NAME: ______________________

STUDENT ID (if any): ______________________

GSI (Daniel or Partha): ______________________

SECTION TIME: ______________________

Stat 2 midterm
10:00 a.m., Friday 18th July, 2008
Time allowed: 60 minutes

Answer all TWENTY questions. All questions are of equal value (two points).

Please keep written answers brief.

You may refer to ONE double-sided sheet of A4 paper with hand-written notes.

You may refer to a standard normal table.

Unless otherwise stated, you may leave numerical answers as fractions, decimal or percentages (where applicable). Show as many decimal places as is necessary to show your understanding.
A researcher wishes to test the effectiveness of an influenza vaccine. A group of patients was randomly divided into a treatment group, which was given the vaccine, and a control group, which received a placebo. Neither the patients nor the nurses and doctors administering the vaccine knew who received the real vaccine and who received the placebo. After 3 months, 60 of 300 patients in the control group had developed influenza, compared to 10 of 300 in the treatment group.

Question 1: Why are the nurses and doctors not told which patients are in each group?

Question 2: From this study alone, can you conclude that the vaccine prevents influenza? If not, why not?
Answer question 3 based on the following histogram of average annual income for a certain population:

Question 3: What percentage of this population earn between $100000 and $200000 dollars per year?

Question 4: Find the SD (not the SD+) of the following list of numbers:

3 4 5 6 7
**Question 5:** I measure the heights of a group of children using a tape measure. A friend says my measurements will be inaccurate because the tape measure is old and has stretched. I reply that I will solve this problem by measuring each child several times and taking the average. Will this solve the problem? Why or why not?

---

**Answer question 6 based on the following data:**

The following table gives the height and weight of Cal football quarterbacks:

<table>
<thead>
<tr>
<th>Height (inches)</th>
<th>Weight (pounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>233</td>
</tr>
<tr>
<td>77</td>
<td>229</td>
</tr>
<tr>
<td>74</td>
<td>224</td>
</tr>
<tr>
<td>73</td>
<td>210</td>
</tr>
</tbody>
</table>

The heights have mean 75.25 inches and SD 1.785 inches. The weights have mean 224 pounds and SD 8.689 pounds.

**Question 6:** What is the correlation between height and weight for these quarterbacks?

---

I calculate two correlations:

(i) Correlation between height and weight for all college football players
(ii) Correlation between average height and average weight of football players at each college that plays football

**Question 7:** Which would you expect to be larger, (i) or (ii)? Why?
Answer questions 8, 9 and 10 based on the following paragraph:

The length of tiger claws is normally distributed with mean 2.5 inches and SD 0.24 inches. The length of tiger paws is normally distributed with mean 24 inches and SD 3.6 inches. The scatter plot of length of tiger paws against length of tiger claws is football-shaped. The correlation between length of tiger claws and length of tiger paws is 0.8.

**Question 8:** What is the upper quartile of tiger paw lengths?

**Question 9:** What do you predict the paw length of a tiger with a 3.0 inch-long claw will be?

**Question 10:** Assuming a normal distribution, what is the upper quartile of tiger paw lengths for tigers with 3.0 inch-long claws?
Answer question 11 based on the following plot of the residuals of a linear regression:

**Question 11:** For this regression, do you think the regression line is a good model of the data? Why or why not?

**Question 12:** I shuffle a deck of cards and deal two cards without replacement. What is the probability that the first card is the queen of spades and the second card is also a spade?
**Question 13:** I shuffle a deck of cards and deal two cards without replacement. What is the probability that the first card is the queen of spades or the second is a spade?

---

*Answer question 14 based on the following paragraph:*

I wish to study the effect of alcohol on liver health. To do this, I gather liver health and other information on a group of alcohol drinkers as well as a group of non-drinkers. Five years later, I check back with the participants to see if liver health and other variables have changed.

**Question 14:** Multiple choice: is this a  
(a) Controlled experiment  
(b) Uncontrolled experiment  
(c) Cohort observational study  
(d) Case-control observational study  
(e) Cross-sectional study

---

**Question 15:** A certain population has mean weight 160 pounds with SD 16 pounds. The weights are not normally distributed. What is the maximum possible percentage of this population who weigh over 200 pounds?

---

**Question 16:** A certain population has mean height 67 inches with SD 3 inches. The heights are normally distributed. I randomly select two people from this population. What is the probability that one of the two people is over six feet, while the other is under five feet?
Question 17: In a 50 question multiple choice test, the number of correct answers and the number of incorrect answers add up to 50. The correlation between GPA and number of correct answers in the test is 0.3. What is the correlation between GPA and the number of incorrect answers? (If it is impossible to give an exact number, give a range.)

Answer question 18 based on the following plot of the residuals of a linear regression:

![Residuals for regression of y against x](image)

Question 18: Are the assumptions of the Gaussian linear model met? Why or why not?
**Question 19:** In one card game, a ace is worth four points, a king is worth three points, a queen is worth two points, and a jack is worth one point. I deal you three cards. What is the probability that you have at least ten points in total (give this answer to four significant figures)?

**Question 20:** What is the minimum number of cards I must deal from a deck (without replacement) before I have a greater than 50% chance of getting at least one queen?