Sample Final Exam 1 – Part A

1. A statistician conducted a test of $\mathrm{H_0} : \mu = 20$ vs. $\mathrm{H_a} : \mu < 20$ for the mean μ of some normally distributed population. Based on the gathered data, the statistician concluded that $\mathrm{H_0}$ 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 $\mathrm{H_0} : \mu = 20$ vs. $\mathrm{H_a} : \mu < 20$ at the 10% level of significance would also lead to rejecting $\mathrm{H_0}$.
   (II) A test of $\mathrm{H_0} : \mu = 19$ vs. $\mathrm{H_a} : \mu < 19$ at the 5% level of significance would also lead to rejecting $\mathrm{H_0}$.
   (III) A test of $\mathrm{H_0} : \mu = 20$ vs. $\mathrm{H_a} : \mu \neq 20$ at the 5% level of significance would also lead to rejecting $\mathrm{H_0}$.

   (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

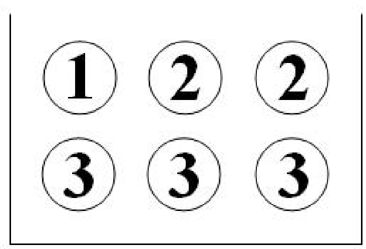
9. A brewery sells its beer in both cans and bottles. The amount of beer per can follows a normal distribution with mean 355 ml and standard deviation 2.0 ml. The amount of beer per bottle follows a normal distribution with mean 341 ml and standard deviation 1.5 ml. If you buy one can and one bottle, 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 ± 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:

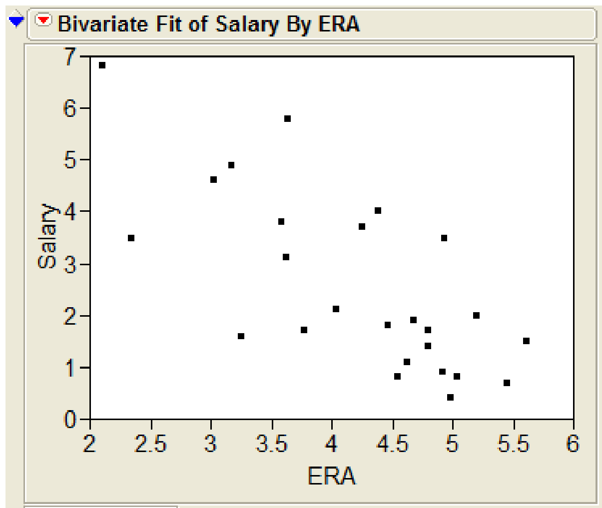
(A) $\sum(y_i - \bar{y})^2$
(B) $\sum(y_i - \hat{y})$
(C) $\sum(\hat{y} - \bar{y})^2$
(D) $\sum(y_i - \bar{y})$
(E) $\sum(y_i - \hat{y})^2$

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:



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:

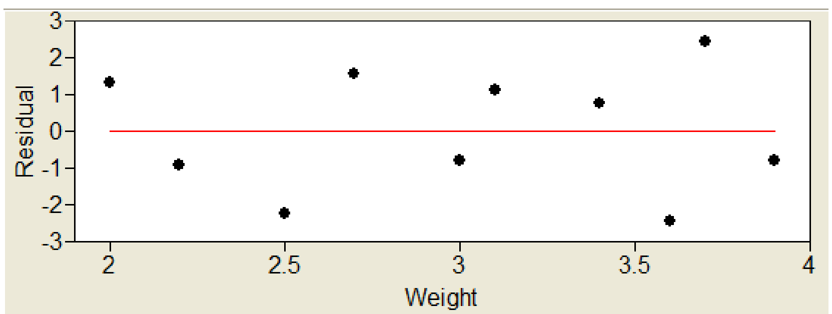
Weight	3.1	3.7	2.2	3.4	2.0	3.9	3.0	2.5	3.6	2.7
Fuel Efficiency	25	21	31	22	35	16	24	27	17	29

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: $\beta_1 \neq 0$ at the 5% level of significance 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.

Part A Answers

1. A
2. E
3. A
4. C
5. A

6. D
7. D
8. C
9. D
10. C

11. E
12. E
13. C
14. A
15. A

16. A
17. D
18. E
19. A
20. A

21. D
22. B
23. E
24. E
25. B

26. A
27. A
28. E
29. C
30. E

31. E
32. A
33. E

Part B Answers

1. (b) $P(X = 0) = 0.144, P(X = 1) = 0.468, P(X = 2) = 0.332, P(X = 3) = 0.056$
   (c) $E(X) = 1.3$, $Var(X) = 0.61$
   (d) 0.3253

2. (a) $(-0.0087, 0.1687)$
   (c) $z = 1.77$, P-value = 0.0768

3. $\chi^2 = 3.13$, P-value = 0.0768

4. (c) $(-10.98, -6.80)$
   (e) $t = -9.82$, P-value less than 0.001
   (g) (18.735, 27.309)

Sample Final Exam 2 – Part A

1. We would like to construct a confidence interval to estimate the true mean systolic blood pressure of all healthy adults to within 3 mm Hg. We have a sample of 36 adults available for testing. Systolic blood pressures of healthy adults are known to follow a normal distribution with standard deviation 14.04 mm Hg. What is the maximum confidence level that can be attained for our interval?

   (A) 80% (B) 90% (C) 95% (D) 96% (E) 98%

2. We would like to conduct a hypothesis test at the 2% level of significance to determine whether the true mean pH level in a lake differs from 7.0. Lake pH levels are known to follow a normal distribution. We take 10 water samples from random locations in the lake. For these samples, the mean pH level is 7.3 and the standard deviation is 0.37. Using the critical value approach, the decision rule would be to reject H_0 if the test statistic is:

   (A) less than -2.326 or greater than 2.326
   (B) less than -2.398 or greater than 2.398
   (C) less than -2.564 or greater than 2.564
   (D) less than -2.764 or greater than 2.764
   (E) less than -2.821 or greater than 2.821

3. In order to estimate the true mean GPA of all students in the University of Manitoba to within 0.05 with 96% confidence, we require a sample of 40 students. GPAs at the University are known to follow a normal distribution with a known standard deviation. How many students would we need to select to estimate the mean GPA of all University of Manitoba students to within 0.02 with 96% confidence?

   (A) 100 (B) 150 (C) 16 (D) 7 (E) 250

4. The GPAs of samples of students from two universities are recorded. Some summary statistics are shown in the table below:

	Sample Size	Sample Mean	Sample Variance
University 1	10	3.57	0.25
University 2	15	2.99	0.09

GPAs for students at both universities are known to follow normal distributions. We would like to conduct a hypothesis test to determine whether the true mean GPA for students at University 1 is greater than that for students at University 2. The value of the test statistic for the appropriate test of significance is:

(A) 3.29 (B) 3.64 (C) 7.04 (D) 7.57 (E) 8.29

5. We measure the time it takes to complete a certain task for samples of females and males. We would like to conduct a hypothesis test to determine whether females (F) can complete the task faster than males (M) on average. We will commit a Type I Error if we conclude that:

(A) $\mu_F > \mu_M$ when in fact $\mu_F = \mu_M$

(B) $\mu_F < \mu_M$ when in fact $\mu_F > \mu_M$

(C) $\mu_F = \mu_M$ when in fact $\mu_F < \mu_M$

(D) $\mu_F < \mu_M$ when in fact $\mu_F = \mu_M$

(E) $\mu_F > \mu_M$ when in fact $\mu_F < \mu_M$

6. We measure the heights (in cm) of a random sample of eight professional basketball players and a random sample of ten professional hockey players. Some JMP output is shown below:

Means and Std Deviations

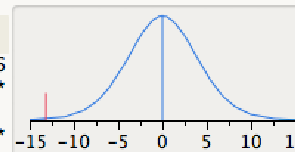
Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
Basketball	8	197.803	11.1494	3.9419	188.48	207.12
Hockey	10	184.658	4.9441	1.5635	181.12	188.19

t Test

Hockey-Basketball

Assuming equal variances

Difference	-13.144	t Ratio	
Std Err Dif		DF	16
Upper CL Dif		Prob > |t|	0.0040*
Lower CL Dif		Prob > t	0.9980
Confidence	0.95	Prob < t	0.0020*

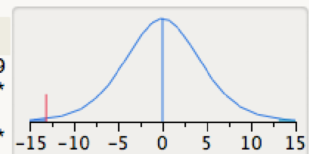


t Test

Hockey-Basketball

Assuming unequal variances

Difference	-13.144	t Ratio	
Std Err Dif		DF	9.198579
Upper CL Dif		Prob > |t|	0.0124*
Lower CL Dif		Prob > t	0.9938
Confidence	0.95	Prob < t	0.0062*



We would like to construct a 90% confidence interval to estimate the difference in the true mean heights of professional basketball and hockey players. The margin of error for the 90% confidence interval is:

(A) 6.71 (B) 6.83 (C) 7.49 (D) 7.68 (E) 7.77

The next **two** questions (**7** and **8**) refer to the following:

We would like to conduct an analysis of variance at the 5% level of significance to compare the mean percentage grades for students in five sections of a first-year university course. We take a simple random sample of students from each section. We assume that percentage scores follow normal distributions for each of the five sections. Some summary statistics are shown below:

Section	Sample Size	Sample Mean	Sample Std. Dev.
A01	3	74	8.9
A02	5	80	8.8
A03	6	59	13.6
A04	2	78	11.3
A05	4	66	15.3

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

(A) 2.71 (B) 2.90 (C) 3.06 (D) 4.56 (E) 5.86

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

(A) 11.58 (B) 11.72 (C) 11.88 (D) 12.10 (E) 12.17

9. A dish contains three cherry candies, four lemon candies and five grape candies. If we randomly select two candies from the dish without replacement, what is the probability that we get one cherry candy and one grape candy?

   (A) 0.1042 (B) 0.1136 (C) 0.1628 (D) 0.2084 (E) 0.2272

10. In a game of poker, you are dealt five cards from a deck of 52 cards. What is the probability that you get a flush (five cards of the same suit)?

   (A) 0.0005 (B) 0.0010 (C) 0.0020 (D) 0.0030 (E) 0.0040

11. In Vancouver, 42% of the days are rainy. Mary takes the bus 65% of the days. On 37% of days, she takes the bus and the weather is rainy. What is the probability that she takes the bus if the weather is not rainy?

   (A) 0.3595 (B) 0.3919 (C) 0.4363 (D) 0.4828 (E) 0.5424

12. A professor gives her class a short multiple choice quiz one day. There are three questions on the quiz. The first question has three possible answers (A, B, C), the second question has four possible answers (A, B, C, D) and the third question has five possible answers (A, B, C, D, E). An unprepared student randomly guesses the answer to each of the three questions. What is the probability that the student gets exactly one of the answers correct?

   (A) 0.3167 (B) 0.4333 (C) 0.5667 (D) 0.6725 (E) 0.7833

13. A swimmer is competing in two events at a swim meet. From past experience, he knows that his times in the 100-meter freestyle event follow a normal distribution with mean 53 seconds and standard deviation 0.7 seconds. His times in the 100-meter buttery follow a normal distribution with mean 55 seconds and standard deviation 0.9 seconds. He also knows that his times for the two events are independent. What is the probability that the swimmer has a faster (lower) time in the freestyle than in the butterfly?

   (A) 0.90 (B) 0.92 (C) 0.94 (D) 0.96 (E) 0.98

14. It is known that 17% of individuals in some population have blue eyes. If we take a random sample of 12 individuals from this population, what is the probability that four of them have blue eyes?

(A) 0.053 (B) 0.063 (C) 0.073 (D) 0.083 (E) 0.093

15. A random variable X has a binomial distribution with parameter $n = 3$. What must be the value of the parameter p in order for $P(X = 2) = P(X = 3)$?

(A) $\frac{1}{4}$ (B) $\frac{1}{3}$ (C) $\frac{1}{2}$ (D) $\frac{2}{3}$ (E) $\frac{3}{4}$

16. A taxi can accommodate anywhere from one to four passengers at a time. The number of passengers X per ride for one taxi has the probability distribution shown below:

x	1	2	3	4
$P(X = x)$	0.3	0.4	0.2	0.1

What is the variance of X?

(A) 0.89 (B) 0.94 (C) 1.05 (D) 1.17 (E) 1.26

17. The number of goals scored in a hockey game follows a Poisson distribution with a mean of 0.1 per minute. What is the probability that at least two goals are scored in a 60-minute game?

(A) 0.9554 (B) 0.9723 (C) 0.9826 (D) 0.9899 (E) 0.9937

18. A car company reports that the number of breakdowns per shift on its machine-operated assembly line follows a Poisson distribution with a mean of 1.5. Assuming that the machine operates independently across shifts, what is the probability of no breakdowns during three consecutive shifts?

(A) 0.0111 (B) 0.0498 (C) 0.0744 (D) 0.1923 (E) 0.2065

19. Suppose it is known that 83% of motorists wear a seatbelt while driving. The police stop a random sample of 200 drivers. What is the probability that more than 80% of them are wearing a seat belt?

(A) 0.8023 (B) 0.8212 (C) 0.8554 (D) 0.8708 (E) 0.8997

20. We would like to estimate the true proportion of students at a large university who are female. What sample size do we require in order to estimate the true proportion to within 0.05 with 98% confidence?

(A) 136 (B) 379 (C) 542 (D) 1127 (E) 2165

21. In a random sample of 250 Winnipeggers, 80 of them said they use transit. A 93% confidence interval for the true proportion of all Winnipeggers who use transit is:

(A) (0.257, 0.383)
(B) (0.262, 0.378)
(C) (0.267, 0.373)
(D) (0.272, 0.368)
(E) (0.277, 0.363)

22. The Conservative Party of Canada received 39.6% of all votes cast in the 2011 federal election. We would like to conduct a hypothesis test to determine whether the proportion of Canadian voters who support the Conservative party has decreased since that election. In a random sample of 200 voters last week, 70 said they support the party. What is the P-value for the appropriate test of significance?

(A) 0.0516 (B) 0.0643 (C) 0.0721 (D) 0.0838 (E) 0.0918

23. The Catholic Church does not allow women to become priests. We will take a random sample of 300 Catholics and ask them if they believe women should be allowed to become priests. We would like to conduct a hypothesis test at the 10% level of significance to determine whether the majority of Catholics support the idea. What is the power of the test if the true proportion of Catholics who favour the idea is actually 0.56?

(A) 0.7357 (B) 0.7881 (C) 0.8025 (D) 0.8238 (E) 0.8599

24. In a random sample of 250 females, 60 of them said they smoke cigarettes. In a random sample of 200 males, 40 of them said they smoke cigarettes. We would like to conduct a hypothesis test to determine whether there is evidence that the true proportion of females who smoke cigarettes is higher than that for males. The value of the test statistic for the appropriate test of significance is:

(A) 1.01 (B) 1.12 (C) 1.23 (D) 1.34 (E) 1.45

25. A clinical trial of Nasonex was conducted, in which 400 randomly selected pediatric patients (ages 3 to 11 years old) were randomly divided into two groups. The patients in Group 1 (experimental group) received 200 mcg of Nasonex, while the patients in Group 2 (control group) received a placebo. We conduct a test of H_0: $p_1 = p_2$ vs. H_a: $p_1 \neq p_2$ to compare the proportions of subjects in the two groups who experienced headaches as a side effect. The test statistic is calculted to be 2.25 and the P-value is 0.0244. Suppose we had instead compared the two proportions using a chi-square test for homogeneity. The value of the test statistic and the P-value would be, respectively:

(A) 5.06 and 0.0244
(B) 2.25 and 0.0006
(C) 1.50 and 0.0244
(D) 5.06 and 0.0006
(E) 2.25 and 0.0244

The next **two** questions (**26** and **27**) refer to the following:

We would like to conduct a test of significance at the 10% level of significance to determine whether smoking behaviour of university students is independent of their parents' smoking behaviour. The data is displayed in the table below, as well as some expected cell counts and cell chi-square values:

Observed Expected Cell Chi-Square	Student Smokes	Student Doesn't Smoke	Row Total
Neither Parent Smokes	17 ??? 0.81	62 ??? 0.29	79
One Parent Smokes	11 ??? 0.51	40 ??? ???	51
Both Parents Smoke	14 7.22 ???	13 19.78 2.32	27
Column Total	42	115	157

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

(A) 4.61 (B) 5.99 (C) 7.78 (D) 9.24 (E) 10.64

27. What is the value of the test statistic for the appropriate test of significance?

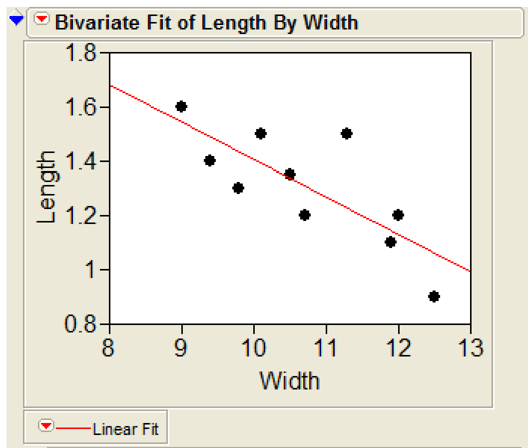
(A) 4.92 (B) 6.54 (C) 8.17 (D) 10.49 (E) 12.31

28. According to the Hershey chocolate company, 50% of its Reese's Pieces candies are orange, 25% are yellow and 25% are brown. Suppose you take a random sample of 160 candies and find that 72 are orange, 56 are yellow and 32 are brown. We would like to conduct a goodness of fit test to verify the company's claim. The value of the test statistic for the appropriate test of significance is:

(A) 8.8 (B) 9.2 (C) 9.8 (D) 10.2 (E) 10.8

The next **two** questions (**29** and **30**) refer to the following:

A study examined the relationship between the sepal width and the sepal length for a certain variety of tropical plant. Some *JMP* output is shown below:



Linear Fit

Length = 2.796 - 0.139 Width

Summary of Fit

RSquare	0.602
RSquare Adj	0.552
Root Mean Square Error	0.141
Mean of Response	1.305
Observations (or Sum Wgts)	10

29. One plant in the sample had a sepal width of 10.7 and a sepal length of 1.2. What is the value of the residual for this plant?

(A) 0.1087 (B) -1.3087 (C) 0.3087 (D) 1.3087 (E) -0.1087

30. 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 sepal width and sepal length. The value of the test statistic for the appropriate test of significance is:

(A) -3.48 (B) -3.14 (C) -2.13 (D) 2.13 (E) 3.48

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

Can the age of a cow be used to predict its milk production? The ages of eight cows (in years) and their milk production (in gallons per week) are shown below:

Age	4	4	6	7	7	8	10	11
Milk Production	37.0	35.4	33.3	35.6	32.3	33.7	32.1	29.6

A regression analysis is run and the equation of the least squares regression line is found to be $\hat{y} = 39.297 - 0.796x$. It is also determined that 73.2% of the variation in milk production can be accounted for by age. We also calculate $\sum(y_i - \hat{y})^2 = 7.90$ and $\sum(x_i - \bar{x})^2 = 44.9$.

31. What is the sample correlation between age and milk production?

(A) -0.536 (B) -0.732 (C) -0.796 (D) -0.856 (E) -0.892

32. A 90% confidence interval for the parameter β_1 in the linear regression model is:

(A) $(-1.018, -0.574)$
(B) $(-1.129, -0.463)$
(C) $(-1.240, -0.352)$
(D) $(-1.351, -0.241)$
(E) $(-1.462, -0.130)$

33. We conduct a hypothesis test of H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 < 0$ to determine whether there exists a negative linear relationship between the age of a cow and its milk production. 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 0.025

Sample Final Exam 2 – Part B

1. (a) A man has a few drinks one night at the bar and is deciding whether to drive home at the end of the night. He is essentially testing the hypoteses

   H_0: My blood-alcohol level is below the legal limit vs.
   H_a: My blood-alcohol level is above the legal limit

   Explain what it would mean in the context of this example to make a Type I error and a Type II error. Explain the potential consequences of each type of error.

   (b) The strengths of prestressing wires manufactured by a steel company have a mean of 2000 N and a standard deviation of 100 N. By employing a new manufacturing technique, the company claims that the mean strength will be increased. To verify this claim, a builder will test a random sample of 36 wires produced by the new process and will conduct a hypothesis test of H_0: $\mu = 2000$ vs. H_a: $\mu > 2000$ at the 10% level of significance. What would be the power of the test if the true mean strength of wires produced by the new process was 2050 N?

2. Measurements of the right-hand and left-hand gripping strengths of six people are measured. The data are shown in the table below with some summary statistics:

Subject	1	2	3	4	5	6	mean	std. dev.
Right	120	97	116	88	126	107	109	14.5
Left	107	92	120	86	116	103	104	13.3
Diff. (Right – Left)	13	5	-4	2	10	4	5	6.0

   (a) We would like to conduct inference procedures to estimate and test for the mean difference in right-hand and left-hand gripping strengths. What assumptions are necessary?

   (b) Assuming all appropriate assumptions are satisfied, calculate a 95% confidence interval for the true mean difference in gripping strength between people's right and left hands.

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

(d) Conduct a hypothesis test at the 5% level of significance to determine whether there is a difference in mean right-hand and left-hand gripping strengths in the population. Use the P-value method.

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

(f) Could you have used the confidence interval in (b) to conduct the test in (d)? Why or why not? If you could have used the interval to conduct the test, what would your conclusion be, and why?

3. A household gets the newspaper delivered every weekday (Monday to Friday). The resident counts the number of times the newspaper is on time for a sample of 80 weeks. The data are shown in the table below:

# of days on time	0	1	2	3	4	5
# of weeks	1	2	5	8	36	28

Conduct a chi-square goodness-of-fit test at the 5% level of significance to determine whether the number of days per week the subscriber gets her newspaper on time follows a binomial distribution. Use the P-value method and show all of your steps.

4. Each year, the Federal Trade Commission tests the tar and nicotine content of various brands of cigarettes in the United States. Data for a sample of eight brands is collected, and we would like to conduct a hypothesis test to determine whether the tar content of cigarettes can be predicted by their nicotine content. All measurements are in milligrams. The data are analyzed using an analysis of variance, and the ANOVA table (with some values missing) is shown below:

Source of Variation	*df*	Sum of Squares	Mean Square
Regression			48.1
Error			
Total		65.3	

(a) Find and enter all missing values in the table. Conduct an analysis of variance at the 5% level of signficance to determine whether there exists a linear relationship between tar content and nicotine content. Use the P-value method. Show all of your steps including the hypotheses, the calculation of the test statistic and the P-value, and a properly worded conclusion.

(b) What is the value of the correlation between tar content and nicotine content?

(c) What is the estimate of the parameter σ in the regression model?

Part A Answers

1. A
2. E
3. E
4. B
5. D

6. E
7. C
8. E
9. E
10. C

11. D
12. B
13. D
14. E
15. E

16. A
17. C
18. A
19. D
20. C

21. C
22. E
23. B
24. A
25. A

26. A
27. D
28. A
29. E
30. A

31. D
32. B
33. A

Part B Answers

1. (b) 0.9573

2. (b) $(-1.3, 11.3)$
   (d) $t = 2.04$, P-value between 0.05 and 0.10

3. $\hat{p} = 0.8$, $\chi^2 = 1.64$, $df = 1$, P-value $= 0.20$

4. (a) $F = 16.78$, $0.001 <$ P-value < 0.01
   (b) 0.8583
   (c) 1.693