Examination Number:____________

Last Name______________First__________

Sign the Honor Pledge Below_____________ PID #________________________________

______________________________________ Write Your Section Number here:_________

University of North Carolina

Economics 400: Economic Statistics and Econometrics

First Midterm Examination

Prof. B. Turchi February 28, 2019

Among his many admirable qualities Coach Smith was an innovator in the use of statistics and mathematics to maximize the competitive advantage for his teams. On the fourth anniversary of his death, this examination will commemorate his contributions with some UNC basketball-based problems.

General Instructions: Answer all **five (5)** questions on this examination, writing your answers on the exam paper itself. Use the scratch page at the end for any extra work, if necessary. Sign the Honor Pledge above. Express all answers to a precision of at least 3 decimal points. Show your work to be eligible for partial credit. **Be sure to note that tables and formulas are on the last 2 pages of the exam.**

1. Dean Smith, the joke goes, was the only person in the world who could hold Michael Jordan under 20 points per game. In fact, since 1953 only six Carolina players out of 320 averaged 20 points or more per game for their entire UNC careers (Jordan wasn't one of them). Averaging double digits (10.0 points or better) for an entire UNC career is really hard; only 20.9 percent of UNC players have done so. Suppose we randomly sample (with replacement) 24 players out of this population of 320 counting as a success each player who averaged 10 or more points per game.

(a) (3 points) What is the appropriate distribution to use in analyzing this sample? Why? <<The Binomial distribution since we are sampling with replacement>>

(b) (3 points) What is the expected number of double-digit scorers we should expect to find in the sample? Explain your work. <<5.016. This is a binomial distribution so $\mu = np = 24 \times 0.209 = 5.016$>>

(c) (12 points) What is the probability of finding between 4 and 10 (inclusive) double-digit scorers? Be sure to show your work and the assumptions you're making.

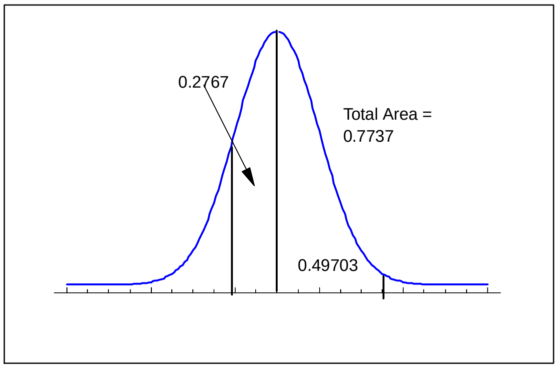
This question asks the student to compute the probability of finding between 4 and 10 successes in a sample of 24. Since this is a binomial problem, one could compute the binomial probabilities for each of the 7 different outcomes, but this is really tedious and time consuming especially in a timed exam. (Using Stata's binomialtail function: binomialtail(24,4,0.209) = 0.7698327 and binomialtail(24,11,0.209) = 0.0054077, we get 0.7698327 - 0.0054077 = **0.764425**. Student should get most credit for this but minor deduction for not using the normal aproximation to the binomial:

1. Apply rules of thumb: (np = 5.016 & n(1-p) = 18.984)
2. Apply continuity correction (4 - .5 = 3.5 & 10 + .5 = 10.5)

3. Compute areas:

$$z_{lower} = \frac{3.5 - 5.016}{1.9919} = -0.7611 \Rightarrow 0.2767 \text{ (area interpolated)}$$

$$z_{upper} = \frac{10.5 - 5.016}{1.9919} = 2.75315 \Rightarrow 0.49703 \text{ (area interpolated)}$$



Total probability of finding 4 - 10 (inclusive) double-digit scoreres is 0.7737

2. (16 points) It is generally agreed that a 3-point shot in basketball is a good shot if a team can make 33 percent of them. Let event B represent "Duke shoots greater than 33 percent" in a game. Reviewing Duke's performance in prior games in his last season (1996-97), suppose Coach Smith found that Duke faced 3 different defenses: zone (35 percent of the time), man-to-man (50 percent of the time) and "junk" defenses (15 percent of the time). He also found that against zone defenses Duke hit 38 percent of their 3-point shots, against man-to-man 27 percent of their shots and against junk defenses 41 percent. In the last game prior to the UNC game Duke hit 25 percent of their shots. What is the probability that Duke faced a "junk" defense in that game?

<< Here's our data: B = Duke hits more than 33 percent of their shots
D1=zone, D2=man-to-man, D3=junk
P(B|zone) = .38
P(B|man-to-man) = .27
P(B|junk) = .41
P(zone) = .35
P(man-to-man) = .50
P(junk) = .15

We want to find P(junk|B-not)

We can infer from the data the following:
P(B-not|zone) = .62
P(B-not|man-to-man)= .73
P(B-not|junk) = .59

$$\mathrm{P}\left(D_3 \mid \bar{B}\right) = \frac{P\left(\bar{B} \mid D_3\right) \bullet P\left(D_3\right)}{\sum_{i=1}^{3} P\left(\bar{B} \mid D_i\right) \bullet P\left(D_i\right)}$$

is Bayes' rule applied to these data. Plugging in the numbers, we get:

$$\mathrm{P}\left(D_3 \mid \bar{B}\right) = \frac{(.59 \times .15)}{(.59 \times .15) + (.62 \times .35) + (.73 \times .50)} = \frac{.0885}{.0885 + .217 + .365} = 0.132$$

So, the probability that Duke faced a "junk" defense fell from 15 to 13.2 percent.

3. Prove the following: (12 points) Prove that the variance (σ^2) of a standardized random variable $(\tilde{z})$ equals one. Show your work and assumptions.

$$\begin{aligned} V[\tilde{z}] &= V\left[\frac{1}{\sigma}\tilde{x} - \frac{\mu}{\sigma}\right] \\ &= V\left[\frac{1}{\sigma}\tilde{x}\right] + V\left[\frac{\mu}{\sigma}\right] \\ &= \frac{1}{\sigma^2} V[\tilde{x}] + 0 \\ &= \frac{1}{\sigma^2} \times \sigma^2 = 1 \end{aligned}$$

4. In Dean Smith's last year of coaching, his team averaged 15.7 personal fouls per 40 minute game. Assuming fouls were spread evenly through the game,

(a) (4 points) What is the expected number of fouls in any 10 minute period?

<<3.925 fouls per 10 minutes>> Ten minutes is 25% of the game, so we use the Poisson function with $\lambda = 15.7$ and *t=0.25*. *Therefore, the mean number of fouls for the 10 minute period is* $\lambda \cdot t = 15.7 \cdot 0.25 = 3.925$.

(b) (8 points) What is the probability that the team would commit 5 or more fouls in a 10 minute period?

$$\begin{aligned} P(x \geq 5) &= 1 - \sum_{i=0}^{4} \frac{e^{-\lambda t}(\lambda \bullet t)^{x_i}}{x_i!} \\ &= \left(1 - [0.0197 + 0.0775 + 0.1521 + 0.1990 + 0.1952]\right) \\ &= 0.3565 \text{ where } \lambda = 15.7 \text{ and } t = 0.25 \end{aligned}$$

Here the student needs to calculate 5 Poisson probabilities for x = 0, 4 and subtract the sum from 1, as above.

(c) (4 points) What is the expected number of fouls in a one minute period?

<<15.7/40 = .3925>> In this case the mean is $\lambda \cdot t = 15.7 \times \frac{1}{40} = 0.3925.$

(d) (8 points) What is the probability that a team could go five or more minutes without a foul?

<<0.1405>> This is a right-tail probability derived from an exponential distribution with

$$P(t \geq 5) = e^{-\frac{\lambda}{40} \bullet 5} = e^{-0.3925 \times 5} = 0.1405 \text{ where } \frac{\lambda}{40} = 0.3925 \text{ and } t = 5.$$

5. "Career Minutes played" is a statistic that was collected for players only since the mid- to late- 1970s. We have minutes played for 189 players. The data below represent a random sample of 25 drawn (without replacement) from those 189 players.

```
 9  54  479 2100 3055
15  74  826 2865 3185
17 123 1135 2889 3315
42 254 1323 2901 3705
49 280 1833 2919 3973
```

a) (6 points) calculate the sample median, mean and standard deviation of these data. Show your work.

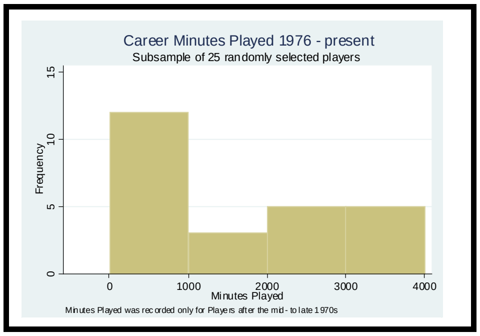
Median: = 1,135

$$\text{Mean: } = \frac{\sum_{i=1}^{25} x_i}{25} = 1,496.8$$

$$\text{Std. Dev:} = \sqrt{\frac{\sum_{i=1}^{25} (x_i - \bar{x})^2}{25-1}} = 1,433.96$$

(b) (6 points) Draw a picture in the box below of the histogram that might represent these data with each bar representing 1,000 minutes. How would you describe the shape of this distribution?

<< Skewed right and slightly bimodal>>

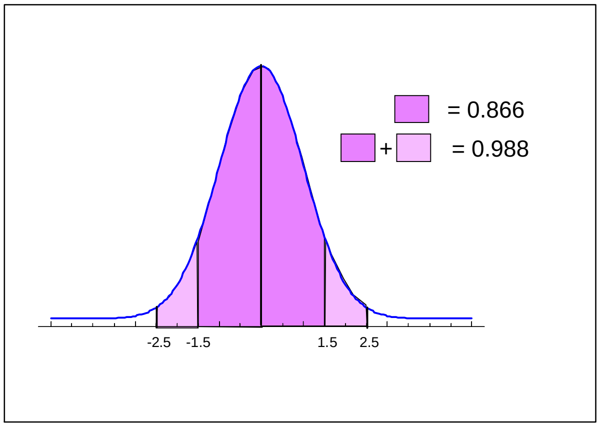


(c) (6 points) Construct intervals around the sample mean of one standard error (on each side) and two standard errors (on each side). What are the intervals? Show how you calculate them.

$(\pm 266.793 \text{ and } \pm 573.586)$

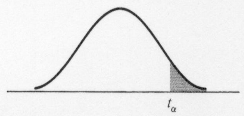
(d) (6 points) Assuming that the distribution of the sample mean is normal, calculate the probability that the true mean is within 1.5 or 2.5 standard errors of the sample mean. Show your work.

<<This is simply finding the areas in the normal table for z = 1.5 and z = 2.5. Those areas are (0.4332x2 = 0.866) and (0.4938x2 = 0.9876) respectively)



(e) (6 points) Are these probabilities likely to represent an over-, under-, or correct- estimate of the true probabilities? Explain.

<<The population is likely to be non-normal and, because we have to use the sample standard deviation to calculate the standard errors it is likely that we overestimate the probability of the true population mean being within the specified intervals. On the other hand the central limit theorem suggests that the sample mean approaches the normal as sample size grows. The question here is whether or not the sample is large enough.



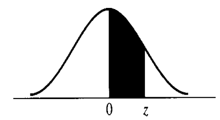
d.f.	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	d.f.
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

Source: From "Table of Percentage Points of the *t*-Distribution." *Biometrika*, Vol. 32 (1941), p. 300. Reproduced by permission of the Biometrika Trustees.

$$\Pr\left(-t_{\alpha/2} < \frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{\alpha/2}\right) = 1-\alpha$$

$$z = \frac{\hat{p}-p_0}{\sqrt{\dfrac{p_0\left(1-p_0\right)}{n}}}$$

Normal Curve Areas



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: This table is abridged from Table 1 of *Statistical Tables and Formulas*, by A. Hald (New York: John Wiley & Sons, Inc., 1952). Reproduced by permission of A. Hald and the publishers, John Wiley & Sons, Inc.

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{D^2}. \qquad n = \frac{z_{\alpha/2}^2}{4D^2}$$

Binomial Coefficients

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
0	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	405	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15304	38760	77520	125970	167960	184756

If necessary, use the identity $\binom{n}{k}=\binom{n}{n-k}$.

$$f(\tilde{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$$

$$p(x) = \frac{C_x^r C_{n-x}^{N-r}}{C_n^N} = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Mean: $\mu = n\left(\frac{r}{N}\right)$

Variance: $\sigma^2 = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$

Standard deviation: $\sigma = \sqrt{\sigma^2}$

$$f(x) = \begin{cases} \frac{1}{(b-a)}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

$$\mu = \frac{1}{2}(b+a) \text{ and } \sigma = \frac{(b-a)}{\sqrt{12}}$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \lambda > 0, x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu = \frac{1}{\lambda} \text{ and } \sigma = \frac{1}{\lambda}$$

$$P(x \ge a) = e^{-\lambda a},\ a \ge 0 \text{ and } \lambda > 0$$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{\text{all } k} P(B|A_k)P(A_k)}$$

$$P(x) = \begin{cases} \frac{e^{-\lambda t}(\lambda t)^x}{x!}, & \text{for } x = 0,1,2,..,\infty, \quad \lambda > 0, \\ 0, & \text{otherwise.} \end{cases}$$

λ = the mean number of events in a given segment of time $(t = 1)$

t = the length of a particular subsegment $(t \le 1)$

$E[x] = \mu_x = \lambda t$ = the expected number of events in one subsegment length t