8

This question is to be answered using *Python*.
In Tutorial 6, we considered the notion of a sensitivity curve for assessing the robustness of an estimator. Suppose we have an estimator based on n observations: $T_n(x_1,x_2,\ldots,x_n)$ The sensitivity curve sc(x) when a new observation x is added is given by $sc(x)=(n+1)\times[T_{n+1}(x_1,\ldots,x_n,x)-T_n(x_1,\ldots,x_n)]$ Let us consider a newly proposed robust estimator called the *broadened median*. Given a sample of size n, here is how it is computed:

- For n odd, the broadened median is
- the average of the three central order statistics for $5\leq n\leq 12$.
- the average of the five central order statistics for $n\geq 13$.
- For n even, the broadened median is
- the weighted average of the central four order statistics for $5\leq n\leq 12$, with weights 1/6,1/3,1/3 and 1/6.
- the weighted average of the central six order statistics for $n\geq 13$, with weights 1/10,1/5,1/5,1/5,1/5 and 1/10.

Suppose we have the following sample of 10 points (n=10):
2,4,6,7,8,10,14,19,21,28

1. Plot the sensitivity curve for $x\in[5,25]$ when T represents the *broadened median*.

2. Which of the robust estimators that we covered in class is most similar to the broadened median? Explain your answer.

Note that your code does not have to handle the general case of estimator. It only has to work for the above dataset.

Item Weight: 6.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.52	-	-	-	-	-

Question #: 9

What is the length of the resulting output in this Python code?
my_list =['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']
result =my_list[2:8:2]
len(result)

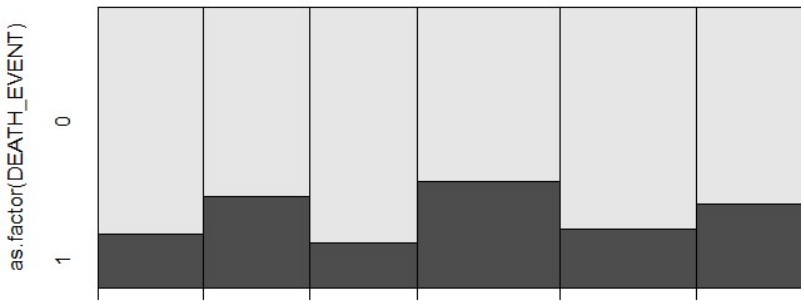
A. 6
B. 4
✓C. 3
D. 2

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.99	-	-	-	-	-0.01

Question #: 10

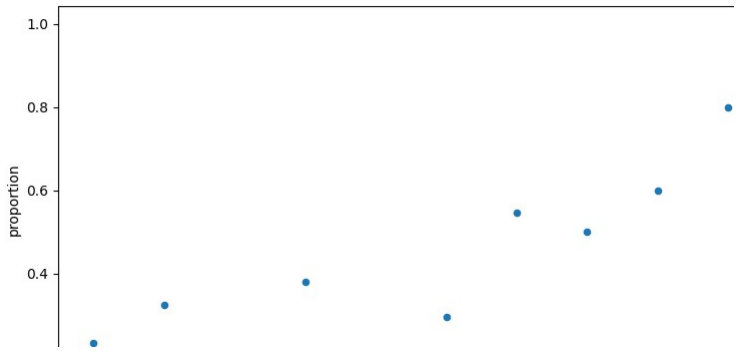
In the topic of categorical data analysis, we encountered the following plot for visualising how a binary variable varies with a continuous one. We had used the data from
`heart_failure_clinical_records_dataset.csv`



Let us try to create the data for such a plot in Python.
1. Read in the dataset as `heart_failure` and create the following table, which contains the binned age column, and the proportion of DEATH_EVENT in the second column:

```
## age_interval proportion
## 0 (40, 45] 0.233333
## 1 (45, 50] 0.324324
## 2 (50, 55] 0.157895
## 3 (55, 60] 0.380000
## 4 (60, 65] 0.208333
## 5 (65, 70] 0.297297
## 6 (70, 75] 0.545455
## 7 (75, 80] 0.500000
## 8 (80, 85] 0.600000
## 9 (85, 90] 0.800000
## 10 (90, 95] 1.000000
```

2. Now create this plot using the pandas dataframe, using kind='scatter'.



Item Weight: 4.0

Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.32	-	-	-	-	-

Question #: 11

Suppose that x is a numpy array with shape (2,2) and y is a numpy array with shape (2,1). What is the name of the numpy function for adding column y as a new column to x (making it have shape (2,3))?

A. np.cbind()
B. np.stack()
✓C. np.hstack()
D. np.concatenate()

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.92	-	-	-	-	0.15

Question #: 12

In many manufacturing processes, the term work-in-progress is often abbreviated to WIP. In a book manufacturing plant, WIP represents the time it takes for sheets from a press to be folded, gathered, sewn, tipped (with glue) on end sheets, and finally bound together.
The data set wip.txt contains samples of 20 books from each of two production plants, and the time for WIP (defined as the time in hours from when the books came off the press till they were packed in cartons).
There are two variables in the data set: time and plant (either 1 or 2).
Consider the following output. For the raw data, when we construct boxplots for each plant, there is a *single* outlier for each plant (the maximum value in each group).

```
## time
##   count mean    std      min  25%    50%    75%    max
## plant
## 1 20.0  9.3820  3.997653 4.42 7.4475 8.515  11.045 21.62
## 2 20.0  11.3535 5.126156 2.33 8.4400 11.960 13.845 25.75
```

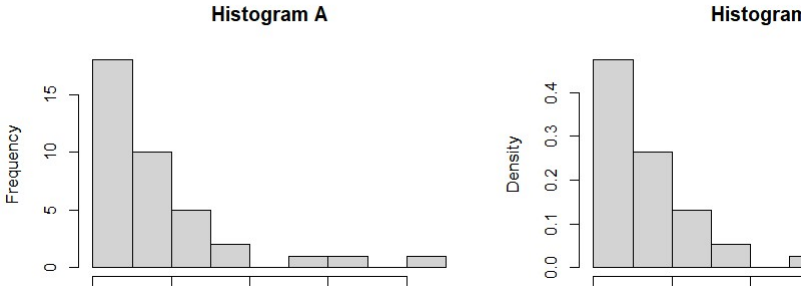
If we had applied a log (base e) transform to time before creating the boxplots, would these two points still be outliers? Explain your answer clearly using the summary statistics above only.

Item Weight: 4.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.34	-	-	-	-	-

Question #: 13

Consider the following two histograms (created with R) from the liverpool dataset used in Tutorial 2:



What is the argument needed to convert the Histogram A into Histogram B?
hist(liverpool$GF, __1__)

1. Choice of: freq=FALSE | density=TRUE | type="density" | type="percent" - Correct Answer:density=TRUE

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.84	-	-	-	-	0.23

Question #: 14

Using the student-mat.csv dataset from our lectures, create a contingency table from the variables address and guardian. Store it as address_guardian.
Write R code that will:

1. Compute the proportion of students whose home address was rural, and whose guardian was their mother.

2. Estimate the probability of students whose home address was rural, and whose guardian was their mother, under the null hypothesis of the chi2-test of independence.

Item Weight: 4.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.64	-	-	-	-	-

Question #: 15

What is the length of the resulting output in this R code?

```
vec1 <- c('a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j')
result <- vec1[2:8][-2]
length(result)
```

A. 3
B. 2
C. 4
✓D. 6

Item Weight: 1.0

Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
1.0	-	-	-	-	0.00

Question #: 16

This question is to be answered using R.
Suppose that daily demand for newspaper is approximately gamma distributed, with mean 10,000 and variance 1,000,000. At present, the newspaper company prints and distributes C=11,000 copies each day. The profit on each newspaper sold is $1, and the loss on each unsold newspaper is $0.25. Formally, the daily profit function h is
$h(X)=\{11000 \text{if } X\geq 11000 \lfloor X \rfloor+(11000-\lfloor X \rfloor)(-0.25) \text{if } X<11000$
where X represents the daily demand. Use simulation to estimate the expected profit per day, for various values of C. Thus recommend the optimal value of C to the company.
Ensure that when you make your case, you include confidence intervals, and that you include a visualisation of your results to assist the company in understanding your recommendation. For this question, set your seed to be 2002.
You will be awarded more marks for

- planning your code well,
- for using functions such as apply instead of for loops.
- for a clean and clear plot,
- and for a clear explanation of your results.

Item Weight: 6.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.7	-	-	-	-	-

Question #: 17

Assuming the appropriate package(s) have been installed on the computer, what command needs to be run in order to resolve the following error?

```
> corPlot(cor(stud_perf[, c("G1", "G2", "G3")]))
Error in corPlot(cor(stud_perf[, c("G1", "G2", "G3")
  could not find function "corPlot"
```

A. install.packages("psych")
B. load(psych)
✓C. library(psych)
D. library(lattice)

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.95	-	-	-	-	0.24

Question #: 18

After working with R, Python and SAS with one semester, you must have realised some of the strengths/limitations of these software. For each of the three software, list one advantage that *you feel* it has over the other two. There is no "right" or "wrong" answer, but your response should be a sincere one, and should be backed up with *examples from our course material*.

Item Weight: 3.0

Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.83	-	-	-	-	-

Question #: 19

The Γ(1.5,1) pdf is given by
f(y)=1Γ(1.5)y1/2e−y,y>0 where Γ(z)=∫0∞tz−1e−tdt is the gamma function. In this question, we shall use R to generate random variables from this pdf. (You are not allowed to use rgamma() in this question, but you can use rexp() and runif().)
The algorithm we shall use is the **rejection algorithm**. Here is pseudo-code for how it works:

1. Generate X~Exp(2/3) and independently, U~Unif(0,1). To be explicit, the pdf for X is: fX(x)=23exp−2/3x,x>0

2. If U<(2eX3)1/2e−X/3 then set Y=X and return Y. This is the acceptance step.

3. Otherwise, return to step 1 and generate a new X and U.

4. Repeat steps 1 - 3 until a Y is accepted. This Y will follow the Γ(1.5,1) distribution.

Note that not all X values will be accepted. The acceptance rate is the number of Y acceptednumber of (X,U) pairs generated

1. Write a function gen_one_Y() that takes **no arguments** and returns a vector of length two each time it is run: a single Y, and the number of X variables that were needed to get it.

2. Use this function to generate 105 random variables from f(y).

3. Create a histogram of the random variables generated, along with the actual pdf.

4. What is the acceptance rate? In other words, on average how many X variables are generated until a Y is accepted?

5. The correctness of the above algorithm can be proved theoretically. However, suppose you were given a vector of random variables, and you were told they were from a gamma distribution. Assuming you can use qgamma, explain how you can create a plot to assess if this claim is true. *Hint: Think about modifying the qq-plots the we learnt about for normality.*

Item Weight: 12.0

Item Psychometrics:

No item psychometrics are available at this time, this item has yet to be scored in any assessment.

Question #: 20

A sequence is generated using the following recursive relation:
$x_n = 2x_{n-1} - x_{n-2}, n \geq 3$
where $x_1 = 0$ and $x_2 = 1$.

Write R code to find x_{30} and $\sum_{i=1}^{30} x_i$.

Item Weight: 3.0

Item Psychometrics:

No item psychometrics are available at this time, this item has yet to be scored in any assessment.

Question #: 21

This question is to be answered using R.
The dataset in taiwan_outlier_removed.csv contains information on housing prices from Taiwan. The dependent variable is price. We had worked with this data in our tutorial; the greatest outlier from there has been removed.
In this question, we shall work with price as the dependent variable, and the following two explanatory variables:

1. ldist: The (natural) log of the distance to the nearest MRT station. The original distance is in meters.

2. num_stores: The number of convenience stores within walking distance.

Refer to the attached SAS output, and answer the questions 1–3. Following that, you will have to re-fit the model in R and run the additional analyses.

1. Interpret the model, when the number of stores is 0. Make sure that your explanation does not involve the logarithm scale.

2. The adjusted R^2 for this model is 0.59. Your colleague has fitted the same model, but *without* the log transform of distance. Are the two adjusted R^2 comparable? Why or why not?

3. Study the SAS plots of residuals versus predictor and assess if there is anything unusual.

4. Use R to return prediction intervals for the following two cases:

1. When the number of stores is 5, and the distance to MRT is 2175m.

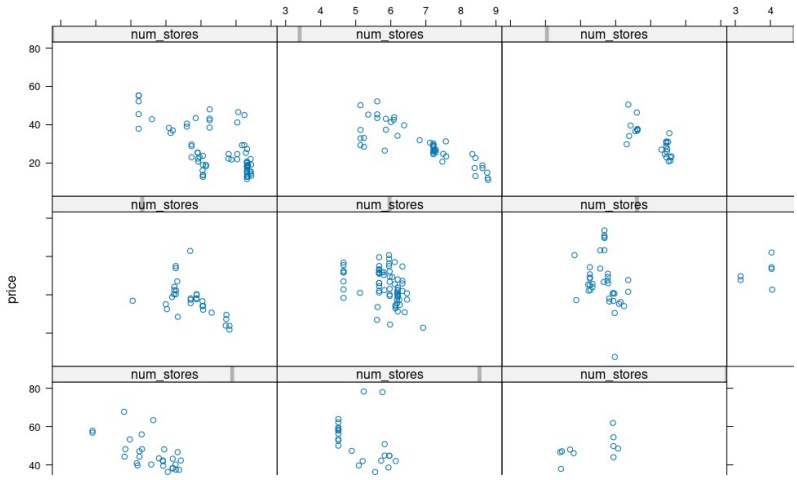
2. When the number of stores is 2 and the distance to MRT is 965m.

5. Identify the most influential point (in absolute value) with regard

to the beta coefficient for the number of stores. You can state the ID number of this point.

6. One of the columns in the influence matrix corresponds to dffit. What is the difference between this column and the unstandardised residuals that we compute?

7. Re-create the following plot. What do you observe about the relationship between price and ldist, in the presence of num_stores? *Hint: what is the association between ldist and num_stores?*



Attachment:

attachment_for_itemid_258138.pdf

Item Weight: 20.0

Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.51	-	-	-	-	-

Question #: 22

A clinical trial is conducted to compare the efficacy of drug A and drug B on lowering blood pressure. Participants are randomly divided into two groups. One group is given drug A and the other drug B. The *average reduction* in blood pressure after taking the drug is measured and compared between the two groups.
Assuming the distributional assumptions hold, the paired-sample t-test is appropriate in the above scenario (instead of the independent-sample t-test).

A. True
✓B. False

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.92	-	-	-	-	0.09

Question #: 23

Which of the following techniques *cannot* be used to assess the Normality of a given dataset?

A. Kolmogorov-Smirnov test
B. Shapiro-Wilk test
C. Skewness
✓D. 1-sample t-test

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.97	-	-	-	-	0.19

Question #: 24

In the One-Way ANOVA, we assume the following model:
$Yij=\mu+\alpha i+eij, i=1,\ldots,k, j=1,\ldots,ni$
The use of contr.sum( ) in R corresponds to the following constraint when performing estimation:

- Setting $\alpha 1=0$.

A. True
✓B. False

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.78	-	-	-	-	0.44

Question #: 25

In the topic on One-Way ANOVA, we worked with the heifers dataset. The following is the SAS output.

Dependent Variable: org

Source	DF	Sum of Squares	Mean Square	F Value
Model	5	0.59082353	0.11816471	7.97
Error	28	0.41500000	0.01482143	
Corrected Total	33	1.00582353		

The value of 0.0148 corresponds to $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y})^2$ where $\bar{Y}$ corresponds to the overall mean of *all* observations.

A. True
✓B. False

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.91	-	-	-	-	0.25

Question #: 26

In the topic on regression, we fitted the following simple linear regression model to the concrete data:

$$Y = \beta_0 + \beta_1 X + e$$

where

- Y: is the Flow, and
- X: is amount of Water in the mixture.

The summary output for the model (from Python) is as follows:

```
##                             OLS Regression Results
## ==============================================================================
## Dep. Variable:                   Flow   R-squared:
## Model:                            OLS   Adj. R-squared:
## Method:                 Least Squares   F-statistic:
## Date:                Wed, 24 Apr 2024   Prob (F-statistic):           8.1
## Time:                        09:39:28   Log-Likelihood:                -4
## No. Observations:                 103   AIC:
## Df Residuals:                     101   BIC:
## Df Model:                           1
## Covariance Type:            nonrobust
## ==============================================================================
##                  coef    std err          t      P>|t|      [0.025      (
## ------------------------------------------------------------------------------
## Intercept    -58.7276     13.286     -4.420      0.000     -85.084     -3
## Water          0.5495      0.067      8.196      0.000       0.416
## ==============================================================================
## Omnibus:                        6.229   Durbin-Watson:
## Prob(Omnibus):                  0.044   Jarque-Bera (JB):
## Skew:                          -0.523   Prob(JB):                     (
## Kurtosis:                       2.477   Cond. No.                    1.9
## ==============================================================================
```

The null hypothesis for the F-test above is H0:β1=β0=0

A. True

✓B. False

Item Weight: 1.0

Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.77	-	-	-	-	0.15

Question #: 27

When assessing robustness of an estimator, one of the properties we consider is the breakdown point.
For a particular parameter of interest, an estimator with a large breakdown point is considered to be better than an estimator with a smaller breakdown point.

✓A. True
B. False

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.78	-	-	-	-	0.41

Question #: 28

Consider the following SAS program:
DATA ex_1;
INPUT subject gender $ CA1 CA2 HW $;
DATALINES;
10 m 80 84 a ;
7 m 85 89 a
;
When the code above was run, there was no output. Which one of the following two steps will fix the error?

✓A. Removing the semi-colon on line 4.
B. Moving the semi-colon on line 6 to the end of line 5.

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.86	-	-	-	-	0.25

Question #: 29

Simulation can be used to estimate expected values of the form $E[g(X)]=\sum x=0\infty g(x)p(x)$. The reason we can assume Normality when computing Confidence Intervals is that it is up to us to choose the number of observations to generate.

✓A. True
B. False

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.64	-	-	-	-	0.08

Question #: 30

Suppose that $X \sim N(\mu=0, \sigma^2=4)$. Which of the following R commands will return $P(X>2)$?

✓A. pnorm(2, mean = 0, sd = 2, lower.tail = FALSE)
B. pnorm(2, mean = 0, var = 4, lower.tail = FALSE)
C. 1 - qnorm(2, mean = 0, sd = 2)
D. 1 - qnorm(2, mean = 0, var = 4)

Item Weight: 1.0
Item Psychometrics:

Difficulty Level/P-value	Upper 27%	Lower 27%	Discrimination Index	Point Biserial(Question)	Point Biserial(Rev)
0.96	-	-	-	-	0.21