STAT 311: Midterm

1 Instructions:

Name:

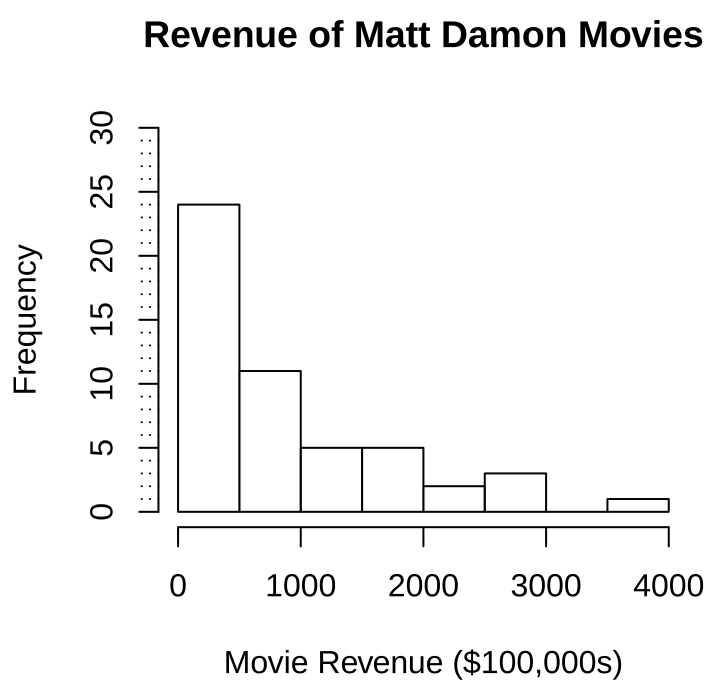
Please read the following statement and sign your name

"I pledge that I have not given or received any unauthorized assistance on this exam."

Signature:

2 Film Distribution

Matt Damon has been credited in **51** movies. The distribution of the movie revenues (adjusted for ticket price inflation) is shown below.



1. *Describe 3 aspects of the distribution of revenue of Matt Damon movies? (6 pts)*
   - Skewed Right
   - Mode is 0-500
   - Range is 0 to 4000

2. *The top 9 highest earning movies are listed below. (9 pts)*

Movie Title	Revenue ($100,000)
Saving Private Ryan	3925
The Bourne Ultimatum	2837
Ocean's Eleven	2752
Good Will Hunting	2533
The Bourne Supremacy	2435
The Martian	2253
Interstellar	1947
True Grit	1856
The Bourne Identity	1797

Calculate the 5 number summary of the revenues for the movies listed above (do not include all the other movies shown in the histogram). Is there an outlier? If so, which movie?

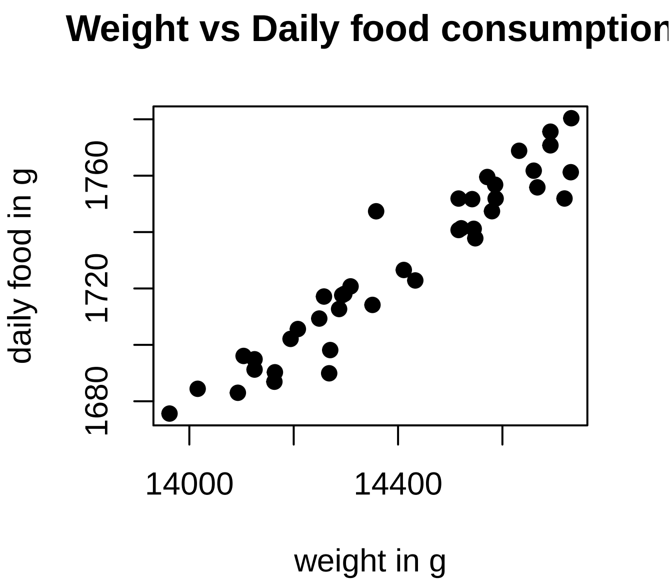
- Min: 1797
- Q1: 1901.5
- Median: 2435
- Q3: 2794.5
- Max: 3925

The IQR is 893.

893 * 1.5 + 2794 is 4133.5 and 1901 - 893 * 1.5 = -562 so there are no outliers.

3 We Bought a Zoo

In the movie "We Bought a Zoo," the recently widowed Benjamin Mee, played by Matt Damon, decides to make a fresh start by buying a zoo. Suppose Benjamin decides to bring in a new animal, a koala bear, but he doesn't know exactly how much food the koala bear will need each day. So, he gathers data on 40 other animals already at the zoo which are roughly the same size as a koala. Specifically, Benjamin decides to use the weight of an animal (in grams) to predict daily food consumption (also in grams). The scatterplot of animal weight vs daily food consumption is shown below.



Benjamin calculates the following quantities from his data

- $s_{weight} = 221.3$
- $s_{food} = 29.92$
- $\bar{x}_{weight} = 14399$
- $\bar{y}_{food} = 1727$
- $cov(food, weight) = 6339.88$

1. *What is the correlation between animal weight and daily food consumption? (10 pts)*

$$r_{food,weight} = \frac{cov(food, weight)}{s_{weight}s_{food}} = \frac{6339.88}{29.92 \times 221.3} = .957$$

2. *Give the best fit line for the relationship between animal weight and daily food consumption. That is calculate the slope and intercept and write the equation for the line. (10 pts)*

$$\hat{b} = r_{food,weight}\frac{s_{food}}{s_{weight}} = .957 \times 29.92/221.3 = .129$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 1727 - .129 \times 14399 = -130.471$$

3. *If* $SS_{total} = 34915$, *calculate* $SS_{regression}$ *and* SS_{error}.

$$SS_{regression} = SS_{total} \times r^2 = 34915 \times .957^2 = 31976.87$$

$$SS_{error} = SS_{total} - SS_{regression} = 34915 - 31976.87 = 2938.13$$

4. *Given that the koala weighs 14500 g, how much food (in grams) would you predict the koala to consume each day? (8 pts)*

$$\hat{y} = -130.471 + .129 \times 300 = .129 \times 14500 - 130.471 = 1740.03$$

5. *After the koala has been at the zoo for a year, Benjamin measures that the koala consumes 300 g of eucalyptus each day (That is really how much koalas eat according to Wikipedia). Because the koala is such a low-energy animal, it is an outlier in our data. Does the koala have high leverage in our regression? Explain why or why not. (4 pts)*

   The koala is an outlier in the Y direction, but not in the X direction. Thus, although it is a very unusual point, it will not have high leverage.

4 Jason Bourne

In the Jason Bourne series, Matt Damon plays a secret agent who has disassociative amnesia and has very little recollection of who he is or his past. Suppose Jason Bourne is trying to remember details about his past missions, but for each major city, he can only remember whether he has visited the city or not. He has put together the following two way table summarizing what he can remember. The counts in each cell represent a city which he has (or has not) visited.

	Visited	Has Not Visited	Total
UK	5	10	15
Germany	2	15	17
France	9	5	14
Italy	10	6	16
Total	26	36	62

1. *What proportion of cities has Jason visited? What type of distribution is this (conditional, joint or marginal)? Write this in probability notation and also give a numeric value. (6 pts)*

$$P(\text{visited}) = \frac{26}{62} = .419$$

This is a marginal probability.

2. *Of the cities in Italy, what proportion has Jason visited? What type of distribution is this (conditional, joint or marginal)? Write this in probability notation and also give a numeric value. (6 pts)*

$$P(\text{visited}|\text{Italy}) = \frac{10}{16} = .625$$

This is a conditional probability.

3. *If there is no association between the country of a city and whether or not Jason has visited the city, what number of cities in France would we expect Jason to have visited? (6 pts)*

$$\text{expected} = \frac{n_{France,+}n_{+,visit}}{n_{++}} = \frac{14 \times 26}{62} = 5.87$$

4. *For cities in France, what are the odds of Jason Bourne has visited?*

$$\text{odds} = P(\text{visit}|\text{France})/P(\text{visit}^c|\text{France}) = \frac{9/14}{5/14} = 1.8$$

5. *What is the risk ratio of Jason having visited a city given that a city is in Germany vs UK? Write this in probability notation and also give a numeric value. (6 pts)*

$$\frac{P(\text{visit}|\text{Germany})}{P(\text{visit}|\text{UK})} = \frac{2/17}{5/15} = .353$$

5 Rounders

In the movie "Rounders," Matt Damon plays Mike McDermott, a talented poker player who is playing against an opponent named Teddy. Mike realizes that when Teddy has a good hand, he often splits Oreos before he eats them. Thus, Mike can use this action as a potential test whether or not Teddy has a good cards or not. Suppose over many hands, Mike observes the following data from Teddy. **For this question, assume that "good cards" is a positive outcome, and "splitting oreos" is a positive test result.**

	Good cards	Bad cards	Total
Splits Oreo	50	10	60
Does not split Oreo	2	350	352
Total	52	360	412

1. *What is the sensitivity of using the Oreo test to test whether or not Teddy has a good hand? Write this in probability notation and give a numeric value. (5 pts)*

$$P(\text{Split Oreos}|\text{Good Hand}) = 50/52 = .962$$

2. *What is the specificity of using the Oreo test to test whether or not Teddy has a good hand? Write this in probability notation and give a numeric value. (5 pts)*

$$P(\text{Does no Split Oreos}|\text{Bad Hand}) = 350/360 = .972$$

3. *What is the Positive Predictive Value of using the Oreo test to test whether or not Teddy has a good hand? Write this in probability notation and give a numeric value. (5 pts)*

$$P(\text{Good Hand}|\text{Split Oreos}) = 50/60 = .833$$

4. *What is the Negative Predictive Value of using the Oreo test to test whether or not Teddy has a good hand? Write this in probability notation and give a numeric value. (5 pts)*

$$P(\text{Bad Hand}|\text{Does not Split Oreos}) = 350/352 = .994$$

6 Ocean's Eleven

In the Ocean's Eleven series, Matt Damon plays Linus Caldwell, a member of a larger team which seeks to rob a casino in Las Vegas. In the movie, the casino stores all its money in a safe in the basement. Suppose we know the following probabilities about how much money is in the casino's safe

- $P(\text{Safe} \leq \$5 \text{ mill}) = .2$
- $P(\$5 \text{ mill} < \text{Safe} \leq \$15 \text{ mill}) = .5$
- $P(\$15 \text{ mill} < \text{Safe}) = .3$

Depending on how much money is in the bank, the Casino randomly decides how many security guards to hire for each day with the following probabilities.

- $P(\text{guards} = 10 | \text{Safe} \leq \$5 \text{ mill}) = .5$
- $P(\text{guards} = 20 | \text{Safe} \leq \$5 \text{ mill}) = .5$
- $P(\text{guards} = 10 | \$5 \text{ mill} < \text{Safe} \leq \$15 \text{ mill}) = .3$
- $P(\text{guards} = 20 | \$5 \text{ mill} < \text{Safe} \leq \$15 \text{ mill}) = .7$
- $P(\text{guards} = 20 | \$15 \text{ mill} < \text{Safe}) = .4$
- $P(\text{guards} = 30 | \$15 \text{ mill} < \text{Safe}) = .6$

1. *Calculate $P(\text{guards} = 20 \cap \text{Safe} \leq \$5 \text{ mill})$. (5 pts)*

$$P(\text{guards} = 20 \cap \text{Safe} \leq \$5 \text{ mill}) = P(\text{Safe} \leq \$5 \text{ mill}) P(\text{guards} = 20 | \text{Safe} \leq \$5 \text{ mill}) = .2 \times .5 = .1$$

2. *Calculate $P(\text{guards} = 20)$. Drawing a probability tree may help, but is not necessary. (10 pts)*

$$\begin{aligned} P(\text{guards} = 20) = {} & P(\text{guards} = 20 \cap \text{Safe} \leq \$5 \text{ mill}) \\ & + P(\text{guards} = 20 \cap \$5 \text{ mill} < \text{Safe} \leq \$15 \text{ mill}) \\ & + P(\text{guards} = 20 \cap \$15 \text{ mill} < \text{Safe}) \\ = {} & .1 + (.5 \times .7) + (.3 \times .4) = .57 \end{aligned} \tag{1}$$

3. *Calculate $P(\text{guards} = 20 \cup \text{Safe} \leq \$5 \text{ mill})$. (10 pts)*

$$\begin{aligned} P(\text{guards} = 20 \cup \text{Safe} \leq \$5 \text{ mill}) = {} & P(\text{guards} = 20) \\ & + P(\text{Safe} \leq \$5 \text{ mill}) \\ & - P(\text{guards} = 20 \cap \text{Safe} \leq \$5 \text{ mill}) \\ = {} & .57 + .2 - .1 = .67 \end{aligned} \tag{2}$$

$$P(\text{guards} = 20 \cup \text{Safe} \leq \$5 \text{ mill}) = P(\text{guards} = 20) + P(\text{Safe} \leq \$5 \text{ mill}) - P(\text{guards} = 20 \cap \text{Safe} \leq \$5 \text{ mill})$$

4. *Suppose you are Linus Caldwell, and you know the probabilities above, but you can only observe the number of security guards on duty, and do not know the amount of money in the safe. Given that you have seen 20 security guards on duty, what probability would you assign to each of the following outcomes*

- $P(\text{Safe} \leq \$5 \text{ mill}|\text{guards} = 20) =$ (3 pts)

  $\frac{P(\text{guards}=20 \cap \text{Safe} \leq \$5 \text{ mill})}{P(\text{guards}=20)} = \frac{.1}{.57} = .175$

- $P(\$5 \text{ mill} < \text{Safe} \leq \$15 \text{ mill}|\text{guards} = 20) =$ (3 pts)

  $\frac{P(\text{guards}=20 \cap \$5 \text{ mill} < \text{Safe} \leq \$15 \text{ mill})}{P(\text{guards}=20)} = \frac{.35}{.57} = .614$

- $P(\$15 \text{ mill} < \text{Safe}|\text{guards} = 20) =$ (3 pts)

  $\frac{P(\text{guards}=20 \cap \$15 \text{ mill} < \text{Safe})}{P(\text{guards}=20)} = \frac{.12}{.57} = .211$

5. *Assuming that the money in the safe from night to night is independent (probably not true, but we'll use it for this example), what is the probability that during a week (7 nights), there is less than \$5 million dollars in the safe at least once. (15 pts)*

$P(\text{Less than \$5 mill at least once}) = 1 - P(\text{Never less than \$5 mill}) = 1 - (1 - .2)^7 = 1 - .201 = .799$

7 Contagion

In the movie "Contagion," Matt Damon plays Mitch Emhoff, a man whose wife is the first person in the world to catch the MEV-1 virus which eventually spreads across the world. In the movie, researchers must find a cure for the disease. Suppose the researchers use the following procedure to test a treatment which they have developed-

The researchers at the Centers for Disease Control (CDC) take a group of 1,000 individuals from the Atlanta suburbs (where the CDC is located) who have already been exposed the MEV-1 virus. In the 1,000 individuals, there are 600 women and 400 men. They randomly split the 600 women into two groups, one group receives a retrovirus injection and one group which does not. They also randomly split the 400 men into two groups, one group receives the retrovirus injection and one group does not. The group which does not receive a retrovirus injection receives a saline injection (essentially salt water) instead. The patients as well as the nurses administering the treatment do not know which injections are the retrovirus and which are saline.

1. *Is this an experiment or observational study? (3 pts)*
   - **Experiment**
   - Observational Study
   - Can't tell

   Explicitly applying treatment (retrovirus) to the subjects
2. *What type of experimental design is this? (3 pts)*
   - Retrospective Analysis
   - Twin Study Design
   - **Blocked Design**
   - Case-Control

   Blocked by sex
3. *What is the control group? (3 pts)*
   - The researchers who designed the experiment
   - **The group of patients who do not receive the retrovirus injection**
   - The group of patients who recover from the MEV-1 virus
   - The group of patients who receive the retrovirus injection

   Control group does not recieve the treatment
4. *What is the most likely purpose of the saline injection? (3 pts)*
   - **To control for a placebo effect**
   - To test whether saline has an effect on the virus
   - To prevent the patients from feeling sick
   - Can't tell

   The saline injection likely doesn't have any real effect, but it makes the control patients believe they are getting a treatment, just like the treatment group
5. *Which type of validity would we be more concerned about in this case? (3 pts)*
   - Internal Validity

- **External Validity**

It is a randomized and double blind study, so the internal validity should be okay. However, we only tested patients from the Atlanta area, so we don't know how well this might generalize to other patients.