SPRING 2016

HUNTER COLLEGE

STAT 213 Section HC1

Introduction to Applied Statistics

Midterm Exam Two

Last Name: ______________________________

First Name: ______________________________

Graduation Year: ________________________

May 2, 2016

1. Please do not leave blank for any question.
2. There are 8 questions, each question is 5 points. A perfect score is 40 points.
3. You have 70 minutes for this exam (4:15 pm - 5:25 pm).
4. Explain briefly = Explain in one sentence or several phrases.

Formulas

$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$z = \frac{x-\mu}{\sigma}$

$x = \mu + \sigma z$

$s_x^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$

$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}}$

$\Pr(|X-\mu| \le k\sigma) \ge 1 - \frac{1}{k^2}$, for any distribution

$\Pr(|X-\mu| \le \sigma) \approx 0.68$, $\Pr(|X-\mu| \le 2\sigma) \approx 0.95$, $\Pr(|X-\mu| \le 3\sigma) \approx 0.997$, for normal distribution

$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{s_x}\right)\left(\frac{y_i-\overline{y}}{s_y}\right)$

$r = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{(n-1)s_x s_y}$

$\widehat{y} = a + bx$

$b = r\frac{s_y}{s_x}$

$b = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$

$a = \overline{y} - b\overline{x}$

$SSTo = \sum_{i=1}^{n}(y_i-\overline{y})^2$, $SSTo = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2$

$SSResid = \sum_{i=1}^{n}(y_i-\widehat{y}_i)^2$, $SSResid = \sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i$

$r^2 = 1 - \frac{SSResid}{SSTo}$

$s_e = \sqrt{\frac{SSResid}{n-2}}$

If events $E_1, \cdots, E_k$ are all mutually exclusive, then $\Pr(E1 \cup \cdots \cup E_k) = \Pr(E_1) + \cdots + \Pr(E_k)$

$\Pr(E \mid F) = \frac{\Pr(E \cap F)}{P(F)}$

The event E and F are independent if and only if $\Pr(E \cap F) = \Pr(E)\Pr(F)$

$\Pr(E \cup F) = \Pr(E) + \Pr(F) - \Pr(E \cap F)$

$\Pr(E \cap F) = \Pr(E|F)\Pr(F) = \Pr(F|E)\Pr(E)$

If events $B_1, \cdots, B_k$ are all mutually exclusive with $\Pr(B_1) + \cdots + \Pr(B_k) = 1$, then for any event E, $\Pr(E) = \Pr(E \mid B_1)\Pr(B_1) + \cdots + \Pr(E \mid B_k)\Pr(B_k)$

If events $B_1, \cdots, B_k$ are all mutually exclusive with $\Pr(B_1) + \cdots + \Pr(B_k) = 1$, then for any event E, $\Pr(B_i \mid E) = \frac{\Pr(E|B_i)\Pr(B_i)}{\Pr(E|B_1)\Pr(B_1) + \cdots + \Pr(E|B_k)\Pr(B_k)}$

For continuous random variable X, $\Pr(a < X \leq b) = \Pr(X \leq b) - \Pr(X \leq a)$

Mean value of a discrete random variable $\mu_X = \sum x \cdot p(x)$

Standard deviation of a discrete random variable $\sigma_X = \sqrt{\sum (x - \mu_X)^2 \cdot p(x)}$

$Y = a + bX$. Mean $\mu_Y = a + b\mu_X$. Standard deviation $\sigma_Y = |b|\sigma_X$

$Y = a_1X_1 + \cdots + a_nX_n$. Mean $\mu_Y = a_1\mu_{X_1} + \cdots + a_n\mu_{X_n}$. Standard deviation (when X_i's are independent) $\sigma_Y = \sqrt{a_1^2\sigma_{X_1}^2 + \cdots + a_n^2\sigma_{X_n}^2}$

Binomial distribution $p(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$, $\mu_X = np$, $\sigma_X = \sqrt{np(1-p)}$

Geometric distribution $p(x) = (1-p)^{x-1}p$, $\mu_X = 1/p$

Continuity correction (X is a discrete variable (integer values), Y is the corresponding Normal random variable): If $\Pr(X \leq m)$ use $\Pr(Y < m + 0.5)$; If $\Pr(X < m)$ use $\Pr(Y < m - 0.5)$

Sampling distribution of $\overline{X}$: $\mu_{\overline{X}} = \mu$, $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

Sampling distribution of $\widehat{p}$: $\mu_{\widehat{p}} = p$, $\sigma_{\widehat{p}} = \sqrt{p(1-p)/n}$

$(1-\alpha) \times 100\%$ confidence interval for p: $\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$

$n = p(1-p)\left(\frac{z_{\alpha/2}}{B}\right)^2$

$(1-\alpha) \times 100\%$ confidence interval for μ: $\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$n = \left(\frac{z_{\alpha/2}\sigma}{B}\right)^2$

$(1-\alpha) \times 100\%$ confidence interval for μ: $\overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$

Finite population correction factor: $\sqrt{\frac{N-n}{N-1}}$

$z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$, where μ is the hypothesized value under H_0

$t = \frac{\overline{x}-\mu}{s/\sqrt{n}}$, where μ is the hypothesized value under H_0

NORMAL CUMULATIVE DISTRIBUTION FUNCTION

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

STUDENT'S t PERCENTAGE POINTS

one-tail	40.0%	33.3%	25.0%	20.0%	12.5%	10.0%	5.0%	2.5%	1.0%	0.5%	0.1%
two-tail	80.0%	66.7%	50.0%	40.0%	25.0%	20.0%	10.0%	5.0%	2.0%	1.0%	0.2%
cum. prob	60.0%	66.7%	75.0%	80.0%	87.5%	90.0%	95.0%	97.5%	99.0%	99.5%	99.9%
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.436	0.684	0.856	1.178	1.316	1.708	2.060	2.485	2.787	3.450
26	0.256	0.436	0.684	0.856	1.177	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.435	0.684	0.855	1.176	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.435	0.683	0.855	1.175	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.435	0.683	0.854	1.174	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.435	0.683	0.854	1.173	1.310	1.697	2.042	2.457	2.750	3.385
35	0.255	0.434	0.682	0.852	1.170	1.306	1.690	2.030	2.438	2.724	3.340
40	0.255	0.434	0.681	0.851	1.167	1.303	1.684	2.021	2.423	2.704	3.307
45	0.255	0.434	0.680	0.850	1.165	1.301	1.679	2.014	2.412	2.690	3.281
50	0.255	0.433	0.679	0.849	1.164	1.299	1.676	2.009	2.403	2.678	3.261
55	0.255	0.433	0.679	0.848	1.163	1.297	1.673	2.004	2.396	2.668	3.245
60	0.254	0.433	0.679	0.848	1.162	1.296	1.671	2.000	2.390	2.660	3.232
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090
conf. level	20.0%	33.3%	50.0%	60.0%	75.0%	80.0%	90.0%	95.0%	98.0%	99.0%	99.8%

1. Bayes' Rule (Chapter 6)
   At a certain gas station 40% of the customers request regular gas, 35% request plus gas, and 25% request premium gas. Of those customers requesting regular gas, 20% fill up their tanks. Of those customers requesting plus gas, 60% fill up their tanks, while of those requesting premium, 45% fill up their tanks. If the next customer fills up the tank, what is the probability that regular gas is requested?

Figure 1: Gas Price at Orlando FL, 2011-08-03

2. Discrete Random Variables (Chapter 7)
Suppose an individual plays a gambling game where it is possible to lose \$1.00, break even, win \$3.00, or win \$5.00 each time she plays. The probability distribution for each outcome is provided by the following table:

Outcome	-\$1.00	\$0.00	\$3.00	\$5.00
Probability	0.30	0.40	0.20	0.10

(1) Verify that the discrete probability distribution above is well-defined.
(2) Find the mean and standard deviation of this discrete random variable.
(3) Suppose that the casino decides that the game does not have an impressive enough top prize with the lower payouts, and decides to change the outcomes, as shown below

Outcome	-\$4.00	-\$2.00	\$4.00	\$8.00
Probability	0.30	0.40	0.20	0.10

Find the relation between new outcome and previous outcomes. Based on the relation, find the mean and standard deviation of the new random variable.

3. Normal Distribution (Chapter 7)

   (You may need a Normal Table for this question.)

   Washington Police officers recorded the speeds of cars driving on a busy street by a school for a one-month period, where the speed limit read 20 mph. The mean of readings was 24.90 mph, with a standard deviation 7.52 mph.

   (data source: www.wtsc.wa.gov)

   What percent of the vehicles were exceeding the posted speed limit in school zones (20 mph)?

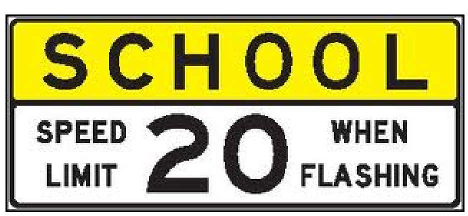


Figure 2: Speed limit 20 MPH.

4. Normal Distribution (Chapter 7)

   (You may need a Normal Table for this question.)

   In Fuel Economy Guide (Model Year 2013), Environmental Protection Agency (EPA) fuel economy estimates for automobile models tested predicted a mean of 23.8 mpg (miles per gallon) and a standard deviation of 6.2 mpg. Assume that a Normal model can be applied.

   (data source: www.fueleconomy.gov)

   An auto dealer introduced you a fuel-efficient car. He told you that this car's gas mileage is higher than 95% of vehicles.

   Find the gas mileage of this car.

Figure 3: Gas Mileage

5. Normal Probability Plot (Chapter 7)
Scientists found two new types of insects in Rocky Mountain, with sample size 4 and 5. For insect A, the body lengths (centimeter) are

$$1.2,\ 1.6,\ 2.4,\ 2.7,$$

for insect B, the body lengths (centimeter) are

$$0.8,\ 1.0,\ 1.1,\ 1.7,\ 3.1.$$

(1) Make Normal Probability Plots for insect A and insect B. You may put two sets of data on one plot with different symbols (for example, $\circ$ and $\times$). (2) Check the normality based on your plots.

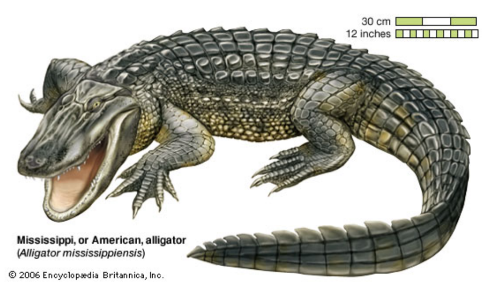
How to make a Normal Probability Plot
1. Sort the data. $x_1 \le x_2 \le \cdots \le x_n$
2. Calculate the corresponding quantile

$$z_i = \Phi^{-1}\left(\frac{i - \frac{1}{2}}{n}\right)$$

3. Make a scatter plot for $\{z_i, x_i\}_{i=1}^{n}$.

6. Confidence Intervals (Chapter 8, 9)

   Biologists measured 8 American crocodiles (male, adult), the average length of this sample is 14.6 feet. The sample standard deviation of the lengths of American crocodiles (male, adult) is 0.74 feet. Find a 95% confidence interval of the mean length for all American crocodiles (male, adult). What is the margin of error?



7. Sample Size (Chapter 9)

   A Company claims its program will allow your computer to download movies quickly. We'll test the free evaluation copy by downloading a movie several times, hoping to estimate the mean download time with a margin of error of only 2 minutes. We think the standard deviation of download times is about 2.5 minutes. How many trial download must we run if we want 98% confidence in our estimate with a margin of error of 2 minutes?

8. Hypothesis Test (Chapter 10)

   Mirex is a chlorinated hydrocarbon that was commercialized as an insecticide and later banned because of its impact on the environment.

   Researchers tested 12 farm-raised salmon for organic contaminants. They found the mean concentration of the carcinogenic insecticide mirex to be 0.0913 parts per million, with sample standard deviation $s = 0.0195$ ppm. As a safety recommendation to recreational fishers, the Environmental Protection Agency's (EPA) recommended "screening value" for mirex is 0.08 ppm. Population distribution is assumed to be normal.

   (1) Is it necessary to use a finite population correction factor to adjust the standard error? Explain.
   (2) Are farmed salmon contaminated beyond the level permitted by the EPA? Use the significance level $\alpha = 0.05$.

More Space

End of the Midterm 2 of Stat 213 Sec HC1 (Instructor: Jiangtao Gou)