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:

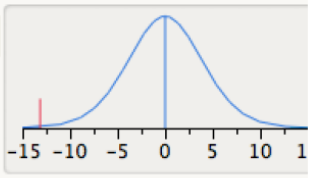
Means and Std Deviations

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

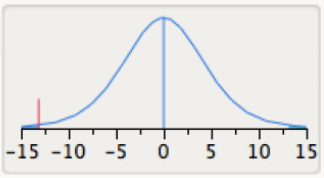
Difference	−13.144	t Ratio	
Std Err Dif		DF	16
Upper CL Dif		Prob > |t|	0.0040*
Lower CL Dif		Prob > t	0.9980
Confidence	0.95	Prob < t	0.0020*



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?

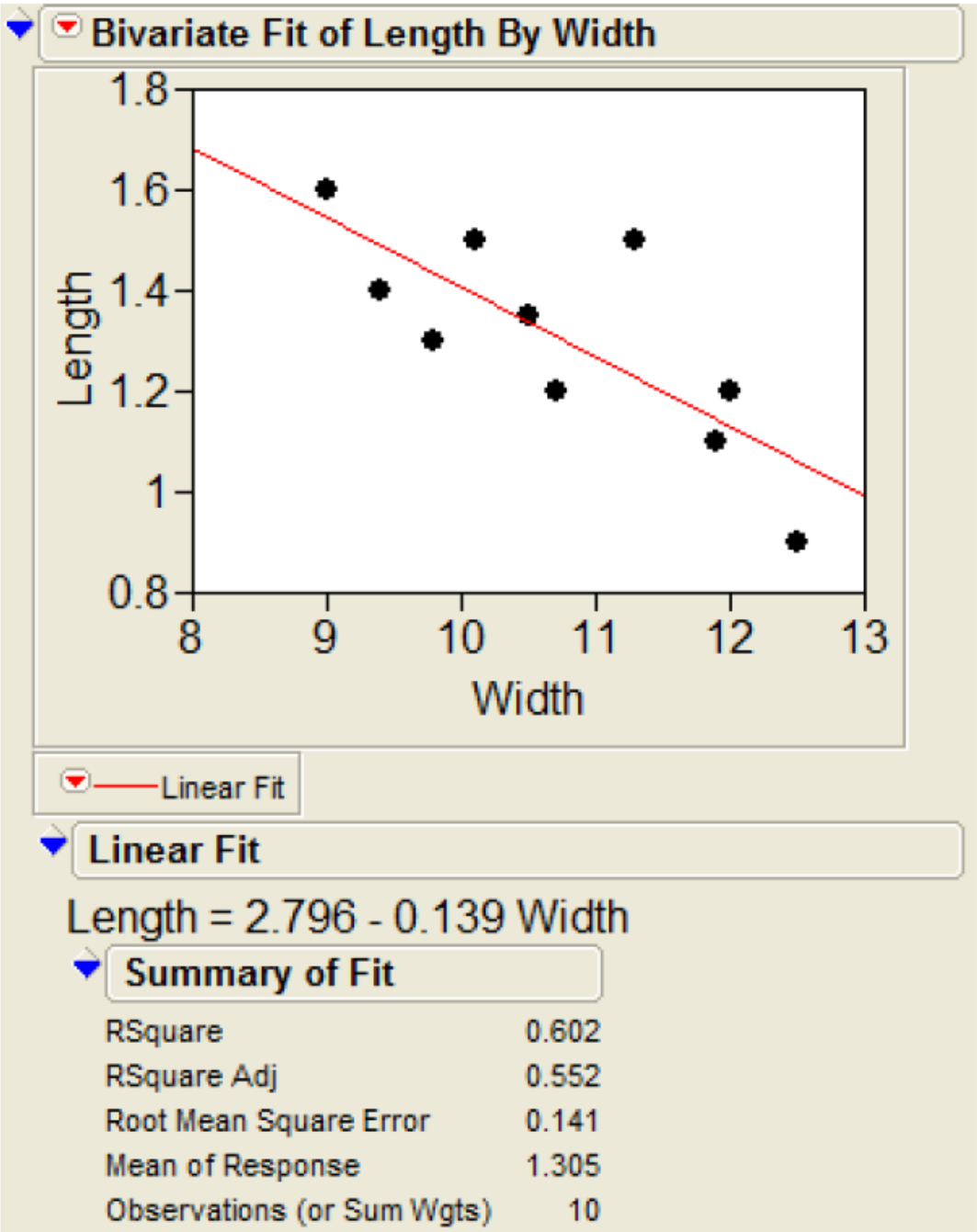
(A) 4.92 (B) 6.54 (C) 8.17 (D) 10.49 (E) 12.31

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:



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?