AP Statistics – Practice FRQ Final Exam **Name:** Solutions

For full credit solve all problems using calculator to find any p-values or confidence intervals and show work as we've done in class. Don't forget to check conditions and give reasons why you believe they are met.

#1) A sports statistician wished to investigate how distance from the basket affect shot accuracy in basketball. A large random sample of basketball games were selected for study and for each game, video footage of every shot attempted was analyzed. For each shot the distance it was recorded whether the shot went in the basket (successful shot) or missed along with the distance the player was from the basket (to the nearest foot) when the ball was thrown. Then for each distance from the basket the percentage of shots that were successful was recorded, for example: at a distance of 3 feet 44% of shots were made, and at a distance of 20 feet 8% of shots were made. The statistician performed a linear regression of the data set which produced the following software output:

Predictor	Coef	SE Coef	T	P
Constant	50.3885	27.143	1.8564	0.0317
Distance	-2.4511	1.6130	-1.5196	0.0643
s = 4.1572	$r^2 = 0.6403$			

(a) Use the computer output above to determine the equation of the sample's least squares regression line. Identify all variables used in the equation.

$\hat{y} = 50.3885 - 2.4511x$

x: distance from basket (ft)

y: % of shots successful

(b) Interpret the slope of the LSRL for the sample in the context of the problem.

$b = -2.4511$ %/ft

For every additional foot the player is from the basket, the percentage of shots that are successful from that distance decreases by 2.45%, on average.

(c) Interpret the value for r^2 in the context of the problem.

About 64% of the variation in percentage of successful shots is explained by the LSRL model which relates percentage of successful shots to distance from the basket.

#2) In order to be a certified welder, students must take an approved course at a vocational school and pass a welding certification test. A researcher wished to determine if there were differences in how effective welding school programs were in different parts of the U.S., so an SRS of students enrolled in welding programs was selected from each of three regions (West Coast, Middle America, East Coast) and for each student it was recorded whether they passed or failed their welding certification test on the first attempt. The data collected are display in the following table:

	U.S. Region			
	West Coast	**Middle America**	**East Coast**	*total*
Passed	**215**	**327**	**298**	*840*
Failed	**27**	**41**	**35**	*103*
total	*242*	*368*	*333*	*943*

(a) If a welding student is selected at random from this sample, what is the probability that the selected student will be from the East Coast?

$P(\text{East Coast}) = \boxed{\frac{333}{943} = .3531}$

(b) If a welding student is selected at random from this sample, what is the probability that the selected student will have passed, given that they are from the West Coast?

$P(\text{passed} \mid \text{West Coast}) = \boxed{\frac{215}{242} = .8884}$

#3) A large tire manufacturer wishes to determine whether there are differences in the stopping distances of the tires manufactured at their 10 factories in Canada and their 10 factories in England. They will select a sample of sets of tires from each country and subject the sets of tires to standardized stop-distance testing which records the number of feet it takes a car to stop on dry pavement with the set of tires under test installed.

(a) What is the name of this sampling technique being used in each of these following sampling scenarios?

(i) All of the sets of tires produced in all factories across Canada in a given month are numbered with a unique serial number and a random number generator is used to generate 100 random serial numbers. The sets of tires with these serial numbers are included in the Canadian sample for testing. This process is repeated with the factories across England to produce the English sample for testing.

Simple Random Sample (SRS)

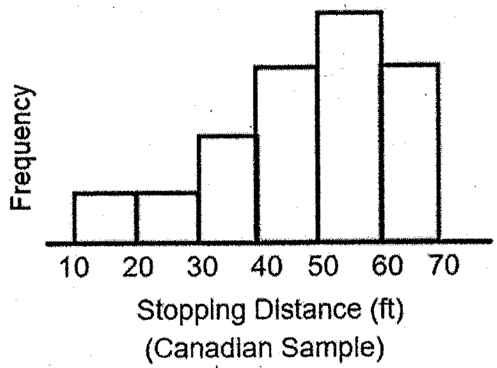
(ii) Three factories are chosen randomly from the factories in Canada, and then all of the tire sets produced in a given month are included in the Canadian sample for testing. Three factories are chosen randomly from the factories in England, and then all of the tire sets produced in a given month are included in the English sample for testing.

Cluster Sampling

(iii) A Simple Random Sample of tire sets from each of the 10 factories in Canada is included in the Canadian sample for testing. A Simple Random Sample of tire sets from each of the 10 factories in England is included in the English sample for testing.

Stratified Random Sampling

(b) The tire sets selected in the sample from each country were tested for stopping distance. A histogram and summary statistics for the stopping distances from the Canadian sample is shown:



$\overline{X} = 41.7$

$s = 14.2$

$Median = 52.4$

$Q1 = 32.6$

$Q3 = 60.2$

$n = 243$

Describe this distribution of stopping distances for the Canadian sample tire sets.

The distribution of stopping distances for the Canadian sample tire sets is skewed left, with a median of 52.4 ft, and an IQR of $60.2 - 32.6 = 27.6$ ft. There are no apparent outliers.

(c) Summary statistics for the science test scores from the samples of students in both districts is provided below:

Canadian Sample	English Sample
$\overline{X} = 41.7$	$\overline{X} = 43.8$
$s = 14.2$	$s = 11.9$
$Median = 52.4$	$Median = 56.3$
$Q1 = 32.6$	$Q1 = 34.4$
$Q3 = 60.2$	$Q3 = 63.1$
$n = 243$	$n = 215$

Conduct a hypothesis test to determine if there is a statistically significant difference in the mean stopping distance between tire sets manufactured in the factories in Canada vs. England.

μ_C = mean stopping distance for all Canadian factory tires
μ_E = mean stopping distance for all English factory tires

$H_0: \mu_C = \mu_E$
$H_A: \mu_C \neq \mu_E$

2 Sample t-test

conditions

- SRS? assume the samples are representative of cars in the country factories
- groups indep? different cars from different countries
- n<10% pop? 243 & 215 < 10% of these countries' cars
- nearly normal? both sample sizes are large so we can assume both samples are nearly normal.

perform a 2-SampTTest in a calculator using:

$\overline{X_1} = 41.7$ $\quad$ $\overline{X2} = 43.8$ $\quad$ $\mu_1 \neq \mu_2$
$Sx_1 = 14.2$ $\quad$ $Sx_2 = 11.9$ $\quad$ no pooling
$n_1 = 243$ $\quad$ $n_2 = 215$

$t = -1.72$
p-value $= .0859$
$df = 454.69$

With $\alpha = .05$, p-value $= .0859$ is high so we fail to reject H_0. We do not have sufficient statistical evidence to conclude that there is a statistically significant difference in the mean stopping distance between tire sets manufactured in the factories in Canada and England.