Name: _____

Statistics 202 - Midterm

10/24/2014

Instructions

There are **five** questions on the exam. The last page is a standard normal table.

1. Do not open the exam until told to do so.
2. Write your name on the exam.
3. You may use a graphing calculator or a standard calculator.
4. You may use the front and back of an 8.5 by 11 inch sheet of paper.
5. You must show your work. Failure to do so may result in zero points for that portion of the question.
6. This exam is to be **only** your own work. Collaboration of any kind is forbidden.

Academic Integrity Statment Suspected violations of academic integrity will be reported to the Dean's Office. For more information on Northwestern's academic integrity policies, see http://www.weinberg. northwestern.edu/handbook/integrity/index.html.

Blank page for scratch work

Question 1 (20 pts)

Here are the results of one of the studies that link high blood pressure to death from cardiovascular disease. The researchers classified a group of white males aged 35 to 64 as having Low blood pressure or High, then followed the subjects for five years. The following two-way table gives the results of the study:

```
                         Blood Pressure
                         ---------------
Cardiovascular death?     Low    High        Total
---------------------------------------------------
     Yes                   21      55           76
      No                 2655    3283         5938
---------------------------------------------------
   Total                 2676    3338         6014
```

(A) What is the response variable? (1 pt)

(B) What is the explanatory variable? (1 pt)

(C) Are cardiovascular death and high blood pressure independent? (8 pts)

(E) Can we claim that high blood pressure is a cause of cardiovascular death from this study? Why or why not? (10 pts)

Question 2 (20 pts)

Recall that the Behavioral Risk Factor Surveillance System is a nationwide survey conducted by the CDC. The BRFSS asks, among other things,

- "How much do you currently weigh?" which they record as `weight` and
- "How much would you like to weigh?" which they record as `wtdesire`.

Here is a scatterplot from the 20,000 person subset of the BRFSS data.
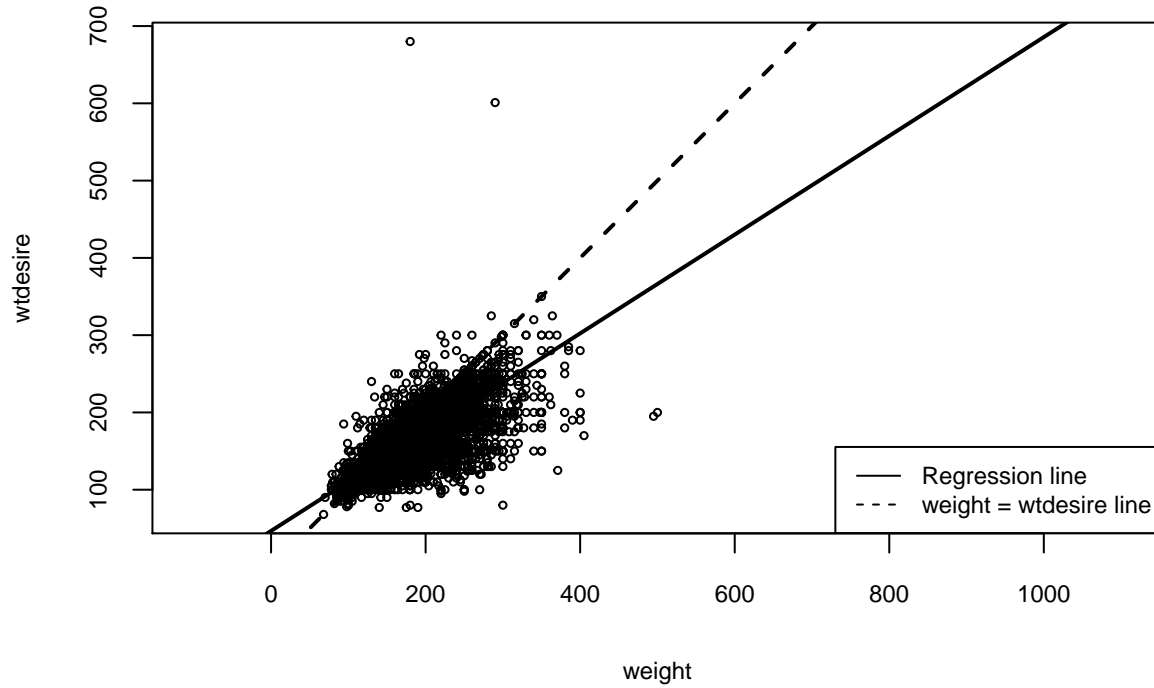
(A) Please circle the outliers in the scatterplot. Do you believe that these outliers will affect an analysis of these data? Why or why not? (3 pts)

Question 2 (continued)

(B) R gives the following output for the regression line: (2 pt)

```
coef( lm(wtdesire ~ weight, data=cdc ) )
```

```
## (Intercept)       weight
##   46.6640146    0.6390143
```

If the regression line is $Y = a + bX$, what is the value of a and what is the value of b?

(C) Predict how much someone who weighs 100 lbs would like to weigh. (2 pts)

(D) Predict how much someone who weights 200 lbs would like to weigh. (2 pts)

(C) The graph implies that people who weigh less than 100 lbs wish to, on average, gain weight, while people who weigh more than 200 lbs wish to, on average, lose weight. According to the regression line, what is the critical weight where people are, on average, satisfied with what they weigh? (8 pts)

(HINT: If you are satisfied with your weight, Y=X.)

(D) The regression line calculated in part (B) used the whole 20,000 person sample. If we calculated a regression line for **only** the men, would you expect the answer to part (C) to change? Why or why not? (3 pts)

Question 3 (16 pts)

According to the information that comes with a certain prescription drug, when taking this drug, there is a 20% chance of experiencing nausea (N) and a 50% chance of experiencing decreased sexual drive (D). The information also states that there is a 15% chance of experiencing both side effects.

(A) What is the probability of experiencing only nausea? (2 pts)

(B) I am currently experiencing nausea. I do not like it. Given that I am nauseated, what is the probability that I will also experience decreased sexual drive? (4 pts)

(C) My friend is also taking the same drug. He is lucky enough to not be experiencing nausea. Given that he is not nauseated, what is the probability that he will experience decreased sexual drive? (4 pts)

(D) Consider a different drug. If, for this different drug, the probability of nausea is 20%, the probability of a rash was 30%, and the probability of both a rash and nausea occurring was 0%, would the events nausea and rash be independent? Why? (2 pts)

Question 4 (24 pts)

The Global Terrorism Database (GTD) at the University of Maryland tracks terrorists attacks throughout the world. According to GTD, there have been 525 terrorist incidents in the United States since 1994.

In June 2013, Edward Snowden leaked various classified documents about the National Security Agency's (NSA) domestic surveillance program. The short version is that the NSA is spying on **everyone** by tracking computer and cell phone usage (including your physical location via your cell phone's GPS!). The goal of the NSA data collection program is to analyze that data to find terrorists.

In this question, we will evaluate the NSA surveillance program according to the conditional probability that a person is, in fact, a terrorist, given that they were flagged by the NSA system as a terrorist.

(A) Let F be the event that the system flags someone as a terrorist. Let T be the event that a person is a terrorist. Which conditional probability will we be using to evaluate the NSA system: P(F|T), P(T|F), P(not F|T), or P(not F|not T)? (1 pt)

(B) Assume there are 1,000 terrorists currently active in the United States. If we increased the number of terrorists in the US by a factor of 10 (now 10,000 terrorists), would we expect the NSA program to work better, the same, or worse? Why? (No calculations are required for this question, just reasoning.) (2 pts)

(C) Since most terrorist acts are committed by individuals, it is likely that the average number of people involved in planning a terrorist incident is no more than 10. If there are 10 terrorists per terrorist attack, 525 terrorist attacks in the last 30 years, 300 million people in living in the US, and *everyone* who once was a terrorist is still an active terrorist, what is the probability that a randomly selected person living in the United States is a terrorist? (1 pts)

(D) Anomalous detection systems are notoriously difficult to design. For the sake of argument, let's assume that the NSA detection system is **unreasonably good** at predicting whether or not someone is a terrorist. Let the probability that a terrorist is flagged by the system be 99.9% and the probability that a person who is not terrorist is flagged by the system is 0.08%. What is the probability that a person is a terrorist given that they are flagged? (20 pts)

Question 5 (20 pts)

Using the information presented in

> McCluney, K. E., Date, R. C. (2008) The effects of hydration on growth of the house
> cricket, Acheta domesticus. *Journal of Insect Science*, 8(32). Retrieved from http://www.
> insectscience.org/8.32/i1536-2442-8-32.pdf

we will model the mass of a common household cricket as normally distributed with a mean of 50 grams and a standard deviation of 5 grams.

Let us assume that a local pet store, Stat Pets, has managed to breed crickets with a (normally distributed) mass of 60 grams and a standard deviation of 2 grams.

(A) Suppose that my 5 year old son loves to catch crickets to feed to our neighbor's pet snake. If, every day, the neighbor's pet snake gets to eat 2 crickets from Stat Pets and 3 (common household) crickets caught by my son, what is average weight of the snake's daily meal? (5 pts)

(B) What is the standard deviation of the snake's daily meal (assuming that weights of the two types of crickets are independent)? (5 pts)

(C) 99.97% of the time the snake's meals fall between which two weights? (2 pts)

(D) If the snake needs at least 230 grams of crickets per day to not be hungry, how many days out of a year does the snake feel hungry? (Assume 365 days per year) (8 pts)