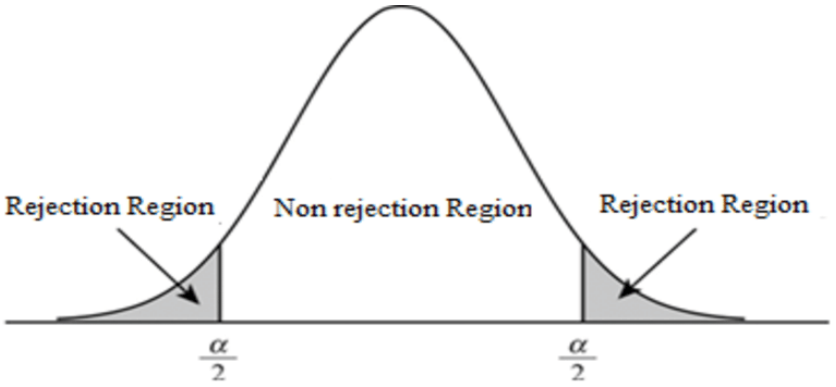
Leman Nur Erkan

COGS 536

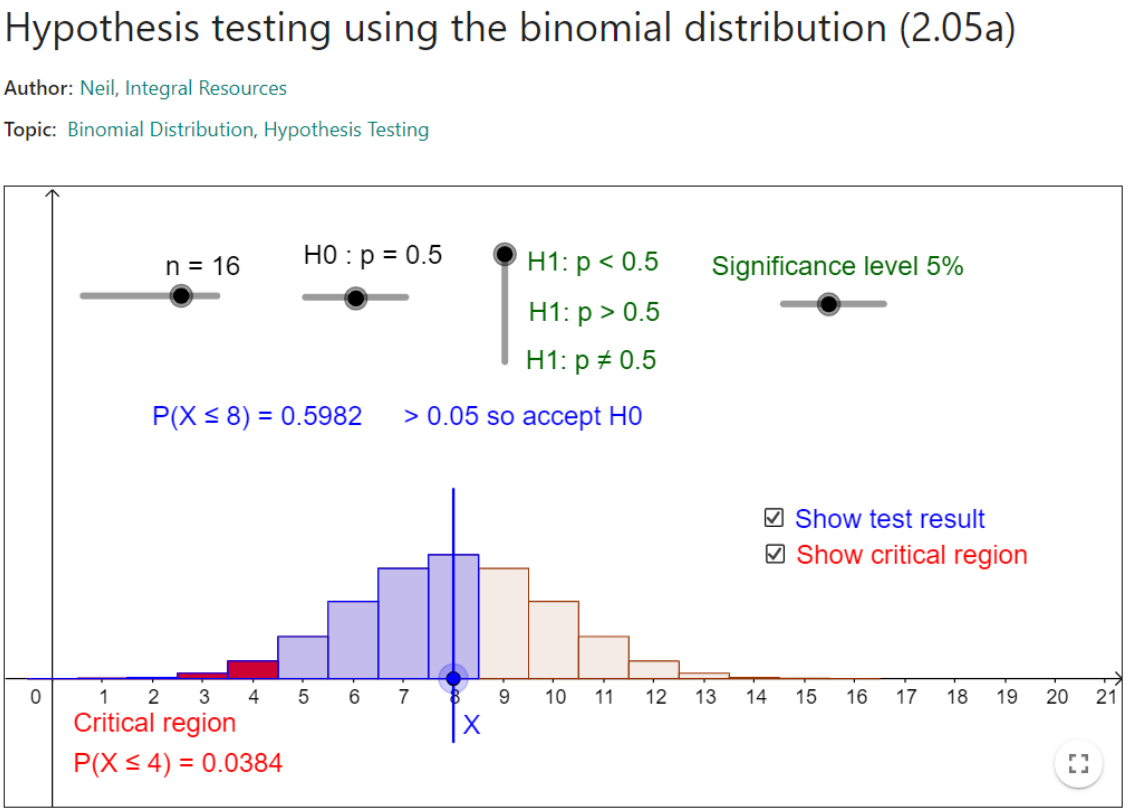
Take Home Midterm Answers

1. **a)** *iii. The histogram shows the probability of each X value when a person is just guessing.* The histogram shows the prob. of X values which is the number of correct guesses when a person is guessing the output. The y axis shows the prob. of X being true and x axis which in the interval of 0-16 - the number of drinks prepared- shows the number of drinks that are predicted. This histogram is like a normal curve which is used for representing natural events.
   b) H0= Mr. Bond cannot distinguish between a shaken Martini from a stirred Martini.

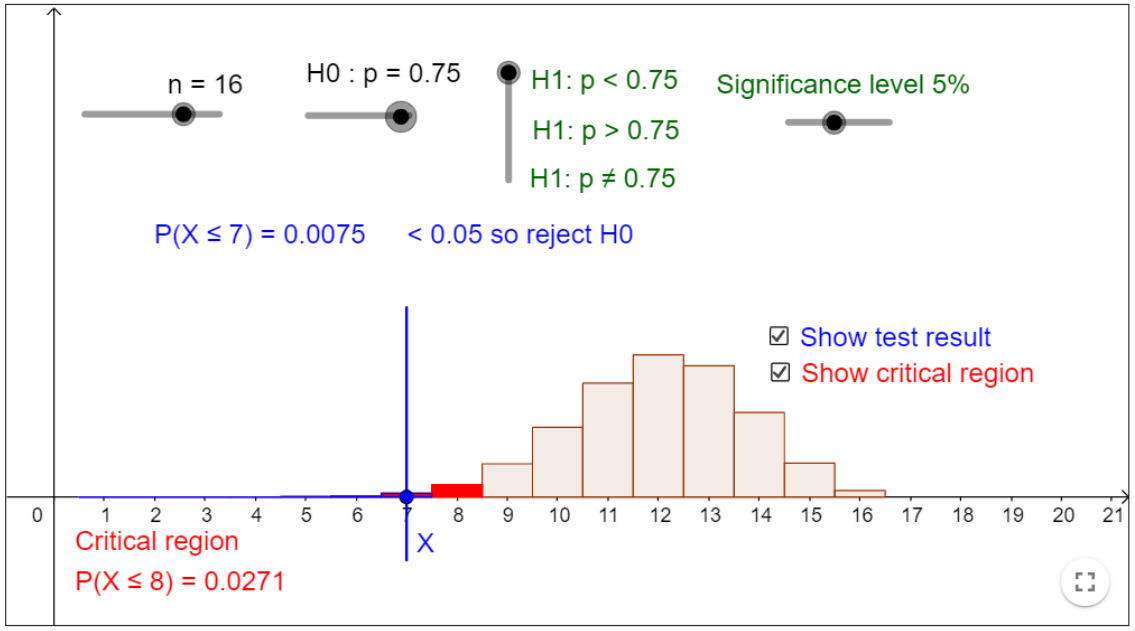


   c)
   The histogram is similar to the normal distribution curve and we can use the rejection process based on normal dist./z score. The 13 out of 16 is in the rejection region, therefore we can reject the null H and accept the HA.
   d) The histogram is similar to normal distribution curve and we can use rejection process based on normal dist./z score. The 9 out of 16 is near the mean of the histogram, which is in the non-rejection region, therefore we cannot reject the Ho and cannot accept the H alternative.
   e) Sd: 5.163977795, mean: 8, Z score=(12-8)/5.16=0.78 P(z>0.78)=.7823

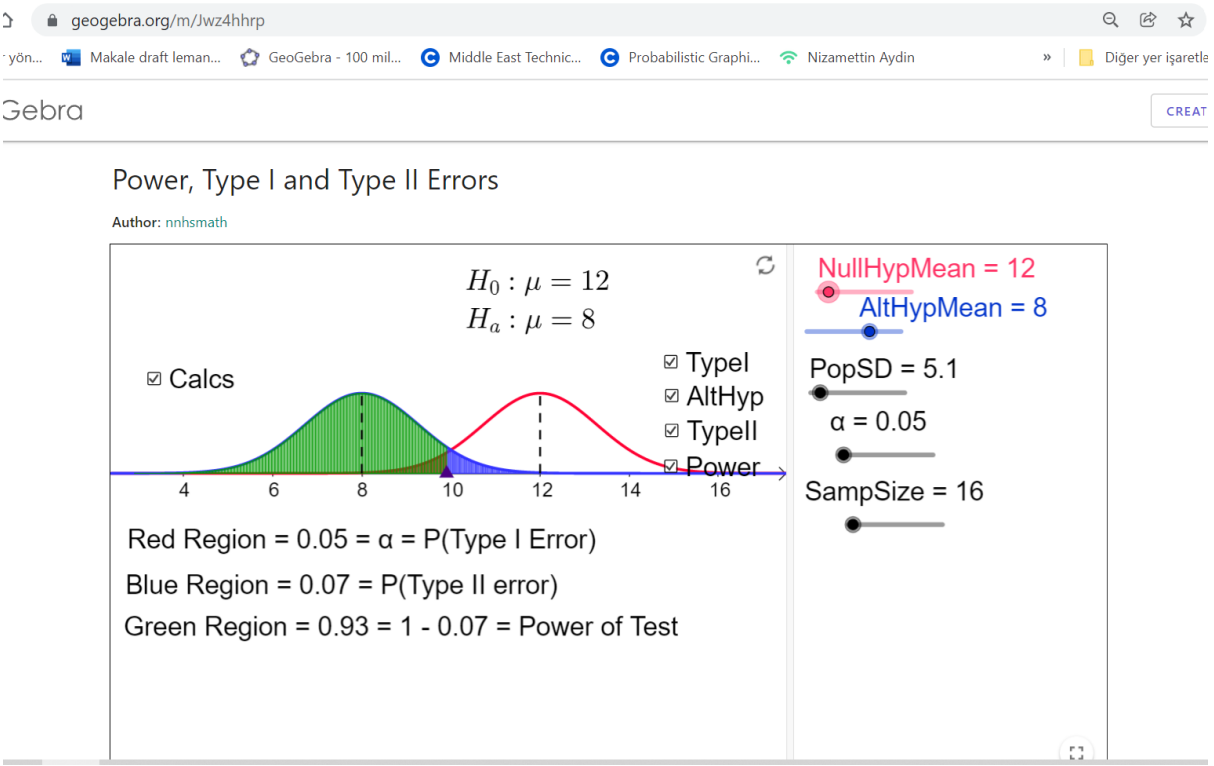
2. a)


Hypothesis testing using the binomial distribution (2.05a)
Author: Neil, Integral Resources
Topic: Binomial Distribution, Hypothesis Testing
n = 16
H0 : p = 0.5
H1: p < 0.5
H1: p > 0.5
H1: p ≠ 0.5
Significance level 5%
P(X ≤ 8) = 0.5982
> 0.05 so accept H0
Show test result
Show critical region
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Critical region
P(X ≤ 4) = 0.0384
X

b)


Hypothesis testing using the binomial distribution (2.05a)
Author: Neil, Integral Resources
Topic: Binomial Distribution, Hypothesis Testing
n = 16
H0 : p = 0.75
H1: p < 0.75
H1: p > 0.75
H1: p ≠ 0.75
Significance level 5%
P(X ≤ 7) = 0.0075
< 0.05 so reject H0
Show test result
Show critical region
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Critical region
P(X ≤ 8) = 0.0271
X

c)



***In this figure the HO and Ha means should be reversed (Ha=12,Ho=8) but it cannot be done on the app, since the calculation is not changed in this situation; I used reverse versions.

d) iii) The statistical power of the test is the probability that the test will reject the null hypothesis correctly when a given alternative hypothesis is true. in other words, detecting correctly the deviations from the H0 through the HA.

e) the power of the experiment == the sensitivity of the experiment. the sensitivity of the experiment is low. We can change the effect size by changing the sample size (as increasing), or within-sample analysis can be used.

3. a) H0: The scores on a lie-scale of children is the same as one would expect as 3.87. HA: The scores on a lie-scale of children is higher than one would expect as 3.87.
   b) 95% Confidence Interval: 4.39 ± 0.853 (3.54 to 5.24) With 95% confidence the population mean is between 3.54 and 5.24, based on 36 samples. Short Styles: 4.39 (95% CI 3.54 to 5.24) 4.39, 95% CI [3.54, 5.24] Margin of Error: 0.853.
   c) Confidence intervals represent the likely range of values of population mean. The interval computed from a given sample in part b contains the true mean (the population mean). If a 95% confidence interval includes the null value, no statistically meaningful difference between the groups. If the confidence interval does not include the null value, then we conclude that there is a statistically significant difference between the groups. Therefore, we cannot reject the H0 (3.54 < 3.87 < 5.24).
4. a) sd=4.35, mean=32, median=32: the middle one is correct since box plots represent the median which is 34, the lines represent the sd. The other two-left and right images are wrong not because of the sd and box plot type, but because of the representing the median.
   b) i. correct: the United states has buildings whose height is more than 540 but the upper bound for China is 540. ii. correct, the cumulative number of buildings that are greater than 390 meters is more than the cumulative height of the US. iii. correct.

This test is used for normality calculation for continuous distribution but the distributions in the histograms are discrete. iv. Correct Power transformation makes the distribution more gaussian(reaching the normal) and skewness is reduced when this method is used.

5. a)

ANOVA – ANALYSIS OF VARIANCE

ONE-WAY ANOVA TABLE

Treatment	Degree of Freedom	Sum of Squares	Mean Square	F-ratio
Between	$\mathbf{DF_B}$ = $B - 1$	$\mathbf{SS_B}$ = $B\ Variance^2 \times N$	$\mathbf{MS_B}$ = $SS_B \div DF_B$	$\mathbf{F}$ = $MS_B \div MS_W$
Within (Error)	$\mathbf{DF_W}$ = $DF_T - DF_B$	$\mathbf{SS_W}$ = $SS_{total} - SS_B$	$\mathbf{MS_W}$ = $SS_W \div DF_W$	
Total	$\mathbf{DF_T}$ = $N - 1$	$\mathbf{SS_T}$ = $Standard\ deviation^2 \times (N - 1)$		

SS Factor A: 369.8, SS Error :152; df Error: 76; MS Error: 2.00; F-ratio Factor A: 3, The equations is derived from the table above, therefore N is 80; Standard deviation is the square root of the variance (to find the variance of Factor A to calculate SS-Factor A). The calculation is based on the table above.

b) One way anova is used to compare two means.

c) The degrees of freedom for this entry is the number of levels minus one. Therefore it is 80.

d) For a one-way ANOVA on k groups it is needed at least k+1 participants (81).

e) The F critical value is a specific value to compare anova f-value. In general, if calculated F value in a test is larger than F critical value, the null hypothesis can be rejected, and since 2.725<3, the Ho can be rejected.