# MATH 1203 – Practice Exam 2

*This is a practice exam only. The actual exam may differ from this practice exam.*

1. Please state the Central Limit Theorem as discussed in class.

2. Please state, in your own words, what the following terms mean

* Contingency Table
* Chi-Square Test
* Least-Square Regression
* Confidence Intervals
* p-value of a Chi-square Test
* Correlation Coefficient r
* Scatter Plot
* Least Square Regression line
* P(event) = 0
* P(z < 1), where z has a standard normal distribution

3. Please decide if the following statements are true or false.

* If the p-value of a Chi-Square test is close to one, the association between two variables is strong.
* Both the p-value of a Chi-Square test and the correlation coefficient r tell you whether two variables are related, but the correlation coefficient r carries even more information.
* A Chi-Square test is appropriate for categorical variables, a regression analysis is appropriate for two numeric variables.
* The expected value in a cell of a contingency table tells you how many items would fall in that cell if the two variables were independent of one another.
* If an expected value in any cell of a contingency table is less than 5, then the two variables are dependent.
* Suppose you compute the equation of a least-square regression line as y = -2 x + 3 and the correlation coefficient r = 0.8, could that be possible?
* If r = 0.8, it means that two variables are strongly related in such a way that as x gets larger, the corresponding y gets smaller.
* P(z < 2) = 0.02211
* If X is N(10, 2) and X = 11, then the corresponding z-value is 2.1.
* If X is N(95, 10) then P(X > 105) = 0.21341
* A 95% confidence interval means that you are 95% certain that the true population mean is contained in the computed interval.
* A 99% confidence interval is *smaller* than a 90% confidence interval.

4. Compute the following probabilities:

* In tossing one coin twice, find P(HH) or P(exactly one head) or P(no head) or P(at least one head).
* In throwing two dice, find P(sum is 4) or P(sum = 1) or P(sum is 4 or more)
* In drawing one card randomly from a standard 52-card deck, find P(card is Ace)

5. A (hypothetical) frequency distribution for the age of people in a survey, the categories have the following probabilities:

|  |  |
| --- | --- |
| **Category** | **Probability** |
| 0 – 18 | 0.15 |
| 19-40 | 0.25 |
| 41-65 |  |
| 65 and older | 0.3 |

* One number is missing – what is that number?
* What is the chance that a randomly selected person is 40 years or younger?

6. Please consider the following data:

**Father’s schooling (years):** 8, 11, 5, 12

**Respondent’s schooling (years):** 8, 12, 5, 15

1. find the mean for both variables
2. create a scatter plot representing this data
3. draw a best-fit line through the scatter plot in part (b)
4. find the exact equation of the least-square regression line
5. compute Pearson’s r
6. predict the highest year of schooling for someone who’s father completed 14 years of school.

Recall the corresponding formulas:

  

1. The following scores were obtained as part of a sample with mean 10 and standard deviation 2. For each score, find the appropriate z-score: X = 10, X = 14, X = 6, X = -1. Then, for each z-score found, use the table at the end to find the probabilities of obtaining a score *less than or equal* to the computed z-score. Note: in mathematical notation this means that we want to find , where  is the computed z-score.

8. Each score listed below comes from a sample with the indicated mean and standard deviation. Convert each one to a z-score and find the indicated probability (in percent). Note that drawing a picture will help to find the indicated probabilities (percentages).

* X is normal with mean 3, standard deviation 1.5, find 
* X is normal with mean 3, standard deviation 3, find
* X is normal with mean 0, standard deviation 2, find
* X is normal with mean 3, standard deviation 1, find

9. Consider the following sample data, selected at random from some population:

12, 16, 5, 19

1. What is your best guess for the unknown population mean?
2. Find the standard error for the sample mean.
3. Find a 95% confidence interval for the unknown population mean.
4. Find a 99% confidence interval for the population mean. Explain why this interval differs from the previous one.
5. If you were to compute a 90% confidence interval, would it be wider or narrower than the previous two?

10. The table on page 592 in our text book can be used to compute probabilities for a variable z, assumed to have the Standard Normal Distribution N(0, 1). Use that table to find the following probabilities and shade the parts in the probability distribution that corresponds to the probability you computed.

   

  

11. When using StatCrunch to draw a scatter plot, it comes up with the following picture:



a) Draw a “best-fit” line through this data.

b) Use the line to estimate the y-intercept and slope of the equation of the least-square regression line

1. Look at the data and your line and estimate whether *r* would be close to -1, close to 0, or close to 1

12. Please consider the following results on a quiz, measuring scores before and after a certain lecture.



**Before lecture:** 5, 6, 7, 9, 3

**After lecture:** 6, 7, 9, 9, 4

* Create a scatter plot representing this data, including a best-fit line for the data
* Find the exact equation of the least-square regression line
* Compute the correlation coefficient r (use back page for computation but show r here)
* Predict the “after lecture” score for a “before lecture” score of 8.

13. When using *StatCrunch* for a linear regression analysis of pre-test versus post-test scores, it computes the output:



1. Find the equation of the least-square regression line
2. What is the correlation coefficient, and what does it mean
3. Predict the post-test score of someone with a pretest score of 77.
4. Do you think your prediction is accurate? Justify your answer using the correlation coefficient

14. The table below shows a contingency table for the variables “DEGREE” by “RACE”. Each cell lists three numbers: the count, the row, and the column percentage.

****

1. Which of the two variables is independent, which is the dependent variable?
2. Which number is the count, the row, and the column percentage
3. Compute the *expected value* for the cell “Whites with a High School degree”
4. How many *Blacks* have a high school degree, in percent?
5. How many people with a college degree, graduate or bachelor, are *White*, in percent?
6. How many *Blacks* have **at most** a junior college degree, in percent?

15. Consider the contingency table for religious preference versus political opinion, using our GSS survey below.

1. Compute the row percentage in the "Liberal and Catholic" cell, as well as column percentage and expected value.

****

1. Using *StatCrunch*, we conducted a Chi-Square test with the output as follows:



What is your conclusion?

1. Why should you compute and double-check all expected values in that table before finalizing your conclusion?

16. To investigate whether a relation exists between affiliation with a particular political party and the opinion on gun permits we used *StatCrunch* to create the following contingency table.

****

1. Based on that table, do you think there is strong evidence that the two variables associated, using common sense?
2. Based on your analysis in part (a), what do you think might be the p-value of a Chi-Square test for this data?

17. Use the GSS survey data to find the average number of siblings for people in the US in 1996 with reasonable accuracy. Note: Using *StatCrunch* we found that the descriptive statistics for the variable ‘sibs’ is as follows:

N 2897

Mean: 3.86

Standard Deviation: 3.52

18. When using *StatCrunch* to draw a “scatter plot, it comes up with the following picture:



a) Draw a “best-fit” line through this data.

b) Use the line to estimate the y-intercept and slope of the equation of the least-square regression line

c) Look at the data and your line and estimate whether r would be close to -1, close to 0, or close to 1.