Sample Final Exam 1 – Part A

1. 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

2. The time it takes a curling rock to travel the length of the ice is known to follow a normal distribution with standard deviation 2.8 seconds. We will measure the times for a random sample of 12 rocks and conduct a hypothesis test at the 5% level of significance to determine whether the true mean time is greater than 20 seconds. What is the probability of making a Type II error if the true mean time is actually 21 seconds?

(A) 0.2358 (B) 0.3409 (C) 0.4078 (D) 0.5922 (E) 0.6591

3. We would like to conduct a hypothesis test to determine whether the true mean rent amount for all one-bedroom apartments in Winnipeg differs from \$850. We take a random sample of 50 one-bedroom apartments and calculate the sample mean to be \$900. A 98% confidence interval for μ is calculated to be **(830, 970)**. The conclusion for our test would be to:

(A) fail to reject H_0 at the 2% level of significance since the value 850 is contained in the 98% confidence interval.

(B) fail to reject H_0 at the 1% level of significance since the value 900 is contained in the 98% confidence interval.

(C) fail to reject H_0 at the 2% level of significance since the value 900 is contained in the 98% confidence interval.

(D) reject H_0 at the 2% level of significance since the value 850 is contained in the 98% confidence interval.

(E) reject H_0 at the 1% level of significance since the value 900 is contained in the 98% confidence interval.

4. 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	$\bar{x}$	s
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
Diff.	−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

5. We record the amount of time (in hours) spent watching TV per week for random samples of children (aged 6 – 12) and teenagers (aged 13 – 19). Some summary statistics are shown below:

	n	$\bar{x}$	s
Children	8	28.3	9.0
Teenagers	15	24.0	4.1

Weekly TV viewing times for both children and teenagers are known to follow normal distributions. We conduct a hypothesis test to determine whether children watch more TV than teenagers on average. The value of the test statistic for the appropriate test of significance is:

(A) 1.28 (B) 1.37 (C) 1.46 (D) 1.59 (E) 1.64

6. 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 (**7** and **8**) refer to the following:

A study was done to compare the fuel consumption (in litres) of different types of hybrid vehicles that were driven a distance of 1000 kilometres. Fuel consumption was measured for samples of seven hybrid cars, eight hybrid trucks and six hybrid SUVs. The ANOVA table is shown below, with some values missing:

Source of Variation	*df*	Sum of Squares	Mean Square	F
Groups				
Error			217	
Total		5292		

7. What is the P-value for 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

8. We would like to estimate the difference in the true mean fuel consumption for hybrid trucks and SUVs. The sample means and standard deviations for trucks and SUVs are shown below:

	mean	std. dev.
Trucks	87.34	10.56
SUVs	81.58	11.92

A 95% confidence interval for the difference in mean fuel consumptions for hybrid trucks and SUVs is:

(A) 5.76 ± 14.96

(B) 5.76 ± 12.65

(C) 5.76 ± 16.71

(D) 5.76 ± 17.34

(E) 5.76 ± 13.12

9. A brewery sells its beer in both cans and bottles. The amount of beer per can follows a normal distribution with mean 355 ml and standard deviation 2.0 ml. The amount of beer per bottle follows a normal distribution with mean 341 ml and standard deviation 1.5 ml. If you buy one can and one bottle, what is the probability that they contain more than 700 ml of beer in total?

(A) 0.0287 (B) 0.0359 (C) 0.0446 (D) 0.0548 (E) 0.0668

10. A sprinter runs two races on a certain day. The probability that he wins the first race is 0.6, the probability he wins the second race is 0.5, and the probability that he loses both races is 0.3. What is the probability that he wins both races?

(A) 0.2

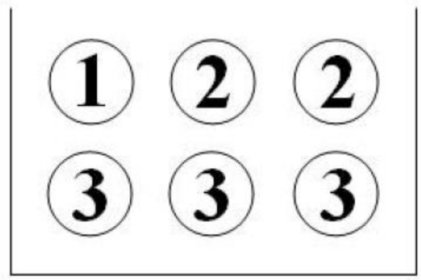
(B) 0.3

(C) 0.4

(D) 0.5

(E) This probability cannot be computed from the information given

11. We will randomly select two balls from the box below **without replacement**.



What is the probability that the sum of the two selected balls is at least 5?

(A) 0.2 (B) 0.3 (C) 0.4 (D) 0.5 (E) 0.6

The next **four** questions (**12** to **15**) 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.

12. 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

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

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

14. 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

15. 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 (**16** and **17**) refer to the following:

16. 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

The expected value of X is calculated to be $E(X) = 1.2$.

What is the variance of X?

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

17. 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

18. When an archer shoots an arrow, he hits the bullseye on the target 78% of the time. Whether he hits the bullseye on any shot is independent of any other shot. If he shoots ten arrows, what is the probability he hits the target at least 8 times?

(A) 0.2984 (B) 0.3185 (C) 0.4437 (D) 0.5689 (E) 0.6169

19. 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

20. A Monday class is held from 8:30 – 9:20 a.m. (50 minutes long). The number of students who fall asleep during the class follows a Poisson distribution with a rate of 0.05 per minute. What is the probability that exactly four students fall asleep in one class?

(A) 0.1336 (B) 0.1457 (C) 0.1592 (D) 0.1664 (E) 0.1748

21. A random variable X follows a Poisson distribution with parameter λ. If we know that $P(X = 1) = P(X = 3)$, then what is the value of λ?

(A) $\sqrt{2}$ (B) $\sqrt{3}$ (C) $\sqrt{5}$ (D) $\sqrt{6}$ (E) $\sqrt{8}$

22. 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.

23. 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

24. 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

25. 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 (**26** and **27**) 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, as well as 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

26. 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

27. 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

28. M & M's chocolate candies come in six different colours. The official M & M's website claims the following colour distribution:

Colour	Blue	Orange	Green	Yellow	Red	Brown
Proportion of M & M's	0.24	0.20	0.16	0.14	0.13	0.13

We take a random sample of 250 M & M's and count the number of candies of each colour. The sample data are shown below, as well as some cell chi-square values:

Colour	Blue	Orange	Green	Yellow	Red	Brown
Count	50	60	35	40	30	35
Expected	???	???	???	???	???	???
Cell χ^2	1.67	???	0.63	???	0.19	0.19

The value of the test statistic for the appropriate test of significance is:

(A) 3.86 (B) 4.04 (C) 4.45 (D) 4.97 (E) 5.39

29. 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

30. We take a random sample of individuals and measure the values of some explanatory variable X and some response variable Y. The least squares regression line is the line that minimizes:

(A) $\sum(y_i - \bar{y})^2$

(B) $\sum(y_i - \hat{y})$

(C) $\sum(\hat{y} - \bar{y})^2$

(D) $\sum(y_i - \bar{y})$

(E) $\sum(y_i - \hat{y})^2$

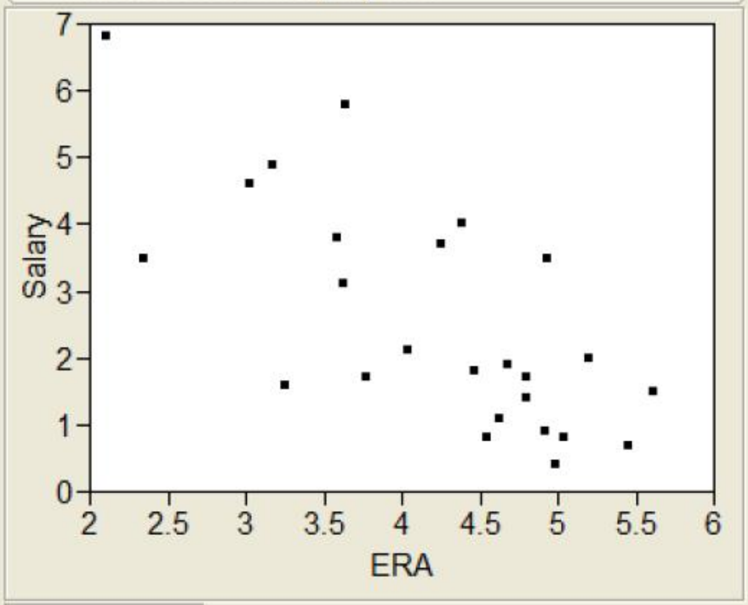
The next **three** questions (**31** to **33**) refer to the following:

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*

31. What is the value of the sample correlation?

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

32. 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)

33. 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.

4. We would like to use the weight of a car to predict its fuel economy. Weights (in 1000s of pounds) and fuel efficiency (in miles per gallon) are shown in the table below for a sample of ten car models:

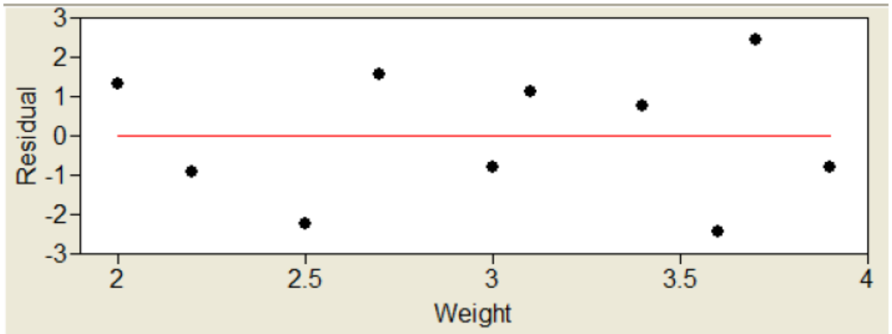
Weight	3.1	3.7	2.2	3.4	2.0	3.9	3.0	2.5	3.6	2.7
Fuel Efficiency	25	21	31	22	35	16	24	27	17	29

It can be shown that $\bar{x} = 3.01$, $s_x = 0.65$, $\bar{y} = 24.70$, $s_y = 6.02$, $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 3.8025$ and $\sum_{i=1}^{n}(y_i - \hat{y})^2 = 24.92$.

The equation of the least-squares regression line is calculated to be $\hat{y} = 51.47 - 8.89x$.

(a) Write out the least squares regression model and define all terms.

(b) The residual plot is shown below:



What does the residual plot tell you about the validity of the regression model in (a)?

(c) Construct a 95% confidence interval for the slope of the least squares regression line.

(d) Provide an interpretation of the confidence interval in (c).

(e) Conduct a test of H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$ to determine whether there exists a linear relationship between the weight of a car and its fuel efficiency.

(f) Provide an interpretation of the P-value of the test in (e).

(g) Calculate a 95% prediction interval for the fuel efficiency of a car that weighs 3200 pounds.