Statistics

Lecture 15

Feb 19-8:47 AM

Among 275 females, 40% of them were in support of . . . . .

$n = 275$, $\hat{p} = .4$ → $x = n\hat{p} = 110$

Among 125 males, 36% of them were in support of . . . . .

$n = 125$, $\hat{p} = .36$ → $x = n\hat{p} = 45$

Females	Males
$x_1 = 110$	$x_2 = 45$
$n_1 = 275$	$n_2 = 125$

1) Pooled Prop.

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{155}{400} = .3875 \approx .39$$

2) Find 98% Conf. int. for the diff. of P_1 & P_2.

2-Prop ZInt

$$-.08 < P_1 - P_2 < .16$$

$$E = \frac{.16 - (-.08)}{2} = .12$$

Jun 5-8:07 AM

No α → .05

3) ~~Test the claim~~ that there is no diff. between P_1 & P_2.

$H_0: P_1 = P_2$ claim

$H_1: P_1 \neq P_2$ TTT

2-Prop Z Test ← Traditional

CTS $Z = .761$

P-Value $P = .447$ ← P-value Method

P-Value > α

.447 .05

H_0 valid H_1 invalid

Valid claim → FTR the claim

Jun 5-8:15 AM

20 female students had a mean age of 28.5 Yrs with Standard dev. of 6.5 Yrs.

15 male students had a mean age of 24.5 Yrs with Standard dev. of 8.2 Yrs.

Females	Males
$\bar{x}_1 = 28.5$	$\bar{x}_2 = 24.5$
$S_1 = 6.5$	$S_2 = 8.2$
$n_1 = 20$	$n_2 = 15$

1) Discuss Pooling options and df.

If we assume $\sigma_1 = \sigma_2$

Pooled: Yes

$df = n_1 + n_2 - 2 = 33$

If we assume $\sigma_1 \neq \sigma_2$

Pooled: No

df = Smaller $n - 1 = 14$

Jun 5-8:20 AM

~~NO C-level ≠ .95~~

Find Conf. int. for the difference of two Pop. means assuming $\sigma_1 = \sigma_2$.

$-1.1 < \mu_1 - \mu_2 < 9.1$

Pooled: Yes

df = 33

$E = \frac{9.1 - (-1.1)}{2} = 5.1$

σ_1 & σ_2 unknown

2-SampTInt

Jun 5-8:25 AM

Use $\alpha = .1$ to test the claim that mean age of all females is greater than the mean age of all males. Assume $\sigma_1 \neq \sigma_2$

$\mu_F > \mu_M$

Group 1 Group 2

$H_0: \mu_1 \le \mu_2$

$H_1: \mu_1 > \mu_2$ claim, RTT

Pooled: NO

df = 14

CTS $t = 1.558$

P-value $P = .066$

σ_1 & σ_2 unknown

2-SampTTest

P-value $\le \alpha$

.066 .1

H_0 invalid, H_1 valid

Valid claim → FTR

If we choose α to be .06, .05, .04, .03, ...

P-Value > α H_0 valid, H_1 invalid → Invalid claim

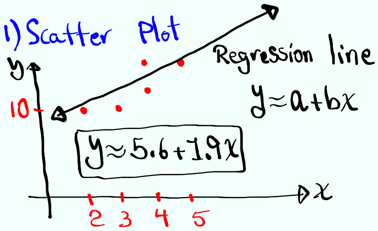
Reject

Jun 5-8:30 AM

Consider the chart below:

x	y
3	10
4	12
4	15
2	10
5	15

1) Scatter Plot



$x \to L1$, $y \to L2$, Use LinReg(a+bx)

$a \approx 5.615 \approx 5.6$ $\quad r^2 \approx .733$

$b \approx 1.885 \approx 1.9$ $\quad r \approx .856$

$r^2(\%) \approx 73\%$

Coef. of determination

73% of Y-values are explained by x-values.

r linear Correlation Coef.

It is close to 1, it appears to be Significant.

Predict Y when $x = 3.5$

1) Assume r is significant

$y \approx 5.6 + 1.9(3.5)$

$\approx \boxed{12.25}$

2) Assume r is not Significant

Use $\bar{y} \approx \boxed{12.4}$

Jun 5-8:41 AM

Testing r:

$H_0: \rho = 0$ (r is not Significant)

ρ: rho

$H_1: \rho \neq 0$ (r is significant) TTT

CTS t

P-Value P

→ STAT TESTS **LinReg T Test**

CTS $t = 2.869$

P-Value $P = .064$

$df = n - 2 = 5 - 2 = 3$

xlist: L1

Ylist: L2

Freq: 1

$\rho \neq 0$

Reg EQ: blank

Calculate

P-Value > α

.064 > .05

H_0 valid, H_1 invalid

r is not Significant

If $\alpha = .1$

P-Value $\leq \alpha$

H_0 invalid, H_1 Valid

r is Significant

CTS $t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}} = .856 \cdot \sqrt{\dfrac{5-2}{1-.733}} = .856 \cdot \sqrt{\dfrac{3}{.267}} = \boxed{2.869}$

Jun 5-8:53 AM

Study time	Exam Score
8	90
7	85
7	80
10	95
5	70
6	75

Study time → x → L1

Exam time → y → L2

Use LinReg (a+bx)

$a \approx 45$

$b \approx 5$ ⇒ $\boxed{y \approx 45 + 5x}$

$r^2 = .926 \approx 93\%$

$r = .962$

Predict exam score for 8 hrs of study time.

1) r is significant

$y \approx 45 + 5x$

$\approx 45 + 5(8)$

$\approx \boxed{85}$

2) r is not significant

use $\bar{y} \approx 82.5$

≈ 83

Jun 5-9:25 AM

Use $\boxed{\alpha = .02}$ to determine whether linear correlation is significant or not.

$H_0: \rho = 0$ (Not Significant)

$H_1: \rho \neq 0$ (Is Significant)

LinReg TTest

xlist: L1

Ylist: L2

Freq: 1

$\rho \neq 0$

RegEQ: blank

Calculate

$\boxed{\text{CTS } t = 7.050}$ ← Due to rounding

$\boxed{\text{P-Value } P = .002}$

$df = n - 2 = 6 - 2 = 4$

P-Value ≤ α

.002 ≤ .02

H_0 invalid, H_1 Valid

Linear Correlation is Significant.

CTS Formula

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}} = .962 \cdot \sqrt{\frac{6-2}{1-.926}} = .962 \cdot \sqrt{\frac{4}{.074}}$$

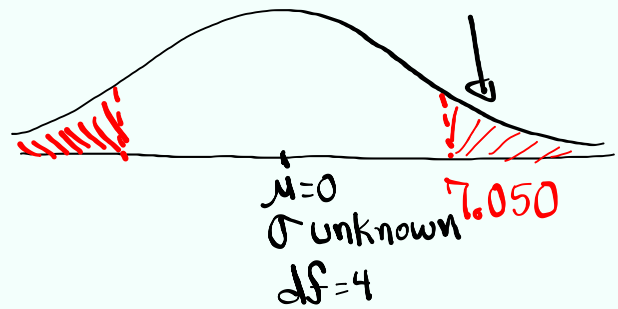
$$\approx \boxed{7.073}$$

Jun 5-9:32 AM

Given CTS $t=7.050$

TTT $df=4$

Find p-Value. → $2 \cdot$ Area



P-Value $= 2 \cdot \text{tcdf}(7.050, E99, 4) \approx \boxed{.002}$

Jun 5-9:42 AM

The linear Correlation Coef. of 10 randomly Selected ordered-Pairs was .6.

Determine whether Linear Correlation Coef is Significant or not.

$n=10$ $r=.6$

No $\alpha \to .05$

$H_0: \rho = 0$ (Not Significant)

$H_1: \rho \neq 0$ (Is Significant)

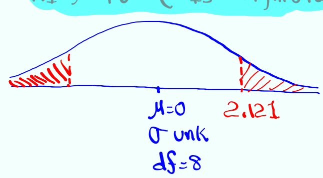
CTS t

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

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

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

$\approx \boxed{2.121}$



P-Value $= 2 \cdot \text{tcdf}(2.121, E99, 8) \approx \boxed{.067}$

P-Value > α

.067 .05

H_0 valid, H_1 invalid

r is not Significant

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

P-Value ≤ α

H_0 invalid, H_1 Valid

r is Significant

SG 31✓

Jun 5-9:46 AM

Comparing at least 3 pop. means. SG 33

$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$

H_1: At least one mean is different. **RTT**

Method: ANOVA (Analysis of Variance)

K → # of groups ⇒ Ndf = K − 1

n → Total Sample Size ⇒ Ddf = n − K

CTS F

P-Value P → STAT → TESTS → ANOVA(

P-Value > α → H_0 Valid, H_1 invalid

P-Value ≤ α → H_0 invalid, H_1 Valid

Jun 5-9:58 AM

I randomly Selected exam Scores from 3 different Schools.

L1 ELAC	L2 Mt. SAC	L3 Citrus
75 88 92	73 89 95	69 100 80
100 80 90	99 82 90	90 75
70 65	75 68	

K = 3 → Ndf = K − 1 = 2

n = 8 + 8 + 5 = 21 → Ddf = n − K = 18

No α → .05

Test the claim that all means are the Same

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

H_1: At least one mean is different. **RTT**

STAT TESTS ANOVA(L1, L2, L3) Enter

CTS F = .029

P-Value P = .971

P-Value > α

H_0 valid, H_1 invalid

Valid claim → **FTR**

Jun 5-10:04 AM

Ages of randomly selected students from 4 different Schools are given below:

L1 ELAC	L2 Mt.SAC	L3 Citrus	L4 Cal Poly Pomona
27 25 30	19 24	19 23	28 42 35
18 28	32 25	27 33	25 20
	28	30	

$K = 4 \rightarrow$ $Ndf = K-1 = 3$

$n = 20$ $Ddf = n-K = 16$

no $\alpha \rightarrow .05$

Test the claim that not all means are the Same.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: At least one mean is different. RTT claim

If a list is missing, STAT Edit 5:SetupEditor Enter

STAT TESTS

ANOVA(L1, L2, L3, L4)

CTS F=.760

P-Value P=.533

P-Value > α

H_0 valid, H_1 invalid

Reject the claim

SG 33√