Statistics 100
Sample Final Questions
(Note: These are mostly multiple choice, for extra practice. Your Final Exam will NOT have any multiple choice!)

Part A - Multiple Choice

Indicate the best choice for each question in the indicated space. Each correct answer is worth 2 marks. There is no penalty for incorrect answers. Select only one answer per question, or your answer will be counted as incorrect.

1. A sample of 400 Regina households is selected and several variables are recorded. Which of the following statements is correct?
    a) Total household income (in $) is interval level data.
    b) Socioeconomic status (recorded as "low income", "middle income", or "high income" is nominal level data.
    c) The number of people living in a household is a discrete variable.
    d) The primary language spoken in the household is ordinal level data.
    e) None of the above.

2. What term would best describe the shape of the given boxplot:

    a) Symmetric
    b) Negatively Skewed
    c) Positively Skewed
    d) Uniform
    e) Normal

3. We have seen that outliers can produce problematic results. Rank the following measures in order or "least affected by outliers" to "most affected by outliers".
    a) mean, median, range
    b) median, mean, range
    c) range, median, mean
    d) median, range, mean
    e) range, mean, median

4. Consider the result of a fictional Stat 100 final exam taken by 120 students, as given in the
   following relative frequency distribution:

Grade	Less than 50	50-59	60-69	70-79	80-89	90-100
Cumulative Frequency	15%	10%	30%	25%	15%	5%

   How many students received at least a 70 on this exam?
   a) 54
   b) 45
   c) 25
   d) 30
   e) 66

5. According to the empirical rule, approximately what percentage of normally distributed data
   lies within one standard deviation of the mean?
   a) 59%
   b) 68%
   c) 72%
   d) 95%
   e) 99.7%

Note: Questions #6-#8 are based on the following sample of ages (in months) of 18 children at a
   day care: 36, 42, 18, 32, 22, 22, 25, 29, 30, 31, 19, 24, 35, 29, 26, 36, 24, 28

6. The median age of the children is...
   a) 29
   b) 28.2
   c) 30.5
   d) 28.5
   e) 31

7. The interquartile range for this data set is...
   a) 8
   b) 12
   c) 16
   d) 20
   e) 24

8. The standard deviation of the age of children is...
   a) 41.24
   b) 11.33
   c) 10.20
   d) 6.42
   e) 6.24

Note: Questions #9-#12 are based on the following grouped frequency table of the income, x, of 30 employees at a local small business (in $1000s).

Income	$26<x\leq$28	$28<x\leq$30	$30<x\leq$32	$32<x\leq$34	$34<x\leq$36
Frequency	2	11	8	5	4

9. The relative cumulative frequency of the $28<x\leq$30 class is
   a) 11
   b) 0.43
   c) 0.06
   d) 13
   e) 0.7

10. The class that contains the 80th percentile is:
   a) $26<x\leq$28
   b) $28<x\leq$30
   c) $30<x\leq$32
   d) $32<x\leq$34
   e) $34<x\leq$36

11. Using class-midpoints as representative values, we can estimate the mean for this data as:
   a) $30,870
   b) $29,790
   c) $31,000
   d) $30,500
   e) $31,340

12. If the boss' income (the "31st employee") is $250,000, the mean income for all 31 workers is *approximately* equal to
   a) $8,000
   b) $30,000
   c) $38,000
   d) $140,000
   e) $220,000

Note: Questions #13-#17 are based on the following information: The cholesterol content of large chicken eggs is normally distributed with a mean of 200 milligrams and standard deviation 15 milligrams.

13. What proportion of these eggs have cholesterol content above 205 milligrams?
    a) 0.2004
    b) 0.6293
    c) 0.7250
    d) 0.3300
    e) 0.3707

14. What is the probability that the mean cholesterol content of a random sample of 25 of these eggs is less than 205 milligrams?
    a) 0.6293
    b) 0.9525
    c) 0.5793
    d) 1.67
    e) 0.0475

15. In sixty-seven percent of the eggs, the cholesterol content is less than a certain value "C". Find the value of "C".
    a) 0.33
    b) 206.6
    c) 210
    d) 0.44
    e) 193.4

16. To test if the claimed 200 mg value is reasonable, a sample of 60 large eggs is analyzed. Their average cholesterol content is 196 mg. What is the P-value based on this sample?
    a) 0.0197
    b) 0.5239
    c) 0.0524
    d) 0.0317
    e) 0.9803

17. We wish to estimate the cholesterol content in duck eggs. How large a sample should be selected if we can assume that σ=15 mg also holds for duck eggs, and we wish our estimate to be correct within 5 mg with 99% confidence?
    a) 8
    b) 99
    c) 35
    d) 75
    e) 60

18. The following data lists eight different investment amounts (X) and the amount of interest they earned (Y):

X	$1000	$2000	$3000	$10000	$500	$5000
Y	$50	$100	$150	$500	$25	$250

What is the best estimate for the regression coefficient r between X and Y?
a) r will be positive and close to zero
b) r will be positive and close to one
c) r will be exactly one
d) r will be larger than one
e) r will be negative

19. In a study of possible correlation between the height in cm (X) and weight in kg (Y) of chimpanzees, a sample of 40 animals produces a correlation coefficient of r=+0.813 and a regression line with equation $\hat{Y} = 0.34X + 19.5$. What is the expected weight of an 80 cm tall chimpanzee?
a) 46.7 kg
b) 177.9 kg
c) 24.0 kg
d) 34.8 kg
e) 57.1 kg

20. The following table lists the number of days that five houses had been up for sale, as well as their selling price.

X (days)	45	12	3	17	32
Y (1000$)	275	401	420	212	365

Calculate the correlation coefficient r between the number of days (X) and the selling price (Y) for this sample.
a) - 0.465
b) - 0.219
c) + 0.219
d) + 0.512
e) + 0.897

21. The Vitamin C content of a particular brand of vitamin supplement pills is normally distributed with mean 490 mg and standard deviation 12 mg. What is the probability that a randomly selected pill contains at least 500 mg of Vitamin C?
a) 0.7967
b) 0.8333
c) 0.0525
d) 0.1123
e) 0.2033

Note: Questions #22-#25 are based on the following information: The fine print on an instant lottery ticket claims that one in nine tickets win a prize.

22. What is the probability that you win at least once if you purchase five tickets?
    a) 0.4451
    b) 0.7919
    c) 0.1778
    d) 0.5549
    e) 0.3468

23. What is the probability that you win at least twice if you purchase ten tickets?
    a) 0.0907
    b) 0.3849
    c) 0.2165
    d) 0.3072
    e) 0.5013

24. What is the approximate probability that you win more than 120 times if you purchase 900 tickets?
    a) 0.5310
    b) 0.2408
    c) 0.0150
    d) 0.0012
    e) 0.9988

25. Of all awarded prizes, 10% are worth $1000, 20% are worth $100, and 70% are worth $10. Find the expected winnings if you purchase a single ticket.
    a) $127
    b) $370
    c) $1.41
    d) $14.11
    e) $42.33

---

26. At a large university, the probability that a student takes calculus and statistics in the same semester is 0.0125. The probability that a student takes statistics is 0.125. Find the probability that a student is taking calculus, given that he or she is taking statistics.
    a) 0.1
    b) 0.1125
    c) 0.0016
    d) 0.1375
    e) 0.4800

27. Suppose the respective probabilities are 0.7, 0.2, and 0.1 that a person applying for a driver's licence in Saskatchewan will require 1, 2, or 3 attempts in order to obtain a licence. Let X be a random variable, the number of attempts in order to obtain a licence. Find the mean (expected value) of X.

    a) 0.63
    b) 2
    c) 1
    d) 1.4
    e) 1.33

28. If $P(-2<Z<k)=.6$, where Z is a standard normal random variable, then k is...

    a) 0.5773
    b) 0.195
    c) 0.73
    d) 0.55
    e) -0.40

29. If you wanted to estimate the true percentage of all voters in Canada who are in favour of abolition of the senate, and if you wanted your maximum error of estimate to be 6% with a confidence level of 95%, what would the required sample size be?

    a) 267
    b) 1068
    c) 131
    d) 4272
    e) 851

30. If $\sigma_{\bar{x}}$ =3 for the sampling distribution of random samples of size 81 from a large population, what sample size (drawn from the same population) would have $\sigma_{\bar{x}}$=2.7?

    a) 27
    b) 100
    c) 108
    d) 73
    e) 90

31. A sample of 40 cows is drawn to estimate the mean weight of a large herd of cattle. If the standard deviation of the sample is 96 kg, what is the maximum error in a 90% confidence interval estimate?

    a) 25 kg
    b) 158 kg
    c) 58 kg
    d) 30 kg
    e) 1.645 kg

32. The P-value for a right-tailed test is P=0.042. Which of the following is INCORRECT?
   a) The P-value for a two-tailed test based on the same sample would be P=0.084
   b) The P-value for a left-tailed test based on the same sample would be P=-0.042
   c) The z-score test statistic is approximately z=1.73
   d) We would reject H0 at α=0.05, but not at α=0.01
   e) We would reject H0 at α=0.1, but not at α=0.04

33. To compare the average amount of time that Canadians and Americans spend commuting, a researcher collects a sample of 50 Canadians and 60 Americans. The Canadians spend an average of 4.6 hours a week commuting, with standard deviation 2.9 hours. The mean and standard deviation for the sample of Americans is 5.2 hours and 1.3 hours, respectively. The standard error of the difference of sample means is
   a) 2.314
   b) 1.023
   c) 0.443
   d) 0.196
   e) Cannot be computed as the sample sizes are different.

34. Which of the following statements about hypothesis tests is INCORRECT?
   a) H0 must always include equality.
   b) In a one-tailed test, H1 involves either ">" or "<", but not "≠".
   c) If the test statistic lies in the rejection region, we reject the claim.
   d) The claim can be either H0 or H1.
   e) The critical value indicates the start of the rejection region.

35. A researcher wishes to compare the proportion of men and women who voted in the last federal election. Which of the following conclusion could NOT be drawn as the result of any hypothesis test. "At the given level of significance, we have evidence..."
   a) "... that the proportion of men who voted is larger than the proportion of women who voted."
   b) "... that the proportion of men who voted differs from the proportion of women who voted."
   c) "... that the proportion of men who voted does not differ from the proportion of women who voted."
   d) "... that the proportion of men who voted is smaller than the proportion of women who voted."
   e) None of the above (each statement a-d is a possible conclusion).

Part B - Worked Problems

Show all of your work and necessary steps for full marks. Each question is worth 10 marks.

1. The manufacturer of a gasoline additive claims that the use of this additive increases gasoline mileage. A random sample of 6 cars was selected, and these cars were driven for one week without the gasoline additive and then for one week with the additive. The following table gives the mileage (in miles per gallon) for these cars with and without the additive:

Car	1	2	3	4	5	6
Mileage Without Additive	24.6	28.3	18.9	23.7	15.4	29.5
Mileage With Additive	26.3	31.7	18.2	25.3	18.3	30.9

Asuming mileages are normally distributed, can you conclude that the use of the gasoline additive increases mileage at the 2.5% significance level?

2. A researcher wished to estimate the difference between the proportion of users of two shampoos who are satisfied with the prodcut. In a sample of 400 users of Shampoo A taken by this researcher, 78 said they are satisifed. In another sample of 500 users of Shampoo B taken by the same researcher, 92 said they were satisfied.

a) Construct a 90% confidence interval for the true difference between the two population proportions.

b) Based on your answer in (a), what would be your response at $\alpha=0.1$ to the claim that there is no difference in customer satisfaction between the two shampoos?

3. A survey recently reported that the mean national annual expenditure for inpatient and outpatient services of all persons over 64 years of age was $5,423 with a standard deviation of $979. A random sample of 352 persons over age 64 living in Sudbury had an average expense of $5,516.

a) Can we conclude that the mean inpatient and outpatient expense of all Sudbury residents over age 64 is higher than the national average of $5,423? Use the .01 level of significance.

b) Calculate the P-value for this test. Would your answer have changed if you had used a .05 level of significance?

Answers:

Multiple Choice:

1C	2B	3B	4A	5B	6D	7A	8D	9B	10D
11A	12C	13E	14B	15B	16A	17E	18C	19A	20A
21E	22A	23D	24C	25D	26A	27D	28B	29A	30B
31A	32B	33C	34C	35C					

Worked Problems:

1.

Let new variable D be the difference, D="mileage with - mileage without".

Values for D are 1.7, 3.4, -0.7, 1.6, 2.9, 1.4, with mean D_{bar}=1.72, sandard deviation s_D=1.43.

Now conduct hypothesis test, assuming that population of such differences is normally distributed. Small sample: we must use t-values with d.f.=5

Step 1:
H0: μD=0
H1: μD>0 (claim)

Step 2:
This is a right-sided rejection region with critical value t=2.571.

Step 3:
Test statistic: t= (1.72 - 0) / (1.43/$\sqrt{6}$) = 2.94.

Step 4:
Reject H0.

Step 5:
At α=0.025 there is sufficient evidence for the claim that the additive increases mileage.

2.

a)

This is a confidence interval for difference of proportions. We use z-value z=1.645.

Here $p1\hat{} = 78/400$ and $p2\hat{}=92/500$, hence $(p1\hat{}-p2\hat{})=0.011$.

The standard error is $\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}} = 0.0263$, the maximum error is 0.0433.

Hence the 90% confidence interval for the difference in proportions is
$$-.0323 < p1-p2 < .0543$$

b)

Since the 90% confidence interval includes the zero value, we would not reject H0: p1=p2 in a two-sided hypothesis test with $\alpha=0.1$.

Hence, at $\alpha=0.1$ we would have no reason to reject the claim that there is no difference in customer satisfaction between the two shampoos.

3.

a)

This is a hypothesis test for one mean (note: just one sample taken and compared to a fixed value, this is NOT a two population problem). We have large samples (and note that $\sigma=\$979$ is given), hence use z-values.

Step 1:
H0: μ=5423
H1: μ>5423 (claim)

Step 2:
This is a right-sided rejection region with critical value z=2.326.

Step 3:
Test statistic is $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = 1.78$

Step 4:
Do not reject H0.

Step 5:
At $\alpha=0.01$ we cannot conclude that expenses are higher for Sudbury residents.

b)
P-value = P(z>1.78) = 0.0375.
t $\alpha=0.05$ we would reject H0 and hence support the claim that expenses are higher for Sudbury residents.