Warm Up

do #6 and #11

and the others if you have time

DATA ANALYSIS 5

Which of the following is a true statement?

(A) Stemplots are useful both for quantitative and categorical data sets.

(B) Stemplots are equally useful for small and very large data sets.

(C) Stemplots can show symmetry, gaps, clusters, and outliers.

(D) Stems should be skipped only if there is no data value for a particular stem.

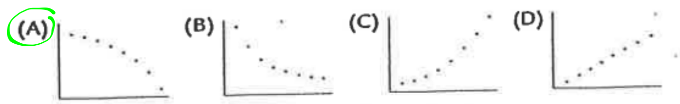
(E) Whether or not to provide a key depends upon the relative importance of the data being displayed.

2	4 7 9 9
3	0 2 2 3 4
4	1 1 2 5
5	2 3

Answer: (C) Stemplots are not used for categorical data sets and are too unwieldy to be used for very large data sets. Stems should never be skipped, even if there is no data value for a particular stem. A key explaining what the stem and leaves represent should always be provided.

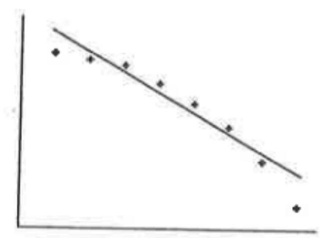
DATA ANALYSIS 6

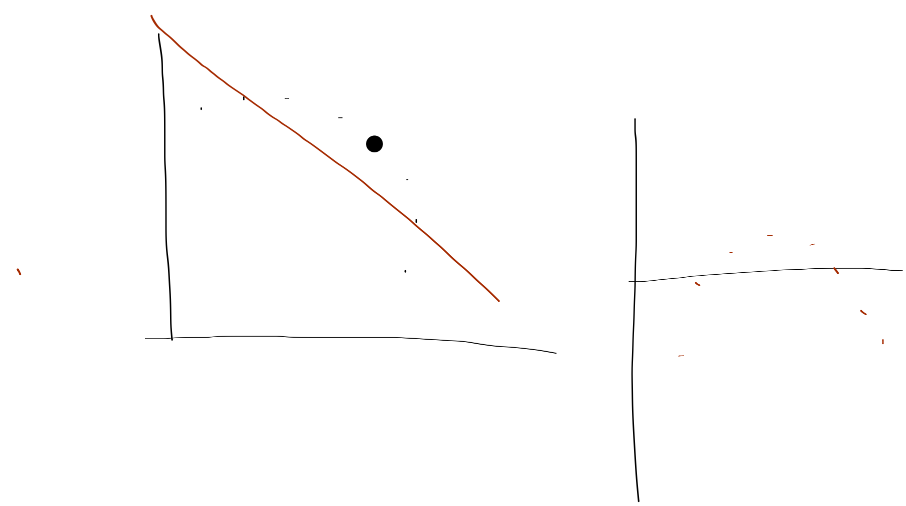
Which of the following scatterplots could have resulted in this *residual* plot? (The *y*-axis scales are not the same in the scatterplots as in the residual plot.)



(E) None of these could result in the given residual plot.

Answer: (A) This is the only scatterplot in which the residuals (actual minus predicted) go from negative to positive and back to negative.





DATA ANALYSIS 7

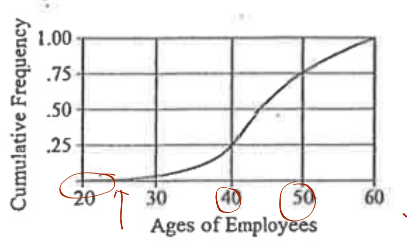
A study is conducted relating AP Statistics exam scores to the total number of study hours for the AP Statistics class put in by students during the academic year, and the correlation is found to be .6. Which of the following is a true statement?

(A) On the average, a 40 percent increase in study time results in a 24 percent increase in exam score.
(B) On the average, a 60 percent increase in study time results in a 100 percent increase in exam score.
(C) Sixty percent of a student's exam score can be explained by the number of study hours.
(D) Sixty percent of the variation in exam scores can be accounted for by this linear regression model.
(E) Higher exam scores tend to be associated with higher numbers of study hours.

Answer: (E) The slope has the same sign, but generally not the same value, as the correlation. The coefficient of determination, giving the proportion of the y-variance that is predictable from a knowledge of x, is equal to r^2, not r. A positive correlation indicates that higher values of x tend to be associated with higher values of y.

DATA ANALYSIS 11

Given this cumulative plot, and using the most commonly accepted definition of outliers, what ages would be considered outliers?



(A) Between 20 and 25
(B) Between 20 and 30
(C) Between 20 and 40
(D) Between 20 and 25, or between 55 and 60
(E) Between 20 and 30, or between 50 and 60

1.5 IQR
Q3 − Q1 10
50 − 40
15

Answer: (A) The interquartile range IQR is $Q_3 - Q_1 = 50 - 40 = 10$. Outliers are values greater than $Q_3 + 1.5(\text{IQR})$ or less than $Q_1 - 1.5(\text{IQR})$. In this case, there are no ages above $50 + 15 = 65$, but there are ages below $40 - 15 = 25$.

Find your 9.1 Day 2 notes from last Friday

The test finds convincing evidence that Ha is true when it really isn't

Type 1 and Type 2 Errors

When we draw conclusions from a significance test we hope our conclusion will be correct. But sometimes it will be wrong. There are two types of errors we can make.

Important ideas:

TYPE I error
rejecting the null when you shouldn't
H_0 is true but we make a wrong decision

occurs by chance
α % of the time

TYPE II error
not rejecting the null when you should
the alternative hypoth (H_a) is true but we make a wrong decision

The test finds convincing evidence that Ha is true when it really isn't

Type 1 and Type 2 Errors

When we draw conclusions from a significance test we hope our conclusion will be correct. But sometimes it will be wrong. There are two types of errors we can make.

Important ideas:

TYPE I error	rejecting the null when you shouldn't. H_0 is true but we make a wrong decision	occurs by chance α % of the time
TYPE II error	not rejecting the null when you should. the alternative hypoth (H_a) is true but we make a wrong decision	

The test does not find convincing evidence that H_a is true when it really is true

Next, pull out your notes from yesterday

9.2 day 1

There are three conditions that must be met for the formula for the standardized test statistic to be valid—one for each of the three components in the formula.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

There are three conditions that must be met for the formula for the standardized test statistic to be valid—one for each of the three components in the formula.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

1. The Random condition

This condition helps ensure that $\hat{p} - p_0$ is a good estimate for the difference between the true value of p and the null value of p_0.

There are three conditions that must be met for the formula for the standardized test statistic to be valid—one for each of the three components in the formula.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

2. The Large Counts condition

This condition allows us to use a Normal distribution to model the sampling distribution of $\hat{p}$. When this condition is met and H_0 is true, the standardized test statistic z has approximately the standard Normal distribution.

There are three conditions that must be met for the formula for the standardized test statistic to be valid—one for each of the three components in the formula.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

3. The 10% condition

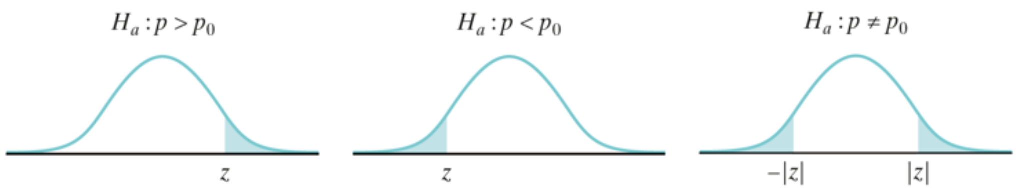
This condition allows us to use the familiar formula for the standard deviation of the sampling distribution of $\hat{p}$ (with p_0 replacing p) when we are sampling without replacement from a finite population.

One-Sample *z* Test for a Proportion

Suppose the conditions are met. To perform a test of H_0: $p = p_0$, compute the standardized test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Find the *P*-value by calculating the probability of getting a *z* statistic this large or larger in the direction specified by the alternative hypothesis H_a using a standard Normal distribution.

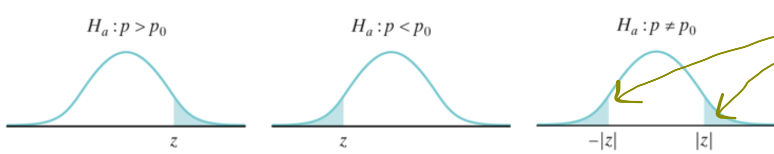


One-Sample *z* Test for a Proportion

Suppose the conditions are met. To perform a test of H_0: $p = p_0$, compute the standardized test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Find the *P*-value by calculating the probability of getting a *z* statistic this large or larger in the direction specified by the alternative hypothesis H_a using a standard Normal distribution.



must **double the** probability to get the P-value

$P(z < -1.3 \text{ OR } z > 1.3)$

$= 2 \cdot P(z > 1.3)$

Significance Test for *p*

Important ideas:

CONDITIONS

Random

10% Sample $< \frac{1}{10}$(pop)

Large Counts

$np_0 \geq 10$

$n(1-p_0) \geq 10$

TEST STATISTIC and P-Value

$N(\mu_{\hat{p}}, \sigma_{\hat{p}})$

$\mu_{\hat{p}} = p_0$

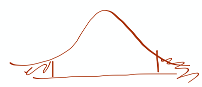
$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$

Standardized Test Statistic $= \frac{\text{Statistic} - \text{Parameter}}{\text{Stand. Deviation of Statistic}}$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Calculate P-Value with Table A or technology

double your probability calculation for a two-sided test when determining your P-Value



Today:

Use the 4-Step Process
For Signicance Tests *For a proportion*

just do the State and Plan steps for now.

But first, let's review the reasons for the conditions once again.

According to the National Institute for Occupational Safety and Health, job stress poses a major threat to the health of workers. A news report claims that 75% of restaurant employees feel that work stress has a negative impact on their personal lives. Managers of a large restaurant chain wonder whether this claim is valid for their employees. A random sample of 100 employees finds that 68 answer “Yes” when asked, “Does work stress have a negative impact on your personal life?” $\alpha = 0.10$

State

[Define parameters, Show evidence for H_a (show the statistic), Hypotheses, and significance level, α]

State

[Define parameters, Show evidence for H_a (show the statistic), Hypotheses, and significance level, α]

$p \rightarrow$ true proportion of employees who say yes about stress.

$H_0: p = .75$

$H_a: p \neq .75$

$\hat{p} = \frac{68}{100} = .68$

evidence for H_a

$\alpha = .10$

Plan [Name the procedure and Check all conditions]

Plan [Name the procedure and Check all conditions]

One sample z test for p

$\boxed{\text{Random}}$

"random sample of 100" ✓

$\boxed{10\%}$

Assumed that $100 < \frac{1}{10}$(all employees in chain) ✓

$\boxed{\text{Large Counts}}$

$100(.75) = 75$
$100(.25) = 25$ ≥ 10 ✓

Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, followed by numbers plugged in, then final answer

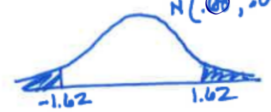
Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, followed by numbers plugged in, then final answer

$\mu_{\hat{p}} = .68$

Standard. test stat = $\dfrac{\text{statistic} - \text{parameter}}{\cancel{\text{SD}}\ \text{Std. error}}$

$N(.68, .043)$



$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$= \frac{0.68 - 0.75}{\sqrt{\frac{.75(.25)}{100}}} = -1.62$$

P-Value:

Table A: .0526

2-tail test

$2(.0526) = .1052$

P-Value

$= 2 \cdot P(z < -1.62)$

$= .1052$

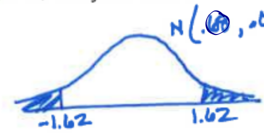
Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, followed by numbers plugged in, then final answer

$\mu_{\hat{p}} = .68$

Standard. test stat = $\dfrac{\text{statistic} - \text{parameter}}{\cancel{\text{SD}}\ \text{Std. error}}$

$N(.68, .043)$



$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$= \frac{0.68 - 0.75}{\sqrt{\frac{.75(.25)}{100}}} = -1.62$$

P-Value:

Table A: .0526

2-tail test

$2(.0526) = .1052$

$P(z < -1.62) = .0526$

$P(z < -1.62 \text{ or } z > 1.62)$

$= 2(.0526) = .1052$

Conclude [Make a conclusion about the hypothesis in the context of the problem, two-sentence structure]

Conclude [Make a conclusion about the hypothesis in the context of the problem, two-sentence structure]

→ Since $.1052 > \alpha = 0.1$, we fail to reject H_0

→ There is not convincing evidence that the true proportion of employees who say yes about stress differs from $p = 0.75$

AP® Exam Tip

When a significance test leads to a **fail to reject H_0 decision**, as in the preceding example, be sure to interpret the results as "We don't have convincing evidence for H_a."

Saying anything that sounds like you believe H_0 is (or might be) true will lead to a loss of credit.

For instance,
it would be *wrong* to conclude, "There is convincing evidence that the true proportion of employees who answer yes is 0.75."

And.......

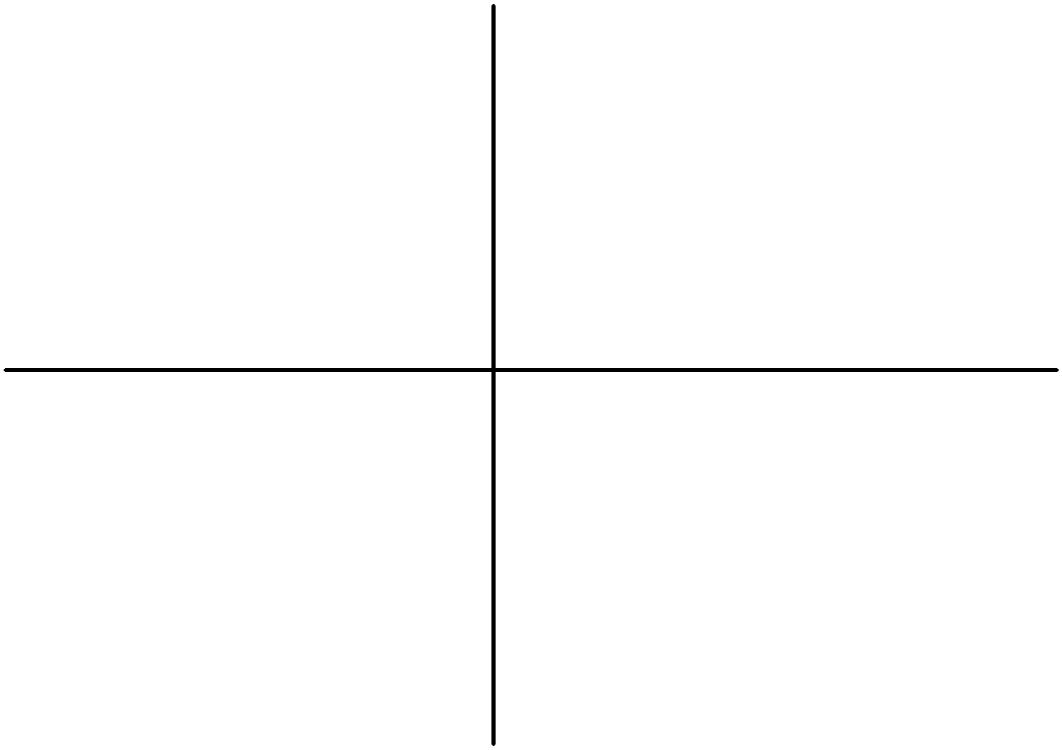
don't write responses as text messages, like "FTR the H_0."

fail to reject

Next example

but this time . . . . . .

You will have revolving partners, meaning a different partner for each of the 4 Steps.



After the "STATE" step, you will rotate to a new group to do the "PLAN". In your new group, compare your work from the previous step to make sure there is agreement... .and then work on the **PLAN** step (but no more than that).

A cw   C ccw

Each person brings their own paper to the next group for the next step.

You will rotate again for the "DO" step, remembering to check the previous step first.

Etc.

Only work on the current section!

Example 2

Does California watch more Netflix? A recent report states that 55% of U.S. adults use Netflix to stream shows and movies. An advertising company believes the proportion of California residents who use Netflix is greater than the national proportion, because Netflix headquarters is located in Los Gatos, California. The company selects a random sample of 600 adults from California and finds that 360 of them use Netflix. Is there convincing evidence at the $\alpha = 0.05$ level that more than 55% of California residents use Netflix?

State

[Define parameters, Show evidence for H_a (show the statistic), Hypotheses, and significance level, α]

State

[Define parameters, Show evidence for H_a (show the statistic), Hypotheses, and significance level, α]

$p \rightarrow$ true proportion of all Calif. residents who use Netflix

$H_0: p = 0.55$ $\hat{p} = \frac{360}{600} = 0.6$ $\alpha = 0.05$

$H_a: p > 0.55$

When making a conclusion in a significance test, be sure that you are describing the parameter
(the proportion of residents in California),

and not the statistic
(the proportion in the sample)

Plan [Name the procedure and Check all conditions]

Plan [Name the procedure and Check all conditions]

One sample z test for p

$\boxed{\text{Random}}$

"random sample of 600 residents" ✓

$\boxed{10\%}$

$600 < \frac{1}{10}$(all residents in Calif.) ✓

$\boxed{\text{Large Counts}}$

$600(.55) = 330 \geq 10$ ✓

$600(.45) = 270 \geq 10$ ✓

Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, followed by numbers plugged in, then final answer

Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, followed by numbers plugged in, then final answer

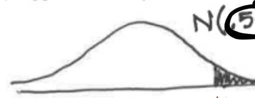
$\hat{p} = 0.6$

$$\text{std. test stat} = \frac{\text{statistic} - \text{parameter}}{\cancel{\text{SD}}\ \text{Std error}}$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$= \frac{0.6 - 0.55}{\sqrt{\frac{.55(.45)}{600}}} = 2.46$$

$N(.55, 0.0203)$



P-value

TABLE A $= .9931$

$P(Z \geq 2.46) = 1 - .9931 = 0.0069$

Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, followed by numbers plugged in, then final answer

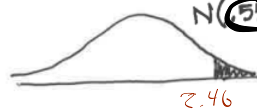
$\hat{p} = 0.6$

$$\text{std. stat} = \frac{\text{statistic} - \text{parameter}}{\cancel{SD}\ \text{Std error}}$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$= \frac{0.6 - 0.55}{\sqrt{\frac{.55(.45)}{600}}} = 2.46$$

$N(.55, 0.0203)$



P-value

TABLE A $= .9931$

$P(Z \geq 2.46) = 1 - .9931 = 0.0069$

P-Value

$P(Z > 2.46) = .0069$

Conclude [Make a conclusion about the hypothesis in the context of the problem, two-sentence structure]

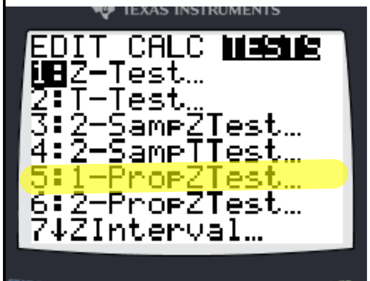
Conclude [Make a conclusion about the hypothesis in the context of the problem, two-sentence structure]

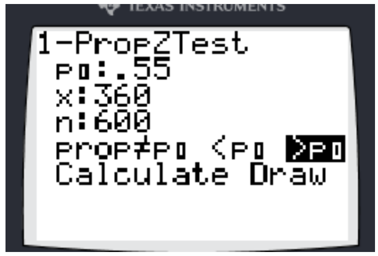
Because the P-Value $0.0069 < \alpha = 0.05$, we reject H_0

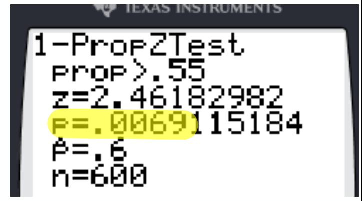
$\therefore$ There is convincing evidence that the true proportion of all California residents who use Netfix is greater than 0.55

1-PropZTest

STAT







AP® Exam Tip

You can use your calculator to carry out the mechanics of a significance test on the AP® Statistics exam. But there's a risk involved. **If you give just the calculator answer with no work, and one or more of your values are incorrect, you will probably get no credit for the "Do" step.** If you opt for the calculator-only method, be sure to name the procedure (one-sample *z* test for a proportion) and to report the standardized test statistic ($z = 1.15$) and *P*-value (0.1243).

For now, I expect you to use the calculator program only to:

- check your answers on the "DO" step
- or to help on M/C questions

See your Unit 3 PPC -FRQ

FOR the PPC

I recommend you re-read the question and look at the solutions (both on the yellow pack)

Then look at your scores (and your answer)

You can double check my scoring with the rubric.

9.2.... 43, 45, 47, 51, 53, 55, 59-62

study pp. 572-580

post