~~Warm Up~~

~~3 AP Review Questions~~

DATA ANALYSIS 52

A study of weekly hours of television watched and SAT scores reports a correlation of $r = -1.18$. From this information, we can conclude that

(A) students who watch more TV tend to have lower SAT scores.

(B) the fewer the hours in front of a TV, the higher a student's SAT scores.

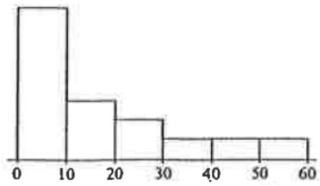
(C) there is little relationship between weekly hours of television watched and SAT scores.

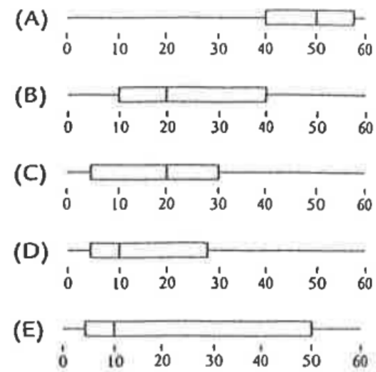
(D) there is strong negative association between weekly hours of television watched and SAT scores, but it would be wrong to conclude causation.

(E) a mistake in arithmetic has been made.

Answer: (E) The correlation r cannot take a value greater than 1 or less than -1.

DATA ANALYSIS 3



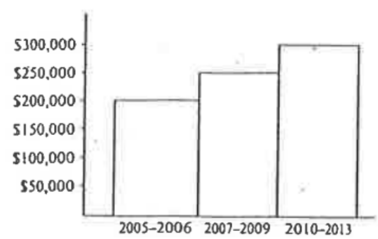


To which of the boxplots can the above histogram correspond?

Answer: (D) The value 10 seems to roughly split the area under the histogram in two, so the median is about 10. The area between 10 and 60 is split in two somewhere between 20 and 30, so Q_3 is between 20 and 30. Finally, boxplot D seems to best pick up the strong right skew.

DATA ANALYSIS 51

Consider the following total sales picture:



Which of the following is a true statement?

(A) Each year since 2005 the total sales has increased.

(B) The average sale has increased during each of the three given time periods.

(C) It is possible that the total sales per year decreased every year between 2005 and 2013.

(D) This picture may be misleading, but it is still a histogram.

(E) To make sales projections, a boxplot would be more informative for this data.

Answer: (C) Labeling the horizontal axis with different year spans results in a misleading picture. Taking into account the number of years represented by each class, the actual total sales per year could be decreasing $\frac{200,000}{2} = 100,000$, $\frac{250,000}{3} \approx 83,333$, and $\frac{300,000}{4} = 75,000$.
This picture is really a bar chart; histograms show relative frequencies through relative areas, and this picture doesn't. A boxplot of the yearly total sales amounts would give no indication of a trend.

9.2
two days

By the end of this section, you should be able to:

✓**STATE and CHECK the Random, 10%, and Large Counts conditions for performing a significance test about a population proportion.**

✓**CALCULATE the standardized test statistic and *P*-value for a test about a population proportion.**

✓**PERFORM a significance test about a population proportion.**

Lesson 9.2: Day 1: Are you sure Mr. Cedarlund isn't a good free throw shooter?

VS

We'll do together

In Lesson 9.1 we used simulation to estimate a P-value to decide whether or not Mr. Cedarlund was exaggerating about his free throw percentage. Today, we will use a formula to find a P-value (somewhat informally)

1. We're going to carry out the significance test from lesson 9.1 again. Here is the hypotheses:

$H_0: p = 0.8$
$H_a: p < 0.8$

2. Suppose Mr. Cedarlund had several sections of AP Stats and each found a different P-Value because each dotplot was different. Would it be appropriate to use a normal distribution to model the sampling distribution of $\hat{p}$?

$\rightarrow$

1. We're going to carry out the significance test from lesson 9.1 again. Here is the hypotheses:

$H_0: p = 0.8$
$H_a: p < 0.8$

2. Suppose Mr. Cedarlund had several sections of AP Stats and each found a different P-Value because each dotplot was different. Would it be appropriate to use a normal distribution to model the sampling distribution of $\hat{p}$?

Yes, if Large counts condition is satisfied.

$np_0 = 50(0.8) = 40 \geq 10$ ✓ $n(1-p_0) = 50(.2) = 10 \geq 10$ ✓

b. Are there any other conditions we should check?

Random

10%

3. **Large Counts Condition – So What?**
We check the Large Counts condition…

So that the sampling distribution of $\hat{p}$ will be approx. Normal so we can use z to estimate the P-value

Random Condition – So What?
We check the random condition….

So we can generalize to the population.

10% Condition – So What?
We check the 10% condition….

so sampling w/o replacement is ok.

4. Now that conditions have been met, find the mean and standard deviation of the sampling distribution of $\hat{p}$.

?

Just watch

A basketball player claims to be an 80% free-throw shooter. You think the player is exaggerating.

H_0: $p = 0.80$
H_a: $p < 0.80$

If H_0 is true (if p = 0.80), then what values of $\hat{p}$ should we expect in a random sample of 50 shots?

Shape: Because the Large Counts condition is met, the sampling distribution of $\hat{p}$ will be approximately Normal.

Center: For random samples, $\mu_{\hat{p}} = p_0 = 0.80$

Variability: $\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.80(1-0.80)}{50}} = 0.0566$

4. Now that conditions have been met, find the mean and standard deviation of the sampling distribution of $\hat{p}$.

$\mu_{\hat{p}} = p_0 = 0.8$

$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{.8(.2)}{50}} \approx 0.0566$

5. Use the mean and standard deviation you found to label the Normal curve.

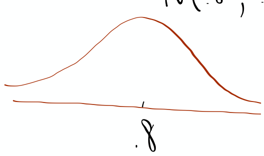
4. Now that conditions have been met, find the mean and standard deviation of the sampling distribution of $\hat{p}$.

$$\mu_{\hat{p}} = p_0 = 0.8$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{.8(.2)}{50}} \approx 0.0566$$

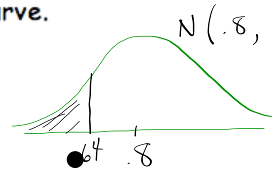
5. Use the mean and standard deviation you found to label the Normal curve.

$N(.8, .0566)$



6. How many standard deviations below the mean (z-score) is $\hat{p} = 0.64$? Label it on the normal curve.

$N(.8, .0566)$



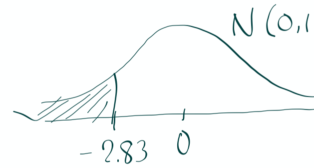
$$Z = \frac{.64 - 0.8}{.0566} = -2.83$$

7. Find the probability of an 80% shooter making 32/50 ($\hat{p} = 0.64$) or less.

$P(Z < -2.83) =$

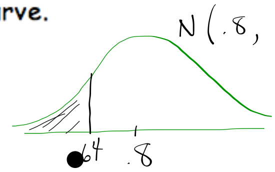
$P(p < .64)$

$N(0,1)$

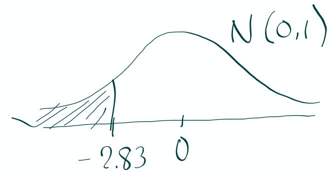


8. What conclusion can we make?

6. How many standard deviations below the mean (z-score) is $\hat{p} = 0.64$? Label it on the normal curve.



$Z = \frac{-0.8}{.0566}$



7. Find the probability of an 80% shooter making 32/50 ($\hat{p} = 0.64$) or less.

$P(Z \le -2.83) =$

normalcdf[-1000 Lower, -2.83 Upper, 0 mean, 1 SD]

= 0.0023

8. What conclusion can we make?

---

making 32/50 ($\hat{p} = 0.64$) or less.

$P(Z \le -2.83) = 0.0023$

8. What conclusion can we make?

normalcdf[-1000 Lower, -2.83 Upper, 0 mean, 1 SD]

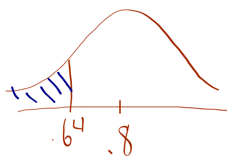
= 0.0023

informally

reject Ho because P-Value < .05

- We have ~~convcing~~ convincing evidence that Mr. is less than a 80% F.T shooter

NOTE: You now have the green light for the shortcut on isolated normal distribution calculations



P(p < .64) = normalcdf[-100, .64, 0.8, .0566]

Lower, Upper, mean, SD

but

for Significance Tests
for now

- you'll be required to show the test statistic formula and values

You may, or may not, have noticed that this is also a binomial situation which could also be calculated binomial probability

binomcdf (n , p, x)

binomcdf (50, 0.8, 32) ← # of successes

≈ 0.0037

Significance Test for *p*

Important ideas:

CONDITIONS

Random

10%

Large Counts

Significance Test for *p*

Important ideas:
CONDITIONS

Random

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

Large Counts

$np_0 \geq 10$

$n(1-p_0) \geq 10$

Significance Test for *p*

Important ideas:
CONDITIONS

Random

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

Large Counts

$np_0 \geq 10$

$n(1-p_0) \geq 10$

TEST STATISTIC and P-Value

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

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

$\sigma_{\hat{p}} =$

Significance Test for *p*

Important ideas:

CONDITIONS

Random

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

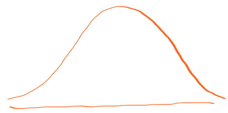
Large Counts

$n p_0 \geq 10$

$n(1-p_0) \geq 10$

TEST STATISTIC and P-Value

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



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

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

Significance Test for *p*

Important ideas:

CONDITIONS

Random

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

Large Counts

$n p_0 \geq 10$

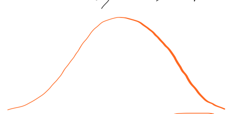
$n(1-p_0) \geq 10$

TEST STATISTIC and P-Value

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

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

Std error of Statistic



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

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

Significance Test for *p*

Important ideas:

CONDITIONS

Random

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

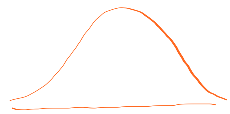
Large Counts

$n p_0 \geq 10$

$n(1-p_0) \geq 10$

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

Conditions for performing a significance test are essentially the same as for Confidence Intervals.

For inference about a proportion, the only difference occurs with the Large Counts condition.

Confidence Intervals	Significance Tests
$n\hat{p} \geq 10$	$n p_0 \geq 10$
$n(1-\hat{p}) \geq 10$	$n(1-p_0) \geq 10$
(don't know p)	Assume hypothesize value is correct

Check Your Understanding

According to the U.S. Census Bureau, the proportion of students in high school who have a part-time job is 0.25. An administrator at a local high school (pop 2500) suspects that the proportion of students at her school who have a part-time job is less than the national figure. She would like to carry out a test at the α = 0.05 significance level. The administrator selects a random sample of 200 students from the school and finds that 39 of them have a part-time job.

(a) State appropriate hypotheses for performing a significance test. Be sure to define the parameter of interest.

$H_0: p = 0.25$

$H_a: p < 0.25$

p = true proportion of students who have a part time job.

(b) Explain why the sample result gives some evidence for the alternative hypothesis.

$\frac{39}{200} = 0.195$ which is less than 0.25

(c) Check if the conditions for performing the significance test are met.

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

$n p_0 = 200(.25) = 50 \geq 10$ ✓

$n(1-p_0) = 200(.75) = 150 \geq 10$ ✓

$\boxed{\text{Random}}$ rand. sample of 200 students ✓

$\boxed{10\%}$

$\frac{1}{10}(2500) = 250$

$200 < 250$ ✓

(d) Calculate the standardized test statistic and P-value.

$\mu_{\hat{p}} = 0.25$

$\sigma_{\hat{p}} = \sqrt{\frac{.25(.75)}{200}}$

$= 0.0306$

(e) What conclusion would you make?

(d) Calculate the standardized test statistic and P-value.

$\mu_{\hat{p}} = 0.25$

$\sigma_{\hat{p}} = \sqrt{\frac{.25(.75)}{200}}$

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

$= 0.0306$

(e) What conclusion would you make?

(d) Calculate the standardized test statistic and P-value.

$\mu_{\hat{p}} = 0.25$

$\sigma_{\hat{p}} = \sqrt{\frac{.25(.75)}{200}}$

$= 0.0306$

$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.195 - .25}{.0306} = -1.797 \approx -1.80$

-1.61

P-Value

Table A

$1 - .9641$

$P(z \leq -1.80) = .036$

$.0361$

(e) What conclusion would you make?

Because the P-Value of $0.036 < \alpha = 0.05$

We reject H_0.

We have convincing evidence ......

See your Ch 8 Test

and

PPC - Unit 3 - FRQ

(a) Construct and interpret a 90% confidence interval for the true mean weight of cereal in the boxes filled by this machine. Show details as done in class.

State:

Parameter: μ = the true mean weight of cereal in the boxes filled by this machine

confidence level: 90%

4

Plan

one sample t-interval for μ

random: SRS of 30 boxes ~~produced~~ by the machine ✓ E

10%: $30 < \frac{1}{10}$(all boxes produced by this machine) ✓

normal: $n \geq 30$ = $30 \geq 30$ ✓

Do

Point est. ± margin of error

$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$ $t^* = 1.699$ df = 29

invT (area: 0.05, df: 29)

$17.92 \pm 1.699 \frac{0.2}{\sqrt{30}}$

E

17.92 ± 0.062 = (,)

= 17.858

= 17.982

Conclude

We are 90% confident that the interval 17.858 ounces to 17.982 ounces captures the true mean weight of cereal in the boxes filled by this machine. E

(a) Construct and interpret a 90% confidence interval for the true mean weight of cereal in the boxes filled by this machine. Show details as done in class.

State: parameter: true mean μ weight of cereal in boxes filled by this machine.

confidence level: 90% confidence

Plan: one sample t interval for μ.

conditions:

Careful

☑ Random: "SRS of 30 boxes" E

☑ 10% n: 30 ≤ 10% of all cereal boxes

☑ Normal: $n \geq 30$ CLT

DO: General formula for CI_*: CI_*: PE ± MOE

Specific formula for CI_*: $CI_* = PE \pm t^* \cdot \frac{s_x}{\sqrt{n}}$ $t^* = 1.699$ area: 0.05 df: 29

$CI_* : 17.92 \pm 1.699 \cdot \frac{0.2}{\sqrt{30}}$

E

$CI_* = 17.92 \pm 0.062$

$CI_* = (17.858, 17.982)$

4

conclude: we are 90% confident that the interval from 17.858oz to 17.982oz captures ~~the~~ true mean weight μ of the cereal ~~in the~~ boxes filled by this machine.

You get good
at what you practice

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.

LCQ 9.1

9.2.... 35 - 41 (odd)

study pp.568-572