AP Stats Sentence Frames

Chapter 1 Descriptive Statistics

Describe/Compare the distribution/s……C.U.S.S.
Center(mean & median), **Unusual Features**(outliers?), **Shape**(skewed Left or Right, fairly symmetric), **Spread**(Range, max vs min)

Describing a single distribution
The_____ has a median of _____ and a mean of _____.
The mean/median of _____ is smaller/larger than the median/mean of _____.
The distribution of _____ has a mean less/greater than the median which supports the strong/weak skew to the left/right.
The distribution of _____ has an outlier at _____ which pulls the mean to the left/right creating the skew to the left/right.
The _____ have no outliers present.
The _____ have _____ that range from _____ to _____.
We have a large spread of _____ due to the fact that the minimum value is _____ and the maximum value is _____.

Comparing more than one distribution
_____ has a larger/smaller mean/median than _____.
Both distributions are similar in shape which are clearly _____.
_____% of the _____'s in the distribution of _____ are more/less than _____% of the _____'s in _____.
_____ which _____'s range from _____ to _____ has a larger/smaller spread than _____ which _____'s range from _____ to _____.
Neither distributions have any outliers.
_____ and _____ share the same _____.
Both _____ and _____ can be described as _____.
By comparing _____ and _____, it became clear that _____.

Chapter 3 Linear Regression

Interpret the correlation "r"

r = _____. There is a **strong/moderate/weak positive/negative** linear relationship between **explanatory variable** and **response variable**.

r = _____. There is a **strong/moderate/weak positive/negative** linear association between **explanatory variable** and **response variable**.

Interpret the coefficient of determination "r^2"

____% of the variation in **response variable** can be explained by the linear relationship with **explanatory variable**

Example…..92.3% of the variation in pulse rate can be explained by the linear relationship with walking speed.

Interpret the slope "b"

For every 1 unit increase in **explanatory variable** our model predicts and average **increase/decrease** of _____ in **response variable**.

Example…..For every additional mph of walking speed our model predicts and average increase of 16.27189 in pulse rate.

Interpret the y-intercept "a"

At an **explanatory variable** of zero units our model predicts _____ in **response variable** .

Example…..At a walking speed of 0 mph out model predicts a pulse rate of 63.235……..A nicer way to write this would be……Our model predicts a resting pulse rate of 63.235.

Comment on a residual plot with a curved pattern

Based on the curved pattern on this residual plot it appears that a linear model may not be a good fit.

The curved pattern in this scatterplot reveals that a linear regression model would not be

appropriate for modeling the relationship between these variables.

Comment on a residual plot with a random scatter

The residual plot shows a random scatter about the regression line, it appears that a linear model is a good fit.

Chapter 5 Experimental Design

Using a random digit table….D.A.R.E.
Determine the # of digits at a time, Assign digits, Repeats, Ending rule.

We will take ___ # of digits at a time starting with line ____ on the random digit table. Digit ___ will represent _____, ……. and digit ____ will represent _____. We will skip/use repeats and #'s larger than _____ will be ignored. We will stop taking digits once ____ unique #'s have been selected.

Assigning treatments or subjects in an experiment
Write names/numbers on pieces of paper and place in a hat. Randomly draw out _____ pieces of paper, these will be assigned to _____. The next _____pieces of paper will be assigned to_____, etc… Compare Results.

Chapter 7 Random Variables.

The phrase that must be written when you are combining or taking the difference between random variables.

The combined _____ also follow a normal distribution.

The difference between _____ also follow a normal distribution.

Chapter 8 Binomial and Geometric Distributions

Identifying a binomial distribution **(P.O.T.I.)**
Probability the same for each trial, Outcome is either Success or Failure, Trials are fixed, Independent Trials.
This is a binomial situation. There are _____ independent trials of _____. The probability of a success is _____& a failure is _____. The probability of _____ is the same for every trial.

Chapter 9 Sampling Distributions

Identifying a sampling distribution
The sampling distribution for _____ has a mean of _____ and a standard deviation of _____.

Chapter 10 Confidence Intervals

Result in context, the R of NASCAR.
I am __?_% confident that the true **mean/proportion/slope** of _____?_____ is between **lower bound** and **upper bound**.

Interpret the meaning of a confidence interval
If many samples were taken of the same size then ___% of those confidence intervals would capture the true **mean/proportion/slope** of _____?_____.

Combining the two sentences above, probably the best way to interpret a confidence interval
I am **90**% confident that the true **mean/proportion/slope** of _____?_____is between **lower bound and upper bound** because I used a method that captures the true **mean/proportion/slope 90** out of every 100 times in repeated sampling.

Chapter 11 Significance Tests, Type I and Type II errors, Power

Type I error
Reject Ho when it's true.
This is when you determine _____ but it was actually _____.

Type II error
Fail to reject Ho when it's false.
This is when you determine _____ but it was actually _____.

Interpreting the power in context
The power in this situation is correctly determining that Ho is false when Ho is actually false.

Example….Ho: It is safe to go surfing today.
Ha: It is not safe to go surfing.
Type I error: This is when you determine that it is unsafe to surf when it actually was safe.
Type II error: This is when you determine that is safe to surf when it was actually not safe.

Power: The power of this situation is correctly determining that it is unsafe to surf when it is actually unsafe to surf.

Hypothesis
H_0: μ = __ vs H_a: μ __ __ , where μ is the true mean _____.

Chapter 12 t-tests and proportion tests.

Hypothesis
H_0: μ = __ vs H_a: μ __ __ , where μ is the true mean _____.
H_0: p = __ vs H_a: p __ __ , where p is the true proportion of _____.

Making Decisions on P-values.

P-value less than α
This p-value of _____ is low enough to reject H_o at the _____% alpha level.

P-value greater than α
This p-value of _____ is too high to reject H_o.

Assumptions/Conditions for inference

S.P.I.N.
Sample, Population 10% rule, Independence, Normal Condition

Sample needs to be random
We can assume that the _____ is a random sample.
There is no reason to believe that the _____ is not a random sample.
We have an independent random sample of _____.

Population must be at least 10 times our sample size.
The population of _____ is more than 10 times our sample size.
We can assume that our sample of _____ is less than 10% of the population of all_____.
Our 10% condition is met since the population of all_____ is well more than 10 times our sample.

Independence, sample must be independent of each other and multiple samples must be independent of each other
Our sample of _____ should be independent of each other.
_____ are independent.
There is no reason to believe that _____ are not independent.

Normal Condition for a mean
Our normal condition is met because our sample of _____is large enough to be considered approximately normal by the CLT.
Because of our large sample of _____ it meets our normal condition by the CLT.
Our box-plot/histogram shows no extreme skew and no outliers, there is no reason to doubt the normality of our sample of _____.

It is given that our sample of _____comes from a normally distributed population.

Normal Condition for a proportion
Our sample size of _____ is large enough to meet our Normal condition, all counts are 10 or more…..show them $n\hat{p} \geq 10 \; and \; n(1-\hat{p}) \geq 10$.

Chapter 14 Ch-Squared Tests

Goodness of Fit Test
H_o: The _____ is/are as expected.
H_a: The _____ is/are not as expected.

Independence/Association
H_o: There is no association between _____ and _____.
H_a: There is an association between _____ and _____.

H_o: _____ and _____ are independent.
H_a: _____ and _____ are dependent.

Homogeneity
H_o: The proportions of _____ are the same.
H_a: The proportions of _____ are different.

H_o: There is no difference between the proportions of _____.
H_a: There is a difference between the proportions of _____.

Normal Condition for Chi-Square

At least 80% of expected counts must be 5 or more.
All expected counts are 5 or more.
100% of the expected counts are 5 or more.
_____% of expected counts are 5 or more.
All expected counts are 5 or more. The smallest expected count is _____.
Only _____% of our expected counts are 5 or more, proceed with caution.

Chapter 15 Inference for Regression

Hypothesis

H_0: $\beta = 0$ vs H_a: β __ 0 , where β is the true slope of the linear relationship between _____and _____.

Normal Condition for Regression

Residuals are normal, residuals are randomly scattered.
Based on our box-plot/histogram the residuals appear to be approximately normal,
There is no reason to doubt normality of the residuals based on the box-plot/histogram.
We will assume that the residuals are approximately normal and scattered about the regression line.
From looking at the residual plot it is clear that the residuals show a random scatter and not a curved pattern.

Conclusions

1-sided < or >
Based on this sample there is evidence to suggest that there is a positive/negative linear relationship between _____ and _____.

There is not enough evidence to suggest that there is a positive/negative linear relationship between _____ and _____.

2-sided ≠
Based on this sample there is evidence to suggest that the slope of the linear relationship between _____ and _____ is different than 0.

There is not enough evidence to suggest that the slope of the linear relationship between _____ and _____ is different than 0.