Stat 2470, Exam #2, Spring 2014 Name KEY

Instructions: You may use your calculator for any functions the TI-84/83 model calculator is capable of using, such as probability distributions and obtaining graphs of any data. To show work on these problems, report the functions and their syntax as entered. Other things, such as integrating, must be done by hand unless specifically directed otherwise. Round means to one more place than the original data, variances and standard deviations to two more than the original data. Round probabilities to three significant figures or use exact values. In order to receive partial credit on any problem, you must show some work or I will have nothing to award partial credit on. Be sure to complete all the requested parts of each problem.

1. Consider the discrete joint probability distribution shown in the table below for a pair of loaded dice.

Die1(x)→ Die2(y)↓ $p(x,y)$	1	2	3	4	5	6
1	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{35}$	$\frac{1}{70}$	$\frac{1}{70}$
2	$\frac{3}{70}$	$\frac{3}{70}$	$\frac{3}{70}$	$\frac{3}{35}$	$\frac{3}{70}$	$\frac{3}{70}$
3	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{35}$	$\frac{1}{70}$	$\frac{1}{70}$
4	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$
5	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{35}$	$\frac{1}{70}$	$\frac{1}{70}$
6	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$

a. What is $p_X(x)$? (5 points)

x	1	2	3	4	5	6
$p_X(x)$	$10/70 = 1/7$	$10/70 = 1/7$	$10/70 = 1/7$	$10/35 = 2/7$	$10/70 = 1/7$	$10/70 = 1/7$

b. What is $p_Y(y)$? (5 points)

y	1	2	3	4	5	6
$p_Y(y)$	$7/70 = 1/10$	$21/70 = 3/10$	$7/70 = 1/10$	$14/70 = 1/5$	$7/70 = 1/10$	$14/70 = 1/5$

c. What is $E(X+Y)$? (8 points)

$2(1/70) + 3(4/70) + 4(5/70) + 5(8/70) + 6(11/70) + 7(11/70) + 8(11/70) +$

$9(7/70) + 10(7/70) + 11(3/70) + 12(2/70) =$

$99/14 \approx 7.07$

from 1-Var-Stats L_1, L_2

d. What is $\rho_{X,Y}$? (10 points)

$\sigma_x = 1.590789818 \quad \mu_x = 3.571428\ 57 = 25/7$

$\sigma_y = 1.688194302 \quad \mu_y = 3.5$

$E(XY) = 1(1/70) + 2(4/70) + 3(2/70) + 4(7/70) + 5(2/70) + 6(7/70)$

$+ 8(8/70) + 9(1/70) + 10(4/70) + 12(9/70) + 15(2/70) + 16(4/70)$

$+ 18(3/70) + 20(4/70) + 24(6/70) + 25(1/70) + 30(3/70) + 36(2/70)$

$= 12.5$

$Cov(X,Y) = E(XY) - \mu_x \mu_y = \frac{25}{2} - \frac{25}{7} \cdot \frac{7}{2} = 0$

$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = 0$ (these are independent)

2. A certain pair of events if governed by the joint probability distribution $f(x,y) = \begin{cases} \frac{6}{7}(xy+y^2), 0 \le x \le 1, 0 \le y \le 2x \\ 0, \quad otherwise \end{cases}$

a. Verify that this is a valid probability distribution. (5 points)

$\frac{6}{7} \int_0^1 \int_0^{2x} xy + y^2 \, dy \, dx = \frac{6}{7} \int_0^1 \frac{xy^2}{2} + \frac{y^3}{3} \Big|_0^{2x} dx =$

$\frac{6}{7} \int_0^1 \frac{x(2x)^2}{2} + \frac{(2x)^3}{3} dx = \frac{6}{7} \int_0^1 2x^3 + \frac{8}{3}x^3 dx = \frac{6}{7} \int_0^1 \frac{14}{3} x^3 dx$

$= \frac{6}{7} \cdot \frac{14}{3} \cdot \frac{1}{4} x^4 \Big|_0^1 = 1$ yes, it is a valid dist.

$E(XY)=$ b. Find $Cov(X,Y)$. (8 points) $E(XY)-E(X)E(Y)$

$$\frac{6}{7}\int_0^1\int_0^{2x}(xy+y^2)xy\,dy\,dx = \frac{6}{7}\int_0^1\int_0^{2x} x^2y^2+xy^3\,dy\,dx = \frac{6}{7}\int_0^1 \frac{x^2}{3}y^3+\frac{xy^4}{4}\Big|_0^{2x}$$

$$=\frac{6}{7}\int_0^1 \frac{8}{3}x^5+4x^5\,dx = \frac{6}{7}\int_0^1 \frac{20}{3}x^5\,dx = \frac{6}{7}\cdot\frac{20}{3}\cdot\frac{1}{6}x^6\Big|_0^1 = \frac{20}{21}$$

$$E(X) = \frac{6}{7}\int_0^1\int_0^{2x} x^2y+xy^2\,dy\,dx = \frac{6}{7}\int_0^1 \frac{x^2y^2}{2}+\frac{xy^3}{3}\Big|_0^{2x}dx = \frac{6}{7}\int_0^1 2x^4+\frac{8}{3}x^4\,dx$$

$$\frac{6}{7}\int_0^1\frac{14}{3}x^4\,dx = \frac{6}{7}\cdot\frac{14}{3}\cdot\frac{1}{5}x^5\Big|_0^1 = \frac{4}{5} \qquad E(Y) = \frac{6}{7}\int_0^1\int_0^{2x} xy^2+y^3\,dy\,dx =$$

$$\frac{6}{7}\int_0^1 \frac{xy^3}{3}+\frac{y^4}{4}\Big|_0^{2x}dx = \frac{6}{7}\int_0^1\frac{8}{3}x^4+4x^4\,dx = \frac{6}{7}\int_0^1\frac{20}{3}x^4\,dx = \frac{6}{7}\cdot\frac{20}{3}\cdot\frac{1}{5}x^5\Big|_0^1$$

$$=\frac{8}{7}$$

$$Cov(X,Y) = \frac{20}{21}-\frac{4}{5}\cdot\frac{8}{7} = \frac{20}{21}-\frac{32}{35} = \frac{4}{105}$$

c. Is the distribution independent? Why or why not? (3 points)

No. $Cov(x,y)\neq 0$ & $f(x,y)\neq f_x(x)\cdot f_y(y)$

d. What is $f_{Y|X}(y|x)$? (4 points)

$$f_x = \frac{6}{7}\int_0^{2x} xy+y^2\,dy = \frac{6}{7}\left[\frac{xy^2}{2}+\frac{y^3}{3}\Big|_0^{2x}\right] = \frac{6}{7}\left[2x^3+\frac{8}{3}x^3\right] = \frac{6}{7}\cdot\frac{14}{3}x^3 = 4x^3$$

$$f_{y|x}(y|x) = \frac{xy+y^2}{4x^3}$$

3. What is the rule of thumb for applying the Central Limit Theorem? What are its limitations? (4 points)

rule of thumb is that it can generally be used for $n\geq 30$. the distribution must be normal or approximately normal (not wildly skewed)

answers may vary

4. What does it mean for an estimator to be unbiased? Give at least one example of an unbiased estimator. (4 points)

an estimator is unbiased if $E(\hat{\theta}) = E(\theta)$

the mean $\bar{x}$ is an unbiased estimator of μ.

answers may vary.

5. What does the term 'point estimate' mean in statistics? What is a 'point estimate' estimating? Give at least one example. (4 points)

a point estimate is a single value used to estimate a parameter (rather than an interval, for instance).

6. Suppose that we have collected 7 pieces of data from an exponential distribution and obtained the following sample results: 2, 11, 17, 23, 29, 31, 58. We'd like to use this data to estimate the value of the parameter λ.

a. What is the maximum likelihood function for this data? (4 points)

$$f(\lambda) = \lambda e^{-\lambda(2)} \cdot \lambda e^{-\lambda(11)} \cdot \lambda e^{-\lambda(17)} \cdot \lambda e^{-\lambda(23)} \cdot \lambda e^{-\lambda(29)} \cdot \lambda e^{-\lambda(31)} \cdot \lambda e^{-\lambda(58)}$$

$$= \lambda^7 e^{-171\lambda}$$

b. Use this function to estimate $\hat{\lambda}$. (You must show the calculus on this problem! You will receive no more than 1 point for correctly calculating the mean by another method.) (8 points)

$$f'(\lambda) = 7\lambda^6 e^{-171\lambda} - 171\lambda^7 e^{-171\lambda} = \lambda^6 e^{-171\lambda}(7 - 171\lambda) = 0$$

$$7 = 171\lambda \Rightarrow \hat{\lambda} = \frac{7}{171} \approx .0409$$

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{171}{7} \approx 24.42857$$

7. A confidence interval is desired for the true average stray-load loss μ (watts) for a certain type of induction motor when the line current is held at 10 amps for a speed of 1500 rpm. Assume that stray-load loss is normally distributed with σ = 3.0.

   a. Compute a 95% CI for μ when n = 50 and $\bar{x}$ = 60. (4 points)

$$60 \pm \frac{1.96(3)}{\sqrt{50}} \approx 60 \pm .83156$$

$$(59.168, 60.832)$$

   b. How large must n be if the width of the 99% interval for μ is to be 1.0? (4 points)

$$n = \left[\frac{2(2.58)\cdot 3}{1}\right]^2 = 239.62 \Rightarrow n = 240$$

8. It was reported that, in a sample of 507 adult Americans, only 142 correctly described the Bill of Rights as the first ten amendments to the U.S. Constitution. Calculate a (two-sided) confidence interval using a 99% confidence level for the proportion of all U. S. adults that could give a correct description of the Bill of Rights. (5 points)

$N = 507$, $X = 142$ $\hat{p} = \frac{142}{507} \approx .28$

using 1 PropZ interval (or use formula)

$X = 142$
$N = 507$
Conf. 99%

$\Rightarrow (.22871, .33145)$

9. Suppose that a sample of 20 statistics students is randomly selected to determine the average level of preparation for their first midterm exam across all statistics classes. The sample has a mean of 74 (percent) and a standard deviation of 12.3 (percent). Calculate a 90% confidence interval for this data. Use your answer to determine whether or not the typical statistics student is well-prepared for their first midterm. (5 points)

$\bar{X} = 74$ $S = 12.3$ TInterval $n = 20$

Stats
$\bar{X} = 74$
$S = 12.3$
$n = 20$
Conf. 90%

$(69.244, 78.756)$

possibly not since the interval contains scores that are not passing.

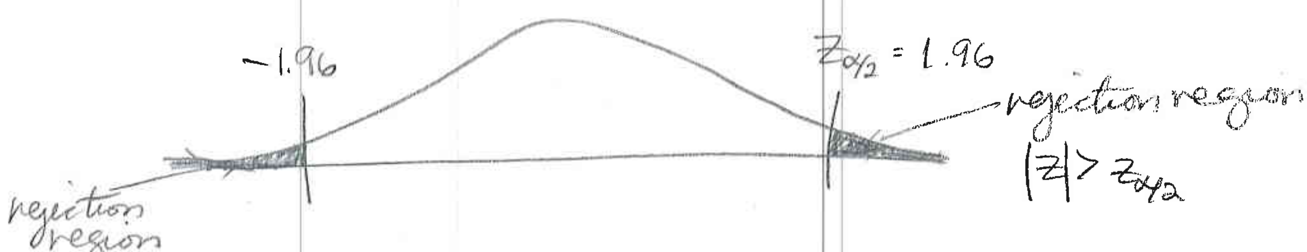
10. Explain what is the difference between a Type I and a Type II error. Give an example of each. (6 points)

a Type I error is the chance of rejecting a null hypotheses when it is true. A type II error is accepting a null hypothesis when it is false.

Type I: Suppose a medicine doesn't really work better than placebo, but your test says it does.

Type II: Saying a medicine does work when it doesn't

11. Suppose that you want to conduct a hypothesis test on a mean from a normally distributed population with a known standard deviation. Sketch the normal curve and the rejection region for a two-tailed hypothesis test with $\alpha = 0.05$. Be sure to label the critical values and the rejection region clearly. (5 points)



12. Light bulbs of a certain type are advertised as having an average lifetime of 800 hours. The price of these bulbs is very favorable, so a potential customer has decided to go ahead with a purchase arrangement unless it can be conclusively demonstrated that the true average lifetime is smaller than what is advertised. A random sample of 50 bulbs was selected, the lifetime of each bulb determined, and the appropriate hypotheses were tested using MINITAB, resulting in the accompanying output. In your analysis of the questions below, be sure to clearly state the test statistic you are comparing your results to.

Variable	N	Mean	St.Dev.	St.Error of Mean	Z	P-Value
Lifetime	50	738.44	38.20	5.4	-2.14	0.016

a. What conclusion would be appropriate for a significance level of .05? (3 points)

Ho: $\mu \geq 800$ Ha: $\mu < 800$ P-value $< .05 \Rightarrow$ reject Ho

b. A significance level of .01? (3 points)

at $\alpha = .01$ we would fail to reject the null

13. A university library ordinarily has a complete shelf inventory done once every year. Because of new shelving rules instituted the previous year, the head librarian believes it may be possible to save money by postponing the inventory. The librarian decides to select at random 1000 books from the library's collection and have them searched in a preliminary manner. If evidence indicates strongly that the true proportion of misshelved or unlocatable books is less than .02, then the inventory will be postponed.

a. Among the 1000 books searched, 15 were misshelved or unlocatable. Test the relevant hypotheses and advise the librarian what to do (use $\alpha = .05$.). (6 points)

$H_0: p \geq .02 \quad H_a: p < .02$

1 PropZtest
$p_0 = .02$
$x = 15$
$n = 1000 \quad < p_0$

P-value = .12936

fail to reject the Ho
conduct inventory

b. If the true proportion of misshelved and lost books is actually .01, what is the probability that the inventory will be (unnecessarily) taken? (3 points)

$$\beta(.01) = \Phi\left(\frac{.02 - .01 + 1.645\sqrt{.02(.98)/1000}}{\sqrt{.05(.95)/1000}}\right) = \Phi(5.49) \approx 1$$

c. If the true proportion is .05, what is the probability that the inventory will be postponed? (3 points)

$$\beta(.05) = \Phi\left(\frac{.02 - .05 + 1.645\sqrt{.02(.98)/1000}}{\sqrt{.05(.95)/1000}}\right) = \Phi(-3.30) = .0005$$

here its highly unlikely Ho will be rejected

14. Let μ denote the mean reaction time to a certain stimulus. For a large-sample z test of $H_0: \mu = 8$ versus $H_a: \mu > 8$, find the P-value associated with each of the given values of the z test statistics. (2 points each)

a. 1.52 $\quad 1 - \Phi(1.52) \approx .0643$

b. 0.95 $\quad 1 - \Phi(.95) \approx .1711$

c. 0.79 $\quad 1 - \Phi(.79) \approx .2148$

15. Give as much information as you can about the P-value of a t test in each of the following situations: (2 points each)

a. Upper-tailed test, df = 8, t = 2.15 $\quad$ 1 − tcdf(−E99, 2.15, 8) = .0318876...

b. Two-tailed test, df = 15, t = -1.69 $\quad$ 1 − tcdf(−1.69, 1.69, 15) = .1116975...

c. Upper-tailed test, df = 19, t = 2.539 $1 - \text{tcdf}(-E99, 2.539, 19) \approx .0100102\ldots$

d. Two-tailed test, df = 40, t = -4.5 $1 - \text{tcdf}(-4.5, 4.5, 40) \approx 5.73 \times 10^{-5}$

16. Consider the large-sample level .01 test for testing $H_0: p = .2$ versus $H_a: p > .2$.

a. For the alternative value p = .21, compute $\beta(.21)$ sample sizes n = 100, 2500, 10,000, and 90,000. (6 points)

$$\beta = \Phi\left(\frac{-.01 + .9320/\sqrt{n}}{.4073/\sqrt{n}}\right) = \Phi\left(\frac{-.01\sqrt{n} + .9320}{.4073}\right)$$

i.e. normalcdf(-E99, meso!)

b. For $\hat{p} = x/n = .21$, compute the p-value when n = 100, 2500, and 10,000. (5 points)

$n = 100$ $\beta = .979$
$n = 2500$ $\beta = .8556$
$n = 10,000$ $\beta = .4337$
$n = 90,000$ $\beta = 1.917 \times 10^{-7}$

$Z = .025\sqrt{n}$

$n = 100$ $P = .4013$
$n = 2500$ $P = .1056$
$n = 10,000$ $P = .0062$