Interleaf 7: Which test should I use?

Data Type	Goal	Test
Categorical	Comparing proportion to hypothesized value	**Binomial test** (or χ^2 goodness of fit when n is too large or binomial test)
	Comparing frequency data to a probability distribution	**χ^2 goodness of fit**
Numerical	Comparing mean to hypothesized value (data are normal)	**One-sample t-test**
	Comparing median to hypothesized value when data are **not** normal	**Sign test**
	Comparing frequency data to a discrete probability distribution	**χ^2 goodness of fit**
	Comparing data to the normal distribution	**Shapiro-Wilk test**

Interleaf 7: Which test should I use?

Response variable	Explanatory variable	
	Categorical	**Numerical**
Categorical	Contingency analysis	Logistic regression
Numerical	T-test ANOVA Mann-Whitney U-test	• Linear and nonlinear regression • linear correlation •Spearman's rank

Interleaf 7: Which test should I use?

Number of treatments	Tests assume normal	Tests do not assume normal
Two treatments (independent)	Two-sample t-test Welch's t-test	Mann-Whitney U-test
Two treatments (paired data)	Paired t-test	Sign Test
> Two treatments	ANOVA	Kruskall-Wallis test

Chapter 13:

How do you detect violations from normality?

1. **Graphical Methods**

2. **Formal Tests of Normality**

How do you detect violations from normality?

1. Graphical Methods

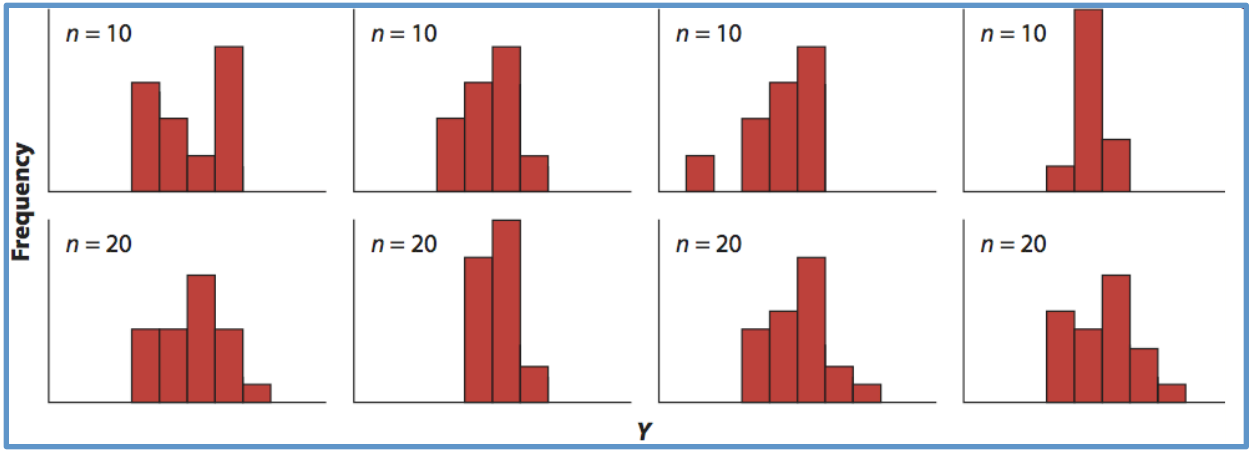
- Histogram
- Normal Quantile Plot

2. Formal Tests of Normality

How do you detect violations from normality?

1. Graphical Methods

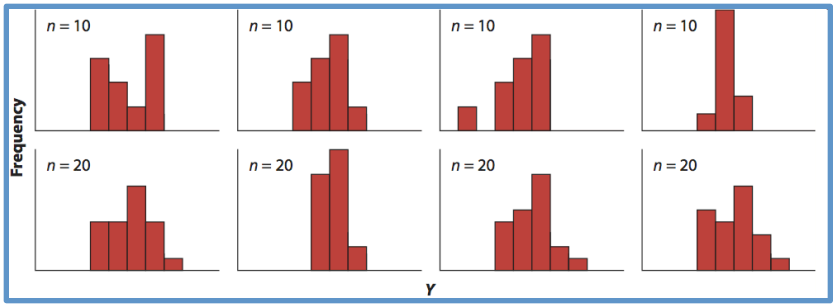
- Histogram

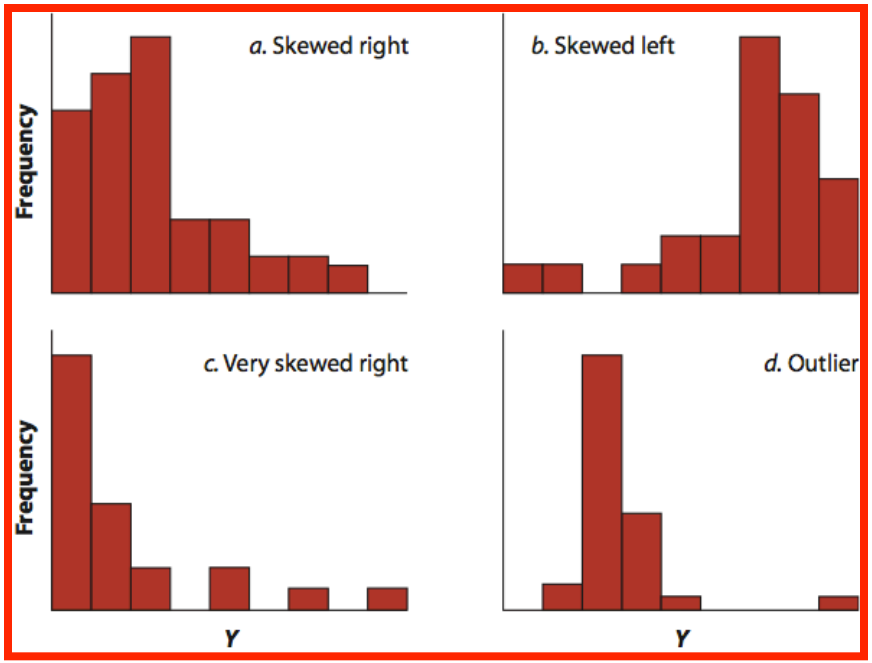


How do you detect violations from normality?

1. Graphical Methods

- Histogram





How do you detect violations from normality?

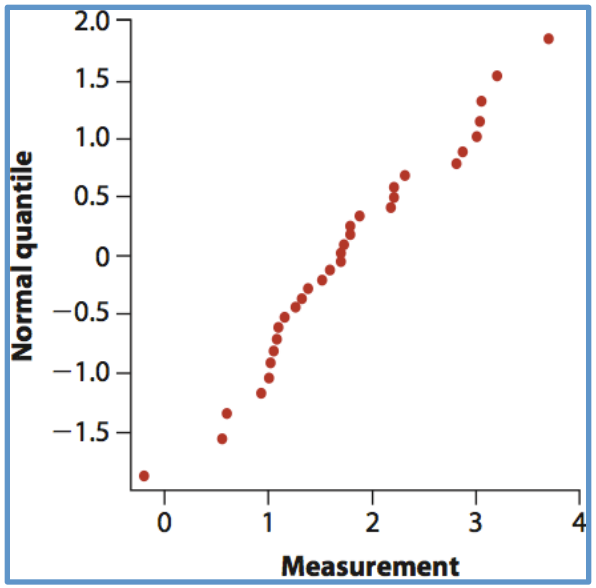
1. Graphical Methods

- Histogram
- Normal Quantile Plot
  - Compares each observation in a sample to the corresponding quantile expected under the **standard normal distribution** Method requires ranking
    - order points from smallest to largest and assign each rank "i"
    - Estimated proportion of distribution is i/n+1. Corresponding normal quantile is the standard normal deviate that has the same area under the curve (use appendix B – although, really, use a computer program that does this automatically)
  - The resultant points should follow a straight line if your observations are from a normal distribution!

How do you detect violations from normality?

1. Graphical Methods

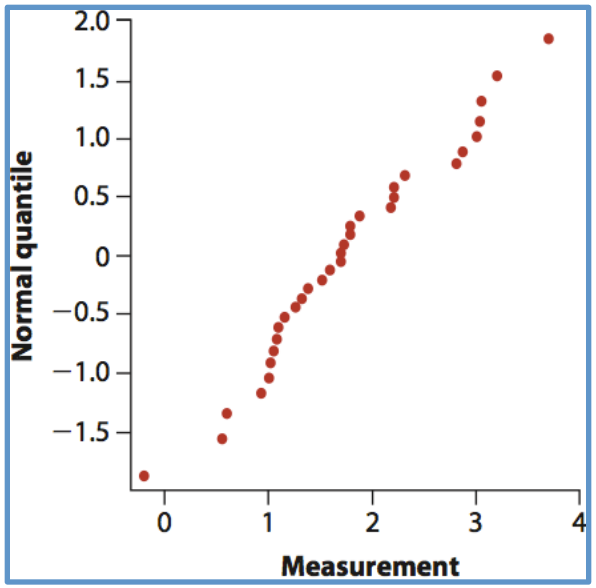
- Histogram
- Normal Quantile Plot
  - Compares each observation in a sample to the corresponding quantile expected under the standard normal distribution
  - The resultant points should follow a straight line if normal distribution

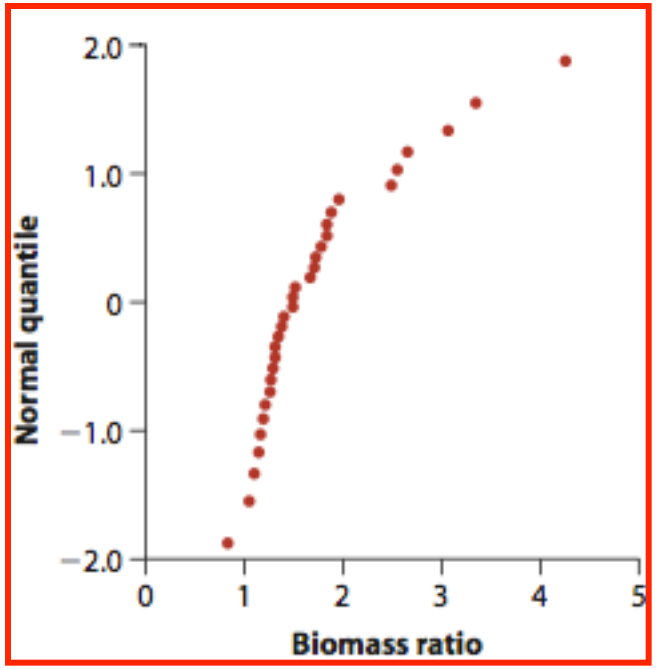


How do you detect violations from normality?

1. Graphical Methods

- Histogram
- Normal Quantile Plot
  - Compares each observation in a sample to the corresponding quantile expected under the standard normal distribution
  - Method requires ranking* - the resultant points should follow a straight line





How do you detect violations from normality?

1. Graphical Methods

- Histogram
- Normal Quantile Plot

2. Formal Tests of Normality

H_0: data are sampled from a population having a normal distribution

H_A: data are sampled from a population which does not have a normal distribution

How do you detect violations from normality?

1. Graphical Methods

- Histogram
- Normal Quantile Plot

2. Formal Tests of Normality

H_0: data are sampled from a population having a normal distribution

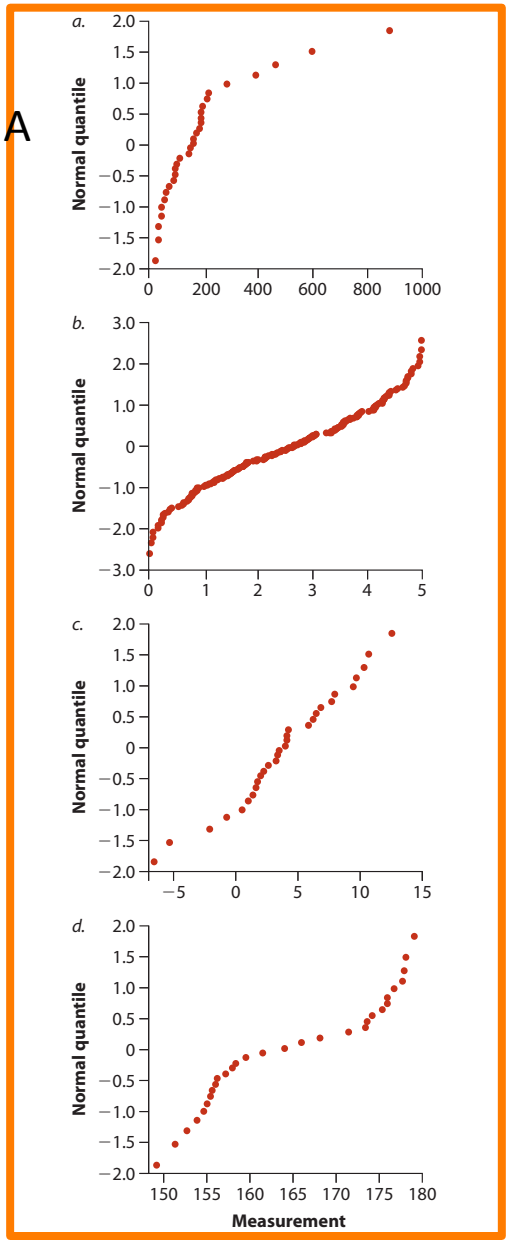
H_A: data are sampled from a population which does not have a normal distribution

- Previous data/theory
- Shapiro-Wilk test
  - General approach: estimates the mean and standard deviation of the population using the sample data
  - Tests goodness-of-fit to the data of a normal distribution that has the same mean and standard deviation

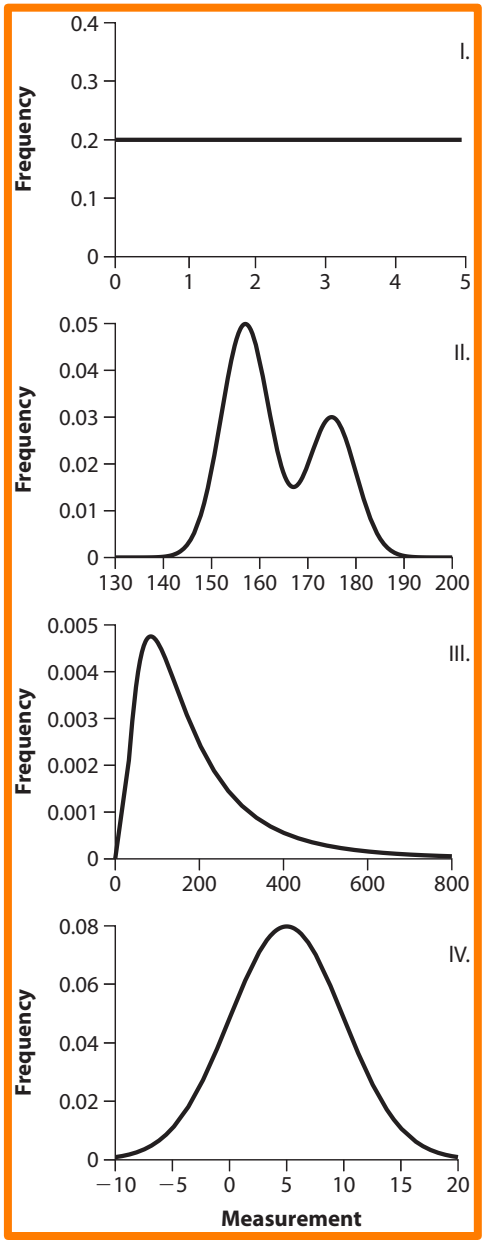
What about tests for unequal variance?

- how to test for ***equal variances*** *–Levene's and F tests*
- We only worry about unequal variances among tested samples if:
  - $n_1 >> n_2$ (or vice versa)
  - $\sigma_1 \geq 3\ \sigma_2$ (or vice versa)
- **Welsh's approximate t-test**
  - Not as powerful as t-test so it should only be used when necessary
- but if $n_1 \approx n_2$ then even ***a 3 fold*** *difference in standard deviations is* ***approximately*** *okay*

A



B



What graphs in A and B match:

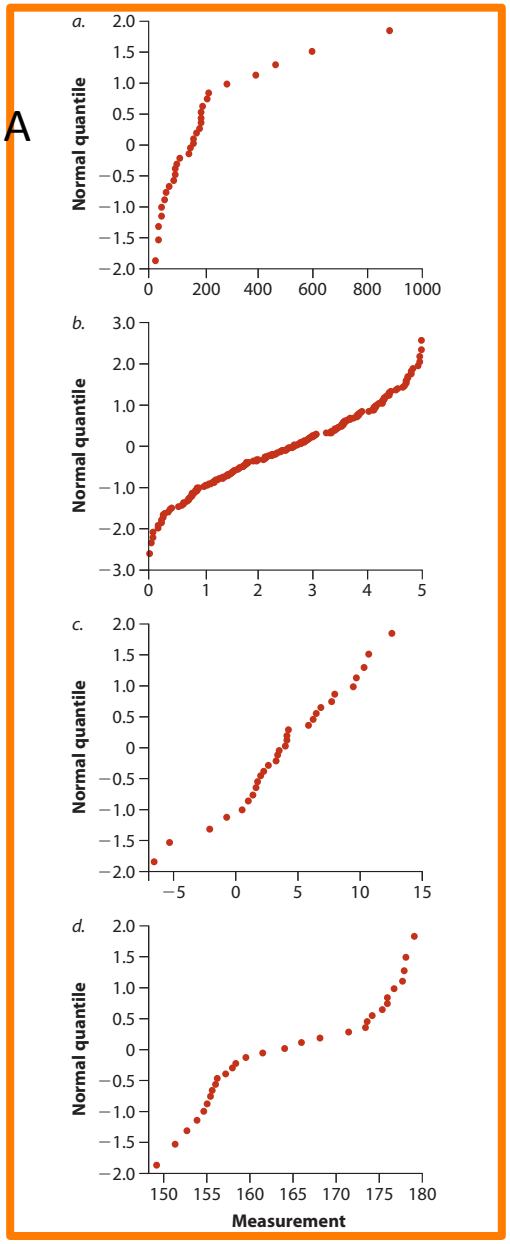
A. i:c, ii:d, iii: a, iv:b

B. i:b, ii:d, iii: a, iv:c

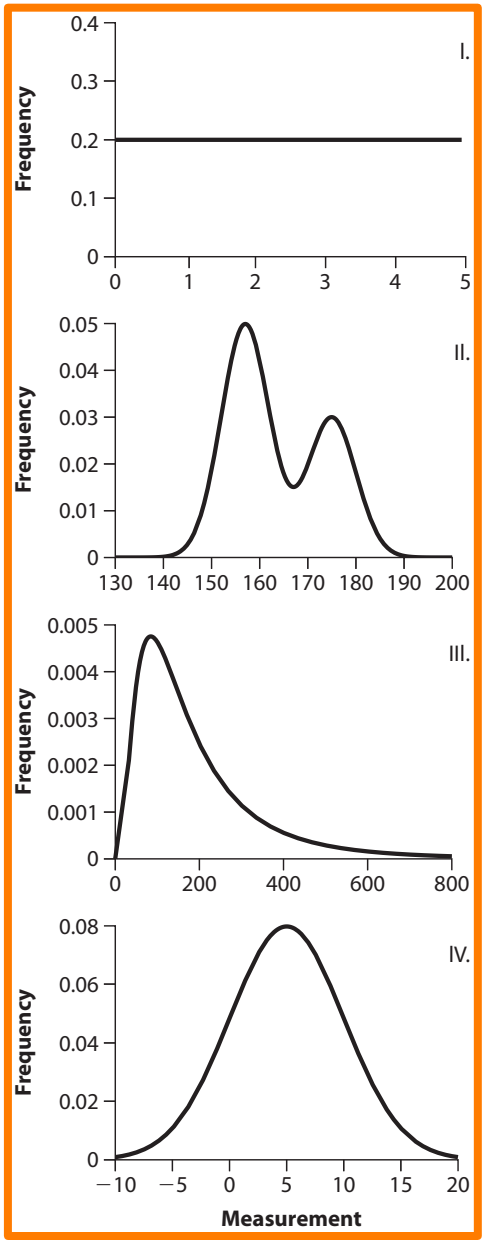
C. i:a, ii:d, iii: b, iv:c

D. i:c, ii:a, iii: d , iv:b

A



B



What graphs in A and B match:

A. i:c, ii:d, iii: a, iv:b

B. i:b, ii:d, iii: a, iv:c

C. i:a, ii:d, iii: b, iv:c

D. i:c, ii:a, iii: d , iv:b

The answer is B

This right-skewed distribution corresponds to plot ‘A’ since most of the weight of the plot happens before 200.

The normal distribution (iv) corresponds to quantile plot ‘C’ since it is approx a straight line.

It’s always a good idea to practice with the logic of these types of plots!