Name: _____ **Final - Part 2 (120 pts)**

For all of the questions below, please show enough work that it is completely clear how your final solution was derived. The part of the exam is open everything. Your solutions can take up more than the page they are on, but **please leave the page breaks in the document**. The page breaks make it much easier to follow everyone's solutions.

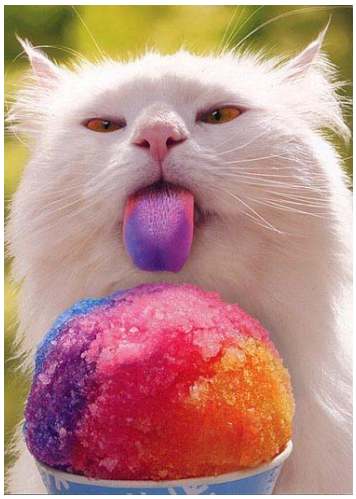

image used without permission from http://ffffound.com/image/7361224f4dd713a39e2aaaff36f190d5258ae960

*1 - 4 pts) Sally decides to see if she can make her cat's saliva non-neutral, i.e. acidic, pH < 7, or basic, pH > 7, by putting her beloved pet on a snow cone diet for a week (see figure to left). At the end of the week she measures the pH of the cat's saliva 10 times. After removing an outlier, she analyzes the following 9 measures.

pH: 5.2, 5.6, 5.8, 6.4, 6.5, 6.8, 6.9, 7.2, 7.5
sample mean = 6.4333
sample sd = 0.7665
sample size, n = 9

Write the null and alternative hypothesis for a **two-sided** test of this question. Remember to use Greek letters for parameters.

*2 - 8 pts) Referring to question 1, **calculate an appropriate test statistic to three decimal places**. Remember, don't round your numbers during calculations. Round at the very end to get three decimals.

*3 - 4 pts) Referring to questions 1-2, what is the appropriate **critical value for a 5% significance level two-sided test**? Provide critical value to 3 decimal places.

*4 - 4 pts) Referring to question 1-3, what is an appropriate **p-value for this test**?

*5 - 8 pts) Referring to questions 1-4, calculate a **95% confidence interval** for the mean pH of Sally's cat's saliva after exposure to the snow cone diet. Write your answer to three decimals.

*6 - 3 pts) Referring to question 1, the outlier Sally dropped was a 2.1. Here is the data with the outlier included and corresponding R output. (The question is at the bottom of the page.)

```
> pH <- c(2.1, 5.2, 5.6, 5.8, 6.4, 6.5, 6.8, 6.9, 7.2, 7.5)
> mean(pH)
[1] 6
> sd(pH)
[1] 1.549193

> t.test(pH, mu=7)

        One Sample t-test

data:  pH
t = -2.0412, df = 9, p-value = 0.07162
alternative hypothesis: true mean is not equal to 7
95 percent confidence interval:
 4.891774 7.108226
sample estimates:
mean of x
        6

> wilcox.test(pH, mu=7)

        Wilcoxon signed rank test with continuity correction

data:  pH
V = 7, p-value = 0.04123
alternative hypothesis: true location is not equal to 7
```

If analyzing all 10 points of data, which test should Sally use and why?

*7 - 4 pts) In a one-sample case, the Wilcoxon signed rank test is for testing the null hypothesis, Ho: F(X) is symmetric about μ_o, which is often simplified to the easier to understand though technically inaccurate Ho: median(X) = μ_o, where μ_o depends on the context of the problem. In the context of question 6, what is an appropriate **null and alternative hypothesis** for the problem?

*8 - 3 pts) Referring to questions 6-7, write a brief **conclusion** for the analysis of all 10 data points using a 5% significance level. **Include an appropriate p-value** in your conclusion.

Type -- Color	**Blue**	Red	Orange	Yellow	Green	Brown	**Total**
Milk chocolate	**86**	51	76	65	57	70	**405**
Dark	**57**	50	95	48	99	68	**417**

image used without permission from http://www.stephengates.com/Blog/uploaded_images/unsignedUser_662.unsignedChar_1170384615063-779906.png

*9 - 8 pts) During class we collected the following M&Ms data. Create a **95% confidence interval for the risk difference (RD)** for the risk of blue for Milk Chocolate M&Ms minus the risk of blue for Dark M&Ms. Write your answer to 3 decimal places, remembering to only round your numbers after you're done with all your calculations.

*10 - 10 pts) Referring to the data in question 9, create a **95% confidence interval for the odds ratio (OR)** for the odds of blue for milk chocolate M&Ms versus the odds of blue for dark M&Ms. Write your answer to 3 decimal places, remembering to only round your numbers after you're done with all your calculations.

image used without permission from http://www.theiphoneblog.com/images/stories/2009/03/app_store_church_lady.jpg

*11 - 4 pts) In order to ensure that patient examinations are of high quality, a hospital employs a team of reviewers to do routine chart reviews on samples of exams. This team of reviewers rate the examinations as passing or failing. Two new reviewers, Pat and Terry, are being testing on their rating accuracy. They both rate 300 exams for which a gold standard rating has been done. Thus, it is known if their ratings are correct for these 300 exams. The results follow.

	Pat Correct	Pat Incorrect
Terry Correct	200	48
Terry Incorrect	32	20

Given that Pat and Terry are both rating the same exams, what is an appropriate **statistical test** with which to test whether Pat or Terry is the better reviewer? **Provide the test's name**.

12 - 10 pts) Calculate a **p-value for a two-sided test of the question in problem 11. Write your answer to 4 decimal places, remembering to only round your numbers after you're done with all your calculations.

**13 - 10 pts) We saw in question 12 that reviewers sometimes give incorrect ratings. In an effort to improve the accuracy of the ratings, the hospital uses the following system. An examination is rated by two reviewers. If the two reviewers agree (both rate pass or both rate fail) then the agreed upon rating is given. If the two reviewers disagree (one rates pass and the other fail), then a third reviewer reviews the examination with their rating breaking the tie.

Suppose the probability of a single reviewer correctly rating an exam is 70% for all reviewers. What is the **probability that the hospital's method will yield the correct exam rating**? Give your answer to three decimal places. Think through all of the scenarios carefully.

14 - 10 pts) Sally's parents form a mini IRB and tell their daughter that any of her future experiments on her pets need to be sufficiently powered. Sally wants to redo her snow cone experiment from question 1. If she assumes that the pH of her cat's saliva after the diet will be normally distributed with a mean of 6.5 and a standard deviation of 1, how many pH measures will she need to take to have her study powered at 80%? In other words, **what sample size will yield a power of 80%? Assume Sally will be doing a two-sided test at a 5% significance level with the assumed standard deviation treated as known.

***15 - 10 pts) Suppose you have two normally distributed samples, both of sample size n, as follows.
X1, X2, …, Xn ~iid Normal(μ=10, σ=1)
Y1, Y2, …, Yn ~iid Normal(μ=50, σ=1)

Consider performing a two-sided t-test on X vs Y and a two-sided Wilcoxon rank sum test on X vs Y. What is the **smallest sample size n such that the t-test p-value and the Wilcoxon test p-value will agree to four decimal places with near certainty**? In other words, choose the smallest n such that $P(|\text{p-value}_{\text{t-test}} - \text{p-value}_{\text{wilcoxon}}| < 0.0001) \geq 0.999999$. Include proof that your n is the correct answer.

***16 - 5 pts) Using Calculus, show that the maximum likelihood estimator for the proportion of blue M&Ms from a sample of size n with x blue M&Ms observed is equal to x/n, i.e. **show that x/n is the maximum likelihood estimator of θ**.

***17 - 5 pts) Referring to question 16, when using a beta prior distribution, B(α, β), show that the **mean posterior estimator for the proportion of blue M&Ms is equal to x/n** only when α = 1 and β = 1.

*18 - 3 pts) Referring to question 9, create a **95% confidence interval for the relative risk (RR)** for the risk of blue for Milk Chocolate M&Ms versus the risk of blue for Dark M&Ms. Write your answer to 3 decimal places.

*19 - 3 pts) Referring to question 9, create a **1/8th support interval for the relative risk (RR)** for the risk of blue for Milk Chocolate M&Ms versus the risk of blue for Dark M&Ms. Write your answer to 2-3 decimal places.

*20 - 3 pts) Referring to question 9, create a **95% credible interval for the relative risk (RR)** for the risk of blue for Milk Chocolate M&Ms versus the risk of blue for Dark M&Ms using independent Beta(1, 1) prior distribution for the proportion of blue M&Ms in Milk Chocolate and Dark Chocolate. Write your answer to 2-3 decimal places.

*21 - 1 pts) Referring to questions 18-20, how do the intervals for the RR from the three statistical paradigms compare to each other?