Hypothesis Tests for Variances and for two means

8.2

Today's topics

Review:

- Hypothesis test steps

New:

- Hypothesis test for variances
- Hypothesis test for two means

Note for 8.1, 8.3

z t σ unknown

In all the examples from Tuesday's lecture, we are assuming that these samples are coming from distributions that are **approximately normal** for these methods to work.

(Otherwise, with small sample sizes, the CLT would not hold) nothing happening

Hypothesis Testing (Steps)

claims

σ

"uninteresting" default

1. Formulate H_0 and H_A (based on the scenario)

$\bar{X}$ s^2

2. Identify a test statistic to use and its distribution under H_0
3. Evaluate the test statistic

$p < \alpha$, Rej H_0

$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim Z$

4. Calculate a p-value, compare to α.
   OR Identify a rejection region
5. Make a decision
   - if $p < \alpha$, reject H_0. $\longleftrightarrow$ Otherwise, (if $p > \alpha$), do not reject H_0.
6. State conclusion **in the context of the original question.**
   - "There is/isn't enough evidence to show that…"

$p = 0.04999 \approx 0.05$

General Form of CI for mean (review)

Estimate $\pm$ (Critical Value * SE of estimate)

$100(1-\alpha)\%$ CI

e.g. if σ is known:

$$\underbrace{\bar{x} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}}$$

if σ is unknown:

$$\underbrace{\bar{x} \pm t_{n-1,\alpha/2} * \frac{s}{\sqrt{n}}}$$

notes

$\alpha = 0.05$
$\alpha = 0.1$

μ_0

$\bar{x} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$

$H_0: \mu = 100$

$H_1: \mu \neq 100 \leftarrow$ 2-sided

test at $\alpha = 0.01$

$\bar{X}$

σ

n

$Z = \underbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}_{\text{value}} \sim Z$

value $\rightarrow$ p-val compare to α

$p < \alpha$, Reject

$100(1-\alpha)\%$ CI

$\alpha = 0.01$ 99% 95% 90%

if $\mu_0 \notin$ interval, Reject H_0 at α

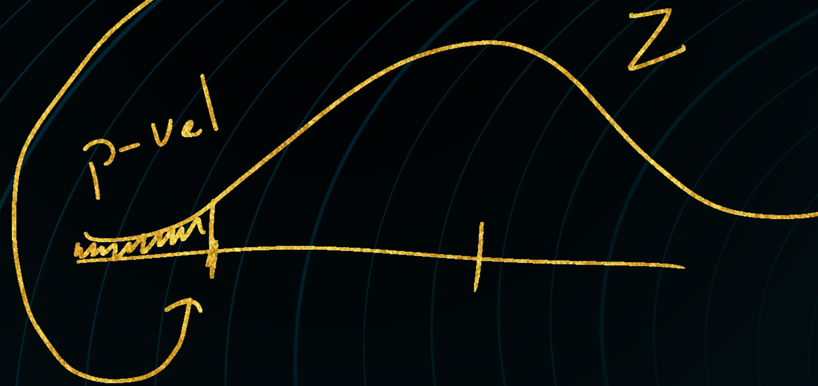
$\boxed{(101.2, 110.5)}$. reject H_0

e.g. (98.5, 103) DNR H_0

$H_A: \mu < 100$

$\bar{X}$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim Z$$



Confidence Interval for σ^2 (review)

Confidence Interval for σ^2 :

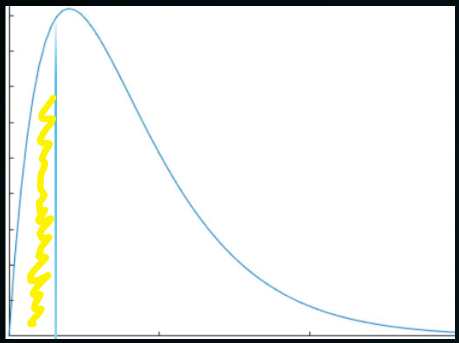
$$\left(\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2}} \right)$$

Hypothesis Tests for Variance (or SDs)

Test statistic:

$$\chi^2 = \boxed{\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}}$$

Rejection Region:



Left-tailed:

$H_0: \sigma^2 \geq \sigma_0^2$ vs $H_1: \sigma^2 < \sigma_0^2$

$H_0: \sigma \geq \sigma_0$ vs $H_1: \sigma < \sigma_0$

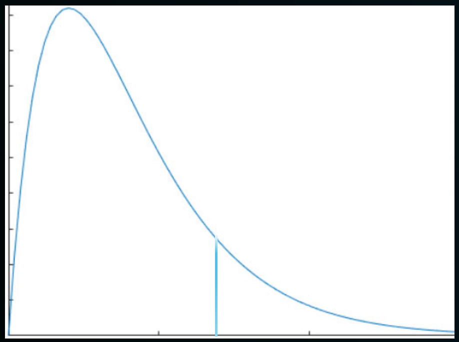
$\sigma = 7$

Rejection Region:

Right-tailed:

$H_0: \sigma^2 \leq \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$

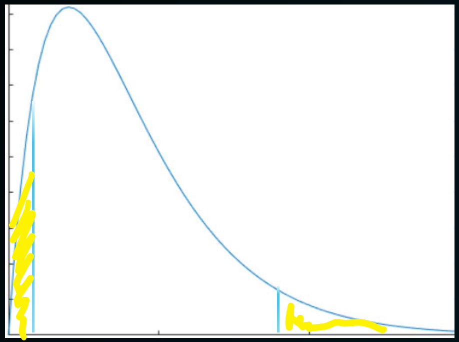
$H_0: \sigma \leq \sigma_0$ vs $H_1: \sigma > \sigma_0$



Two-tailed:

$H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$

$H_0: \sigma = \sigma_0$ vs $H_1: \sigma \neq \sigma_0$



notes

Hypothesis Tests for Variance (or SDs)

Left-tailed:

$H_0: \sigma^2 \geq \sigma_0^2$ vs $H_1: \sigma^2 < \sigma_0^2$

$H_0: \sigma \geq \sigma_0$ vs $H_1: \sigma < \sigma_0$

Right-tailed:

$H_0: \sigma^2 \leq \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$

$H_0: \sigma \leq \sigma_0$ vs $H_1: \sigma > \sigma_0$

Two-tailed:

$H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$

$H_0: \sigma = \sigma_0$ vs $H_1: \sigma \neq \sigma_0$

Example

< left

> right

A sample of 12 random exam scores from the final exam was collected: 78, 81, 82 ,77, 86, 79, 84, 87, 86, 91, 88, 88. (s = 4.48). Last year's final had $\sigma = 7$. Is there enough evidence to suggest that the population standard deviation for this year is lower than 7? Test at alpha = 0.05

$H_0: \sigma = 7$

$H_A: \sigma < 7$

$H_0: \sigma^2 = 49$

$H_A: \sigma^2 < 49$

11

Reject H_0

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

$$= \frac{11 \cdot 4.48^2}{7^2} = 4.5056 \sim \chi^2_{11}$$

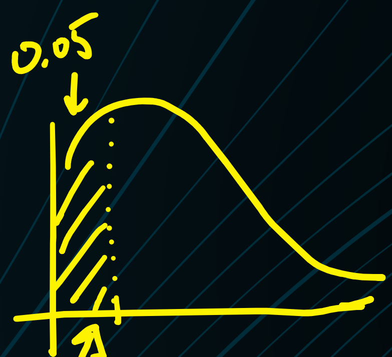
$p = 0.047 < \alpha$

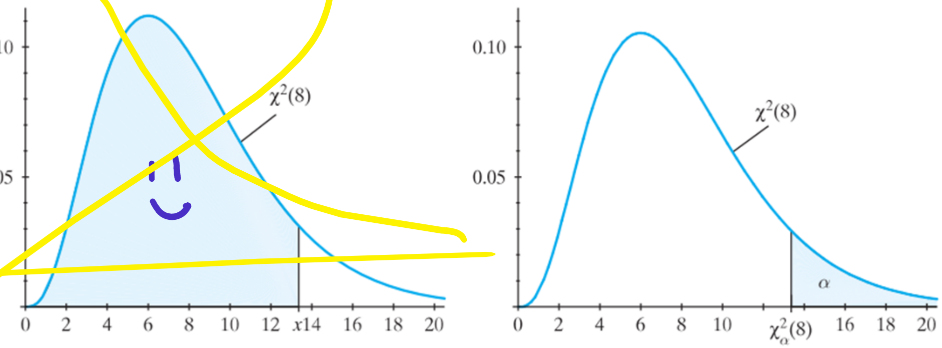
There is significant evidence to suggest that the population SD this year is lower than 7 at alpha =0.05.

$H_A: \sigma < 7$

$\alpha = 0.05$

test. stat $\boxed{4.5056} \sim \chi^2_{11}$





$$P(X \le x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2} dw$$

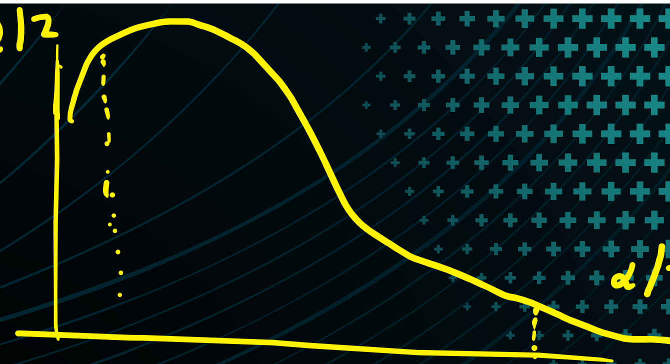
	$P(X \le x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21
11	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72

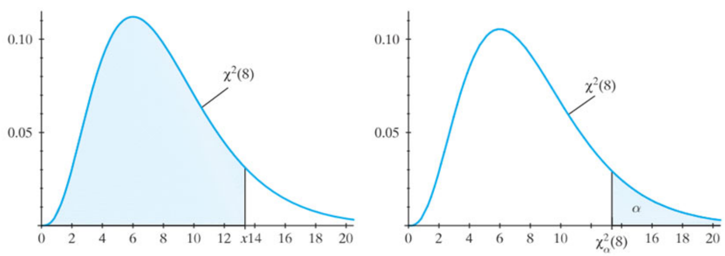
What if it's two sided?

$\alpha = 0.05$

- H_0 $\sigma = 7$

  $H_1 : \sigma \neq 7$



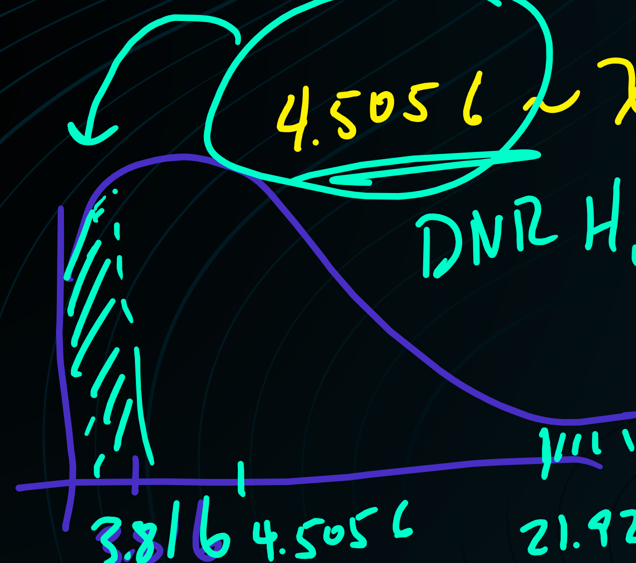


$$P(X \leq x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2} dw$$

	$P(X \leq x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21
11	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72

4.5056 $\sim \chi^2_{11}$

DNR H_0



G1 CS G2 CE

Two Sample tests for means

Say we have two populations: with mean μ_1, variance σ_1^2, and mean μ_2, variance σ_2^2.

Let $X_1, X_2, \ldots, X_{n1}$ and $Y_1, Y_2, \ldots, Y_{n2}$ be two independent samples from each of these populations.

If n_1 and n_2 are large, or these populations are both approximately normal then a confidence interval for $\mu_1 - \mu_2$ is:

$$\boxed{(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$Var[X-Y] = Var[X] + Var[Y]$

Review – Functions of Normal Distributions

Assuming i.i.d:

If each $X_i \sim N(\mu, \sigma^2)$, what is the distribution of $\bar{X}$?

* $\boxed{\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$

n_1 If each X_i has mean μ_1, and variance σ_1^2, $\bar{X} \sim$? $\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$

n_2 If each Y_i has mean μ_2, and variance σ_2^2, $\bar{Y} \sim$? $\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$

$\bar{X} - \bar{Y} \sim$? $N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

General form of CI for mean

CI: Estimate $\pm$ Critical Value * SE(estimate)

$\mu_1 - \mu_2$

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$SE_{\bar{X}-\bar{Y}}$

if σ_1^2 & σ_2^2 unknown

2-sample confidence interval

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

1 sample

df = n - 1

Two approaches for df:

1. Conservative ←

   df = min(n_1, n_2) -1

e.g. $n_1 = 6$ $n_2 = 50$

df = 6 - 1 = 5

2. Welch's T:

default

$$\text{df} = \left\lfloor \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2} \right\rfloor$$

floor function

$\lfloor 7.9 \rfloor \to 7$

$\lceil \ \rceil$ ceiling function

pop 1 σ_x^2 = variance σ_y^2 pop 2 $y_1, y_2, \ldots$

$x_1, x_2 \ldots$

Two sample means – pooled variance

$\bar{x}$ $\bar{y}$

s_x^2 s_y^2

If we can assume that population 1 and population 2 standard deviations are equal, we can use s_{pooled} as an estimator for both.

$$\left(\overline{X}-\overline{Y}\right) \pm t_{\alpha/2} \cdot s_{\text{pooled}} \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

where

$$s_{\text{pooled}}^2 = \frac{(n_1-1)\cdot s_1^2+(n_2-1)\cdot s_2^2}{n_1+n_2-2}$$

σ^2

and df = $n_1 + n_2 - 2$

Two sample means – Matched pairs

Assume that the differences $D_i = X_i - Y_i$ are a random sample from normal distribution with mean δ_D and standard deviation σ_D.

Pair			Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = X_n - Y_n$

A confidence interval for δ_D is

$$\boxed{\bar{D} \pm t_{\alpha/2} \frac{s_D}{\sqrt{n}}. \qquad \text{df} = n-1}$$

The test statistic for testing H_0: $\delta = \delta_0$ is

$$\boxed{T = \frac{\bar{D} - \delta_0}{s_D/\sqrt{n}}}$$

	B	A	Diff
	100	90	

2 populations

Hypothesis test for 2 means

σ known

- Steps
  - Determine which statistic to use t/z
  - Are the data paired? before/after "dependent"
  - Can the variances be pooled?
  - Follow the usual procedures for hypothesis testing:
    - Come up with H_0 and $H_{A,}$
    - Create and evaluate test statistic, identify the null distribution
    - Find p-value (or rejection region)
    - Make decision and conclusion.

Example

n m

n_1 n_2

$n_2 = 16$

Among the data collected for the World Health Organization air quality monitoring project is a measure of suspended particles, in μg/m^3. Let X and Y equal the concentration of suspended particles in μg/m^3 in the city centers of Melbourne and Houston, respectively. Using $\boxed{n = 13}$ observations of X and $\boxed{m = 16}$ observations of Y, we obtain $\bar{x}$=72.9, s_x = 25.6, $\bar{y}$=81.7, s_Y = 28.3. Perform a test to see if Houston has worse air quality at 0.05 α significance and state the result.

$n_1 = 13$

H_0: $\boxed{\mu_X = \mu_Y}$ $\quad \underbrace{\mu_x - \mu_y}{} = 0$

H_1: $\mu_X < \mu_Y$. $\quad \mu_x - \mu_y < 0$

Continued:

"Welch's t"

2-sample t-test

$$\boxed{\frac{est - \mu_0}{SE_{est}}}$$

$\bar{x}$=72.9, $\boxed{s_x = 25.6,}$ $\bar{y}$=81.7, $\boxed{s_Y = 28.3.}$ Perform a test to see if Houston has worse air quality at 0.05 significance and state the result.

$$\frac{\bar{X} - \mu}{s/\sqrt{n}}$$

H_0: $\mu_X = \mu_Y$ $\qquad$ H_1: $\mu_X < \mu_Y$. $\quad \mu_x - \mu_y \leq 0$

$$\boxed{H_0 : \mu_x - \mu_y = 0}$$

$$t = \frac{\bar{x} - \bar{y} - 0}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}} \sim t_{26}$$

$$\sqrt{s^2} = s$$

$\boxed{\text{Welch's}}$

t_{26}

$$t = \frac{72.9 - 81.7}{\sqrt{\frac{25.6^2}{13} + \frac{28.3^2}{16}}} = -.878 \sim t_{26}$$

$p = 0.194 > \alpha$ DNR H_0

σ SD

SD: population parameter

- Usually refers to variable like X

σ_x std. of x

SE: Standard error

- If I have something like X_bar

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

SE of $\bar{X}$

$n_1 - 1$

$n_2 - 1$

$$\left(\overline{X} - \overline{Y}\right) \pm t_{\alpha/2} \cdot s_{\text{pooled}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$s^2_{\text{pooled}} = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}$$

What if we want to pool the variance?

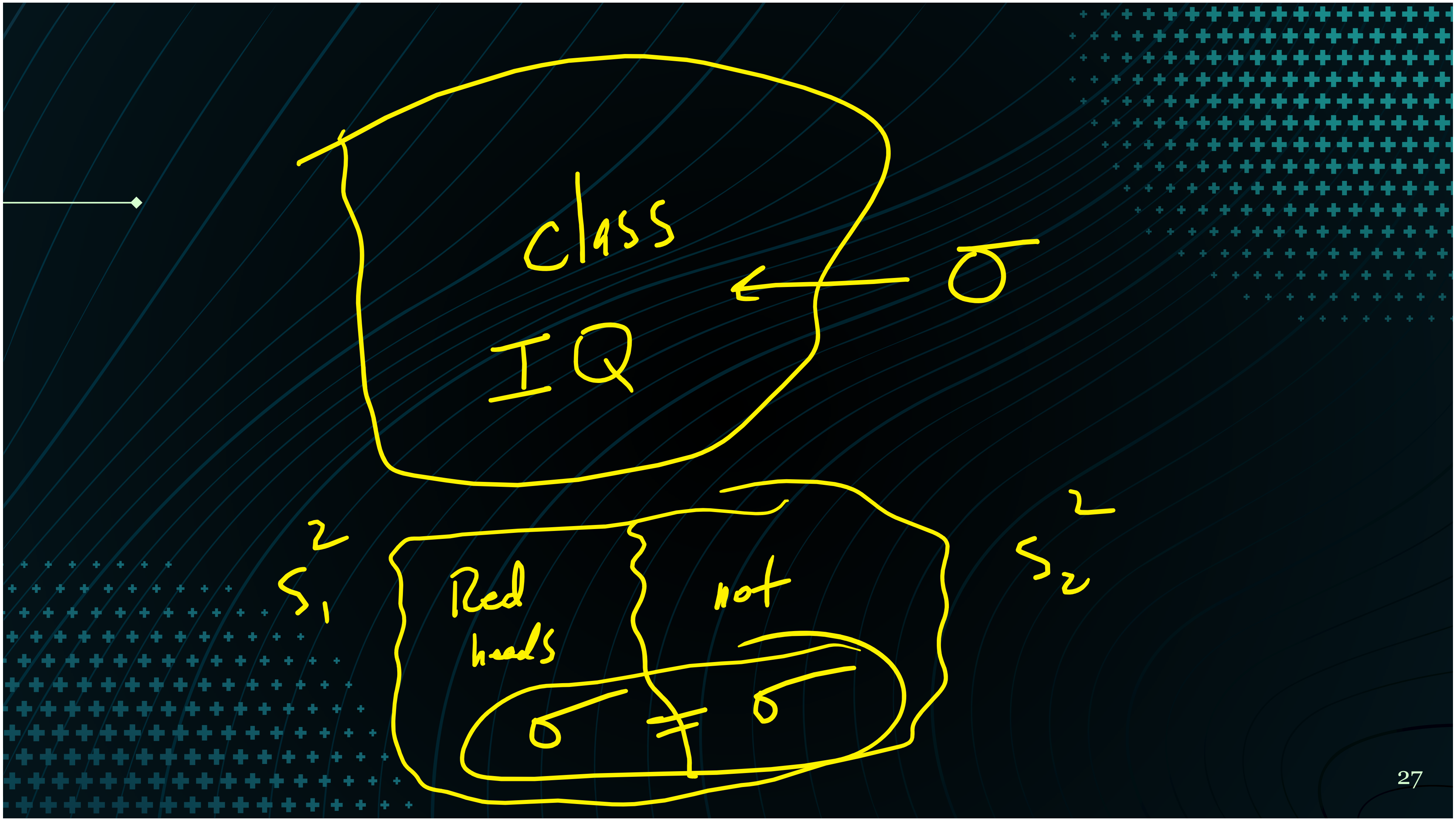
$\bar{x}$=72.9, $\boxed{s_X = 25.6}$, $\bar{y}$=81.7, $\boxed{s_Y = 28.3}$. Perform a test to see if Houston has worse air quality at 0.05 significance and state the result.

H_0: $\mu_X = \mu_Y$

H_1: $\mu_X < \mu_Y$.

$$S_p^2 = \frac{(13-1)\,25.6^2 + (16-1)\,28.3^2}{13+16-2} =$$

$$\frac{\bar{x} - \bar{y} - 0}{\sqrt{\frac{S_p^2}{13} + \frac{S_p^2}{16}}} \sim t_{\underbrace{n_1+n_2-2}}$$


class
IQ
σ
s_1^2
Red heads
not
σ ≠ σ
s_2^2

- Degrees of freedom: 2 sample

- If pairing: → n-1
- If not paring:
  - If pooling: → n1 + n2 – 2
  - If not pooling:
    - Conversative: min(n1,n2) – 1
    - Or Welch's ← default

$$P(X \le x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2} dw$$

	$P(X \le x)$					
	0.010	$\boxed{0.025}$	$\boxed{0.050}$	0.100	0.900	0.950
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$
1	0.000	0.001	[illegible]	0.016	2.706	3.841
2	0.020	0.051	[illegible]	0.211	4.605	5.991
3	0.115	0.216	[illegible]	0.584	6.251	7.815
4	0.297	0.484	[illegible]	1.064	7.779	9.488
5	0.554	0.831	[illegible]	1.610	9.236	11.07
6	0.872	1.237	[illegible]	2.204	10.64	12.59
7	1.239	1.690	[illegible]	2.833	12.02	14.07
8	1.646	2.180	2.733	3.490	13.36	15.51
9	2.088	[illegible]	3.325	4.168	14.68	16.92
10	2.558	3.247	3.940	4.865	15.99	18.31
11	3.053	3.816	4.575	5.578	17.28	19.68
12	3.571	4.404	5.226	6.304	18.55	21.03
13	4.107	5.009	5.892	7.042	19.81	22.36
14	4.660	5.629	6.571	7.790	21.06	23.68
15	5.229	6.262	7.261	8.547	22.31	25.00

2.5

$2.5 \sim \chi^2_8$

$.025 < p < .05$