Final Exam

STAT 305, Section 3
Fall 2018

Instructions

- The exam is scheduled for 120 minutes, from 12:00 to 2:00pm. At 2:00pm the exam will end.
- A forumula sheet is attached to the end of the exam. Feel free to tear it off.
- You are allowed to use a self-produced one-page (front and back) formula sheet during this exam.
- You may use a calculator during this exam.
- Answer the questions in the space provided. If you run out of room, continue on the back of the page.
- If you have any questions about, or need clarification on the meaning of an item on this exam, please ask your instructor. No other form of external help is permitted attempting to receive help or provide help to others will be considered cheating.
- **Do not cheat on this exam.** Academic integrity demands an honest and fair testing environment. Cheating will not be tolerated and will result in an immediate score of 0 on the exam and an incident report will be submitted to the office of the dean.

Name: ______________________________

Student ID: ______________________________

1. Portable pneumatic pumps have long been used by hobby machinists, but are gaining renewed interest as robotics are becoming a more common part of our daily lives. Specifically, drones can make use of such pumps due to their weight benefits over hydraulic pumps with equivalent strength. Amazon is experimenting with using such pumps in its delivery drones and is interested in factors effecting the lifetime of the pumps.

   A team of engineers working for Amazon has been tasked with developing pumps for delivery drones that improve both time-in-flight (measured in hours) and the number of pump-actions (the number of times the drone uses the pump). With those goals in mind, the team designed three pumps: a single-piston/single-stage design, a dual-piston/single-stage design, and a dual-piston/dual-stage design. The engineers are testing the pumps with three types of compressed gas: OFN (oxygen-free nitrogen), Carbon dioxide, and Oxygen. Concerned about environmental effects, the engineers selected four locations to test the drones in - San Antonio, Seattle, Lexington, and Amherst in which to test the drones.

   In all four locations they tested the every design-gas combination in five drones (45 drones in each location). The drones delivered 5 pound packages from an initial location to a location 2000 meters away, retrieved the package, and returned it to the original point. The number of deliveries and the flight time were recorded for each drone.

   Note: (3 designs) · (3 gas types) · (5 drones) · (4 locations) = 180 total drones

   The goal is to determine is to identify how to maximize both flight time and number of deliveries.

   (a) (2 points) Is this an experiment or an observational study? Explain.

   Experimental study. The experimenter has an active rule in manipulation of the system under study by intentionally changing pumps & gas.

   (b) Identify each of the following and describe them as numeric (in which case, identify whether it is continuous or discrete) or categorcial (in which case list the possible levels).

      i. (2 points) Identify the response variable(s):

      The number of deliveries & the flight time.

      ii. (2 points) Identify the experimental variable(s):

      3 different pumps, 3 gas types

      iii. (2 points) Blocking variable(s):

      4 locations.

   (c) (2 points) Identify two controlled variables used in this process.

   designs (pumps) & gas types.

2. The quality of a diamond is often described in terms of the four C's: - carat (the diamonds mass, with 1 carat = 200 milligrams), - cut (a description of how well the diamond has been shaped), - color (the less color the more rare), - clarity (a description of the flaws in the diamond). Of these, carat is the most easily understood in terms of its impact on the diamonds value: all things being equal, the larger the diamond the higher its value.

   The following plot shows the sale price of 375 diamonds (in thousands of dollars), which were appraised by experts as having an ideal cut, the depth, internally flawless clarity, and being colorless or essentially colorless.

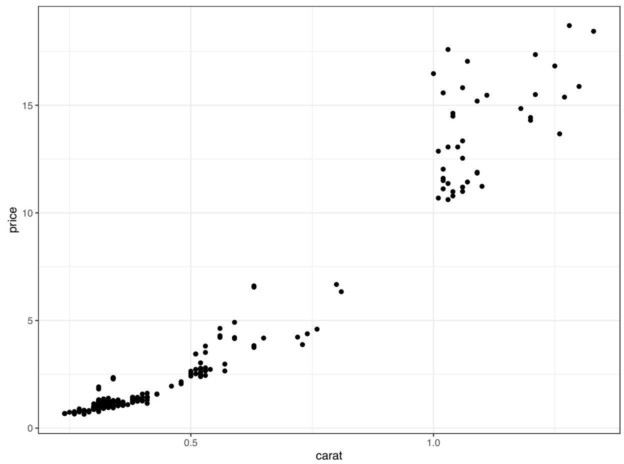


Figure 1: Plot depicting the sale price of 375 diamonds with the same quality of cut, clarity, and color. There is a general pattern indicating that higher carat (i.e., the mass) is associated with higher price.

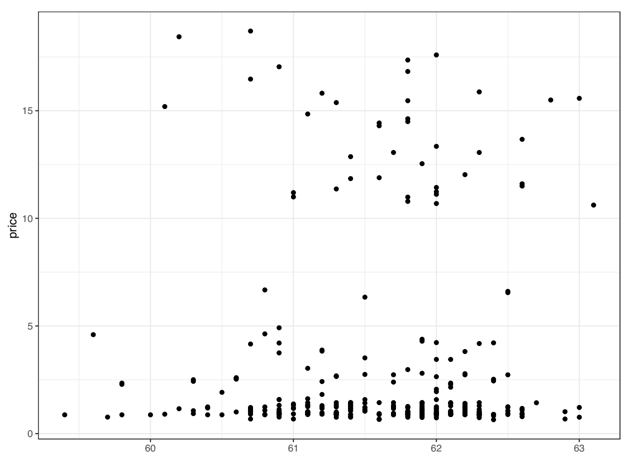


Figure 2: Plot depicting the sale price of 375 diamonds with the same quality of cut, clarity, and color. There is a general pattern indicating that higher carat (i.e., the mass) is associated with higher price.

The JMP output below comes from fitting a this quadratic relationship using price as the response (`price`) and carat (`carat`) and depth (`depth`) as the model variables.

Summary of Fit

RSquare	0.943787
RSquare Adj	0.943485
Root Mean Square Error	$\boxed{0.95508}$
Mean of Response	2.716083
Observations (or Sum Wgts)	375

$\sqrt{MSE} = \sqrt{S^2_{SF}}$

n

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	5697.2089	2848.60	3122.859
Error	372	339.3303	$\boxed{0.91}$	Prob > F
C. Total	374	6036.5393		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-3.039366	4.689684	-0.65	0.5173
carat	15.653187	0.198074	79.03	<.0001*
depth	-0.017474	0.076096	-0.23	0.8185

$p = 3$ $\hat\beta_0$ $\hat\beta_1$ $\hat\beta_2$

Effect Tests

(a) (3 points) Using the JMP output, write the equation of the fitted linear relationship between carat, depth and price.

$$\hat{y} = -3.039 + 15.6531 \text{ carat} - 0.01747 \text{ depth}$$

(b) (3 points) Using the fitted line, what do we suppose the price would be for a 1.07 carat and 62 depth diamond?

$$\hat{y} = -3.039 + 15.6531(1.07) - 0.01747(62)$$

$$= 12.62668$$

(c) (3 points) The actual price of a 1.07 carat with 62 depth diamond in the data is 11.434 thousand dollars. What is the residual for this specific diamond using the linear model?

$$e = y - \hat{y} = 11.434 - 12.62668$$

(d) (3 points) For the linear relationship, find r, the sample correlation coeffecient and R^2, the coeffecient of determination.

$$R^2 = 94.37\%$$

in linear regression: $$r = \sqrt{R^2} = \sqrt{0.9437} = 0.9714$$

(e) (3 points) Provide an estimate for σ^2.

$$\hat{\sigma}^2 = MSE = S^2_{SF} = 0.91$$

(f) (3 points) Provide an estimate for the variance of the coefficient of *depth*.

$$Var(b_2) = (SE(b_2))^2 = (0.076)^2 = 0.005776$$

(g) (5 points) Calculate and intepret the 95% two-sided confidence interval for the coefficient of *depth*.

$$b_2 \pm t_{(n-p,\, 1-\alpha/2)} SE(b_2)$$

$$= -0.017474 \pm t_{(375-3,\, 0.975)} (0.076)$$

$$= -0.017474 \pm (1.96)(0.076)$$

$$= (-0.166434, 0.131486)$$

(h) (10 points) Conduct a formal hypothesis test at the $\alpha = 0.05$ significance level to determine if there is significance relationship between price (y) and **carat** (x_1), holding depth constant.
Note: Write down all six steps for full credit.

① $H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

② $\alpha = 0.05$

③ I'll use $K = \dfrac{b_1 - 0}{SE(b_1)} \sim t_{(n-p)}$ assuming that 1

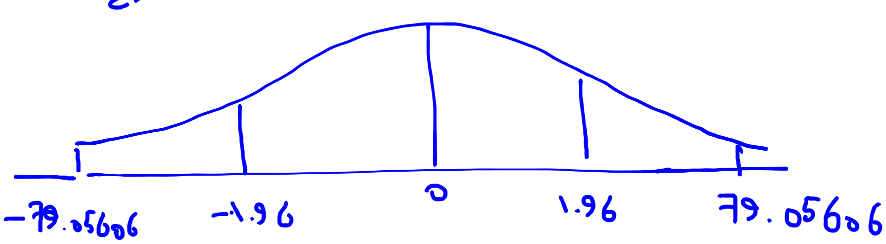
1- H_0 is true

2- $y = \beta_0 + \beta_1 \text{ carat} + \beta_2 \text{ depth} + \epsilon$ is valid.

④ $K = \dfrac{15.6531 - 0}{0.1980} = 79.05606 \sim t_{(375-3)}$

p-value = $P(|T| > K) = P(|T| > 79.05606)$

$t_{(375-3,\, 1-\alpha/2)} = t_{(372,\, 0.975)} = 1.96$



⑤ p-value < α; So we reject H_0.

⑥ There's enough evidence to reject H_0 concluding that the linear relationship between the price & carat is significant holding depth constant.

3. An arctic research station recently did a major overhaul to their server system hardware and the technicians are checking to make sure that there has been no loss in download speed. The previous download speed had an average of 63.4 Mbps. A systems analyst took 10 readings on the download speeds during the course of a day to check. Her results are below (in Mbps):

63.15, 62.85, 63.28, 63.03, 63.18, 62.91, 62.61, 62.98, 62.73, 63.03

The sample average is 62.98 and the sample variance is 0.043.

(a) (5 points) Provide a 90% confidence interval for the mean download speed.

$n < 25$; σ^2 unknown: $\bar{x} \pm t_{(n-1, 1-\alpha/2)} \, s/\sqrt{n}$

$t_{(9, 0.95)} =$ $\quad 62.98 \pm 1.833 \sqrt{\frac{0.043}{10}} = (62.8598, 63.1002)$

(b) (5 points) Provide a 95% lower confidence bound for the mean download speed.

$t_{9, .975} = 2.262 \quad (\bar{x} - t_{(9, 0.975)} \cdot s/\sqrt{n}, +\infty) = (62.98 - 2.262 \sqrt{\frac{0.043}{10}}, +\infty)$

$= (62.83167, +\infty)$

(c) (10 points) Conduct a hypothesis test at the 95% confidence level for the null hypothesis $\mu \geq 63.4$ against the alternative $\mu < 63.4$. Include your hypothesis statement, the choice of test statistic, the p-value, and your conclusion.

① $H_0: \mu = 63.4$ vs. $H_a: \mu < 63.4$

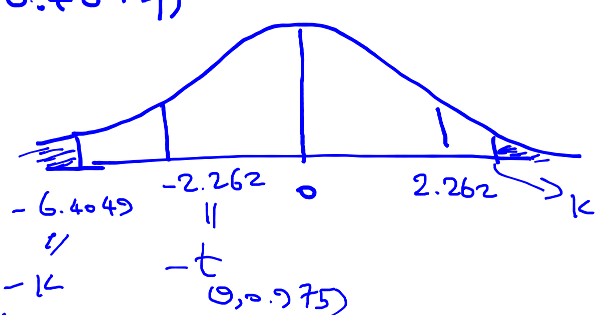
② $\alpha = 0.05$

③ Since $n < 25$, σ^2 unknown, I'll use $K = \frac{\bar{x} - 63.4}{s/\sqrt{n}} \sim t_{(n-1)}$ under the assumption that ① H_0 is true ② x_i's are iid $N(0,1)$.

④ $K = \frac{62.98 - 63.4}{\sqrt{\frac{0.043}{10}}} = -6.40494 \sim t_{(n-1)} = t_{(9)}$

p-value $= P(|T| > K) = P(|T| > |-6.4049|)$

$t_{(9, 0.975)} = 2.262$



⑤ Since $|K| > |t_{9, .975}|$,

$P(|T| > |K|) < \alpha = 0.05 \Rightarrow$ we reject H_0

⑥ There's enough evidence to reject H_0 concluding that the download speed is 63.4 Mbps.

4. A group of scientists are trying to understand the effects of temperature on two O-ring designs for a rocket. By placing the O-ring (attached to a valve) in a chamber and slowly lowering the chamber's temperature, the scientists are able to record the temperature at which the O-ring fails by monitoring when the valve begins to leak. After testing 10 O-rings for each type, the scientists find the mean failure temperature for the first O-ring design sample to be 50 K with a sample variance of 10 and the mean failure temperature of the second O-ring sample to be 53 K with a sample variance of 20.

$\bar{x}_1 = 50$, $s_1^2 = 10$

$\bar{x}_2 = 53$, $s_2^2 = 20$

two indep. samples

(a) (4 points) Provide a 95% confidence interval for the true failure temperature of the first O-ring design.

$$\bar{x}_1 \pm t_{(n-1,1-\alpha/2)} \cdot s_1/\sqrt{n_1}$$

$$= 50 \pm t_{(9,0.975)} \cdot \sqrt{\frac{10}{10}} = 50 \pm 2.262(1)$$

$$= (47.738, 52.262)$$

(b) (4 points) Provide a 95% confidence interval for the true failure temperature of the second O-ring design.

$n_1 = n_2 = 10$

$$\bar{x}_2 \pm t_{(n-1,1+\alpha/2)} \, s_2/\sqrt{n_2}$$

$$53 \pm t_{(9,.975)} \sqrt{\frac{20}{10}} = 53 \pm 2.262(\sqrt{2})$$

$$(49.80105, 56.19895)$$

(c) (10 points) Conduct a hypothesis test at the 90% confidence level to see if the true failure temperature of the first sample is equal to that of second sample.

① $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$

② $\alpha = 0.1$

③ I'll use $K = \frac{\bar{x}_1 - \bar{x}_2 - 0}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t_{(n_1+n_2-2)}$ assuming that ① the First sample $\overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ ② indep of second sample $\overset{iid}{\sim} N(\mu_2, \sigma_2^2)$ ③ and ④ $\sigma_1^2 \approx \sigma_2^2$ and ⑤ H_0 is true.

④ $$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{9(10)+9(20)}{20-2}} = 3.8729$$

$$K = \frac{50-53-0}{3.8729\sqrt{\frac{1}{10}+\frac{1}{10}}} = -1.7320$$

p-value $= P(|T| > |-1.7320|)$

$t_{(18,\ .95)} = 1.734$

⑤ since $|K| < t_{(18,\ .95)}$ => p-value > α, so we fail to reject H_0

⑥ There's not enough evidence to reject H_0 Concluding that the there's not significant difference between the true means of two O-rings.

5. Let X be a normal random variable with a mean of 5 and a varaince of 4 (i.e., $X \sim N(5, 4)$) and let Z be a random variable following a standard normal distribution. Find the following probabilities (note: the attached standard normal probability table may be helpful):

   (a) (2 points) $P(Z \leq 1.5)$ $= \Phi(1.5) = 0.93319$

   (b) (2 points) $P(|Z| \geq 1.25)$ $= P(Z > 1.25) + P(Z < -1.25)$

   $= 2\Phi(-1.25) = 2(0.10564) = 0.21129$

   (c) (2 points) $P(1 \leq X < 9)$ $= P\left(\frac{1-5}{\sqrt{4}} < Z < \frac{9-5}{\sqrt{4}}\right) = P(-2 < Z < 2)$

   $= \Phi(+2) - \Phi(-2) = 0.9772 - 0.0227$

   $= 0.9544$

   (d) (2 points) $P(|X| \leq 5)$

   $= P(-5 < X < 5) = P\left(\frac{-5-5}{\sqrt{4}} < Z < \frac{+5-5}{\sqrt{4}}\right) = P(-5 < Z < 0)$

   $\Phi(0) - \Phi(-5) = \frac{1}{2} - 0$

   $= \frac{1}{2}$

6. Suppose that X is a continuous random variable with probability density function (pdf):

$$f(x) = \begin{cases} 0 & x < 0 \\ 2e^{-2x} & x \geq 0 \end{cases}$$

   (a) (3 points) Find F(x), the cumulative probability function.

   $F(x) = P(X \leq t) = \int_0^t 2e^{-2x}\,dx = -e^{-2x}\Big|_0^t = 1 - e^{-2t}$

   $\Rightarrow F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-2t} & x \geq 0 \end{cases}$

   (b) (3 points) What is the probability that X takes a value less than 1?

   $P(X < 1) = F(1) = 1 - e^{-2(1)} = 1 - e^{-2} = 0.8646$

   (c) (3 points) What is the probability that X takes a value greater than 2?

   $P(X > 2) = 1 - P(X \leq 2) = 1 - F(2)$

   $= 1 - (1 - e^{-2(2)}) = e^{-4} = 0.01831$

7. Suppose that X is a discrete random variable with the following probability function:

$$f(x) = \begin{cases} \dfrac{x^2}{c} & x = -3, -2, -1, 0, 1, 2, 3 \\ 0 & o.w \end{cases}$$

where c is a constant.

(a) (4 points) Find the value of c that makes $f(x)$ a valid probability function.

$\sum f(x_i) = 1 \longrightarrow \frac{1}{c}(-3^2 + (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2) = 1$

$\Rightarrow c = 28$

(b) (3 points) Find $P(X \geq 2)$.

$P(X \geq 2) = P(X=2) + P(X=3) = f(2) + f(3)$

$= \frac{1}{28}(2)^2 + \frac{1}{28}(3)^2 = \frac{4}{28} + \frac{9}{28}$

$= \frac{13}{28}$

(c) (3 points) Find $E(X)$.

$EX = \sum x \cdot f(x) = \sum x \cdot \frac{x^2}{28}$

$= \frac{1}{28} \sum x^3 = \frac{1}{28}((-3)^3 + (-2)^3 + (-1)^3 + 0^3 + 1^3 + 2^3 + 3^3)$

$= 0$

8. Let X be a random variable with Binomial distribution $x \sim Binomial(n = 10, p = 0.2)$ and Y be a Normally distributed random variable with mean $\mu = 1$ and variance $\sigma^2 = 4$. (X, Y are indep.)

   (a) (2 points) Find the expected value of $2^{-1}X + 3Y - 4$

$$E(2^{-1}X + 3Y - 4) = 2^{-1}EX + 3EY - 4$$
$$= 2^{-1}(10(0.2)) + 3(1) - 4 = 0$$

   (b) (4 points) Find the **standard deviation** of $3X + 2^{-1}Y + 1/2$

$$SD(3X + \tfrac{1}{2}Y + \tfrac{1}{2}) = \sqrt{var(3X + \tfrac{1}{2}Y + \tfrac{1}{2})}$$
$$= \sqrt{Var(3X) + Var(\tfrac{1}{2}Y)}$$
$$= \sqrt{9var(X) + \tfrac{1}{4}var(Y)}$$
$$= \sqrt{9 \cdot 10(0.2)(0.8) + \tfrac{1}{4}(4)}$$
$$= \sqrt{15.4}$$
$$= 3.9242$$