STAT235 FINAL EXAM VERSION 111 SOLUTION

Complete the following (please **print**):

Section Number: ___________ Last name, First name: ______________________________

Instructions

1. Enter your lecture section number, your last name (the same one that is illustrated on your One Card), and first name (the same one that is illustrated on your One Card) into the space provided above. Enter your student ID # in the upper right hand corner on all the *other* pages of the exam.

2. Make sure you use ONLY PENCIL to put and mark your name, student ID #, and your exam version number on the Scantron sheet. Your name (last name and first name) should be entered into the NAME block, your student ID # into the IDENTIFICATION NUMBER block, and finally, the three-digit exam version into the SPECIAL CODES block. The exam version is specified at the top of the exam. Make sure to shade in the circle that corresponds to the letter, digit, or empty space in the box at the top of each column.

3. This is a multiple-choice closed book exam. There are **50** questions in the exam. For each question, carry out the appropriate analysis and put your answer on the Scantron sheet by shading the letter A, B, C, D, or E that corresponds to your chosen answer. Make sure your answers are clearly marked with ONLY PENCIL. Otherwise, no marks will be given. For each question exactly one of the five answers is correct. If you fill in more than one answer to a question, the question will be scored incorrect. Each question is worth 1 mark.

4. Please note that only the answers in the Scantron sheet will be considered. If you initially mark your answers in the exam sheet, make sure that you copy them correctly into the Scantron sheet. If you change an answer on the Scantron sheet, be sure that you erase your first mark completely and then blacken the circle of the answer choice you prefer. Your score will be based on the number of questions you answer correctly. No credit will be given for omitted answers and no credit will be lost for wrong answers. You must mark all your answers on the Scantron sheet during the allotted time. No additional time will be allowed at the end of the session for this purpose.

5. You are permitted to use a **non-programmable** calculator approved by the Faculty of Engineering. During the lab exam you are forbidden to use any devices with communication capabilities including cell phones and pagers. You are also forbidden to use any photographically capable devices in the exam room. Copying questions or answers on paper to take from the exam room is prohibited.

6. Note that the formula sheet and the table of the cumulative distribution function of the standard normal distribution are attached to the exam. There are **10** pages in the exam. The exam is graded out of **50** points.

7. You must return your Scantron and exam booklet when you finish the exam. You have 180 minutes to complete the exam.

8. Sign the exam booklet in the space provided below.

SIGNATURE: ________________________________

PART 1 (MULTIPLE-CHOICE QUESTIONS, 1 POINT EACH)

1. Let the probability of a microwave in the engineering faculty surviving more than five years (M) be 0.15. Also, the probability of the microwave being used more than 40 times a day (F) is 0.75. If the probability of the microwave surviving more than five years, given it is being used more than 40 times a day, is 0.05, one can conclude that

   (a) M and F are independent events.
   (b) P(M or F) is greater than 0.5.
   (c) P(M and F) is greater than 0.5.
   (d) $P(F \mid M)$ is greater than 0.5.
   (e) M and F are mutually exclusive events.

   Solution: Note that $P(M) = 0.15$, $P(F) = 0.75$, and $P(M \mid F) = 0.05$.
   Since $P(M) \neq P(M \mid F)$, M and F are not independent. → (a) is wrong.
   P(M and F) = $P(F) \times P(M \mid F) = 0.75 \times 0.05 = 0.0375 < 0.5$ → (c) is wrong.
   P(M or F) = $P(M) + P(F) - P$(M and F) = $0.15 + 0.75 - 0.0375 = 0.8625 > 0.5$ → (b) is right.
   $P(F \mid M) = P$(M and F) / $P(M) = 0.0375 / 0.15 = 0.25 < 0.5$ → (d) is wrong.
   P(M and F) > 0, M and F are not mutually exclusive. → (e) is wrong

2. Back in the mid-twentieth century, grain elevators were common in Saskatchewan. In one region, each grain elevator achieved their monthly quota with a probability of 0.87. If each grain elevator worked independently of the others, what is the probability of checking 15 grain elevators before finding 13 that have met their monthly quota?

   (a) 0.0028 (b) 0.1635 (c) 0.2516 (d) 0.1904 (e) 0.2903

   Solution: Let X be the number of grain elevators to be checked. Then X follows a negative binomial distribution with the parameters $r = 13$ and $p = 0.87$. Thus,

$$P(X=15)=\binom{15-1}{13-1}(1-0.87)^{15-13}(0.87)^{13}=0.2516$$

3. Suppose the number of times a particular brand of calculator experiences a malfunction follows a Poisson distribution with an average of 0.235 times an hour. What is the probability of the calculator experiencing two malfunctions in 90 minutes?

   (a) 0.0218 (b) 0.0437 (c) 0.0873 (d) 0.1239 (e) 0.2478

   Solution: Let X be the number of times the calculator experiences malfunctions in an hour. As the number of times that the student checks the clock follows a Poisson distribution with the mean of 0.235 per hour, the variable Y follows a Poisson distribution with a mean of 0.3525 per 1.5 hours (90 minutes). Thus,

$$P(Y=2)=\frac{e^{-0.3525}0.3525^{2}}{2!}=0.0437$$

4. Refer to the previous question. If X is the time it takes until a calculator experiences a malfunction, what is the probability that it takes more than 3 hours to experience a malfunction?

   (a) 0.4941 (b) 0.1206 (c) 0.7029 (d) 0.3619 (e) 0.5059

   Solution: Based off the definition of X, it follows an exponential distribution with λ = 0.235/hour.

$$P(X>3)=1-F(3)=1-(1-e^{-(0.235)(3)})=e^{-0.705}=0.4941$$

5. A quality control technician at Cyberdyne Systems Corporation is checking each machine in a row until he finds one that does not operate, but he will check no more than three machines, even if they all operate. Assume each machine operates independently and has a probability of operating of 0.9. What is the expected number of operating machines?

   (a) 1.791 (b) 2.439 (c) 2.710 (d) 1.110 (e) 0.111

Solution: There are 4 possible scenarios for operating machines: 0, 1, 2, or 3.

x	0	1	2	3
$P(X = x)$	0.1	0.09	0.081	0.729

Thus, for X = # operating machines,

$E(X) = \sum x_i P(X = x_i) = 0(0.1) + 1(0.09) + 2(0.081) + 3(0.729) = 2.439$

6. Suppose a continuous random variable X has the following probability density function:

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & elsewhere \end{cases}$$

Which of the following probabilities approximately equals 0.1353?

(a) $P(1 \leq X \leq 4)$ (b) $P(2 < X \leq 3)$ (c) $P(2 \leq X < 4)$ (d) $P(X \leq 0.5)$ (e) $P(X > 2)$

Solution: Determine $F(x) = 1 - e^{-x}$. Use this for every interval to speed up calculations.

7. Three independent observations are selected randomly from a standard normal distribution. The probability that their total is smaller than 2 is

(a) 0.8413 (b) 0.7486 (c) 0.9582 (d) 0.8749 (e) 0.9207

Solution: If $Y = Z_1 + Z_2 + Z_3$, where $Z_i \sim N(0, 1)$, $E(Y) = 0 + 0 + 0 = 0$; $Var(Y) = 1 + 1 + 1 = 3$.

$$P(Y < 2) = P\left(\frac{Y-\mu}{\sigma} < \frac{2-0}{\sqrt{3}}\right) = P(Z < 1.15) = 0.8749$$

8. Regarding the general properties of the sampling distribution of the sample mean, it is TRUE that

(a) the standard deviation of the sample mean increases, as the sample size increases.

(b) no matter the shape of the population distribution, the sampling distribution of the sample mean is approximately normal if the sample size is greater than 5.

(c) the spread of the sampling distribution of the sample mean is larger than the spread of the population distribution.

(d) the sampling distribution of the sample mean is normal for any sample size, if the population distribution is normal.

(e) the standard deviation of the population decreases, as the sample size increases.

Solution: Review Chapter 7 notes.

9. Chris Broderick runs a company that manufactures digital video cameras: Mega-Def. Suppose the true proportion of defective cameras is 0.06. If a shipment consists of 250 "random" cameras, what is the probability of the sample proportion being no more than 0.055?

(a) 0.3707 (b) 0.6368 (c) 0.3300 (d) 0.3632 (e) 0.6293

Solution: Before the Fall 2021 term:

$$P(\hat{p} \leq 0.055) \rightarrow P\left(\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{0.055-0.06}{\sqrt{\frac{0.06(1-0.06)}{250}}}\right) = P(Z \leq -0.33) = 0.3707$$

For the Fall 2021 term and after:

$$P(\hat{p} \leq 0.055) \approx P\left(\frac{\hat{p}+\frac{0.5}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{0.055+0.5/250-0.06}{\sqrt{\frac{0.06(1-0.06)}{250}}}\right) = P\left(\frac{\hat{p}+\frac{0.5}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{-0.0030}{0.01502}\right)$$

$= P(Z \leq -0.20) = 0.4207$

10. Refer to the previous question. If shipments are packaged independently of each other, what is the probability that exactly 8 of 10 shipments have less than 20 defective cameras each?

(a) 0.3820 (b) 0.8246 (c) 0.9082 (d) 0.0918 (e) 0.2241

Solution: Use normal approximation to binomial for $P(X < 20)$.

$$P(X<20)=P(X\le 19)\approx\Phi\left(\frac{19+0.5-250(0.06)}{\sqrt{250(0.06)(1-0.06)}}\right)=\Phi(1.20)=0.8849$$

Let Y be the number of shipments with less than 20 defective cameras each. Then, Y follows a binomial distribution with n = 10 and p = 0.8849. Thus,

$$P(Y=8)=\binom{10}{8}(0.8849)^8(1-0.8849)^{10-8}=45(0.8849)^8(0.1151)^2=0.2241$$

11. Interpolation methods are used to estimate heights above sea level for locations where direct measurements are unavailable. Errors caused by interpolation methods for estimating heights are normally distributed with a standard deviation of 4.8 cm. A sample of 16 measurement errors gave a mean of 3.8 cm. The margin of error for the 98% confidence interval is

(a) 0.12 (b) 2.79 (c) 1.61 (d) 3.12 (e) 3.10

Solution: Since σ is given, use z. At the 98% confidence level, $z_{\alpha/2} = z_{0.02/2} = 2.326$.

$$E=z_{\alpha/2}\times\left(\frac{\sigma}{\sqrt{n}}\right)=2.326\times\left(\frac{4.8}{\sqrt{16}}\right)=2.79$$

12. A 95% confidence interval for a population mean is computed from a sample of size 50. Another 95% confidence interval will be computed from a sample of size 200, drawn from the same population. The interval from the sample of size 50 will be approximately ___________________ as the interval from the sample of size 200.

(a) one eighth as wide (b) one fourth as wide (c) half as wide
(d) twice as wide (e) four times as wide

Solution: Based off the formula for the margin of error, use $n/4$ for 50 to relate to 200.

$$E_{50}=z_{\alpha/2}\times\left(\frac{\sigma}{\sqrt{n/4}}\right)=z_{\alpha/2}\times\left(\frac{\sigma}{(1/2)\sqrt{n}}\right)=2\left[z_{\alpha/2}\times\left(\frac{\sigma}{\sqrt{n}}\right)\right]=2E_{200}.$$

13. Stainless steels can be susceptible to stress corrosion cracking under certain conditions. A materials engineer is interested in determining the proportion of stainless steel alloy failures that that are due to stress corrosion cracking. In the absence of preliminary data, **how large of a sample** must be taken so that the 95% confidence interval will specify the proportion to be within a margin of error of 0.05?

(a) 416 (b) 385 (c) 514 (d) 476 (e) 325

Solution: Proportions require using z. At the 95% confidence level, $z_{\alpha/2} = z_{0.05/2} = 1.960$.

$$E=z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\rightarrow n=\hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2=(0.5)(1-0.5)\left(\frac{1.960}{0.05}\right)^2=384.16$$

14. A supplier sells synthetic fibres to a manufacturing company. A simple random sample of 25 fibres is selected from a shipment and their breaking strength is noted. The average breaking strength is found to be 29 lbs. with a standard deviation of 3.79 lbs. The one-sided lower confidence bound on the mean is 28 lbs. With what approximate level of confidence is this bound constructed?

(a) 80% (b) 85% (c) 90% (d) 95% (e) 98%

Solution: Since s is given, use t.

$$\text{Lower bound} = \bar{x}-t_{\alpha,n-1}\times\frac{s}{\sqrt{n}} \rightarrow 28=29-t_{\alpha,n-1}\times\frac{3.79}{\sqrt{25}}\rightarrow t_{\alpha,24}=1.319$$

Examining the t-table, the approximate confidence level is $1 - \alpha = 1 - 0.10 = 0.90$, or 90%.

15. A scientist computes a 90% confidence interval to be (4.38, 6.02). Using the same data, she also computes a 95% confidence interval to be (4.22, 6.18) and a 99% confidence interval to be (3.91, 6.49). If she wants to test H_0: $\mu = 4$ versus H_A: $\mu \neq 4$, which one of the following statements is true regarding the *P*-value?

    (a) *P*-value > 0.10
    (b) 0.05 < *P*-value < 0.10
    (c) 0.01 < *P*-value < 0.05
    (d) *P*-value < 0.01.
    (e) None of the above statements are true.

    Solution: The hypothesized value is inside the 99% confidence level, but it is not inside the 95% and 90% confidence intervals. Thus, the *P*-value is somewhere between 1 – 0.99 = 0.01 and 1 – 0.95 = 0.05.

16. An automobile manufacturer is considering using robots for part of its assembly process. Converting to robots is an expensive process, so it will be undertaken only if there is strong evidence that the proportion of defective installations is lower for the robots than the known defective proportion of human assemblers of 0.02. The test results in a *P*-value of 0.102. In reality, the robots are producing 99% good parts. What happens as a result of our testing? ($\alpha = 0.05$)

    (a) We say that the proportion of defective installation by robots may be more than or equal to 0.02, making a Type I error.

    (b) We say that the proportion of defective installation by robots may be more than or equal to 0.02, making a Type II error.

    (c) Fail to reject H_0 saying the proportion of defective installation by robots is less than 0.02, making a Type I error.

    (d) Fail to reject H_0 saying that the proportion of defective installations by robots is less than 0.02, making a Type II error.

    (e) Reject H_0, saying the proportion of defective installations by robots is greater than 0.02.

    Solution: In reality, the null hypothesis should have been rejected since the proportion of defective installations with robots was 1 – 0.99 = 0.01, which is noticeably less than the hypothesized proportion ($p_0 = 0.02$). However, by random chance the data collected led to a *P*-value (0.102) that is greater than $\alpha = 0.05$, so the null hypothesis would not be rejected. Failing to reject a null hypothesis, when in reality it should have been rejected because it is false, is a Type II error.

17. One hundred random observations are taken from a normal distribution with mean μ and variance $\sigma^2 = 4$. To test H_0: $\mu = 3$ versus H_A: $\mu > 3$, a rejection region of the form $\bar{X} > c$ is to be used. What is the value of *c* such that the probability of a Type I error is 0.10?

    (a) 3.17 (b) 3.26 (c) 3.33 (d) 3.51 (e) 5.56

    Solution: Since σ is given, use *z*. Examining the critical points in the *z*-table, $z_{0.10} = 1.282$.

    $$z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \Rightarrow 1.282 = \frac{c - 3}{2 / \sqrt{100}} \Rightarrow c = 3.26$$

18. A machine manufactures bolts that are supposed to be 3 inches in length. Each day, a quality engineer selects a random sample of 36 bolts from the day's production, measures their lengths, and performs a hypothesis test of H_0: $\mu = 3$ versus H_A: $\mu \neq 3$, where μ is the mean length of all the bolts manufactured that day. Assume that the population standard deviation for bolt lengths is 0.1 inches. H_0 is rejected, forcing the machine to be shut down and recalibrated, if the sample mean falls outside the range of $2.98 < \bar{X} < 3.02$. Assume that on a given day, the true mean length of bolts is 3 inches. What is the (approximate) probability that the machine will be shut down?

    (a) 0.81 (b) 0.19 (c) 0.23 (d) 0.77 (e) 0.89

Solution: Since σ is given, use z. Also, probability is mentioned, so use Chapter 7.

$$P(2.98 \le \overline{X} \le 3.02) \rightarrow P\left(\frac{2.98-3}{0.1/\sqrt{36}} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \frac{3.02-3}{0.1/\sqrt{36}}\right) = P(-1.20 \le Z \le 1.20)$$

$= 1 - 2P(Z \le -1.20) = 1 - 2(0.1151) = 0.7698$

P(machine will be shut down) = $1 - P(2.98 \le \overline{X} \le 3.02) = 1 - 0.7698 = 0.2302 \approx 0.23$

19. Paint used to paint lines on roads must reflect enough light to be clearly visible at night. Let μ denote the true average reflectometer reading for a new type of paint under consideration. A test of H_0: $\mu = 20$ versus H_A: $\mu > 20$, from a sample of 15 observations, gave $t_0 = 23.2$. Using a confidence interval to test the hypothesis,

(a) Construct a two-sided confidence interval and see whether 23.2 is in the interval. If so, reject the null hypothesis.

(b) Construct a two-sided confidence interval and see whether 20 is in the interval. If so, fail to reject the null hypothesis.

(c) Construct a lower one-sided confidence bound and see whether 20 is equal to or exceeds the bound. If so, fail to reject the null hypothesis.

(d) Construct an upper one-sided confidence bound and see whether 20 is equal to or exceeds the bound. If so, fail to reject the null hypothesis.

(e) Construct an upper one-sided confidence bound and see whether 23.2 is equal to or exceeds the bound. If so fail to reject the null hypothesis.

Solution: A right-tailed hypothesis test can be tested with a lower, one-sided confidence bound. If the hypothesized value (μ_0), which in this case is 20, is less than the lower bound, this would mean that the null hypothesis would be rejected; otherwise, we would fail to reject the null hypothesis

20. Suppose a manufacturer of steel rods is experimenting with a new technology that is supposed to increase the mean strength of the rods. One group of 31 random rods was obtained with the old technology, while another group of 31 was obtained with the new technology. The breaking strength of each rod in the two groups was recorded (in psi). We test whether the new technology rods are stronger than the old technology rods on the average. Assume that the population standard deviations of strength for the old and new technology rods are the same. The Excel output for the test with some missing entries is given below:

t-Test: Two-Sample Assuming Equal Variances

	NEW TECH	*OLD TECH*
Mean	55	54
Variance	2.89	3.24
Observations	31	31
Pooled Variance		
Hypothesized Mean Difference	0	
df		
t Stat		
P(T<=t) one-tail		

Thus, based on the above output, the absolute value of the test statistic is approximately

(a) 2.048 (b) 2.125 (c) 2.189 (d) 2.249 (e) 2.372

Solution: $s_p = \sqrt{\dfrac{(31-1)2.89+(31-1)3.24}{31+31-2}} = 1.7507 \Rightarrow t_0 = \dfrac{(55-54)-0}{1.7507\sqrt{(1/31)+(1/31)}} = 2.249$

21. Refer to the previous question. What is the distribution of the test statistic?

(a) t distribution with 30 degree of freedom,
(b) t distribution with 31 degrees of freedom,
(c) t distribution with 32 degrees of freedom,
(d) t distribution with 60 degrees of freedom,
(e) t distribution with 62 degrees of freedom.

Solution: The test statistic follows a t-distribution with df = $n_1 + n_2 - 2 = 31 + 31 - 2 = 60$.

22. Refer to the previous two questions. What is the P-value range of the test statistic?

(a) 0.001 and 0.0005 (b) 0.001 and 0.0025 (c) 0.0025 and 0.005
(d) 0.005 and 0.01 (e) 0.01 and 0.025

Solution: Based on the t-table and the df above, the P-value is between 0.01 and 0.025.

23. Refer to the previous three questions. The 95% confidence interval for $\mu_{\text{NEW TECH}} - \mu_{\text{OLD TECH}}$ is

(a) 54.5 ± 0.66812 (b) 1 ± 0.8894 (c) 1 ± 0.9043 (d) 1 ± 0.9228 (e) 1 ± 0.9545

Here $\mu_{\text{NEW TECH}}$, $\mu_{\text{OLD TECH}}$ denote the mean strength of the new and old technology rods, respectively.

Solution: At the 95% confidence level, $t_{\alpha/2,\, df} = t_{0.05/2,\, 60} = 2.000$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,df} \times s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \Rightarrow (55-54) \pm 2.000 \times 1.7507\sqrt{\frac{1}{31}+\frac{1}{31}} \Rightarrow 1 \pm 0.8894$$

Use the output below to answer Questions 24 – 29:

An article studied the ozone levels on the South Coast air basin of California for the 16-year period between 1976 and 1991 (1 observation per year). The author believes that the number of days (*DAYS*) that the ozone level exceeds 0.20 parts per million depends on the seasonal meteorological index (*INDEX*). The following linear regression model was considered:

DAYS = $\beta_0 + \beta_1 \cdot$INDEX $+\varepsilon$,

Incomplete Excel regression ANOVA output for the data is given below:

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3827.43	3827.43		
Residual		5592.01	399.43		
Total		9419.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-217.94	93.90				
Index	16.79	5.42				

24. In 1983, there were 82 days where the ozone level exceeded 0.20 parts per millions. The meteorological index that year was 17.2. The residual for this observation is:

(a) -11.152 (b) 11.901 (c) 11.152 (d) –11.901 (e) 13.432

Solution: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -217.94 + 16.79(17.2) = 70.848$

$e = y_i - \hat{y}_i = 82 - 70.848 = 11.152$

25. The estimate of the standard deviation of the errors (σ) is:

(a) 5.421 (b) 13.906 (c) 19.986 (d) 61.877 (e) 74.782

Solution: $\hat{\sigma} = \sqrt{MSE} = \sqrt{399.43} = 19.986$

26. The proportion of the variation in the number of days that can be explained by the meteorological index is

(a) 0.4063 (b) 0.4936 (c) 0.5937 (d) 0.6841 (e) 0.9357

Solution: The phrasing refers to R^2; thus, $R^2 = \frac{SSR}{SST} = \frac{3827.43}{9419.44} = 0.4063$.

27. The estimated mean change in number of days that the ozone level exceeds 0.20 parts per million as the seasonal meteorological index raises by 3 is:

(a) 16.79 (b) –210.15 (c) –167.57 (d) 50.37 (e) –653.82

Solution: Increasing the x-value by 3 will increase y by $3\beta_1$; thus, $3\hat{\beta}_1 = 3(16.79) = 50.37$

28. Suppose the researchers wish to test the claim that for each unit increase in index, the mean increase in number of days the ozone level exceeds 0.20 parts per million is more than 10. The value of the appropriate test statistic in this case is

(a) 1.253 (b) 1.675 (c) 1.890 (d) 2.345 (e) 3.098

Solution: For a one-tailed test in regression, the *t*-test must be used.
H_0: $\beta_1 = 10$ H_A: $\beta_1 > 10$

$$t_0 = \frac{\hat{\beta}_1 - b_1}{s.e.(\hat{\beta}_1)} = \frac{16.79 - 10}{5.42} = 1.253$$

29. A 95% confidence interval for the slope of the regression line is

(a) 16.79 ± 5.42 (b) 16.79 ± 8.79 (c) 16.79 ± 9.26 (d) 16.79 ± 10.80 (e) 16.79 ± 11.63

Solution: n = 1991 – 1975 = 16, so for a 95% confidence level, $t_{\alpha/2,\, n-2} = t_{0.05/2,\, 16-2} = 2.145$

$$\hat{\beta}_1 \pm t_{\alpha/2,v} \times s.e.(\hat{\beta}_1) \Rightarrow 16.79 \pm 2.145 \times 5.42 \Rightarrow 16.79 \pm 11.63$$

30. In order to compare the tensile strength of three alloys labelled A, B and C, 23 specimens were randomly selected: five alloy A specimens, six alloy B specimens, and twelve alloy C specimens. The tensile strength was measured in each specimen. The following is incomplete output of one-way ANOVA for the tensile strength obtained with Excel.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
A	5	880	176	484
B	6	1020	170	400
C	12	2160	180	441

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	100					
Within Groups	600					
Total	700					

Based on the output, what is the value of the *F*-statistic?

(a) 0.957 (b) 1.667 (c) 2.141 (d) 2.974 (e) 3.547

Solution: $F = \frac{SSTr/(k-1)}{SSE/(N-k)} = \frac{100/(3-1)}{600/(23-3)} = \frac{50}{30} = 1.667$

Use the following information for questions 31 – 37:

A mining company is recommending a new alternative for natural resources: unobtainium. Unfortunately, it can only be used properly with a minimum 90% purity level when mined. A random sample of four pieces of rock containing unobtainium was found in one region (*Tskxe*), with the purity level being measured for each rock. To improve the purity, each rock was then subjected to a purification technique. The data provided in the first two columns of the following table represent the rocks from *Tskxe* before and after purification. A second random sample of rocks from the *Utral* region is provided in the third column. Previous studies determined that purity level follows a normal distribution.

TSKXE (T)	PURIFICATION (P)	UTRAL (U)
89.9	93.2	98.8
90.2	95.1	97.2
89.7	92.8	95.2
83.7	91.9	94.7

Summaries for each of the three columns are as follows: (Note: std. dev. = standard deviation)

	TSKXE	**PURIFICATION**	**UTRAL**
MEAN	88.4	93.3	96.5
STD. DEV.	3.1	1.3	1.9

Summaries for the differences between each pair of columns are as follows: (Note still applies.)

	T – P	**T – U**	**P – U**
MEAN	-4.9	-8.1	-3.2
STD. DEV.	2.4	2.4	1.6

31. Do the data provide evidence that the purification technique improves the purity level of unobtainium in rocks? Select the most appropriate test.

(a) A two independent sample *t*-test (general procedure).
(b) A paired *t*-test.
(c) A two independent sample *t*-test (pooled procedure).
(d) A two independent sample test (*z*-procedure).
(e) ANOVA test for several means.

Solution: Since it is the same rocks before and after the purification technique, the data are paired.

32. Refer to the previous question. What are the appropriate hypotheses to answer this test?

(a) H_0: $\mu_d = 0$ H_A: $\mu_d < 0$ where $d = x_T - x_P$
(b) H_0: $\mu_d = 0$ H_A: $\mu_d > 0$ where $d = x_T - x_P$
(c) H_0: $\mu_d = 0$ H_A: $\mu_d \neq 0$ where $d = x_T - x_P$
(d) H_0: $\mu_d = 0$ H_A: $\mu_d < 0$ where $d = x_P - x_T$
(e) H_0: $\mu_d = 0$ H_A: $\mu_d \neq 0$ where $d = x_P - x_T$

Solution: If the purification technique intends to improve the purity level, then the Tskxe (T) sample should be lower than the Purification (P) sample. This also means it will be a one-tailed test, so carefully read the defined differences on the right to get the correct one-tailed alternative.

33. Refer to the previous two questions. If the test statistic is -4.083, what are the respective degrees of freedom and *p*-value range for the test?

(a) $df = 3$ and (0.10, 0.50)
(b) $df = 6$ and (0.002, 0.01)
(c) $df = 3$ and (0.02, 0.05)
(d) $df = 6$ and (0.001, 0.005)
(e) $df = 3$ and (0.01, 0.025)

Solution: $df = n - 1 = 4 - 1 = 3$
Since $-4.541 < t_0 = -4.083 < -3.182$, the *p*-value range is (0.01, 0.025).

34. Refer to the previous three questions. With a 5% significance level, what is the appropriate conclusion?

(a) At $\alpha = 0.05$, the purification technique does not improve the purity level of unobtainium in rocks.
(b) At $\alpha = 0.05$, the purification technique may improve the purity level of unobtainium in rocks.
(c) At $\alpha = 0.05$, the purification technique improves the purity level of unobtainium in rocks.
(d) At $\alpha = 0.05$, the purification technique may not improve the purity level of unobtainium in rocks.
(e) At $\alpha = 0.05$, the purification technique decreases the purity level of unobtainium in rocks.

Solution: Since the *p*-value < $\alpha = 0.05$, reject H_0. In conclusion, the purification technique improves the purity level of unobtainium in rocks.

35. Is there any evidence that rocks mined from the *Tskxe* region have a different purity level before purification than rocks mined from the *Utral* region? Select the most appropriate data structure.

(a) A two independent sample *t*-test (general procedure).
(b) A paired *t*-test.
(c) A two independent sample *t*-test (pooled procedure).
(d) A two independent sample test (*z*-procedure).
(e) ANOVA test for several means.

Solution: There are two samples that are not related, so they are independent. Though this situation involves independent samples where the standard deviation ratio is 3.1/1.9 = 1.63 < 2 with equal sample sizes, the sample sizes are too small, so the general procedure applies.

36. Refer to the previous question. What is the standard error of the appropriate estimate?

(a) 1.200 (b) 3.305 (c) 3.532 (d) 1.118 (e) 1.818

Solution: $S.E.(\bar{x}_T - \bar{x}_U) = \sqrt{\frac{s_T^2}{n_T} + \frac{s_U^2}{n_U}} = \sqrt{\frac{3.1^2}{4} + \frac{1.9^2}{4}} = 1.818$

37. Refer to the previous question. Construct an appropriate 90% confidence interval.

(a) (-13.885, -2.315)
(b) (-11.632, -4.568)
(c) (-11.091, -5.109)
(d) (-12.378, -3.822)
(e) (-11.078, -5.122)

Solution: $df = \min\{n_T - 1, n_U - 1\} = \min\{3, 3\} = 3$, so for 90% confidence, $t_{0.05,\,3} = 2.353$
$\bar{x}_T - \bar{x}_U \pm t_{\alpha/2,v} \times S.E.(\bar{x}_T - \bar{x}_U) \rightarrow (88.4 - 96.5) \pm (2.353)(1.818) \rightarrow (-12.378, -3.822)$

38. Refer to the previous question. What is an appropriate conclusion?

(a) Since zero is included in the interval, there is enough evidence of a difference in average purity level between the two samples.
(b) Since zero is included in the interval, there is not enough evidence of a difference in average purity level between the two samples.
(c) Since zero is not included in the interval, there is not enough evidence of a difference in average purity level between the two samples.
(d) Since zero is not included in the interval, there is enough evidence of a difference in average purity level between the two samples.
(e) Since zero is included in the interval, there is enough evidence that rocks from the Tskxe region have a higher average purity level.

Solution: There is no zero in the interval, so there is enough evidence to reject it as a value. There is enough evidence of a difference in average purity level between the two samples.

Use the following information for questions 39 – 46:

In a study of 3,400 oil and gas structures gathered from the Gulf of Mexico, there were 2,175 active and 1,227 idle (inactive) structures at the end of 2003. The table given below breaks down these oil and gas structures by type (caisson, well protector, or fixed platform) and status (active or inactive). Assume that the 3,400 structures are a representative sample of all oil and gas structures worldwide.

	Caisson	Well Protector	Fixed Platform	Total
Active	503	225	1447	2175
Inactive	598	177	450	1225
Total	1101	402	1897	3400

39. Conduct a test to determine if the proportion of inactive caisson structures exceeds the proportion of inactive well protector structures. Let p_C represent the proportion of inactive caisson structures and p_W represent the proportion of inactive well protector structures. What are the appropriate hypotheses to answer this test?

(a) H_0: $p_C - p_W = 0$ H_A: $p_C - p_W > 0$
(b) H_0: $p_C - p_W = 0$ H_A: $p_C - p_W < 0$
(c) H_0: $p_C - p_W = 0$ H_A: $p_C - p_W \neq 0$
(d) H_0: $p_C - p_W < 0$ H_A: $p_C - p_W = 0$
(e) H_0: $p_C - p_W > 0$ H_A: $p_C - p_W = 0$

Solution: The question ("exceeds") suggests a one-tailed test where mean breaking strength of new technology rods should be higher. Based off the defined symbols, p_C should be higher, so the defined difference should be positive in the alternative.

40. Refer to the previous question. What is the value of the appropriate test statistic?

(a) 0.103
(b) 0.0291
(c) 0.516
(d) 3.531
(e) 0.484

Solution: $\hat{p} = \dfrac{x_C + x_W}{n_C + n_W} = \dfrac{598+177}{1101+402} \approx 0.516$

$$z_0 = \frac{(\hat{p}_C - \hat{p}_W) - (p_C - p_W)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_C} + \frac{1}{n_W}\right)}} = \frac{0.543 - 0.440 - 0}{\sqrt{0.516(1-0.516)\left(\frac{1}{1101} + \frac{1}{402}\right)}} = \frac{0.103}{0.0291} = 3.53$$

41. Refer to the previous two questions. What assumptions are required to make the inferences valid?

I. The samples are paired.
II. The samples are independent.
III. The samples are random.
IV. The sample sizes are at least 30.
V. The successes/failures of each sample are greater than 5.
VI. The samples come from a normal population.
VII. The samples have unknown yet equal population variances.

(a) II, III, and V
(b) II, III, IV, and VI
(c) II, III, and IV
(d) II, III, IV, and V
(e) I, III, and IV

Solution: The samples must be random and independent. Since the data are categorical, then the success/failure condition is important.

42. Refer to the previous three questions. What are the respective distribution and *p*-value (range) for the test?

(a) *t*-distribution and $(0, \infty)$
(b) *t*-distribution and (0, 0.001)
(c) *t*-distribution and (0, 0.0005)
(d) *z*-distribution and (0, 0.001)
(e) *z*-distribution and (0, 0.0005)

Solution: The two sample proportion *z*-test is used, so the distribution is *z* (or standard normal). *p*-value = $P(Z > 3.53)$ but the value is cannot be found on the *z*-table, so *t*-table must be used. Since $3.291 < z_0 = 3.53 < \infty$, the *p*-value range is (0, 0.0005).

43. Refer to the previous four questions. With a 10% significance level, what is the appropriate conclusion?

(a) Since $\alpha = 0.10 > p$-value, the proportion of inactive caisson structures does not exceed the proportion of inactive well protector structures.
(b) Since $\alpha = 0.10 < p$-value, the proportion of inactive caisson structures does not exceed the proportion of inactive well protector structures.
(c) Since $\alpha = 0.10 > p$-value, the proportion of inactive caisson structures exceeds the proportion of inactive well protector structures.
(d) Since $\alpha = 0.10 < p$-value, the proportion of inactive caisson structures exceeds the proportion of inactive well protector structures.
(e) Since $\alpha = 0.10 > p$-value, the proportion of inactive caisson structures may exceed the proportion of inactive well protector structures.

Solution: Since the p-value $< \alpha = 0.10$, reject H_0. The proportion of inactive caisson structures exceeds the proportion of inactive well protector structures.

44. Construct a 99% confidence interval for the proportion of inactive well protector structures.

(a) 0.440 ± 0.0815
(b) 0.440 ± 0.0016
(c) 0.440 ± 0.0248
(d) 0.440 ± 0.0576
(e) 0.440 ± 0.0638

Solution: $\hat{p}_W \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_W(1-\hat{p}_W)}{n_W}} \rightarrow 0.440 \pm z_{0.005}\sqrt{\dfrac{0.440(1-0.440)}{402}}$
$\rightarrow 0.440 \pm (2.576)(0.0248) \rightarrow 0.440 \pm 0.0638$

45. Construct a 99% confidence interval for the proportion of inactive fixed platform structures.

(a) (0.205, 0.269)
(b) (0.227, 0.245)
(c) (0.215, 0.260)
(d) (0.212, 0.262)
(e) (0, 0.260)

Solution: $\hat{p}_F \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_F(1-\hat{p}_F)}{n_F}} \rightarrow 0.237 \pm z_{0.005}\sqrt{\dfrac{0.237(1-0.237)}{1897}} \rightarrow 0.237 \pm (2.576)(0.00977)$
$\rightarrow 0.237 \pm 0.0252 \rightarrow (0.212, 0.262)$

46. Refer to the previous two questions. Does the proportion of inactive well protector structures differ from the proportion of inactive fixed platform structures?

(a) Since the intervals do not overlap at all, the two proportions are different.
(b) Since the intervals overlap, the two proportions are different.
(c) Since the intervals do not overlap at all, the two proportions may be equal.
(d) Since the intervals overlap, the two proportions may be equal.
(e) Since the intervals do not overlap at all, the two proportions are equal.

Solution: The first interval is (0.358, 0.522), which does not overlap at all with (0.212, 0.262). Thus, the two proportions are different.

Use the following information for questions 47 – 49:

A circuit consists of 30 resistors connected in series and operating independently. The resistance of each resistor is exponentially distributed with a mean of 10 ohms. As the resistors are connected in series, the total resistance in the circuit is the sum of the individual resistances.

47. What is the probability distribution of the total resistance in a randomly selected circuit? In other words, name the distribution and specify its parameters.

(a) The total resistance follows an approximately normal distribution with a mean of 300 and a variance of 3000.
(b) The total resistance follows an exponential distribution with $\lambda = 1/10$.
(c) The total resistance follows an approximately normal distribution with a mean of 300 and a variance of 300.
(d) The total resistance follows an exponential distribution with $\lambda = 1/300$.
(e) The total resistance follows an approximately normal distribution with a mean of 300 and a variance of 90,000.

Solution: Let X_i denote the resistance of the i^{th} resistor for $i = 1, 2, \ldots, 30$. Let the variable *TOTAL* be defined as follows: $TOTAL = X_1 + X_2 + \ldots + X_{30}$. As each X_i follows an exponential distribution with a mean of 10 ohms and $E(X_i) = 1/\lambda$, then $\lambda = 1/10$. Note $V(X_i) = 1/\lambda^2 = 100$.

$$E(TOTAL) = E(X_1 + X_2 + \ldots + X_{30}) = E(X_1) + E(X_2) + \ldots + E(X_{30}) = 30 \cdot 10 = 300 \text{ ohms}$$

$$V(TOTAL) = V(X_1) + V(X_2) + \ldots + V(X_{30}) = 30 \cdot 100 = 3000 \text{ ohms}^2$$

Notice that as $n = 30$, the sampling distribution of the sample mean $\overline{X}$ approximately follows a normal distribution based on the Central Limit Theorem. Thus $TOTAL = 30\overline{X}$ also approximately follows a normal distribution with a mean of 300 and variance of 3000.

48. Refer to the previous question. Suppose you want to compare the total resistance of two randomly selected circuits. Respectively, what are the mean and standard deviation of the difference between the two circuits?

(a) Mean = 0 and standard deviation = 77.460.
(b) Mean = 300 and standard deviation = 3000.
(c) Mean = 0 and standard deviation = 154.919.
(d) Mean = 0 and standard deviation = 6000.
(e) Mean = 0 and standard deviation = 24,000.

Solution: Let $TOTAL_1$ and $TOTAL_2$ denote the total resistances of the two randomly selected circuits. As each $TOTAL_i$ ($i = 1, 2$) follows an approximately normal distribution, the difference of the variables ($TOTAL_1 - TOTAL_2$) also follows an approximately normal distribution with

$$E(TOTAL_1 - TOTAL_2) = E(TOTAL_1) - E(TOTAL_2) = 300 - 300 = 0$$

$$V(TOTAL_1 - TOTAL_2) = V(TOTAL_1) + V(TOTAL_2) = 3000 + 3000 = 6000$$

$$\sigma = \sqrt{6000} \approx 77.460$$

49. What is the probability that the total resistance of two randomly selected circuits differs by more than 100 ohms?

(a) 0.0167
(b) 0.9840
(c) 0.0985
(d) 0.1970
(e) 0.9015

Solution:

$$P(|TOTAL_1 - TOTAL_2| > 100) = 1 - P(|TOTAL_1 - TOTAL_2| \le 100)$$

$$= 1 - P(-100 < TOTAL_1 - TOTAL_2 \le 100) = 1 - P\left(\frac{-100}{\sqrt{6000}} < Z = \frac{(TOTAL_1 - TOTAL_2) - 0}{\sqrt{6000}} < \frac{100}{\sqrt{6000}}\right)$$

$$= 1 - \left(1 - 2 \cdot P\left(Z < -\frac{100}{\sqrt{6000}}\right)\right) = 2 \cdot P(Z < -1.29) = 2 \cdot 0.0985 = 0.1970$$

50. Three independent observations are selected randomly from a standard normal distribution. The probability that their average is greater than 0.5 is

(a) 0.6915
(b) 0.8660
(c) 0.3085
(d) 0.1922
(e) 0.8078

Solution: Let $\bar{X} = \dfrac{Z_1 + Z_2 + Z_3}{3}$, where Z_i ~ N(0, 1). Then, $E(\bar{X}) = 0$ and $V(\bar{X}) = \dfrac{\sigma^2}{n} = \dfrac{1}{3}$

→ $P(\bar{X} > 0.5) = P\left(\dfrac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \dfrac{0.5 - 0}{\sqrt{1/3}}\right) = P(Z > 0.87) = 1 - 0.8078 = 0.1922$