

Stat 6601 Quiz #1 (50 minute), Name: _____

- Open book and open note.
- Each problem is 1 points unless stated otherwise
- Write answers on this paper. But you may want to 'backup' to do question #3 (take-home) This question itself will be posted on the class webpage.

#1. (4 pt) Suppose $X_1, \dots, X_k \sim \text{iid } N(0,1)$. We're interested in the distribution of the following statistic for $k=5$.

$$U = \sum_{i=1, \dots, k} X_i^2.$$

1) Write R codes for simulating 1,000 independent samples of the statistic U and put them in the numeric vector 'u.stats'. Note that each sample requires generating k X_j s.

2) Write R codes that plot the probability histogram of the simulated distribution above. In particular, let the number of bins to be 50. Be careful the histogram needs to be scaled correctly to be 'probability histogram' rather than 'frequency histogram'.

3) Write R code for adding the density curves of $\chi^2(m)$ (chi-squared distribution with m degrees of freedom) distribution for $m=3,5,7$, and 9 , to the above histogram. Using 'for' loop is recommended.

4) Write R codes that i) draws the Q-Q plot of the $\chi^2(5)$ (chi-squared distribution with 5 degrees of freedom) distribution against the simulated samples in 'u.stats' and ii) add the straight line $y=x$ on top of it.

#2. (4 pt) Suppose $X_1, \dots, X_n \sim \text{iid}, \{N(0, 1) \text{ with } 95\% \text{ probability and } N(0, 10^2) \text{ with } 5\% \text{ probability}\}$ (Contaminated normal distribution).

1) Write R codes that

- i) generate a sample of $n=100$ observations ($X_i, i=1, \dots, 100$) and put it in a vector 'x.vals' and
- ii) compute the sample median $X\text{-tilde}$ and sample mean $X\text{-bar}$ of the sample and put them in 'x.tilde' and 'x.bar' respectively.

2) Modify the code in 1) so that we repeat the simulation $m=1,000$ times. At each of 1,000 simulations, you need to generate a sample of ($X_i, i=1, \dots, 100$) and compute $X\text{-tilde}$ and $X\text{-bar}$ from the sample. Put the 1,000 results in vectors (of lengths 1,000) named 'my.medians' and 'my.means' respectively.

3) Write R-codes that draw two normal Q-Q plots, one for each vector 'my.medians' and 'my.means'. Also write the codes to compute and display the mean and SD of the vectors.

4) The means of the vectors my.medians and my.means are 0.002461129 and 0.01992826 respectively. Their standard deviations (SD) are 0.1303055 and 0.2406342 respectively. The normal Q-Q plot show that both distributions are close to normal distribution. From these results, what can you say about the sampling distributions of the two competing estimators, sample median and sample mean, when one wants to estimate the unknown mean of contaminated normal distribution with unknown population mean (here it's zero). In particular, which of the two estimators would you use if you believe your data comes from such contaminated normal distribution (with unknown mean)?

#3. **Take Home.** Run all the codes above and

- 1) Cut-and-paste the R codes and results (R-outputs, plots) to an MS Word document with name "last_name,first_name.doc". (e.g. "kwon, jaimie.doc")
- 2) you should attach it to an email with subject lines "stat 6601, quiz 1" (nothing else please; no message body)
- 3) Email it to me no later than Midnight of next Monday (10/18). Results submitted later than then won't be graded.
- 4) Minimize comment in the MS Word document. Typical headings (title, date, and author) and question numbers are enough. (Please keep in mind that I have to go over 50 of them!) See the template available on the course website.