CS2810 Day 18
Mar 28 2022

Admin:
Quiz3 is Friday
Review session tomorrow (see piazza)
stop by my OH on Thursday too!
how to request a zoom link

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

FAMILY WISE ERROR RATE
EXPERIMENTAL BIAS

Which song is preferred by students?

X

MR BRIGHTSIDE

3 5 2 2 3 3 4 3

$N_X = 8$

Y

FEEL THIS MOMENT

5 5 2 5 3 4 3 2 4

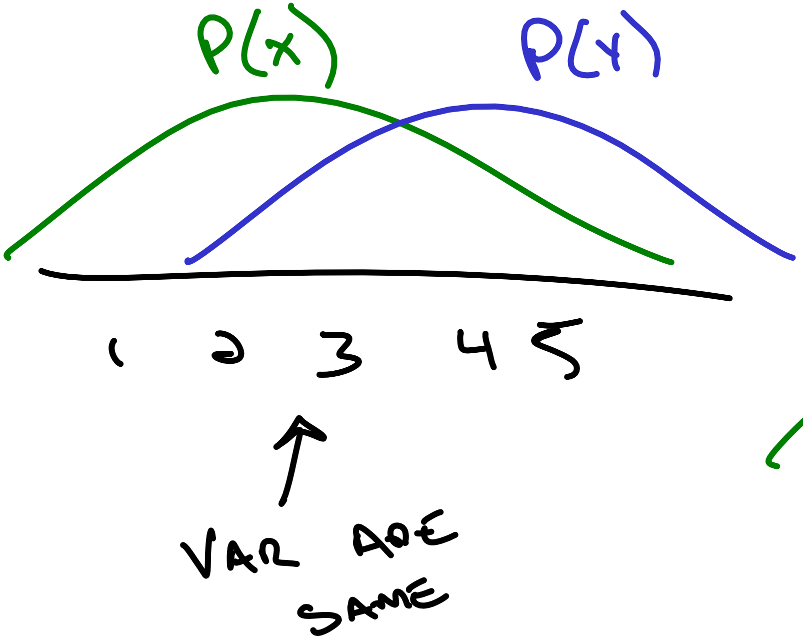
$N_Y = 9$

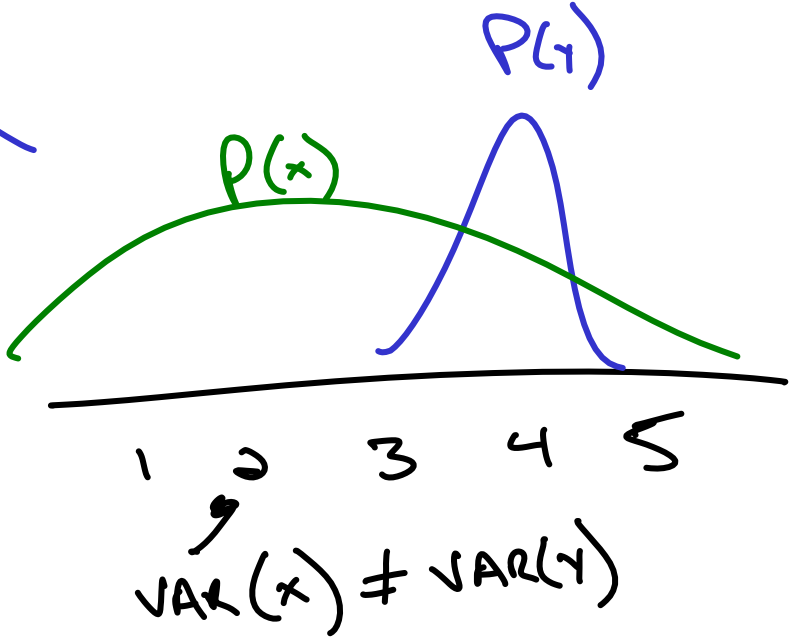
Testing Difference of means:

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

$$H_0: \mu_x = \mu_y \qquad H_1: \mu_x \neq \mu_y$$

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
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


P(x)
P(Y)
1 2 3 4 5
VAR ARE SAME


P(x)
P(Y)
1 2 3 4 5
VAR(x) ≠ VAR(Y)

In Class Assignment 1

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

1. Each observation is independent of all others

- sample from a group of similar people (at same concert) the variance might be lower than another sample set
- if one respondee listens to the responses given before their own

2. Variance of each distribution is the same

- sample from a group of similar people (at same concert) the variance might be lower than another sample set

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

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

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

ALL SAMPLES INDEP

$VAR(CX) = C^2 VAR(X)$

$$
\begin{aligned}
VAR(\bar{X} - \bar{Y}) &= VAR(\bar{X}) + VAR(-1\bar{Y}) \\
&= VAR(\bar{X}) + (-1)^2 VAR(\bar{Y}) \\
&= VAR(\bar{X}) + VAR(\bar{Y}) \\
&= VAR\left(\frac{X_1 + X_2 + \ldots}{N_X}\right) + VAR\left(\frac{Y_1 + Y_2 + \ldots}{N_Y}\right)
\end{aligned}
$$

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

$$= \frac{1}{N_x^2}\mathrm{VAR}(x_1) + \frac{1}{N_x^2}\mathrm{VAR}(x_2) + \ldots$$

$$\frac{1}{N_y^2}\mathrm{VAR}(y_1) + \frac{1}{N_y^2}\mathrm{VAR}(y_2) + \ldots$$

$$= \frac{\mathrm{VAR}(x)}{N_x} + \frac{\mathrm{VAR}(y)}{N_y}$$

Bessel's VAR

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

# STEP 1: COMPUTE Z-STATISTIC

$$Z = \frac{\bar{X} - \bar{Y}}{\hat{S}} = \frac{25/8 - 33/9}{\sqrt{.289}} \approx -1$$

| X | $\bar{X} = \frac{25}{8}$ | Y | $\bar{Y} = 33/9$ |
|---|---|---|---|
| MR BRIGHTSIDE | | FEEL THIS MOMENT | |
| 3 5 2 2 3 3 4 3 | | 5 5 2 5 3 4 3 2 4 | |

$X$ $\qquad$ $Y$

MR BRIGHTSIDE $\qquad$ FEEL THIS MOMENT

3 5 2 2 3 3 4 3 $\qquad$ 5 5 2 5 3 4 3 2 4

$$\hat{\sigma}_x^2 = \frac{1}{8-1}\left[\left(3-\frac{25}{8}\right)^2 + \left(5-\frac{25}{8}\right)^2 + \left(2-\frac{25}{8}\right)^2 + \ldots\right] \hat{=} .98214$$

$$\hat{\sigma}_y^2 \hat{=} 1.5$$

$$\hat{S}^2 = \frac{\hat{\sigma}_x^2}{N_x} + \frac{\hat{\sigma}_y^2}{N_y} = \frac{.982}{8} + \frac{1.5}{9} \hat{=} .289$$

## 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 (<u>ALWAYS</u>)

UNDER THE NULL HYPOTHESIS $Z \sim N(\overset{\mu}{0}, \overset{\sigma^2}{1})$

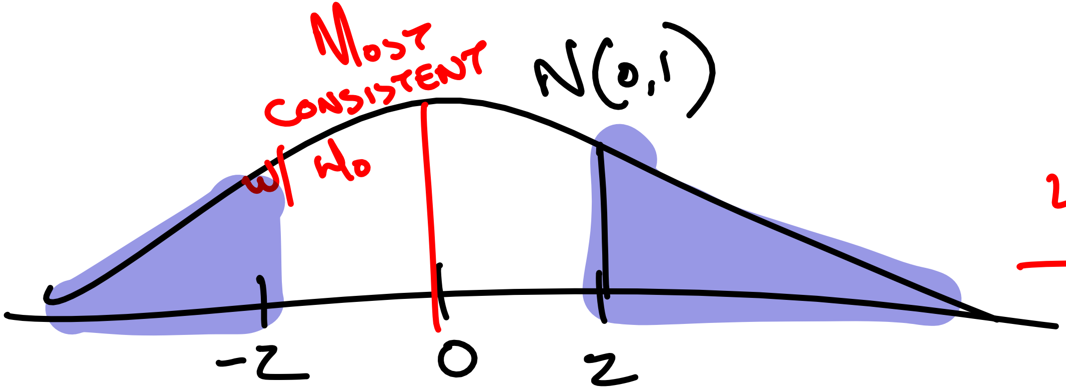
(Z IS OFTEN CALLED "STANDARD NORMAL")

# Computing P value Step 3

$H_0$: Same mean

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

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




Less consistent w/ $H_0$

Less consistent w/ $H_0$

2 tailed test

Computing P value

Step 3

$H_0$: X is bigger than Y

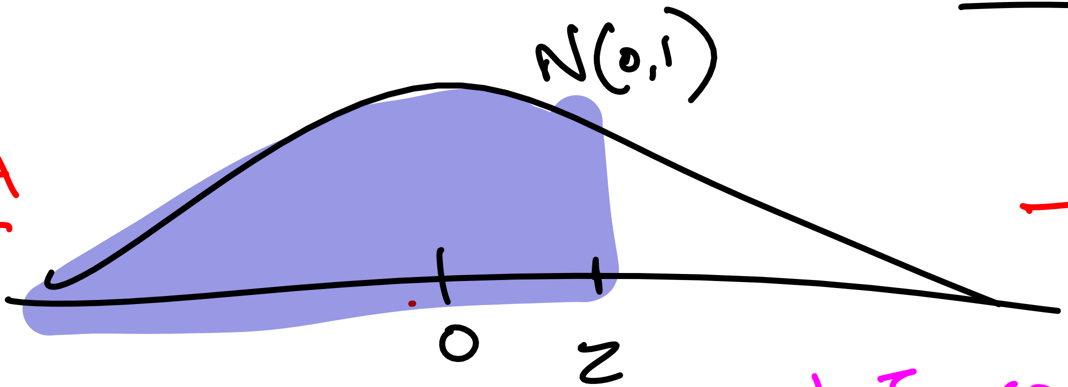
Prob of all outcomes less consistent w/ $H_0: \mu_x > \mu_y$

$H_1: \mu_x \leq \mu_y$

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

A

Least consistent w/ $H_0$




Most consistent w/ $H_0$

1-Tailed Test

Computing P value

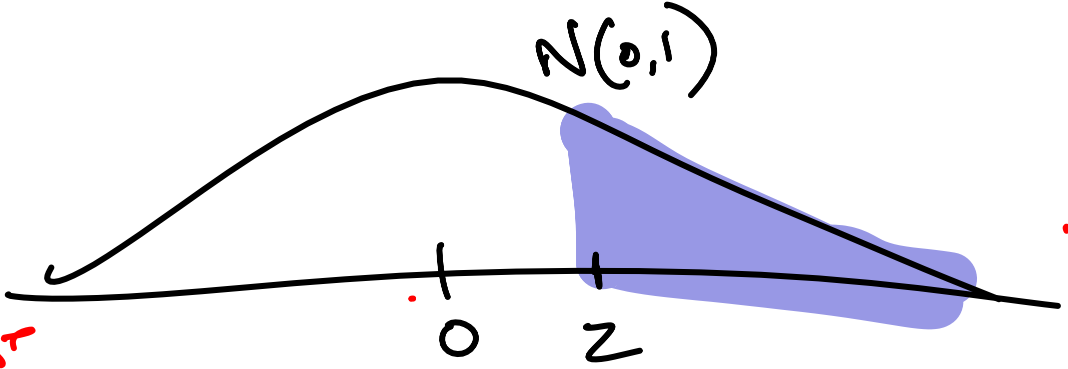
Step 3

X smaller than Y

Prob of all outcomes less consistent w/ $H_0: \mu_X < \mu_Y$

$H_1: \mu_X \geq \mu_Y$

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




Most consistent w/ $H_0$

Least consistent $H_0$

1-tailed test

Testing Difference of Mean: Overview (Z STAT VERSION)

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

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

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 $\alpha = .05$

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

X

MR BRIGHTSIDE

3 5 2 2 3 3 4 3

Y

FEEL THIS MOMENT

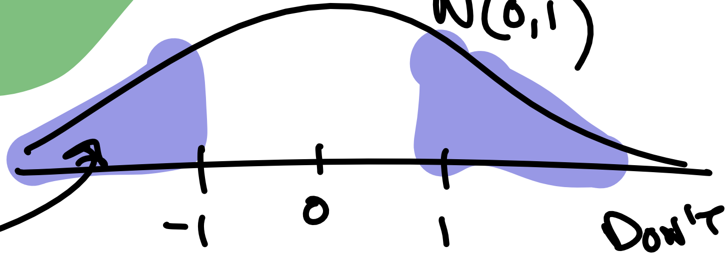
5 5 2 5 3 4 3 2 4

$H_0: \mu_x = \mu_y$

$H_1: \mu_x \neq \mu_y$

$z \hat{=} -1$ (SEE PREVIOUS COMPUTE)

N(0,1)



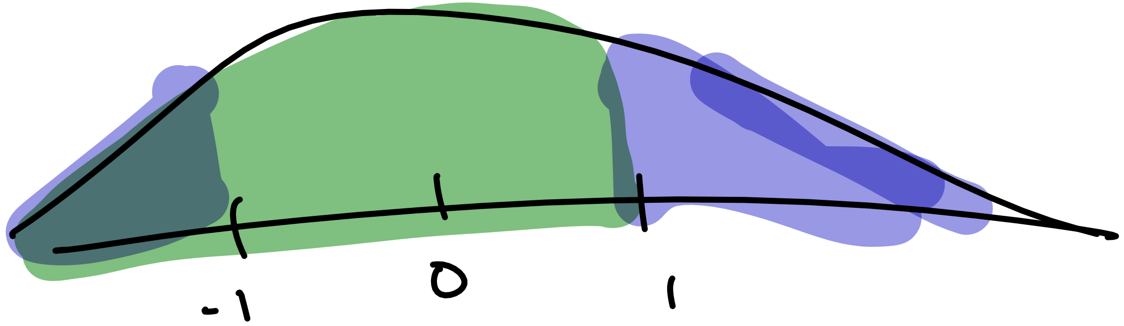

PVAL = .317

$2 \times CDF(-1)$

DON'T REJECT $H_0$

SONGS SEEM EQUALLY POPULAR

# COMMON MISTAKE




$2 \cdot CDF(1) \neq PVALUE$

# Testing Difference of means:

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

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

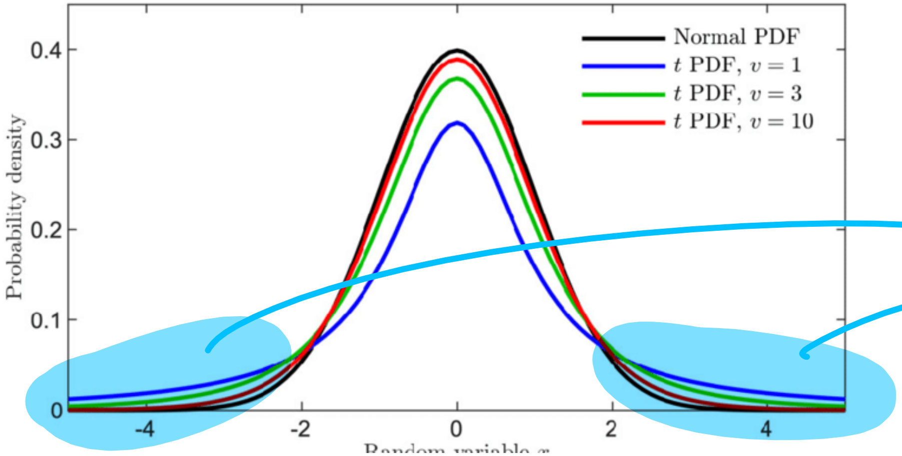
Assumptions:

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



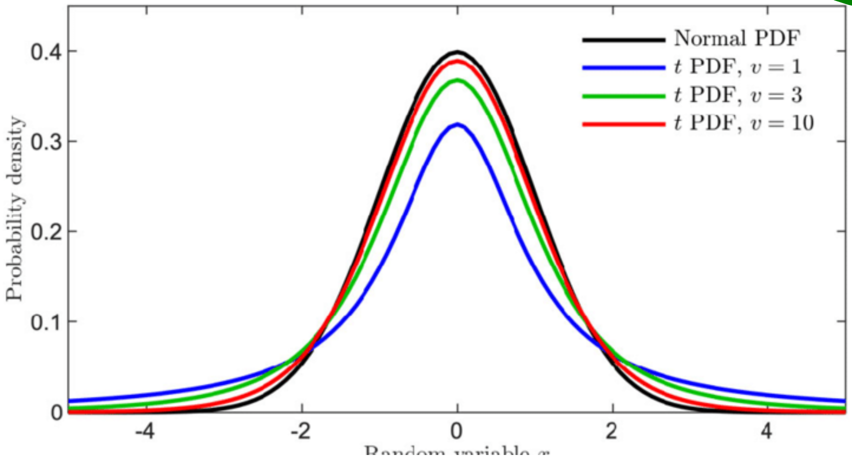

"Fat Tails"

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

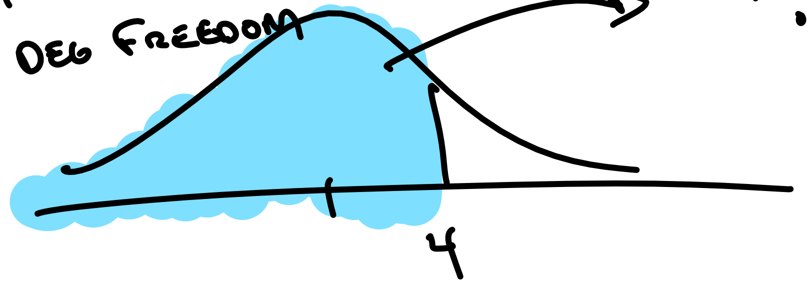



As DF increases T distribution approaches Normal ($\hat{S}^2$ is unbiased + law large numbers)

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:

| | | | |
|---|---|---|---|
| Hypotheses | \mu_x < \mu_y | \mu_x != \mu_y | \mu_x > \mu_y |
| Null Hypotheses | \mu_x >= \mu_y | \mu_x = \mu_y | \mu_x <= \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)

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

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

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

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:

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

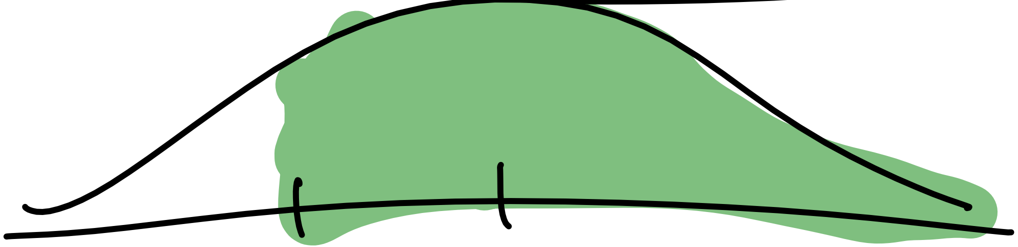
Perform a two-sample T test (as shown) which is able to claim that the icy start to the day improves test scores at the alpha = .05 signifigance level.

$\mu_x > \mu_y$ $\qquad$ $\mu_x \le \mu_y$

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

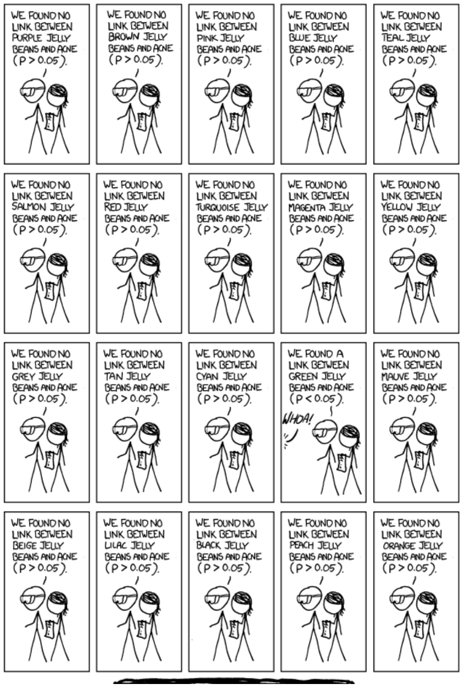
$H_1$: $\mu_x > \mu_y$ $\quad$ $\mu_x \neq \mu_y$ $\quad$ $\mu_x < \mu_y$

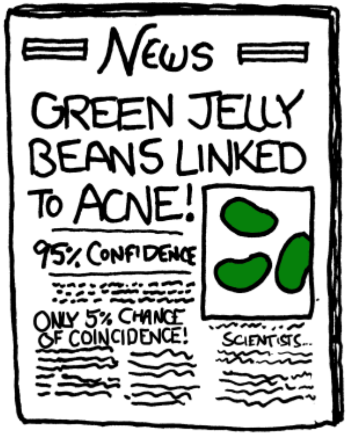
$H_0$: $\mu_x \leq \mu_y$ $\quad$ $\mu_x = \mu_y$ $\quad$ $\mu_x \geq \mu_y$



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)

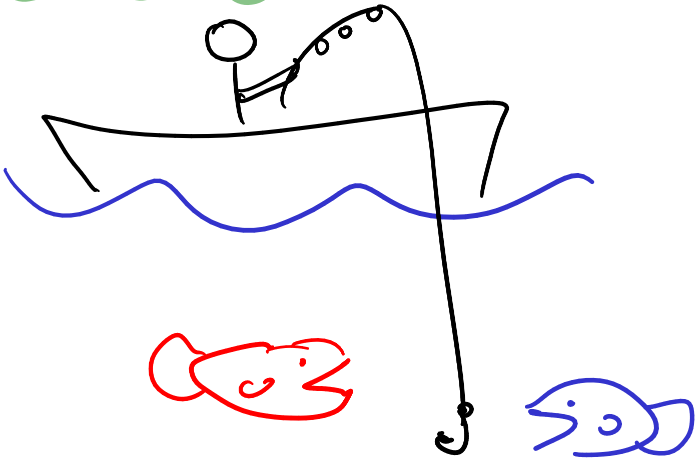
# FAMILY WISE ERROR RATE

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

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

$N \rightarrow$ 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