Questions from: <http://www.math.wustl.edu/~jmding/math3200/chw/hw3.html>

1. **(a)** *See Output pg. 1 for partial SAS output of Z1, Z2 … Z4 terms from an N (0, 1) distribution and X= Z1+ Z2+ Z3+ Z4 terms.***(b)**  *See Output pg. 2 for the percentiles of the simulated sample. See Output pg. 3 for a reference table of the corresponding percentiles of the X42distribution.*  For this simulated sample, the X= Z1+ Z2+ Z3+ Z4 terms had the following percentiles: the max (100%) was 11.211; the min (0%) was 0.456; the 25th percentile (Q1) was 1.757; the 50th percentile (Median) was 3.334; and the 90th percentile was 7.033. As reported on the reference table for a X42distribution, the 25th percentile is 1.923. This number is relatively close to the first quartile of the data set. It has a difference of magnitude 0.166 which is only 8.632% of the value on the reference table. As reported on the reference table, the 50th percentile is 3.357. This number is very close to the median of the data set; furthermore, it has a difference of magnitude 0.023 which is only 0.685% of the reference table value. Finally, the 90th percentile of a chi-squared distribution is 7.779. This is relatively close to the 90th percentile of the simulated sample. It has a difference of magnitude 0.776 which is 9.976% of the reference table value. Generally, the simulated sample of X terms does follow a Chi-squared distribution; however, all percentiles are slightly lower than the reference table value for the Chi-squared distribution. If the sample size was bigger than 100, the chi-squared distribution would better approximate the simulated sample. For reiteration,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Percentile** | **Simulated Sample** | **Chi-Squared4 Distribution** | **Difference (Chi Squared minus Sample)** | **Difference as Percentage of Reference Table Value** |
| **25th** | **1.757** | **1.923** | **0.166** | **8.632%** |
| **50th** | **3.334** | **3.357** | **0.023** | **0.685%** |
| **90th** | **7.003** | **7.779** | **0.776** | **9.976%** |

1. **(a)** *See Output pg. 4 for partial SAS output of Z ~ N (0, 1), U ~ X42 and T = Z/sqrt(U/4)***(b)** *See Output pg. 5 for the percentiles of the simulated sample. See Output pg. 6 for a reference table of the corresponding percentiles of the t4 distribution.* For this simulated sample, the T values had the following percentiles: the max (100%) was 4.492; the min (0%) was -6.070; the 25th percentile (Q1) was -0.819; the 50th percentile (Median) was -0.141; and the 90th percentile was 1.020. As reported on the reference table for a t4 distribution, the 25th percentile is -0.741. This number is larger than the first quartile of the data set. It has a difference of magnitude 0.078 which is 10.53% of the value on the reference table. For a t4 distribution, the median is 0. This number is very close to the median of the data set; furthermore, it only has a difference of magnitude 0.141. Finally, the 90th percentile of the t4 distribution is 1.533. This value is much higher than the 90th percentile of the simulated sample. It has a difference of 0.513 which is 33.46% of the reference table value. Generally, the simulated sample of T values does follow a Chi-squared distribution although the percentiles of the sample are all much lower than the reference table values for the corresponding percentiles. If the sample size was larger than 100, the t4 distribution would better approximate the simulated sample. For iteration,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Percentile** | **Simulated Sample** | **t4 Distribution** | **Difference  (t4 minus Sample)** | **Difference as Percentage of Reference Table Value** |
| **25th** | **-0.819** | **-0.741** | **0.078** | **-10.53%** |
| **50th** | **-0.141** | **0** | **0.141** | **NA** |
| **90th** | **1.020** | **1.533** | **0.513** | **33.46%** |

1. **(a)** *Partial SAS output of 25 samples not included. See Output pg. 7 for a 95% confidence interval of each sample. The values for zlower represent the lower bounds of the confidence interval while the values for zupper represent the upper bounds for the confidence interval. If CI\_50=1, then the confidence interval contains the true mean, 50. If CI\_53=1, then the confidence interval contains the wrong mean 5.* All confidence intervals except for sample 24 contained 50; thus, 96% of the confidence intervals contained the true mean. However, only 12 of the confidence intervals contained 53; thus, 48% of the confidence intervals contained the wrong mean, 53.

**(b)** If the sample size were increased to n=100, the width of the 95% confidence interval would shrink. For sample size n=20, we computed the width of the confidence interval as follows: 2\*(zmean-1.96\*6/sqrt(20)). If sample size was increased to n=100, width would be computed as follows: the 2\*(1.96\*6/sqrt(100)). Here we are now dividing by a larger number which will reduce the width of the confidence interval. Now, with a greater sample, I expect more of these intervals to contain the true mean 50 (since there are more intervals in general and the sample will be more representative of the population); however, since the width of the 95% confidence interval will be smaller, I expect less of these intervals to contain the wrong mean 53.

**(c)** If I did not know that the true mean is 50, I cannot tell for any particular CI whether it included the true mu. A confidence interval is interpreted as follows: in an infinitely long series of trials in which repeated samples of size n are drawn from the same population and 95% confidence intervals for mu are calculated using the same method, the proportion of intervals that actually include mu will be 95%. For any particular CI, it is not known whether or not that CI includes mu.

**4.      (a)** *See Table A3 in textbook.* Alpha risk is the probability that we reject H0 given that H0 is true. Therefore for this particular rule, alpha risk is 0.05 equal to the level of significance. To determine beta risk (the probability that we fail to reject H0 given that Mu= 1), I first calculated the xmean at which we would reject H0 (i.e. the xmean corresponding to an upper zscore of 1.645) and found this to be xmean = 0.5483. Beta risk can be calculated as the probability that xmean is less than 0.5483 (we accept H0) given that mu=1. Thus, the beta risk is prob (z< (0.5483-1/(1/sqrt(9))) = 0.087 for a sample size n=9.   
  
**(b)** *Partial SAS output of 100 samples not included. See Output pg.* 8 *for the frequencies of Type I error. See Output pg. 9 for the frequencies of Type II error given the true mu=1.* For these simulated 100 samples of size 9, type I error was committed 5 times or in 5% of the samples. For mu=1, type II error was committed 9 times or in 9% of the samples. The proportion of type I error is exactly the alpha risk (0.05), and the proportion of type II error is almost exactly the beta risk (0.087) with a difference of magnitude 0.003.  
  
**(c)** *Partial SAS output of 100 samples not included. See Output pg. 10 for the frequencies of Type II error after increasing sample size. See Output pg. 11 for the frequencies of Type II error given the true mu=1.* For these simulated 100 samples of size 50, type I error was committed 4 times or in 4% of the samples. For mu=1, no type II error was committed. The proportion of type I error is very slightly less than the alpha risk (0.05); however, we since the sample size has changed, we must recalculated the expected beta risk. I first calculated the xmean at which we reject H0 (i.e. the xmean corresponding to an upper zscore of 1.645) and found this to be xmean = 0.232638. The beta risk can be calculated as the probability that xmean is less that 0.232638 (where we would we accept H0) given that mu =1. Thus, the beta risk is prob (z< (0.232638-1/(1/sqrt(50)))= 0. This is exactly the proportion of type II errors committed for the simulated sample; however, this proportion is much less than the proportion of type II error when the sample size was 9.

**SAS CODE**

TITLE "Homework 3 Question 5.22a and b";

**data** hw3q1a;

do i=**1** to **100**;

array z(**4**);

do j=**1** to **4**;

z[j]=rand('NORMAL',**0**,**1**);

X= z[**1**]\*\***2** + z[**2**]\*\***2** + z[**3**]\*\***2** + z[**4**]\*\***2**;

end;

output;

end;

drop i j;

**run**;

\*Only first page of output included;

**proc** **print** data=hw3q1a;

**run**;

**proc** **univariate** data=hw3q1a;

var X;

**run**;

TITLE "Homework 3 Question 5.22b";

**data** dataset;

p\_25=cinv(**.25**,**4**);

p\_50=cinv(**.5**,**4**);

p\_90=cinv(**.9**,**4**);

**proc** **print** data=dataset;

**run**;

TITLE "Homework 3 Question 5.28a and b";

**data** hw3q2;

call streaminit(**12345**);

do i=**1** to **100**;

Z=rand('NORMAL',**0**,**1**);

U=rand('CHISQUARE',**4**);

T=Z/sqrt(U/**4**);

output;

end;

drop i;

**run**;

\*Only first page of output included;

**proc** **print** data=hw3q2;

**run**;

**proc** **univariate** data=hw3q2;

var T;

**run**;’

TITLE "Homework 3 Question 5.28b";

**data** hw3q2b;

p\_25=tinv(**.25**,**4**);

p\_50=tinv(**.5**,**4**);

p\_90=tinv(**.9**,**4**);

**proc** **print** data= hw3q2b;

**run**;

TITLE "Homework 3 Question 6.12a";

**data** hw3q3a;

\*Optional: For part B, change 20 to 100;

do i=**1** to **25**;

array z(**20**);

do j=**1** to **20**;

z[j]=rand('NORMAL',**50**,**6**);

zmean=mean(of z1-z20);

zlower=zmean-**1.96**\***6**/sqrt(**20**);

zupper=zmean+**1.96**\***6**/sqrt(**20**);

CI\_50=(zlower<=**50** and zupper>=**50**);

CI\_53=(zlower<=**53** and zupper>=**53**);

end;

output;

end;

drop i j;

**run**;

**proc** **print** data=hw3q3a;

\*Optional: If CL\_50= 1, then the intervals contain the true mean;

\*Optional: If CL\_53= 1, then the intervals contain the wrong mean 53;

var zmean zlower zupper CI\_50 CI\_53;

**run**;

TITLE "HW 3 Question 6.23";

**data** hw3q4;

call streaminit(**12345**);

do i=**1** to **100**;

array z(**9**);

do j=**1** to **9**;

z[j]=rand('NORMAL',**0**,**1**);

end;

zmean=mean(of z1-z9);

zstat=zmean/(**1**/sqrt(**9**));

\*Type=1 means a type 1 error has been committed;

\*Type=0 means a type 1 error has not been committed;

if zstat>**1.645** then Type=**1**;

else Type=**0**;

output;

end;

drop i j;

**run**;

TITLE "HW 3 Question 6.23: Frequency of Type I Error";

**proc** **freq** data=hw3q4;

tables Type;

**run**;

TITLE "Homework 3 6.23b";

**data** hw3q4b;

call streaminit(**12345**);

do i=**1** to **100**;

array x(**9**);

do j=**1** to **9**;

x[j]=rand('NORMAL',**1**,**1**);

end;

xmean=mean(of x1-x9);

\*Type=2 means a type 2 error has been committed;

\*Type=0 means a type 2 error has not been committed;

if xmean<**0.548333333** then Type=**2**;

else Type=**0**;

output;

end;

drop i j;

**run**;

TITLE "HW 3 Question 6.23b: Frequency of Type II Error";

**proc** **freq** data=hw3q4b;

tables Type;

**run**;