Outline

Example: Pairie Dogs

Table: Bark distance (m)

urban	rural
29	40
10	47
15	38
41	59
18	45
18	52
12	57
45	50
34	50
30	49
22	50
26	43
18	

Notation

- Y_{1i}: Random variable of the ith response in the first sample for $i = 1, \ldots, n_1$.
- Y_{2i}: Random variable of the ith response in the second sample for $i = 1, \ldots, n_2$.
- $\mu_1 = E(Y_{1i})$: Population mean response of the first group.
- $\mu_2 = E(Y_{2i})$: Population mean response of the second group.
- Our goal is to test

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_A : \mu_1 \neq \mu_2.$$

Assumptions

#1 The first sample $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ is an i.i.d. sample of size n_1 from $N(\mu_1, \sigma_1^2)$.

#2 The second sample $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ is an i.i.d. sample of size n_2 from $N(\mu_2, \sigma_2^2)$.

#3 The two samples $\{Y_{1i}\}$ and $\{Y_{2i}\}$ are independent.

#4 The (unknown) variances are the same $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Test Statistic

- To test
$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_A : \mu_1 \neq \mu_2.$$
- The main idea is to consider the difference in mean
$$\bar{Y}_1 - \bar{Y}_2$$
and construct a T-type test statistic
$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - \mathbb{E}_0(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{Var(\bar{Y}_1 - \bar{Y}_2)}}$$

Test Statistic

Under the null hypothesis:

- What is the distribution of $\bar{Y}_1$?

$$\bar{Y}_1 \sim N(\mu_1, \frac{\sigma^2}{n_1})$$

- What is the distribution of $\bar{Y}_2$?

$$\bar{Y}_2 \sim N(\mu_2, \frac{\sigma^2}{n_2})$$

- What is the expectation of $\bar{Y}_1 - \bar{Y}_2$?

$$\mu_{\bar{Y}_1 - \bar{Y}_2} = E(\bar{Y}_1 - \bar{Y}_2) = E(\bar{Y}_1) - E(\bar{Y}_2) = \mu_1 - \mu_2$$

- What is the variance of $\bar{Y}_1 - \bar{Y}_2$?

$$\sigma^2_{\bar{Y}_1 - \bar{Y}_2} = Var(\bar{Y}_1) + Var(\bar{Y}_2) = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$$

Estimated Variance of $\bar{Y}_1 - \bar{Y}_2$

- If S_p^2 is an estimator of $\sigma^2 = \text{Var}(Y_1) = \text{Var}(Y_2)$, then we can estimate $\sigma^2_{\bar{Y}_1 - \bar{Y}_2}$ by

$$\text{Var}(\bar{Y}_1 - \bar{Y}_2) = S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

- Consider a **pooled variance estimate** of σ^2:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, \quad \text{where}$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (Y_{1i} - \bar{Y}_1)^2, \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_{2i} - \bar{Y}_2)^2$$

- Interpretation of S_p^2: A weighted average of the two sample variances, weighted by the d.f.'s.

T Test Statistic

- Our goal is to test

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_A : \mu_1 \neq \mu_2$$

- Under the H_0, the test statistic follows t-distribution with df=$n_1 + n_2 - 2$.

$$t = \frac{\bar{Y}_1 - \bar{Y}_2 - 0}{\sqrt{S_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim T_{n_1+n_2-2}.$$

- A $(1-\alpha)$ CI for $\mu_1 - \mu_2$ is

$$\bar{Y}_1 - \bar{Y}_2 \pm t_{n_1+n_2-2,\alpha/2} \times \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

where $t_{n_1+n_2-2,\alpha/2}$ is the critical value for which $\mathbb{P}(|T_{n_1+n_2-2}| \geq t_{n_1+n_2-2,\alpha/2}) = \alpha$.

Example: Bark Distance

- From the bark distance data, we have
  $\bar{y}_1 = 48.33, s_1^2 = 38.97, n_1 = 12$ and $\bar{y}_2 = 24.46, s_2^2 = 118.77, n_2 = 13.$
- The pooled sample variance is:

$$s_p^2 = \frac{(12-1)\times 38.97 + (13-1)\times 118.77}{12+13-2} = 80.60$$

- The observed test statistic is:

$$t = \frac{48.33 - 24.46 - 0}{\sqrt{80.60\left(\frac{1}{12}+\frac{1}{13}\right)}} = 6.642$$

- The degrees of freedom are: df = $12 + 13 - 2 = 23$.
- The p-value is: The p-value is $2 \times P(T_{23} \geq 6.642)$, which is less than 0.002.
- The conclusion is: Reject H_0 at the 5% level. There is very strong evidence that the mean bark distances for rural and urban prairie dogs are different.

Outline

Levene's Test

An example:

- Consider two independent samples Y_1 and Y_2:

  Sample 1: 4, 8, 10, 23
  Sample 2: 1, 2, 4, 4, 7

- Test $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_A : \sigma_1^2 \neq \sigma_2^2$.

Levene's test for comparing two population variances:

(1) Find the median for each sample.

(2) Subtract the median from each observation.

(3) Take absolute values of the residual.

(4) Perform an unpaired T test on the modified samples, denoted as Y_1^* and Y_2^*.

Main idea: Test for equal variances $\rightarrow$ test for the equal absolute deviation.

Example: Levene's Test

(1) Find the median for each sample.
Sample 1:
Sample 2:

(2) Subtract the median from each observation.
Sample 1:
Sample 2:

(3) Take absolute values of the results.
Sample 1*:
Sample 2*:

Example: Levene's Test

(4) Perform an unpaired T test on the modified samples, denoted as Y_1^* and Y_2^*.

Here $\bar{y}_1^* = 5.25, \bar{y}_2^* = 1.60, s_1^{2*} = 37.58, s_2^{2*} = 2.30$.

Outline

Assumptions: Overview

- Statistical methods for statistical inference (e.g., hypothesis testing and confidence intervals) rely on assumptions.
- A statistical method is said to be **robust** against a failure in an assumption if the method does not depend critically on that assumption.
- That is, if the assumption does not hold exactly, a **robust method** gives results that are still a good approximation to the correct answer.
- Otherwise, we say that the method is **not robust** against a failure in an assumption or is **sensitive** to the assumption.

One-sample T Test

- Assumptions: The observations Y_i are independent and are from a normal population.
- Independence:
    - The one-sample T test is sensitive to the independence assumption.
    - Possible outcomes: The correct p-value is substantially larger than the calculated p-value.
- Normality:
    - The one-sample T test is robust against non-normality.
    - If the population is not exactly normal but symmetric, the T test still gives a good approximation to the correct result. [Why?]
- Outlier: The one-sample T test is sensitive to outliers.

Paired T Test

- The paired T test can be viewed as a one-sample T test based on the differences of the original paired two samples.
- Assumptions of D_i: Same as the one-sample T test (i.e., independence and normality).
- The test is robust against non-normality, but not against dependence.
- Assumptions of (Y_{1i}, Y_{2i}):
  - The pairs are independent of each other pair, but not between Y_{1i} and Y_{2i}.
  - No explicit assumptions about the distributions of Y_{1i} and Y_{2i}.
  - Assumption made on $Y_{1i} - Y_{2i}$.

Unpaired Two Sample T Test

- Assumptions: Normality for each sample, equal variance, independence within each sample, and independence between the two samples.
- The test is robust against non-normality, especially if the two populations being sampled are symmetric.
- The test is not robust against dependence.
- The relative difference in sample sizes plays a role in the robustness of the equal variance assumption.
- If the sample sizes are approximately equal, then the T test is robust against differences in variance.
- Outlier: not resistant against outliers

Assessment of Assumptions

- Assessing independence
- Assessing normality
- Assessing equal variance

Assessing Independence

- Consider how the data were collected; that is, the study or experimental design.
- Paired two sample versus unpaired two sample studies.
- The notion of independent observations is closely related to the idea of an i.i.d. sample.
- Examples of dependence:
- Not always obvious, as the answer depends on the kinds of scientific questions of interest and the corresponding populations under study.

Assessing Normality

- Histogram
- Difficulties
  - Small sample size:
  - How to distinguish between bell-shaped and mound-shaped, but non-normal distribution?

QQ Plot

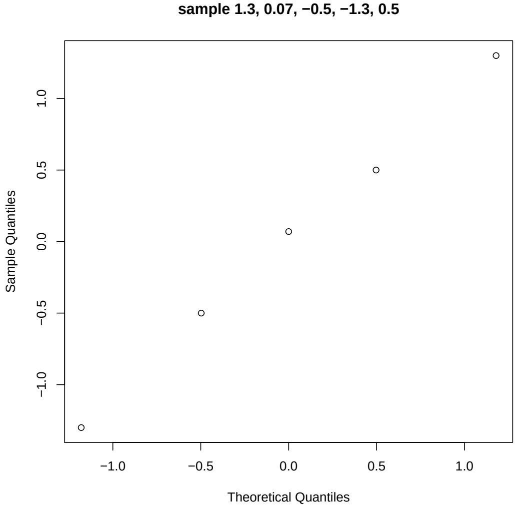
- A **quantile-quantile (QQ) plot** is more reliable tool to assess normality.
- A QQ plot is also known as a **quantile comparison plot.**
- How to construct a QQ plot for a sample $y_1, y_2, \ldots, y_n$?
  (1) Sort the observations in ascending order:
  $$y_{(1)}, y_{(2)}, \ldots, y_{(n)}.$$
  (2) Compute quantiles of $N(0,1)$:
  $$z_{[(1-.5)/n]}, z_{[(2-.5)/n]}, \ldots, z_{[(n-.5)/n]}.$$
  (3) Plot pairs of
  $$(z_{[(1-.5)/n]}, y_{(1)}), (z_{[(2-.5)/n]}, y_{(2)}), \ldots, (z_{[(n-.5)/n]}, y_{(n)}).$$

Example: One Sample of Size 5

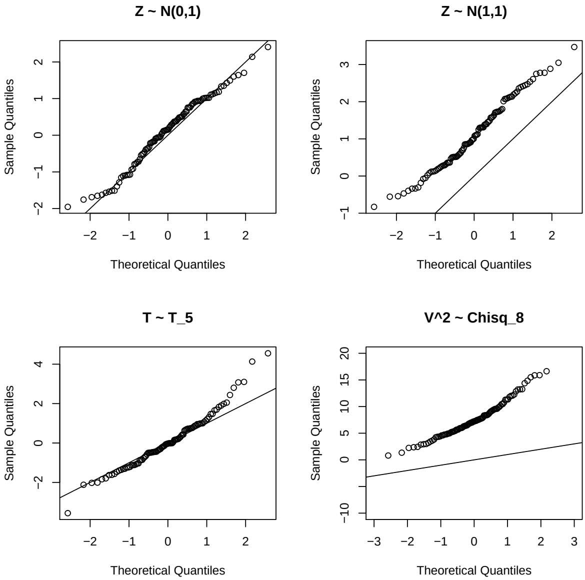
- A small sample of size 5: $1.3, 0.07, -0.5, -1.3, 0.5$.
  (1) Sort the observations in ascending order:
  $$-1.3, -0.5, 0.07, 0.5, 1.3$$
  (2) Compute quantiles of $N(0,1)$:
  $$z_{[0.1]} = -1.28, z_{[0.3]} = -0.52, z_{[0.5]} = 0, z_{[0.7]} = 0.52, z_{[0.9]} = 1.28$$
  - R code: qqnorm( ); compare to 45^o line.

Example: One Sample of Size 5

(3) Plot the quantile pairs:



Example: Samples of Size 100

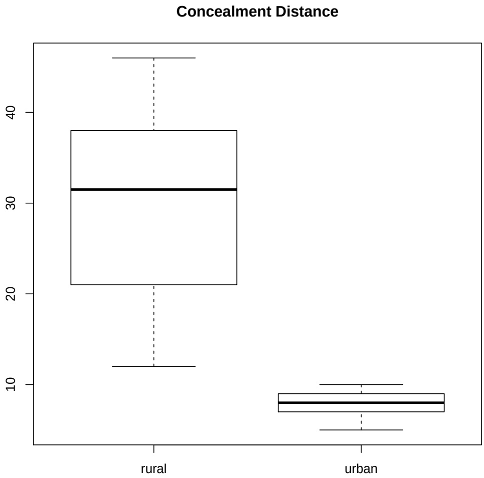


Assessing Equal Variance

- Equal variance is also called **homoscedasticity** and unequal variance **heteroscedasticity**.
- For an unpaired T test:
  - Robust against unequal variances if the sample sizes are roughly equal.
  - Rule of thumb: If the ratio of s_1^2 and s_2^2 is within two or three, there is probably little need to pursue a formal test.
- Levene's test.
- Graphical methods: histogram, (side-by-side) box plot.

Outline

Example: Concealment Distance



Notation

- Y_{1i}: Random variable of the ith response in the first sample for $i = 1, \ldots, n_1$.
- Y_{2i}: Random variable of the ith response in the second sample for $i = 1, \ldots, n_2$.
- $\mu_1 = E(Y_{1i})$: Population mean response of the first group.
- $\mu_2 = E(Y_{2i})$: Population mean response of the second group.
- Our goal is to test

$$H_0 : \mu_1 = \mu_2$$

vs.

$$H_A : \mu_1 \neq \mu_2.$$

Assumptions

- #1 The first sample $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ is an i.i.d. sample of size n_1 from $N(\mu_1, \sigma_1^2)$.
- #2 The second sample $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ is an i.i.d. sample of size n_2 from $N(\mu_2, \sigma_2^2)$.
- #3 The two samples $\{Y_{1i}\}$ and $\{Y_{2i}\}$ are independent.
- #4* The variances are not the same $\sigma_1^2 \neq \sigma_2^2$.

Test Statistic

- To test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$, we test

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_A : \mu_1 - \mu_2 \neq 0$$

- The main idea is still to use $\bar{Y}_1 - \bar{Y}_2$ and construct a T-type test statistic

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - \mu_{\bar{Y}_1 - \bar{Y}_2}}{S_{\bar{Y}_1 - \bar{Y}_2}}$$

- Under the assumptions above:
  - Distribution of $\bar{Y}_1$: $\bar{Y}_1 \sim N(\mu_1, \sigma_1^2/n_1)$
  - Distribution of $\bar{Y}_2$: $\bar{Y}_2 \sim N(\mu_2, \sigma_2^2/n_2)$
  - Expectation of $\bar{Y}_1 - \bar{Y}_2$: $\mu_1 - \mu_2$
  - Variance of $\bar{Y}_1 - \bar{Y}_2$: $\sigma_1^2/n_1 + \sigma_2^2/n_2$

Test Statistic

- The main idea is still to use a T-type test statistic, but now

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

- Under the H_0, the test statistic

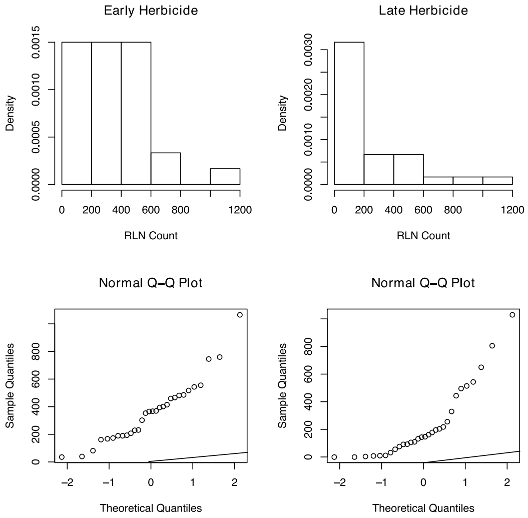
$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

does not follow $T_{n_1+n_2-2}$!

Test Statistic

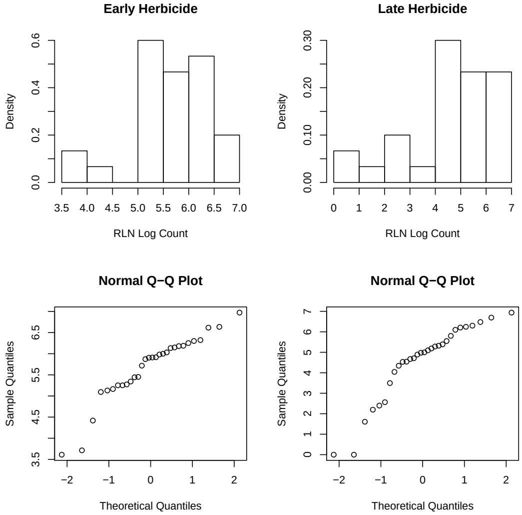
- However, the distribution can be approximated by a T_{adf} where
$$\text{adf} \approx \frac{(r_1 + r_2)^2}{\frac{r_1^2}{n_1-1} + \frac{r_2^2}{n_2-1}}$$
where
$$r_1 = \frac{s_1^2}{n_1}$$
and
$$r_2 = \frac{s_2^2}{n_2}.$$
- This test is known as the **Welch's** T **test.**

Example: RLN Count



Normality holds?

Example: RLN ln(Count)



Example: RLN ln(Count)

- Apply a log transformation to the count data: $n_1 = n_2 = 30$ and

$$\bar{y}_1^{\dagger} = 6.67, \bar{y}_2^{\dagger} = 4.55, (s_1^{\dagger})^2 = 0.60, (s_2^{\dagger})^2 = 3.32$$

- Expected log counts: $\mu_1^{\dagger}$ versus $\mu_2^{\dagger}$.
- Levene's test:

$$f = 7.34 \text{ on df = 58; p-value = 0.008849.}$$

- Welch's T test (why not unpaired T-test?):

$$t = 3.10 \text{ on df = 39.165; p-value = 0.003549.}$$

- 95% CI: $[0.39, 1.85]$.

Remarks on Transformation

- Transform the data so that the transformed data might align better with the assumptions than the original data.
- Transformation affects normality and equal variance.
- For count data, log or square root transformations are common.
- Caution: Transform so that the assumptions of the analysis are better met, not that the p-value is the smaller.

Example: Weight Gain

Subject	Diet A (Y_1)	Diet B (Y_2)
1	37.8	12.3
2	27.5	14.3
3	41.2	19.2
4	26.5	4.0
5	28.6	25.9
Mean	32.32	15.14
Variance	44.96	66.28
SD	6.70	8.14

Example: Weight Gain

- Levene's test:

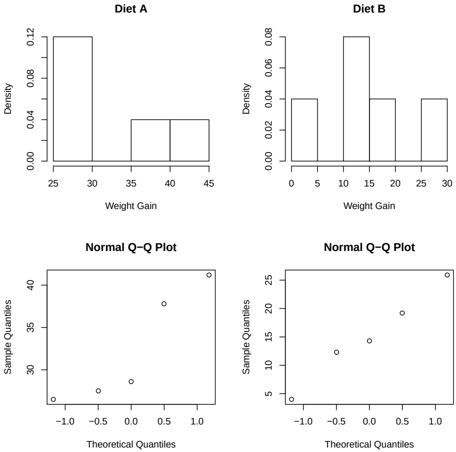
$$f = 0.0509 \text{ on df = 8; p-value = 0.8271.}$$

- Unpaired T test (why not Welch's T test?):

$$t = 3.64 \text{ on df = 8; p-value = 0.006567.}$$

- 99% CI: $[1.35, 33.01]$.

Example: Weight Gain



Randomization Test

- Goal: $H_0 : \mu_1 = \mu_2$ vs $H_0 : \mu_1 \neq \mu_2$.
- Re-arrange the data to look like:

Wt Gain	Diet
37.8	A
27.5	A
41.2	A
26.5	A
28.6	A
12.3	B
14.3	B
19.2	B
4.0	B
25.9	B

- Sample mean difference: $\bar{y}_1 - \bar{y}_2 = 32.32 - 15.14 = 17.18$.

Randomization Test

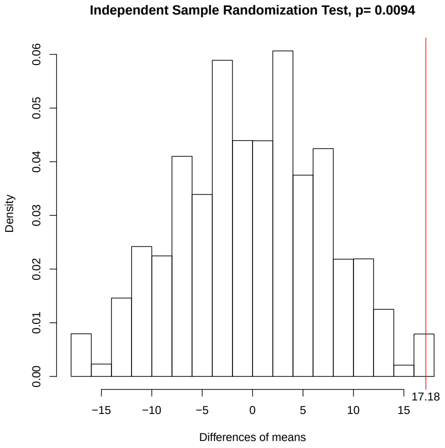
- Under $H_0 : \mu_1 = \mu_2$, randomly assign diet types A and B.
- Some possibilities are:

	Data	Random		
Wt Gain	Diet	1	2	3
37.8	A	B	A	B
27.5	A	A	B	A
41.2	A	A	B	A
26.5	A	A	B	B
28.6	A	A	A	A
12.3	B	B	A	A
14.3	B	A	B	B
19.2	B	B	B	B
4.0	B	B	A	B
25.9	B	B	A	A

Randomization Test

- Data: $\bar{y}_1 - \bar{y}_2 = 32.32 - 15.14 = 17.18.$
- Random 1: $\bar{y}_1 - \bar{y}_2 = 27.62 - 19.84 = 7.78$
- Random 2: $\bar{y}_1 - \bar{y}_2 = 21.72 - 25.74 = -4.02$
- Random 3: $\bar{y}_1 - \bar{y}_2 = 27.10 - 20.36 = 6.74$

Example: Weight Gain



Randomization Test

- Repeat 10,000 times, count the number of times that the sample mean difference is as large as or larger than the observed sample mean difference 17.18, and compute a p-value as
$$2 \times \frac{47}{10000} = 0.0094.$$
- Conclusion: There is strong evidence against the H_0 that the population mean weight gain is the same for the two diet types. Reject the H_0 at the $\alpha = 0.05$ level.
- The test is then known as a **permutation test**. Suitable for small-sample size problems.

Example: Blood Pressure

Subject	Before (Y_1)	After (Y_2)	Diff ($D = Y_1 - Y_2$)
1	90	88	2
2	100	92	8
3	92	82	10
4	96	90	6
5	96	78	18
6	96	86	10
7	92	88	4
8	98	72	26
9	102	84	18
10	94	102	-8
11	94	94	0
12	102	70	32
13	94	94	0
14	88	92	-4
15	104	94	10
Mean	95.87	87.07	8.80
Variance	21.41	75.92	120.46
SD	4.63	8.71	10.98

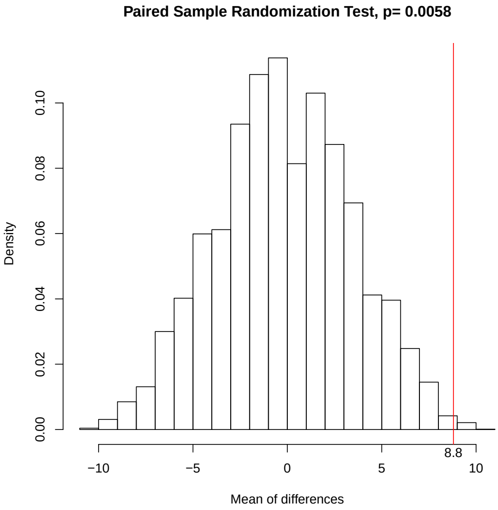
Example: Blood Pressure

- Paired T test:

$$t = 3.10 \text{ on df} = 14\text{; p-value} = 0.007749.$$

- 95% CI: $8.80 \pm 6.08 = [2.72, 14.88]$.
- How to perform randomization test?
    - Apply the randomization test idea as before, but when switching labels, preserve the pairing.

Example: Blood Pressure



Nonparametric Test

- In a **nonparametric test**, the distribution of the population can be arbitrary.
- The primary nonparametric test for comparing two independent samples is the **Wilcoxon test.**
- Our goal is to test

H_0 : The two populations have the same distribution,

vs.

H_A : The two populations have the same shape but different locations.

Wilcoxon Test (for Unpaired Two-Sample Test)

(1) Merge the two samples into one combined sample.

(2) Sort the observations in the combined sample in ascending order and assign each observation a **rank.**

(3) Add the ranks corresponding to observations from **the first treatment group** denoted as R_1 (Let r_1 be the observed test statistic).

(4) Using R_1 as a test statistic. Based on the observed r_1, conduct a randomization or approximate Z test.

Example: Weight Gain

(1) Merge the two samples into one combined sample.

Wt Gain	Diet
37.8	A
27.5	A
41.2	A
26.5	A
28.6	A
12.3	B
14.3	B
19.2	B
4.0	B
25.9	B

Example: Weight Gain

(2) Sort the observations in the combined sample in ascending order and assign each observation a rank.

Wt Gain	Diet	Rank
37.8	A	9
27.5	A	7
41.2	A	10
26.5	A	6
28.6	A	8
12.3	B	2
14.3	B	3
19.2	B	4
4.0	B	1
25.9	B	5

Example: Weight Gain

(3) Add the ranks corresponding to observations from the first treatment group. Observed rank sum is $r_1 = 40$

(4) Using R_1 as a test statistic and based on the observed r_1, conduct a randomization or approximate Z test. From R, p-value = 0.007937.

- Note that $R_1 + R_2 = (n_1 + n_2)(n_1 + n_2 + 1)/2$.

Wilcoxon Signed Rank Test

For paired two-sample test or one-sample test.

(1) Take the **sample of differences** and compute their absolute values.

(2) Sort the observations in the sample of absolute differences in ascending order and assign each observation a rank.

(3) Add the ranks corresponding to the positive values in the original data set and this is the observed test statistic.

(4) Perform a randomization test on this sum of the positive ranks.

Example: Blood Pressure (Simplified)

(1) Take the sample of differences and compute their absolute values.

d_i	$\lvert d_i \rvert$
2	2
8	8
-4	4

(2) Sort the observations in the sample of absolute differences in ascending order and assign each observation a rank.

d_i	$\lvert d_i \rvert$	Rank
2	2	1
8	8	3
-4	4	2

Example: Blood Pressure (Simplified)

(3) Add the ranks corresponding to the **positive values in the original data set** and this is the observed test statistic.

(4) Perform a permutation test on this sum of the positive ranks.

Subject	Data	Rank	Permutation							
1	+2	1	-2	-2	-2	-2	+2	+2	+2	+2
2	+8	3	-8	-8	+8	+8	-8	-8	+8	+8
3	-4	2	-4	+4	-4	+4	-4	+4	-4	+4
Rank sum		4	0	2	3	5	1	3	4	6

Remedial Measures

Methods/Robustness	**Independence**	**Normality**	**Equal Variance**	**Outlier**
Independent two sample or paired	Sensitive	Robust under conditions	Robust under conditions	Sensitive
Welch's T test	Sensitive	Robust under conditions	No need to assume	Sensitive
Independent two sample or paired randomization test	Sensitive	Robust	Robust	Sensitive
Wilcoxon rank sum or signed rank test	Sensitive	Robust	Robust	Robust