

SOLUTIONS

Instructor: M. Pivovarova

Duration: 100 minutes.

Format: The test has **32** multiple-choice questions. Record answers on pink SCANTRON form. Total possible points: **88**

- Questions with two possible answers, **(A)-(B)**, are worth 1 point
- Questions with three possible answers, **(A)-(C)**, are worth 2 points
- Questions with five possible answers, **(A)-(E)**, are worth 5 points

Allowed aids: A calculator and aid sheets attached to this test.

Instructions for completing your SCANTRON form:

- Answers must be properly recorded on the pink SCANTRON form to earn marks
- Print your **LAST NAME** and **INITIALS** in the boxes AND darken each letter in the corresponding bracket below each box; Sign your name in the **SIGNATURE** box
- Print your 9 digit **STUDENT NUMBER** in the boxes AND darken each number in the corresponding bracket below each box
- Your **FORM NUMBER** is **01**
- Use only a pencil or blue or black ball point pen
 - **Pencil strongly recommended** because it can be erased
- Make dark solid marks that fill the bubble completely
- Erase completely any marks you want to change
- Crossing out a marked box is not acceptable and is incorrect
- If more than one answer is marked then that question earns 0 points

Instructions: Choose the best answer for each question. For questions with numeric answers that require rounding, round your final answer to be consistent with the choices offered. Use standard rounding rules.

DEFEND YOUR ACADEMIC INTEGRITY: Make sure to cover your answers and do not write your letter answers to each question in large font next to each question. Providing assistance to another student writing a test is as bad as receiving assistance and is treated equally harshly.

(1) If random variables X and Y are positively related which one of the inequalities below is true?

- ✓(A) $V[X+Y] > V[X] + V[Y]$
- (B) $V[X+Y] < V[X] - V[Y]$
- (C) $V[X+Y] < V[X] + V[Y]$

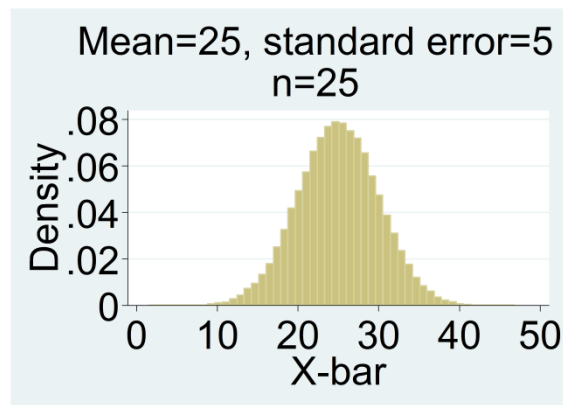
(2) Consider a random variable X that has a triangle distribution with parameters 10 and 20. The sampling distribution of the sample mean when sample size is equal to 36 is: (I did not count this question because of extremely poor performance)

- (A) $T(15, 2.08)$
- (B) $T(15, 4.17)$
- (C) $N(15, 4.17)$
- ✓(D) $N(15, 0.12)$
- (E) $N(15, 0.34)$

(3) In which of the following circumstances we could use Standard Normal table to estimate the odds of winning for the following candidates (n is a number of randomly selected voters):

- (A) Mr John Doe: 55% favour him and $n=10$
- (B) Mr Doe's opponent # 1: 30% favour him and $n=25$
- ✓(C) Mr Doe's opponent # 2: 15% favour him and $n=100$

► **For Question (4):** Consider the following histogram of the sampling distribution of sample mean.



(4) Population mean and standard deviation are:

- (A) 25 and 1
- (B) 25 and 5
- ✓(C) 25 and 25

► **For Questions (5)-(7):** Consider the following description of a sample drawn from the population of undergraduate students at the Big State University. Variable “hours” measures the number of hours spent on studying statistics in a regular week. Some of the output is deliberately erased.

hours

	Percentiles	Smallest		
1%	0	0		
5%	1	0		
10%	3.5	1	Obs	50
25%	6	1	Sum of Wgt.	50
50%	9		Mean	10.04
		Largest	Std. Dev.	
75%	14	19		
90%	18	19	Variance	32.65143
95%	19	20	Skewness	
99%	20	20	Kurtosis	

(5) The point estimate of the number of hours spent studying statistic among the undergraduate students at the Big State University is ____

- (A) 9
 (B) 10

(6) What percent of these data are more than two standard deviations below the mean?

- (A) 0 percent
 (B) at least 2.5 percent
 (C) about 5 percent

(7) The 95% confidence interval for the average number of hours spent studying statistics among undergraduate students at the Big State University is:

- (A) (7.42, 11.62)
 (B) (8.42, 11.66) (use critical values from t-table because population st.dev. is unknown)
 (C) (8.46, 11.70)
 (D) (8.71, 11.37)
 (E) not enough information provided

► **For Questions (8)-(10):** Suppose that the household income in a population is distributed with mean equal to \$25,000 and standard deviation equal to \$10,000.

(8) It is reasonable to conclude that the histogram of the household income is _____

- (A) Positively skewed
 (B) Bell-shaped
 (C) Negatively skewed

(9) If the poverty level is set at \$10,000, what fraction of households lives in poverty?

- (A) 0.0000
 (B) 0.0668
 (C) 0.4332

(D) 0.5668**(E)** 0.6680

(10) If you randomly sample 36 households, what is the chance that the average household income in a sample is greater than \$30,000?

✓**(A)** 0.0013**(B)** 0.4987**(C)** 0.5000**(D)** 0.5013**(E)** 0.0000

► **For Questions (11)-(14):** Below is a Stata summary of 225 randomly drawn records from the library at the Big State University.

X	Freq.	Percent	Cum.
0	52	23.11	23.11
1	71	31.56	54.67
2	32	14.22	68.89
3	48	21.33	90.22
4	10	4.44	94.67
6	9	4.00	98.67
10	3	1.33	100.00
Total	225	100.00	

(11) The random variable X counts the number of checked out books. The interquartile range of the distribution is _____

✓**(A)** 2**(B)** 3

(12) The range of the distribution is _____

(A) 9✓**(B)** 10

(13) These data are _____

✓**(A)** cross-sectional**(B)** longitudinal**(C)** time-series

(14) The median number of checked out books is _____

(A) 0✓**(B)** 1

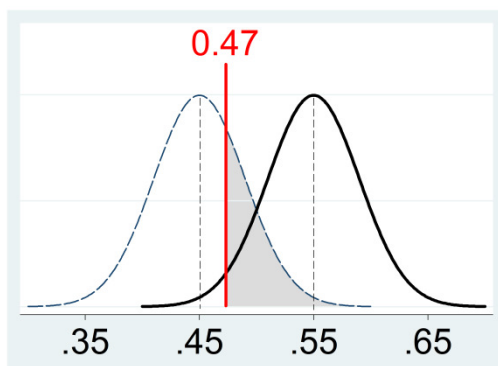
(15) In a family with 11 children what is the probability that there are more boys than girls (assume boys and girls are equally likely)?

- (A) 0.025
- (B) 0.250
- (C) 0.450
- ✓(D) 0.500
- (E) 0.750

(16) You've been hired by a union of workers who claim that there is a discriminatory practice against female employees; specifically, the average wage of female employees is lower than that of male employees for the same amount of work performed. You decide to use the method of hypothesis testing to shed the light on the problem. Your null hypothesis in this case is _____

- ✓(A) The wage is the same for female and male employees
- (B) The wage for male employees is different from the wage of female employees
- (C) The wage of male employees is higher than the wage of female employees

(17) The following graph displays the probability of Type II error (shaded area). The value of the parameter under the null hypothesis is _____



- (A) 0.45
- (B) 0.47
- ✓(C) 0.55

(18) A poll will be conducted to predict the winner of the upcoming mayoral election. The pollsters want to be 98 percent certain that their prediction will be within 2 percentage points of the actual results (that is ± 0.02). The sample size they need in order to reduce their margin of error to this level is (round intermediate values involved in the computations to the second decimal point):

- (A) 264
- (B) 2653
- ✓(C) 3364
- ✓(D) 3393
- (E) cannot be calculated because population proportion is unknown

(19) The latest poll results revealed that among consumers who were going to buy a new smart phone, 54% would buy an i-Phone. If the number of respondents who participated in the poll was 225, the 96% confidence interval of the fraction of population who would buy an i-Phone as their new smart phone is:

- ✓(A) (0.47, 0.61)
- (B) (0.48, 0.62)
- (C) (0.49, 0.63)

(20) When conducting hypothesis test, a researcher does not have the freedom to set:

- (A) significance level
- (B) probability of type I error
- (C) confidence level
- ✓(D) probability of Type II error
- (E) all of the above

(21) The TTC streetcars are known for their tardiness. Assuming that the average delay on route 510 is 5 minutes with standard deviation of 4 minutes, what proportion of streetcars is late by more than 9 minutes? (I did not count this question because of extremely poor performance)

- (A) about 16%
- (B) cannot be estimated because population proportion is unknown
- (C) cannot be estimated because sample size is unknown
- ✓(D) cannot be estimated because distribution is not normal
- (E) cannot be estimated because Central Limit Theorem does not hold

(22) In 2010, TTC lost \$22 million due to fare cheaters. Assume that on average, 4 percent of TTC riders do not pay the proper fare. Also assume that among 60 riders randomly selected for checks by transit enforcement officials more than 8 percent turned out to be "fare cheaters". Such a high percentage can be explained by: (I did not count this question because of extremely poor performance)

- (A) sampling error because the chance of this happening is more than 5%
- (B) non-sampling errors
- ✓(C) incorrectly estimated population parameter

(23) An average Canadian has a credit card balance of \$900 per month with standard deviation of \$400, and a mortgage payment of \$400 with standard deviation of 100 dollars. If credit card balance and mortgage payment are negatively related with correlation equal to -0.2, what is the total and standard deviation of what an average Canadian has to pay per month?

- (A) 1300 and 520
- ✓(B) 1300 and 392
- (C) 1300 and 125

(24) According to the statistics in question (23), mortgage payments are ___ volatile compared to the credit card balances.

- ✓(A) less
- (B) more

(25) A researcher wants to estimate a 90% confidence interval for unknown population mean. He computed sample mean and sample standard deviation from a sample of size 45. The critical value he is going to use for construction of his confidence interval is:

- (A) 1.301
- (B) 1.645
- ✓(C) 1.680

(26) “The survey polled 1,508 Canadians and is considered accurate within plus or minus 2.5% 19 times out of 20”. The underlined text implies that:

- (A) The confidence interval estimator has a significance level of 5%
- (B) The confidence interval estimator will miss population parameter 5 out of 100 times
- (C) The margin of error of the estimate is 0.025
- (D) None of the above
- ✓(E) All of the above

(27) The net weight of cereal boxes is normally distributed with the mean of 454 grams (this is what producer claims and states on the cereal box). 10 cereal boxes have been randomly selected and the average weight was found to be 430 grams with the sample standard deviation of 50 grams. What is the chance of observing such a low average weight among the sample of 10 boxes if the actual average weight is 454 grams?

- (A) 6 percent
- ✓(B) between 5 and 10 percent (value from t-table, not standard normal)
- (C) between 10 and 20 percent

(28) An Innovation Research Poll done for Maclean’s magazine shows that 27.1% of Canadians consider 9/11 to have been the most important international development of the last decade. If the margin of error for this estimate is 3% and the confidence level is 95%, how many people the online poll had?

- (A) 594
- ✓(B) 843
- (C) 1068

► **For Questions (29)-(30)** identify what kind of probability the underlined text indicates:

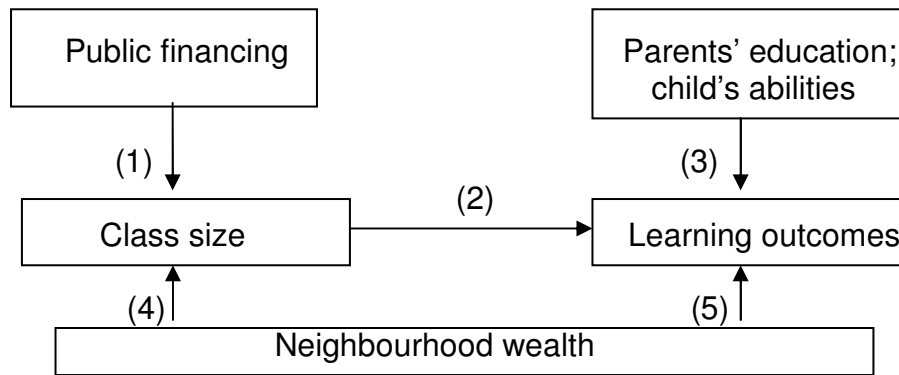
(29) The M&M’s Consumer Affair Department announced in 1995 that on average the new mix of colors of M&M’s Plain Chocolate Candies would contain 30% browns, 20% each of yellows and reds, and 10% each of oranges, greens and blues.

- (A) Conditional
- ✓(B) Marginal

(30) “The recent study of paramedics from Ontario and Nova Scotia involved 1381 respondents. Of respondents surveyed: 67.4 percent reported verbal abuse, with 62.9 percent of this group saying it was perpetrated by patients, 36.4 percent saying it was perpetrated by the patient’s family or friends and 20.8 percent by colleagues.”

- (A) Joint
- (B) Marginal
- ✓(C) Conditional

(31) Consider the following flow-chart similar to the one you saw in Lecture 5 describing a research question. Which arrows will be missing with experimental data?



- (A) 4, 5 and 1
- (B) 4 and 5 only
- (C) 1, 3, 4 and 5
- (D) 1 and 3 only
- ✓(E) 1 and 4 only

(32) In the question above, class size is an example of:

- ✓(A) independent variable
- (B) exogenous variable
- (C) lurking/confounding variable

Your **FORM** number is **01**. Complete the **FORM** box at the top right of your pink SCANTRON form. Do not forget to fill in your name and student number before the end of the exam is announced. No extra time is permitted.