Statistics

Lecture 30

Feb 19-8:47 AM

(SG 31)

12 randomly selected students from ELAC had a mean age of 28.5 yrs with standard deviation of 7.5 yrs. ELAC $n=12$ $\bar{x}=28.5$ $s=7.5$

15 randomly selected students from UCLA had a mean age of 36.2 yrs with standard deviation of 10.5 yrs. UCLA $n=15$ $\bar{x}=36.2$ $\boxed{s=10.5}$

Use $\boxed{\alpha=.02}$ to test the claim that there is a difference between two pop. standard deviations. $\sigma_1 \neq \sigma_2$

$H_0: \sigma_1 = \sigma_2$

$H_1: \sigma_1 \neq \sigma_2$ claim, TTT

UCLA → Sample 1 | Sample 2 ← ELAC

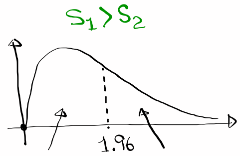
Sample 1	Sample 2
$n_1=15$	$n_2=12$
$s_1=10.5$	$s_2=7.5$

$s_1 > s_2$

NDF $= n_1 - 1 = 14$

DDF $= n_2 - 1 = 11$

CTS $F = \frac{s_1^2}{s_2^2} = \frac{10.5^2}{7.5^2} = 1.96$ ✓



P-value = 2 · smaller area

$= 2 \cdot (.134) = \boxed{.268}$ ✓

Fcdf(0, 1.96, 14, 11) $= \boxed{.866}$

Fcdf(1.96, E99, 14, 11) $= \boxed{.134}$

Total Area = 1

P-value > α

.268 .02

H_0 valid

H_1 invalid

Invalid claim

Reject the claim

Dec 13-7:20 AM

Sample 1	Sample 2
$n_1 = 15$	$n_2 = 12$
$S_1 = 10.5$	$S_2 = 7.5$

$S_1 > S_2$

Now let's use 2-SampFTest

$\boxed{\text{STAT}}$ → TESTS ↓

$\boxed{\text{2-SampFTest}}$

Inpt: $\boxed{\text{Stats}}$

$S_1 = 10.5$

$n_1 = 15$

$S_2 = 7.5$

$n_2 = 12$

$\sigma_1 \neq \sigma_2$ TTT

$\boxed{\text{Calculate}}$

CTS $F = 1.96$ ✓

P-value $P = .267$ ✓

P-value $> \alpha$

If we choose α to be .27, .28, .29, .30, ...

then $\boxed{\text{P-value} \leq \alpha}$ ⇒ H_0 invalid

$.267 \leq \alpha$

H_1 valid → Valid claim

FTR the claim

Dec 13-7:36 AM

SG 35

I randomly selected cars on different freeways, chart below shows speed of those cars

FWY 10			FWY 101			FWY 405		
72	75	80	70	65	68	60	65	70
68	70	78	75	82	78	55	58	48
	75			75			60	

No α → use .05

Test the claim that $\boxed{\text{all pop. means}}$ are $\boxed{\text{equal}}$

H_0: $\mu_1 = \mu_2 = \mu_3$ claim

H_1: At least one mean is different. $\boxed{\text{RTT}}$

Since we are comparing at least 3 pop. means, we use ANOVA.

FWY 10 → L1

FWY 101 → L2

FWY 405 → L3

$K = 3$

$n = 7 + 7 + 7 = 21$ ⇒ $\boxed{\begin{array}{l} NDF = K - 1 = 2 \\ DDF = n - K = 18 \end{array}}$

$\boxed{\text{STAT}}$ → TESTS ↑ $\boxed{\text{ANOVA(}}$ L1, L2, L3 $\boxed{\text{Enter}}$

CTS $F = 13.776$ ✓

P-value $P = 2.349 \times 10^{-4}$ ✓

P-value $\leq \alpha$

2.349×10^{-4} $.05$

H_0 invalid

H_1 valid

Reject the claim ← Invalid claim

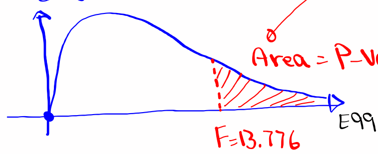
Dec 13-7:43 AM

Suppose You are Comparing 3 pop. means with total Sample Size of 21 with $\boxed{\text{CTS } F = 13.776}$

→ ANOVA
→ RTT

Find the Corresponding P-Value.

$K = 3$
$n = 21$
→ $NdF = K - 1 = 2$
$DdF = n - K = 18$



P_value = Fcdf(13.776, E99, 2, 18)

$= \boxed{2.349 \times 10^{-4}}$ ✓

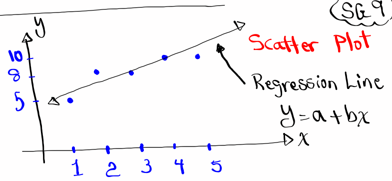
Dec 13-7:58 AM

Testing Linear Correlation Coef.

SG 33

SG 9

x	y
1	5
2	8
3	8
4	10
5	10



x→L1
y→L2

$\boxed{\text{STAT}}$ → CALC ↓ $\boxed{\text{LinReg}(a+bx)}$

x list : L1
Y list : L2

Clear

$a = 4.6$
$b = 1.2$
⇒ $\boxed{y = 4.6 + 1.2x}$

$r^2 = .857$
$r = .926$

Linear Correlation Coef.

Calculate → Coef. of determination

$r^2 \approx 86\%$

86% of all Y-Values are explained by x-Values

If r is close to 1 or -1, Linear Correlation is Significant.

If r is close to 0, Linear Correlation is not Significant.

$r = .926$
is close to 1 ⇒ Linear Correlation appears to be Significant.

Dec 13-8:04 AM

$H_0: \rho = 0$ not Significant

rho

$H_1: \rho \neq 0$ Is Significant TTT

STAT → TESTS ↓ LinRegTTest

x list: L1

Y list: L2

Freq: 1

$\rho \neq 0$ TTT

Reg EQ: Clear

Calculate

CTS $t = 4.243$

P_Value $P = .024$

$df = 3$

↓ $a =$

↓ $b =$

↓ $r^2 =$

↓ $r =$

P_Value $< \alpha$ — H_0 invalid, H_1 Valid

$.024 < .05$

Linear Correlation is Significant.

Suppose $\alpha = .05$

If we choose $\alpha = .01$

P_Value $> \alpha$, — H_0 Valid → Linear Correlation is not Significant.

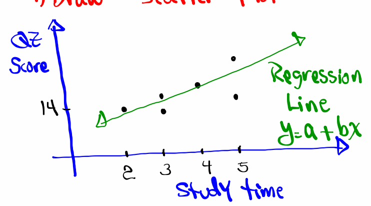
$.024 > .01$ — H_1 invalid

Dec 13-8:15 AM

Consider the chart below

Study time	QZ Score
2	14
3	15
3	14
4	16
5	15
5	18

1) Draw Scatter Plot



Study time → x → L1

QZ Score → y → L2

Lin Reg (a+bx)

$a = 12$

$b = .909$ ⇒ $\boxed{y \approx 12 + .9x}$

$r^2 = .535$ → $r^2 \approx 53.5\%$

$r = .731$

53.5% of QZ Scores are explained by study time.

r is close to 1, Linear Correlation appears to be Significant.

Dec 13-8:45 AM

Use $\alpha = .1$ to test the claim that linear correlation is significant.

$H_0: \rho = 0$ Not Significant

$H_1: \rho \neq 0$ is Significant TTT claim

Rho

STAT TESTS LinRegTTest

Xlist: L1
Ylist: L2
Freq: 1
$\rho \neq 0$
RegEQ: Clear
Calculate

$\boxed{\text{CTS } t = 2.144}$
$\boxed{\text{P-Value } P = .099}$ ✓

P-value $\leq \alpha$
$.099 \leq .1$

H_0 invalid

H_1 valid

Valid claim

Linear Correlation is significant.

FTR the claim

If we choose $\alpha = \boxed{.08, .07, .06, .05, .04, \ldots}$

then P-value $\geq \alpha \rightarrow H_0$ valid, H_1 invalid

.099

Reject the claim

Linear Correlation is not significant

Dec 13-8:52 AM

CTS Formula for t

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

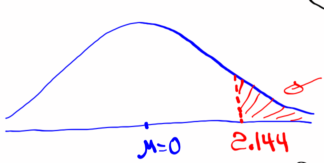
From last example
$n = 6$, $r = .731$, $r^2 = .535$

$$t = .731 \cdot \sqrt{\frac{6-2}{1-.535}}$$

$$= .731 \cdot \sqrt{\frac{4}{.465}} \approx \boxed{2.144}$$

P-value → t-dist, TTT, $df = n-2$

$df = 6-2 = 4$



Area ·2 = P-Value

σ unknown

df=4

P-value $= 2 \cdot \text{tcdf}(2.144, E99, 4)$

$= \boxed{.099}$

Dec 13-9:01 AM

Given $\boxed{n=10}$, $r=.6$, $\boxed{\alpha=.02}$

1) find CTS t

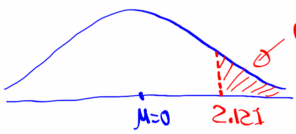
$$t=r\cdot\sqrt{\frac{n-2}{1-r^2}}$$

$$=.6\cdot\sqrt{\frac{10-2}{1-.6^2}}$$

$$=.6\cdot\sqrt{\frac{8}{.64}}=\boxed{2.121}$$

2) find P-value for TTT

Area $\cdot 2$ = P-value



σ unknown

$df=n-2=8$

P-value =

$2\cdot tcdf(2.121, E99, 8)$

$=\boxed{.067}$

P-value > α H_0 valid → Linear Correlation

.067   .02 H_1 Invalid   is not Significant

If we choose $\alpha=.07, .08, .09, .1, \ldots$

P-value $\leq\alpha$ → H_0 invalid

H_1 Valid → Linear Correlation is Significant.

Dec 13-9:08 AM

How to make Predictions:

If r is Significant ⇒ Use regression line

H_1 is valid   Plug in x to find y

If r is not Significant ⇒ Use $\bar{y}$

H_0 is Valid

Dec 13-9:16 AM

Given $\boxed{\hat{y} = 12 + 5x}$, $\bar{y} = 18$, $\boxed{r = .8}$, $\boxed{n = 10}$

$\boxed{\alpha = .01}$

Predict y when $x = 2$.

If r is Significant $\Rightarrow \hat{y} = 12 + 5(2)$
$= 12 + 10 = \boxed{22}$

If r is not Significant $\Rightarrow \hat{y} = \bar{y} = \boxed{18}$

Test r:

$H_0: \rho = 0$ Not Significant

$H_1: \rho \neq 0$ Is Significant

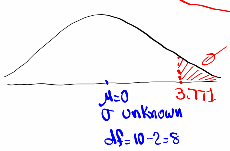
CTS t

$t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$

$= .8 \cdot \sqrt{\dfrac{10-2}{1-.8^2}}$

$= .8 \cdot \sqrt{\dfrac{8}{.36}} = \boxed{3.771}$

Find P-value TTT $df = n-2$



Area $\cdot 2$ = P-value

P-value $= 2 \cdot \text{tcdf}(3.771, E99, 8)$

$= \boxed{.005}$

P-value $< \alpha$ → H_0 invalid
.005 .01 → H_1 Valid → Linear Correlation is Significant

use regression line
$\hat{y} = 12 + 5x$

Prediction Value

for $x = 2$ → $\hat{y} = 12 + 5(2) = 12 + 10 = \boxed{22}$

SG 33 ✓

Dec 13-9:19 AM