Quiz II

STAT 305, Section 3
FALL 2019

Instructions

- The quiz is scheduled for 80 minutes, from 09:30 to 10:50 AM. At 10:50 AM the exam will end.
- Total points for the exam is 69. Points for individual questions are given at the beginning of each problem. Show all your calculations clearly to get full credit. Put final answers in the box at the right (except for the diagrams!).
- A forumula sheet is attached to the end of the exam. Feel free to tear it off.
- 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 dean's office.

Name: ______________________________

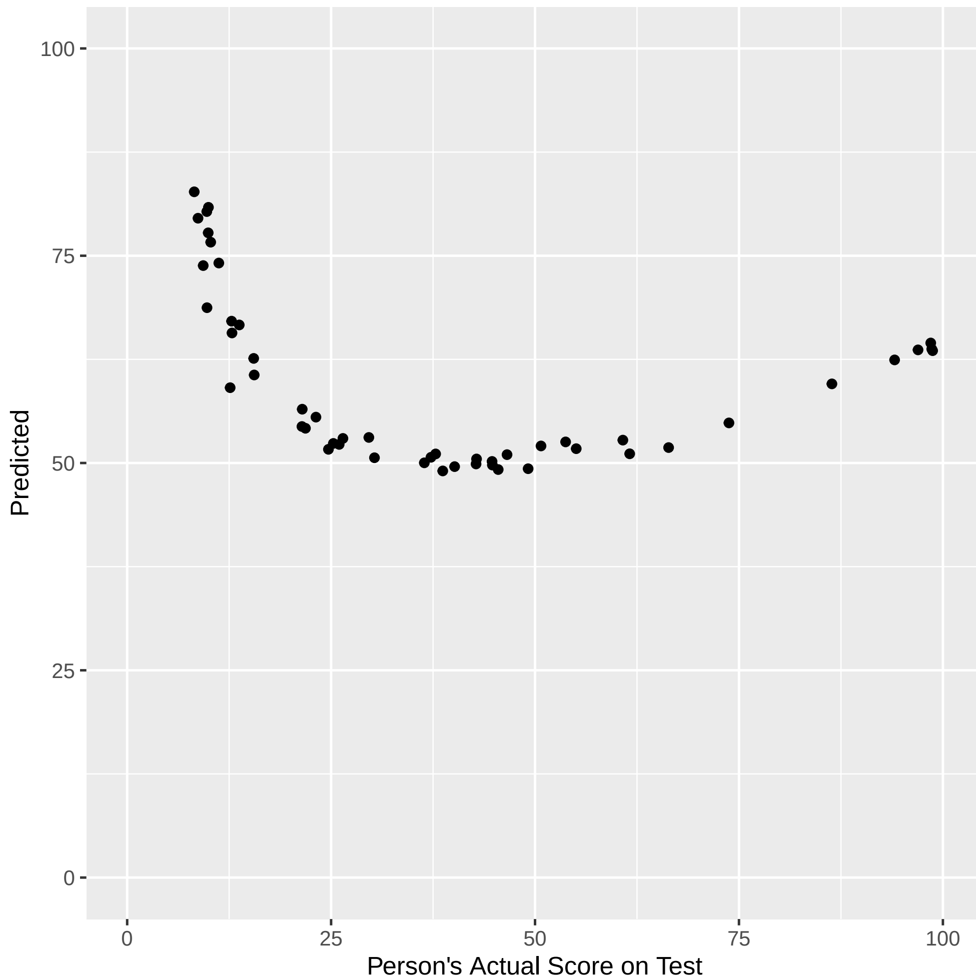
Student ID: ___________________________

1. Professional engineers often encounter issues relating to *human resources* as they advance in their careers (building a better team of employees is after all not too different than improving any other system, at least on paper). However, many of the "laws" governing human behavior are very different than the strict laws of physics. For instance, a phenomenon known as the Dunning-Kruger effect states that for a given skill incompetent people will

   - fail to recognize their own lack of skill
   - fail to recognize genuine skill in others
   - fail to recognize the extremity of their inadequacy
   - recognize and acknowledge their own lack of skill, after they are exposed to training for that skill

   A group of 50 job applicants are asked to estimate their skill in technical writing. They are told they will be taking a test with a mean score of 50 and asked to guess what their score will be. Then they are given the test and get an actual score.

   The results are depicted below (using the actual score on the x-axis):



Here are some summaries of the data (again using the actual score as the x-value and the person's evaluation of their score as the y-value):

$$\sum_{i=1}^{50} x_i = 1922 \qquad \sum_{i=1}^{50} x_i^2 = 110659$$

$$\sum_{i=1}^{50} y_i = 2954 \qquad \sum_{i=1}^{50} y_i^2 = 179606$$

$$\sum_{i=1}^{50} x_i y_i = 108893$$

(a) Using the summaries above, fit a linear relationship between **the actual score** (x) and **the guessed score** (y).

i. (5 points) Write the equation of the fitted linear relationship.

$\hat{y} =$

ii. (5 points) Find and interpret the value of R^2 for the fitted linear relationship. $R^2 =$

iii. (5 points) Using the fitted line, what do we suppose a person will guess their score will be if they actually scored a 40.14.

$\hat{y} =$

iv. (2 points) A person who scored a 40.14 on the test predicted that they would score 49.56. What is the residual for this person using the linear relationship?

$e =$

(b) The JMP output below comes from fitting a quadratic model using the actual score ("`actual_score`") and the square of the actual score (`actual_score^2`).

Response guess_score

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	4031.5440	2015.77	93.5569
Error	47	1012.6597	21.55	**Prob > F**
C. Total	49	5044.2038		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	82.626343	1.864461	44.32	<.0001*
actual_score	-1.246181	0.091868	-13.56	<.0001*
actual_score^2	0.011013	0.000873	12.62	<.0001*

i. (5 points) Write the equation of the fitted quadratic relationship.

$\hat{y} =$

ii. (5 points) Find and interpret the value of R^2 for the fitted quadratic relationship. $R^2 =$

iii. (5 points) Using the fitted quadratic relationship, what do we suppose a person will guess their score will be if they actually scored a 98.74.

$\hat{y} =$

iv. (2 points) A person who scored a 98.74 on the test predicted that they would score 63.55. What is the residual for this person using the quadratic relationship?

$e =$

2. Suppose the following is the probability distribution for X.

x	-2	0	1	2	3
f(x)	0.1	a	0.2	0.2	0.3

i. (3 points) Find the value of a that makes this a valid probability distribution.

$a =$

ii. (6 points) Calculate the expected value and the standard deviation of X.

E(X)=

SD(X)=

iii. (2 points) Find the probability that $P(|X| = 2)$

$P(|X| = 2) =$

3. As the discussion in the class, King Joffrey was poisoned and Tyrion was found guilty for his death. Tyrion then decided to let the gods decide his fate and demanded a trial by combats. Then it was supposed to have five completely independent combats to decide on his guilt or innnocence. The one who fought for him (his combatant) had $p = 0.7$ probability of success in each combat. Let X be the random variable associated with the number of combats his combatant wins out of the five combats.

   i. (2 points) Precisely specify the probability distribution associated with this trial.

   ii. (3 points) What is the probability that his combatant wins all combats? $\boxed{p =}$

   iii. (3 points) What is the probability that his combatant wins at least two combats? $\boxed{p =}$

   iv. (2 points) What is the expected value of this random variable X? $\boxed{E(X) =}$

   v. (2 points) What is the variance of this random variable X? $\boxed{Var(X) =}$

4. Suppose that an eddy current nondestructive evaluation technique for identifying cracks in critical metal parts has a probability of about .20 of detecting a single crack of length .003in. in a certain material. Let Y be the number of specimens inspected in order to obtain the first crack detection.

   i. (2 points) Precisely state the distribution of Y, giving the values of any parameters necessary.

   ii. (3 points) Find the probability that $P(Y = 5)$

   $\boxed{P(Y = 5) =}$

   iii. (3 points) Find the probability that $P(Y \leq 4)$

   $\boxed{P(Y \leq 4) =}$

   iv. (2 points) What is EY?

   $\boxed{E(Y) =}$

   v. (2 points) What is $SD(Y)$ (the standard deviation of Y)?

   $\boxed{SD(Y) =}$

STAT 305 Quiz II
Reference Sheet

Numeric Summaries

mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
population variance	$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
population standard deviation	$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
sample variance	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
sample standard deviation	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Linear Relationships

Form	$y \approx \beta_0 + \beta_1 x$
Fitted linear relationship	$\hat{y} = b_0 + b_1 x$
Least squares estimates	$b_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$
	$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$
	$b_0 = \bar{y} - b_1\bar{x}$
Residuals	$e_i = y_i - \hat{y}_i$
sample correlation coeffecient	$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$
	$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}}$
coeffecient of determination	$R^2 = (r)^2$
	$\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$

Multivariate Relationships

Form	$y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$
Fitted relationship	$\hat{y} \approx b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$
Residuals	$e_i = y_i - \hat{y}_i$
Sums of Squares	$\text{SSTO} = \sum_{i=1}^{n}(y_i - \bar{y})^2$
	$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
	$\text{SSR} = SSTO - SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
coeffecient of determination	$R^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$
	$R^2 = \frac{\text{SSTO}-\text{SSE}}{\text{SSTO}}$
	$R^2 = \frac{\text{SSR}}{\text{SSTO}}$
	$\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$

Functions

Quantile Function $Q(p)$ For a univariate sample consisting of n values that are ordered so that $x_1 \leq x_2 \leq \ldots \leq x_n$ and value p where $0 \leq p \leq 1$, let $i = \lfloor n \cdot p + 0.5 \rfloor$. Then the quantile function at p is:

$$Q(p) = x_i + (n \cdot p + 0.5 - i)(x_{i+1} - x_i)$$

Basic Probability

Definitions

Random experiment	A series of actions that lead to an observable result. The result may change each time we perform the experiment.
Outcome	The result(s) of a random experiment.
Sample Space (S)	A set of all possible results of a random experiment.
Event (A)	Any subset of sample space.
Probability of an event ($P(A)$)	the likelihood that the observed outcome of a random experiment is one of the outcomes in the event.
A^C	The outcomes that are not in A.
$A \cap B$	The outcomes that are both in A and in B.
$A \cup B$	The outcomes that are either A or B.

General Rules

Probability A given B	$P(A\|B) = P(A \cap B)/P(B)$
Probability A and B	$P(A \cap B) = P(A\|B)P(B) = P(B\|A)P(A)$
Probability A or B	$P(A \text{ or } B) = P(A) + P(B) - P(A, B)$

Independence

Two events are called independent if $P(A, B) = P(A) \cdot P(B)$. Clever students will realize this also means that if A and B are independent then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Joint Probability

Joint Probability	The probability an outcome is in event A and in event $B = P(A, B)$.
Marginal Probability	If $A \subseteq B \cup C$ then $P(A) = P(A \cap B) + P(A \cap C)$.
Conditional Probability	For events A and B, if $P(B) \neq 0$ then $P(A\|B) = P(A \cap B)/P(B)$.

Discrete Random Variables

General Rules

Probability function	$f_X(x) = P(X = x)$
Cumulative probability function	$F_X(x) = P(X \leq x)$
Expected Value	$\mu = E(X) = \sum_x x f_X(x)$
Variance	$\sigma^2 = Var(X) = \sum_x (x - \mu)^2 f_X(x)$
	or, $Var(X) = \sum_{i=1}^{n} x_i^2 \cdot f(x_i) - \mu^2$
	or, $Var(X) = \sum\limits_x (x - \mathrm{E}X)^2 f(x) = \mathrm{E}(X^2) - (\mathrm{E}X)^2$
Standard Deviation	$\sigma = \sqrt{Var(X)}$

Joint Probability Functions

Joint Probability Function	$f_{XY}(x, y) = P[X = x, Y = y]$
Marginal Probability Function	$f_X(x) = \sum_y f_{XY}(x, y)$ $f_Y(y) = \sum_x f_{XY}(x, y)$
Conditional Probability Function	$f_{X\|Y}(x\|y) = f_{XY}(x, y)/f_Y(y)$ $f_{Y\|X}(y\|x) = f_{XY}(x, y)/f_X(x)$

Geometric Random Variables

X is the trial count upon which the first successful outcome is observed performing independent trials with probability of success p.

Possible Values	$x = 1, 2, 3, \ldots$
Probability function	$P[X = x] = f_X(x) = p(1 - p)^{x-1}$
Expected Value	$\mu = E(X) = \frac{1}{p}$
Variance	$\sigma^2 = Var(X) = \frac{1-p}{p^2}$

Binomial Random Variables

X is the number of successful outcomes observed in n independent trials with probability of success p.

Possible Values	$x = 0, 1, 2, \ldots, n$
Probability function	$P[X = x] = f_X(x) = \frac{n!}{(n-x)!x!} p^x (1 - p)^{n-x}$
Expected Value	$\mu = E(X) = np$
Variance	$\sigma^2 = Var(X) = np(1 - p)$

Poisson Random Variables

X is the number of times a rare event occurs over a predetermined interval (an area, an amount of time, etc.) where the number of events we expect is λ.

Possible Values	$x = 0, 1, 2, 3, \ldots$
Probability function	$P[X = x] = f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$
Expected Value	$E(X) = \lambda$
Variance	$Var(X) = \lambda$