Solutions

STAT 651 Midterm Test (60 minutes) - 10:20am - 11:20am November, 2014

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 both 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]

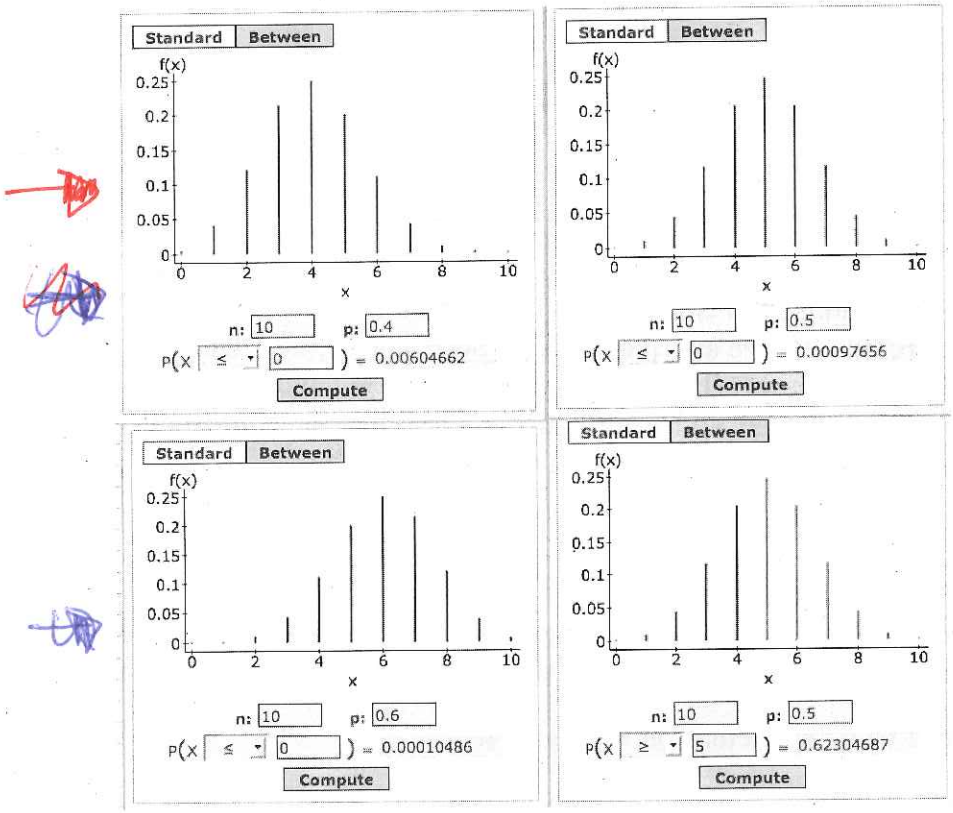


Figure 1: Binomial probabilities

Probability no males get the prize is

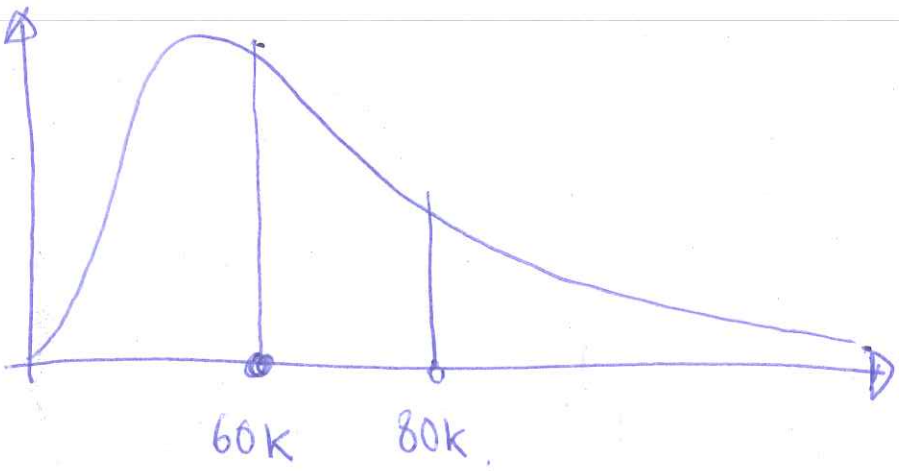
$(0.6)^{10} = 0.006$

Alternatively you can use the plots above.

The probability a male does not get a prize in a given year is a Bernoulli trial, with probability of getting the prize 40% and probability of not getting the prize 60%. $P(\text{male}) = \$_n \sim Bin(10, 0.4)$

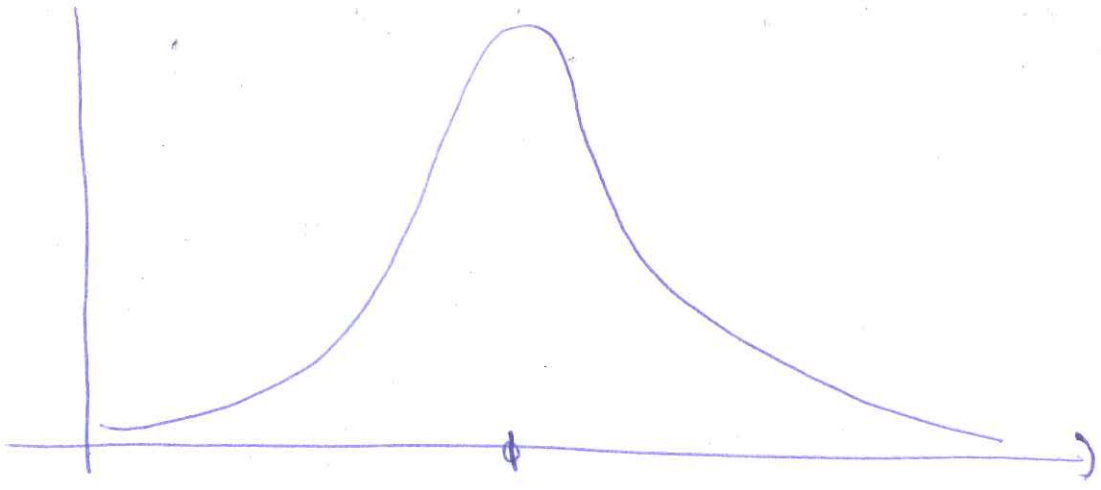
(2) The income per household in the US is known to be heavily right skewed. The average is known to be 80K, the median income is 60K and the standard deviation is 30K.

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



sd = 30

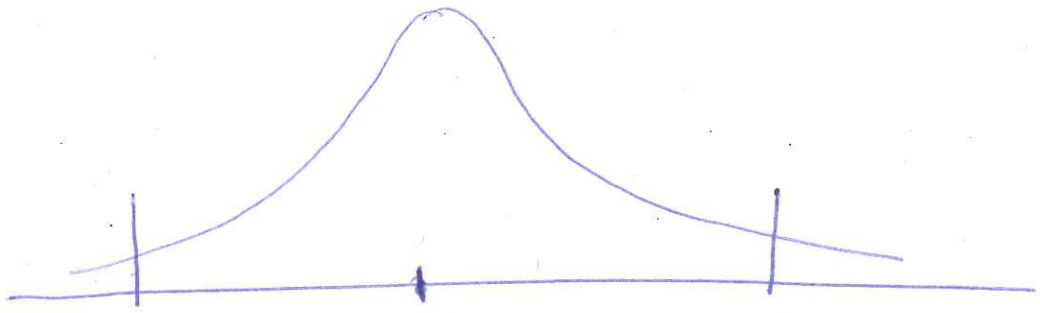
(b) Suppose a 100 households are randomly sampled. The average is taken. Make a plot of the distribution of the average (based on a sample of 100), using the figures given above (remember to indicate on your plot the mean and standard error). [2]



$S.e = \frac{30}{\sqrt{100}}$

(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]



$45 - 2.57 \times \frac{4}{\sqrt{100}}$ $\quad \mu = 45$ $\quad 45 + 2.57 \times \frac{4}{\sqrt{100}}$

$$\left[45 - 2.57 \times \frac{4}{10},\ 45 + 2.57 \times \frac{4}{10}\right]$$

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

$$100 \times \left[45 - 2.57 \times \frac{4}{10},\ 45 + 2.57 \times \frac{4}{10}\right]$$

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

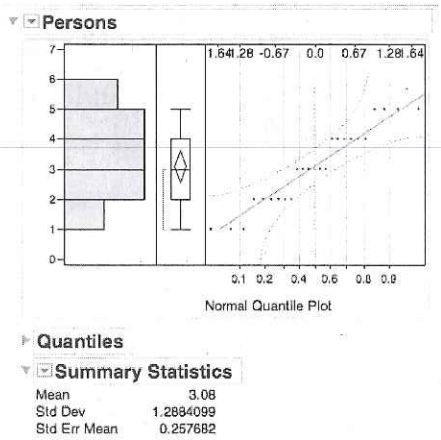
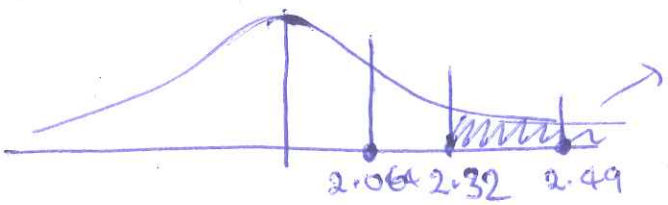


Figure 2: Summary statistics of the outcome of 25 households.

(a) Use the t-distribution to calculate the probability that the average in the sample is greater than 3.08 given 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.

$t = \frac{3.08 - 2.5}{0.25} = 2.32$ $t = 24df$



probability between 1% - 2.5%

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

The probability is based on the assumption normality of the average. The data is numerical discrete (taking integer values), which is fit for normal. The average is taken over $n = 25$, which is not enough for normality to kick in. ~~The probabil~~ Based on this the probability is not that reliable.

(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.

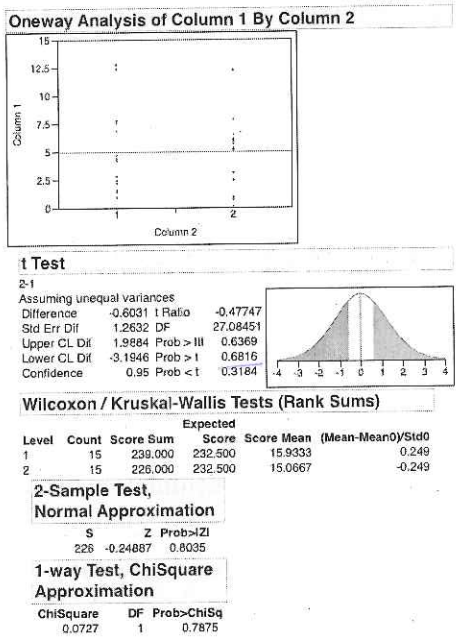


Figure 3: Summary statistics of the outcome of 25 households.

(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]

$H_0: \mu_2 - \mu_1 \leq 0 \quad H_A: \mu_2 - \mu_1 > 0$

p-value = 68% clearly we cannot reject the null.

(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]

Rank for Placebo group = 239
Rank for Placebo group = 226
→ Since If Placebo Hops was better we would expect that the rank of Hops is less than Placebo

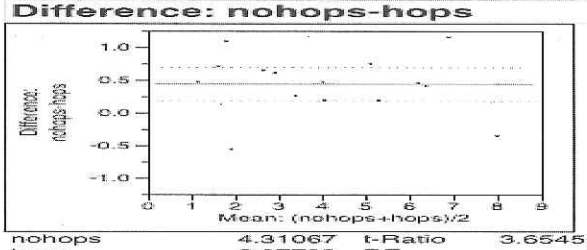
Thus (eg) This is not true. Cannot reject null.

(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]

Paired studies are better at reducing sample variation. This means smaller standard errors which means easier to reject the null.

(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]

$H_0: \mu_n - \mu_h \leq 0 \quad H_A: \mu_n - \mu_h > 0$

p-value $= 0.0013^* = 0.13\% < 5\%$

We reject the null at the 5% level

(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]

We are under alternative T_+ will be very large and T_- will be small. Choose T_-. Look adaptable $n = 15$ Threshold is 30. Since $T_- < 30$ we reject the null at 5% level.

Anand Gupta ↦ (2) ↦ $\frac{19}{25}$ (mod)