Stat310

More tests

Hadley Wickham

1. Revision
2. Example
3. T-test

Assessment

Assessment left

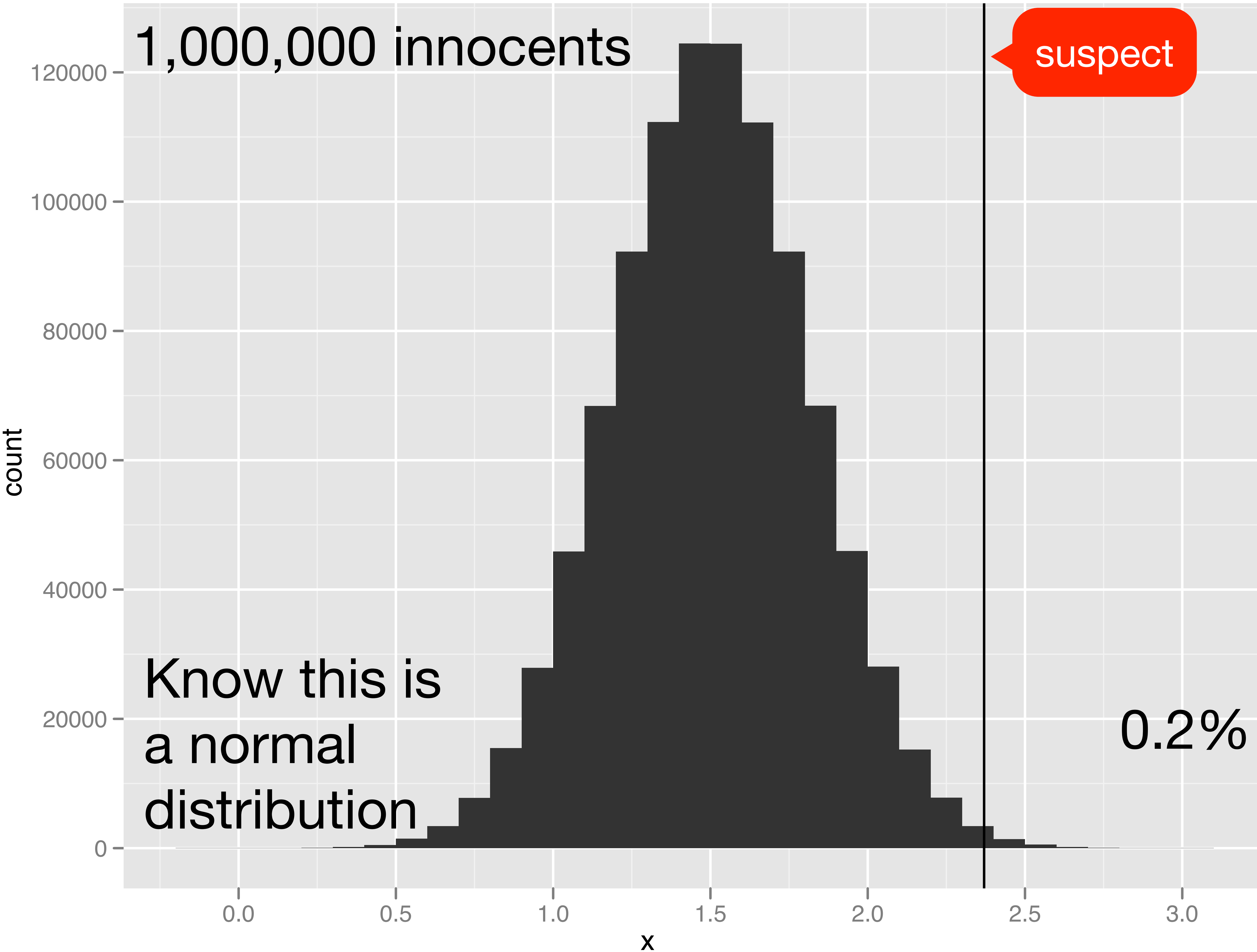
- Stats in real life (bonus): due May 2
- Stats in practice: due May 2
  (Revised grading will be up on Friday.
  All regrades in owlspace)
- Final: due May 2
  (available April 25. Now 3 hours. Info online)

Study help sessions

- M, T, W, R, F?
- Morning, afternoon, evening?
- http://www.doodle.com/8m6s9w82ku9s5avr

Revision

A **suspect** is accused of a **crime**. The suspect is declared guilty or not guilty based on a **trial**. Each trial has a **defence** and a **prosecution**. On the basis of how **evidence** compares to a **standard**, the judge makes a decision to **convict** or **acquit**.


1,000,000 innocents
suspect
Know this is
a normal
distribution
0.2%
count
x
0
20000
40000
60000
80000
100000
120000
0.0
0.5
1.0
1.5
2.0
2.5
3.0

1. Write down null and alternative hypotheses (positions of defence and prosecution)

2. Figure out good test statistic (what numeric summary captures useful evidence)

3. Work out null distribution (distribution of innocents)

4. Calculate p-value by comparing actual value to null distribution (what proportion of true innocents look more guilty than the suspect)

5. Reject H_o if p-value smaller than cutoff

Hypotheses

Null hypothesis (H_o)
Always of form $\mu = 0$

Alternative hypothesis (H_a / H_1)
One sided: $\mu < 0$, $\mu > 0$
Two sided: μ != 0

X_i iid, and n large:

$$\frac{\bar{X}_n - E(\bar{X})}{sd(\bar{X})} \dot{\sim} Z$$

$$\frac{\bar{X}_n - E(\bar{X})}{\hat{sd}(\bar{X})} \dot{\sim} Z$$

$$X_i \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$$

$$\frac{\bar{X}_n - E(\bar{X})}{sd(\bar{X})} \sim Z$$

$$\frac{\bar{X}_n - E(\bar{X})}{\hat{sd}(\bar{X})} \sim t_{n-1}$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

P-value

Standardised measurement of evidence.

Low p-value = low probability of innocent looking this guilty = **reject the null**

High p-value = high probability of innocent looking this guilty = **don't reject**

Don't need to know anything else about the test!

	Reject H_0	Fail to reject H_0
H_0 false	Correct	**Type II error**
H_0 true	**Type I error**	Correct

Rates

For a given test,

P(false conviction) = significance level

P(true conviction) = power

Example

Grade difference

I'm interested in whether or not there is a difference between this year's average stat310 grade and last year's average grade.

Unfortunately owlspace is broken and I can only access 20 randomly selected grades from each year.

Is there a difference in the average grade?

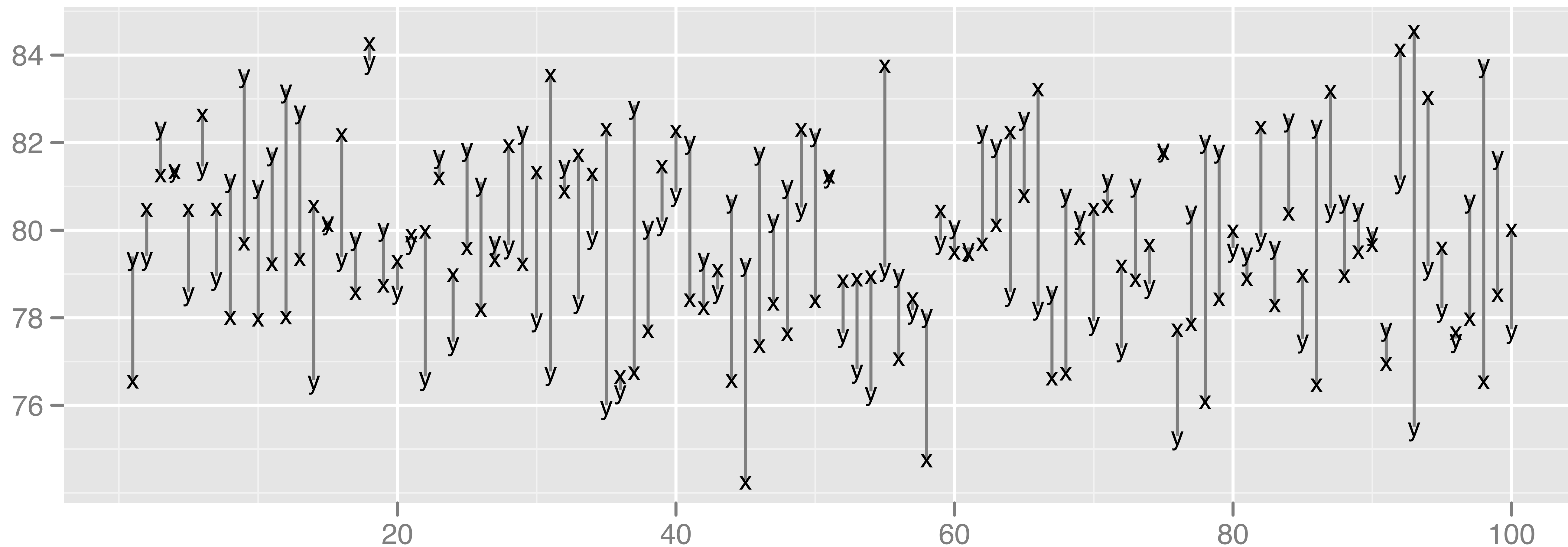
2010 $$X_i \overset{iid}{\sim} \text{Normal}(\mu_x, 80)$$

2011 $$Y_i \overset{iid}{\sim} \text{Normal}(\mu_y, 80)$$

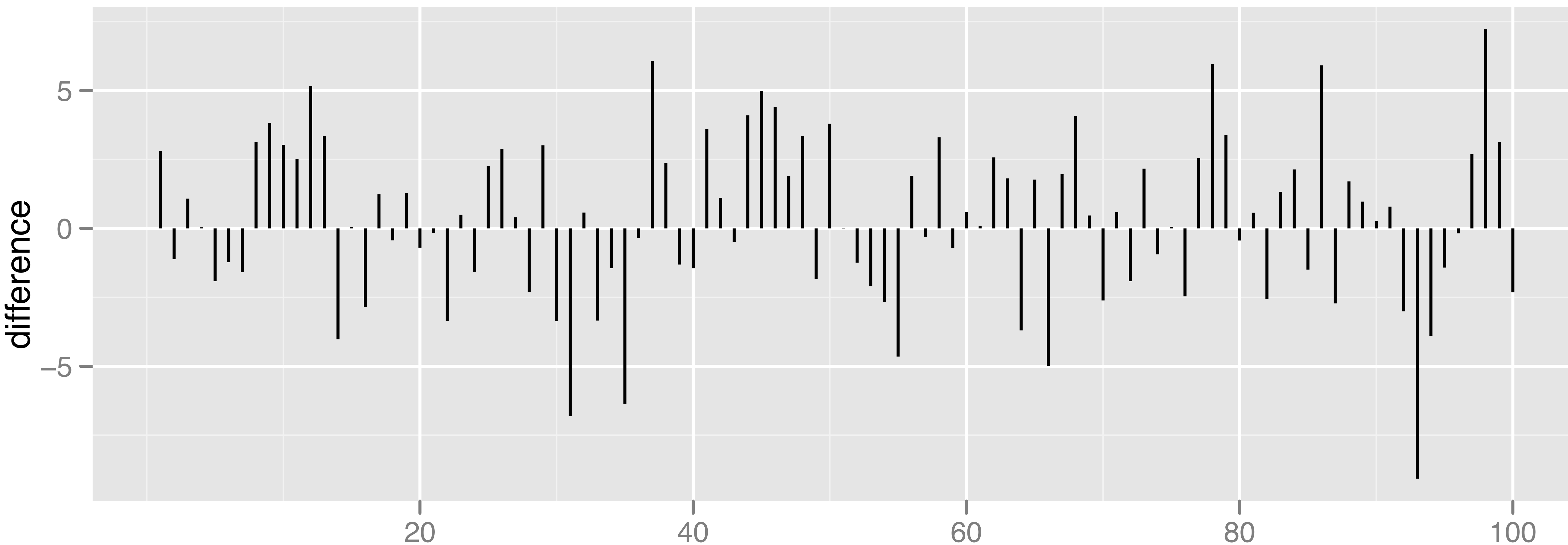
$$i = 1, \cdots, 20$$

Null hypothesis

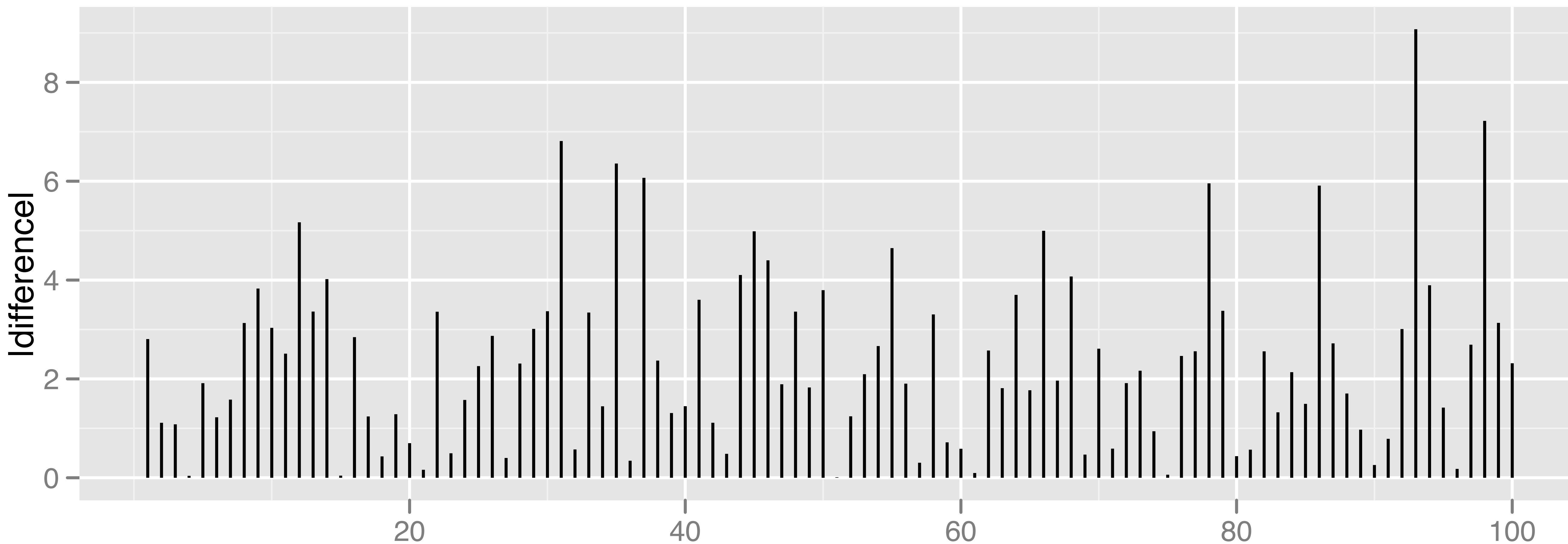
$\mu_x = \mu_y = 80$



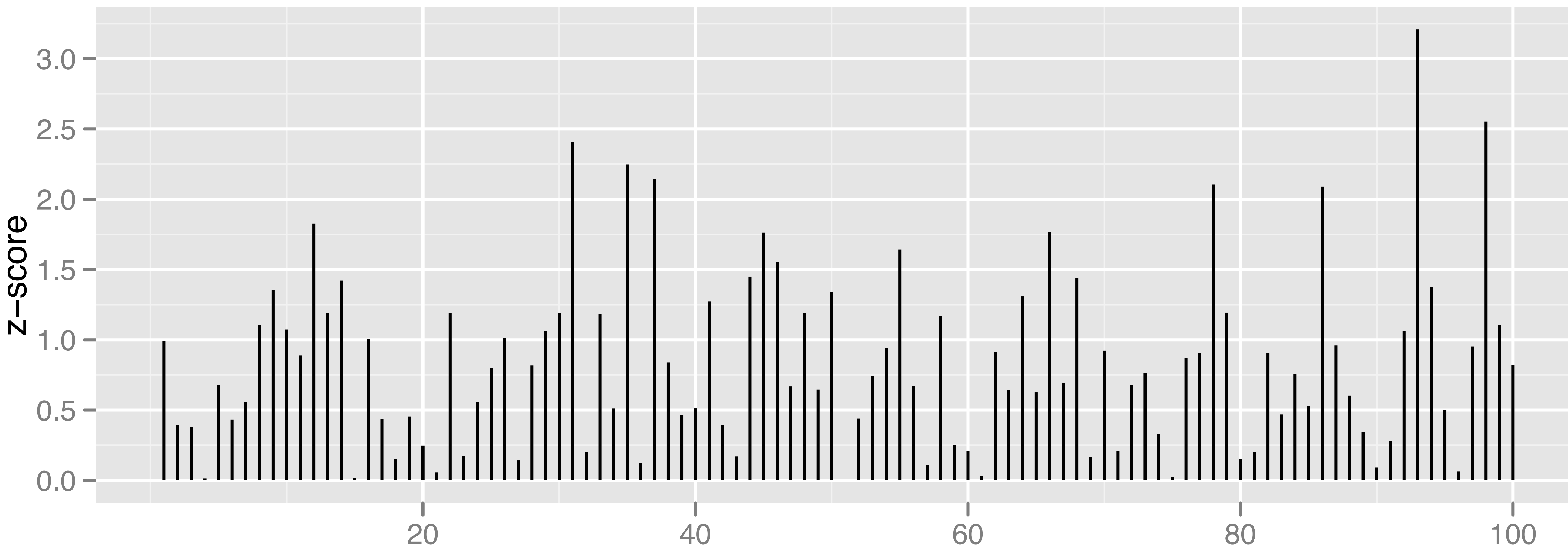
$\mu_x = \mu_y = 80$



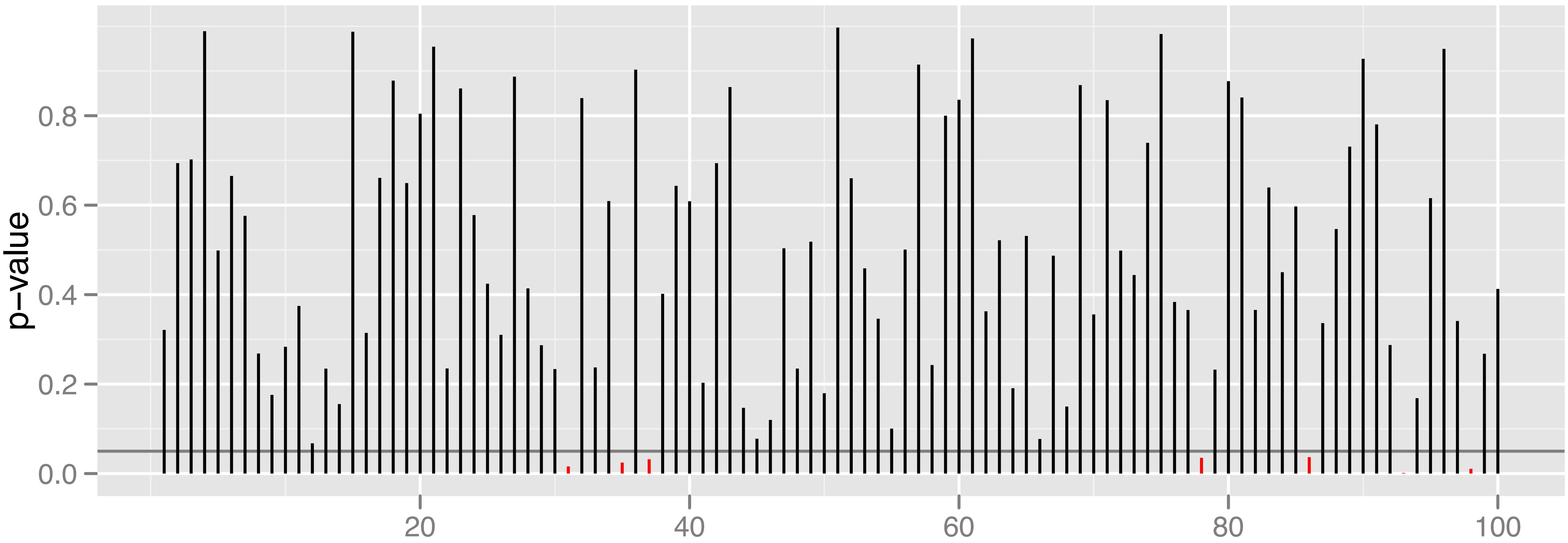
$\mu_x = \mu_y = 80$



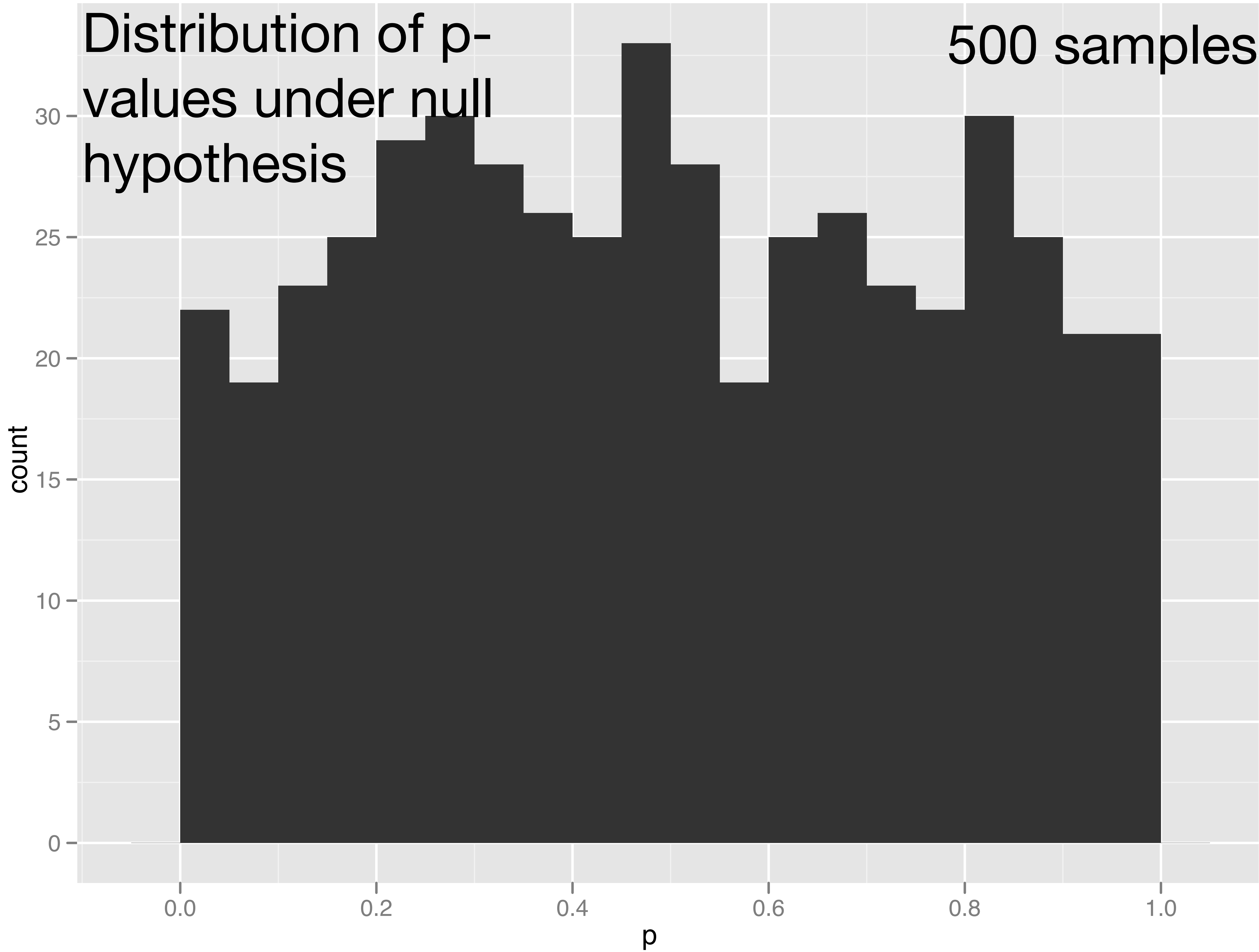
$\mu_x=\mu_y=80$



$\mu_x = \mu_y = 80$



Incorrectly reject null 7% of the time


Distribution of p-values under null hypothesis
500 samples
count
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
p

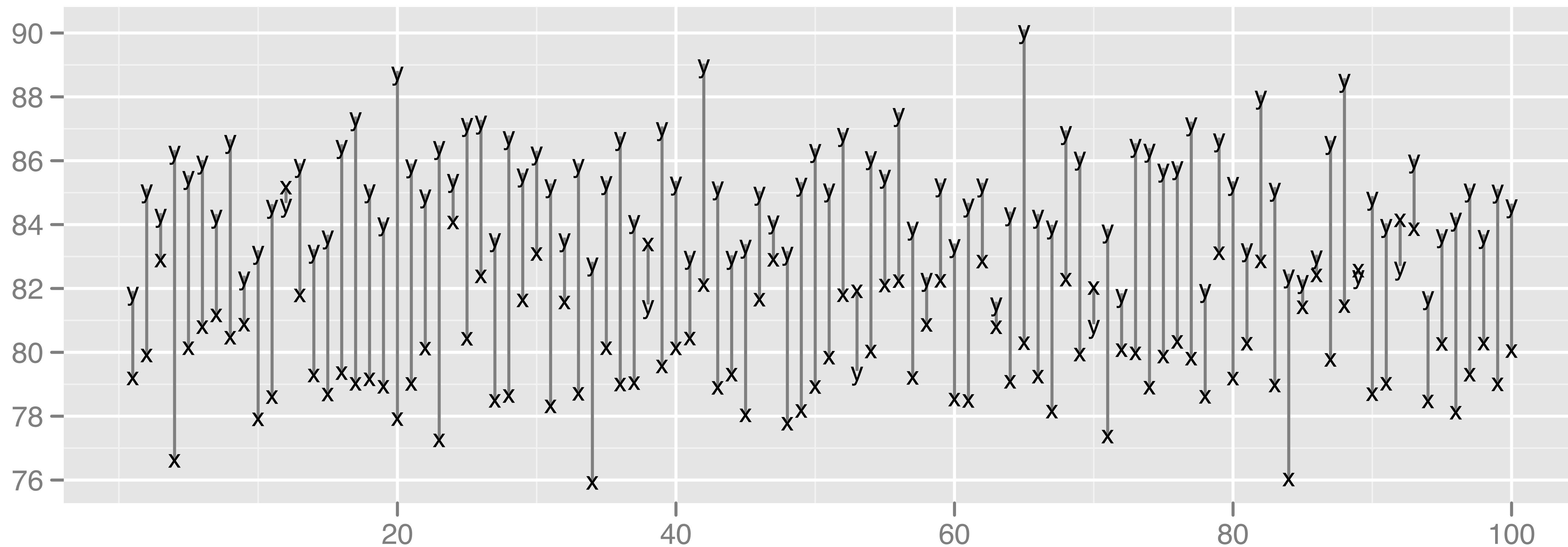
Your turn

Under the null hypothesis, what's the distribution of the p values?

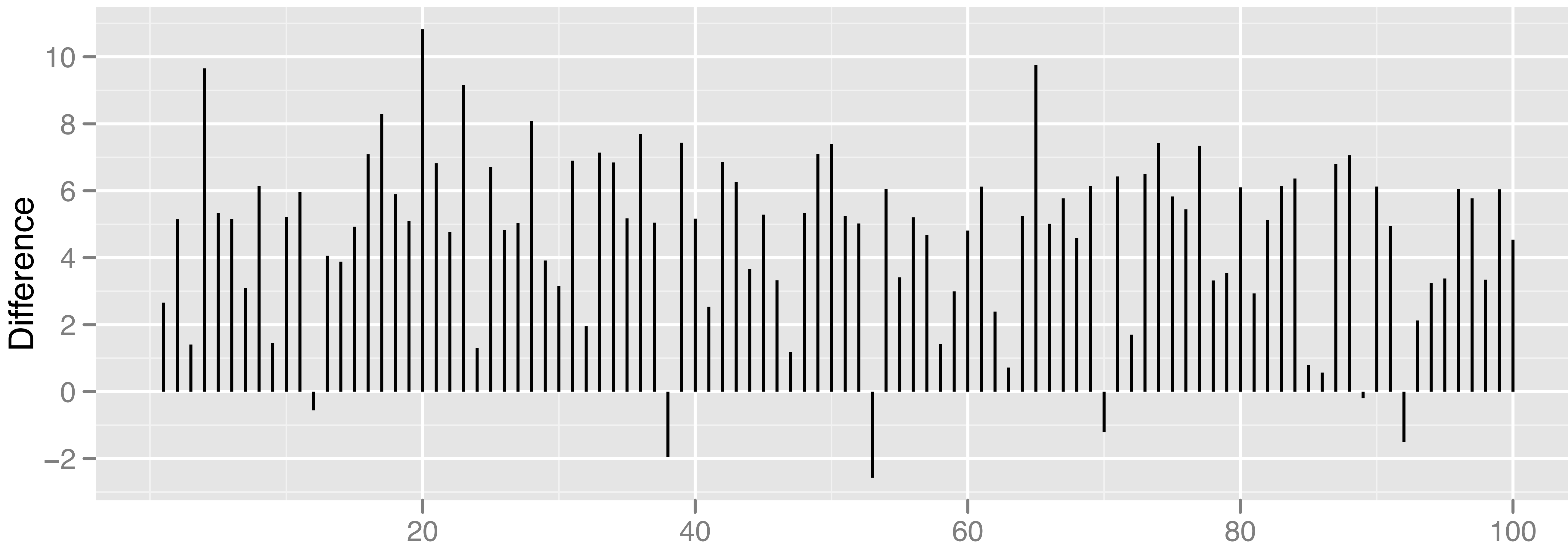
(Think about the probability that the p-value is less than 0.05, 0.1, 0.2, etc)

Alternative hypothesis

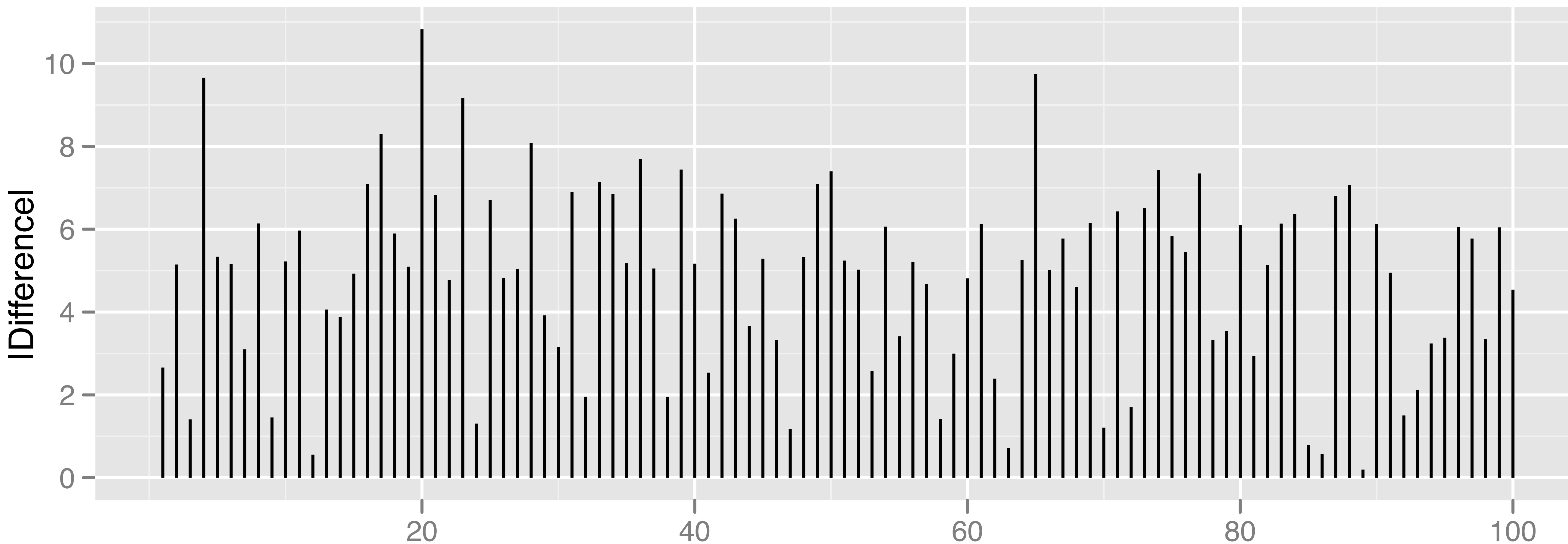
$\mu_x=80,\ \mu_y=85$



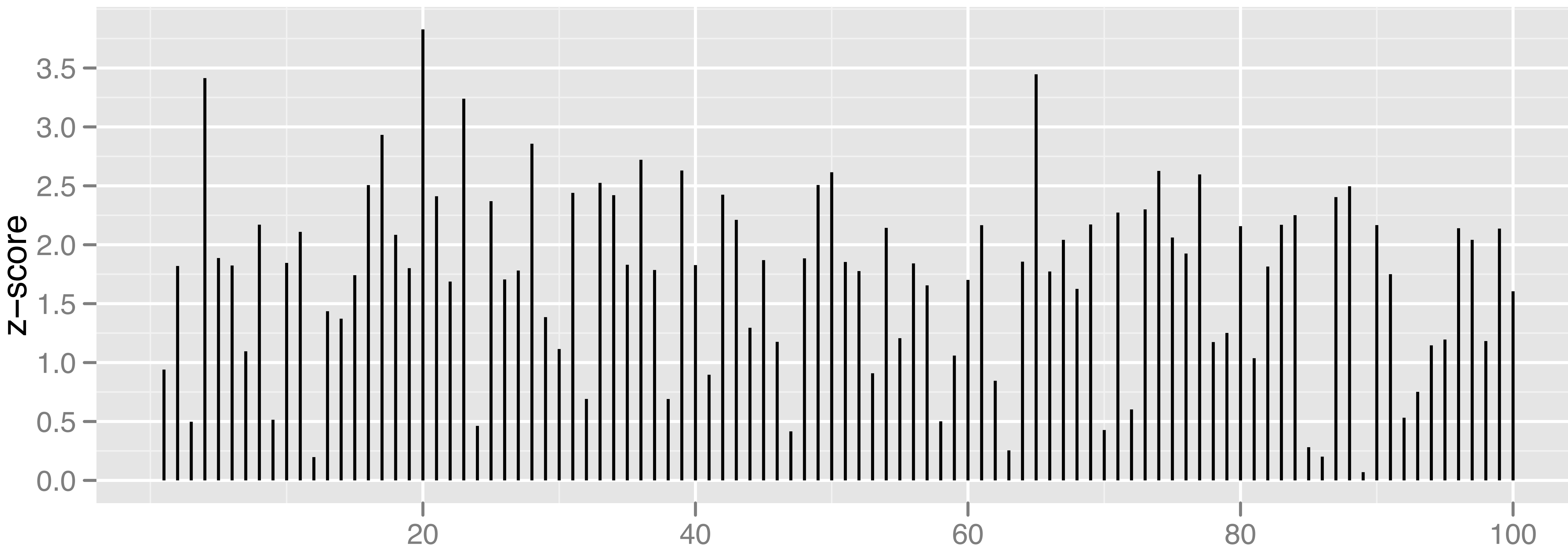
μ_x=80, μ_y=85



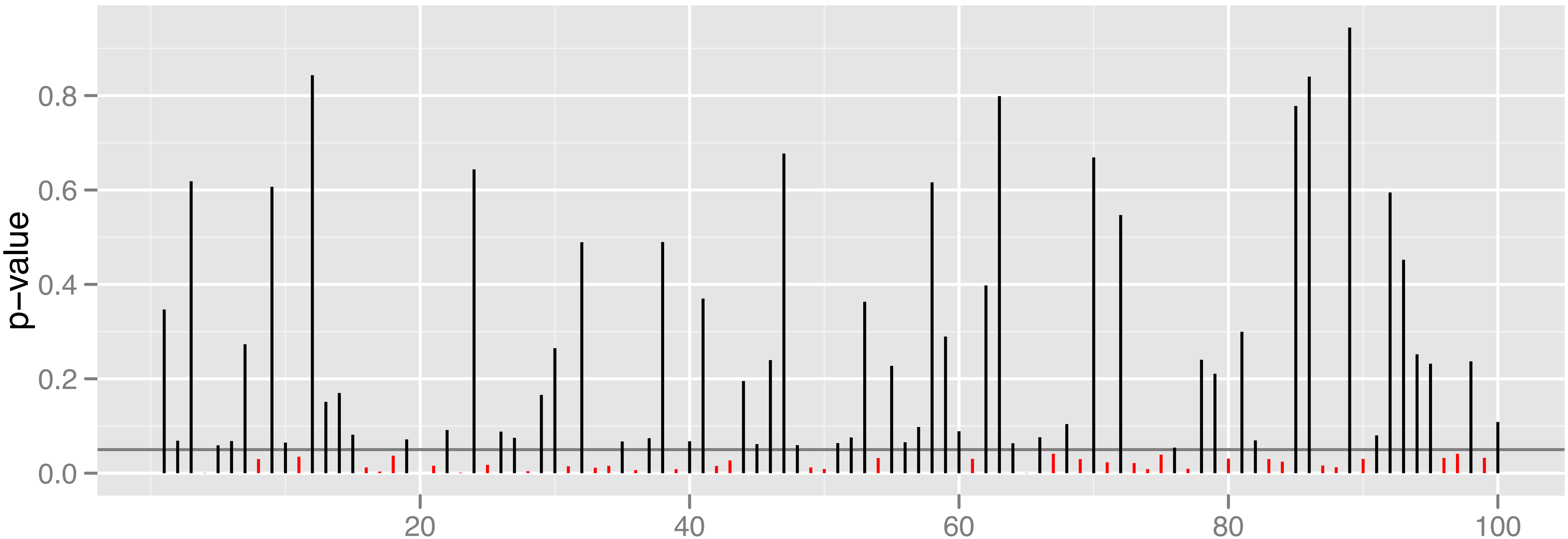
μ_x=80, μ_y=85



μ_x=80, μ_y=85

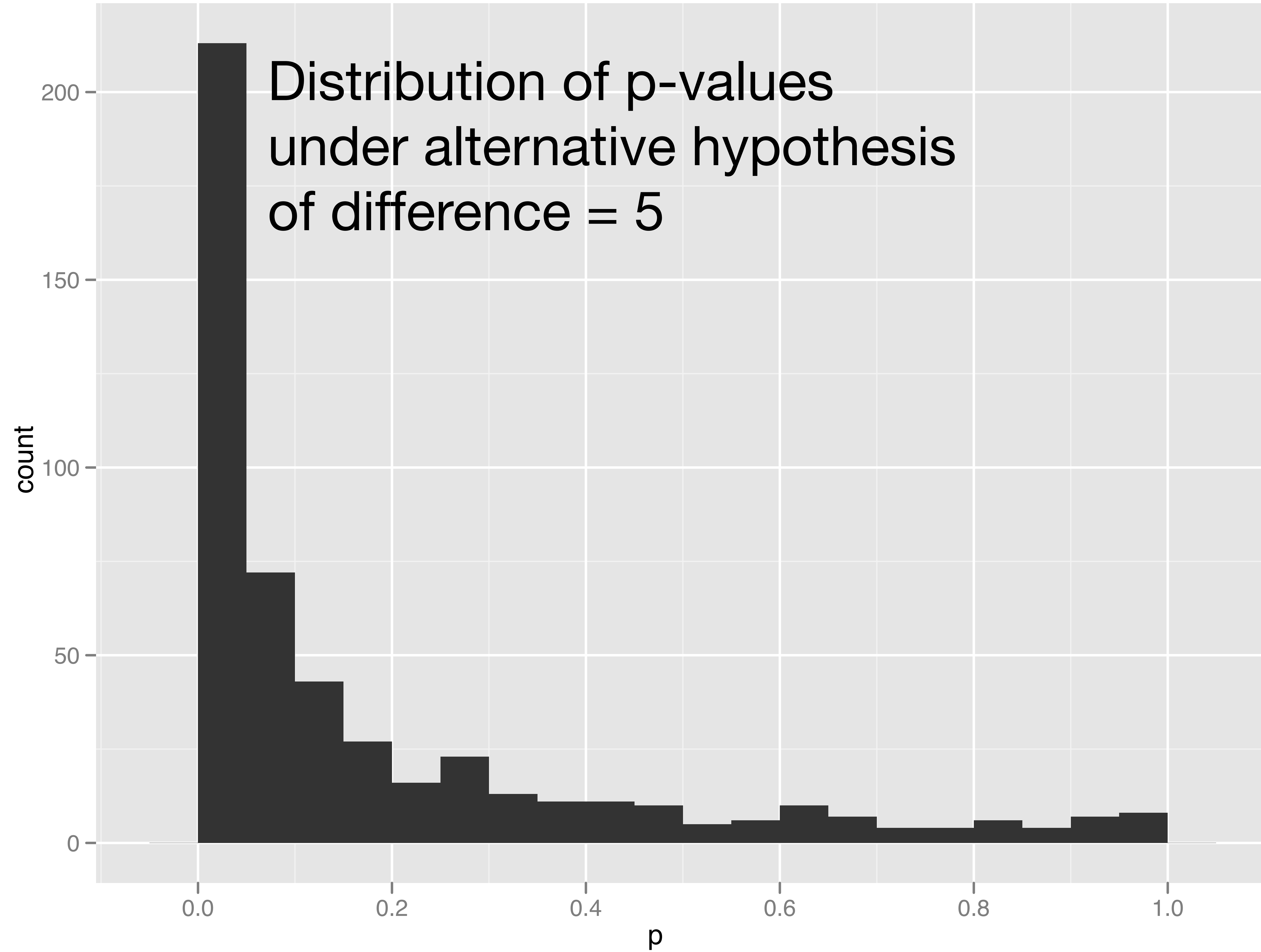


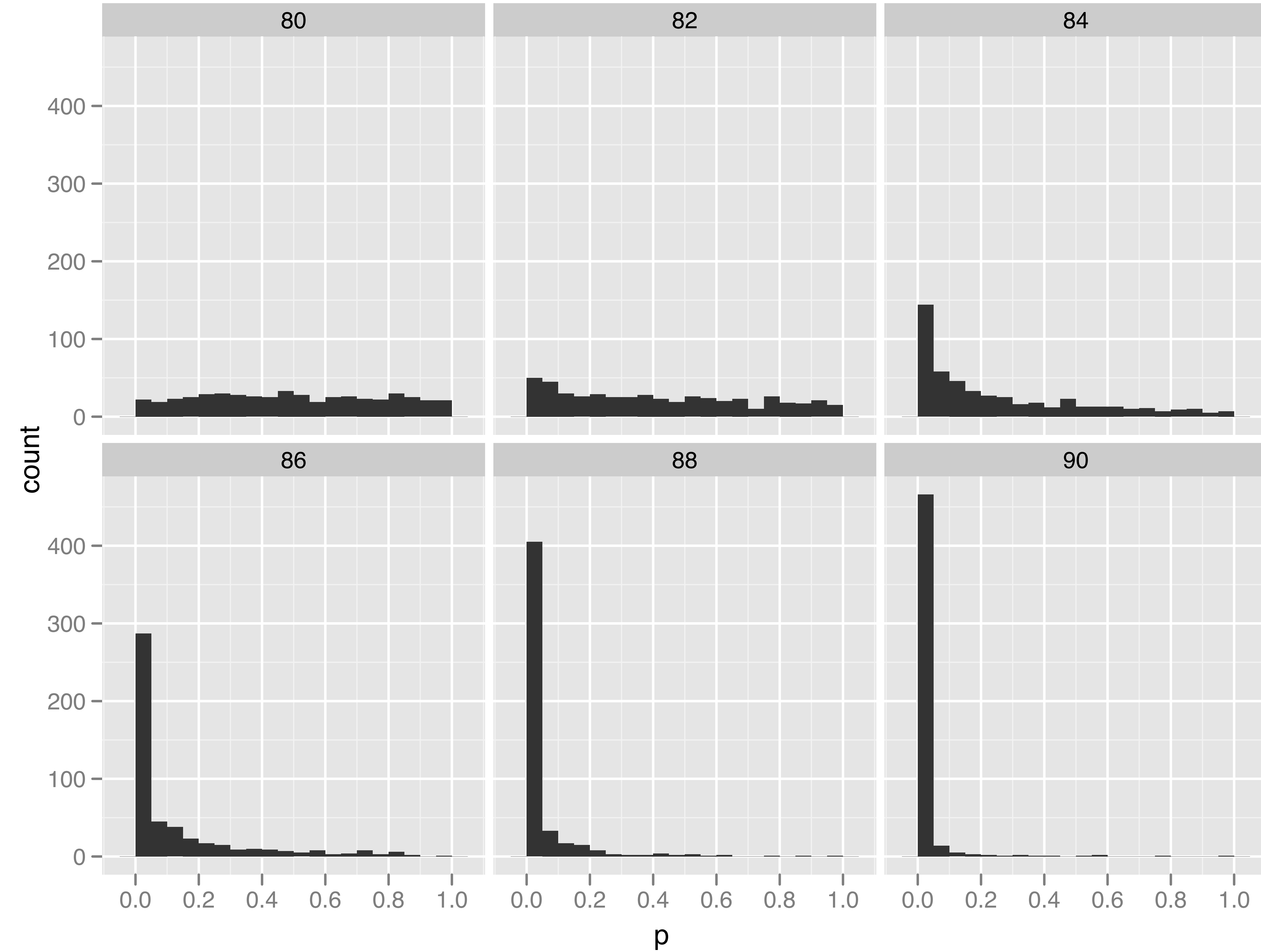
μ_x=80, μ_y=85

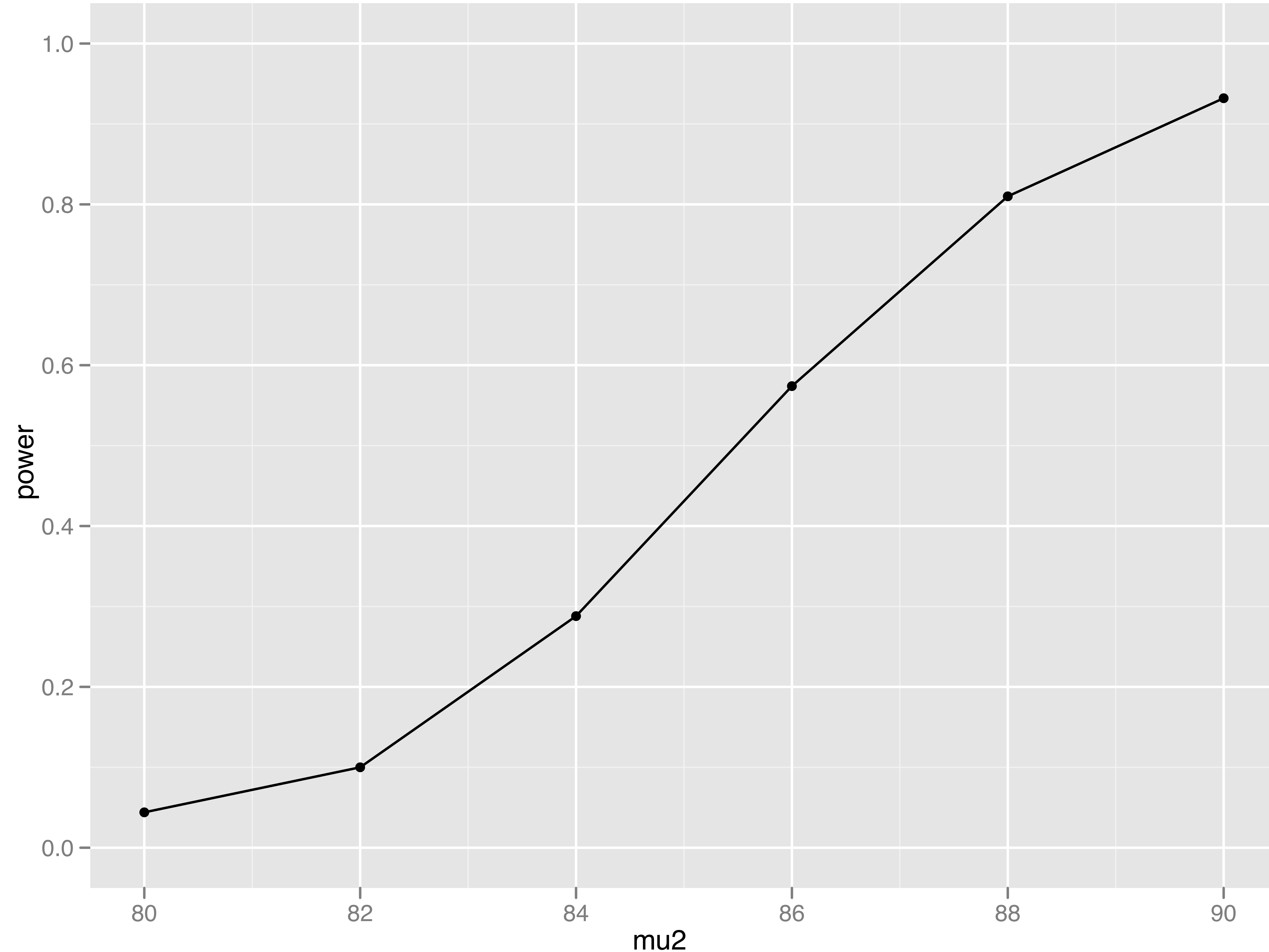


Correctly reject null 39% of the time

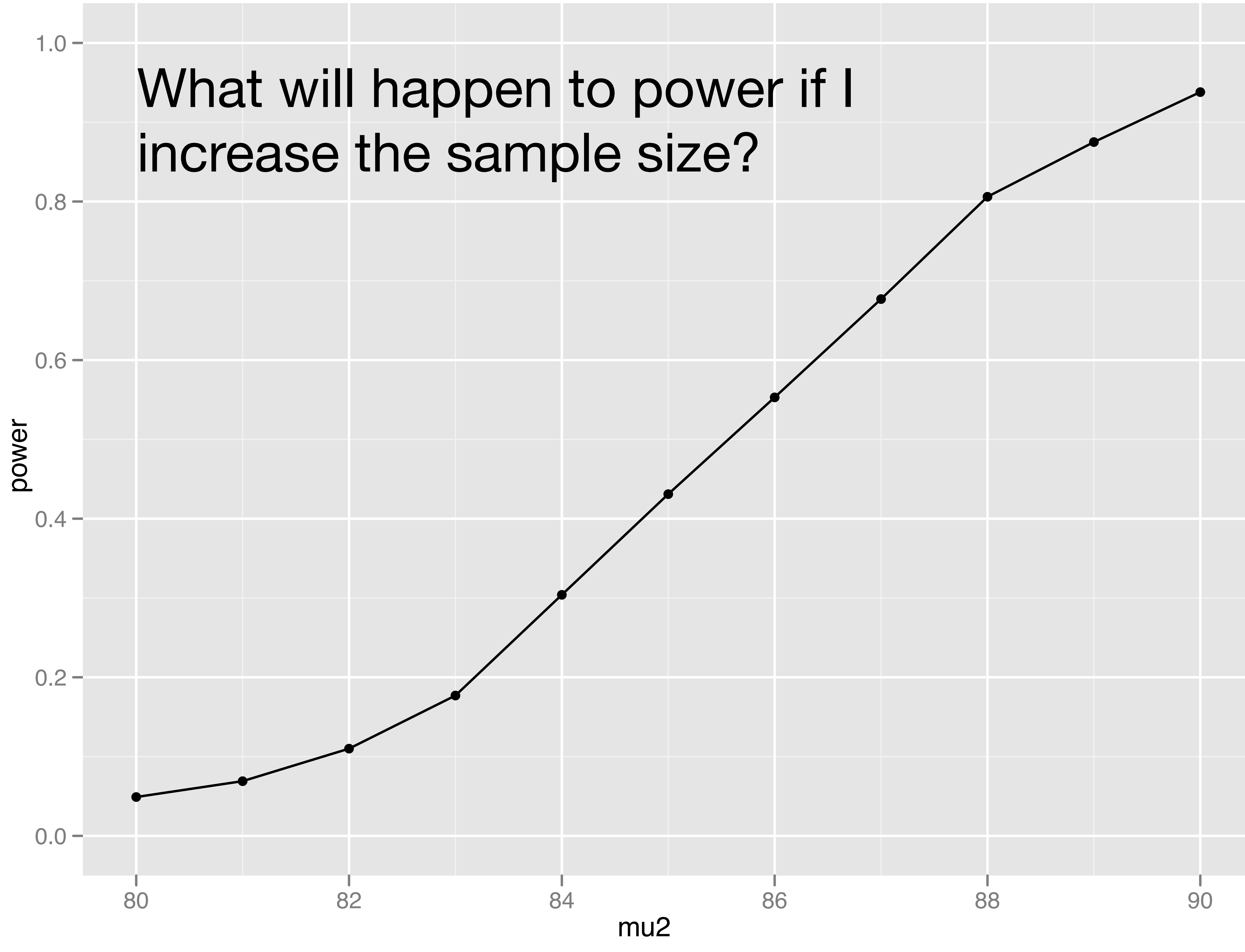
Can only work out power when alternative is known

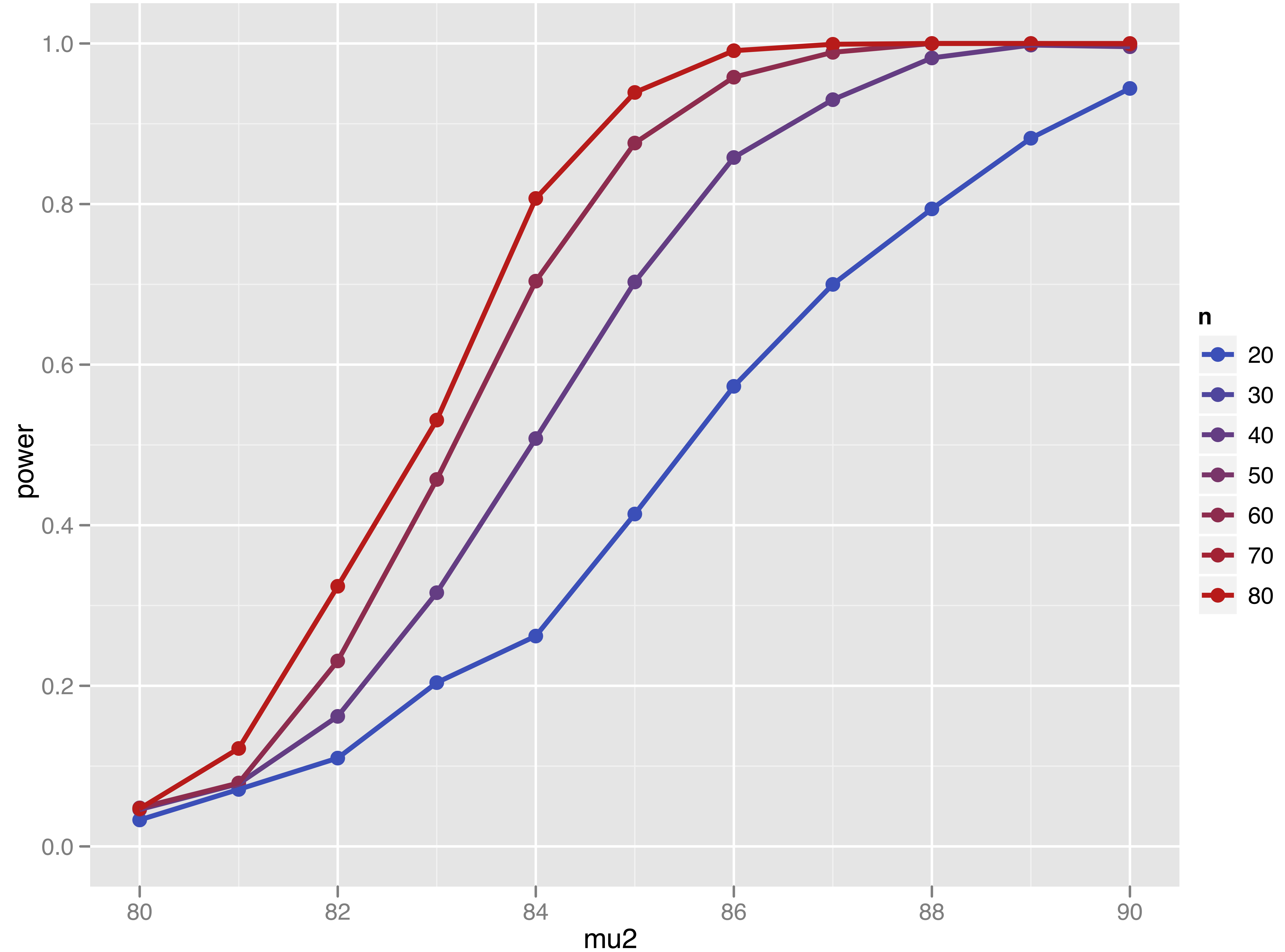

Distribution of p-values
under alternative hypothesis
of difference = 5
count
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
p


80
82
84
86
88
90
count
0
100
200
300
400
0.0
0.2
0.4
0.6
0.8
1.0
p

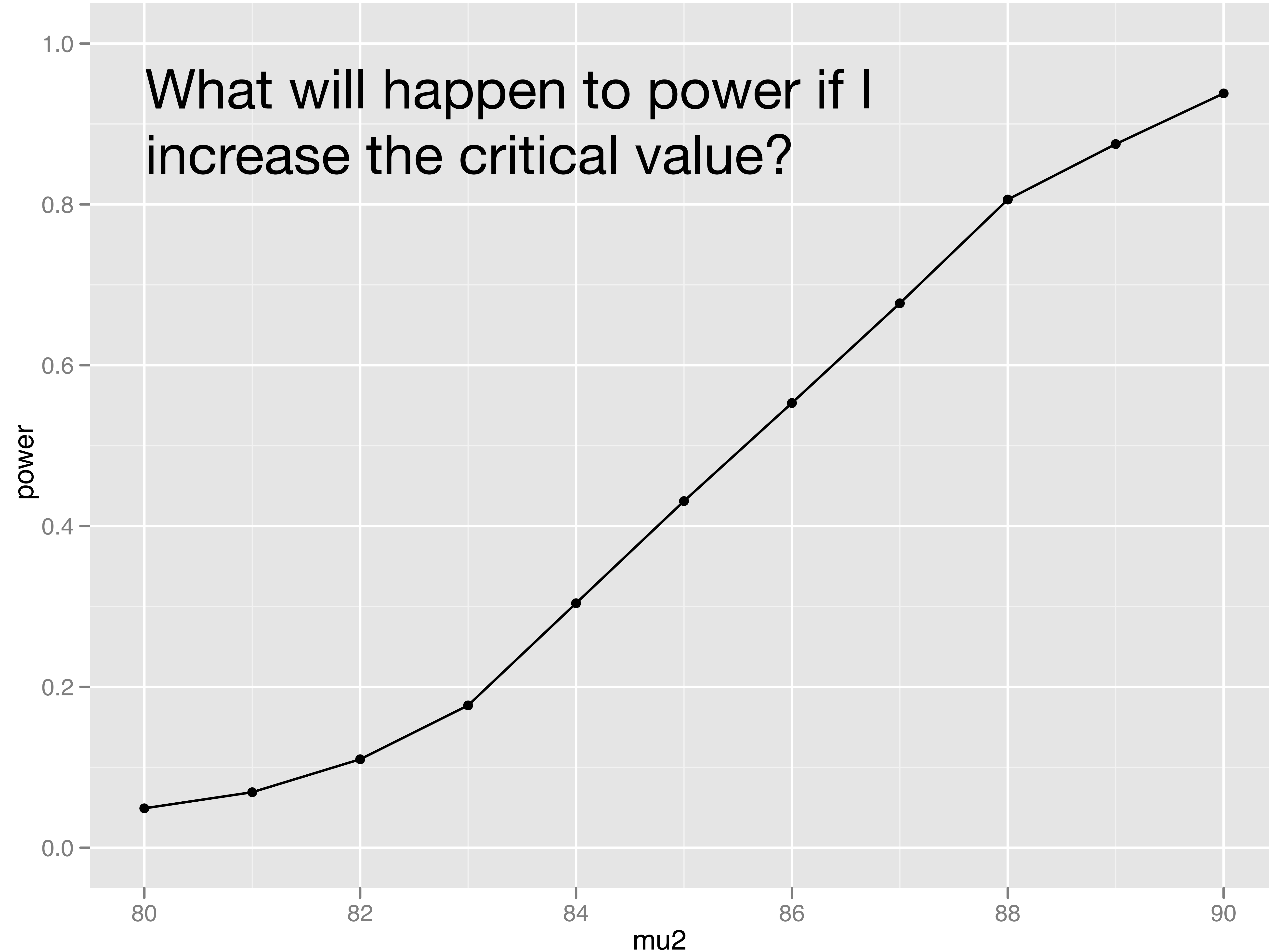


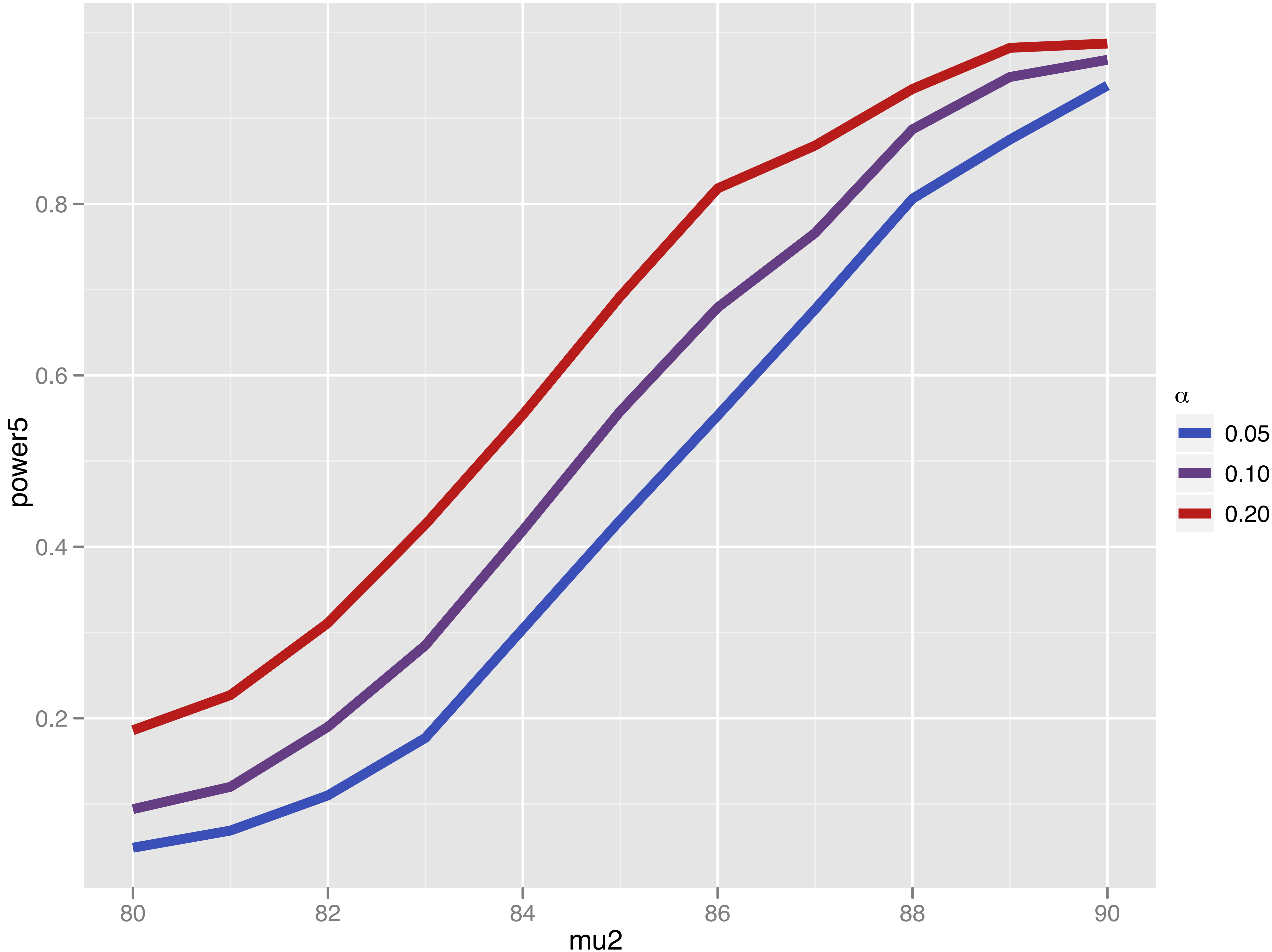
1.0
0.8
0.6
0.4
0.2
0.0
power
80
82
84
86
88
90
mu2


What will happen to power if I increase the sample size?
1.0
0.8
0.6
0.4
0.2
0.0
power
80
82
84
86
88
90
mu2



power
mu2
n
20
30
40
50
60
70
80


What will happen to power if I increase the critical value?
power
mu2
0.0
0.2
0.4
0.6
0.8
1.0
80
82
84
86
88
90



power5
0.2
0.4
0.6
0.8
80
82
84
86
88
90
mu2
α
0.05
0.10
0.20

Your turn

Compare and contrast the the impact of increasing the sample size of the test to changing the alpha level.

t-test

Motivation

In previous example, we assumed that the variance was known. What if it's not?

$$X_i \overset{iid}{\sim} \text{Normal}(\mu_x, \sigma^2)$$

$$Y_i \overset{iid}{\sim} \text{Normal}(\mu_y, \sigma^2)$$

$$i = 1, \cdots, n$$

$$\text{H}_o : \mu_x = \mu_y$$

Your turn

What test statistic might we use?

What distribution would it have?

$$X_i \stackrel{iid}{\sim} \text{Normal}(\mu_x, \sigma^2)$$

$$i = 1, \cdots, n$$

$$Y_j \stackrel{iid}{\sim} \text{Normal}(\mu_y, \sigma^2)$$

$$j = 1, \cdots, m$$

What's the distribution of $\bar{X}_n - \bar{Y}_m$?

Your turn

What's a good test statistic?

What's the distribution of that test statistic?

What if?

Variances are not equal?

Things get complicated!

End up with something that's approximately t-distributed with a non-integer degrees of freedom