Math 361

Inv. 1.7 – Power

Last time – the Binomial Test

Research question involves parameter π from a Binomial Process

H_0: π = some number

H_a: π ≠ some number

Collect a binary variable from a sample of size n

Verify that the data collection is modelled well by a binomial process

Compute a binomial p-value, either through simulating a coin toss or the exact formula for a Binomial probability, assuming
π = some number

If p-value is large,
there's no evidence against **H_0**.
If p-value is small,
there's evidence against **H_0**.

Using a p-value to make a conclusion

A **p-value** is the probability of seeing a sample results at least as extreme as our sample result assuming that the null hypothesis H_0 is true.

$P(X \geq 21) = 0.001$ $\pi = 0.5$, $n = 26$

If we see a ***small*** p-value (say less than 0.05) then we conclude *H_0 must not be true*.

people aren't just randomly choosing tim on left.

Could we be making a mistake?

Consider a trial by jury

What do we assume about the defendant? not guilty = null model is true

How do we decide whether to reject this assumption?

Consider a trial by jury

null

Alternative

Ha

What do we assume about the defendant? H_0 : innocent

How do we decide whether to reject this assumption?

P-value = weight of evidence presented during the trial

Could the jury make a mistake?

Consider a trial by jury

H_0 : innocent

P-value = weight of evidence presented during the trial

Could the jury make a mistake in their decision based on the p-value?

		Reality	
		Defendant is innocent	Defendant is guilty
Decision of the Jury	Defendant is guilty	I	
	Defendant is innocent		II

individual

public

Consider a trial by jury

H_0 : innocent

P-value = weight of evidence presented during the trial

Could the jury make a mistake in their decision based on the p-value?

		Reality	
		Defendant is innocent	Defendant is guilty
Decision of the Jury	Defendant is guilty	wrong	Correct
	Defendant is innocent	Correct	wrong

Consider a trial by jury as a test of significance

H_0 : innocent

P-value = weight of evidence presented during the trial

Could the jury make a mistake in their decision based on the p-value?

		Reality	
		H_0 : is true	H_a : is true
Decision of the Jury	H_a : is true	Type I Error	Correct
	H_0 : is true	Correct	Type II Error

Inv. 1.7 – Improved Baseball Player

- A baseball player who has been a 0.250 hitter suddenly improves over one winter to the point where he is now a 0.333 hitter.
- In order to get a raise in his salary, he needs to convince his manager that he really has improved.

Let's set this up as a test of significance and that his manager will compute a p-value to determine if the player improved.

Inv. 1.7: parts a, b,

Part a) Define the parameter of interest using appropriate notation

π = probability gets a hit

Part b) State the null and alternative hypotheses

$H_0: \pi = 0.250$

$H_a: \pi > 0.250$

Inv. 1.7: parts a, b,

Part a) Define the parameter of interest using appropriate notation

π = probability that the player gets a hit.

Part b) State the null and alternative hypotheses using appropriate notation using the manager's perspective

$H_0 : \pi = 0.250$

$H_a : \pi > 0.250$

Inv. 1.7 part c

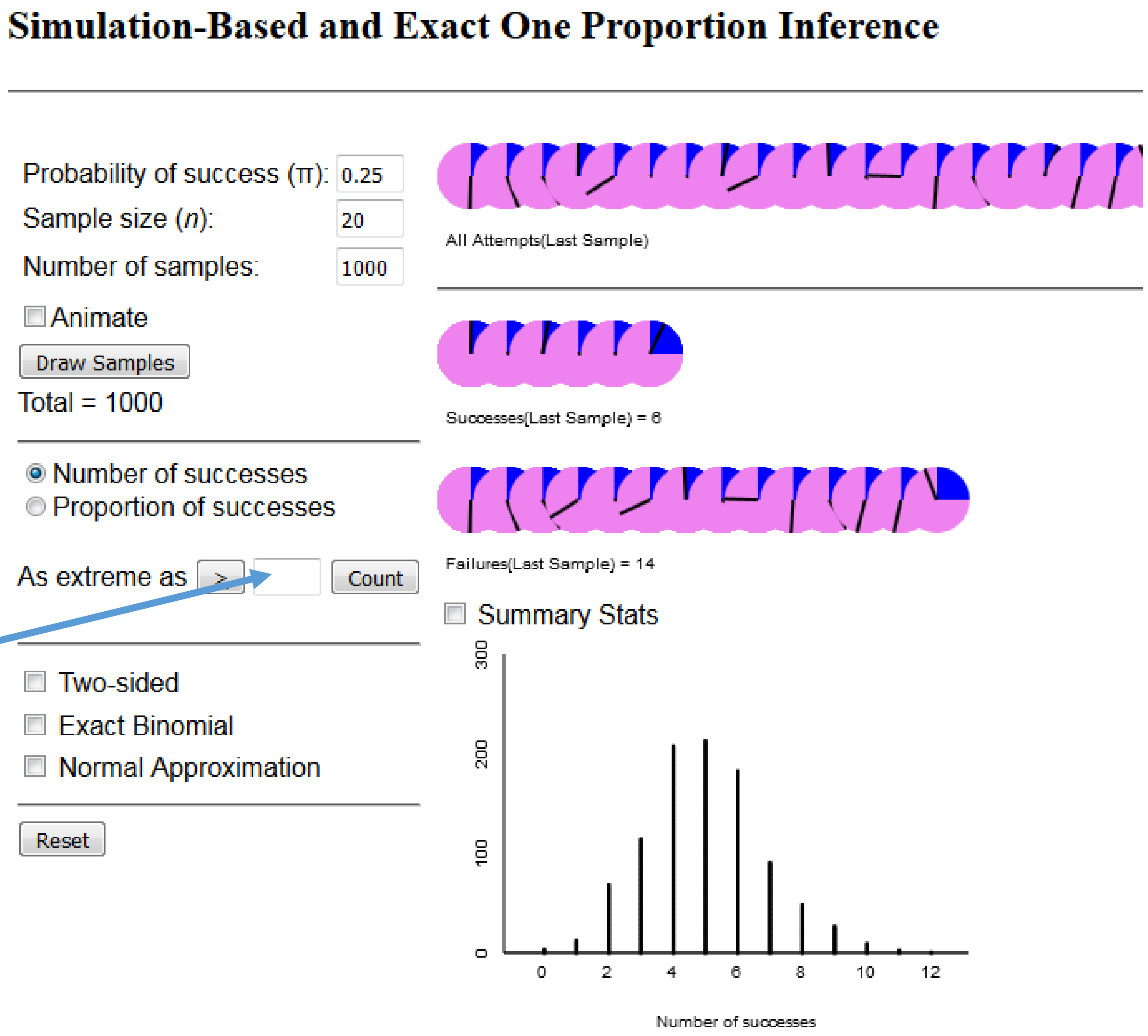
Suppose the manager decides to give the player 20 at-bats in which to prove his improvement.

How many hits would the player need to make out of 20 at-bats in order to convince the manager that he has improved to the point of being in the top 5% of 0.250 hitters?

In statistics jargon, what is the ***rejection region*** for the null hypothesis that corresponds to a ***level of significance*** of 0.05?

Inv. 1.7 part (e): How many hits would the player need to make out of 20 at-bats in order to convince the manager that he has improved to the point of being in the top 5% of 0.250 hitters?

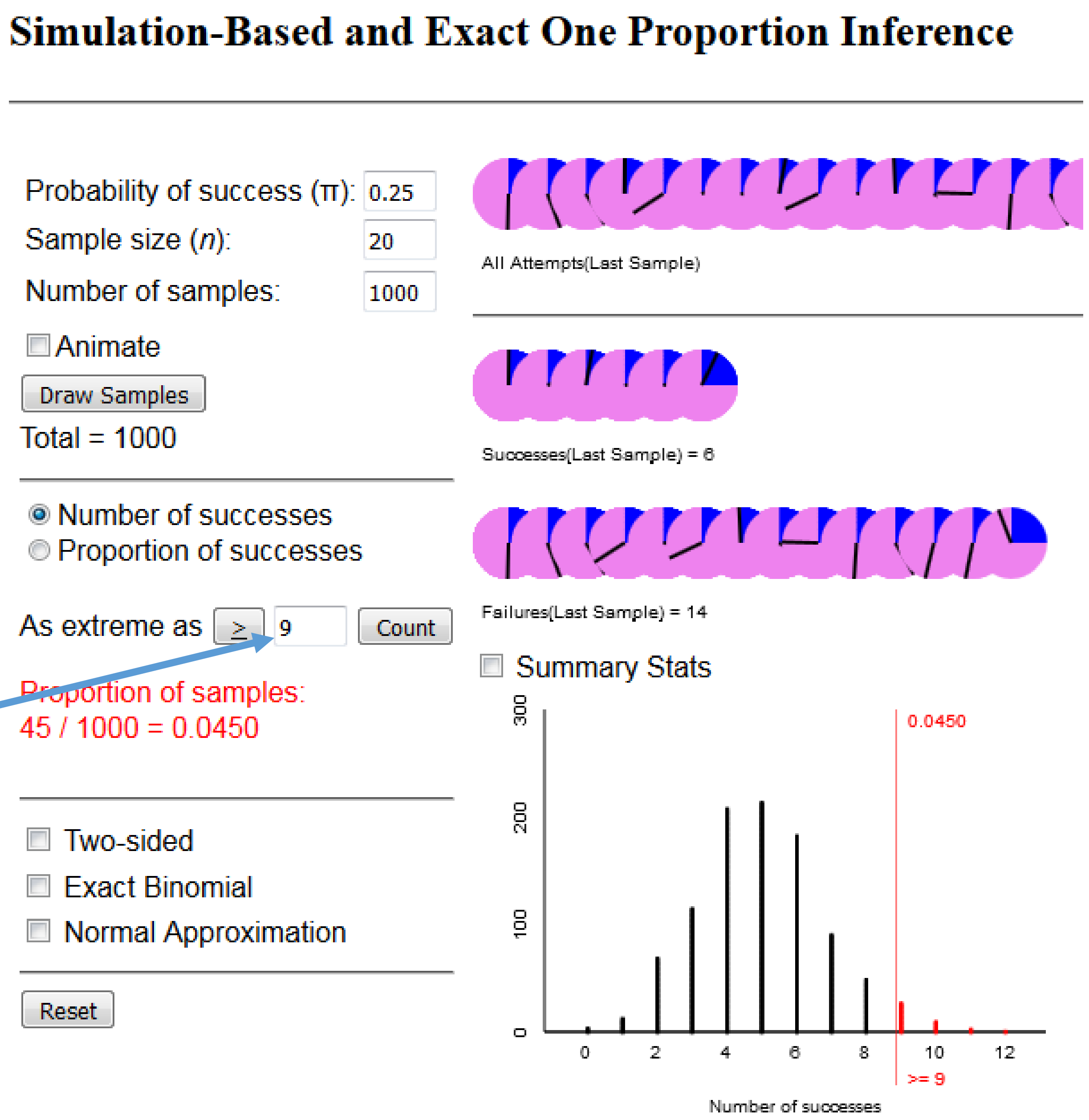
How many hits should we put in the box so that the probability is no more than 0.05?



Inv. 1.7 part (e): How many hits would the player need to make out of 20 at-bats in order to convince the manager that he has improved to the point of being in the top 5% of 0.250 hitters?

9 hits

If we put 9 in the box so then the probability is no more than 0.05



Terminology

Rejection Region: the values of the statistic that correspond to rejecting the null hypothesis.

Example: the rejection region is **hits ≥ 9**

*If the manager observes **more than 9** hits he will reject the idea that the player is a typical 0.250 hitter and in fact is in the top 5%.*

More terminology

Level of significance (notation α) is the value such that

- If p-value ≤ α we "reject" $\mathbf{H_0}$
- If p-value > α we "fail to reject" $\mathbf{H_0}$

Example: α = 0.05

the manager rejected H_0*:π = 0.250 if the player appeared to be in the top 5%*

Back to the types of errors

Recall that the player really did improve and is now a 0.333 hitter.

Which type of error is the player worried about the manager making?

I player didn't improve but manager believes he did

		Reality	
		$H_0: \pi = 0.250$ is true	$H_a: \pi > 0.250$ is true
Decision of the Jury	$H_a: \pi > 0.250$ is true	Type I Error	Correct
	$H_0: \pi = 0.250$ is true	Correct	Type II Error

power

II player improved but manager doesn't believe it

Inv. 1.7: Power

power + β = 1
β = 1 − power

Probability of a Type II error:

the probability of incorrectly rejecting H_0

Notation: β

Power: probability of correctly rejecting H_0 when H_a is true

Notation: $1-\beta$

The player wants to minimize the manager's probability of a type II error and therefore maximize power.

How can we compute the **Power** of a test?

Power: probability of correctly rejecting H_0 when a specific H_a is true

1. Find the rejection region for a given level of significance.
2. Simulate the distribution assuming the *alternative* hypothesis is true.
3. Compute the probability of the rejection region assuming H_a is true.

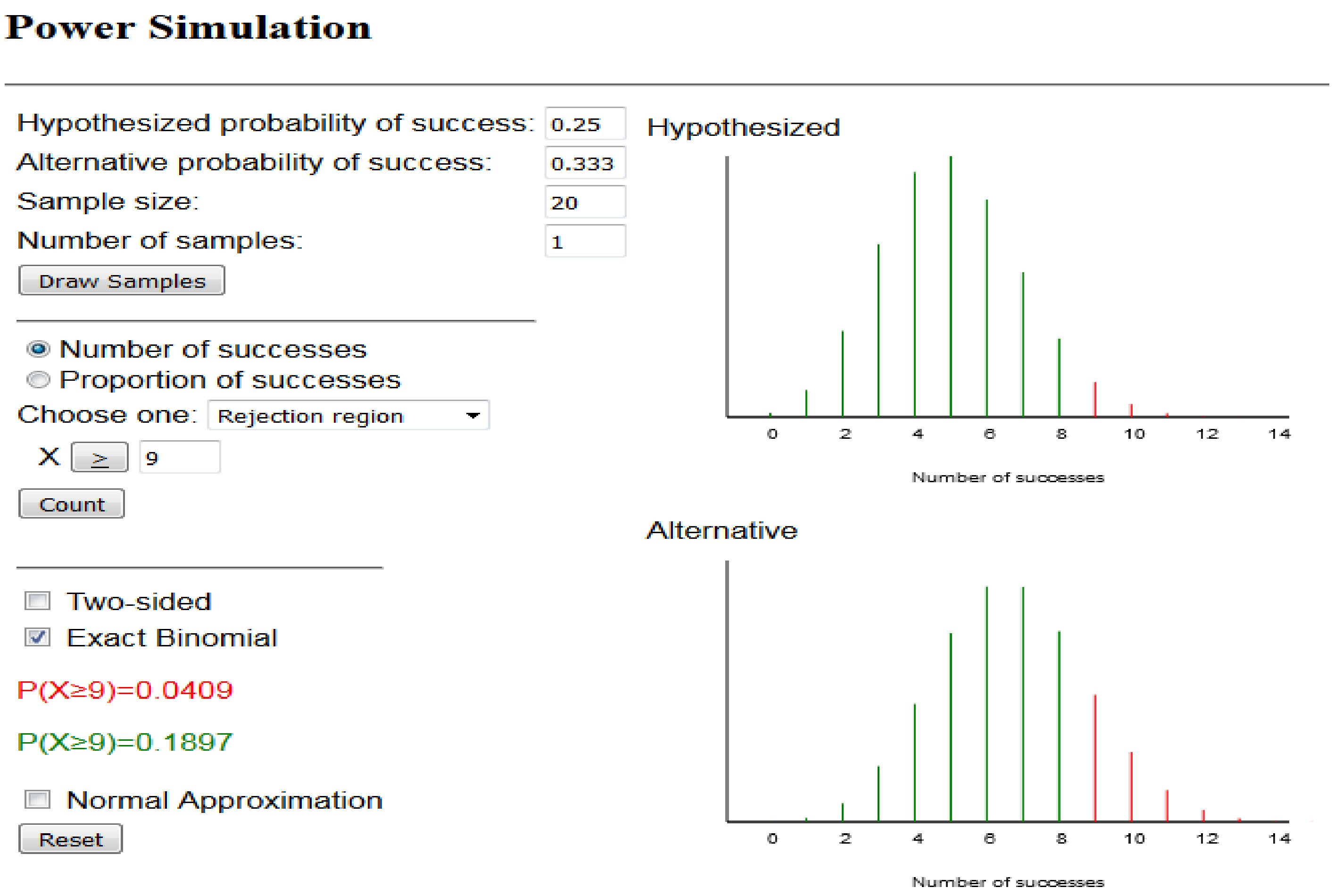
Sound hard? It's easy when you use an applet!

Power Simulation Applet (batting averages)

Power is given in green

The probability that the manager correctly decides the 0.333 player improved is 0.19…

…not very likely.

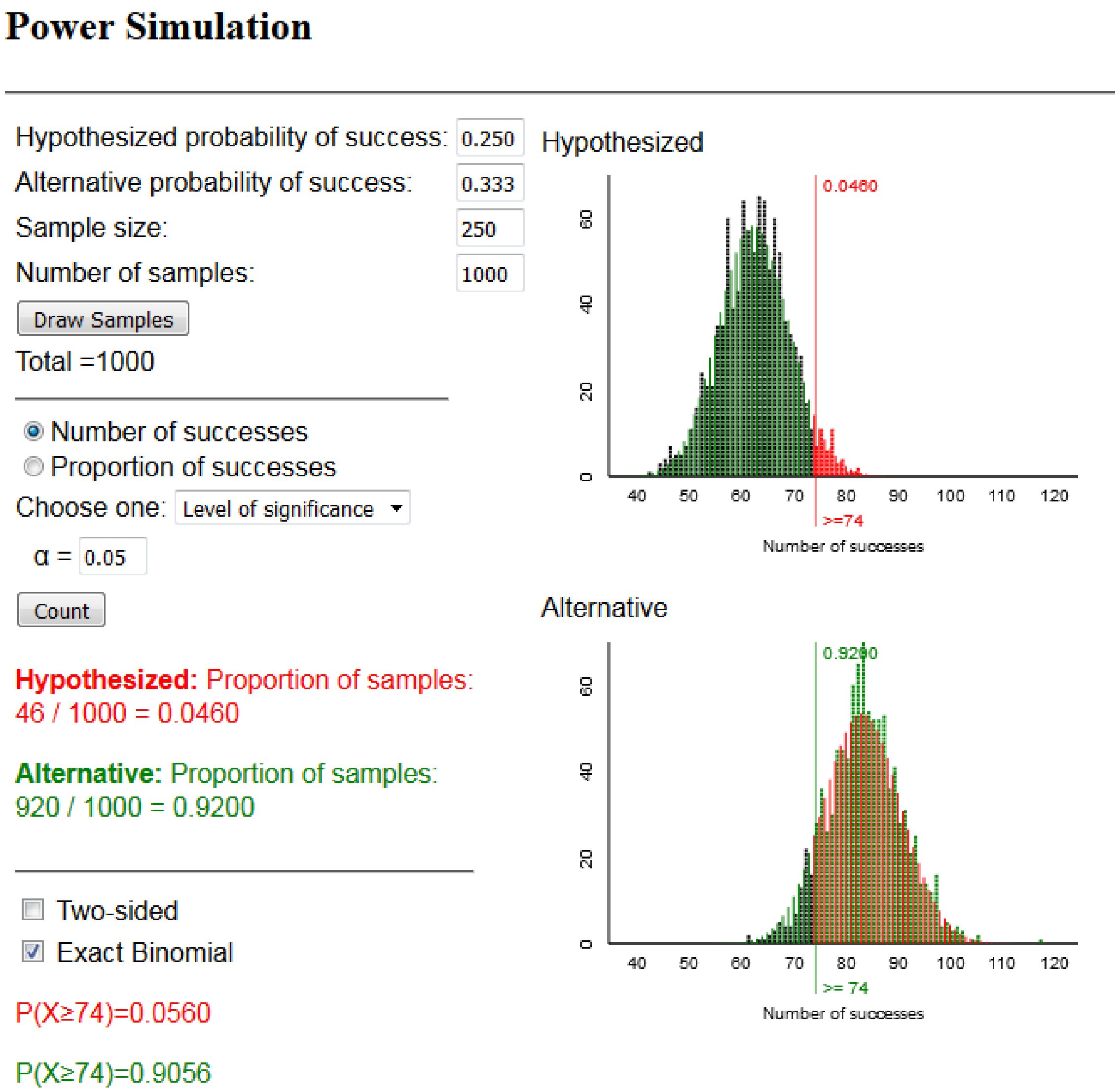


How can **power** be improved?

Try increasing number of at-bats the manager observes.

If the manager watches 250 at-bats then the probability he will decide the 0.333 hitter has improved is about 0.91…

… much better from the player's perspective



What is the probability of a **Type I error**?

The manager is worried about making a **Type I error**, that is, deciding the player improved when really he hasn't.

The probability Type I error is set by choosing a cutoff value so that if the p-value is below this value then the null hypothesis will be rejected.

This value is called the **level of significance** and is notated α.

The manager could decrease the probability of a type I error by using $\alpha = 0.01$ instead of $\alpha = 0.05$.

Controlling Type I & II error

Control the probability of Type I error by fixing the level of significance α:

If you only reject H_0 when the p-value is less than α then the probability of making a type I error is at most α

Control the probability of Type II error (β) **by adjusting your study design**:

Design your study so that power = 1-β is high (close to 1).

Try increasing sample size or decreasing α

Try practice problem 1.7D on page 61

Practice Problem 1.7D

Suppose you want to test a person's ability to discriminate between two types of soda. You fill one cup with Soda A and two cups with Soda B. The subject tastes all 3 cups and is asked to identify the odd soda. You record the number of correct identifications in 10 attempts. Assume a one-sided alternative.

(a) If the subject's actual probability of a correct identification is 0.50, what is the power of this test for a level of significance of $\alpha = 0.50$? [*Hint*: What is the null hypothesis?]

(b) Write a one-sentence interpretation of the power you calculated in (a) in context.

(c) What is the power if you give the subject 20 attempts?