Math 142
Part One Exam

March 17, 2014

Figure 1 is a scatter plot of midterm vs. final exam test scores for a class.

Questions 1-3 refer to Figure 1.
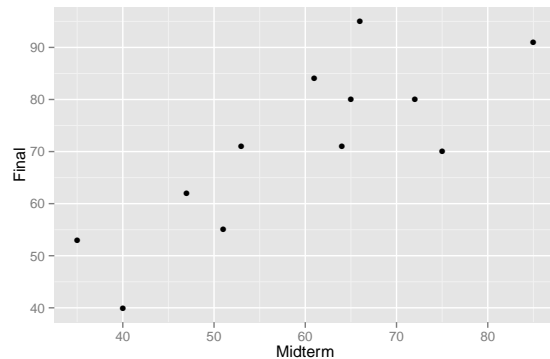

Figure 1: Questions 1-3

1. (3 points) Estimate the mean score on the final for people who scored more than 60 on the midterm:

   (a) about 70

   (b) about 80

   (c) about 90

   > **Solution:** The students who scored more than 60 on the midterm are the students in the top right quadrant of the graph. The mean on the final for these students is about 80.

2. (3 points) Estimate the correlation coefficient:

   (a) about 0.5

   (b) about 0.8

   (c) about 0.99

   > **Solution:** 0.5 would be a more widely dispersed cloud of points and 0.99 would be a nearly straight line. The actual coefficient is around 0.8.

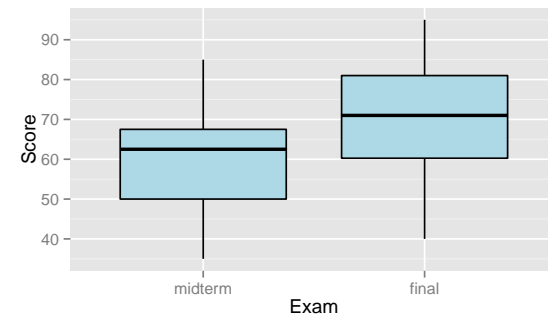3. (a) (8 points) Draw box plots for the midterm and final

   > **Solution:**
   >
   >
   >
   > Figure 2: Question 3 (a)

(b) (2 points) Compared to the midterm, was the final exam

  (a) harder

  (b) easier

  (c) about the same difficulty

> **Solution:** The scores were generally higher on the final, so the test was probably easier than the midterm.

4. Here are the average points per game for some NBA starting players:

Player	Points	Player	Points
Trey Burke	12.3	LeBron James	27.2
Kentavious Caldwell-Pope	6.0	Michael Kidd-Gilchrist	7.4
Tyson Chandler	9.3	Robin Lopez	10.8
Glen Davis	12.1	Kevin Martin	19.2
Jared Dudley	7.3	Ben McLemore	7.6
Kevin Durant	31.8	Chandler Parsons	16.6
Raymond Felton	10.0	Kendrick Perkins	3.4
Randy Foye	12.6	Zach Randolph	17.1
Marc Gasol	13.7	Terrence Ross	10.7
Paul George	22.2	Jared Sullinger	12.9
Spencer Hawes	13.0	P.J. Tucker	9.3
Dwight Howard	18.9	Evan Turner	16.6

Table 1: Points per game

.

(a) (10 points) Draw a histogram with a bin width of 5 points for these players.

> **Solution:** Raymond Felton has an average of exactly 10, so you have a choice of what group to include him

> in. I made the each bar include the lower boundary and exclude the upper boundary, but it is equally valid to do it the other way around also.
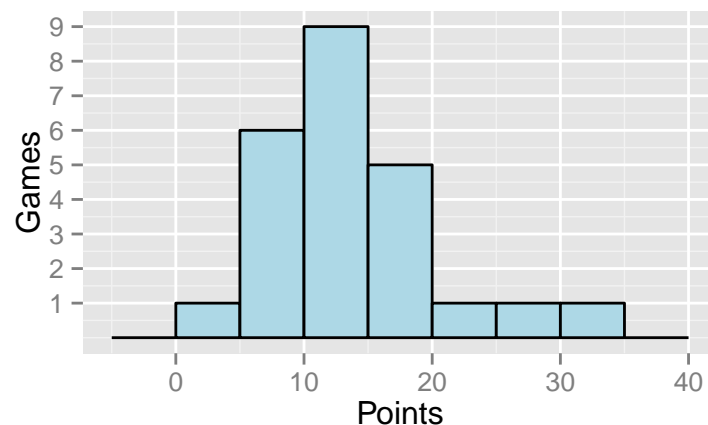>
> Figure 3 contains the histogram.

Figure 3: Question 4

(b) (2 points) Is the distribution symmetrical, left-skewed, or right-skewed?

> **Solution:** right-skewed

(c) (3 points) Is the mean larger or smaller than the median? Explain how you can know the answer without calculating the actual values.

> **Solution:** Because the distribution is right-skewed, the mean while be larger than the median.

(d) (3 points) Is the median or mean a better measure of the center of this distribution? Why?

> **Solution:** Since this is a skewed asymmetrical distribution, the median is a better measure of the center. The median indicates what a typical player is likely to score, while the mean is heavily influenced by a few high scoring players.

5. There is heated debate this year about whether Kevin Durant or LeBron James should be the MVP.

   Here is a representative selection of the number of points they scored in selected games this season:

Durant	{ 13, 17, 19, 24, 26, 28, 29, 31, 32, 32, 32, 37, 41, 42, 42 }
James	{ 13, 15, 17, 19, 25, 25, 25, 26, 29, 30, 32, 33, 35, 36, 36 }

   (a) (8 points) Find the 5-number summary for both players

   > **Solution:**
   >
   > | Player | Min. | Q1 | Median | Q2 | Max. |
   > |---|---|---|---|---|---|
   > | Durant | 13 | 24 | 31 | 37 | 42 |
   > | James | 13 | 19 | 26 | 33 | 36 |

(b) (6 points) Draw a box plot for both players
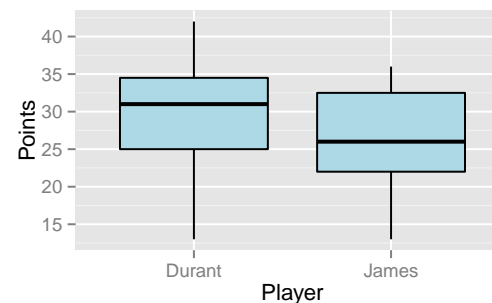
> **Solution:**
>
>
>
> Figure 4: James vs. Durant Points

(c) (3 points) Which player seems to be the MVP, based solely on points scored in these games?

> **Solution:** All of Durant's 5-number summary numbers except the minimum are better, and the minimum is the same. The sample is representative of all the games. If the award was based solely on scoring Durant would be the MVP.

6. Three point shooting accuracy in 2013 is a roughly Normal distribution with a mean of 0.36 and a standard deviation of 0.05.

   (a) (5 points) Carmelo Anthony's three point shooting accuracy is 0.42 this year. What is the z-score for this value?

   > **Solution:**
   > $$\frac{0.42 - 0.36}{0.05} \approx \boxed{1.2}$$

3

(b) (5 points) What percentage of the three point shooters in the NBA are less accurate than Melo?

> **Solution:** If you look up Carmelo's z-score in table A, you find that about 0.8849 or $\boxed{88.5\%}$ of players have scores lower than 1.2.

(c) (5 points) What is the three point shooting accuracy for someone who is better than 20% of the three point shooters?

> **Solution:** For this problem you need to reverse the process. From table A, the z-score that corresponds to 0.2 is -0.8416.
>
> You have to convert this back to a shooting percentage:
>
> $$-0.8416 = \frac{x - 0.36}{0.05}$$
> $$x \approx \boxed{0.3179}$$

Table 2 shows the shots made and missed by position for NBA players in the 2013-14 season.

Use Table 2 to answer questions 7-9.

position	missed	made
Center	11105	11372
Power Forward	18972	17601
Small Forward	19583	15090
Shooting Guard	20721	15617
Point Guard	20791	15087

Table 2: Shots by position

7. (5 points) Make a bar plot showing shots made for each position. Order the bars by number of shots made, with the most productive position first.
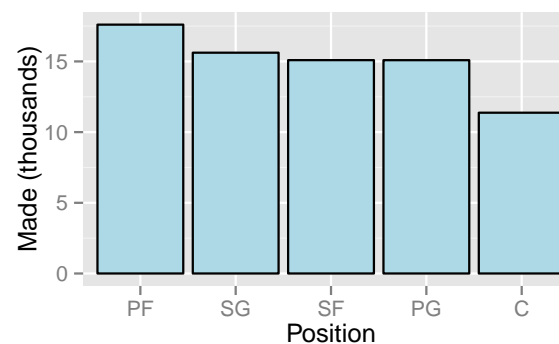
> **Solution:**
>
>
>
> Figure 5: Question 7: shots made by position

8. (5 points) Fill in the table showing the marginal distributions of shots taken by position, shots missed, shots made, and total number of shots taken.

Solution:

position	made	missed	total
Center	11372	11105	22477
Power Forward	17601	18972	36573
Small Forward	15087	20791	35878
Shooting Guard	15090	19583	34673
Point Guard	15617	20721	36338
total	74767	91172	165939

Table 3: Marginal distributions

9. (a) (5 points) Make a table showing the conditional distributions of percentage of shots made and missed for each position.

Solution:

position	made	missed
Center	51%	49%
Power Forward	48%	52%
Small Forward	44%	56%
Shooting Guard	43%	57%
Point Guard	42%	58%

(b) (2 points) Which position seems to contain the most accurate shooters? Is this is because they are the best shooters or is there a lurking variable that explains the shooting percentage differences?

Solution: The centers are the most accurate shooters. This is because most of them are 7 feet tall and never venture more than 3 feet from the basket or attempt a shot more difficult than a dunk.
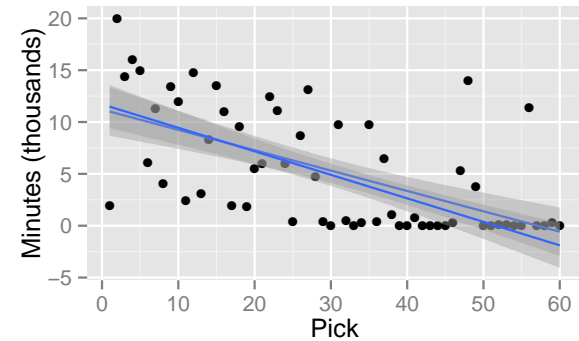

Figure 6: 2008 NBA draft order vs. minutes played

	mean	s
Pick	30.5	17.5
Minutes (thousands)	5219	5698

Table 4: 2008 NBA Draft

10. Figure 6 shows the total number of minutes played vs. draft order for the 2008 draft. The better players are generally drafted first and also usually play more, so there is a negative correlation of $r = -0.6020$ between draft order and minutes played.

The first two selections in the 2008 draft were Greg Oden (bottom left corner) and Kevin Durant (top left corner). As you can see in the chart, Greg Oden has had an injury plagued career

and has only played 1940 minutes while Kevin Durant has had a successful and so far mostly injury free career and has played nearly 20,000 minutes

Many of the players after pick 30 never played in the NBA at all. Any of them that were underclassmen declaring for the draft early should have stayed in school.

(a) (7 points) Find the equation for the regression line using the draft pick as the explanatory variable and minutes played as the response variable.

> **Solution:** Use the correlation coefficient and the standard deviations to find the slope of the regression line:
>
> $$b = -0.6020 \cdot \frac{5698}{17.5} \approx -196$$
>
> Use the slope and the means to find the y-intercept:
>
> $$\bar{y} = a - 196\bar{x}$$
> $$5219 = a - 196 \cdot 30.5$$
> $$a \approx 11,197$$
>
> The equation is:
>
> $$\boxed{\hat{y} = 11,197 - 196x}$$

(b) (5 points) Graph the regression line.

> **Solution:** See Figure 6 (slightly less steep regression line).

(c) (2 points) What percentage of the variation in the minutes played is accounted for by the linear regression line?

> **Solution:** $r^2 \approx 0.3625$

(d) (3 points) Mark Gasol and Ramon Sessions (top right corner) are outliers because they were late round draft picks (pick 48 and 56) who have had a successful NBA careers with over 11,000 minutes each so far.

Would removing these data points from the graph have a significant effect on the regression line? Why or why not?

> **Solution:** It depends on what you mean by "significant." You could say "yes" because they are away from the center of the graph, so they are sort of outliers on the x axis. On the other hand, you could say "no" because there are 7 other points in the same x-range right around zero, so the line can't follow the outliers very well.
>
> Either answer is correct, as long as you provided a reasonable explanation.
>
> The actual numbers without those two players are:
>
> $$r \approx -0.6975$$
> $$\hat{y} \approx 11,708 - 227x$$
>
> Both regression lines are in Figure 6. The line without the outliers is slightly steeper (lower on the right side).