

*Solutions*  
*Actual*  
**MATH 10, SAMPLE MIDTERM**

**Thursday, 4 November 2021<sup>1</sup>**

Name: \_\_\_\_\_

Student ID: \_\_\_\_\_

**Instructions:** You are allowed to have handwritten notes on the notecard that was distributed in class (both sides). Your name must be at the top of both sides of the notecard. No other resources are allowed. Your work will be graded on clarity as well as correctness; if your code works correctly but is significantly more complicated than necessary, that will not receive full points. Points will not be deducted for small syntax errors if your meaning is clear. Cross out incorrect work. Do work in the space provided. Good luck.

Question	Score	Maximum
1		8
2		6
3		6
4		6
5		10
6		8
7		6
Total		50

---

<sup>1</sup>Version A

1. (8 points)

a. Why will the following code cause an error?

```
my_tuple = (3,1,4,1)
my_set = {3,1,4,1}
my_list = [3,1,4,1]
new_set = {my_tuple, my_set, my_list}
new_list = list(new_set)
z = new_list[-1]
```

Can't put a set or a list  
into a set.

b. What is an advantage of a dictionary like

```
{"first name": "yasmeen", "last name": "baki"}
over a list of tuples, like
[("first name", "yasmeen"), ("last name", "baki")]
```

Easy to access values like

`d["first name"]`

Would be a lot harder with  
a list of tuples.

2. (6 points) Rewrite the following code to define `new_list` using list-comprehension instead of the for loop. (Assume `f` is already defined.)

```
new_list = []
for i in range(7):
    if f(i) > 4:
        new_list.append(i-2)
```

`new_list = [i-2 for i in range(7) if f(i) > 4]`

3. (6 points) Why would it be incorrect to start a Streamlit app with the following code?

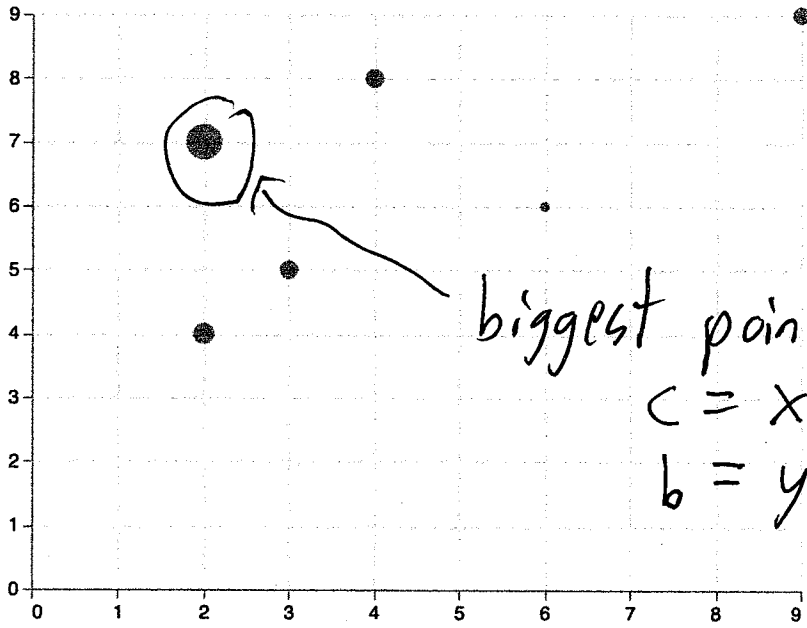
```
import streamlit as st

if "name" in st.session_state:
    st.session_state["name"] = "Default name"
else:
    name = st.session_state["name"]
```

We reach this line if "name" is not in `st.session_state`, but then `st.session_state["name"]` will raise an error.

4. (6 points) An Altair chart and the corresponding DataFrame are shown below, except the a column is not shown. In which row of df do you think the value in the a column is biggest? Explain your answer.

```
chart = alt.Chart(df).mark_circle().encode(  
  x = "c",  
  y = "b",  
  size = "a",  
)  
  
chart
```



```
df[["b", "c"]]
```

	b	c
0	8	4
1	5	3
2	7	2
3	9	9
4	6	6
5	4	2

Row 2

5. (10 points) Write code using NumPy or pandas (but no for loops) to estimate the following probability. (Try to make your code produce a number which is accurate to two decimal places, and try to make your code efficient.)

- Assume you choose 3 random real numbers (floats)  $x, y, z$  between 1.2 and 5. What is the probability that their product  $xyz$  is strictly greater than 20?

(You can use a function like `A.prod(axis = ???)` or `df.prod(axis = ???)` or you can just multiply 3 things together.)

```
import numpy as np
rng = np.random.default_rng()
n = 10**6
A = rng.random(size=(n,3))*3.8 + 1.2
p = (A.prod(axis=1) > 20).sum()/n
```

6. (8 points) Here is an example from Stack Overflow of one method to check if a column in a pandas DataFrame is numeric or not.

▲ In pandas 0.20.2 you can do:

```
127 import pandas as pd
    from pandas.api.types import is_string_dtype
    from pandas.api.types import is_numeric_dtype

    df = pd.DataFrame({'A': ['a', 'b', 'c'], 'B': [1.0, 2.0, 3.0]})

    is_string_dtype(df['A'])
    >>> True

    is_numeric_dtype(df['B'])
    >>> True
```

Share Edit Follow Flag

edited Aug 8 '17 at 12:39

answered Aug 8 '17 at 12:23



danthelion

2,740 ●2 ●12 ●9

Assume `df2` is some pandas DataFrame. Using the above method, make a new pandas DataFrame `df3` which consists of only the numeric columns from `df2`. For example, if we started with the DataFrame `df` defined above, then `df3` should be `df[["B"]]` (or equivalently `df.loc[:, ["B"]]` or `df.iloc[:, [1]]`). Try to do this without using a for loop.

$$\text{good\_cols} = [c \text{ for } c \text{ in } df2.columns$$
  
$$\text{if } is\_numeric\_dtype(df2[c])$$

$$df3 = df2[\text{good\_cols}]$$

7. (6 points) The DataFrame `df` shown below represents 100 data-points, where  $x_1$  and  $x_2$  are thought of as input variables, and where  $y$  is thought of as an output. Based on the following computation, what would you predict as the output  $y$  for input  $(x_1, x_2) = (1, 2)$ ? Show your work, but you don't need to simplify your answer.

```
df.shape
```

```
(100, 3)
```

```
df.head(4)
```

	<code>x_1</code>	<code>x_2</code>	<code>y</code>
0	11.740749	-21.353653	-394.468465
1	20.474649	-39.216277	-1402.911413
2	34.515304	-4.544947	-152.443622
3	-20.527154	3.335480	-243.297814

```
df["x_1 x_2"] = df["x_1"] * df["x_2"]
```

```
reg = LinearRegression()
```

```
reg.fit(df[["x_1", "x_2", "x_1 x_2"]], df["y"])
```

```
LinearRegression()
```

```
reg.coef_
```

```
array([ 4.11229395, -2.99405312,  2.00364286])
```

```
reg.intercept_
```

```
-1.2101020656248949
```

$$x_1 \cdot x_2 = 2$$

$$y \approx -1.21 + 1 \cdot 4.11 + 2 \cdot (-2.99) + 2 \cdot 2.00$$