Sample Final Exam 1 – Part A

1. We would like to estimate the true mean amount (in $) consumers spent last year on Christmas gifts. We record the amount spent for a simple random sample of 30 consumers and we calculate a 95% confidence interval for μ to be (500, 545), i.e., the length of the interval is 45. The standard deviation σ of the amount spent by consumers is known. Suppose we had instead selected a simple random sample of 90 consumers and calculated a 95% confidence interval for μ. What would be the length of this interval?

   (A) 5.00 (B) 12.99 (C) 15.00 (D) 25.98 (E) 77.94

2. A statistician conducted a test of $H_0 : \mu = 20$ vs. $H_a : \mu < 20$ for the mean μ of some normally distributed population. Based on the gathered data, the statistician concluded that H_0 could be rejected at the 5% level of significance. Using the same data, which of the following statements **must** be true?

   (I) A test of $H_0 : \mu = 20$ vs. $H_a : \mu < 20$ at the 10% level of significance would also lead to rejecting H_0.

   (II) A test of $H_0 : \mu = 19$ vs. $H_a : \mu < 19$ at the 5% level of significance would also lead to rejecting H_0.

   (III) A test of $H_0 : \mu = 20$ vs. $H_a : \mu \neq 20$ at the 5% level of significance would also lead to rejecting H_0.

   (A) I only

   (B) II only

   (C) I and II only

   (D) I and III only

   (E) I, II and III

3. Prior to distributing a large shipment of bottled water, a beverage company would like to determine whether there is evidence that the true mean fill volume of all bottles differs from 600 ml, which is the amount printed on the labels. Fill volumes are known to follow a normal distribution with standard deviation 2.0 ml. A random sample of 25 bottles is selected. The sample has a mean of 598.8 ml and a standard deviation of 3.0 ml. What is the value of the test statistic for the appropriate test of significance?

(A) $t = -0.50$ (B) $z = -2.00$ (C) $t = -2.00$ (D) $z = -3.00$ (E) $t = -3.00$

4. Yields of apples per tree in a large orchard are known to follow a normal distribution with standard deviation 30 pounds. We will select a random sample of 25 trees and conduct a hypothesis test at the 1% level of significance to determine whether the true mean yield per tree is greater than 275 pounds. What is the probability of making a Type II error if the true mean yield per tree is actually 300 pounds?

(A) 0.0329 (B) 0.0537 (C) 0.0606 (D) 0.0838 (E) 0.0985

5. Nine runners are asked to run a 10-kilometer race on each of two consecutive weeks. In one of the races, the runners wear one brand of shoe and in the other a different brand (with the order randomly determined). All runners are timed and are asked to run their best in each race. The results (in minutes) are given below, with some sample calculations that may be useful:

Runner	1	2	3	4	5	6	7	8	9	Mean	Std. Dev.
Brand 1	31.2	29.3	30.5	32.2	33.1	31.5	30.7	31.1	33.0	31.40	1.22
Brand 2	32.0	29.0	30.9	32.7	33.0	31.6	31.3	31.2	33.3	31.67	1.31
Difference	−0.8	0.3	−0.4	−0.5	0.1	−0.1	−0.6	−0.1	−0.3	−0.27	0.35

We wish to conduct a hypothesis test to determine if there is evidence that average running times for the two brands differ. Suppose it is known that differences in times for the two brands follow a normal distribution. The P-value for the appropriate test of significance is:

(A) between 0.01 and 0.02

(B) between 0.02 and 0.04

(C) between 0.04 and 0.05

(D) between 0.05 and 0.10

(E) between 0.10 and 0.20

6. The blood cholesterol levels for randomly selected individuals were compared for two diets – one low fat and one normal. Some summary statistics are given in the table below:

	Sample Size	Sample Mean	Sample Std. Dev.
Low Fat	9	170	12
Normal	14	193	25

We would like to construct a confidence interval to estimate the difference between the true mean blood cholesterol level of all people on a low fat diet (μ_1) and all people on a normal diet (μ_2). The standard error of $\bar{X}_1 - \bar{X}_2$ is:

(A) 2.95 (B) 3.12 (C) 6.08 (D) 7.79 (E) 8.98

7. The following are the percentages of alcohol found in samples of two brands of beer, along with some sample statistics:

					$\bar{x}$	s
Brand A	5.4	5.6	5.7	5.3	5.5	0.183
Brand B	4.9	5.4	5.5	5.0	5.2	0.294

A 90% confidence interval for the difference in mean alcohol content for the two brands is:

(A) (0.015, 0.585)

(B) (0.002, 0.598)

(C) $(-0.022, 0.622)$

(D) $(-0.036, 0.636)$

(E) $(-0.051, 0.651)$

The next **two** questions (**8** and **9**) refer to the following:

A service centre for electronic equipment is conducting a study on three of their technicians: Joe, Bill, and John. The manager of the service centre wishes to assess if the average service times for their three technicians are equal. Each technician was given a random sample of disk drives, and the service time (in minutes) for each was recorded. Some sample statistics are given in the table below:

Technician	Sample Size	Sample Mean	Sample Std. Dev.
Joe	7	18	4
Bill	4	14	3
John	9	20	5

Service times for each of the three technicians are known to follow a normal distribution.

8. One assumption required in conducting an ANOVA F test is that all population variances are equal. The estimate of this common population variance is:

(A) 16.00 (B) 16.67 (C) 18.44 (D) 18.65 (E) 19.00

9. The value of the test statistic is calculated to be 2.63. What is the P-value of the appropriate test of significance?

   (A) between 0.001 and 0.01
   (B) between 0.01 and 0.025
   (C) between 0.025 and 0.05
   (D) between 0.05 and 0.10
   (E) greater than 0.10

10. The time it takes a student to drive to university in the morning follows a normal distribution with mean 28 minutes and standard deviation 4 minutes. The time it takes the student to drive home from university in the afternoon follows a normal distribution with mean 25 minutes and standard deviation 3 minutes. Morning and afternoon commuting times are known to be independent. What is the probability that it takes the student longer to drive to university than to drive home?

(A) 0.5517 (B) 0.6664 (C) 0.7257 (D) 0.8707 (E) 0.9987

The next **four** questions (**11** to **14**) refer to the following:

A hockey player compiles the following facts:

- Her team wins (W) 60% of their games.
- She scores a goal (G) in 45% of her games.
- She gets a penalty (P) in 40% of her games.
- In 76% of her games, her team wins or she scores a goal.
- In 24% of her games, her team wins and she gets a penalty.
- In 15% of her games, she scores a goal and gets a penalty.
- In 6% of her games, her team wins, she scores a goal and she gets a penalty.

11. In any given game, what is the probability that exactly two of the three events (win, goal, penalty) occur?

(A) 0.46 (B) 0.47 (C) 0.48 (D) 0.49 (E) 0.50

12. Which of the following statements is **true**?

   (A) W and G are independent.
   (B) G and P are mutually exclusive (disjoint).
   (C) W and P are independent.
   (D) W and G are mutually exclusive (disjoint).
   (E) G and P are independent.

13. What is the probaiblity the team wins if the player does not score a goal and does not get a penalty?

(A) 0.433 (B) 0.367 (C) 0.525 (D) 0.475 (E) 0.317

14. In a random sample of 60 games, what is the approximate probability that the player scores a goal in at least 30 of them?

(A) 0.2177 (B) 0.2215 (C) 0.2358 (D) 0.2420 (E) 0.2578

The next **two** questions (**15** and **16**) refer to the following:

A small deck of cards contains five red cards, four blue cards and one green card. We will shuffle the deck and select three cards without replacement. Let X be the number of blue cards that are selected. The probability distribution of X is shown below:

x	0	1	2	3
$P(X = x)$	0.167	0.500	0.300	0.033

15. What is the variance of X?

(A) 0.56 (B) 0.61 (C) 0.68 (D) 0.72 (E) 0.85

16. What would be the variance of X if we had instead selected the three cards **with replacement**?

(A) 0.56 (B) 0.61 (C) 0.68 (D) 0.72 (E) 0.85

17. Ten statisticians took separate random samples and each calculated a 90% confidence interval to estimate the value of some population parameter. What is the probability that exactly eight of the intervals contain the true value of the parameter?

(A) 0.1937

(B) 0.2166

(C) 0.2408

(D) 0.2691

(E) depends on which parameter they are trying to estimate

18. Harvey the clumsy waiter is in trouble at work because he has been breaking too many dishes. Suppose that the number of dishes he breaks follows a Poisson distribution with a rate of 0.11 per hour. Harvey works an 8-hour shift today, and his boss has warned him that if he breaks any dishes during the shift, he will be fired. What is the probability that Harvey will be fired today?

(A) 0.324 (B) 0.415 (C) 0.585 (D) 0.676 (E) 0.896

19. A random variable X follows a Poisson distribution with parameter λ. We are conducting a hypothesis test of H_0: $\lambda = 5$ vs. H_a: $\lambda < 5$. We decide to reject the null hypothesis if $X \leq 2$. What is the power of the test if $\lambda = 1$?

(A) 0.1912 (B) 0.3679 (C) 0.5518 (D) 0.7358 (E) 0.9197

20. The number of undergraduate students at the University of Winnipeg is approximately 9,000, while the University of Manitoba has approximately 27,000 undergraduate students. Suppose that, at each university, a simple random sample of 3% of the undergraduate students is selected and the following question is asked: "Do you approve of the provincial government's decision to lift the tuition freeze?" Suppose that, within each university, approximately 20% of undergraduate students favour this decision. What can be said about the sampling variability associated with the two sample proportions?

(A) The sample proportion from U of W has less sampling variability than that from U of M.

(B) The sample proportion from U of W has more sampling variability that that from U of M.

(C) The sample proportion from U of W has approximately the same sampling variability as that from U of M.

(D) It is impossible to make any statements about the sampling variability of the two sample proportions without taking many samples.

(E) It is impossible to make any statements about the sampling variability of the two sample proportions because the population sizes are different.

21. A local newspaper would like to estimate the true proportion of Winnipeggers who approve of the job being done by the mayor. A random sample of Winnipeggers is selected and each respondent is asked whether they approve of the job being done by the mayor. When the newspaper reports the results of the poll, it states, "results are accurate to within $\pm$ 3%, 19 times out of 20". What sample size was used for this poll?

(A) 672 (B) 720 (C) 894 (D) 936 (E) 1068

22. A Tim Horton's manager claims that more than 75% of customers purchase a coffee when they visit the store. We take a random sample of 225 customers and find that 180 of them purchased a coffee. We would like to conduct a hypothesis test to determine whether there is significant evidence to support the manager's claim. The P-value for the appropriate hypothesis test is:

(A) 0.0179 (B) 0.0256 (C) 0.0301 (D) 0.0359 (E) 0.0418

23. A random sample of voters is selected from each of Canada's three prairie provinces (Manitoba, Saskatchewan and Alberta) and respondents are asked which federal political party they support (Conservative, Green, Liberal or NDP). We would like to conduct a test at the 10% level of significance to determine whether voters in Canada's prairie provinces are homogeneous with respect to which political party they support. What is the critical value for the appropriate test of significance?

(A) 9.24 (B) 10.64 (C) 12.59 (D) 14.68 (E) 18.55

The next **two** questions (**24** and **25**) refer to the following:

A game of bowling consists of ten frames. A bowler scores a strike in a frame if he or she knocks down all the pins on the first shot. A bowler counted the number of strikes he has scored in his last 120 games. We would like to conduct a chi-square goodness-of-fit test at the 5% level of significance to determine whether the number of strikes per game for this bowler follows a binomial distribution. The data are shown in the table below, together with some expected cell counts:

# of strikes	0	1	2	3	4	5	6	7	8	9	10
# of games	6	10	16	28	27	18	11	3	0	1	0
Expected Count	1.62	8.70	21.08	30.26	???	???	8.27	2.54	0.52	0.06	0.00

24. Under the null hypothesis, what is the expected number of games in which the bowler scores five strikes?

(A) 18.43 (B) 19.78 (C) 20.92 (D) 21.65 (E) 22.37

25. What is the critical value for the appropriate test of significance?

(A) 9.49 (B) 12.59 (C) 14.07 (D) 15.51 (E) 16.92

26. The table below displays the number of accidents recorded at a particular intersection during each of the four seasons last year:

Season	Spring	Summer	Fall	Winter
# of accidents	13	24	18	25

We would like to conduct a chi-square goodness-of-fit test to determine whether accidents are uniformly distributed over the four seasons. The value of the test statistic for the appropriate test of significance is:

(A) 3.8 (B) 4.7 (C) 5.9 (D) 6.4 (E) 7.6

27. The height (in feet) and trunk diameter (in inches) are measured for a sample of 14 oak trees. The sample correlation between height and trunk diameter is calculated to be 0.68. We would like to conduct a test of H_0: $\rho = 0$ vs. H_a: $\rho \neq 0$ to determine whether there exists a linear relationship between the two variables. The P-value for the appropriate test of significance is:

(A) between 0.001 and 0.0025

(B) between 0.0025 and 0.005

(C) between 0.005 and 0.01

(D) between 0.01 and 0.02

(E) between 0.02 and 0.04

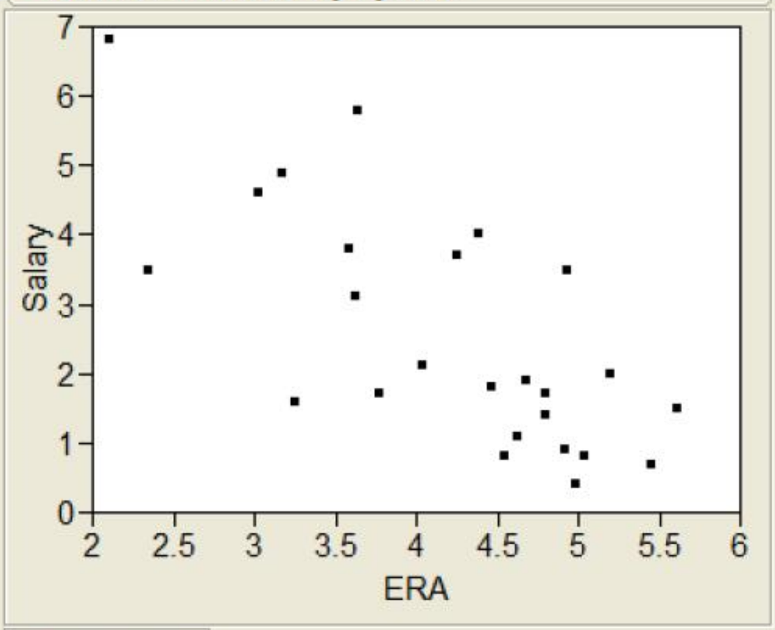
The next **three** questions (**28** to **30**) refer to the following:

Earned Run Average (ERA) is a common statistic used for pitchers in Major League Baseball. A pitcher's ERA is the average number of runs he gives up per game. The lower a pitcher's ERA, the better he is, and so we might expect him to be paid a higher salary. A sample of 25 pitchers is selected and their ERA X and Salary Y (in \$millions) are recorded.

From these data, we calculate $\bar{x} = 4.21, \bar{y} = 2.56$ and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 20.7$.

A least squares regression analysis is conducted. Some *JMP* output is shown below:

Bivariate Fit of Salary By ERA



Linear Fit

Linear Fit

Salary = 8.20 - 1.34 ERA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1		37.09	25.38
Error	23		1.46	Prob > F
C. Total	24			<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	8.20	1.14	7.16	<.0001*
ERA	-1.34	0.27	-5.04	<.0001*

28. What is the value of the sample correlation?

(A) -0.275 (B) -0.384 (C) -0.525 (D) -0.605 (E) -0.724

29. A 95% confidence interval for the mean salary of all Major League Baseball pitchers with an ERA of 4.00 is:

(A) (2.33, 3.35)
(B) (1.97, 3.71)
(C) (1.42, 4.26)
(D) (1.13, 4.55)
(E) (0.29, 5.39)

30. Which of the following intervals would be **wider** than the 95% confidence interval in the previous question?

(I) A 95% prediction interval for the salary of a pitcher with an ERA of 4.00
(II) A 99% confidence interval for the salary of all pitchers with an ERA of 4.00
(III) A 95% confidence interval for the salary of all pitchers with an ERA of 2.50

(A) I only

(B) II only

(C) I and II only

(D) I and III only

(E) I, II and III

Sample Final Exam 1 – Part B

1. The following three games are scheduled to be played at the World Curling Championship one morning. The values in parentheses are the probabilities of each team winning their respective game.

Game 1:	Finland (0.2)	vs.	Canada (0.8)
Game 2:	USA (0.3)	vs.	Switzerland (0.7)
Game 3:	Germany (0.4)	vs.	Japan (0.6)

   (a) The outcome of interest is the set of winners for each of the three games. List the complete sample space of outcomes and calculate the probability of each outcome.

   (b) Let X be the number of European teams that win their respective games. Find the probability distribution of X.

   (c) Find the expected value and variance of X.

   (d) If two European teams win their games, what is the probability that Finland is one of them?

2. A candidate for political office would like to determine whether his support differs among male and female voters. In a random sample of 300 male voters, 180 indicated they support the candidate. In a random sample of 200 female voters, 104 said they support the candidate.

   (a) Calculate a 95% confidence interval for the difference in proportions of male and female voters who support the candidate.

   (b) Provide an interpretation of the confidence interval in (a).

   (c) Conduct a two-sample z test to determine whether there is a significant difference between the proportions of male and female voters who support the candidate. Use the P-value approach and a 5% level of significance.

   (d) Provide an interpretation of the P-value of the test in (c).

3. Suppose we want to instead conduct the test in the previous question using a chi-square test for homogeneity. Conduct the test using the P-value approach and a 5% level of significance. Determine an **exact** P-value.