Student t Test & Two Population

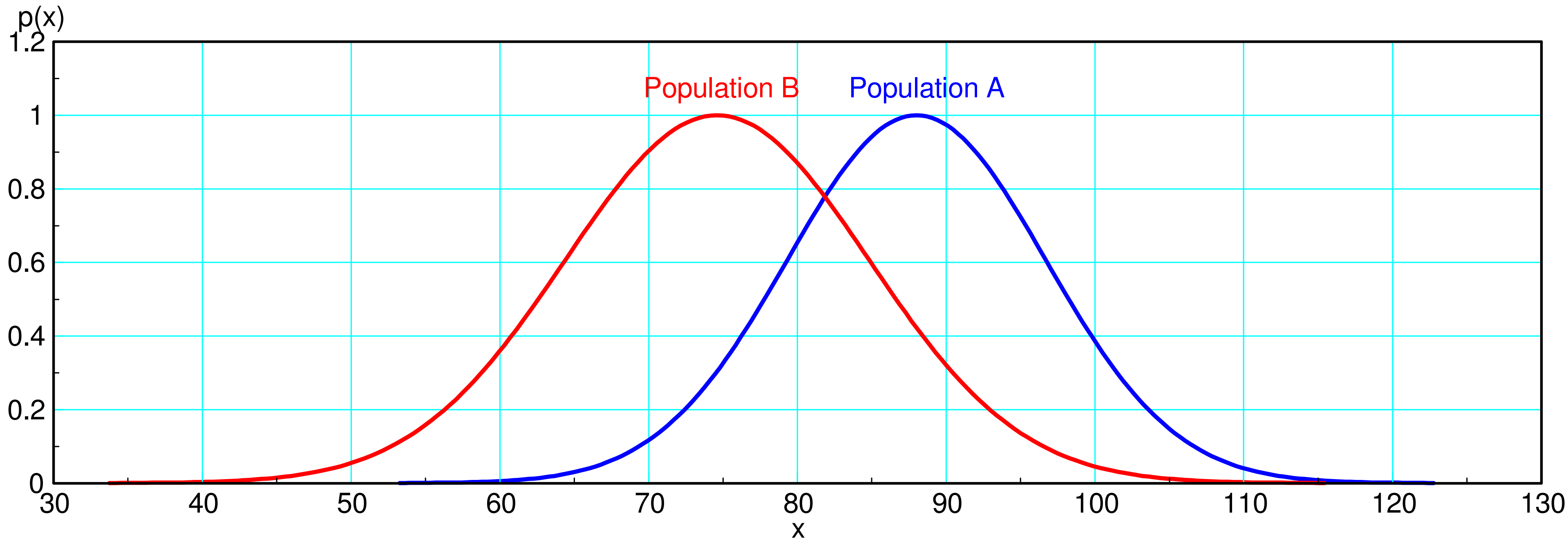
ECE 376 Embedded Systems

Jake Glower - Lecture #16b

Please visit Bison Academy for corresponding lecture notes, homework sets, and solutions

Student t Distribution with 2 Populations

- Use a student-t distribution to determine which sample has the higher mean



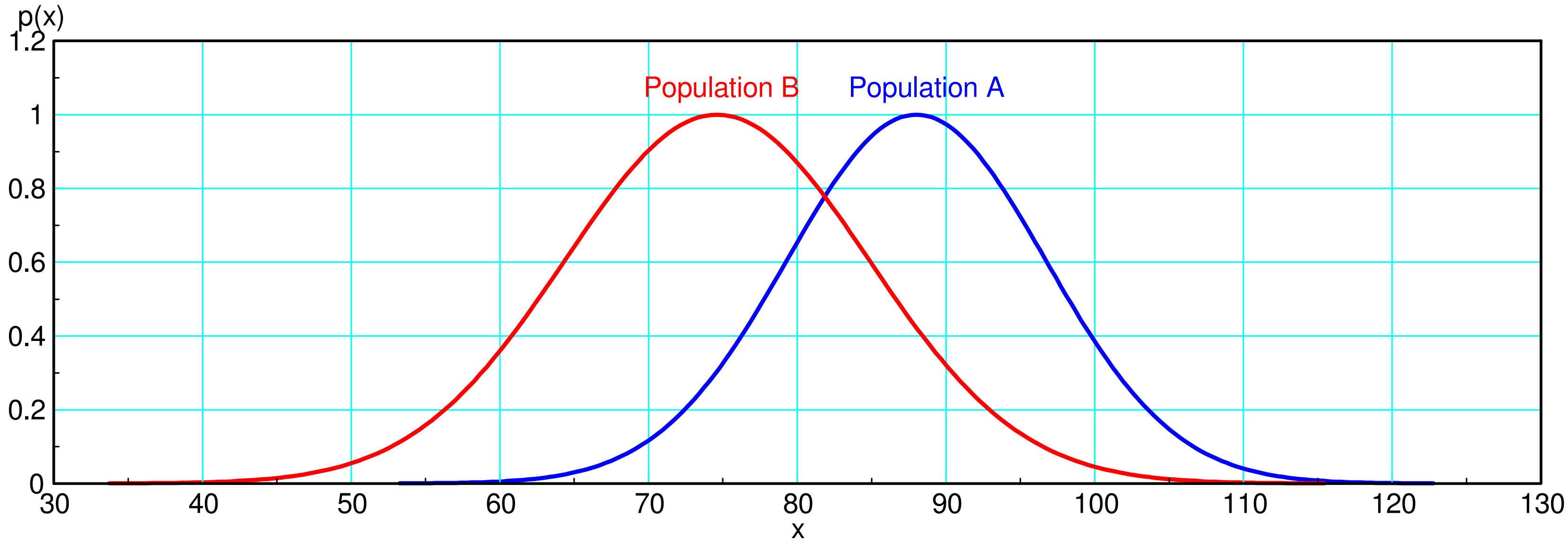
Comparison of Means (Individual)

Assume:

- A and B be normally distributed distributions with unknown means and variances.
- na and nb samples are taken from population A and B respectfully.

What is the probability that the next value from A will be larger than B?

$$p(x_a > x_b) = ?$$



Example: Hungry-Hungry-Hippo

- Two players, A and B
- Press a button as fast as you can for 5 seconds
- The winner is the person who hit their button the most number of times.

Data: Play N games:

A:	78	79	95	94	94	
B:	77	68	72	86	75	82

What is the chance that A will win the next game?

Find the mean and standard deviation:

$\bar{x}_a = 88.0$ $\quad$ $s_a = 8.69$

$\bar{x}_b = 76.6$ $\quad$ $s_b = 6.56$

Create a new variable, W = A - B. If A and B were Normal distributions

$$\mu_w = \mu_a - \mu_b$$
$$\sigma_w^2 = \sigma_a^2 + \sigma_b^2$$

A and B have estimated means and variances:

$$\bar{x}_w = \bar{x}_a - \bar{x}_b = 11.33$$
$$s_w^2 = s_a^2 + s_b^2 = 188.56$$
$$s_w = 10.89$$

Form the t-score

$$t = \left(\frac{\bar{x}_w}{s_w}\right) = \left(\frac{\bar{x}_a - \bar{x}_b}{\sqrt{s_a^2 + s_b^2}}\right) = 1.041$$

Degrees of Freedom (Wikipedia)

$$d.f. = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}} = 7.37 \approx 7$$

or, approximately, the degrees of freedom is the smaller of the degrees of freedom for A and B

$$d.f. \approx \min(5 - 1, 6 - 1) = 4$$

Convert the t-score to a probability using a t-table

$t = 1.041$

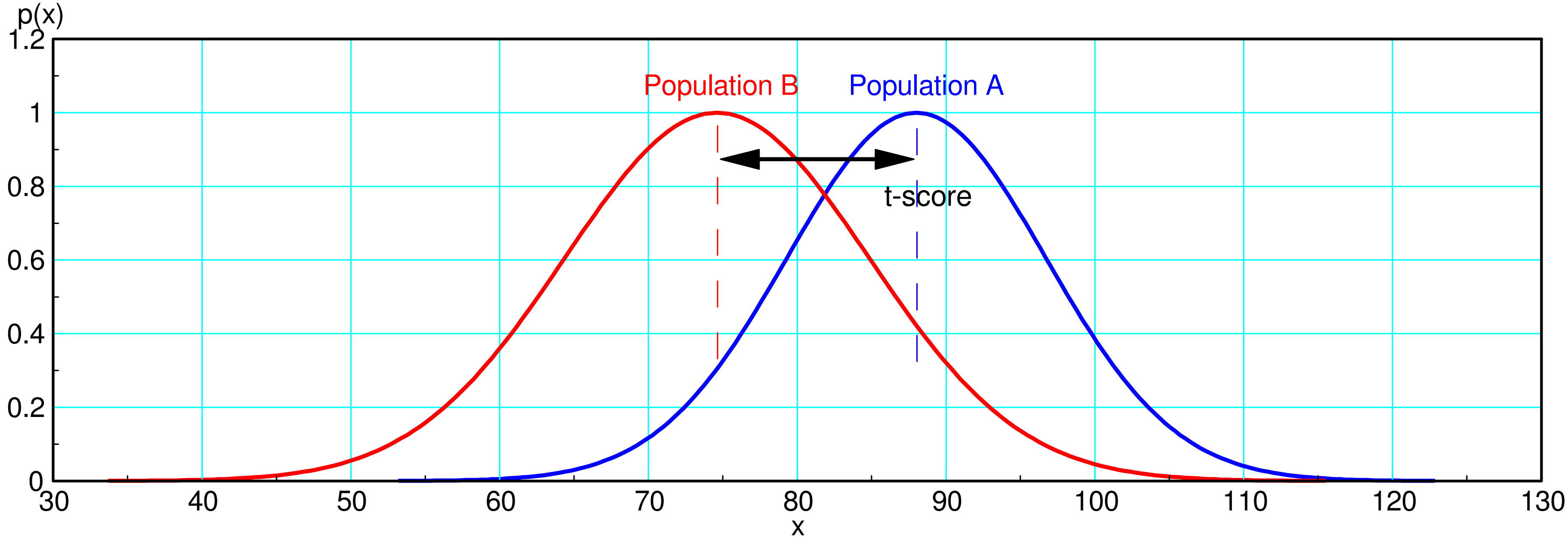
$p = 0.1634$

The tail has an area of 16.34%

- Player B has a 16.34% chance of winning the next game
- Player A has a 83.66% chance of winning the next game.

Student-t Table									
df \ p	0.001	0.0025	0.0005	0.01	0.0025	0.05	0.1	0.15	0.2
1	-636.62	-318.31	-63.66	-31.82	-12.71	-6.31	-3.08	-1.96	-1.38
2	-31.6	-22.33	-9.92	-6.96	-4.3	-2.92	-1.89	-1.39	-1.06
3	-12.92	-10.21	-5.84	-4.54	-3.18	-2.35	-1.64	-1.25	-0.98
4	-8.61	-7.17	-4.6	-3.75	-2.78	-2.13	-1.53	-1.19	-0.94
5	-6.87	-5.89	-4.03	-3.36	-2.57	-2.02	-1.48	-1.16	-0.92
6	-5.96	-5.21	-3.71	-3.14	-2.45	-1.94	-1.44	-1.13	-0.91
7	-5.41	-4.79	-3.5	-3	-2.36	-1.89	-1.41	-1.12	-0.9

Comparison of Means (Individual)



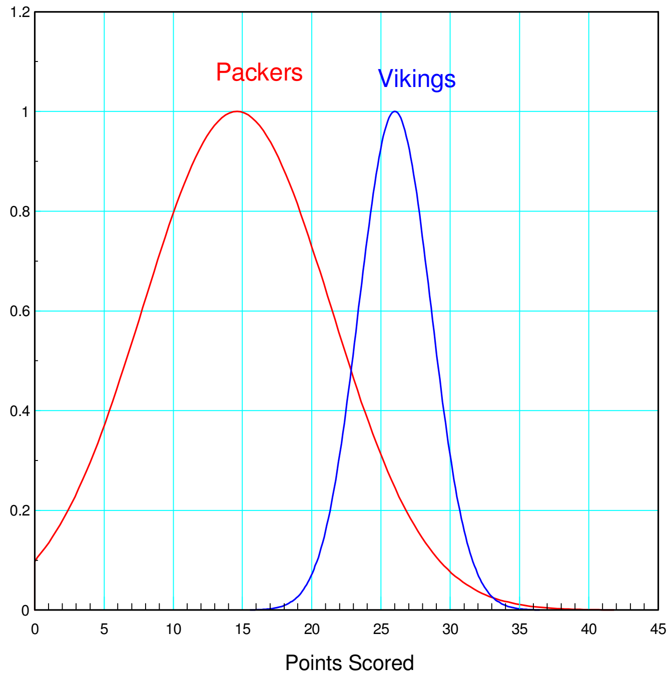
The probability that the next sample from A will be larger than the next sample from B

Handout: The number of points the Vikings scored in the first six weeks are:

- A: Points Scored by Vikings: {23, 24, 28, 28, 29, 24}
- Mean = 26.00
- Standard Deviation = 2.61

The number of points the Packers scored in the first five weeks are:

- B: Points Scored by Packers: {7, 14, 22, 21, 9}
- Mean = 14.60
- Standard Deviation = 6.80



What is the probability that the Vikings will beat the Packers the next time they play?

- Vikings win a one-game series

Solution: Create a new variable, W

W = A - B

$$\bar{x}_w = \bar{x}_a - \bar{x}_b = 11.4$$

$$s_w^2 = s_A^2 + s_B^2 = 53.05$$

$$s_w = 7.28$$

Find the probability W>0 (A>B)

$$t = \left(\frac{11.4-0}{7.28}\right) = 1.565$$

Degrees of freedom

$$d.o.f. \approx \min(5,4) = 4$$

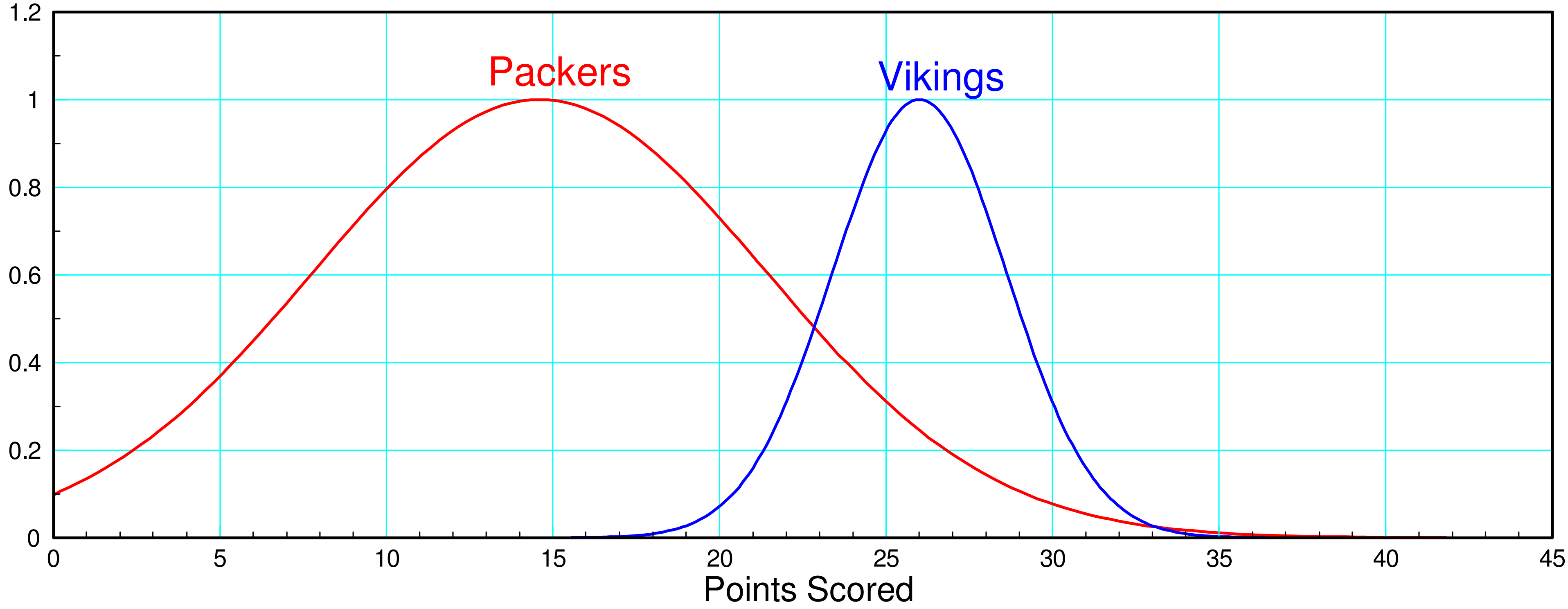
Student-t Table									
df \ p	0.001	0.0025	0.0005	0.01	0.0025	0.05	0.1	0.15	0.2
1	-636.62	-318.31	-63.66	-31.82	-12.71	-6.31	-3.08	-1.96	-1.38
2	-31.6	-22.33	-9.92	-6.96	-4.3	-2.92	-1.89	-1.39	-1.06
3	-12.92	-10.21	-5.84	-4.54	-3.18	-2.35	-1.64	-1.25	-0.98
4	-8.61	-7.17	-4.6	-3.75	-2.78	-2.13	-1.53	-1.19	-0.94
5	-6.87	-5.89	-4.03	-3.36	-2.57	-2.02	-1.48	-1.16	-0.92

Result:

p = 10% (area of tail)

The Packers have a 10% chance of winning the next game

The Vikings have a 90% chance of winning the next game



Example: Hungry-Hungry Hippo (take 2)

- Two players, A and B
- Press a button as fast as you can for 5 seconds
- The winner is the person who hit their button the most number of times.

Data: Play N games:

A:	78	79	95	94	94	
B:	77	68	72	86	75	82

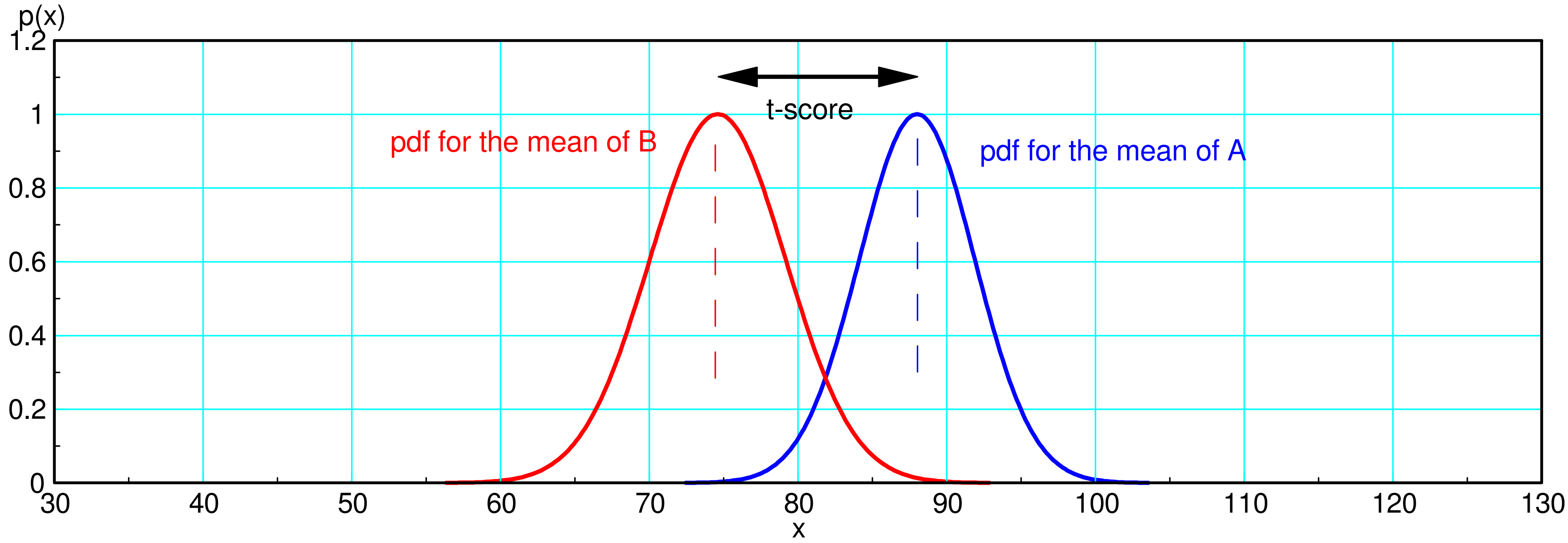
What is the chance that A has a higher average then B?

- What is the chance that A will win an infinite-game match?

Comparison of Means (population)

Essentially,

- The previous question was who would win the next game of hungry-hungry hippo.
- This question is who would win a match with an infinite number of games?



Slightly different question: Which population has the larger mean?

Assume first that A and B are normally distributed. The statistic $\bar{x}$ then has a normal distribution

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

As the sample size goes to infinity, the estimate of the mean converges to the true mean.

If you want to compare two means, μ_a and μ_b, then create a new variable:

$$W = \mu_a - \mu_b \qquad W \sim N(\mu_w, \sigma_w^2)$$

where

$$\mu_w = \mu_a - \mu_b$$

$$\sigma_w^2 = \frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}$$

Now, suppose you estimate the variance, creating a student-t distribution. Then

$$W = \bar{x}_a - \bar{x}_b$$

will have a student t-distribution with

$$s_w^2 = \frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}$$

The t-score is then

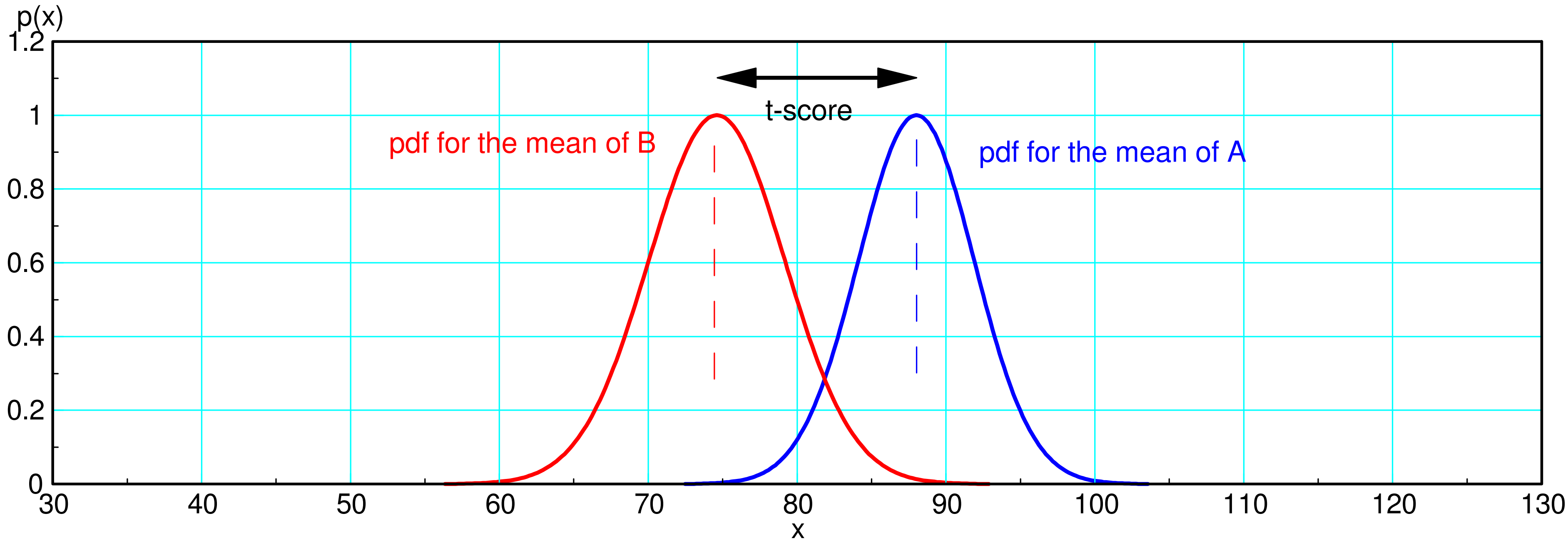
$$t = \left(\frac{\bar{x}_w}{s_w}\right) = \left(\frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}\right) = 2.4012$$

Degrees of freedom are the same as before (7)

- or apprixImately min(4, 5) = 4
- doesn't affest the results that much

Use a t-table to convert this to a probability

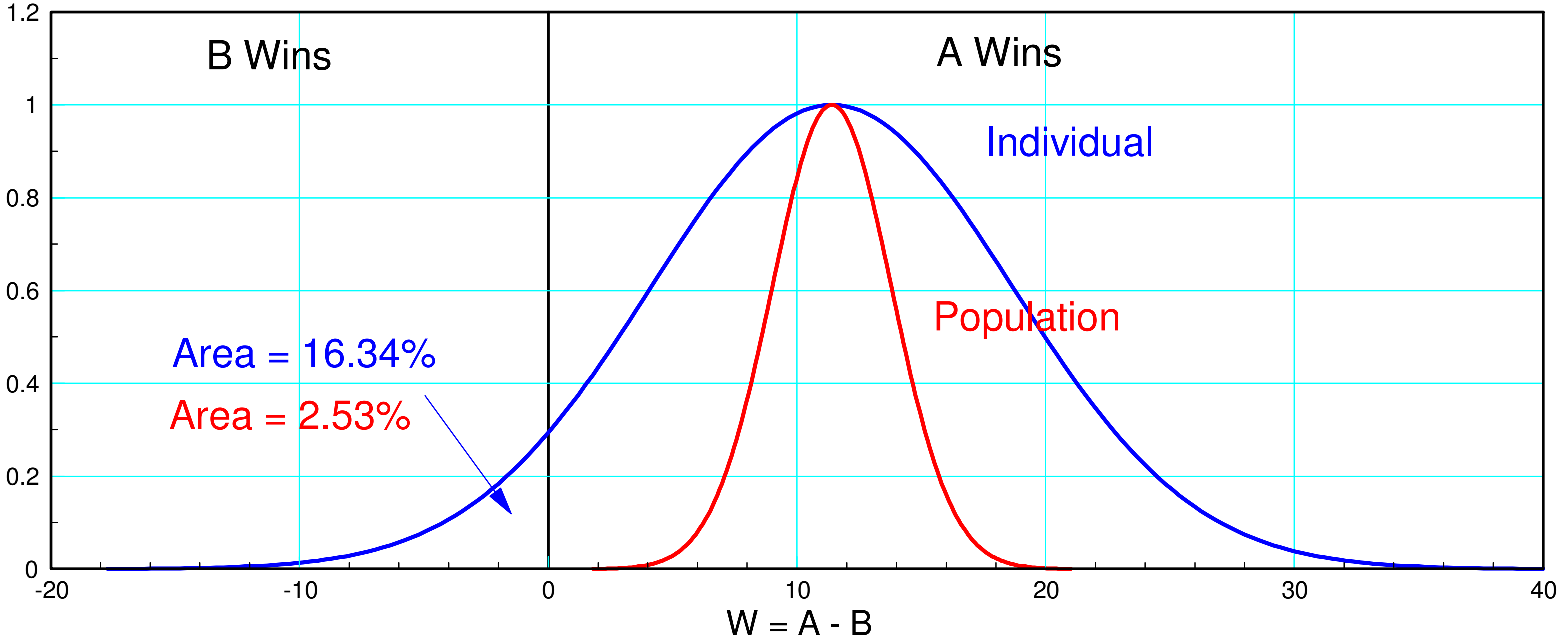
- t = 2.4012
- 7 degrees of freedom, p = 0.9763
- 4 degrees of freedom, p = 0.9629



The probability that the mean of A is greater than the mean of B
t-score = 2.4012, p = 0.9763

Note: You know more about populations than individuals:

- B has a 16.34% chance of winning any given game (previous calculation)
- B only has a 2.53% chance of winning a match



How large does the sample size have to be for being 99.5% certain?

- t-score for 99.5% certain = 3.499
- (assume 7 degrees of freedom)

The t-score is

$$t = 3.499 = \left(\frac{\bar{x}_w}{s_w}\right) = \left(\frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}\right)$$

If we assume na = nb = n

$$3.499 = \left(\frac{\bar{x}_a - \bar{x}_b}{\sqrt{s_a^2 + s_b^2}}\right)\sqrt{n} = (1.054)\sqrt{n}$$

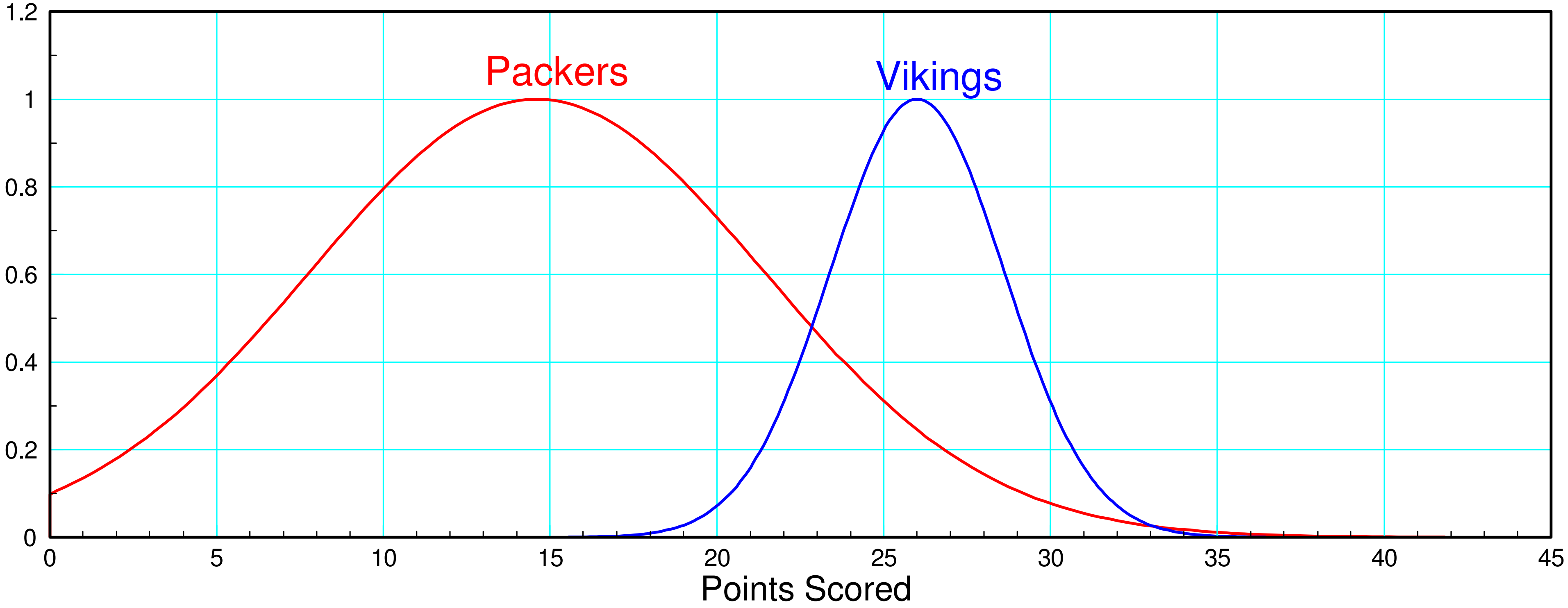
$$n = 11.02$$

Round up: n = 12

- A little conservative: > 7 d.o.f. now
- In theory, with a large enough sample size, you can detect minute differences

Handout: What is the probability that the Vikings are the better team?

- Vikings win an infinite-game series
- Vikings: $\bar{x}_a = 26,\ s_a = 2.61,\ n_a = 6$
- Packers: $\bar{x}_b = 14.6,\ s_b = 6.80,\ n_b = 5$



Solution: Create a new variable, W

- Population question so divide the variance by the sample size

W = A - B

$$\bar{x}_w = \bar{x}_a - \bar{x}_b = 11.4$$

$$s_w^2 = \frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b} = 10.38$$

$$s_w = 3.22$$

Find the probability W > 0 (A > B) (4 d.o.f.)

$$t = \left(\frac{\bar{x}_w - 0}{s_w}\right) = \left(\frac{11.4}{3.22}\right) = 3.54$$

Student-t Table									
df \ p	0.001	0.0025	0.0005	0.01	0.0025	0.05	0.1	0.15	0.2
1	-636.62	-318.31	-63.66	-31.82	-12.71	-6.31	-3.08	-1.96	-1.38
2	-31.6	-22.33	-9.92	-6.96	-4.3	-2.92	-1.89	-1.39	-1.06
3	-12.92	-10.21	-5.84	-4.54	-3.18	-2.35	-1.64	-1.25	-0.98
4	-8.61	-7.17	-4.6	-3.75	-2.78	-2.13	-1.53	-1.19	-0.94
5	-6.87	-5.89	-4.03	-3.36	-2.57	-2.02	-1.48	-1.16	-0.92

Example: Hungry-Hungry Hippo (take 3)

- Three players, A, B, and C
- Press a button as fast as you can for 5 seconds
- The winner is the person who hit their button the most number of times.

Data: Play N games

A:	78	79	95	94	94		
B:	77	68	72	86	75	82	
C:	85	90	87	77	83	72	89

What is the chance that A will win a given game?

What is the chance that A has the highest mean?

Summary

A Student t-Test is a test of a mean

- Which is designed for a single population

You *can* use a t-Test with two populations

- Create a new variable: W = A - B
- mean(W) = mean(A) - mean(B)
- var(W) = var(A) + var(B)
- dof(W) ≈ min(dof(A), dof(B))

You now have a single population

With 3+ populations, a Student t-Test doesn't really work

- You need to use a different statistical test