NAME: **Total number of Marks: /25**

There are 6 questions in this paper, do not be deterred, they are all straightforward. Read each question carefully. There are questions on <u>both</u> side of the page. The number of marks for each question are given in brackets. Be smart about how you answer. If you can't answer one question move on the to next and return to the questions you could not do after answering all the other questions! There are 4 Figures in this paper.

Rubric

(1) Do each test at the 5% level.

(2) No marks will be given for answering only reject null or don't reject null. Explain your conclusion by stating your answer with respect to the p-value or rejection region.

(3) If you need to do a calculation then give it.

(4) This exam is a closed book exam, but you can use a 4-sided cheat sheet, normal, t-tables, Wilcoxon Tables (5 and 6) and a calculator.

Write your solutions in the question paper.

GOOD LUCK!

(1) Suppose that over the past 10 years the ratio of males to females is 2:3 every year (in other words 40% male and 60% female).

Every year a course prize is given and for the past 10 years **only females** have won this prize. Calculate the probability that this happened by random chance. You can use the probabilities given in Figure 1. [2]
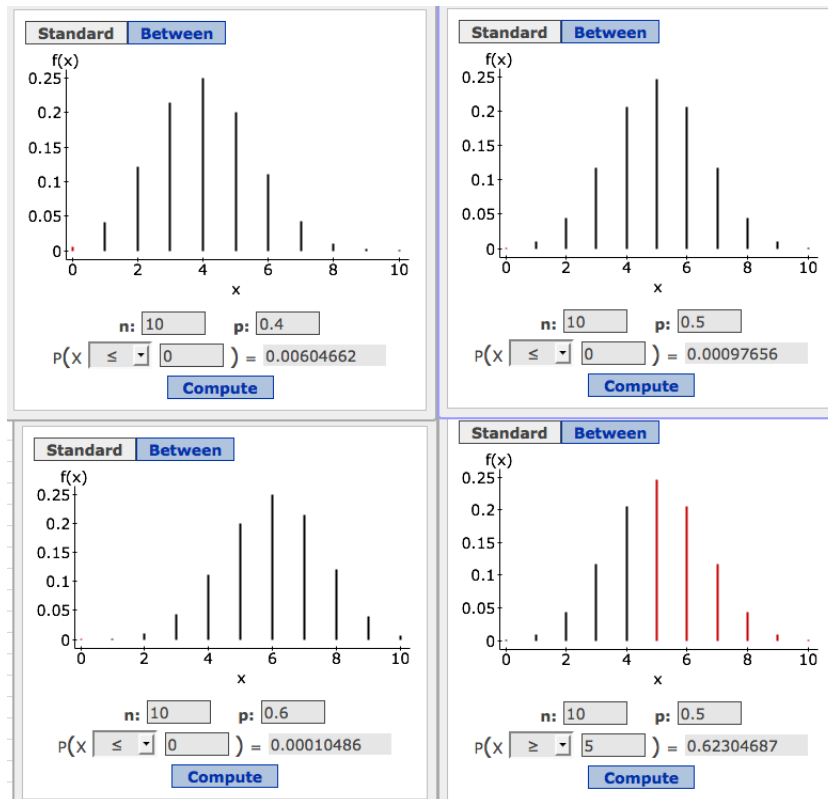

Figure 1: Binomial probabilities

(2) The income per household in the US is known to be heavily right skewed. The average is known to be 80K, the median income is 60K and the standard deviation is 30K.

    (a) Using these figures, plot the distribution of household income.   [2]

    (b) Suppose a 100 households are randomly sampled. The average is taken. Make a plot of the distribution of the average (based on a sample of 100), using the figures given above (remember to indicate on your plot the mean and standard error).   [2]

(3) The amount of sap (in a season) produced by a maple tree is known to be normally distributed with mean $\mu = 45$ litres and standard deviation $\sigma = 4$ litres.

(a) A farmer has 100 maple trees on her farm. She wants to know the average (sample mean) amount of sap her 100 trees will produce this year. Construct an **interval** centered about the population mean ($\mu = 45$), where she can be 99% certain that this years average (sample mean) of sap will lie. [3]

(b) Use your answer in (a) to construct an interval where she can be 99% certain that this years **total** sap will lie. [2]

(4) **25** households were randomly sampled and asked how many people lived in their house. The summary statistics is given in Figure 2
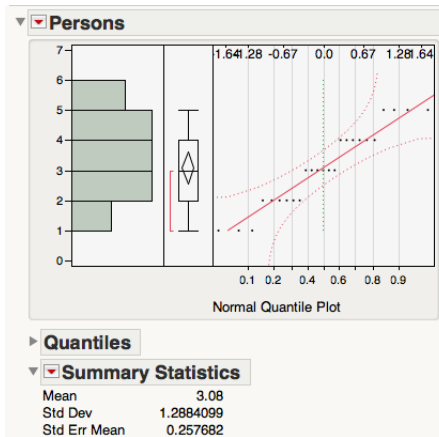
Figure 2: Summary statistics of the outcome of 25 households.

(a) Use the t-distribution to calculate the probability that the average in the sample is greater than 3.08 <u>given</u> that the population mean (number of people in a household) is 2.5.

Note: You may not be able to give the exact probability but give the best bounds that you can using the t-tables.

(b) Comment on the reliability the probability calculated in (a) given the distribution of the data.

(5) Hops is a herb that is generally believed to aid sleep. To see whether there is any truth in this a random sample of 30 patients with sleep problems was taken. They were partitioned into two groups of 15.

Group 1 is given Hops before went to bed and the time taken to go to sleep was noted.

Group 2 is given a placebo before went to bed and the time taken to go to sleep was noted.

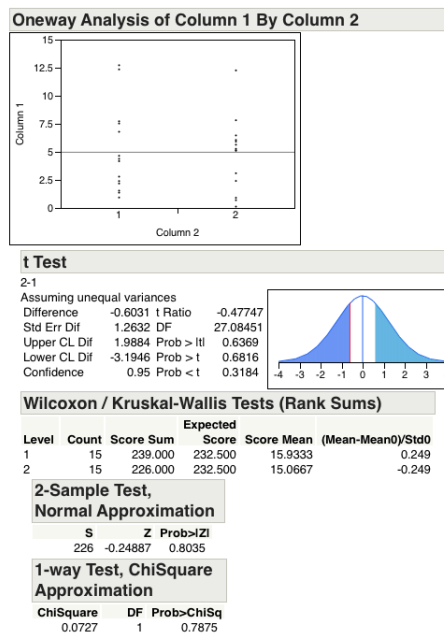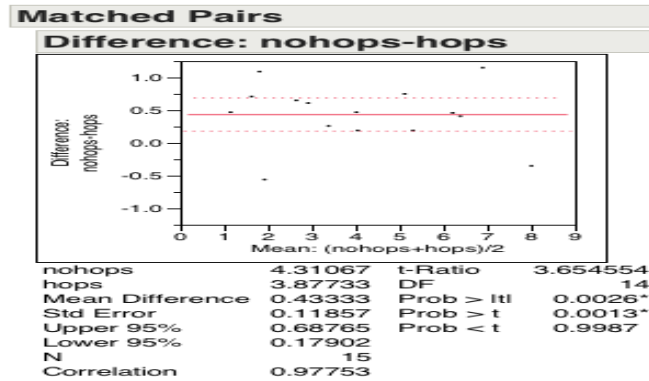The output from an independent sample t-test is given in Figure 3.

Oneway Analysis of Column 1 By Column 2

t Test

2-1
Assuming unequal variances

Difference	-0.6031	t Ratio	-0.47747
Std Err Dif	1.2632	DF	27.08451
Upper CL Dif	1.9884	Prob > |t|	0.6369
Lower CL Dif	-3.1946	Prob > t	0.6816
Confidence	0.95	Prob < t	0.3184

Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
1	15	239.000	232.500	15.9333	0.249
2	15	226.000	232.500	15.0667	-0.249

2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
226	-0.24887	0.8035

1-way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
0.0727	1	0.7875

Figure 3: Summary statistics of the outcome of 25 households.

(a) We want to see whether hops reduces the time needed to fall asleep. State the hypothesis of interest and the results of the independent sample t-test at the 5% level. [2]

(b) Your Wilcoxon sum Rank tables do not go up to $n_1 = 15$ and $n_2 = 15$. However, by comparing the numbers in **Score Sums** what do you believe is the outcome of the Wilcoxon sum rank test, giving a reason for your answer? [2]

(6) In the previous question the researchers used two different groups of individuals. This time they decided to use the same set of patients for both the placebo trial and the hops trial.

pre	4.25	5.40	6.59	3.51	4.14	7.83	7.47	1.38	1.64	1.98	2.97	2.35	6.44	5.49	3.22
hop	3.78	5.21	6.18	3.25	3.95	8.18	6.32	0.91	2.20	1.27	2.32	1.26	5.98	4.74	2.61
diff	0.47	0.19	0.41	0.26	0.19	-0.35	1.15	0.47	-0.56	0.71	0.65	1.09	0.46	0.75	0.61

Matched Pairs
Difference: nohops-hops

nohops	4.31067	t-Ratio	3.654554
hops	3.87733	DF	14
Mean Difference	0.43333	Prob > |t|	0.0026*
Std Error	0.11857	Prob > t	0.0013*
Upper 95%	0.68765	Prob < t	0.9987
Lower 95%	0.17902		
N	15		
Correlation	0.97753		

(a) Explain why this study may be better suited to detecting a difference (if there exists one) than the study described in question (5)? [2]

(b) We want to see whether hops reduces the time needed to fall asleep. Stating precisely the null and alternative and do the test at the 5% level. [2]

(c) A Wilcoxon Sign Rank test is done, with $T_- = 13$ and $T_+ = 107$. What are the results of the Wilcoxon Sign Rank test? [1]