Stat 342 Final -

Student Name _____

Student Number _____

You have exactly 180 minutes to complete this exam.

This test has 11 pages including this one

No calculators or other electronics are necessary or allowed.

You are allowed to bring a two-sided A4 (standard page) size aid sheet into this exam.

Protips:

- Show your work whenever appropriate. It shows understanding, and that's what's being tested.

- Use the backs of pages if space is an issue.

- If you get stuck on a part, don't abandon the question. Often later parts can be answered without earlier ones.

- Ask about any words you don't know. If it's not related to statistics, you will likely get an answer.

- Trust yourself. You have studied; you are an expert in this.

- Even experts check their work.

- Try not to panic, it rarely helps.

Good luck!

Question	1	2	3	4	5	6	7	
Out of	8	5	5	4	4	8	9	

Question	8	9	10	11	12			TOTAL
Out of	5	6	4	6	9			72

Problem 1, Total   / 8   2 points each.

In a PROC IML, what is the result of each of these print statements. Draw it in the space on the right.

A = (1 2 3, 4 5); print A;	
B = (1 2 3, 4 5 .); print B;	
C = (8, 11); C = shape(C,3,3); print C;	
D = (20, 40, 60, 80); E = (10 30 50); print D E;	

Problem 2, Total   / 5

Write a GLM procedure from the dataset payroll in the work library that...
- Models salary as a function of age and age squared
    … in such as way that a curve of best fit will be drawn.
- Produces all possible plots.

Age is a continuous variable.

Problem 3, Total   / 5

Write a procedure that…
- performs a linear regression
- of Y as a function of X1 X2 and X3,
- where X1 and X2 are categorical variables, and X3 is continuous.

Problem 4, Total   / 4

Write a PROCEDURE that imports a file from cheeses.csv and saves to the 'cheeses' dataset in the work library. Make sure to specify the DBMS. Use the variable names from the file.

Problem 5, Total   / 4

Write a univariate procedure that takes data from horses, and makes a histogram of the variable 'speed', with a overlay of a normal distribution on top.

Problem 6, Total   / 8

(5 pts) Write a procedure that…
- performs a logistic regression from the insurance data in the records library.
- that models the log-odds of claim being a 'yes' response (specify this),
- as a function of car_type, years_driving, and prev_accidents.
- The variable car_type is categorical.

(1 pt) What option is added to the model statement to find the 95% confidence interval of the slopes and intercept parameters?

(1 pt) If the response was not a binary yes/no, but an unordered categorical variable like 'no claim', 'theft', 'property damage', and 'injury', what option allow you to model this?

(1 pt) If SAS gave you a warning about an unstable result / failure to converge, what link option could you try that may give you a better result?

Problem 7, Total   / 9

(9 pts) Write a glm procedure that takes the data from the mtcars database in the work library...
- makes a model of 'mpg' as a function of 'weight' (continuous), and 'cylinders' (category),
- outputs the parameter estimates, standard errors, and p-values,
- outputs the predicted values and residuals,
- outputs the **99%** confidence limits of the mean responses, and
- saves the dataset with all these extras as mtcars2.

Problem 8, Total   /5

Given the dataset Olympics.Marathon below:

Name_First	Name_Last	Country	Age	Hours	Drugs
Bob	Mackenzie	Canada	48	8.92	No
Usian	Bolt	Kenya	22	2.83	No
Vladmir	Putin	Russia	65	1.11	No
…	…	…	…	…	...

(and so on for hundreds of rows)

Make a data step that adds the following variables to the dataset, and saves it as the same dataset.

Seconds, instead of hours.
Name_Full, a concatenation of name_first and name_last

Also replace the value 'drugs' to a missing value for all competitors from Russia.

Problem 9, Total   / 6

Explain what each line in the following code does. (Except run;)

```
proc sgplot data=warpbreaks;

    vbox breaks / category=tension;

    scatter y=breaks x=age;

    loess y=breaks x=age / smooth = 0.5;

run;
```

Problem 10, Total   / 4

Explain what each line in the following code does. (Except run;)

```
proc kde data=ds;

          bivar x y / plots(only) = contour;

run;
```

Explain what each line in the following code does. (Except run;)

```
proc glmselect data = mtcars;

    class cylinders;

    model mpg = weight hp cylinders

 weight*cylinders hp*cylinders

/ selection=stepwise(select = SBC);

run;
```

Explain what each part in the following code does.

```
DATA copenhagen;
     INFILE DATALINES TRUNCOVER;




     INPUT
          housing $
          income
          satisfaction $
          n;




DATALINES;
 tower  45000    low 21
 tower  32000 medium 21
 tower  77000   high 28;
```