

Comparing Means in Two Paired Samples: An Example

Brady T. West

Example: Comparing Means in Paired Samples

Background:

NHANES researchers want to make sure that measures of blood pressure are reliable across subgroups
→ each NHANES respondent had two measures collected

Example: Comparing Means in Paired Samples

Research Question:

For female Hispanic adults living in U.S. in 2015-2016, did two measures of systolic blood pressure *differ significantly*?

Expectation = no!

Inference Approaches:

- Form a confidence interval for the **mean difference**
- Perform a paired t-test for the **mean difference**
- Be sure to check assumptions!

Approach 1: Form a Confidence Interval

Compute difference in SBP measures for each woman

$$\text{difference} = \text{SBP2} - \text{SBP1} \rightarrow$$

mean difference: -0.977, standard deviation = 4.848, $n = 911$

- **Best Point Estimate:** sample mean difference is -0.977 mmHg
- **Interpretation:** In 2015-2016, we estimate the mean difference in systolic blood pressures for all female Hispanic adults was -0.977 mmHg.

Approach 1: Form a Confidence Interval

Compute difference in SBP measures for each woman

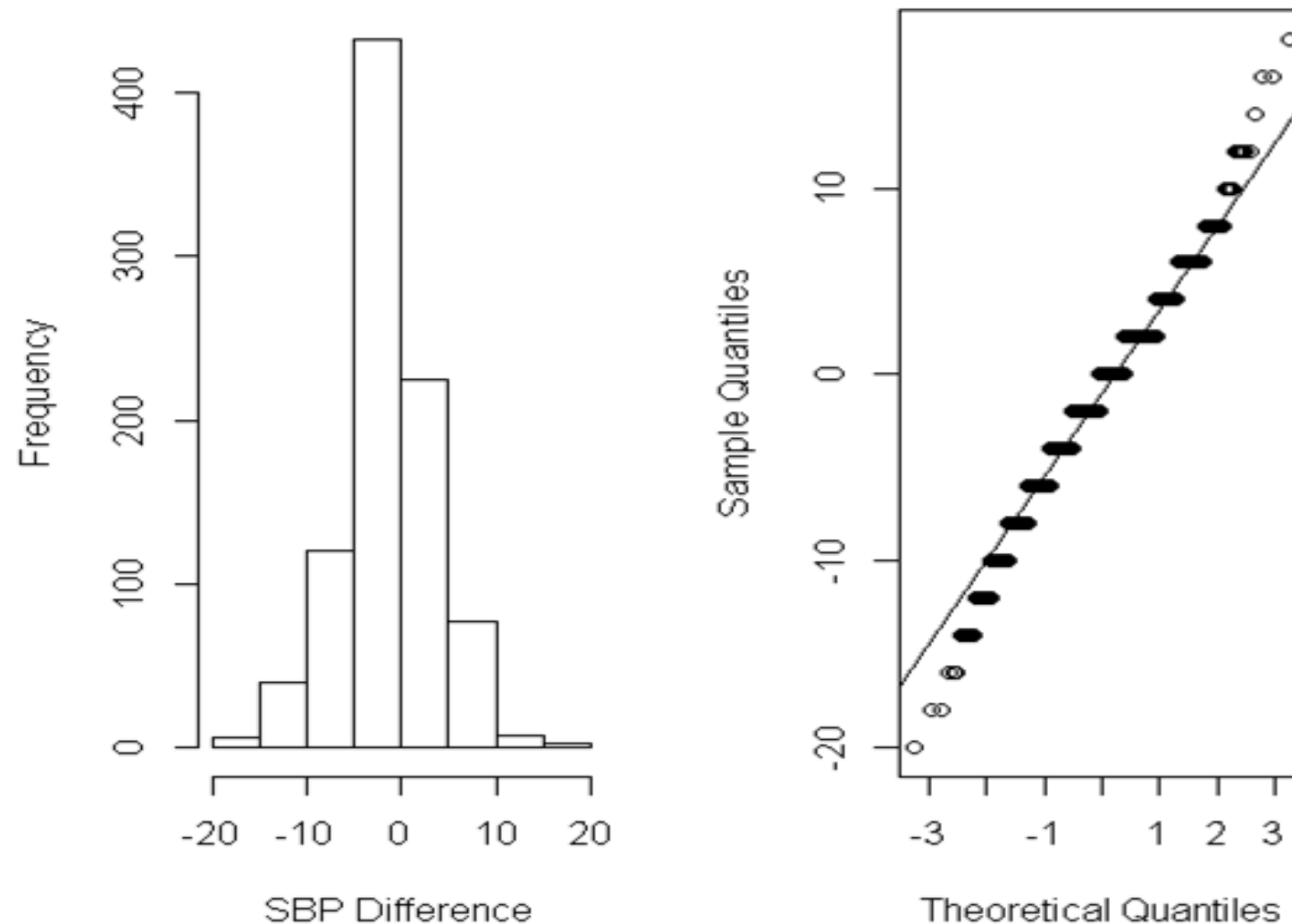
$$\text{difference} = \text{SBP2} - \text{SBP1} \rightarrow$$

mean difference: -0.977, standard deviation = 4.848, $n = 911$

Note: on average, the *first measurements* were larger,
by nearly 1 mmHg!

Let's examine the data more and
check some assumptions.

Check Assumptions: Normality

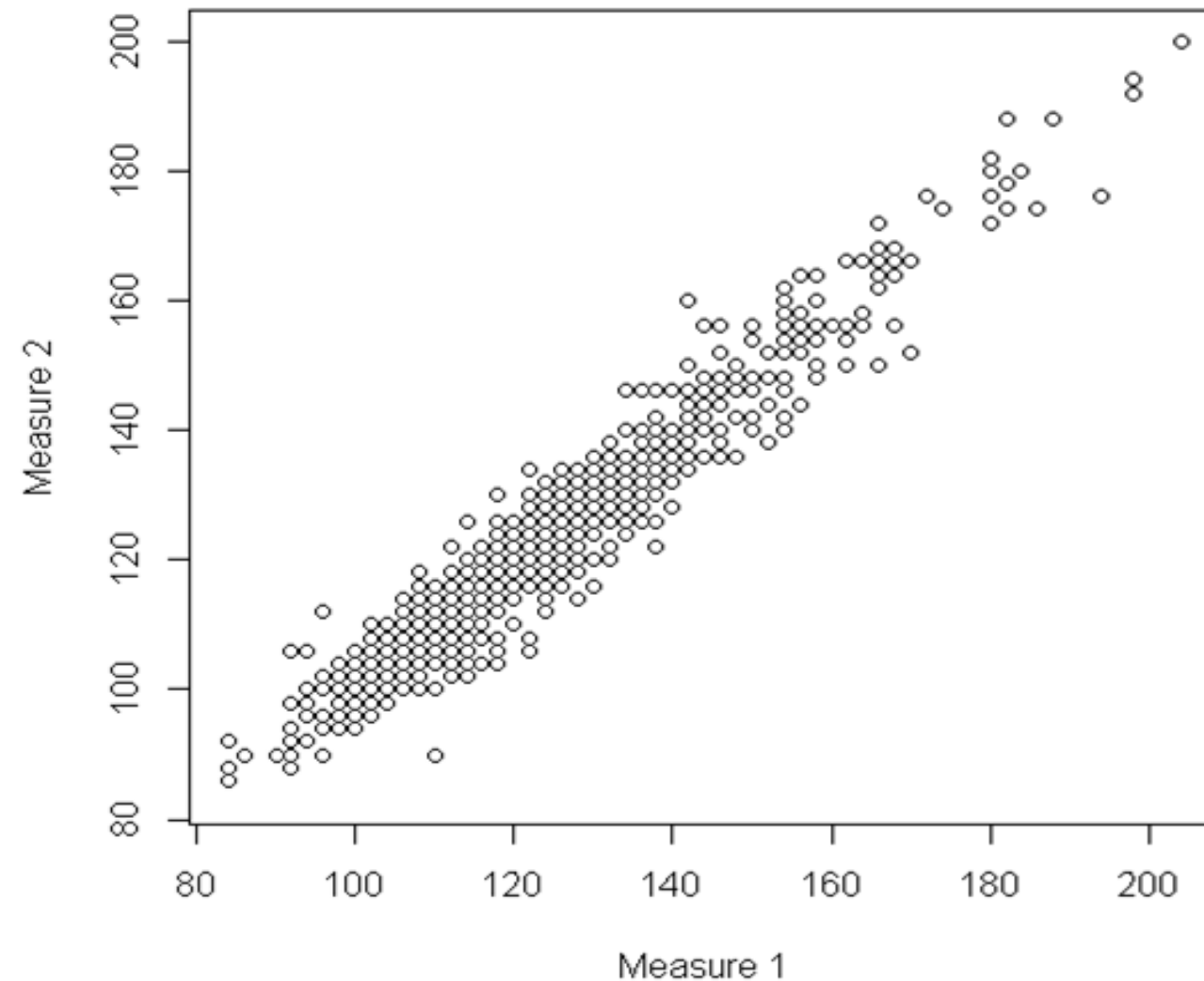


Histogram and Normal Q-Q plot suggest slight deviations from **normality**

(Recall: If distribution of differences in population was normal, expect all points to lie near 45-degree line in Q-Q plot)

Note: large sample size + CLT → normality assumption not critical!

Examine the Data: Correlation



Strong evidence of a **correlation** between two measures of SBP

Pearson correlation coefficient = 0.966

Clear evidence of these two measures being **paired** → supports paired-sample t-test procedures as appropriate!

Approach 1: Form a Confidence Interval

Best Estimate \pm Margin of Error

Best Estimate \pm “a few” (estimated) standard errors

- **Sample Mean** of the 911 differences in SBP = -0.977 mmHg
 - **Sample standard deviation** of the 911 differences in SBP = 4.848 mmHg
- Estimated standard error*** = $\frac{4.848}{\sqrt{911}} = 0.161$ mmHg

Note: Sample mean difference seems **quite large** relative to its standard error

Approach 1: Form a Confidence Interval

95% confidence interval for the population mean difference in systolic blood pressure of all female Hispanic adults living in U.S. in 2015-2016 is:
(-1.292 mmHG, -0.662 mmHg)

- Interval doesn't include 0 → **Significant difference!**
- **Inference:** Evidence that the *first measure tends to be significantly larger than the second measure* (for this subgroup)

Why might this be?

Approach 2: Paired Samples t-test

- **Null:** Population mean difference in measurements is 0
(*two measurements are identical to each other, on average*)
- **Alternative:** Population mean difference is not 0
(*two measurements are different, on average*)

Alternative allows first measurement to be
either greater or less than the second (on average)
→ **two-tailed test**

**Significance
Level = 5%**

Approach 2: Paired Samples t-test

Assumptions:

- Sample of **differences** considered a **simple random sample**
- **Normal** distribution of differences in blood pressure
(not as critical given large sample size)

Examine data: Assess if **paired measures** are in fact **correlated**
(recall that the previous graph supports this assumption!)

Approach 2: Paired Samples t-test

Result under stated assumptions:

$$t = -6.082, df = 910 (911 - 1), p\text{-value} < 0.001$$

We reject the null hypothesis →
support the population mean difference in SBP not equal to 0
Evidence the *first SBP measure tends to be*
significantly different than the SBP second measure on average
(for the population represented by this sample)

What if Normality Doesn't Hold?

- Not convinced that the differences follow a normal distribution?
→ **non-parametric test** that does not assume normality
- Non-parametric analogue of the paired samples t-test
= **Wilcoxon Signed Rank Test**
~ uses median to examine location of distribution of differences

What if Normality Doesn't Hold?

Wilcoxon Signed Rank Test Result: $p\text{-value} < 0.001$

- We reject the null that both measures have identical medians

Conclusion is robust to potential violations of normality!

Consistent evidence the two measures of systolic blood pressure *differ significantly* for the population of interest
~ regardless of assumptions made and inference approach used
→ appears the two measures aren't reliable!

What's Next?

How to compare **two proportions**
based on ***independent samples***