

## STATISTICS 641 - Exam 2

Total time is 120 minutes. The exam is available for 24 hours window starting from noon (CST) April 03, 2014 to 11:59 am (CST) April 04, 2014.

### Instructions

1. The exam is for two hours.
2. Students are allowed to bring two pages of cheat sheets (front&back and front&back).
3. Students are allowed to use calculator (need to use a calculator that can at least do normal probability calculation)

Name \_\_\_\_\_

Email Address \_\_\_\_\_

Please put your answers in the following table.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

## STATISTICS 641 - Exam #2

In a genetic experiment, investigators investigated 300 chromosomes of a particular type and counted the number of sister-chromatid exchanges on each chromosome.

# of exchanges	0	1	2	3	4	5	6	7	$\geq 8$	Total
Frequency	7	24	42	59	62	44	41	14	7	300

Assume that the average number of exchanges is  $\bar{X} = 3.85$  (take this value even if this is not a correct average). Answer the next two questions.

- Assuming that the number of exchanges follow a Poisson distribution, obtain the expected frequency for  $X = 1$ .
  - 24.57 •
  - 6.38
  - 1.97
  - 28.89
- What is the degrees of freedom of the Chi-square goodness of fit statistic for testing  $X$  follows a Poisson distribution?
  - 6
  - 7 •
  - 8
  - 9

Let  $W$  be the weights of 21-day-old Leghorn Chickens and let  $F$  be the cdf of  $W$ . Suppose that we have a random sample of size  $n = 20$  from  $F$ . Answer the next three questions.

156	162	168	182	186	190	190	196	202	210
214	220	226	230	230	236	236	242	246	270

- Let  $W$  follows a Normal distribution, and  $F(x)$  be the corresponding CDF. Then  $Y = F(X)$  follows
  - Gamma distribution
  - Normal distribution
  - Beta distribution
  - Uniform distribution •
- To test if  $W$  follows a Normal distribution, the best test statistic is
  - Anderson-Darling
  - Kolmogorov-Smirnov
  - Cramer-von Mises
  - Shapiro-Wilk •
- What are the empirical estimate of  $\text{pr}(W < 190)$  and its standard error?
  - 0.25 and 0.097 •
  - 0.25 and 0.433
  - 0.3 and 0.102
  - 0.3 and 0.1

6. The purpose of Box-Cox transformation is to make the distribution of the data

- (a) Log-normal
- (b) Beta distribution
- (c) Uniform
- (d) Normal •

7. Suppose that  $\hat{\theta}$  is an unbiased estimator for  $\theta$ . Hence,

- (a) the distribution  $\hat{\theta}$  must be normal with variance  $\theta/n$ .
- (b) the median of the distribution  $\hat{\theta}$  must be  $\theta$ .
- (c) the mean of the distribution  $\hat{\theta}$  must be  $\theta$ . •
- (d) the distribution  $\hat{\theta}$  must be normal with mean  $\theta$ .

Suppose that  $X_1, \dots, X_n$  are independent and identically distributed random variables drawn from the Beta distribution with mean 0.25 and variance 0.28. Answer the next two questions.

8. The distribution of the sample average  $\bar{X}$  when  $n = 16$  is

- (a) Beta distribution
- (b) Uniform
- (c) Log-normal
- (d) Normal
- (e) none of the above. •

9. The variance of the sample average  $\bar{X}$  when  $n = 16$  is

- (a)  $0.28/16$  •
- (b) 0.28
- (c)  $0.28/4$
- (d) 0.25
- (e) none of the above.

10. What is probability that the sample average  $\bar{X}$  is 0.25?

- (a) 1 as the the sample average  $\bar{X}$  is unbiased for the population average
- (b) 0 as the the sample average  $\bar{X}$  is not unbiased for the population average
- (c) 0.5 as the the sample average  $\bar{X}$  is most likely to be larger than the population average
- (d) 0 as the the sample average  $\bar{X}$  is a continuous random variable •
- (e) Can be determined only when we know the value of the sample average  $\bar{X}$

11. To estimate the location parameter of a symmetric distribution with finite variance we should use

- (a) mean or median depending on which of them follows an exact normal distribution
- (b) mean or median depending on the variance of the estimator •
- (c) use the sample median as median always exists
- (d) use the sample mean as the sample mean is always an unbiased estimator

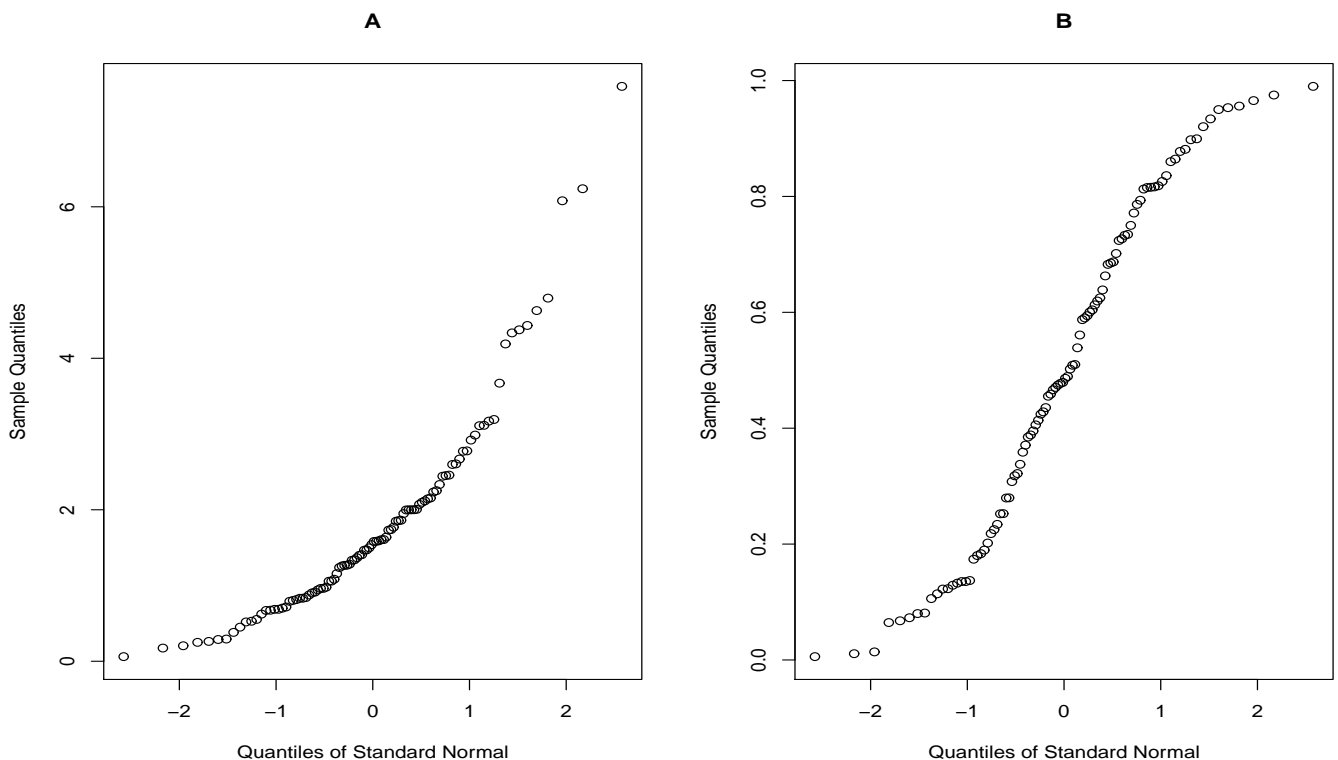
12. Suppose that a random variable  $Y$  follows the Bionomial( $n = 100, p = 0.2$ ). Obtain  $\text{pr}(X \leq 25)$  using the continuity correction.

- (a) 0.7340
- (b) 0.8697
- (c) 0.8943
- (d) 0.9154 •

Suppose that  $X_1, \dots, X_n$  are iid Bernoulli(0.28) random variables (Bernoulli(0.28) = Bionomial( $n = 1, p = 0.28$ )). Answer the next two questions.

13. When  $n = 8$ , the exact distribution of  $\bar{X}$
- (a) can be estimated via bootstrap samples. •
  - (b) is a Normal, therefore, no need to use bootstrap samples.
  - (c) is a Binomial, therefore, no need to use bootstrap samples.
  - (d) is a Uniform, therefore, no need to use bootstrap samples.
14. Suppose that we decide to apply the parametric bootstrap method to get the sampling distribution of  $\bar{X}$ . In each bootstrap sample, we draw
- (a)  $n$  independent Bernoulli observations with success probability 0.28. •
  - (b)  $n$  observations with replacement from the original sample  $(X_1, \dots, X_n)$ .
  - (c)  $n$  observations without replacement from the original sample  $(X_1, \dots, X_n)$ .
  - (d)  $n/2$  observations without replacement from the original sample  $(X_1, \dots, X_n)$  and  $n/2$  independent Bernoulli observations with success probability 0.28.

The following figure shows the plot of sample quantiles against the quantiles of the Standard Normal distribution for two datasets each with sample size  $n = 100$ . Answer the next three questions.



15. The observations in Figure A are most likely from the

- (a) Normal(2, 1)
  - (b) Uniform(0, 1)
  - (c) Beta(2, 1) with variance 0.05
  - (d) Gamma(2, 1) with variance 2 •
  - (e) Gamma(2, 1.5) with variance 0.88
16. For Figure A, the sample median of the observed data is
- (a) close to 0
  - (b) close to 0.5
  - (c) close to 2 •
  - (d) close to 3.5
  - (e) Cannot be said unless we see the actual observations
17. The observations in Figure B are most likely from a
- (a) Normal(0.5, 0.1<sup>2</sup>)
  - (b) Uniform(0, 1) •
  - (c) Gamma(2, 4) with variance 0.125
  - (d) mixture distribution, (2/3)Normal(0.5, 0.1<sup>2</sup>) + (1/3)Normal(0.8, 0.1<sup>2</sup>)
- Suppose that  $X_1, \dots, X_n$  are iid from the Normal(2, 1) distribution. Answer the next two questions.
18. What is the probability that a randomly drawn observation from Normal(2, 1) distribution is an outlier?
- (a) 0.0886
  - (b) 0.1773
  - (c) 0.0007
  - (d) 0.0069 •
  - (e) 0.0443
19. Suppose that probability of an outlier is 0.02 (Take this value even it is not the correct answer). What is the probability that at least one of the  $n$  observations is an outlier when  $n = 10$ .
- (a) 0.1829 •
  - (b) 0.3771
  - (c) 0.4582
  - (d) 0.1667
  - (e) 0.8860
20. The following figure shows the boxplot of 80 observations. The figure indicates that the data distribution
- (a) is left skewed. •
  - (b) is right skewed.
  - (c) is symmetric with few outliers, as determined by the box plot.
  - (d) has the median = 0.75.

