Math 205
Spring 2015
Dr. Lily Yen

Final Exam

Show all your work

Name: ____________________
Number: ____________________
Signature: ____________________
Score: ___/100

Problem 1: Express your answer for each problem as a product of binomial coefficients, factorials, or exponents. One mark each. No steps required.

a. Find the number of ways you can arrange all the letters in **engineering**.

$$\boxed{\frac{11!}{3!\,3!\,2!\,2!} = 277\,200}$$

b. How many different ways can 10 multiple-choice questions each with 5 possible choices be answered?

$$\boxed{5^{10} = 9\,765\,625}$$

c. Roll a fair coin and a fair six-sided die simultaneously. How many events are in the sample space.

$$\boxed{12}$$

d. In a group of 12 students, one needs to be a class representative, and two need to plan social events. If one person can perform at most one task, how many ways can you form such a team with a designated representative and two social planners?

$$\boxed{\binom{12}{1}\binom{11}{2} = 660}$$

e. A deck of cards has thirteen cards of each suit. How many 5-card hands exist that have exactly 1 ace and 1 face card?

$$\boxed{\binom{4}{1}\binom{12}{1}\binom{36}{3} = 342\,720}$$

Score: /5

Problem 2: Express all your answers as simplified fractions. Two marks each. Short steps required.

a. A universe contains two events A and B. The probability of neither A nor B is 2/3. What is the probability of either (including both)?

$$\boxed{1/3}$$

$P(\overline{A \cup B}) = 1 - P(A \cup B)$, so $2/3 = 1 - P(A \cup B)$. Solving for $P(A \cup B)$ yields 1/3.

b. Roll a fair coin three times. What is the probability that you get at least 2 tails?

$$\boxed{1/2}$$

The sample space has 8 equally likely events. At least 2 tails in three rolls means $\{TTH, THT, HTT, TTT\}$. So $4/8 = 1/2$.

c. Roll two fair six-sided dice. What is the probability that you get a total of 3 given that at least one of the dice shows a 1?

$$\boxed{2/11}$$

The sample space contains 36 equally likely events, but 11 of which have at least one die showing a 1. Under this condition, 2 of them, that is, $\{(1,2),(2,1)\}$ have a total of 3 with at least one of the dice showing 1. Thus 2/11.

d. Draw five cards out of a 52-card poker deck. What is the probability that you have exactly 2 queens?

$$\boxed{\frac{47\times 46}{13\times 17\times 49} \approx 0.1996}$$

$\binom{4}{2} \times \binom{48}{3} / \binom{52}{5}$.

Score: /8

Problem 3: For each of the following random variables X, determine the type of distribution that best models X. Where possible, give values for the parameters of the distribution chosen. Give reasons for your choice of distribution.

a. At Capilano University, 200 students are enrolled in an introductory psychology course. The term papers in this course are graded by a team of teaching assistants: however, a sample of the papers is examined by the course professor for grading consistency. Experience suggests that 1% of all papers will be improperly graded. The professor selects 10 papers at random from the 200 submitted and examines them for grading inconsistencies. X is the number of papers in the sample that are improperly graded.

binomial random variable X where success is an improperly graded paper. The number of trials is 10, the number of papers selected at random. Since $10 = 5\% \times 200$, we can treat this sampling without replacement from a dichotomous population of size 200 as though it were exactly a binomial experiment. $b(x; 10, 0.01) = b(x; n, p)$ where $p = 0.01$, the percentage of improperly graded papers. If not approximated, then use exact hypergeometic distribution.

Score: /2

b. A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let p be the probability that a randomly selected couple agrees to participate. X is the number of couples who do not agree to the regimen before the fifth couple joins the study.

Negative binomial described by $nb(x; 5, p)$ since X is the number of failures before the fifth success.

Score: /2

c. Five orcas from the Pacific Ocean have been caught, tagged, and released to mix into the population. After they have had an opportunity to mix, a random sample of 10 orcas is selected. Let X be the number of tagged animals in the second sample. Assume that there are actually 25 orcas in this region.

Hypergeometric $h(x; 10, 5, 25)$ where $n = 10$, the sample size drawn from a population with 5 successes and a total population size of 25. Being tagged is a success.

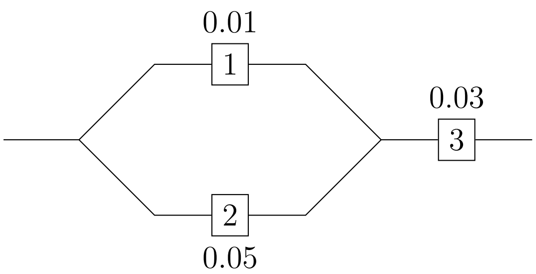
Score: /2

d. Customers arrive at an automated teller machine independently and at random. During lunch hour, customers arrive at the machine at a rate of one per minute on average. X is the number of people who arrive between 12:15 pm and 12:30 pm.

Poisson distribution with $\lambda = 1\,/\text{min}$.

Score: /2

Problem 4: Calculate the reliability of the system described in the figure below. The numbers beside each component represent the probabilities of failure for this component. Assume independence of components. Provide 4-decimal place accuracy.



Now $\boxed{1}$ and $\boxed{2}$ fail simultaneously with probability $0.01 \times 0.05 = 0.0005$. Therefore the left part of the system works with probability $1 - 0.0005 = 0.9995$.

For the whole system to work, both the left part and $\boxed{3}$ must work. Therefore the reliability of the system is $0.9995 \times (1 - 0.03) = 0.9695$.

Score: /4

Problem 5: A certain disease is known to affect 1% of the population. A test for the disease has the following features: If the person tested is contaminated, the test is positive with probability 0.98. On the other hand, if the person is healthy, the test is negative with probability 0.95.

a. What is the probability of a positive test when applied to a randomly chosen subject? Provide 4-decimal place accuracy.

Let B be the event that the test is positive, and let C be the event that the subject is contaminated. Then (recall that the symbol $\wedge$ means "and")
$P(B) = P(B \wedge C) + P(B \wedge \bar{C}) = P(C) \cdot P(B \mid C) + P(\bar{C}) \cdot P(B \mid \bar{C}) =$
$0.01 \cdot 0.98 + 0.99 \cdot 0.05 = 0.0593$.

b. What is the probability that an individual is affected by the disease after testing positive? Provide 4-decimal place accuracy.

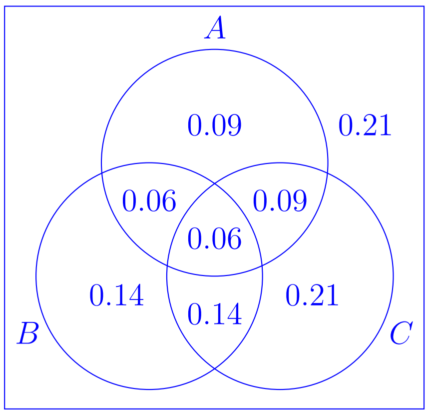
$P(C \mid B) = \frac{P(C \wedge B)}{P(B)} = \frac{P(B \wedge C)}{P(B)} = \frac{P(B|C) \cdot P(C)}{P(B)} = \frac{0.98 \cdot 0.01}{0.0593} = 0.1653$

Score: /4

Problem 6: Suppose that events A, B, and C are *mutually independent* with

$$P(A) = 0.3, \quad P(B) = 0.4, \quad P(C) = 0.5.$$

a. Draw a Venn Diagram labelling all probabilities.



Score: /4

b. Determine the following.

(1) $P(A \mid \bar{C})$

$P(A \mid \bar{C}) = \frac{P(A \cap \bar{C})}{P(\bar{C})} = \frac{0.09+0.06}{1-0.5} = 0.3$

(2) $P(A \cap \bar{B})$

$P(A \cap \bar{B}) = 0.09 + 0.09 = 0.18$

(3) $P(B \cup \bar{C})$

$P(B \cup \bar{C}) = (1 - 0.5) + 0.14 + 0.06 = 0.7$

Score: /3

Problem 7: Directly from the definitions of expected value and variance, compute $E(X)$ and $V(X)$ when X has probability mass function given by the following table:

X	-2	-1	0	1	2
$P(X)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{5}$	$\frac{4}{15}$	$\frac{1}{3}$

$$E(X) = -2 \times \tfrac{1}{15} + (-1) \times \tfrac{2}{15} + 0 \times \tfrac{1}{5} + 1 \times \tfrac{4}{15} + 2 \times \tfrac{2}{3} = \frac{2}{3}$$

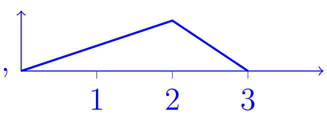
$$V(X) = (-2 - \tfrac{2}{3})^2 \tfrac{1}{15} + (-1 - \tfrac{2}{3})^2 \tfrac{2}{15} + (0 - \tfrac{2}{3})^2 \tfrac{1}{5} + (1 - \tfrac{2}{3})^2 \tfrac{4}{15} + (2 - \tfrac{2}{3})^2 \tfrac{2}{3} = \frac{14}{9}$$

Score: /4

Problem 8: The probability density of a random variable is given by

$$f(x) = \begin{cases} kx, & \text{if } 0 < x < 2, \\ 2k(3-x), & \text{if } 2 \le x < 3, \\ 0, & \text{otherwise.} \end{cases}$$

a. Find the exact value of k such that f is a probability density function.

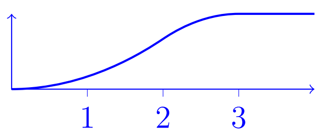
The graph of f is a triangle with base 3 and height $2k$, ,

therefore the area under the graph is $\frac{6k}{2} = 3k$. For f to be a probability density function, the area has to be 1, so $k = \frac{1}{3}$.

Score: /1

b. Find the corresponding cumulative distribution function with exact coefficients.

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0, & x \le 0 \\ \frac{1}{2}kx^2 = \frac{1}{6}x^2, & 0 < x < 2 \\ 2k(3x - \frac{1}{2}x^2) - 6k = 2x - \frac{1}{3}x^2 - 2, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$



Score: /3

c. Find the median exactly.

Now $F(2) = \frac{2}{3} > \frac{1}{2}$, so $F(x) = \frac{1}{2}$ when $\frac{1}{6}x^2 = \frac{1}{2}$, so $x = \sqrt{3}$.

Score: /2

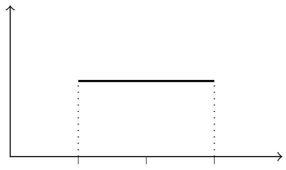
d. Find $P(1 \le x < 2)$.

$P(1 \le x < 2) = F(2) - F(1) = \frac{2}{3} - \frac{1}{6} = \frac{1}{2}$.

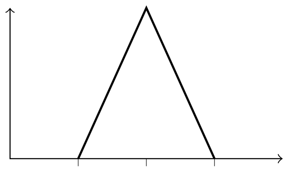
Score: /1

Problem 9: Match each of these probability density functions

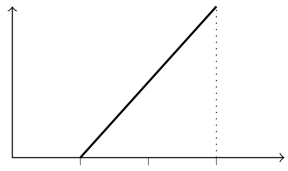
Answer: E



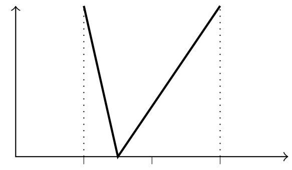
Answer: F



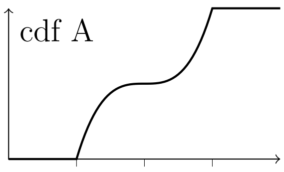
Answer: B

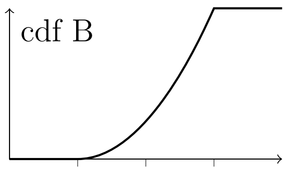


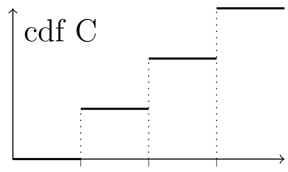
Answer: D

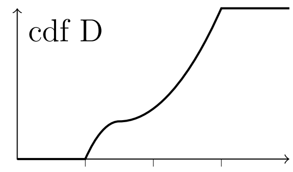


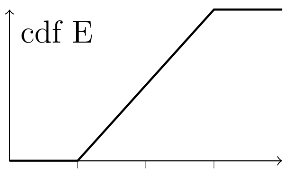
to one of these cumulative distribution functions

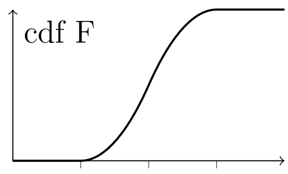












Score: /4

Problem 10: On average, 3 traffic accidents per month occur at Willingdon and Hastings. Find each probability in any given month. Clearly state the distribution with its corresponding parameters for each of the following.

a. Exactly 5 accidents will occur. Answer accurate to 4 decimal places.

You need a Poisson distribution with parameter $\mu = 3$. Then $f(5;3) = \frac{3^5 e^{-3}}{5!} = \frac{81}{40e^3} = \text{poissonpdf}(3,5) \approx 0.100\,82$

Score: /2

b. Fewer than 3 accidents will occur. Answer accurate to 4 decimal places.

$\text{poissoncdf}(3,2) \approx 0.423\,19$

Score: /2

c. At least 2 accidents will occur. Answer accurate to 4 decimal places.

$1 - \text{poissoncdf}(3,1) \approx 0.800\,85$

Score: /2

Problem 11: State the Central Limit Theorem.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then if n is sufficiently large, both $\bar{X}$ and $T_0 = \sum X_i$ have approximately a normal distribution with $\mu_{\bar{X}} = \mu$, $\sigma^2_{\bar{X}} = \sigma^2/n$ and $\mu_{T_0} = n\mu$, $\sigma^2_{T_0} = n\sigma^2$. The larger the value of n, the better the approximation.

Score: /2

Problem 12: Pure and Applied Science Division purchases large lots of erasable pens. The departmental assistant rejects a lot if 2 or more defective pens are found in a random sample of 100 pens.

a. Find the probability of rejecting a lot that is 1% defective? Use the Poisson approximation to the binomial. Specify all parameters used. Answer accurate to 4 decimal places.

Here $p = 0.01$ and $n = 100$, so $\mu = np = 1$. The probability of rejection is then $P(X \geq 2) = 1 - \text{poissoncdf}(1, 1) = 1 - \frac{2}{e} \approx 0.2642$.

Score: /3

b. Find the probability of accepting a lot that is 5% defective. Use the Normal approximation to the binomial. Answer accurate to 4 decimal places.

Here $p = 0.05$ and $n = 100$, so $\mu = np = 5$ and $\sigma = \sqrt{np(1-p)} = \frac{\sqrt{19}}{2} \approx 2.1794$.
With continuity correction the probability of rejection is then
$\text{normalcdf}(-\infty, 1.5, 5, 2.1794) \approx 0.0541$

Score: /3

Problem 13: Bags of Nonni's Focaccia Crouton have contents listed as 737 grams. Due to variability in production processes, the contents of these bags are actually normally distributed with a mean of 735 grams and a standard deviation of 20 grams. Give all answers to 6 decimal places.

a. What is the probability a randomly selected bag of Nonni's Focaccia Crouton has contents weighing between 730 and 745 grams?

$\text{normalcdf}(730, 745, 735, 20) = 0.2902$

Score: /1

b. Suppose you buy 10 bags of Nonni's Focaccia Crouton, find the probability at least 4 of them weigh between 730 and 745 grams?

This is a binomial experiment with $n = 10$ and $p = 0.2902$ (see the previous part). Therefore the answer is $1 - \text{binomialcdf}(10, 0.2902, 3) = 0.3244$.

Score: /2

c. Retail stores receive their Nonni's Focaccia Crouton in cartons of 30 bags. If you randomly select a carton and weigh the contents of each bag,

(1) what is the probability that the bags will have an average weight between 730 and 745 grams?

Here $\mu = 735$ and $\sigma = \frac{20}{\sqrt{30}} = 3.651$. Therefore the answer is $\text{normalcdf}(730, 745, 735, 3.651) = 0.9115$

Score: /3

(2) What is the probability the carton will have a total weight less than 22.5 kg?

Here $\mu = 30 \times 735 = 22050$ and $\sigma = 20\sqrt{30} = 109.5$. Therefore the answer is $\text{normalcdf}(-1 \times 10^{99}, 22500, 22050, 109.5) = 0.999\,980$

Score: /3

Problem 14: Osteoarthritis most frequently affects weight-bearing joints such as the knee. An article studying load redistribution at the knee joint in the elderly when ascending stairs and ramps presented the following summary data on stance duration (ms) for samples of both older and younger adults.

Age	Sample Size	Sample Mean	Sample SD
older	28	801	117
younger	16	780	72

Assume that both stance duration distributions are normal.

a. Calculate and interpret a 99% confidence interval for true average stance duration among elderly individuals. Answer accurate to 4 decimal places.

These are small samples, so you must use the t-distribution. Now $\alpha = 0.01$ and $n = 28$, so you have 27 degrees of freedom. Let μ_e be the true average stance duration among the elderly. Then $\mu_e \in \left(801 - t_{0.0025,27} \cdot \frac{117}{\sqrt{28}}, 801 + t_{0.0025,27} \cdot \frac{117}{\sqrt{28}}\right) = \left(801 - 2.7707 \cdot \frac{117}{\sqrt{28}}, 801 + 2.7707 \cdot \frac{117}{\sqrt{28}}\right) = (739.7376, 862.2624)$.

Score: /3

b. Perform a test of hypotheses at significance level 0.05 to decide whether true average stance duration is larger among elderly individuals than among younger individuals.

$\boxed{\text{STAT}}$ TESTS 4:2-SampTTest with Input Data as Stats, $\bar{x}_1 = 801$, $s_{x_1} = 117$, $n_1 = 28$, $\bar{x}_2 = 780$, $s_{x_2} = 72$, $n_2 = 16$, and $H_0 : \mu_1 > \mu_2$ yields that $t = 0.7366$ and $p = 0.2328$ with 41.6883 degrees of freedom.

Now $t_{0.05,41} = 1.6829 \not< 0.7366$, so the data does not provide enough reason to reject the hypothesis that the means are equal.

Score: /6

Problem 15: Six samples of each of four types of cereal grain grown in a certain region were analyzed to determine thiamin content, resulting in the following data (μg/g).

Wheat	5.2	4.5	6.0	6.1	6.7	5.8
Barley	6.5	8.0	6.1	7.5	5.9	5.6
Maize	5.8	4.7	6.4	4.9	6.0	5.2
Oats	8.3	6.1	7.8	7.0	5.5	7.2

a. Complete the following ANOVA table. Provide 4 decimal place accuracy.

Score: /3

$\boxed{\text{STAT}}$ <u>TESTS</u> H:ANOVA(

Source of Variation	df	Sum of squares	Mean Square	f	p
Treatments	3	8.9833	2.9944	3.9565	0.022 934
Error	20	15.137	0.756 83		
Total	23	24.120	3.7513		

You store data for each grain in lists in your graphing calculator: L_1, L_2, L_3, L_4.

b. State hypothesis and test statistic.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, where the μ_i's are the true mean thiamin content in wheat, barley, maize, and oats, respectively.
The test statistic is $p = 0.0229$ as listed above, or use $f = 3.9565$.

Score: /2

c. At level $\alpha = 0.05$, is there sufficient evidence to conclude that a difference in the mean thiamin content exist between two of the grains? Provide reason.

Since $p = 0.0229 < 0.05$, yes, at level 0.05, there is sufficient evidence to reject H_0 and conclude that at least two cereal grains have different mean thiamin content.
You can also use $F_{\alpha,3,20}$ by finding the intersection of the following two functions.

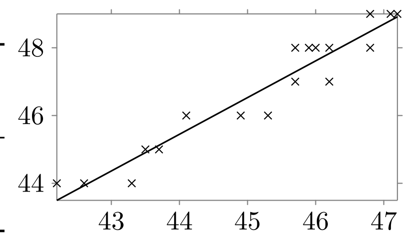
$$Y_1 = \text{Fcdf}(0, X, 3, 20), \quad \text{and} \quad Y_2 = 0.95$$

We obtain $F_{\alpha,3,20} = 3.0984$, and our test statistic $f > F$, in rejection region, thus we reject H_0, and conclude that there is sufficient evidence to conclude that a difference in the mean thiamin content exists between at least two of the grains at level 0.05.

Score: /2

Problem 16: Mopeds are very popular in Europe due to cost and ease of operation. However, they can be dangerous if performance characteristics are modified. One of the features commonly manipulated is the maximum speed. An article included a simple linear regression analysis of the variables x and y for test track speed (km/h) versus rolling test speed (km/h).

x	42.2	42.6	43.3	43.5	43.7	44.1	44.9	45.3	45.7
y	44	44	44	45	45	46	46	46	47
x	45.7	45.9	46.0	46.2	46.2	46.8	46.8	47.1	47.2
y	48	48	48	47	48	48	49	49	49



a. Does a scatter plot of the data support the strategy of a simple linear regression analysis? Sketch your plot next to the table and state the equation of the linear regression line accurate to 6 decimal places for each coefficient.

$\boxed{\text{STAT}}$ TESTS F:LinRegTTest (on the two lists and $\beta \& \rho \neq 0$) yields that $t = 13.879\,991$, $p = 2.432\,108 \times 10^{-10}$,

$$y = 1.083\,417x - 2.224\,140,$$

$s = 0.506\,890$, and $r = 0.960\,894$, so the correlation is strong.

Score: /2

b. What proportion of rolling test speed can be attributed to the simple linear regression relationship between test track speed and rolling test speed? Give 6-decimal place accuracy.

$r^2 = 0.923\,318$

Score: /1

c. Give a 95% confidence interval for the slope β_1 of the true regression line. Give 6-decimal place accuracy.

$\boxed{\text{STAT}}$ CALC 1:1-Var Stats on the list of x-values yields that $\sum x_i^2 = 36780.74$ and $\bar{x} = 45.1778$, so
$S_{XX} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 = 36780.74 - 18 \times 45.1778^2 = 42.134\,970$. Therefore $S_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{XX}}} = \frac{0.506\,890}{\sqrt{42.134\,970}} = 0.078\,089$
The confidence interval is then $\beta_1 \in \left(\hat{\beta}_1 - t_{0.025,16} S_{\hat{\beta}_1}, \hat{\beta}_1 + t_{0.025,16} S_{\hat{\beta}_1}\right) =$
$(1.083\,417 - 2.119\,905 \times 0.078\,089, 1.083\,417 + 2.119\,905 \times 0.078\,089) =$
$(0.917\,874, 1.248\,959)$.

Score: /4

d. Perform a model utility test by stating your null and alternative hypotheses, specifying the test statistic. State your conclusion.

$H_0 : \beta_1 = 0$.
In part a above, we found that $t = 13.879\,991$. This value is well over $t_{0.025,16} = 2.119\,905$, so the evidence resoundingly refutes H_0, and we conclude that $\beta_1 \neq 0$.

Score: /3