MATH 1342

Review for Exam 2

24 Multiple Choice Questions

Terms and Vocabulary:

Know sampling techniques

Experiment vs. Observational Study

Uniform Distribution vs. Normal Distribution

Normal Distribution (x) vs. Standard Normal Distribution (z)

A density curve consists of two line segments, going from (0,5) to (0.1,5) and from (0.1,5) to the x-axis. *

a. Sketch the distribution.

b. Find where it crosses the x-axis

c. Determine P(X>0.1) $A_2 = 0.5$

d. Find the median. $X = 0.1$

b) $A_1 = b \cdot h = (0.1)(5) = .5$

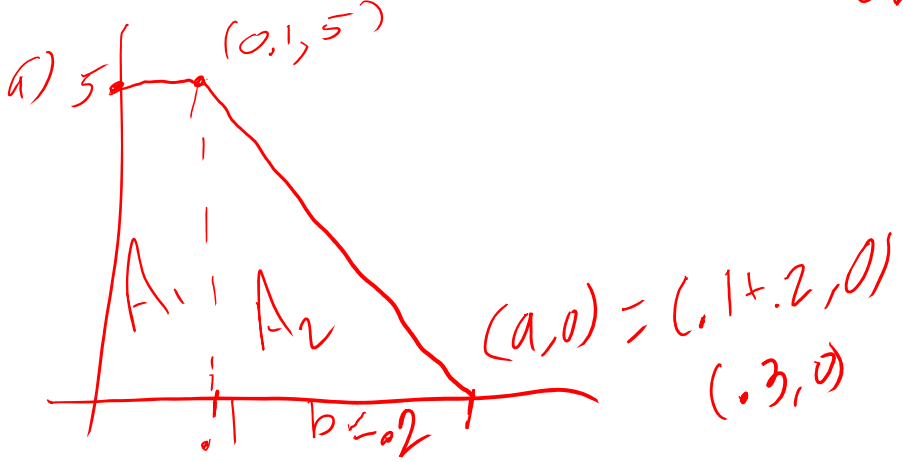
$A_2 = 1 - A_1 = 1 - .5 = .5$

$A_2 = \frac{1}{2} \cdot b \cdot h$

$.5 = \frac{1}{2} \cdot b \cdot 5$

$\frac{.5}{2.5} = \frac{2.5b}{2.5}$

$.2 = b$

a) 

The weight of persons of a certain age is normally distributed with a mean of 160 pounds and a standard deviation of 15 pounds. What range of weights should the middle 68% contain?

$68\% \rightarrow [\mu - \sigma, \mu + \sigma]$

$[160 - 15, 160 + 15]$

$[145, 175]$

Empirical Rule: 68%, 95%, 99.5%

If X is normally distributed with a mean of 150 and a standard deviation of 15, what is P(X>155)?

Greater Than: Keep "1-"

```
> 1-pnorm(155,150,15)
[1] 0.3694413
```

If you are looking for probability, it will be pnorm command (regardless of z or x)

Find the value of c so that P(Z>c) = 0.32.

Greater Than: "1-"

```
> qnorm(1-0.32)
[1] 0.4676988
```

If you are given a probability and looking for a c-value, use a qnorm command, regardless of x or z)

A sample of 60 data points is selected from a population with mean of 140 and variance of 13. Determine the mean and standard deviation for the sample.

$n = 60$

$\mu = 140$

$\sigma^2 = 13$

$\downarrow$

$\sigma = \sqrt{13}$

$\mu_{\bar{x}} = \mu = 140$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{13}}{\sqrt{60}}$

```
sqrt(13)/sqrt(60)
L] 0.4654747
```

Regression Analysis was done on a data set with the following results:

$y = 3x - 1$ ($r^2 = 0.55$) Linear

$y = 2x^2 + 4x - 3$ ($r^2 = 0.88$) Quadratic

$y = 0.33e^{.22x}$ ($r^2 = 0.75$) Exponential

$y = 2.3 + \ln(x)$ ($r^2 = 0.62$) Logarithmic

Which model is the best fit for this data set?

Quadratic (r^2 is closest to 1)

In a large city, 75% of adults drive to work. A sample of 50 adults are selected. What is the mean and standard deviation of this sample?

$p = .75$ $\quad \mu_{\hat{p}} = p = .75$

$n = 50$

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

```
sqrt(.75*(1-.75)/50)
] 0.06123724
```

What is the probability that the sample has less than 30 people drive to work? $x = 30, \quad \hat{p} = 30/50$

$\hat{p} = \frac{x}{n}$ $\quad x/n, p, \sqrt{\frac{p(1-p)}{n}}$

```
pnorm(30/50,0.75,sqrt(.75*(1-.75)/50))
L] 0.007152939
```

Based on the two-way table describing car-type to driver's age, determine what percent of 40-65 year olds drive pick up trucks.

	18-25	26-40	40-65	Over 65
2-Door	25	40	54	32
4-Door	30	55	40	26
MiniVan	15	50	43	28
PickUp Truck	60	40	37	22

$$P(\text{Truck} \mid 40\text{-}65) = \frac{37}{54+40+43+37} = 0.2126$$

A researcher is running a study to determine the proportion of men that wear neckties to work. To collect data, he keeps track of how a group of men are dressed as the enter an office building. *

Is this an observational study or an experiment? Why? No control

Identify the sample and population of this study. sample: group that are watched pop: all men

In order to match the demographics of the company, the researcher wants 10% of the sample to be management and 90% to not be. How can this be done using stratified sampling? Divide population into management and non-management, select from each.

The researcher also decides to keep track of subjects' age. This would be considered: stratifying, bias, blocking, control?

If the researcher removed all employees of Asian descent from the sample, what is this considered? under-representation bias

A researcher wishes to understand if there is a relationship between room lighting and quality of work. To do this, he gathers 300 people and places them in three separate rooms (one with dim light, one with normal light, and one with bright light). They are given one hour to proofread an essay, and afterwards, their completion and accuracy is recorded. *

Is this an observational study or an experiment? Why? Active control by the researcher

Identify the sample and population of this study. sample: 300 people tested
pop: all people that work

Describe how matching, based on previous work, can be used to create these groups. Each group will contain good workers, average workers, poor workers based on previous exp.

Identify the control and treatment groups. normal light dim, bright light

The subjects did not know what the purpose of the study was, and the researcher did not know who was assigned to which group until after data was collected. What feature does this lack of knowledge describe? blindness (double-blindness)

Describe Convenience Sampling

Non-probability sampling based off of proximity or ease of selection.

What type of sampling is used here:
The customer service score of a store is determined by the results of a survey attached the bottom of the store receipt.

✓ voluntary sampling

A normal distribution has a mean of 150 and a standard deviation of 12. Determine the z-score for a data point that has a value of 133.

$$Z = \frac{x - \mu}{\sigma}$$

$$Z = \frac{133 - 150}{12}$$

```
(133-150)/12
] -1.416667
```

Base your answers to the following on the table comparing age and average commuting distance. *

explanatory response

Ages(Decade)	Distance (Miles)
20	10
30	25
40	35
50	40
60	30
70	20

a. Construct a scatter plot

b. Find the LSRL

c. Interpret the slope

d. Find the correlation coefficient

e. Find the coefficient of determination

f. Find the residual value for a person in their 40s. $y - \hat{y}$

$35 - (.20(40) + 17.67)$

```
34-(.20*40+17.67)
.] 8.33
```

g. Construct the residual plot

h. Is this a good model for this data?

No.

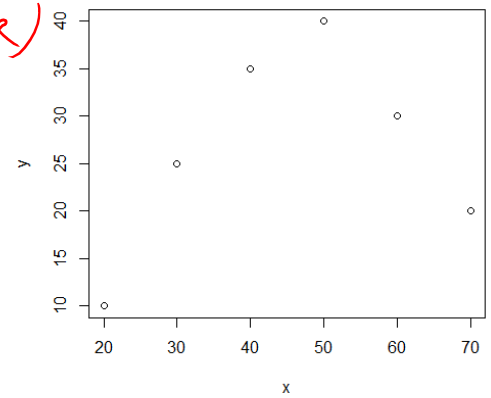
```
> assign("x",c(20,30,40,50,60,70))
> assign("y",c(10,25,35,40,30,20))
> plot(x,y)
> regline=lm(y~x)
> regline

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
      17.67         0.20
```

b) $\hat{y} = .20x + 17.67$

or

$\hat{y} = 17.67 + .20x$

a)



c. There is an increase of 0.20 miles for every year increase of age.
For every year of age, there is an increase of 0.2 mile.

d

```
> cor(x,y)
[1] 0.3464102
```

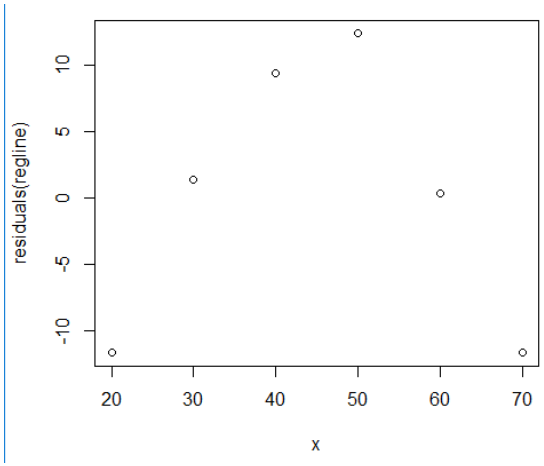
e

```
> cor(x,y)^2
[1] 0.12
```

r: There is a weak relationship between the variables.
r^2: 12% of the variation in y is explained by the LSRL
Not a good model

g)

```
plot(x,residuals(regline))
```



Use a random digit table to simulate the following experiment:

A pet store stocks merchandise based on the primary pet their customer owns. It has been seen that 40% of customers are cat-owners, 50% are dog owners, and 10% own other pets.

How would you assign the digits on a Random Number Table to match these specifications?

Cat Owners: 40% (need 4 digits): 0, 1, 2, 3
Dog Owners: 50% (need 5 digits): 4, 5, 6, 7, 8
Other Pets: 10% (need 1 digit): 9

The value of the digit is not the important part to this assignment. The important part is the number of digits assigned matching the percents given in the question.

Popper 21

Fill out choice D for questions 1 – 5.