<u>Statistics Final Exam Questions</u>

Number of questions: 75 (45 theory questions + 30 Exercises)

1. Describe the differences and similarities between the population mean and the sample mean.
2. Describe the differences and similarities between the numerical variable and the categorical variable.
3. Describe the differences and similarities between the categorical variable and the dummy variable.
4. Describe the differences and similarities between the categorical and the ordered categorical variable.
5. Describe the differences and similarities between the discrete variable and the continuous variable.
6. Mean and Median: Definition, calculation methods, advantages and disadvantages
7. The Description of the Normal Distribution. Areas under the normal distribution
8. The characteristics of the standard normal distribution. The transformation of the normal distribution to the standard normal distribution
9. Discrete random variables. Discrete uniform distribution
10. Compare discrete frequency distribution with continuous frequency distribution
11. Standard deviation and coefficient of variation: calculation and interpretation of results
12. The Normal Distribution: The Empirical Rule
13. Skewness: Main types and their graphical representation
14. Kurtosis: Main types and their graphical representation
15. Variance and standard deviation: Definition, calculation methods
16. Covariance and Correlation: Definition and calculation methods
17. Mean and Variance of the Bernoulli random variable.
18. Expectation, Variance and Standard deviation of a discrete random variable.
19. Describe the differences and similarities between the sample mean and the expected value.
20. Description of Probability. Conditional Probability
21. Description of Probability. Joint Probability
22. Description of Probability. Marginal Probability

23. Description of Probability. Bayes Theorem
24. The Bernoulli Distribution
25. The Binomial Distribution
26. The IID Random variable. The IID Bernoulli Model
27. The Uniform Distribution
28. The Random Walk Model
29. The cumulative distribution function. The empirical Rule
30. The Central Limit theorem
31. Estimator and Estimate
32. Building Confidence Intervals for the IID Bernoulli random variable
33. Building Confidence Intervals for the mean of a numeric variable
34. Standard Error: Definition and calculation methods
35. The chi-square Distribution. Degrees of Freedom
36. The t-distribution
37. The "Plug-in" Predictive interval for the IID Normal
38. Hypothesis Tests and p-values for IID Bernoulli Data
39. Type I and Type II Errors
40. Hypothesis Tests: The rejection rule
41. The Simple Linear Regression Model
42. Introduction to hypothesis testing. Definition of null and alternative hypothesis. Steps in hypothesis testing.
43. Significance testing and confidence intervals
44. Ordinary Least Squares (OLS) Estimates
45. The Simple Regression: Confidence Intervals and Hypothesis Tests

1. Suppose the probability distribution of the random variable X is given by the following table:

X	0.02	0.04	0.07	0.1
P(X)	0.1	0.3	0.4	??

We might think of X as describing the return on an asset over the next year.

(a) What is Pr(X=0.1)?

(b) What is Pr(X>0.05)?

(c) What is E(X)?

(d) What is Var(X)?

(e) What is the standard deviation of X?

2. Suppose you roll a standard six-sided dice. Let X be the random variable which is a 1 if the dice comes up "6" and 0 otherwise.

   Note. (You can think of this random variable as a dice game where you need to roll a "6" to win. If you don't roll a "6", you don't really care whether you rolled a "1", "2", "3", "4", or "5", because you lose anyway.)

   (a) What is the distribution of the random variable X? (You can write out the probabilities, OR you can recognize that this is a 'special' distribution we talked about…)

   (b) What are the mean and variance of X?

3. Suppose X~Bernoulli(p).
   For what value of p is the mean and variance the smallest? And the largest?

4. Suppose we have a group of 10 voters: 5 Democrats and 5 republicans. We are about to pick two voters from this group of ten. Let Y1 be 1 if the first voter we pick is a democrat, and 0 otherwise. Let Y2 be 1 if the second voter we pick is a democrat, and 0 otherwise. Suppose we sample WITHOUT replacement. This means that after we ask the first voter whether s/he is a democrat or republican, we tell her/him to go home and take the second voter from the nine people who are left.

(a) What is the probability distribution of Y1? (Hint: It has a name!)
(b) What is the conditional probability distribution of Y2 given Y1=1?
(c) Find the joint distribution of Y1 and Y2. Write in a 2x2 table like we did in the notes.
(d) Find the marginal probability distribution of Y2 using your table.
(e) Are Y2 and Y1 identically distributed? Why?

5. Suppose we have a group of 10 voters: 5 Democrats and 5 republicans. We are about to

pick two voters from this group of ten. Let Y1 be 1 if the first voter we pick is a democrat, and 0 otherwise. Let Y2 be 1 if the second voter we pick is a democrat, and 0 otherwise. Suppose we sample WITHOUT replacement. This means that after we ask the first voter whether s/he is a democrat or republican, we tell her/him to go home and take the second voter from the nine people who are left.

(a) Now suppose we randomly choose a third voter (again without replacement) and let Y3 = 1 if the third voter is a democrat, and 0 otherwise.

Find the joint distribution of (Y1,Y2,Y3) by determining the probabilities p(y1,y2,y3) of each possible combination of values and entering them in the table below:

(y1, y2, y3)	p(y1, y2, y3)
(0,0,0)	
(0,0,1)	
(0,1,0)	
(0,1,1)	
(1,0,0)	
(1,0,1)	
(1,1,0)	
(1,1,1)	

Hint: (0,0,1) means the first two voters were republican and the third was democrat. Now suppose we sample WITH replacement. This means that after we ask the first voter whether s/he is a democrat or republican, we ask her/him to rejoin the group, and that person may then be selected again when we randomly choose the second voter.

6. The tables below give the joint distributions of a pair of random variables, (X,Y).

	X	
	5	**15**

Y	**5**	0.45	0.05
	15	0.05	0.45

(a) Compute the covariance and correlation between X and Y.

(b) Are X and Y independent?

(e) Are X and Y identically distributed? Are X and Y iid?

7. Suppose we have a casino game with random winnings, W where E(W)=0 and Var(W)=10. Also, if you play the game more than once, the winnings each time you play are i.i.d.

(a) Suppose that each time you play the game, the casino makes you pay a $1 "cover charge". So each time you play your winnings are Y = W – 1. If you play the game three times, your winnings are Z = Y1 + Y2 + Y3. What is E(Z)? What is Var(Z)?

(b) Now suppose you are given the option of tripling your bet. If you triple your bet, your winnings are given by V = 3W – 1. What is E(V)? What is Var(V)? Compare playing the game once and tripling your bet versus playing three times. On average, which strategy is more profitable? Which is riskier?

<u>Note. Starting with (c) and for the rest of the question, forget about the cover charge.</u>

(c) Suppose you play the game twice. Your TOTAL winnings are T = W1 + W2.

Your AVERAGE winnings are A = .5W1 + .5W2 = (1/2)(W1 + W2). What are E(T) and Var(T)? What are E(A) and Var(A)?

8. Calculate the results taking the cumulative distribution function into account:

(a) F(0)

(b) F(-1)

(c) F(1)

(d) F(-2)

(e) F(2)

9. Y is distributed N(500,10000). Find Pr(Y>696 or Y<304).

10. It is election night and you are working for a TV news service. You are covering a race between two candidates, Brown and Cameron. Assume that all voters in the UK vote for one of these two candidates and let p be the proportion of people who vote for Cameron.

Votes are now being counted, and the network executives want you to make a projection for which candidate will win. Though of course they want this projection as soon as possible, they are concerned with the network's credibility and are willing to accept at most a 5% chance that the projection is wrong.

(a) Suppose that 500 votes have been counted. The results are 263 votes for Cameron (the other 237 voted for Brown). Construct a 95% confidence interval for p. Can you project a winner in the election?

(b) Now suppose that 10,000 votes have been counted and 5,150 of them were votes for Cameron. Can you project a winner in the election?

(c) Now suppose you learn that the 10,000 voters in part (b) came from parts of the UK known to support Cameron's party, and that some parts of the country known to support Brown have not yet reported their results. Are you still willing to make a projection based on your confidence interval from (b)? Explain.

11. Suppose we had 200 parts, 40 of which were defective. We decided that the model is:

$$X1,2,\ldots,Xn \sim Bernoulli(p)$$

Where Xi is 1, if the ith part is defective and 0 otherwise.

Suppose $\hat{p}$=.2 and n=100.

(a) Calculate its standard error (s.e.).

(b) If I want half the standard errors, by what factor do I have to increase the sample size?

12. Suppose you are in charge of the "cereal box filling process" from the notes. You have lots of data on your process and are sure that the amount (in grams) of cereal put in each box can be

described as i.i.d. draws from the normal distribution with mean $\mu = 345$ and standard deviation $\sigma = 25$.

You are about to be audited by an inspector who will take a sample of 25 boxes and compute the average amount put in the 25 boxes. If that average is greater than 350 or less than 340, you will fail the audit and be reprimanded by your superiors.

(a) The inspector is about to look at the average amount in 25 boxes. Before the inspector begins looking at boxes, what is the probability distribution of this average? (This is the "sampling distribution" we talked about in class.)

(b) What is the probability that you will pass the audit?

(c) How would your answer to (b) change if the inspector took a sample of 100 boxes?

(d) If you are right about μ and σ, do you want the inspector to look at more boxes or fewer? Explain intuitively.

13. Consider the following simple model of an asset return, R. Let R be the random variable with the following possible values and corresponding probabilities

R	-0.05	-0.01	0.1	0.05	0.1	0.15
P(R)	0.1	0.1	0.2	0.3	0.2	0.1

(a) What is $P(0<R<0.15)$?

(b) What is the probability that the return is NEGATIVE?

(c) What is the expected return, E(R)?

(d) What is σ_R?

(e) Plot P(R) against R.

14. Suppose we have observations on the variable weight (in pounds) for 10 people:
100, 120, 210, 175, 155, 110, 160, 185, 200, 220.

(a) Calculate the sample mean and sample standard deviation.

(b) Suppose the scales underreport weight by 10 pounds. What is the mean and standard deviation of the correct weight in pounds?

(c) What is the mean and standard deviation of weight in kilograms? Note 1 kilogram equals 2.2046 pounds.

15. The Philadelphia office of Price Waterhouse Coopers LLP hired five accounting trainees this year. Their monthly starting salaries were: $3,536; $3,173; $3,448; $3,121; and $3,622.

(a) Compute the population mean.

(b) Compute the population variance.

(c) Compute the population standard deviation.

(d) The Pittsburgh office hired six trainees. Their mean monthly salary was $3,550, and the standard deviation was $250. Compare the two groups.

16. As you walk into your econometrics exam, a friend bets you £10 that she will outscore you in the exam. Let X be a random variable denoting your winnings. The variable X can take on the values 10, -10 or 0 (you tie on the exam). You know that the distribution function for X, P(X) depends on whether she studied for the exam or not. Let Y=0 if she studied and Y=1 if she did not. Consider the following joint distribution table:

	Y=0	**Y=1**	**P(X)**
X=-10	0.18	?	**?**
X=0	0	?	**0.3**
X=10	?	0.45	**?**
P(Y)	**?**	**0.75**	**?**

(a) Fill in the missing elements in the table

(b) Compute E(X) and E(Y). Should you take a bet? Why?

17. Consider the following joint probability distribution for the discrete random variables X and Y:

	Y=10	**Y=20**	**Y=30**
X=10	0.2	0.1	0.1
X=20	0.3	0.2	0.1

(a) What are the marginal distributions for Y and X?

(b) Calculate E(X) and E(Y).

(c) Calculate Var(X) and Var(Y).

18. The first card selected from a standard 52-card deck is a king.

(a) If it is returned to the deck, what is the probability that a king will be drawn on the second selection?

(b) If the king is not replaced, what is the probability that a king will be drawn on the second selection?

(c) What is the probability that a king will be selected on the first draw from the deck and another king on the second draw (assuming that the first king was not replaced)?

19. You are organizing an outdoor concert and believe that attendance will depend on the weather. You believe the following possibilities are appropriate:

Weather	**Probability**	**Attendance**
Terrible Weather	0.2	500
Mediocre Weather	0.6	1,000
Great Weather	0.2	2,000

(a) What is the expected attendance?

(b) Suppose each ticket costs £5 and the fixed costs (tents, bands, etc.) are £2,000. What are the expected profits?

20. Consider the following cumulative density distribution function:

X	**P(X<=x)**
5	0.23
10	0.34
15	0.41
20	1

If this distribution is based on 1000 observations, then what is the frequency in the second interval, i.e. (10, 15]?

21. Suppose that the price of a given company's common stock follows the random walk model,

$$P_{t+1}=P_t+U_{t+1}$$

Where P_{t+1} – the price at the close of trading tomorrow is, P_t - is today's closing price, and U_{t+1} – is the change in price during tomorrow's trading. Both P_t and U_{t+1} are in dollars ($).

Unlike the discrete random walk model we studied in class, this time we assume the price change each day, U_{t+1}, is normally distributed:

$$U_{t+1}\sim(1,4)\ i.i.d.\ t=0,1,2,3,\ldots$$

Notice that here $E(U_{t+1})=1$ so that the stock price has a tendency to increase over time. (This is called a random walk with drift). Trading just closed on the stock today at a price of P_0=$70.

(a) On a given day, what is the probability that the stock goes up OR down $3 or more, $P\ (U_{t+1}>3\ or\ U_{t+1}<-3)$?

(b) Let $P_1=P_0+U_1$ be the price of the stock at the close of trading tomorrow. Assuming that we observe today's closing price of P_0=$70, what is ($P_1$>$75)?

22. Suppose that the price of a given company's common stock follows the random walk model,

$$P_{t+1}=P_t+U_{t+1}$$

Where P_{t+1} – the price at the close of trading tomorrow is, P_t - is today's closing price, and U_{t+1} – is the change in price during tomorrow's trading. Both P_t and $U_{t+1}$are in dollars ($).

Unlike the discrete random walk model we studied in class, this time we assume the price change each day, U_{t+1}, is normally distributed:

$$U_{t+1}\sim(1,4)\ i.i.d.\ t=0,1,2,3,\ldots$$

Notice that here $E(U_{t+1})=1$ so that the stock price has a tendency to increase over time. (This is called a random walk with drift). Trading just closed on the stock today at a price of P_0=$70.

(a) What is the variance of the price the day after tomorrow, (P_2)?

(b) What is the probability that, 4 days from today, the stock is trading below its current price of $70, ($P_4<70$)?

23. Suppose we have observations on the variable weight (in pounds) for 10 people:

200, 320, 410, 75, 55, 210, 260, 285, 400, 120.

(a) Calculate the sample mean and sample standard deviation.

(b) Suppose the scales underreport weight by 10 pounds. What is the mean and standard deviation of the correct weight in pounds?

(c) What is the mean and standard deviation of weight in kilograms? Note 1 kilogram equals 2.2046 pounds.

24. Suppose we had 200 parts, 60 of which were defective. We decided that the model is:

$$X1,2,\ldots,Xn\sim Bernoulli(p)$$

Where Xi is 1, if the ith part is defective and 0 otherwise.

Suppose $\hat{p}$=.3 and n=200.

(a) Calculate its standard error (s.e.).

(b) If I want half the standard errors, by what factor do I have to increase the sample size?

25. Suppose the probability distribution of the random variable X is given by the following table:

X	0.02	0.04	0.07	0.1
P(X)	0.2	0.15	0.45	??

We might think of X as describing the return on an asset over the next year.

(a) What is Pr(X=0.1)?

(b) What is Pr(X>0.05)?

(c) What is E(X)?

(d) What is Var(X)?

(e) What is the standard deviation of X?

26. Y is distributed N(50,100). Find Pr(Y>70 or Y<30).

27. Consider the following joint probability distribution for the discrete random variables X and Y:

	Y=10	**Y=20**	**Y=30**
X=10	0.2	0.1	0.1
X=20	0.3	0.2	0.1

(a) What are the marginal distributions for Y and X?

(b) Calculate E(X) and E(Y).

(c) Calculate Var(X) and Var(Y).

28. Suppose that the price of a given company's common stock follows the random walk model,

$$P_{t+1}=P_t+U_{t+1}$$

Where P_{t+1} – the price at the close of trading tomorrow is, P_t - is today's closing price, and U_{t+1} – is the change in price during tomorrow's trading. Both P_t and $U_{t+1}$are in dollars ($).

Unlike the discrete random walk model we studied in class, this time we assume the price change each day, U_{t+1}, is normally distributed:

$$U_{t+1}\sim(1,4)\ i.i.d.\ t=0,1,2,3,\ldots$$

Notice that here $E(U_{t+1})=1$ so that the stock price has a tendency to increase over time. (This is called a random walk with drift). Trading just closed on the stock today at a price of P_0=$70.

(c) On a given day, what is the probability that the stock goes up OR down $5 or more, $P\ (U_{t+1}>5\ or\ U_{t+1}<-5)$?

(d) Let $P_1=P_0+U_1$ be the price of the stock at the close of trading tomorrow. Assuming that we observe today's closing price of P_0=$70, what is ($P_1$>$75)?

29. Suppose that the price of a given company's common stock follows the random walk model,

$$P_{t+1}=P_t+U_{t+1}$$

Where P_{t+1} – the price at the close of trading tomorrow is, P_t - is today's closing price, and U_{t+1} – is the change in price during tomorrow's trading. Both P_t and $U_{t+1}$are in dollars ($).

Unlike the discrete random walk model we studied in class, this time we assume the price change each day, U_{t+1}, is normally distributed:

$$U_{t+1}\sim(1,4)\ i.i.d.\ t=0,1,2,3,\ldots$$

Notice that here $E(U_{t+1})=1$ so that the stock price has a tendency to increase over time. (This is called a random walk with drift). Trading just closed on the stock today at a price of P_0=$70.

(c) What is the variance of the price the day after tomorrow, (P_3)?

(d) What is the probability that, 4 days from today, the stock is trading above its current price of \$70, ($P_4>70$)?

30. The tables below give the joint distributions of a pair of random variables (W,V).

		W	
		5	**15**
V	**5**	0.05	0.05
	15	0.45	0.45

(a) Compute the covariance and correlation between W and V.

(b) Are W and V independent?

(e) Are W and V identically distributed? Are W and V iid?