CS2810 Day 18

Mar 28 2022

Admin:
Quiz3 is Friday
Review session tomorrow (see piazza)
stop by my OH on Thursday too!
final swap sections (lottery)

SBJ FINAL LOTTERY
BY TOMORROW NIGHT

Content:
Big goal: T-Tests (difference of mean of two distributions)
Pooled Covariance
~~One and two tailed~~ hypothesis tests

Which song is preferred by students?

$X$

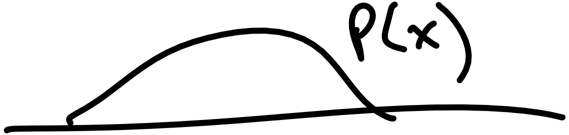
LET IT GO

2 2 3 5 3
5 4 3 1 3

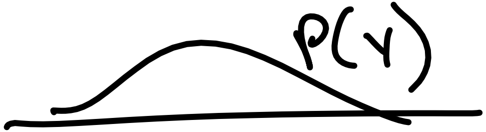
$Y$

NIGHTS AVICII

1 5 3 1 4 4
2 4 4 1 1 1




$\mu_x$ $\mu_Y$



$S^2 = VAR(\bar{x} - \bar{y})$

Which song is preferred by students?

ICA 2

COMPUTE $\hat{S}^2 = \frac{\hat{\sigma}_x^2}{N_x} + \frac{\hat{\sigma}_y^2}{N_y}$

Let it go (X): 2, 2, 3, 5, 3, 5, 4, 3, 1, 3 (Nx = 10)
Nights (Y): 1, 5, 3, 1, 4, 4, 2, 4, 4, 1, 1 (Ny = 11)

$$\bar{X} = \frac{2+2+3+5+3+5+4+3+1+3}{10} = \frac{31}{10} = 3.1 \qquad \bar{Y} = \frac{30}{11}$$

$$\hat{\sigma}_x^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2 = \frac{1}{10-1}\left[(2-3.1)^2 + (2-3.1)^2 + (3-3.1)^2 + \dots\right.$$

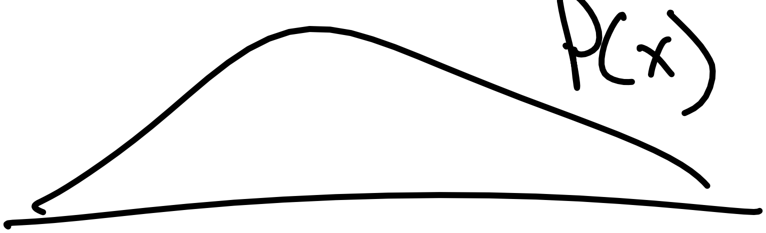
$$\simeq 1.65$$

$$\hat{\sigma}_y^2 \simeq 2.418$$

$$\hat{S}^2 \simeq \frac{1.65}{10} + \frac{2.418}{11} \simeq .385$$

$$\bar{Z} = \frac{\bar{X} - \bar{Y}}{\hat{S}} = \frac{3.1 - \frac{30}{11}}{\sqrt{.385}} \hat{=} .6$$

SINCE $\hat{S}^2 = .385$




$$E[x] = \mu$$

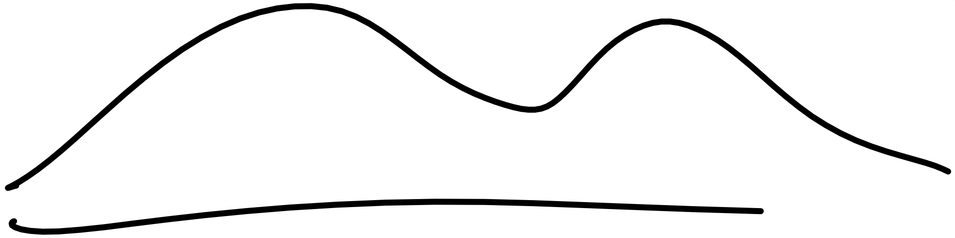
$$VAR(x) = \sigma^2$$

COLLECT SAMPLES

$x_1 \quad x_2 \quad x_3 \quad \ldots$

$$\hat{\mu} = \bar{x} = \frac{1}{N_x} \sum x_i$$

$$\hat{\sigma}^2 = \frac{1}{N_x - 1} \sum_i (x_i - \bar{x})^2$$



BIMODAL

Testing Difference of means: $\bar{X}$ $\bar{Y}$

Given samples from two distributions, we seek to test if the mean of one is different than other

$H_1: \mu_x > \mu_y$ $H_1: \mu_x \neq \mu_y$ $H_1: \mu_x < \mu_y$

Assumptions:
1. Each observation is independent of all others
2. Variance of each distribution is the same $\rightarrow VAR(X) = VAR(Y)$
3. Either
   - Each distribution is normal
   - There are sufficiently many observations that we can claim mean of distribution is normal
     - Central Limit Theorem: mean of a set of indep observations gets closer to normal with more samples
4. Our variance estimates equals the ground truth variance
   - This assumption is too strong to make approach practical ... we'll modify to remove it later

## In Class Assignment 1

Describe a circumstance which explicitly breaks assumption 1 ~~and 2~~ in our music preference example.

friends sit together and have similar music interests, so sampling neighbors in class might introduce dependancy

temporal dependancy between samples (earlier samples influence the later ones)
- can be solved by "blinding" everyone to other's survey responses

temporary dependancy between songs (earlier songs surveyed have higher / lower score)
- have two groups and mix the order in which songs are presented

shouldn't survey the person who selected the song itself

## Testing Difference of Mean: Overview (Z STAT VERSION)

Step 0: Compute \hat{S}^2, the sample variance of \bar{x} - \bar{y}

$\hat{S}^2 = \mathrm{VAR}(\bar{x} - \bar{y})$

Step 1 : Compute Z statistic

$Z = (\bar{x} - \bar{y}) / \hat{S}$

Step 2: Build distribution of Z statistic under the null hypothesis

$N(0,1)$

Step 3: Compute p-value

Step 4: Compare p-value to alpha threshold

- If p-value < alpha:
    - reject null hypothesis, claim hypothesis is true
- If pvalue >= alpha:
    - don't reject null hypothesis (no claims made)

## STEP 2: COMPUTE $S^2$

FROM ASSUMPTIONS

$\bar{X} \sim N(\mu_X, \sigma^2)$

$\bar{Y} \sim N(\mu_Y, \sigma^2)$

EACH SAMPLE IS INDEPENDENT

$$VAR(\bar{X} - \bar{Y}) = VAR(\bar{X}) + VAR(-\bar{Y})$$

$VAR(cX) = c^2 VAR X$

$$= VAR(\bar{X}) + (-1)^2 VAR(\bar{Y})$$

$$= VAR(\bar{X}) + VAR(\bar{Y})$$

$$= VAR(\bar{x}) + VAR(\bar{y})$$

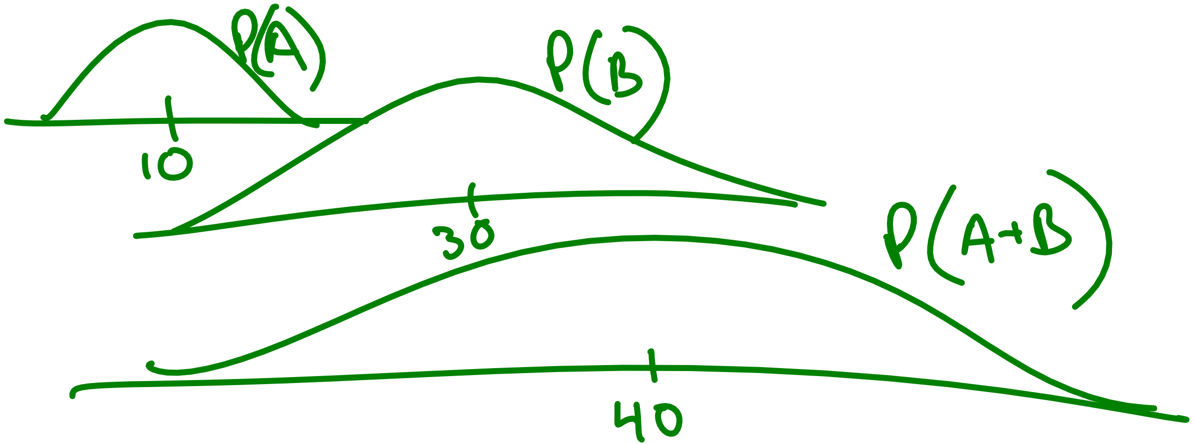
$$= VAR\left(\frac{x_1 + x_2 + x_3 + \ldots}{N_x}\right) + VAR\left(\frac{y_1 + y_2 + y_3 + \ldots}{N_y}\right)$$

$$= \frac{1}{N_x^2}\left(VAR(x_1) + VAR(x_2) + \ldots\right) + \frac{1}{N_y^2}\left(VAR(y_1) + VAR(y_2) + \ldots\right) = \frac{VAR(x)}{N_x} + \frac{VAR(y)}{N_y}$$

$$\hat{S}^2 = \frac{\hat{\sigma}_x^2}{N_x} + \frac{\hat{\sigma}_y^2}{N_y}$$

ASSUME X, Y INDEP

$$\mathrm{VAR}(A)+\mathrm{VAR}(B) = \mathrm{VAR}(A+B) = \mathrm{VAR}(A-B)$$



# STEP 1: COMPUTE Z-STATISTIC

$$Z = \frac{\bar{X} - \bar{Y}}{\hat{S}}$$

## STEP 2: BUILD DISTRIBUTION OF Z STATISTIC UNDER NULL HYPOTHESIS ($H_0: \mu_x = \mu_y$)

→ SINCE $\bar{X}, \bar{Y}$ ARE NORMAL SO IS $\dfrac{\bar{X} - \bar{Y}}{S}$

$$Z = \frac{\bar{X} - \bar{Y}}{S}$$

→ UNDER NULL HYPOTHESIS $E[\bar{X}] = E[\bar{Y}]$

SO $E[Z] = 0$

$$\mathrm{VAR}(z) = \mathrm{VAR}\left(\frac{\bar{x} - \bar{y}}{\hat{S}}\right)$$

$$= \frac{1}{\hat{S}^2} \mathrm{VAR}(\bar{x} - \bar{y})$$

$$= \frac{S^2}{\hat{S}^2} = 1$$

ASSUME 4: VAR ESTIMATE IS EXACT

# STEP 2 (ALWAYS)

$\mu \quad \sigma^2$

UNDER THE NULL HYPOTHESIS $Z \sim N(0,1)$

($Z$ IS OFTEN CALLED "STANDARD NORMAL")

# Computing P value

H0: Same mean

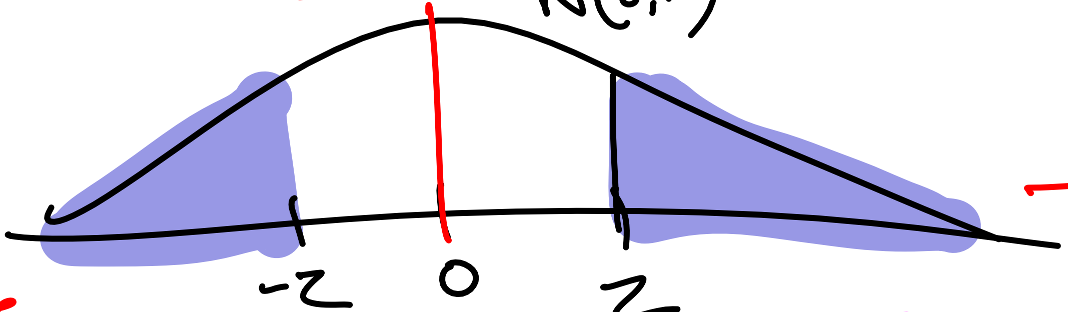
Prob of all outcomes less consistent w/ $H_0: \mu_x = \mu_y$

$H_1: \mu_x \neq \mu_y$

Most consistent w/ $H_0$

$N(0,1)$

$$z = \frac{\bar{x} - \bar{y}}{\hat{s}}$$

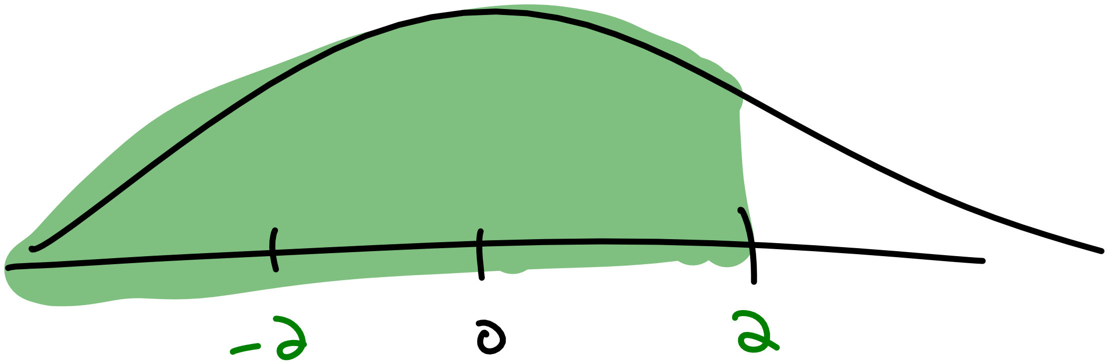



Less consiste nt w/ $H_0$

Less consistent w/ $H_0$

$P_{val} = 2 \cdot CDF(-z)$

2-tailed test




$2 \cdot \mathrm{CDF}(2)$

# Computing pvalue

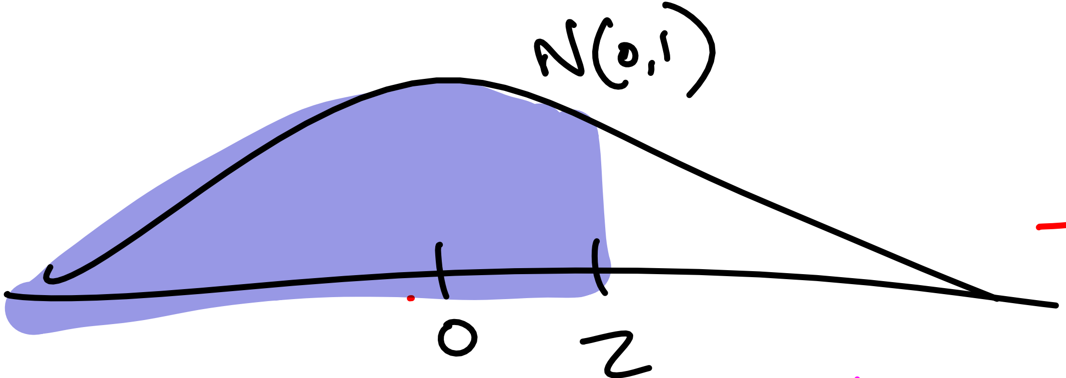
**Prob of all outcomes less consistent w/ $H_0: \mu_x > \mu_y$**

$H_1: \mu_x \leq \mu_y$

$H_0$: x is bigger than y

$$z = \frac{\bar{x} - \bar{y}}{\hat{s}}$$

Least consistent w/ $H_0$




Most consistent w/ $H_0$

pval = CDF(z)

1-tailed test

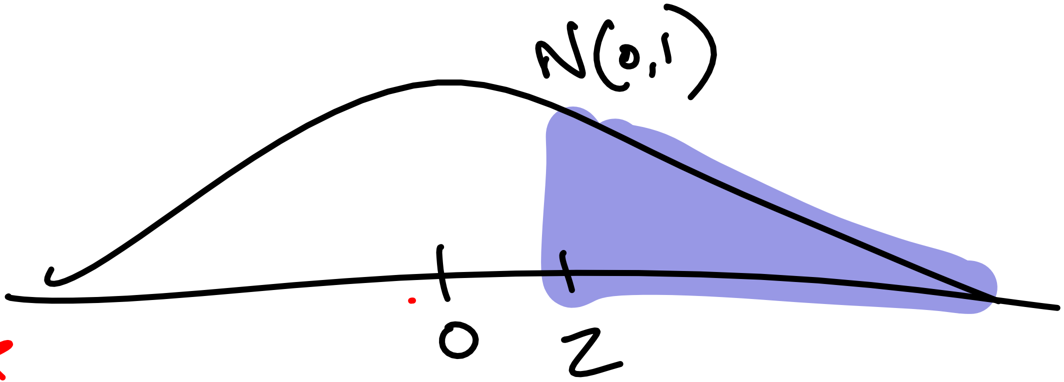
# Computing P value

Prob of all outcomes less consistent w/ $H_0: \mu_x < \mu_y$

X smaller than Y

$H_1: \mu_x \geq \mu_y$

$$z = \frac{\bar{x} - \bar{y}}{\hat{s}}$$




Most consistent $H_0$

Least consistent $H_0$

1-tailed

$$P_{VAL} = 1 - CDF(z)$$

Testing Difference of Mean: Overview (Z STAT VERSION)

$H_0: \mu_X = \mu_Y$

Step 0: Compute \hat{S}^2, the sample variance of \bar{x} - \bar{y}

$$\hat{S}^2 = \frac{\hat{\sigma}_x^2}{N_x} + \frac{\hat{\sigma}_Y^2}{N_Y}$$

$H_1: \mu_X \neq \mu_Y$

Step 1 : Compute Z statistic

$$Z = (\bar{x} - \bar{y}) / \hat{S} = .6$$

Step 2: Build distribution of Z statistic under the null hypothesis

$N(0,1)$

Step 3: Compute p-value


$PVAL = 2 \cdot CDF(.6) = .548$

Step 4: Compare p-value to alpha threshold $\alpha = .05$

If p-value < alpha:

reject null hypothesis, claim hypothesis is true

If pvalue >= alpha:

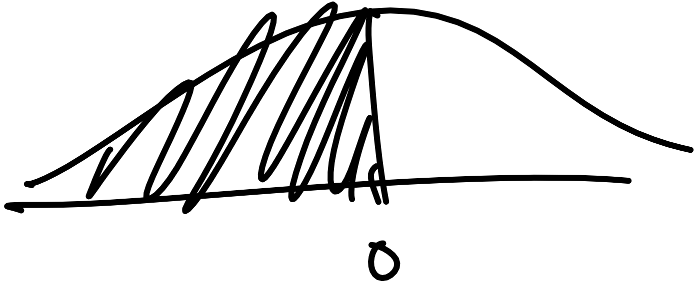
don't reject null hypothesis (no claims made)

ICA 1:
Compute a final p-value and summarize the results of our analysis about song preference

WHAT IF $Z = 0$

$P_{VAL} = 2 \cdot CDF(0)$

$= 2 \cdot (.5) = 1$




$H_0: \mu_x = \mu_y$

$Z = \frac{\bar{X} - \bar{Y}}{\hat{S}}$

$Z = 0 \Rightarrow \bar{X} = \bar{Y}$

Testing Difference of means:

Given samples from two distributions, we seek to test if the mean of one is different than other

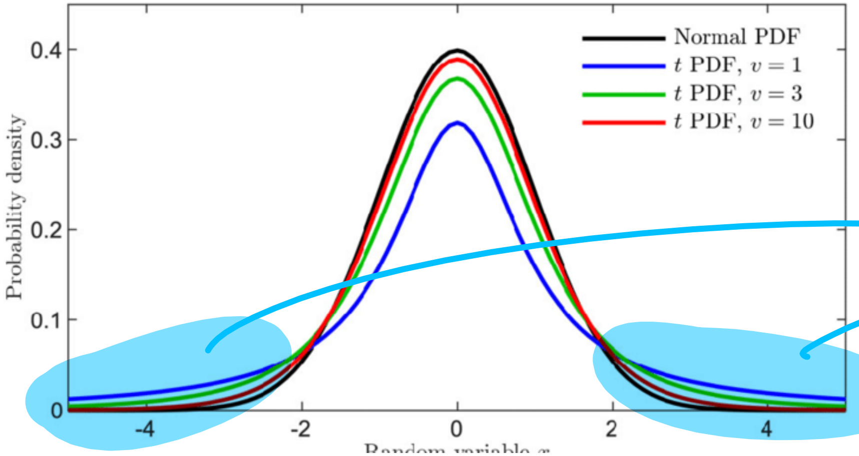
$\hat{S}$ ESTIMATES VAR$(\bar{x}-\bar{y})$

Assumptions:
1. Each observation is independent of all others
2. Variance of each distribution is the same
3. Either
    - Each distribution is normal
    - There are sufficiently many observations that we can claim mean of distribution is normal
        - Central Limit Theorem: mean of a set of indep observations gets closer to normal with more samples
4. Our variance estimates equals the ground truth variance
    - This assumption is too strong to make approach practical ... we'll modify to remove it later

LETS REMOVE THIS!

# T Distribution

Using a T Distribution in Place of Normal Accounts For uncertainty in variance estimate



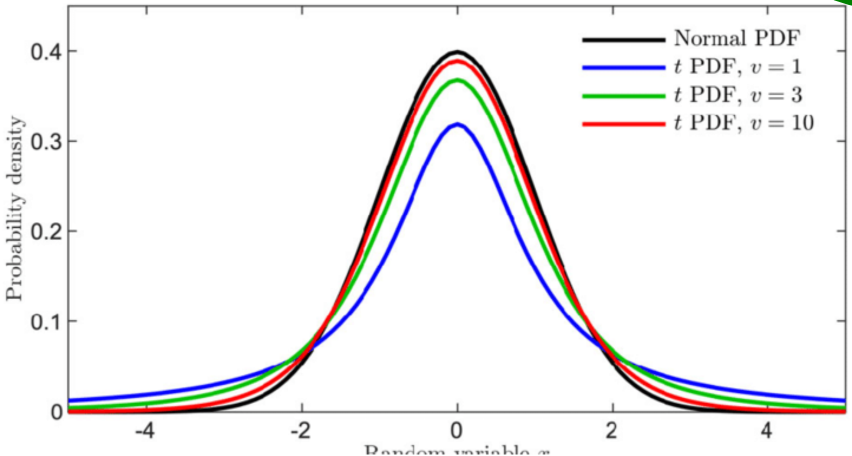

"Fat Tails"

T Distribution is less confident than Normal

# T Distribution

T Distribution has 1 Parameter: Degrees of Freedom (DF). $DF = N_x + N_y - 2$

$\#$ observations of $X$

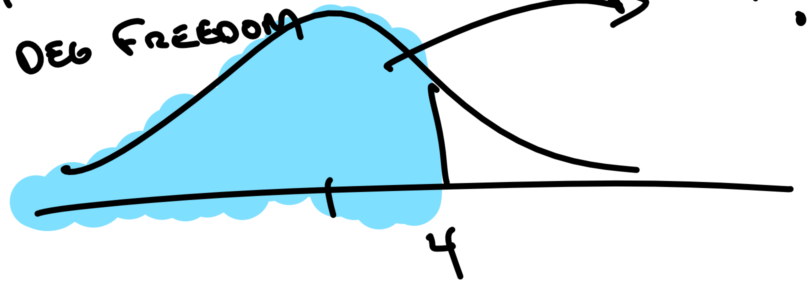



As DF increases T Distribution Approaches Normal ($\hat{S}^2$ is unbiased + Law Large Numbers)

IN PYTHON (ALL LOWERCASE)

FROM SCIPY.STATS IMPORT T

T DISTRIBUTION W/
11 DEG FREEDOM




T.CDF(4, DF=11)

Let's summarize ...

Testing Difference of means (T-Test version ... use this one, Z-test only for exposition)

Given samples from two distributions, we seek to test if the mean of one is different than other

3 different hypotheses we can investigate:

$H_0: \mu_x > \mu_Y$ $H_0: \mu_x < \mu_Y$ $H_0: \mu_x = \mu_Y$

$H_1: \mu_x \leq \mu_Y$ $H_1: \mu_x \geq \mu_Y$ $H_1: \mu_x \neq \mu_Y$

Assumptions:

1. Each observation is independent of all others
2. Variance of each distribution is the same
3. Either
    - Each distribution is normal
    - There are sufficiently many observations that we can claim mean of distribution is normal
        - Central Limit Theorem: mean of a set of indep observations gets closer to normal with more samples

## Testing Difference of Mean: Overview (T STAT VERSION)

$$\hat{S}^2 = \frac{\hat{\sigma}_x^2}{N_x} + \frac{\hat{\sigma}_y^2}{N_y}$$

Step 0: Estimate \hat{S}^2, the sample variance of \bar{x} - \bar{y}

Step 1 : Compute T statistic $T = (\bar{x} - \bar{y}) / \hat{S}$

Step 2: Build distribution of T statistic under the null hypothesis $T \sim T(DF = N_x + N_y - 2)$

Step 3: Compute p-value

Step 4: Compare p-value to alpha threshold
- If p-value < alpha:
  - reject null hypothesis, claim hypothesis is true
- If pvalue >= alpha:
  - don't reject null hypothesis (no claims made)

ICA 3

$H_1: \mu_x > \mu_y$ $H_0: \mu_x \leq \mu_y$

Somebody (somewhere) thinks starting each day at 4 AM with an ice cold shower will increase student performance. They conduct an experiment where a group of students wakes up at 4 AM with an icy shower while another group of students does not. Their test scores are listed below:

$X = 90, 95, 90, 80, 70$ $Y = 100, 80, 70, 90, 95$

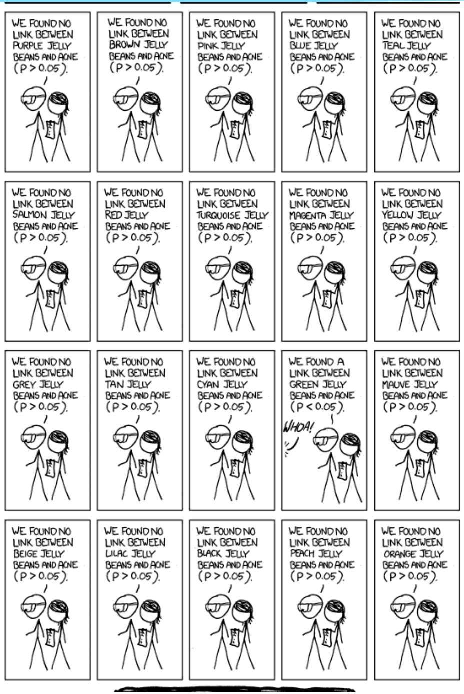
Perform a two-sample T test (as shown) which is able to claim that the icy start to the day improves test scores at the alpha = .05 signifigance level.

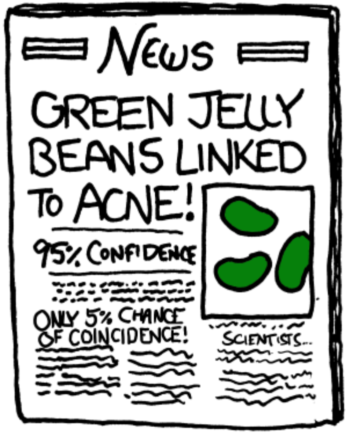
1. Express hypotheses (algebraically: H_0: \mu_x > \mu_y while H_1: \mu_x <= \mu_y or similar)
2. Compute \hat{S}^2
3. Compute T statistic
4. Compute P-value
5. Synthesize your analysis with a one sentence summary

$$\hat{S}^2 = \frac{\hat{\sigma}_x^2}{N_x} + \frac{\hat{\sigma}_y^2}{N_y}$$

If time in class …

# Family Wise Error Rate




CREDIT : XKCD

# FAMILY WISE ERROR RATE

EACH TEST WAS $P(\text{TYPE 1 ERROR}) < .05 = \alpha$

... BUT IF WE RUN MANY TESTS WE INCREASE PROBABILITY OF AT LEAST 1 TEST GETTING TYPE 1 ERROR ACROSS ALL EXPERIMENTS

→ FAMILY WISE ERROR RATE (FWER)

# FAMILY WISE ERROR RATE

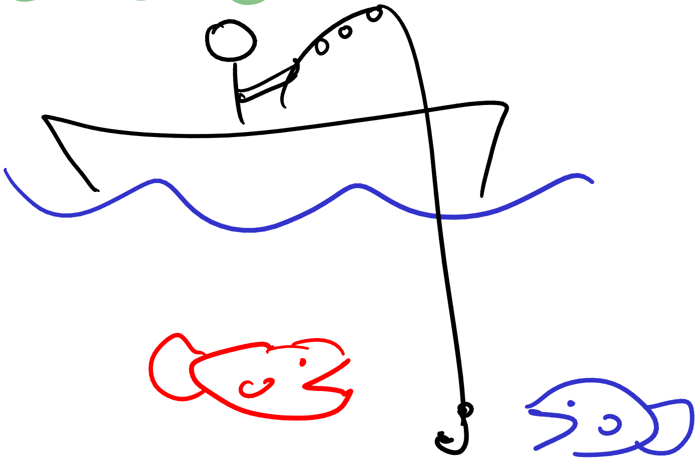
BONFERRONI

TO ENSURE $P(\underline{FWER}) < \alpha$ USE SIGNIFICANCE

THRESHOLD $\frac{\alpha}{N}$ IN EACH EXPERIMENT

TOTAL # EXPERIMENTS

# Experimental Bias



Fish caught

| Red | Blue |
|---|---|
| 卌 卌 | 卌 |

"There are twice as many red fish in water"

What if blue fish are less likely to be caught?

★ How do my observations depend on reality? ★

# BIAS EXAMPLES

SUBJECT BIAS (PLACEBO EFFECT)

EXPERIMENTER BIAS (COUNTING HORSE)

SELECTION BIAS (CHARTER SCHOOL)

TEMPORAL BIAS (LEARNING CURVE)

FUNDING BIAS (NFL CONCUSSION RESEARCH)

**SINGLE BLIND**

SBJS DO NOT KNOW TEST CONDITION

**DOUBLE BLIND**

SBJS + EXPERIMENTER DO NOT KNOW TEST CONDITION