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 ± 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

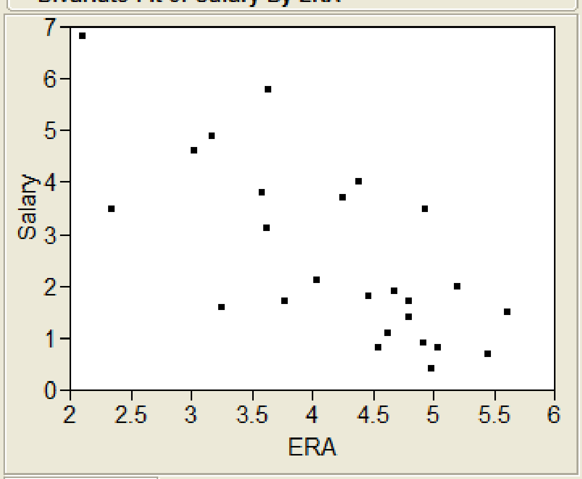
The next **three** questions (**28** to **30**) refer to the following:

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.

Part A Answer Key

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

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. 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.

The next **two** questions (**4** and **5**) refer to the following:

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.

4. We would like to conduct a hypothesis test to determine whether the true mean GPA of all students at University 1 differs from 3.00. What is the P-value of the appropriate test of significance?

(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

5. 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

6. We would like to conduct a hypothesis test to determine whether the mean final exam score μ_1 for section A01 in a large course is greater than the mean final exam score μ_2 for section A02. We will make a Type I Error if we conclude that:

   (A) $\mu_1 = \mu_2$ when in fact $\mu_1 > \mu_2$.

   (B) $\mu_2 > \mu_1$ when in fact $\mu_1 > \mu_2$.

   (C) $\mu_2 > \mu_1$ when in fact $\mu_1 = \mu_2$.

   (D) $\mu_1 > \mu_2$ when in fact $\mu_2 > \mu_1$.

   (E) $\mu_1 > \mu_2$ when in fact $\mu_1 = \mu_2$.

7. We would like to conduct a matched pairs t test to determine whether premium gasoline is more efficient than regular gasoline. We take a sample of ten different models of car. Each car drives for one tank on premium gasoline and one tank on regular gasoline (with the order randomly determined) and the mileage is recorded for each. Which of the following statements about the matched pairs t test are **true**?

   (I) For each car, mileage on premium gas and mileage on regular gas are independent.
   (II) For each car, mileage on premium gas and mileage on regular gas are dependent.
   (III) For any two cars, mileages on premium gas are independent.
   (IV) We must assume that mileage on premium gas and mileage on regular gas are both normally distributed.
   (V) We must assume that the differences in mileage (premium − regular) follow a normal distribution.

   (A) I and IV only

   (B) II and V only

   (C) I, III and IV only

   (D) I, III and V only

   (E) II, III and V only

8. We conduct an experiment to compare the effectiveness of four different headache medications. Each medication is randomly assigned to three patients, and the time (in minutes) until patients experience relief is recorded. Suppose it is known that relief times for the four medications are normally distributed. The test statistic is calculated to be 3.97. 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

The next **two** questions (**9** and **10**) 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

9. 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

10. 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

The next **two** questions (**11** and **12**) refer to the following:

11. 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

12. Now suppose we repeatedly select candies with replacement. Let X be the number of selections required to get the first cherry candy. What is $P(X = 5)$?

    (A) 0.079 (B) 0.086 (C) 0.093 (D) 0.104 (E) 0.117

13. Suppose we have two unfair coins. Coin 1 lands on Heads 37% of the time and Coin 2 lands on Heads 54% of the time. What is the probability that Coin 1 lands on Heads or Coin 2 lands on Tails?

    (A) 0.49 (B) 0.57 (C) 0.66 (D) 0.71 (E) 0.83

14. A hat contains three gold coins, two silver coins and one copper coin. We will select coins without replacement until the first silver coin is selected. The outcome of interest is the sequence of coins that are selected during this process. How many outcomes are in the appropriate sample space for this experiment?

    (A) 9 (B) 12 (C) 13 (D) 14 (E) 15

15. 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 the student's total travel time to and from school one day exceeds one hour?

(A) 0.0040 (B) 0.0808 (C) 0.1587 (D) 0.2296 (E) 0.3897

16. 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) $\frac{3}{4}$

17. A car company reports that the number of breakdowns per 8-hour 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 8-hour shifts?

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

18. 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) $\sqrt{8}$

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 94% confidence interval for the true proportion of all Winnipeggers who use transit is:

(A) (0.260, 0.380)
(B) (0.265, 0.375)
(C) (0.270, 0.370)
(D) (0.275, 0.365)
(E) (0.280, 0.360)

22. 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

23. 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 **three** questions (**24** to **26**) 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 21.13 0.81	62 57.87 0.29	79
One Parent Smokes	11 13.64 0.51	40 37.36 0.19	51
Both Parents Smoke	14 ??? 6.37	13 ??? ???	27
Column Total	42	115	157

24. 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

25. What is the expected count for the number of non-smoking university students who have two parents who smoke?

(A) 12.37 (B) 14.22 (C) 16.45 (D) 19.78 (E) 21.09

26. 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

The next **two** questions (**27** and **28**) refer to the following:

We would like to determine whether the number of errors on income tax forms processed by an accounting firm has a Poisson distribution. An employee selects a random sample of 100 tax returns and determines the number of errors on each. The data are shown in the table below, as well as some expected cell counts:

# of errors	0	1	2	3	4	5
# of forms	36	28	23	8	3	2
Expected	30.12	???	???	8.67	2.60	0.62

27. Under the null hypothesis, what is the expected number of forms with two errors?

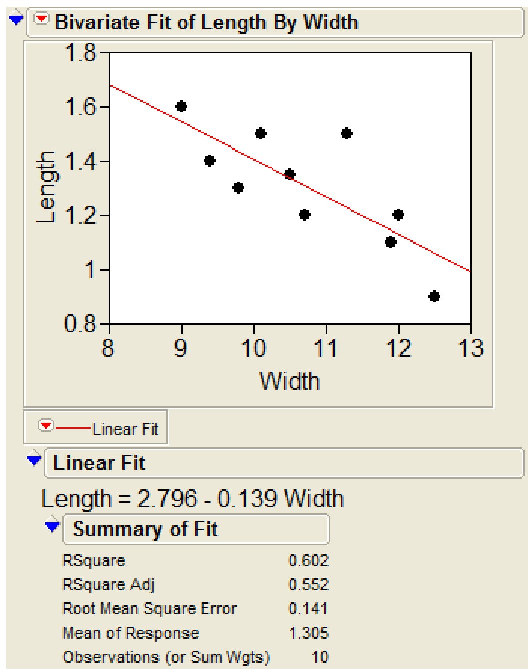
(A) 11.53 (B) 17.46 (C) 21.69 (D) 24.22 (E) 26.01

28. What are the degrees of freedom for the appropriate test statistic?

(A) 2 (B) 3 (C) 4 (D) 5 (E) 6

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

Sample Final Exam 2 – Part B

1. (a) The owner of a small store suspects that one of his employees is stealing, as the amount of money in the cash register at the end of the employee's shift is often less than expected. The employee agrees to take a polygraph (lie detector) test. During the test, the employee claims that he has not stolen any money. The polygraph is essentially testing the hypotheses

   H_0: The employee is telling the truth. vs. H_a: The employee is lying.

   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. Three contestants compete on each episode of the TV game show *Jeopardy!* The number of female contestants for a random sample of 200 shows are shown in the table below:

# of females	0	1	2	3
# of shows	50	94	52	4

   Conduct a chi-square goodness-of-fit test at the 5% level of significance to determine whether the number of female contestants per episode follows a binomial distribution. Use the P-value method and show all of your steps.

3. 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:

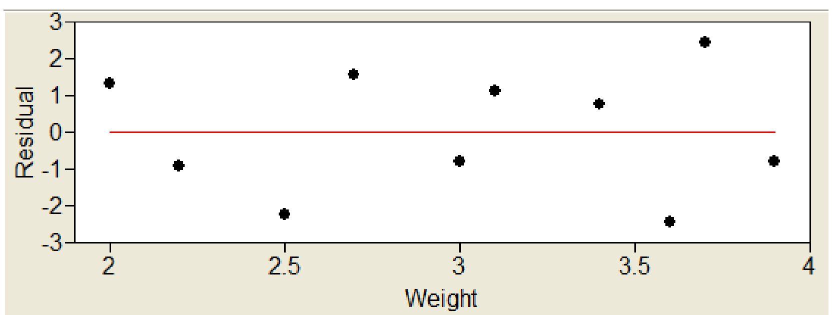
Weight	3.1	3.7	2.2	3.4	2.0	3.9	3.0	2.5	3.6	2.7
Fuel Efficiency	25	21	31	22	35	16	24	27	17	29

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.

Part A Answer Key

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

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

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

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

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

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

Part B Answers

1. (b) Power = 0.9573
2. $\hat{p} = 0.8, \chi^2 = 1.63$, df = 1, $0.20 <$ P-value < 0.25
3. (c) $(-10.977, -6.803)$
   (e) $t = -9.82$, P-value < 0.001
   (g) (18.735, 27.309)