Pre/Post data analysis – simple or is it?

SMAR 2011_1 – Mar 3, 2011

Al Feiveson
James Fiedler
Robert Ploutz-Snyder

Today: March 3, 2011

- Paradigm for pre/post data.
- How should we describe change?
- Common analysis methods for comparing post to pre results.
- What do we mean by “% change”?
- What are we testing when we compare % changes?
- Some simulation results
- Conclusions

Characteristics of Pre/Post data

repeated-measures design

subject	pre	post
1	36.4	31.5
2	48.8	40.1
3	25.9	25.6
etc	...	...

subject	session	meas.
1	L-60	50.3
1	L-10	48.6
1	R+0	35.7
1	R+3	39.4
1	R+10	46.0
2	L-60	75.0
2	L-10	78.5
2	R+0	69.6
2	R+3	73.5
2	R+10	77.9
etc	...	...

Characteristics of Pre/Post data (cont.)

partially repeated-measures design

subject	session	meas.	group
1	L-60	50.3	C
1	L-10	48.6	C
1	R+0	35.7	C
1	R+3	39.5	C
1	R+10	46.0	C
2	L-60	75.0	C
2	L-10	78.5	C
2	R+0	69.6	C
2	R+3	73.5	C
2	R+10	77.9	C
etc	...	...	C

subject	session	meas.	group
9	L-60	50.3	E
9	L-10	48.6	E
9	R+0	35.7	E
9	R+3	39.4	E
9	R+10	46.0	E
10	L-60	75.0	E
10	L-10	78.5	E
10	R+0	69.6	E
10	R+3	73.5	E
10	R+10	77.9	E
etc	...	...	E

How should we express the effect of an intervention?

- mean change?
- % change?
- mean change in log measurements?

How should we express the effect of an intervention?

- mean change

pre	post	dif	% ch
75	85	10	13.3%
100	111	11	11.0%

How should we express the effect of an intervention?

- % change

pre	post	dif	% ch
10	18	8	80%
30	51	21	70%

How should we express the effect of an intervention?

- mean change in log measurements

pre	post	dif	% ch
10	18	8	80%
1000	1700	700	70%

How should we express the effect of an intervention?

- mean change in log measurements

pre	post	dif	% ch
10	18	8	80%
1000	1700	700	70%

- Observed outcome measurements are > 0
- They can range over one or more orders of magnitude

How should we express the effect of an intervention?

- mean change?
- % change?
- mean change in log mesaurements?

pre	post	dif	% ch
75	85	10	13.3%
100	110	10	10.0%
130	140	10	7.7%

How should we express the effect of an intervention?

- mean change?
- % change?
- mean change in log mesaurements?

pre	post	dif	% ch
75	85	10	13.3%
100	110	10	10.0%
130	140	10	7.7%

Ans. Correct interpretation should be driven by clinical relevance.

pre and post- bedrest data

sub	pre	post	dif
1	88.7	88.2	-0.5
2	85.1	102.1	16.9
3	106.3	98.6	-7.6
4	115.6	96.2	-19.4
5	62.6	77.3	14.8
6	85.4	82.5	-2.9
7	93.1	97.8	4.7
8	87.1	36	-51.1
9	80.7	64.6	-16.1
10	138.6	111.5	-27.1

Typical approaches to inference on the effect of bedrest:

subject	pre	post	diff
1	36.4	31.5	-4.9
2	48.8	40.1	-8.7
3	25.9	25.6	-0.3
etc	...	...	...

RPM ANOVA with phase (pre/post) as a factor

t-test of H_0 : mean diff = 0

t-test of H_0 : pct change of means = 0

t-test of H_0 : mean pct change = 0

Typical approaches to inference on the effect of bedrest:

subject	pre	post	diff
1	36.4	31.5	-4.9
2	48.8	40.1	-8.7
3	25.9	25.6	-0.3
etc	...	...	...

RPM ANOVA with phase (pre/post) as a factor ($\mu_{post} = \mu_{pre}$)
t-test of H_0 : mean diff = 0 ($\mu_{post} = \mu_{pre}$)
t-test of H_0 : pct change of means = 0 ($\mu_{post} = \mu_{pre}$)
t-test of H_0 : mean log pre = mean log post ($\zeta_{post} = \zeta_{pre}$)
t-test of H_0 : mean observed pct change = 0 (???)

ANOVA

```
                          Number of obs =      20     R-squared     =  0.7828
                          Root MSE      =  14.526     Adj R-squared =  0.5414

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  6843.84439    10  684.384439       3.24     0.0455
                         |
                    isub |  6453.49865     9  717.055406       3.40     0.0414
                    post |  390.345734     1  390.345734       1.85     0.2069
                         |
                Residual |  1899.05337     9  211.005931
              -----------+----------------------------------------------------
                   Total |  8742.89776    19  460.152514
```

t-test on differences

```
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
     dif |      10   -8.835675    6.496244    20.54293   -23.5312     5.85985
------------------------------------------------------------------------------
    mean = mean(dd)                                               t =  -1.3601
Ho: mean = 0                                     degrees of freedom =        9

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.1034       Pr(|T| > |t|) = 0.2069          Pr(T > t) = 0.8966
```

Percent change in means

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
post	10	85.4812	6.958522	22.00478	69.73992	101.2225
pre	10	94.31687	6.662214	21.06777	79.24589	109.3878
diff	10	-8.835675	6.496244	20.54293	-5.85985	23.5312

$$\hat{r}_\mu = 100 \frac{\bar{y}_{post} - \bar{y}_{pre}}{\bar{y}_{pre}}$$

Testing $r_\mu = 0$ is the same as testing $\mu_{post} = \mu_{pre}$. ✓

t-test on differences of logs

```
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  log AES|      10   -.1180426     .097688    .3089166   -.3390283     .102943
------------------------------------------------------------------------------
    mean = mean(z)                                                t =  -1.2084
Ho: mean = 0                                     degrees of freedom =        9

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.1288        Pr(|T| > |t|) = 0.2577          Pr(T > t) = 0.8712
```

t-test on observed percent changes

```
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
%ch AES  |      10   -.0775129    .0741236    .2343994   -.2451921    .0901664
------------------------------------------------------------------------------
    mean = mean(pch)                                              t =  -1.0457
Ho: mean = 0                                     degrees of freedom =        9

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.1615         Pr(|T| > |t|) = 0.3230          Pr(T > t) = 0.8385
```

t-test on observed percent changes

```
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
%ch AES  |      10   -.0775129    .0741236    .2343994   -.2451921    .0901664
------------------------------------------------------------------------------
    mean = mean(pch)                                              t =  -1.0457
Ho: mean = 0                                     degrees of freedom =        9

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.1615        Pr(|T| > |t|) = 0.3230          Pr(T > t) = 0.8385
```

What are we testing when we do this?

Objective: compare pre/post means

subject	pre	post	diff
1	36.4	31.5	-4.9
2	48.8	40.1	-8.7
3	25.9	25.6	-0.3
etc	...	...	...

- ✓ RPM ANOVA with phase (pre/post) as a factor
- ✓ t-test of H_0 : mean diff = 0
- ✓ t-test of H_0 : pct change of means = 0
- ✓ t-test of H_0 : mean log pre = mean log post
- ? t-test of H_0 : mean observed pct change = 0

“Bone density is decreased by 6% after ISS missions.”

“Bone density is decreased by 6% after ISS missions.”

% change in means = -6% ?

"Bone density is decreased by 6% after ISS missions."

% change in means = -6% ?

mean % change = -6% ?

"Bone density is decreased by 6% after ISS missions."

% change in means = -6% ?

mean % change = -6% ?

What's the difference?

pct change in means $$100 \times \frac{\mu_{post} - \mu_{pre}}{\mu_{pre}}$$

estimated by $$100 \times \frac{\bar{y}_{post} - \bar{y}_{pre}}{\bar{y}_{pre}}$$?

pct change in means $$100 \times \frac{\mu_{post} - \mu_{pre}}{\mu_{pre}}$$

estimated by $$100 \times \frac{\bar{y}_{post} - \bar{y}_{pre}}{\bar{y}_{pre}}$$?

---

mean pct change $$\mu_{PC} = E\left(100 \times \frac{\mu_{i,post} - \mu_{i,pre}}{\mu_{i,pre}}\right)$$

estimated by $$\frac{1}{n}\sum_i \left\{100 \times \frac{(y_{i,post} - y_{i,pre})}{y_{i,pre}}\right\}$$??

(average <u>observed</u> pct change)

An Example

subject	pre	post	diff	% ch
1	30	25	-5	-16.7
2	40	36	-4	-10.0
3	25	19	-6	-24.0
4	53	49	-4	-7.5
ave	37	32.25	-4.75	-14.6

estimated % change in means = 100 $\times$ (-4.75/37) = -12.8

estimated mean observed % change = -14.6

Interpretation of mean obs pct change

mean <u>obs</u> pct change $\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$

$$\approx \frac{1}{n}\sum_i \left\{100 \times \frac{(y_{i,post} - y_{i,pre})}{y_{i,pre}}\right\}$$ for “large” n

Interpretation of mean obs pct change

mean obs pct change $\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$

$$\approx \frac{1}{n}\sum_i \left\{100 \times \frac{(y_{i,post} - y_{i,pre})}{y_{i,pre}}\right\}$$ for “large” n

If it exists, Ψ is a population characteristic, but what does it mean?

Interpretation of mean obs pct change

mean obs pct change $\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$

$$\approx \frac{1}{n}\sum_i \left\{100 \times \frac{(y_{i,post} - y_{i,pre})}{y_{i,pre}}\right\}$$ for "large" n

If it exists, Ψ is a population characteristic,
but what does it mean?

Is Ψ the actual mean pct change $\mu_{PC} = E\left(100 \times \frac{\mu_{i,post} - \mu_{i,pre}}{\mu_{i,pre}}\right)$?

Interpretation of mean obs pct change

mean obs pct change $\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$

$$\approx \frac{1}{n}\sum_i \left\{100 \times \frac{(y_{i,post} - y_{i,pre})}{y_{i,pre}}\right\}$$ for "large" n

If it exists, Ψ is a population characteristic, but what does it mean?

Is Ψ the actual mean pct change $\mu_{PC} = E\left(100 \times \frac{\mu_{i,post} - \mu_{i,pre}}{\mu_{i,pre}}\right)$?

If not, how well does Ψ describe the effect of spaceflight/bedrest over a population of subjects?

Simulation 1

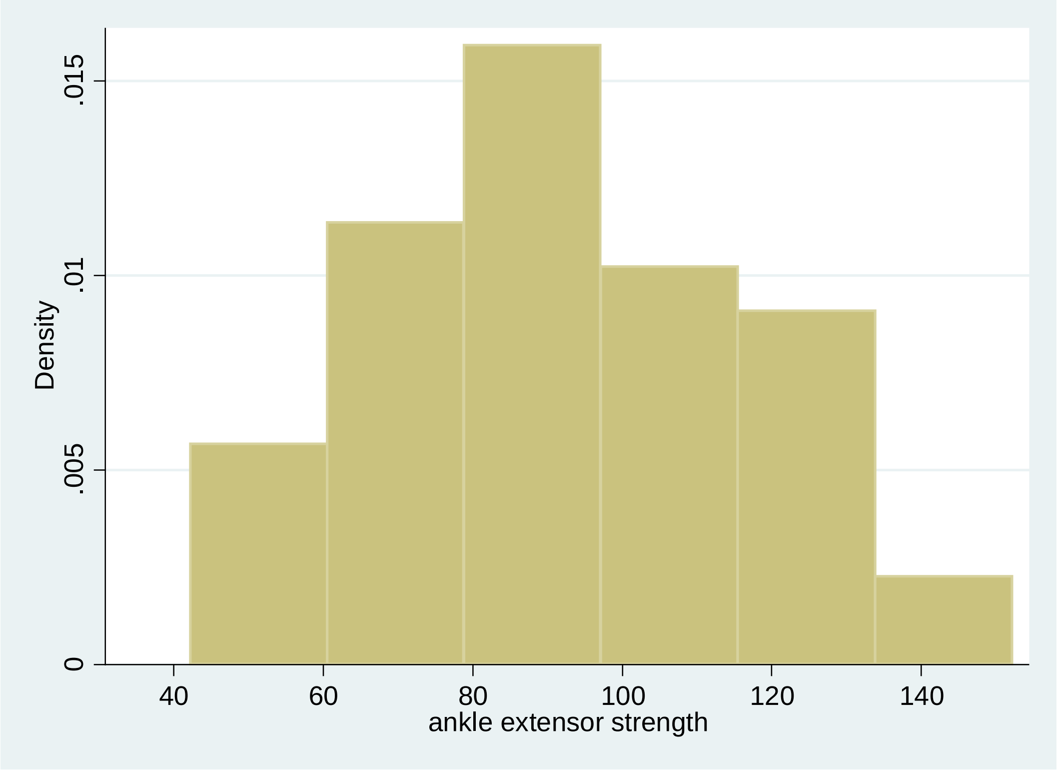
- 10,000 "experiments"

- N = 20 simulated paired pre-post measurements of ankle extensor strength (AES) per experiment
- mean effect of "bedrest" : $\Delta = \mu_{post} - \mu_{pre}$

- t-test #1 – test H_0: $\Delta = 0$.

- t-test #2 – test H'_0: $\Psi = 0$.

actual AES data (N = 24, pre/post)



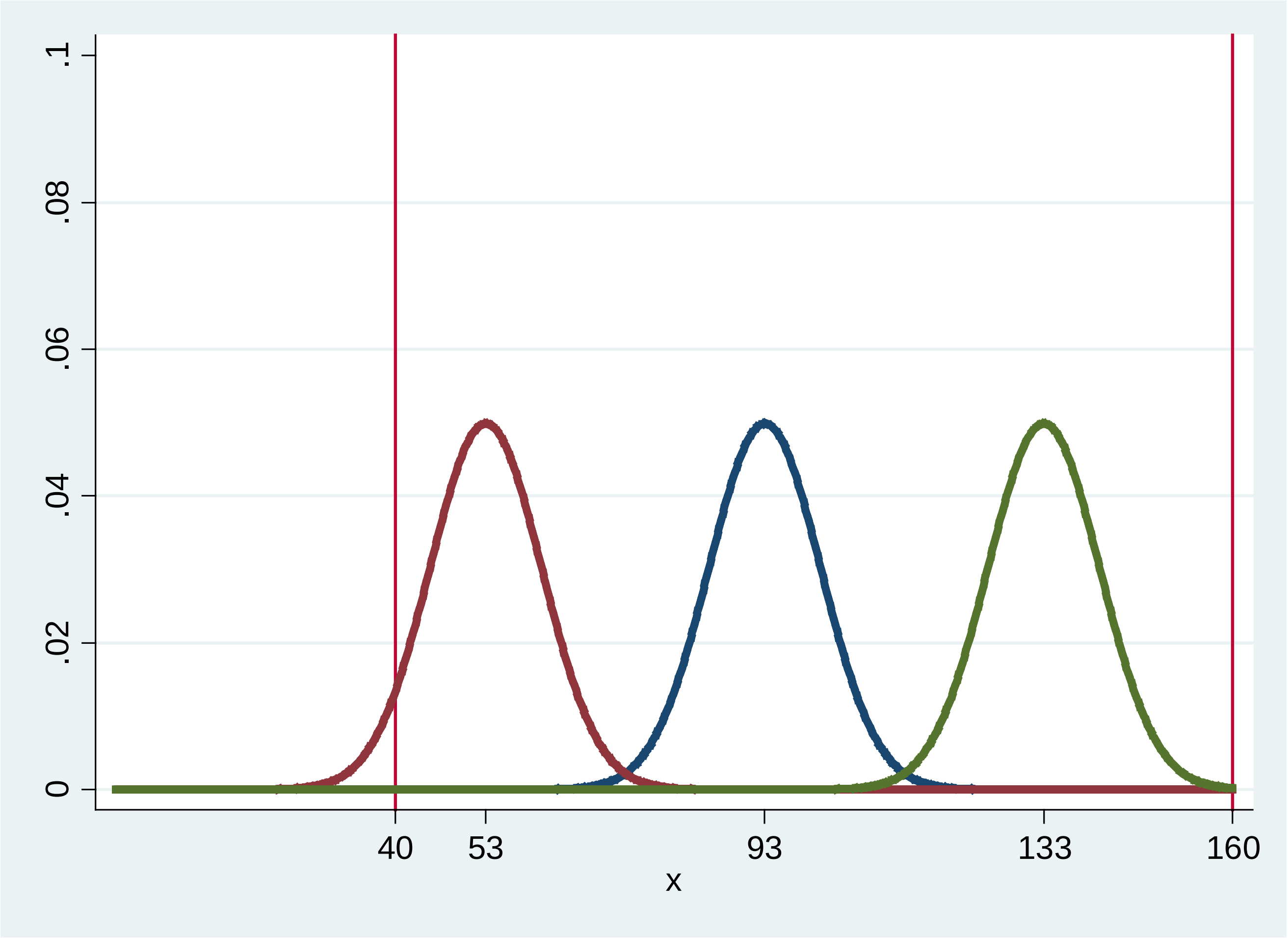
Variable	Obs	Mean	Std. Dev.	Min	Max
AES (pre)	24	103.0	23.0	67.8	152.1
AES (post	24	82.8	22.9	42.2	128.1

Simulation 1 (cont.)

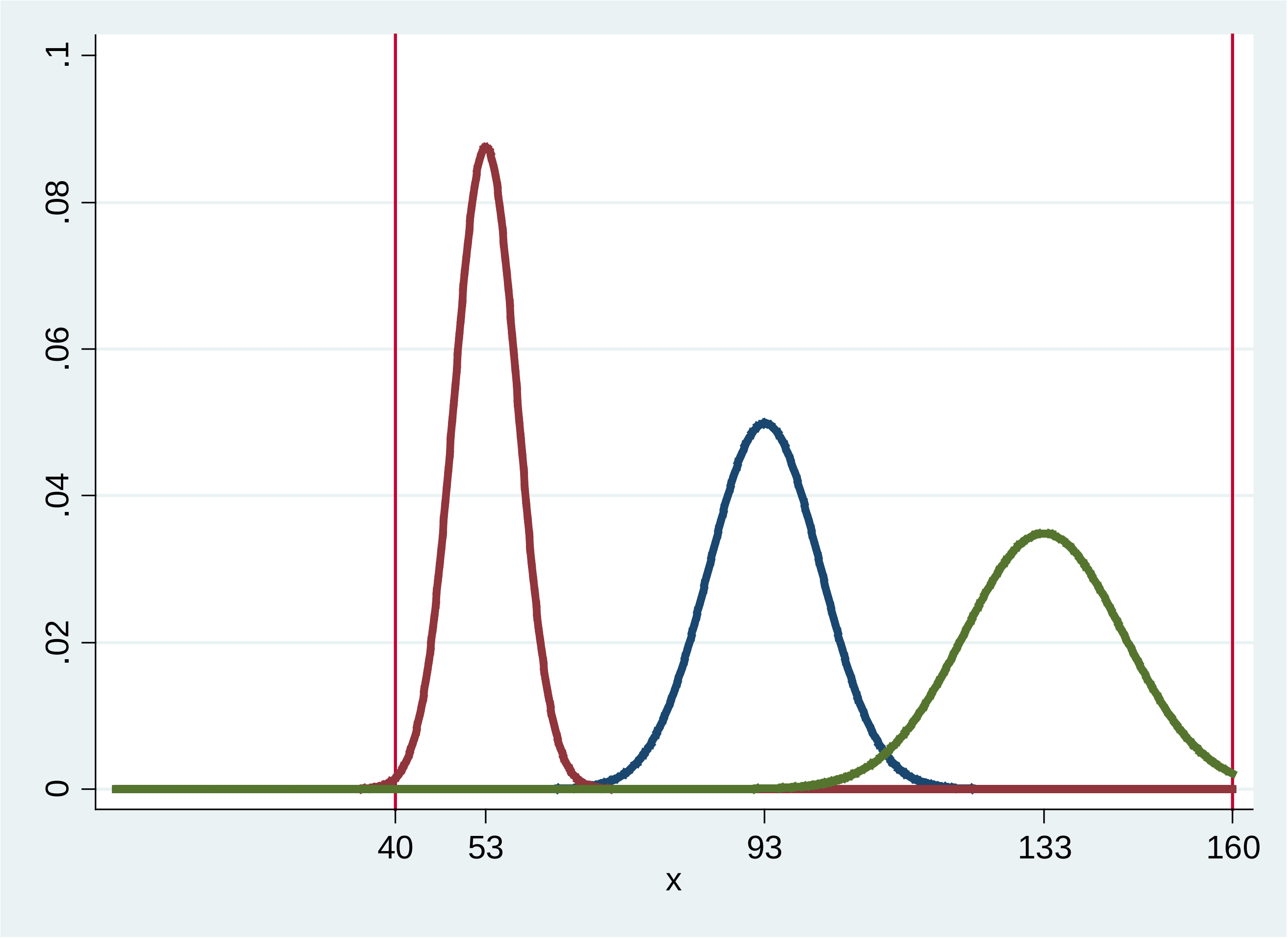
- distribution of pre-means (between subjects)
- distribution of post-means (between subjects)
- distribution around subject-specific means (within subjects)

Simulation 1 (cont.)

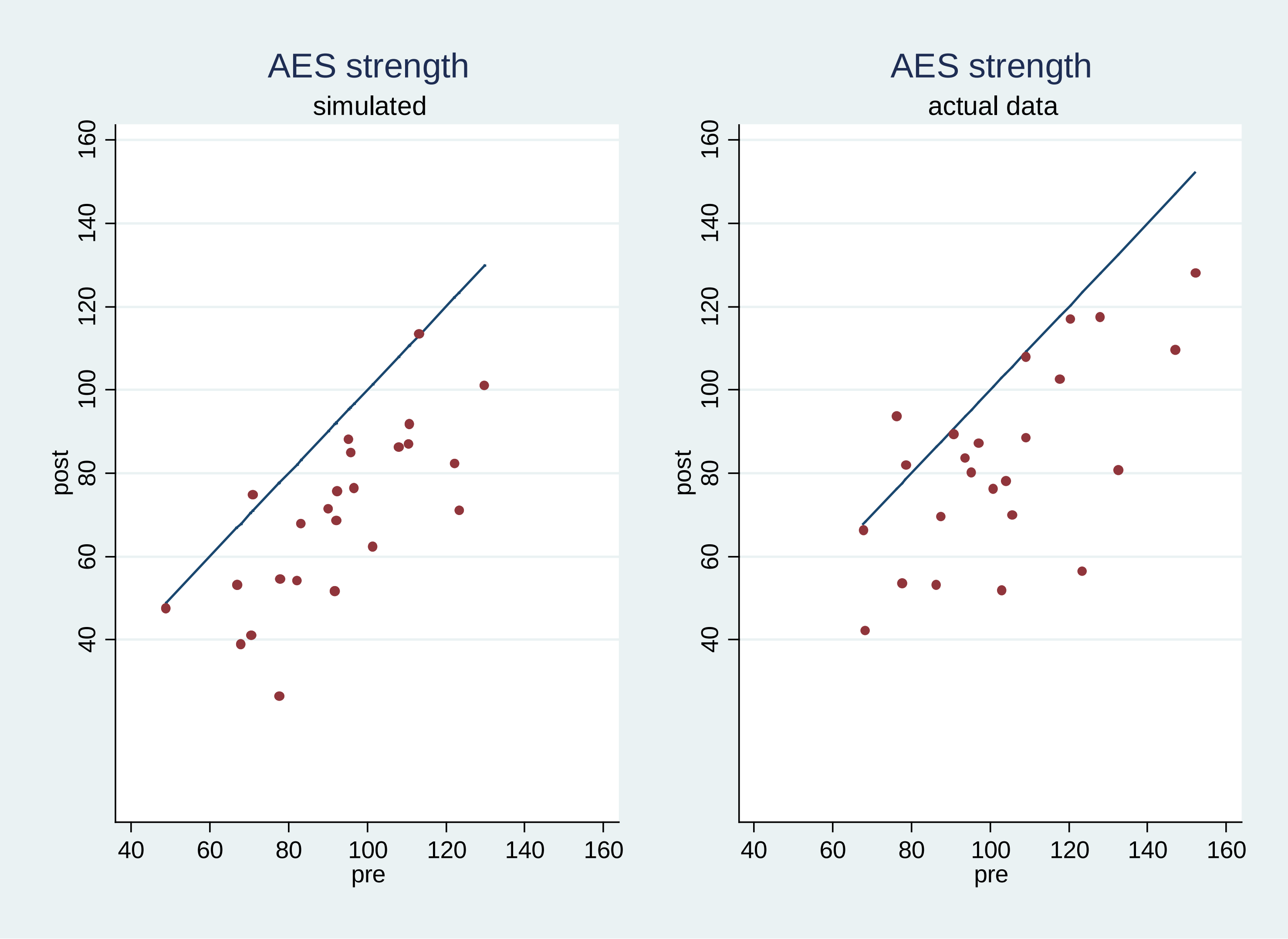
- distribution of pre-means (between subjects)
- distribution of post-means (between subjects)
- distribution around subject-specific means (within subjects)
    - constant SD
    - constant CV (SD/mean)
    - other ?

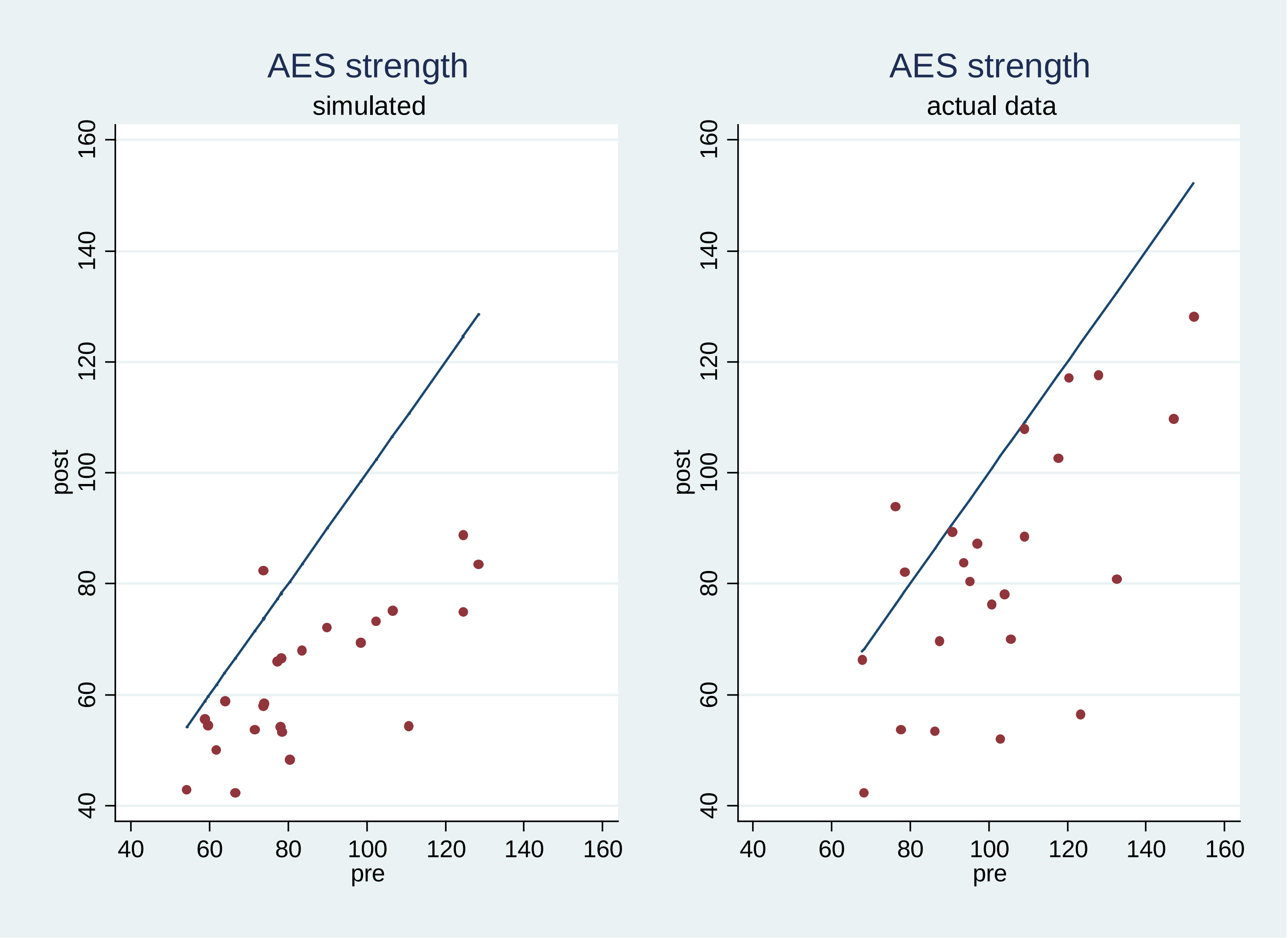


.1
.08
.06
.04
.02
0
40
53
93
133
160
x



0
.02
.04
.06
.08
.1
40
53
93
133
160
x


AES strength
simulated
AES strength
actual data
post
pre
40
60
80
100
120
140
160


AES strength
simulated
post
pre
AES strength
actual data
post
pre
40
60
80
100
120
140
160

Simulation 1 (cont.)

- 10,000 "experiments"

- N = 20 simulated paired pre-post measurements of ankle extensor strength (AES) per experiment
- mean effect of "bedrest" : $\boxed{\Delta = \mu_{post} - \mu_{pre} = 0}$

- t-test #1 – test H_0: $\Delta = 0$.

$$\Delta = E\left(y_{i,post} - y_{i,pre}\right)$$

- t-test #2 – test H'_0: $\Psi = 0$.

$$\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$$

Simulation 1 (cont.)

- 10,000 "experiments"

- N = 20 simulated paired pre-post measurements of ankle extensor strength (AES) per experiment
- mean effect of "bedrest" : $\boxed{\Delta = \mu_{post} - \mu_{pre} = 0}$

- t-test #1 – test H_0: $\Delta = 0$. $\Delta = E\left(y_{i,post} - y_{i,pre}\right)$

- t-test #2 – test H'_0: $\Psi = 0$. $\Psi = E\left(100 \times \dfrac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$

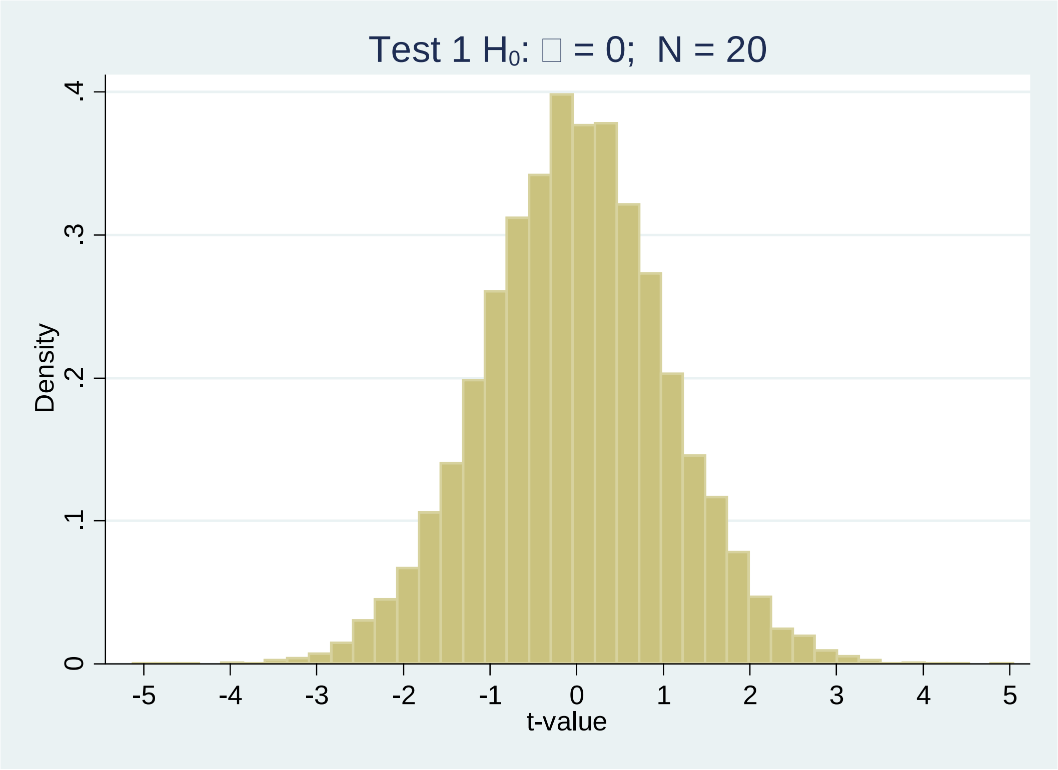
Output:

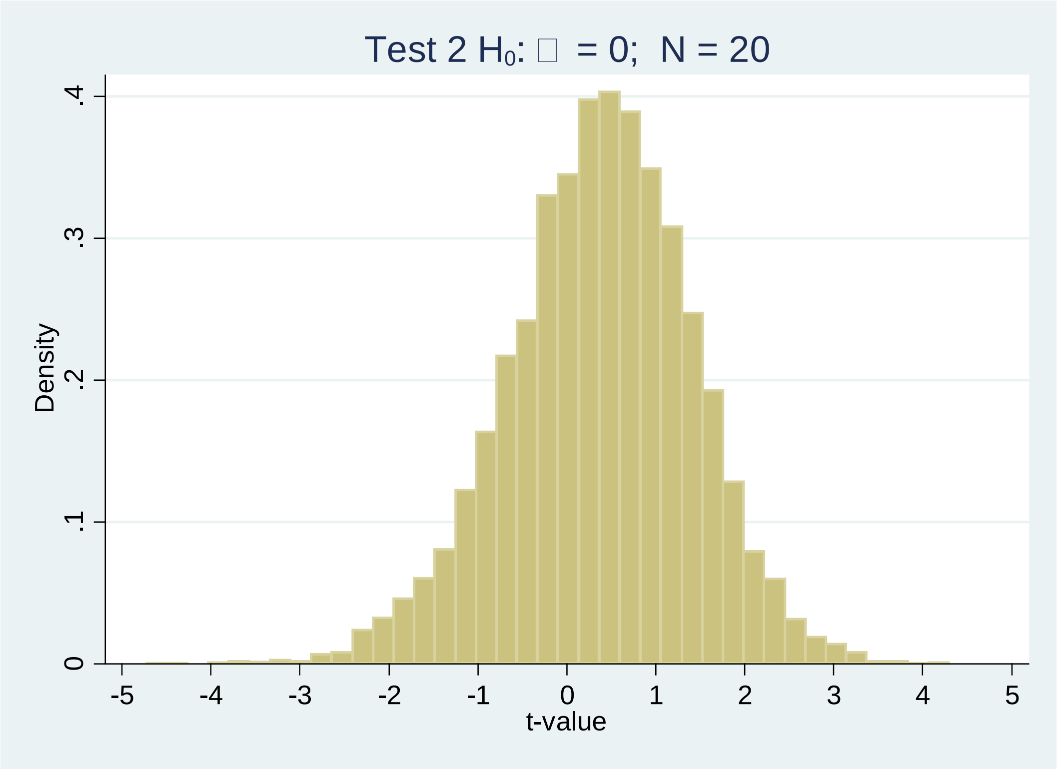
- 10,000 "t"-values and p-values for t-test #1
- 10,000 "t"-values and p-values for t-test #2

Expected Results

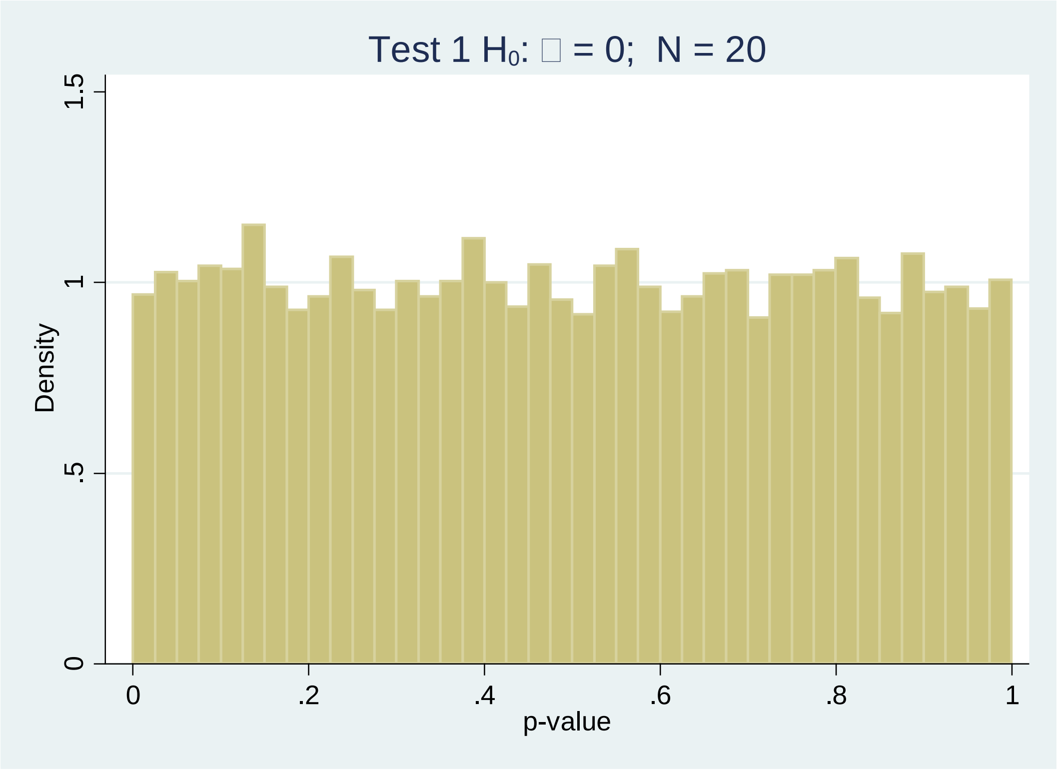
If H_0 ($\Delta = 0$) is true:
t-values for test 1 should be distributed as t(19)
p-values for test 1 should be distributed as U(0,1)

If H'_0: ($\Psi = 0$) is true:
t-values for test 2 should be distributed as t(19)
p-values for test 2 should be distributed as U(0,1)

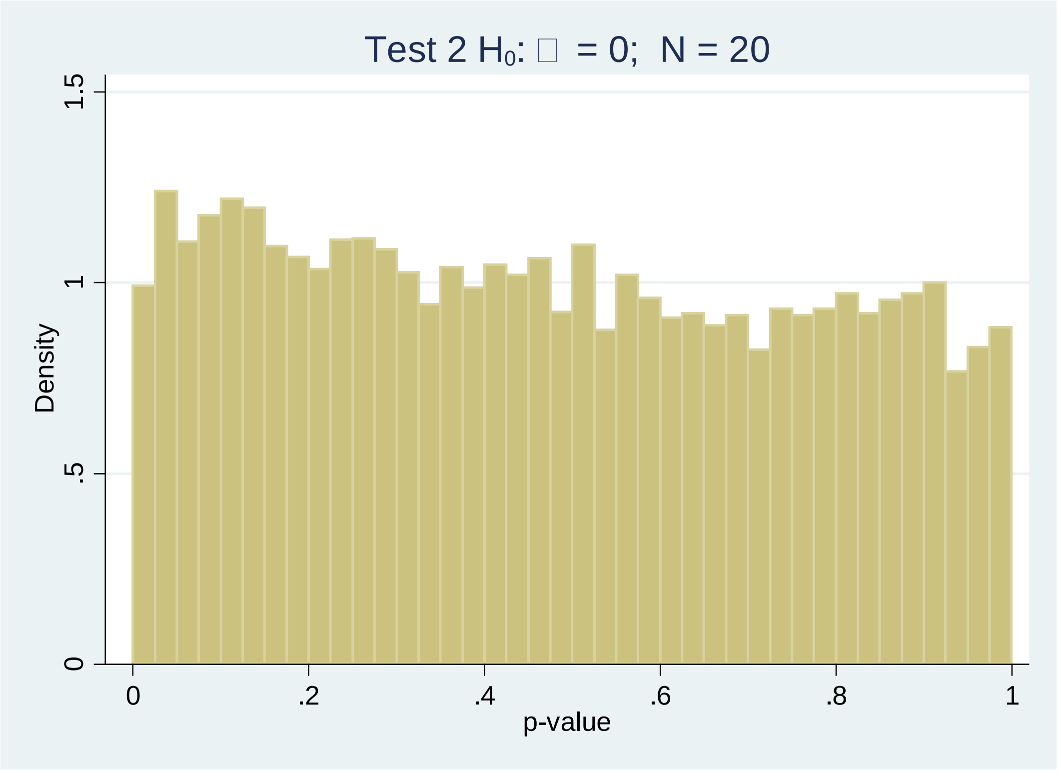

Test 1 H0: □ = 0; N = 20
Density
0
.1
.2
.3
.4
-5
-4
-3
-2
-1
0
1
2
3
4
5
t-value


Test 2 H0: □ = 0; N = 20
Density
0
.1
.2
.3
.4
-5
-4
-3
-2
-1
0
1
2
3
4
5
t-value

P < .05 = 499



P < .05 = 558



Simulation 2

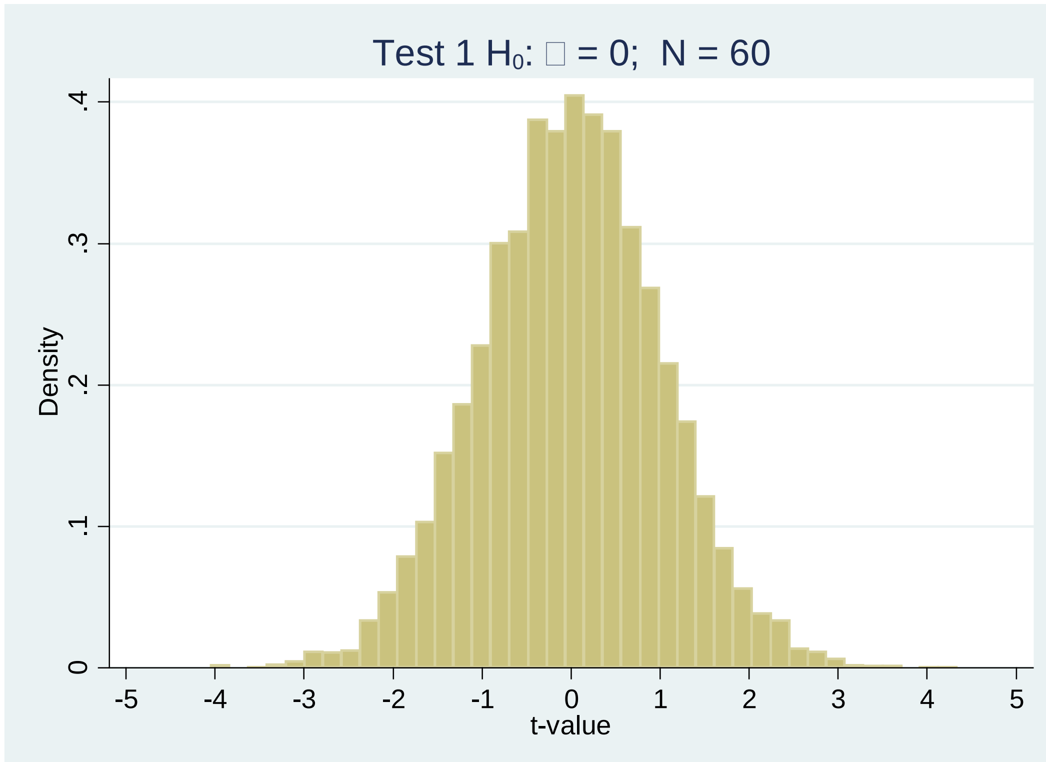
- 10,000 “experiments”

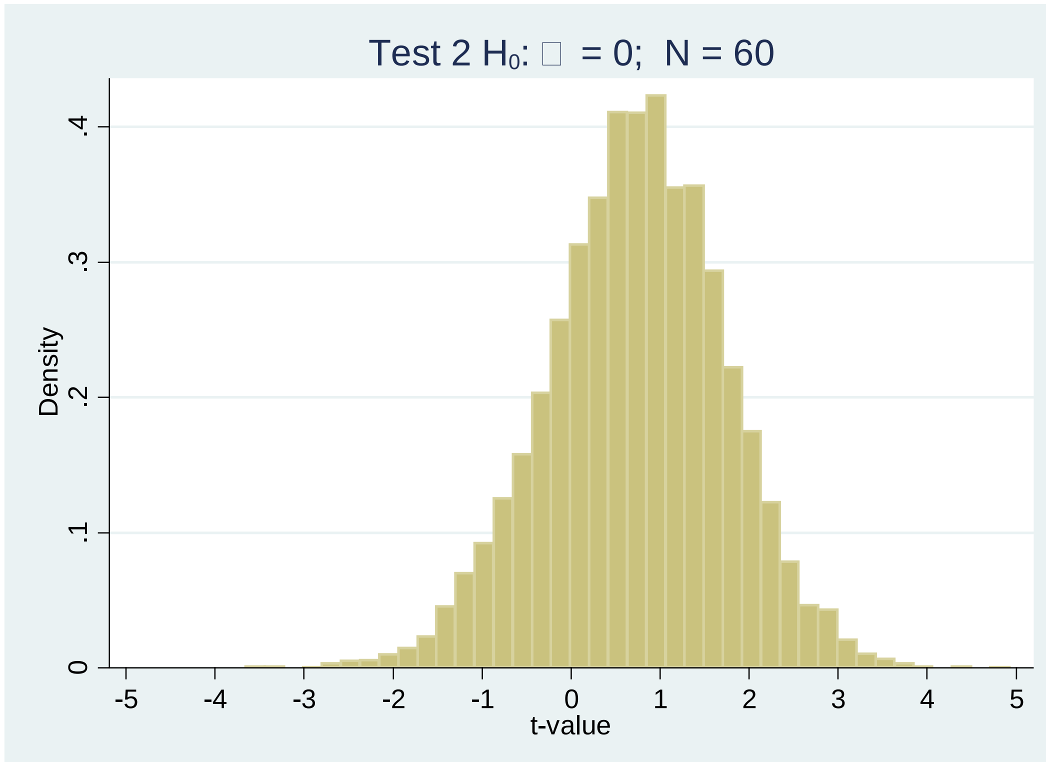
- N = 60 simulated paired pre-post measurements of ankle extensor strength (AES) per experiment
- mean effect of “bedrest” : $\boxed{\Delta = \mu_{post} - \mu_{pre} = 0}$

- t-test #1 – test H_0: $\Delta = 0$. $\Delta = E\left(y_{i,post} - y_{i,pre}\right)$

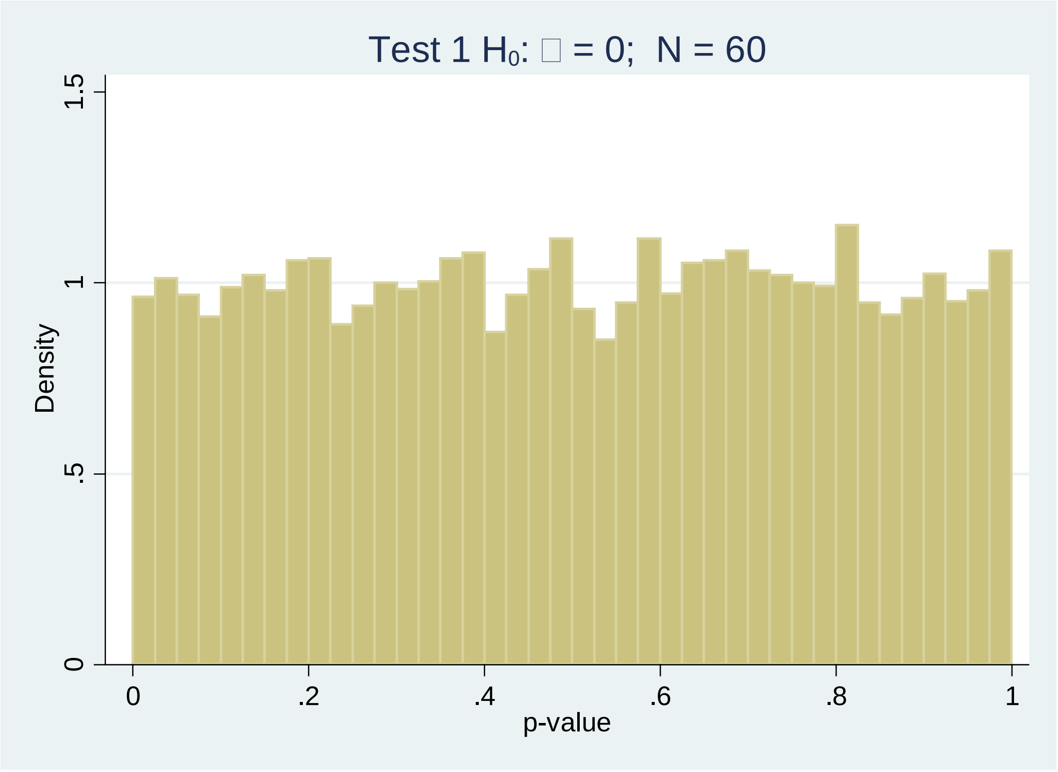
- t-test #2 – test H'_0: $\Psi = 0$. $\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$

Output:
- 10,000 “t”-values and p-values for t-test #1
- 10,000 “t”-values and p-values for t-test #2

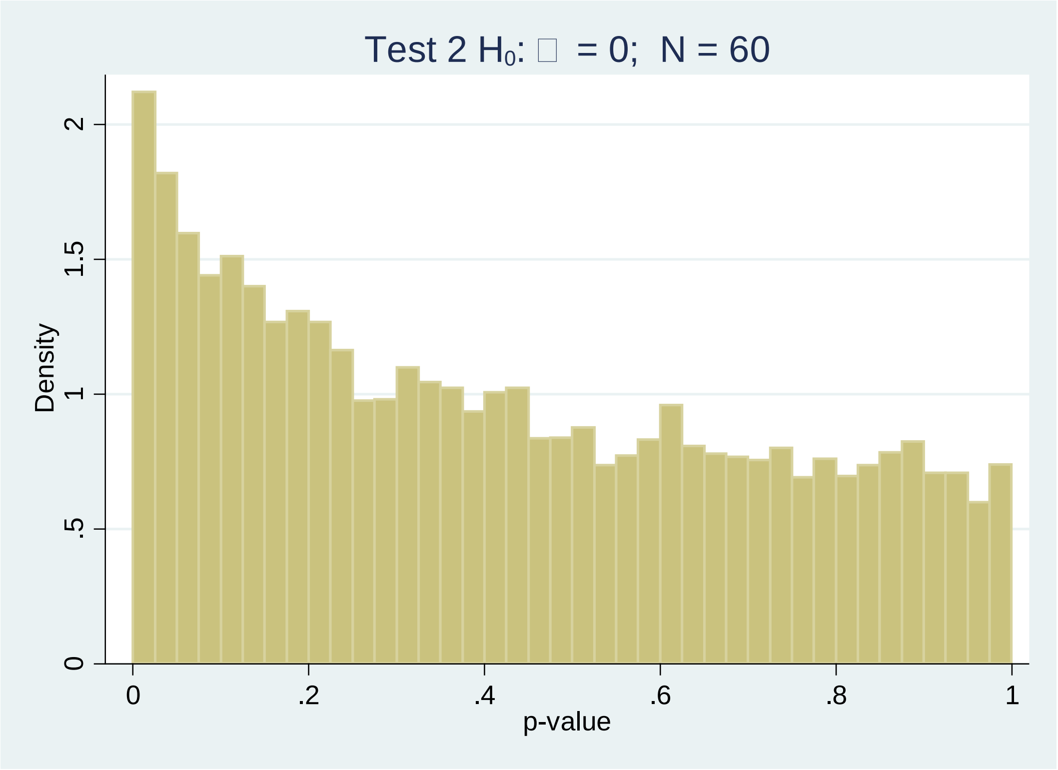

Test 1 H0: □ = 0; N = 60
Density
0
.1
.2
.3
.4
-5
-4
-3
-2
-1
0
1
2
3
4
5
t-value


Test 2 H_0: □ = 0; N = 60
Density
0
.1
.2
.3
.4
-5
-4
-3
-2
-1
0
1
2
3
4
5
t-value

P < .05 =492



P < .05 = 985



What’s going on ??

What's going on ??

Ans: $\Psi \neq 0$

Estimating Ψ

Simulation of pre, post AES values with $\Delta = \mu_{post} - \mu_{pre} = 0$

$$\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$$

N = 10^3, 10^4, 10^5, and 10^6

Estimating Ψ

Simulation of pre, post AES values with $\Delta = \mu_{post} - \mu_{pre} = 0$

$$\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$$

N	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
1000	1.950	.639	20.2	0.696	3.204
10000	2.028	.207	20.7	1.623	2.435
100000	2.079	.066	20.9	1.949	2.209
1000000	2.048	.021	20.9	2.007	2.089

Estimating Ψ

Simulation of pre, post AES values with $\Delta = \mu_{post} - \mu_{pre} = 0$

$$\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$$

N	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
1000	1.950	.639	20.2	0.696	3.204
10000	2.028	.207	20.7	1.623	2.435
100000	2.079	.066	20.9	1.949	2.209
1000000	2.048	.021	20.9	2.007	2.089

Δ = 0, μ_{PC} = 0, but Ψ = 2.05 % !

Estimating Ψ

Simulation of pre, post AES values with $\Delta = \mu_{post} - \mu_{pre} = 0$

$$\Psi = E\left(100 \times \frac{y_{i,post} - y_{i,pre}}{y_{i,pre}}\right)$$

N	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
1000	1.950	.639	20.2	0.696	3.204
10000	2.028	.207	20.7	1.623	2.435
100000	2.079	.066	20.9	1.949	2.209
1000000	2.048	.021	20.9	2.007	2.089

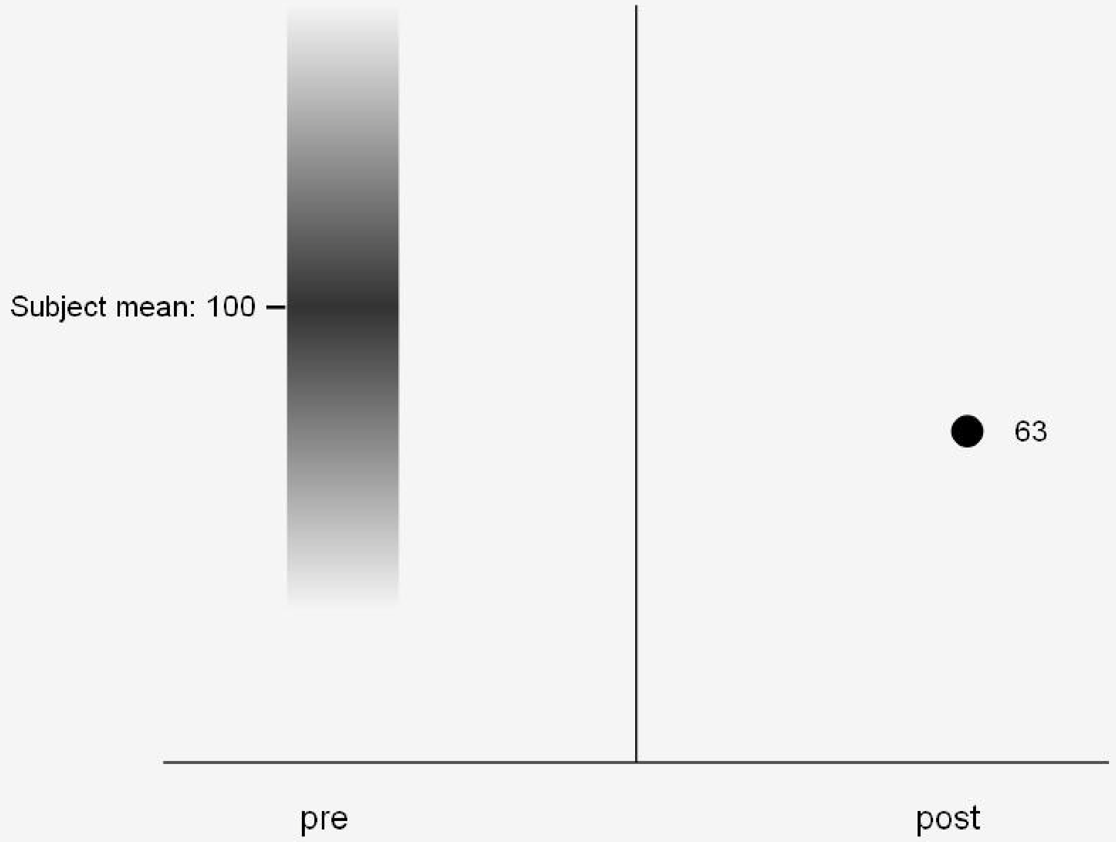
$\Delta = 0$, $\mu_{PC} = 0$, but $\Psi = 2.05$ % !

Does this mean that sham bedrest causes AES to increase by 2.05% ???

Ψ depends on the within-subject CV (sd/mean)

CV	Ψ (%)
0.01	0.01
0.05	0.23
0.10	1.04
0.12	1.55
0.14	2.05
0.16	2.70
0.18	3.53
0.20	4.53

A single pre value is not a stable baseline ...



next

What if $\Delta \neq 0$?

(mean post $\neq$ mean pre)

pre mean = 91.75

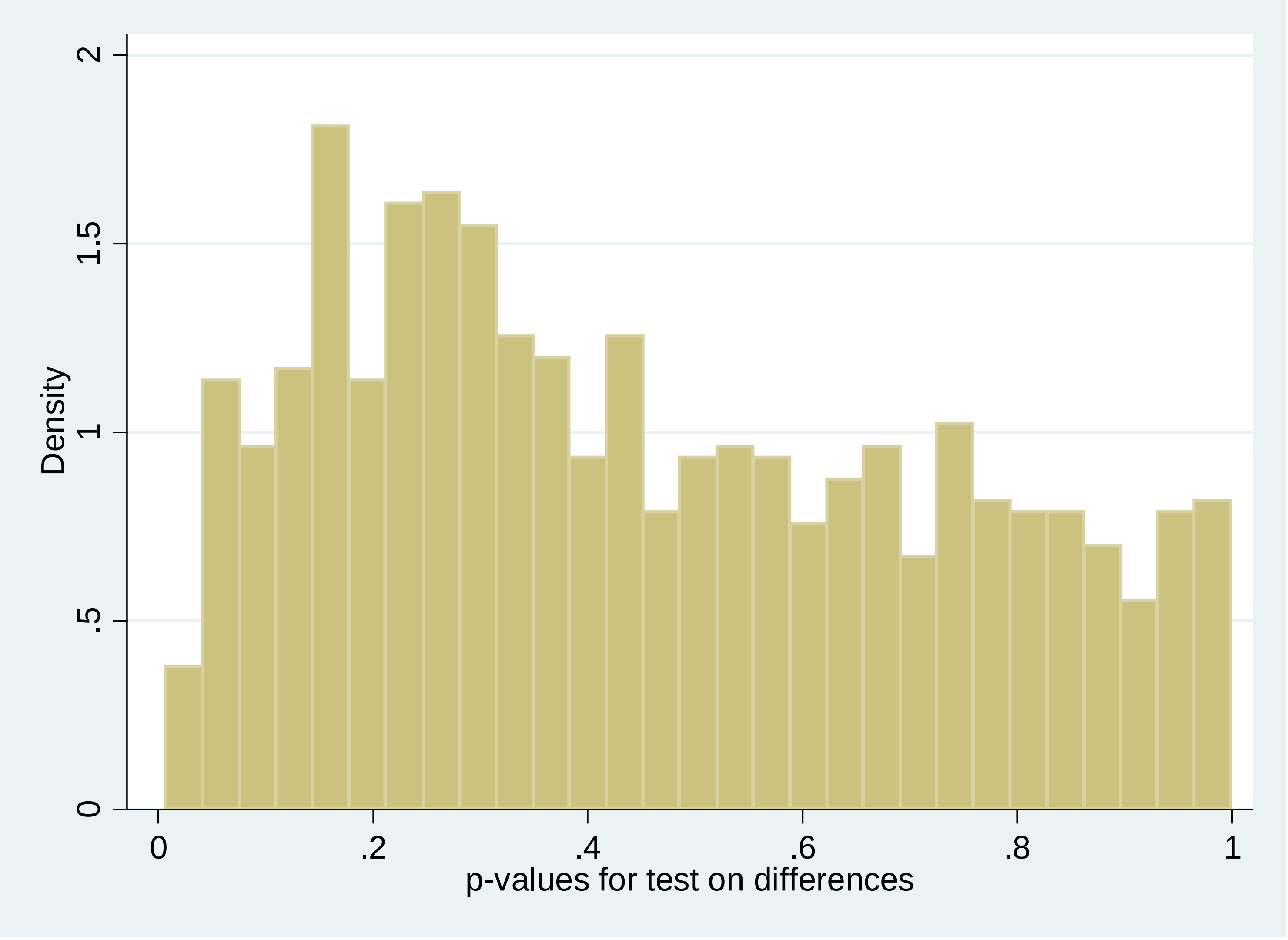
post mean = pre mean + Δ

n	Δ	pow(d)	pow(pct ch)	% ch mean	Ψ
20	-11.5	0.82	0.74	-24.5	-11.1
30	-9.2	0.80	0.65	-10.0	-8.3
40	-7.8	0.78	0.59	-8.5	-6.5
60	-6.4	0.79	0.56	-7.0	-5.5

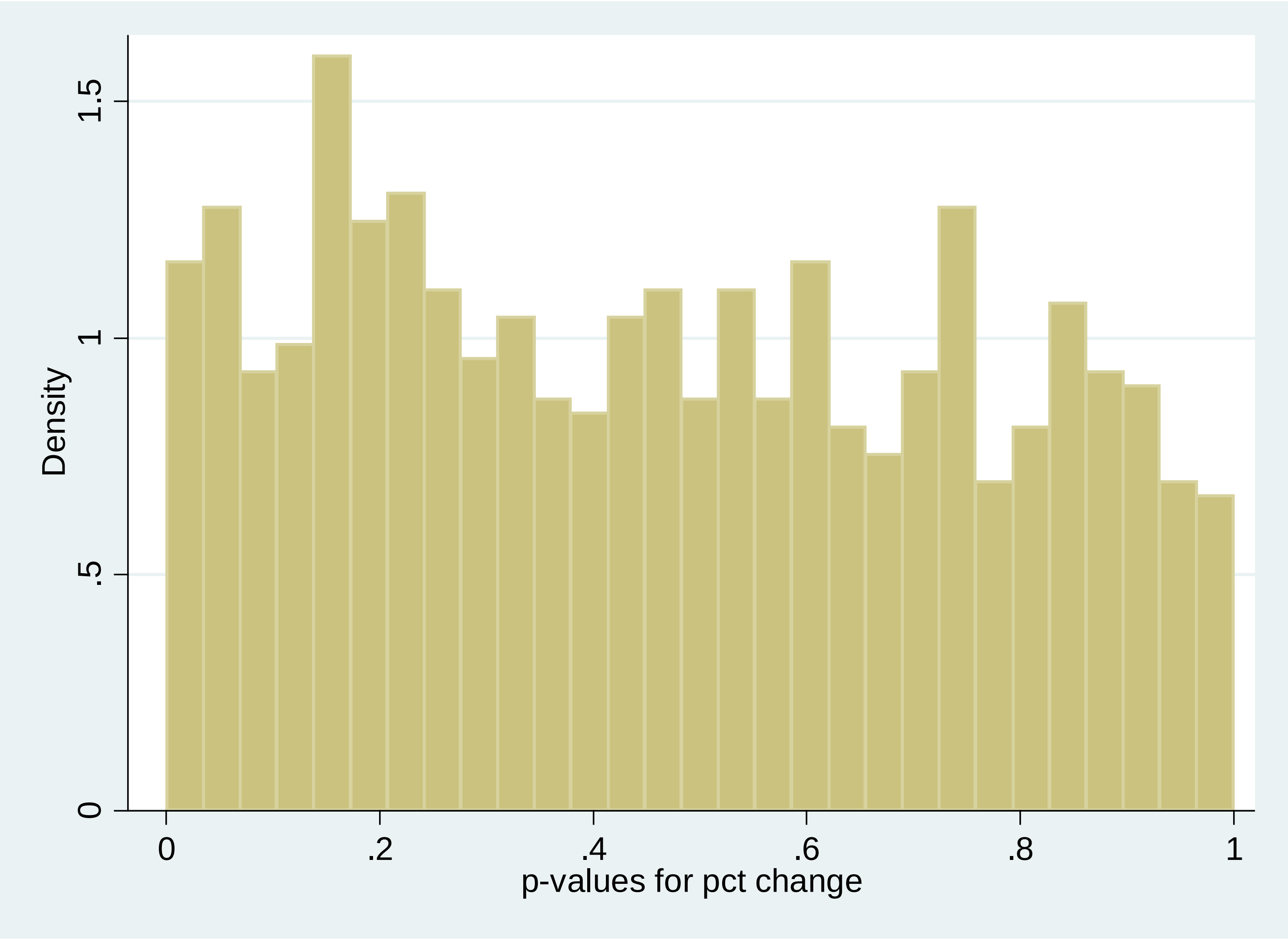
What about data varying over several orders of magnitude?

Should we analyze logs or percent changes?

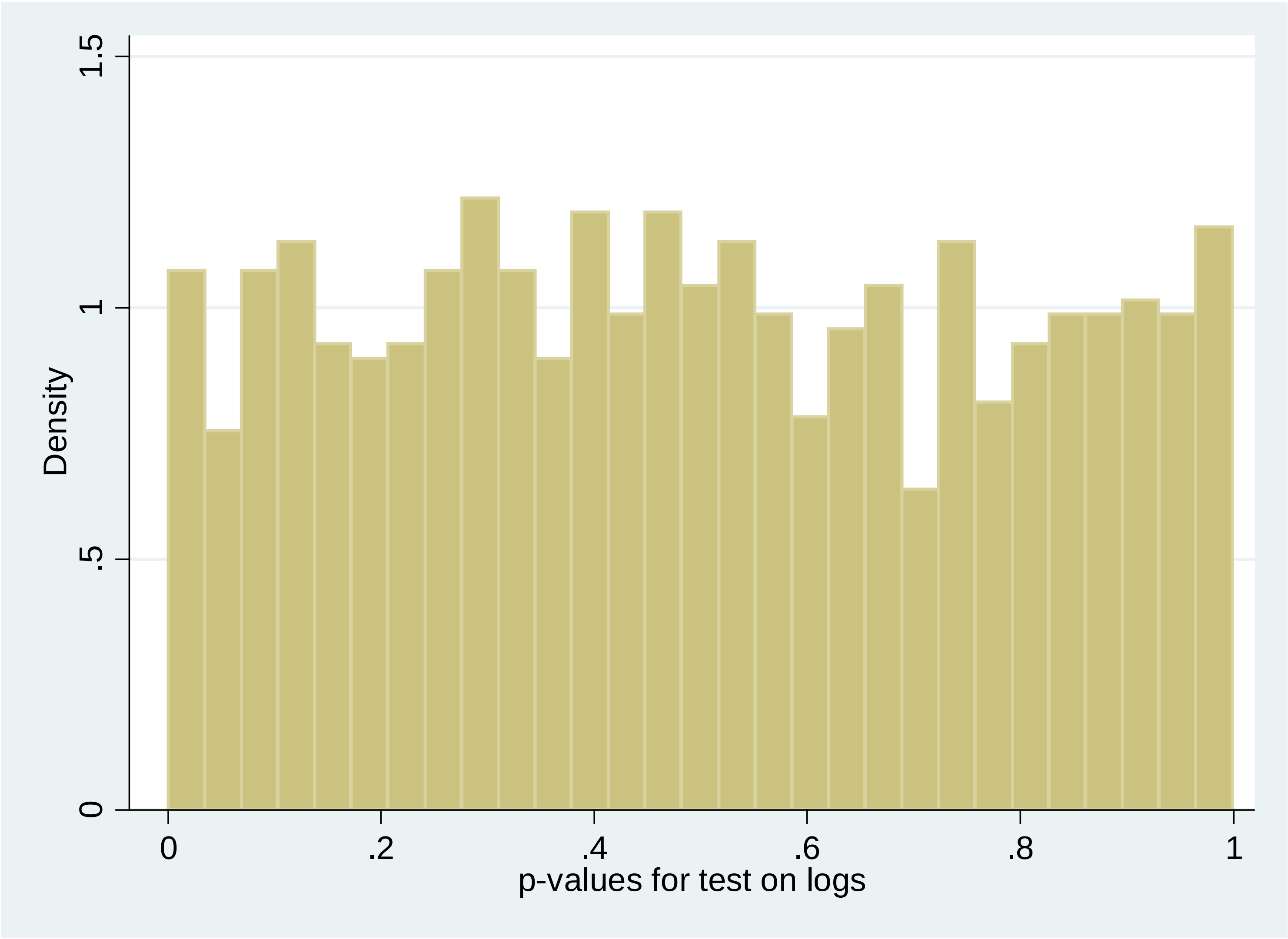
t-test on differences



t-test on pct change



t-test on differences of logs



Comparing experimental groups of subjects

sub	pre	post	diff	group
1	36.4	31.5	-4.9	C
2	48.8	40.1	-8.7	C
3	25.9	19.6	-6.3	C
…	…	…	…	C
10	40.5	38.2	-2.3	C
1	36.4	31.5	-4.9	E
2	48.8	46.1	-2.7	E
3	25.9	25.6	-0.3	E
…	…	…	…	E
15	37.7	37.0	-0.7	E

Ψ_1

Ψ_2

Comparing experimental groups of subjects

Equal CV’s

Bias in average observed % change approximately cancels so result of inference using % change as data and comparing means are about the same as t-test on differences.

Comparing experimental groups of subjects

Equal CV’s

Bias in average observed % change approximately cancels so result of inference using % change as data and comparing means are about the same as t-test on differences.

But what does it mean to report average values of % change for each group?

Comparing experimental groups of subjects

Unequal CV’s

Biases in averages of observed % change do not cancel and thus inference itself is biased (size of test not 0.05)

Conclusions

If outcome measure >0 and does not vary over several orders of magnitude, analyze original data.

You can estimate or make inference on the % change in the mean response from the above
analysis.

Conclusions (cont.)

Within an experimental group, there is a positive bias when analyzing individual percent changes as data.

- Average observed percent change does not estimate either the population mean percent change nor the percent change in the population means.
- Type I error rate > α (e.g. 0.05) when no actual effect
- Reduced power when there really is an effect
- Effect of bias on inference becomes more evident as sample size increases.

Conclusions (cont.)

When comparing experimental groups, if CV's are similar, much of the bias cancels, thus the results of inference on mean % change is similar to results of inference on differences or ANOVA.

However difficult to characterize each group in terms of average observed % change because of the individual biases.

If CV's differ, inference on % change is biased (α-level not 0.05).

Conclusions (cont.)

If outcome measure >0 and *does* vary over several orders of magnitude, analyze logs of original data.

- stabilizes variance
- ANOVA, t-tests perform as advertised
- ANOVA, t-tests perform poorly on original data
- analysis of pct change is also biased as before

Next time:

- Improving estimation of mean percent change (μ_{PC})
- Using regression models that use pre-data to help explain post-data.
- Regression to the mean and how it can lead to misleading conclusions.
- Instrumental variables to correct bias in regression models with random errors in predictors.