Name: Key   **Math 127 – Exam 2 – Spring 2016**

Oath: ***"I will not discuss the exam contents with anyone until it is returned to me by my instructor".***

Sign Name: Key

The penalty for cheating on this exam is a grade of 0% for Math 127 Exam 2.

Testing Center Staff Instructions

1. **One sheet of handwritten or typed notes is OK.**

   Students may not use the "pink sheet" or any copied or scanned answer keys or Math 127 department documents.

2. **Collect the sheet of notes and staple it to the test when submitted.**

3. **Any calculator is OK. No cell phone calculators.**

4. **www.statcrunch.com is required. All other webpages are prohibited.**

5. **Test must be completed in one sitting, but it is untimed. Very short bathroom breaks are permitted.**

Student Instructions

1. **This test is graded out of 100 points and counts for 20% of your Math 127 grade.**

2. **Show work when necessary or points will be deducted. If you only report an answer and it is wrong, you will receive no credit. You may use the StatCrunch calculators for any probability calculations when appropriate.**

3. **Points are in parentheses for each question.**

4. **Good luck, do your best.**

1. Use the "**Calendar Year 20116 Grocery Prices**" dataset. Pretend we work for Wal-Mart. Based on the "***Supermarket Price***", we'd like to predict "***Wal Mart Price***" to see if our prices are in line with the marketplace.

1a. (2) Using the data, give the linear regression equation. Use the whole dataset. Do not delete any data points.

$\widehat{Walmart} = 0.28 + 0.77(\text{Supermarket Price})$

1b. (3) The slope is approximately 77 cents. Interpret this value with a sentence in the context of the problem.

For every \$1 increase at the supermarket, we expect an increase of 77¢ at Walmart

1c. (5) Any products with large positive Studentized residuals could possibly be over-priced at Wal-Mart. Circle all of these products:

"36 count Donut Shop K- Cups 0.39 oz" "Clif Bar Energy Bars 14.40oz"

"Double Stuffed Oreo, 15.4 oz" "Gallon of milk"

"Honey Nut Cheerios Oats 12.25 oz" "Jif Peanut Butter 28 oz"

"Oscar Meyer Beef Bologna, 16 oz" "Silk Almond mil, 8 fl oz, 6 ct"

"TGI Fridays Honey BBQ Boneless wings, 15 oz" "UTZ Cheese Balls, 28 oz"

1d. (5) Any products with large negative Studentized residuals could possibly be under-priced at Wal-Mart. Circle all of these products:

"21.8 oz. Strawberry Nesquick" "36 count Donut Shop K- Cups 0.39 oz"

"Chex Mix Traditional Snack Mix 15 oz" "Chips Ahoy Original Chocolate Chip Cookies, 13 oz"

"Coca-Cola 2ltr." "Ellio's pepperoni pizza 18.9 oz"

"Honey Nut Cheerios Oats 12.25 oz" "Hormel Black Label Maple Bacon, 12oz"

"Silk Almond milk, 8 fl oz, 6 ct" "TGI Fridays Honey BBQ Boneless wings, 15 oz"

1e. (2) Any products with large Cook's Distances are influential and deserve a bit of scrutiny. For this particular dataset, what is the cutoff for a large Cook's Distance? Show calculation.

$4/128 = 0.03125$

1f. (2) For this dataset, how many products have a large Cook's Distance? 6

2. (2) Open up the "**Mold Colonies**" dataset on StatCrunch. Make a scatterplot to predict "***Mold Count***" using "***Hours***". Which condition for linear regression is clearly not met?

Answer: Linearity

3. Open up the "**NHL Skaters Raw Data 2013-2014**" dataset on StatCrunch.

Use x = "***Time on Ice***" to predict y = "***Points***" scored. It would seem reasonable that if a hockey player is on the ice for more time, he should end up scoring more points, all things considered.

"***Time on Ice***" is measured in minutes.

"***Points***" is number of points scored for the season by that player.

We will run an analysis for just the players who are Right Wingers. <u>"***Position***" = "***RW***" only!</u>

You can Group By "***Position***" <u>**or**</u> you can do a Where Box: Position = RW.

When you run the analysis, your correlation should be $r = 0.83436757$.

3a. (2) Why is the y-intercept of -34.55 not an interpretable value? There are three reasons. Give two of them.

1. Not possible to score y=-34.55 points
2. X=0 minutes played is silly

X=0 minutes - No data anywhere close by.

3b. (2) With a sentence in the context of the problem, interpret the value of the slope of 0.062: For each additional minute played, we expect an additional 0.062 points scored.

3c. (2) With a sentence in the context of the problem, interpret the value of $R^2 = 69.62\%$: 69.62% of the variation in Points scored can be explained by knowing Time on Ice. 30.38% is still unexplained.

3d. (2) With a sentence in the context of the problem, interpret the value of $S_e = 9.89$: On average, our predicted Points are off by 9.89 when using x= Time on Ice

3e. (2) Craig Adams, row 105, has a residual of –17.44. Interpret the residual with a sentence in context.

He scored 17.44 fewer points than predicted, based on having 1022 minutes on the ice.

3g. (2) Jaromir Jagr, row 191, has a residual of 4.72. Show the calculation to arrive at this value:

$e = y - \hat{y} = 67 - 62.28 = 4.72$

$62.28 \approx -34.55 + 0.0616(1571)$

3h. (2) Which Right Winger in the dataset is the biggest overachiever in terms of scoring way more "***Points***" than we would expect for his "***Time on Ice***"? Circle the correct choice:

Corey Perry Gustav Nyquist Martin St. Louis Phil Kessel

3i. (2) Blake Wheeler had "***Time on Ice***" of 1532 minutes. Show the calculation for his predicted "***Points***".

$\widehat{\text{Points}} = -34.55 + 0.0616(1532)$

$= 59.8212$

(59.87 on StatCrunch)

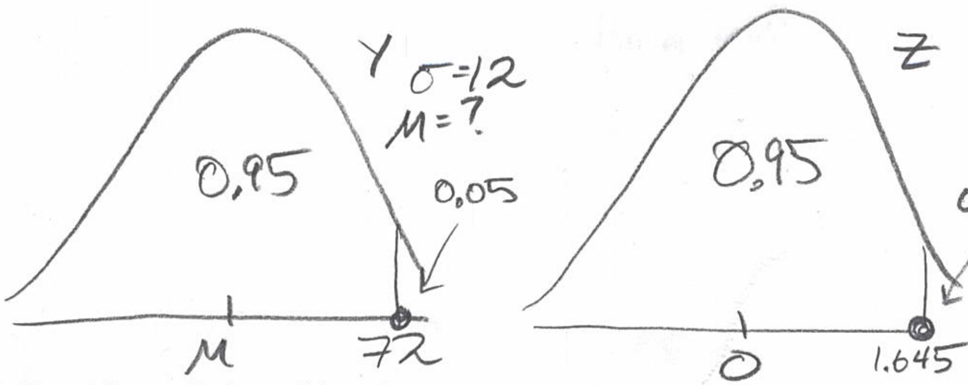
3j. (2) An extra 200 minutes of "***Time on Ice***" will lead to an additional 12.32 "***Points***" as predicted by our linear regression equation. Show calculation to arrive at this value:

$200(0.0616) = 12.32$

3k. (1) How do you feel about the equal spread condition for this analysis? Circle the correct choice:

I feel perfectly fine and the condition is met I feel very anxious and the condition is not met

4. (2) Concert ticket prices for Ryan Adams shows follow a Normal model with an unknown mean and a known standard deviation of $12.00. If 5% of tickets cost $72 or more, show the calculations to determine the mean ticket price.



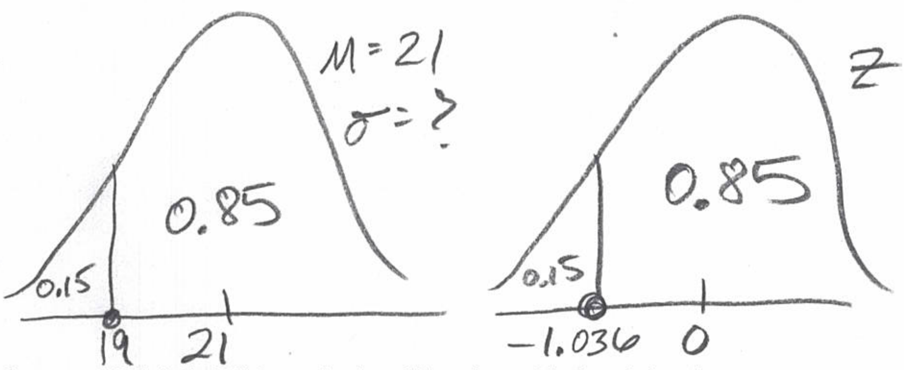
$$z = \frac{y - \mu}{\sigma}$$

$$1.645 = \frac{72 - \mu}{12}$$

$$\mu = 72 - 1.645(12)$$

$$\boxed{\mu = 52.26}$$

5. (2) A jogger feels confident she averages 21 miles each week but hasn't a clue on the standard deviation for her miles. Her records indicate that for 15% of the weeks, she jogged 19 or fewer miles. Assuming a Normal model applies, calculate her standard deviation for mileage. Show calculations.



$$z = \frac{y - \mu}{\sigma}$$

$$-1.036 = \frac{19 - 21}{\sigma}$$

$$\boxed{\sigma = 1.93 \text{ miles}}$$

6. ***Midnight White*** paint has blue dye added to it by the Home Depot associate. The amount of blue dye follows a N(5 ml, 0.03 ml) model.

6a. (2) Determine the probability that any can's blue dye amount falls within 3 standard deviations of the mean.

Answer: 0.9973

6b. (2) If you buy 25 cans to paint your gymnasium, what's the probability that all 25 cans fall within three standard deviations of the mean? Show calculation.

$$P(25 \text{ for } 25 \text{ w/in } 3 \text{ SD}) = (0.9973)^{25}$$

$$= 0.9346$$

6c. (2) For those 25 cans, what's the probability that at least one can has less than 4.91 ml or more than 5.09 ml of blue dye? Show calculation.

P(at least one can not inside) =

1 − P(all 25 are inside) =

1 − 0.9346 = 0.0654

7. The following table shows the outcomes, prizes, and probabilities for a carnival gambling game that uses a standard deck of 52 cards plus the two jokers. For each turn, the cards are shuffled and we use the full 54 cards. Show calculations for full credit.

Outcome	Prize	Probability
Club	\$0	0.24074
Spade	\$50	0.24074
Diamond	\$100	0.24074
Heart	\$200	0.24074
Joker	\$500	0.03704

7a. (2) P(Win at least \$100) = $0.24074 + 0.24074 + 0.03704 = 0.51852$

7b. (2) P(Win nothing three times in a row) =

$P(3 \text{ for } 3 \text{ win } \$0) = (0.24074)^3 = 0.01395$

7c. (2) Show calculation. Determine the expected winnings on each turn.

$$\mu = 0(0.24074) + 50(0.24074) + 100(0.24074) + 200(0.24074) + 500(0.03704) = \$102.78$$

Bonus (3) You play twice. P(Win at least \$700 total) = 0.0192.

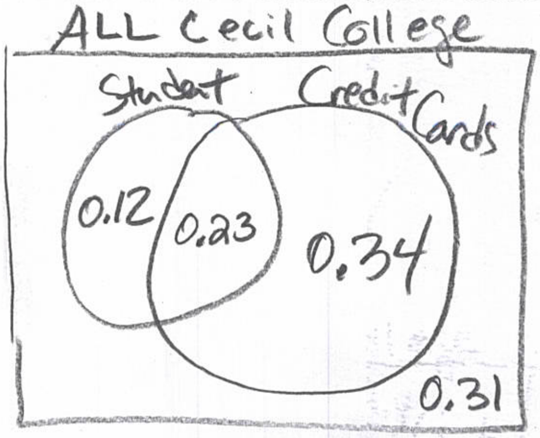
Show calculation for any credit:

$P(\text{Win at least } \$700) = P(\text{Win } \$1000) + P(\text{Win } \$700) =$

$(0.03704)^2 + 2(0.24074)(0.03704) =$

0.0192

8a. (2) Draw the Venn diagram and label everything properly. Suppose 35% of Cecil College students have student loans, 57% of our students have credit cards, and 23% have both.



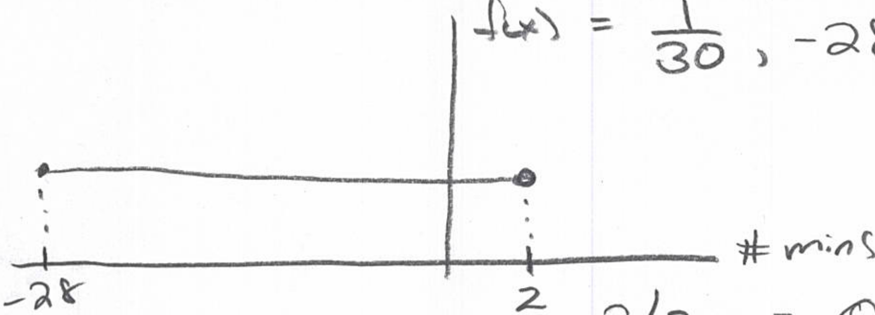
8b. (2) What's the probability that a student has neither student loans nor credit cards? 0.31

8c. (2) P(Student Loans | Have a Credit Card) = $\dfrac{P(\text{Both})}{P(\text{Credit Card})} = \dfrac{0.23}{0.57}$

$= 0.4035$

9. A math professor almost always ends class early. Class can end up to 28 minutes early or end 2 minutes late or anything in between. All times are equally likely, so the [–28, +2] Uniform model will apply. Negative numbers mean that class let out early and positive numbers mean class let out late.

9a. (2) Draw the model, label entirely, and give the probability function $f(x)$ = $\frac{1}{30}, \; -28 \le x \le 2$.



9b. (2) Determine the P(class runs late) = $2/30 = 0.0667$

9c. (2) Determine P(both classes this week get out early) =

$P(\text{any class gets out early}) = \dfrac{28}{30}$

$P(2 \text{ for } 2 \text{ early}) = \left(\dfrac{28}{30}\right)^2 = 0.8711$

10. A computer is programmed to produce random numbers on the interval +9 to +12. A Uniform model applies.

10a. (2) Calculate the mean of this probability model. Show calculation.

$$\mu = \frac{9+12}{2} = 10.5$$

10b. (2) Calculate the 80th percentile of this probability model. Show calculation.

$L = 3$. 80% of $3 = 2.4$

$P_{80} = 9 + 2.4 = 11.4$

11. The website for the game Uncle Web Knows Best gets a new visitor every so often. Suppose that time between visits to the site follows an Exponential model with a mean of 70 minutes.

11a. (2) What's the probability that the site's next visitor isn't for at least 3 hours? 0.0764

11b. (2) What's the probability that the site's next visitor is in the next 20 minutes? 0.2485

11c. (2) What's the 99th percentile for this probability model? 322.36 minutes

11d. (2) Interpret the 99th percentile with a sentence in the context of the problem: 99% of the time, UWKB gets a visitor within the next 322.36 minutes. 1% of the time it is longer.

12. (2) Hot Pockets runs a promotion aimed at college students where "1 in 3 wins" is printed on the boxes. Presuming winning boxes are randomly distributed, you stop at Acme and buy four boxes to tide you over during the weekend. What's the probability you win at least once? Show calculation.

P(win) = 1/3 , P(lose) = 2/3

P(win at least once in four) =

1 − P(4 for 4 lose) =

$$1 - \left(\tfrac{2}{3}\right)^4 = \frac{65}{81} = 0.8025$$

13. Suppose the heights of the men playing in the NCAA basketball tournament (i.e. March Madness) follow a Normal model with a mean of 74.5 inches and a standard deviation of 2.4 inches.

13a. (2) Determine P(Random player is over 77 inches) = 0.1488

13b. (2) Presuming independence, take a team of 12 players. P(all 12 are over 6 foot tall). Show calculation.

6 ft = 72 inches.

P(any player over 72") = 0.8512

P(12 for 12 over 72") = $(0.8512)^{12}$

= 0.1447

13c. (2) The tallest 10% of players are 77.58 inches tall or taller.

13d. (2) The central 70% of players are between 72.01 inches and 76.99 inches tall.

14. A trivia card game has different categories (sports, entertainment, etc…) and 15.9% of cards are about geography. A typical game has $n = 19$ cards played.

14a. (2) How many geography cards do we expect to be played in a typical game? Show calculation.

$\mu = np = 19(0.159) = 3.021$

14b. (2) What is the standard deviation for this probability model? Show calculation.

$\sigma = \sqrt{np(1-p)} = \sqrt{19(0.159)(0.841)} = 1.59$

14c. (2) P(At least 5 geography cards played in a game) = 0.1729

14d. (2) P(At most 2 geography cards played in a game) = 0.3987

14e. (2) The Normal model approximation for this problem is a faulty idea. Explain why: np = 3.021 does not exceed 10. Shape is skewed right.