Assignment 2

This assignment will give some practice with interpreting studies and computing and interpreting descriptive statistics

1. (Taken from Bland and Peacock) Gardosi compared giving birth in the conventional recumbent manner versus upright squatting using a birth cushion. All participants were full term mothers with a singleton pregnancy and no known antenatal risk factors for complicated delivery. Women were initially randomized to the recumbent versus upright treatment with 209 assigned to the recumbent treatment versus 218 randomized to the upright treatment. However, 22 women in the recumbent group spontaneously requested and were allowed to have a cushion and deliver in the upright position and 39 assigned to the upright position were comfortable lying down and delivered in the recumbent position.

The authors compared the two groups by intent to treat, that is, by their original randomized assignment. The table below shows a primary outcome, whether the delivery required forceps or not, by intent to treat

	Upright	Recumbent	
Num forceps Deliveries (pct)	19/218=8.7%	34/209=16.3% (plus 2 others required C section)	

1a. This result implies that Upright has fewer forceps deliveries. In a few sentences, briefly discuss the differences between intent to treat and on-treatment analysis. What is the directionality in the biases they have, if any? For example - if no one had changed groups, would we expect to see a bigger or smaller difference between upright and recumbent? Does treatment group "crossover" cause the groups to appear to be more similar or more distinct?

1b. Can this study be blinded and does this matter? Briefly discuss.

2. The table below shows the mean and SD for duration of the second stage of labor from this study

second stage of labor duration in minutes

	Upright	Recumbent	
Median	31	45	
Mean	39	50	
SD	26	29	
IQR	23-50	28-67	(Interquartile range)

2a. Compute the mean difference and the corresponding SD of the difference.

2b. Do these results imply that there is little or no overlap in the distribution of second stage duration between the two groups? Hint, if there was no overlap, would all the differences be the same "sign" (positive or negative)?

2c. Do the results imply that the distribution of duration time may be well modeled by a symmetric Gaussian (normal bell curve)? Explain why or why not or why one cannot determine this based on the table. Is a mean or median the best choice as a measure of "typical" duration or is this not relevant?

3. The data below compares TB mortality in rats given a placebo vaccine versus rats given a live agent vaccine in a randomized trial

Group	n	num dead	mean follow up in days
Placebo	25	10	7
Agent	25	15	30

3a. Based on statistics that you can derive from this data, which treatment has a lower mortality rate? (Show work)

3b. In another randomized trial comparing the same two treatments in a new set of rats, survival after 14 days was reported as below. Note that the follow up times were more than 14 days for some or all animals in each group and some animals may die after 14 days. Some may have also died before 14 days. Follow up did NOT end at 14 days. You may assume that the mortality (hazard) **rate** in each group is a constant.

Group	n	pct alive at 14 days (not last follow up time)
Placebo	50	45%
Agent	50	76%

Overall, do the findings of this second study seem to agree or disagree with the findings of the previous study?

4. The hazard rate ratio for post treatment breast cancer (non) survival is HR=2.0 in those who drink more than 4 oz of alcohol per day compared to non drinkers. If the 12 month survival in non drinkers is 90%, what is the survival in alcohol drinkers?

5. Below are serum log 10 IgG (immunoglobulin) titers in children with and without Pertussus.

Titers in those with no Pertussus
0.41, 0.43, 0.47, 0.51, 0.59, 0.67, 0.88, 0.88, 0.91, 0.92, 0.97, 1.06, 1.24, 1.27, 1.27

Titers in those with Pertussus
0.93, 0.95, 1.01, 1.13, 1.17, 1.18, 1.21, 1.22, 1.22, 1.23

We wish to use the serum log IgG level as a medical test for Pertussus. We defined a log IgG level of 1.1 as the threshold. That is, patients with log IgG > 1.1 test "positive" for Pertussus and those < 1.1 test negative.

Based on the above, compute the estimated sensitivity and specificity of log IgG. You do not have to make any distribution assumptions.

If this data is from a case-control study, should you use this data to compute the positive and negative predictive values? If so, do so. If not, briefly explain why.

JMP data assignment 2

A "fat" dataset has been given where "TFAT" is total dietary fat in grams and "PFAT" is the percent of body weight that is composed of fat. There are two times, time 0 baseline and 36 week follow up. There are two treatment groups. Group 1 was given a dietary education and intervention and Group 2 is a control group.

1a. Using JMP, compute and report the sample size, mean and SD for the following groups:

a) TFAT0 Group 1 (treated at baseline)
b) TFAT0 Group 2 (control at baseline)
c) TFAT36 Group 1 (treated at 36 weeks)
d) TFAT36 Group 2 (control at 36 weeks)

1b. Using this data from 1a, manually calculate the mean difference and the SD of the differences for TFAT0 versus TFAT36 in each group. Assume that the subjects in TFAT0 and TFAT36 are NOT the same people.

2. Repeat these computations if TFAT0 and TFAT36 are from the same persons measured at two different times (baseline and after 36 weeks of dieting). How can one do this in JMP?

Is there any difference in the results if this is data on the same subjects measured twice instead on two sets of subjects measured once?

Be sure to report tables of the results that have the appropriate titles and units (grams).

3. For TFAT0, determine using JMP if the distribution of TFAT0 on the original scale or on the log scale (ie log TFAT0) is closer to following a normal (Gaussian) distribution. A normal quantile plot diagnostic and goodness of fit test (not covered in class yet) can be reported although you may skip the goodness of fit test if we have not yet covered it.