MATH1180
Homework 3
Due Friday, February 10

Please (re)read Chapter 3 (Tables, Graphs, & Numerical Summaries), Section 4.1 (Experiments & Sample Spaces), and Section 4.2 (Events, Operations, and Probability) through the bottom of page 68. Please submit solutions to the following problems. You *must* use graph paper for the scatter plot in problem 5. You may download a graph paper PDF file from the course website.

When submitting homework, please remember the following:

- Show all work leading to each solution.
- *Staple* all sheets together. A paper clip is not acceptable.
- Do not submit crossed-out or sloppy work.
- Do not submit ripped or torn pages.
- Be sure to submit your own work.

PROBLEM 1 (RG). This problem gives a bit more practice with *sigma notation* (Σ) discussed in class as a way to describe sums. Let $x_1 = 4$, $x_2 = 3$, $x_3 = 11$, $y_1 = 8$, $y_2 = 3$, and $y_3 = 7$. Find the numerical value of these quantities:

(*a*) $\sum_{i=1}^{3} x_i$. (*b*) $\sum_{i=1}^{3} (y_i - 2)^2$. (*c*) $\sum_{i=1}^{3} x_i y_1$. (*d*) $\sum_{i=1}^{3} x_i^2 y_1^2$.

PROBLEM 2 (RG). The Super Bowl will be played on February 5 between the New England Patriots and the Atlanta Falcons. Suppose that 45% of the viewers will root for the Patriots (P), and the remainder will root for the Falcons (F). Suppose also that 55% of the viewers are men (M) and 45% are women (W), and that 20% of the viewers are both women and Patriots fans.

(*a*) Consider an experiment in which a viewer is randomly selected (with each viewer equally likely to be chosen). Use this information to construct a contingency table of probabilities for the experiment:

	W	M
Patriots (P)	$P(P \cap W)$	$P(P \cap M)$
Falcons (F)	$P(F \cap W)$	$P(F \cap M)$

NOTE: Your job is to fill the table with 4 probabilities that sum to 1.00. Work to 4 decimal places.

(*b*) Describe in words the sets $F^c \cap W^c$ and $M \cup F$.

PROBLEM 3 (SK). Luge competitions consist of a descent down an iced track in the prone position, feet pointing down the run. The best teams practice intensely, have dedicated coaches, and are helped by natural weather conditions. The x-values below are average times in seconds for men, and the y-values are average times for women, in random samples of single competitor luge runs for 10 countries who competed in the 2008 World Championships in Oberhof, German.

x (secs)	45.00	45.53	45.53	45.69	45.74	45.86	45.87	46.14	46.38	46.84
y (secs)	43.02	43.63	43.81	43.47	43.70	45.01	43.62	43.57	45.05	46.50

For these data, the mean time for male competitors is $\overline{x} = 45.86$ seconds, with a standard deviation of $s_x = 0.51$ seconds, and the mean time for female competitors is $\overline{y} = 44.14$ seconds, with a standard deviation of $s_y = 1.05$ seconds.

(a) Construct a chart of (z_x, z_y) pairs. Use 2 decimal places of accuracy in your chart.
(b) Use your answer to part (a) to compute the correlation between men's and women's times. Is there a strong relationship between the variables?

PROBLEM 4 (RG). In class, in the text, and in the notes, we have discussed a *randomized controlled experiment*. Explain in your own words:
(a) what is *random* in a randomized controlled experiment.
(b) what the word *controlled* refers to.

PROBLEM 5 (SK). Infant mortality and life expectancy are often used to characterize the overall health of a country. The x values below are infant mortality rates (deaths per 1,000 live births), and the y values are life expectancies in years, for 6 randomly chosen countries in 2008.

x (deaths per 1000 live births)	3.4	4.3	4.5	5.1	5.6	9.0
y (years)	80.9	78.6	79.4	81.2	80.1	77.4

For these data, the mean infant mortality rate is $\overline{x} = 5.32$ deaths per 1000 live births, with a standard deviation of $s_x = 1.95$ deaths per 1000 live births, and the mean life expectancy is $\overline{y} = 79.60$ years, with a standard deviation of $s_y = 1.44$ years. Further, the correlation between infant mortality and life expectancy is $r = -0.68$.
(a) Use this information to find the slope and intercept of the least squares regression line. Explain what the slope in the least squares regression equation says about the relationship between infant mortality and life expectancy. Be specific.
(b) What is the predicted life expectancy for a country whose infant mortality is 6.2 deaths per 1000 live births?
(c) Use your answer to part (a) to construct a chart of $(\widehat{y_i}, e_i)$ pairs, where $\widehat{y_i}$ is the predicted response and e_i is the residual for the i^{th} country. Use 2 decimal places of accuracy.
(d) Construct a scatter plot of (x_i, y_i) pairs. Include a graph of the least squares regression line from part (a) in your plot.
(e) Use your answer to part (c) to construct a residual plot of $(\widehat{y_i}, e_i)$ pairs.

PROBLEM 6. List all the outcomes in the sample space, $\mathcal{S}$, of the following experiment:

> A budget-minded family follows the amounts they spend each week for milk, bread and meat. At the end of 2014, they will compare the average weekly amount spent in 2014 to the average weekly amount spent in 2013 for each item, and record whether that average has gone up, gone down, or remained the same.

PROBLEM 7. List all the outcomes in the sample space, $\mathcal{S}$, of the following experiment:

> A market researcher asks a consumer the following questions:
> (a) What is your sex (`M`, `F`)?
> (b) Which flavor of ice cream do you prefer (`Chocolate`, `Vanilla`, `Other`)?
> (c) How many pints of ice cream do you eat weekly (`0`, `1`, `MoreThan1`)?

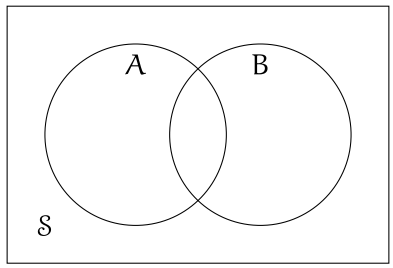
PROBLEM 8. List all the outcomes in the sample space, $\mathcal{S}$, of the following experiment:

> The plans of a college student can be regarded as a random experiment. Such a student might choose to go to a Montreal resort, go to Sydney for the Australian Open tennis tournament, go on a boat trip to save the whales, or go home to visit parents. If the student goes to Montreal, then he or she could choose to visit bars (and drink alcohol legally) or not. Similarly, if the student goes to Sydney, then he

or she could choose to visit bars or not. However, the student cannot legally drink alcohol on the boat trip or while staying with his or her family. Finally, on any of these trips, the student may bring along textbooks to study, or may leave them at school.

The outcomes of this experiment must indicate the location of the trip, whether or not the student chooses to drink alcohol, and whether or not the student brings along textbooks.

PROBLEM 9. Venn diagrams are useful ways to visualize two events, say A and B, contained in a sample space $\mathcal{S}$.



(a) Draw separate Venn diagrams in which you shade the events

$$A^c \cap B \text{ and } A^c \cup B^c.$$

Use your diagrams to determine if the event $A^c \cap B$ is a subset of the event $A^c \cup B^c$.

(b) Draw separate Venn diagrams in which you shade the events

$$A \cup B^c \text{ and } A^c \cap B^c.$$

Use your diagrams to determine if the event $A \cup B^c$ is a subset of the event $A^c \cap B^c$.

PROBLEM 10. The *National Center for Health Statistics* routinely publishes information on deaths in the United States. The following table gives the probabilities that a randomly selected citizen who died in 2010 was in each of eight groups:

White		*Black*		*Native American*		*Other*	
Male	Female	Male	Female	Male	Female	Male	Female
0.4260	0.4307	0.0591	???	0.0034	0.0029	0.0108	0.0100

(a) What number should be used to replace "???" in the table?

(b) Use your answer to part (a) to find the probabilities of the following events:

- (i) A citizen who died in 2010 was Male.
- (ii) A citizen who died in 2010 was Female.
- (iii) A citizen who died in 2010 was White.
- (iv) A citizen who died in 2010 was Black.
- (v) A citizen who died in 2010 was neither White nor Black.

PROBLEM 11 (WT). AM-PAC (Activity Measure for Post-Acute Care) is a patient-reported health survey to measure patients' general physical functioning after physical therapy or post acute care services. AM-PAC uses a four-point scale (1, 2, 3, 4) to rate a patient's performance on the tasks described in the survey, where 1 indicates complete dependence and 4 indicates complete independence.

(a) Suppose that a patient is evaluated on two occasions. If the patient receives a score of 2 on the first evaluation and a score of 4 on the second evaluation, then the patient's outcome can be symbolized by "24". List the sample space of possible outcomes for patients who are evaluated on two occasions.

(b) Each individual in a certain population of patients was given the survey 1 month after therapy and again 6 months after therapy. The following table gives the probabilities that a randomly selected patient from this population had each possible score on each occasion:

	6-*Month Score* 1	6-*Month Score* 2	6-*Month Score* 3	6-*Month Score* 4
1-*Month Score* 1	0.011	0.028	0.071	0.031
1-*Month Score* 2	0.003	0.100	0.280	0.112
1-*Month Score* 3	0.001	0.023	0.153	0.127
1-*Month Score* 4	0.001	0.003	0.015	0.041

For this population of patients, find the probability distribution of 1-Month Scores and find the probability distribution of 6-Month Scores. That is, fill-in each of the following tables

1-*Month Score*

1	2	3	4
???	???	???	???

6-*Month Score*

1	2	3	4
???	???	???	???

What, if anything, can you say about the changes from 1-month to 6-months?

(c) For the population of patients given in part (b), find the probability of each possible difference in scores, where the difference is defined as follows:

$$\text{Difference} = \text{6-Month Score} - \text{1-Month Score}.$$

(The possible differences are -3, -2, -1, 0, 1, 2, and 3.) Display your probabilities in a chart.

(d) For the population of patients given in part (b), find the probability that a patient's 6-Month Score was greater than his or her 1-Month Score.