Biostatistics
Week IX

Ege Ülgen, M.D.

2 December 2021

ACIBADEM
MEHMET ALİ AYDINLAR
ÜNİVERSİTESİ

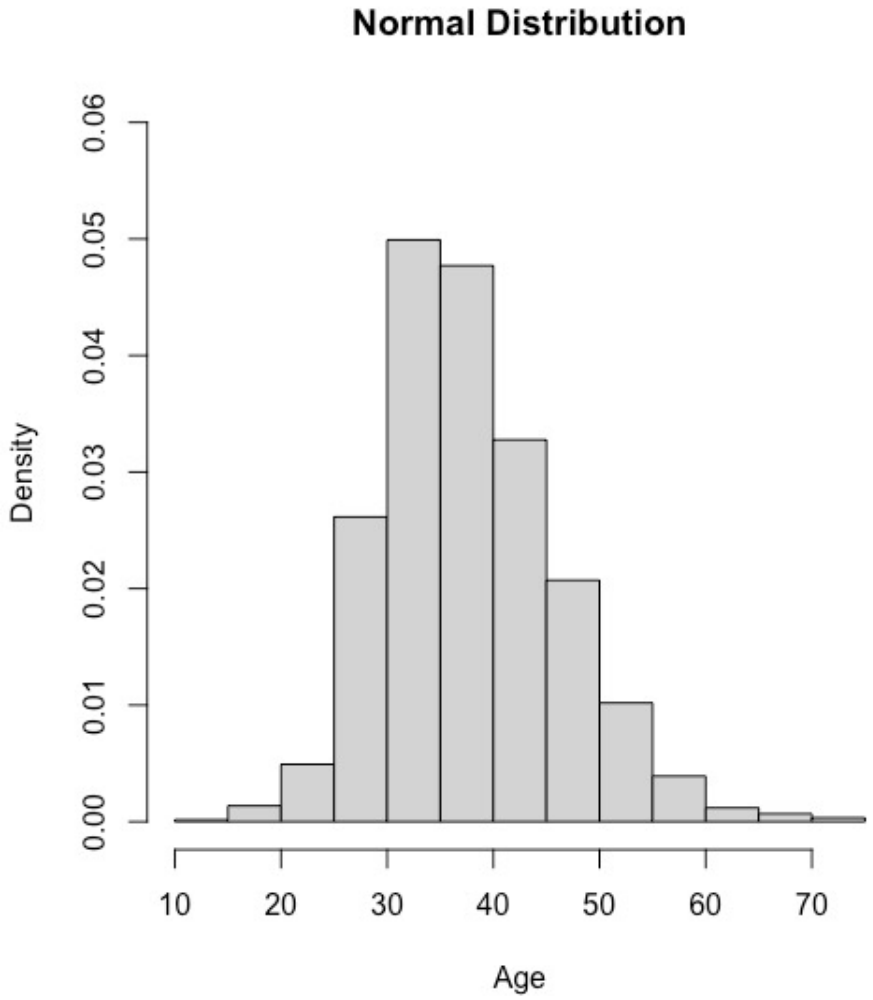
General Assumptions of Parametric Tests

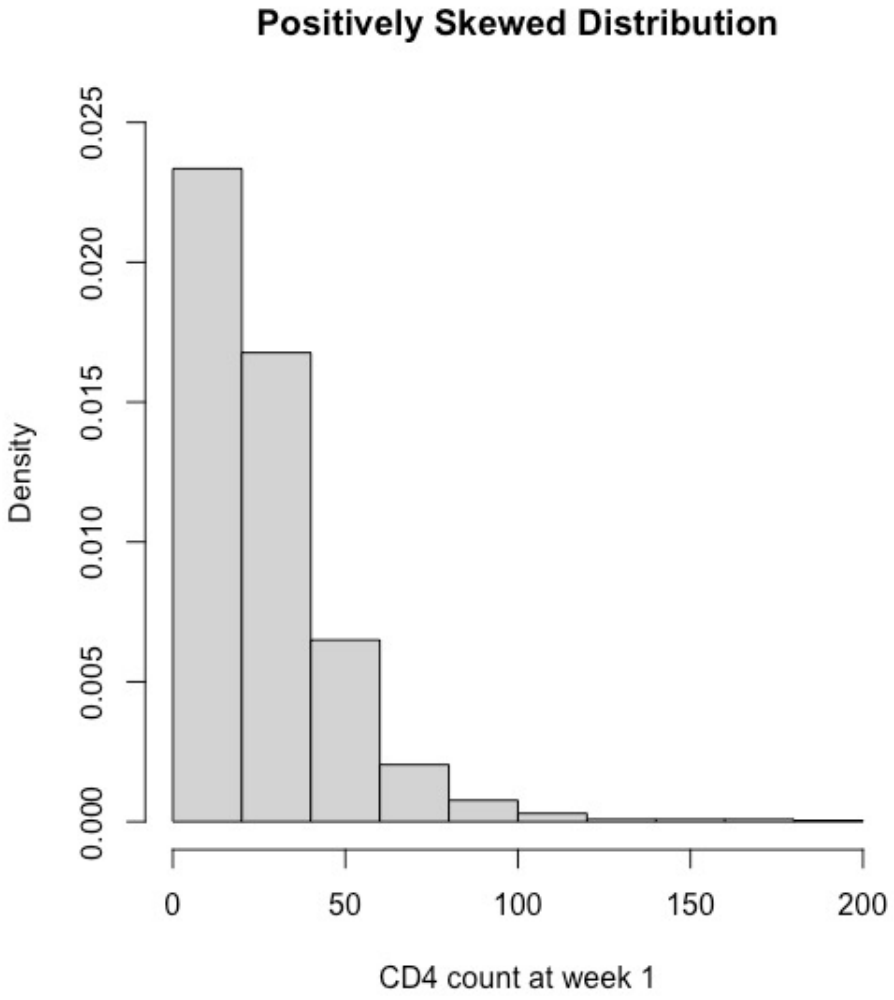
- The population(s) are **normally distributed**
- The selected sample is **representative of general population**

Assessing Normality

- Inspecting the **histogram** of the variable
- **Quantile-quantile plots**
- **Shapiro-Wilk test**
  - $p < 0.05$ indicates normal distribution
- …

Inspecting Histogram



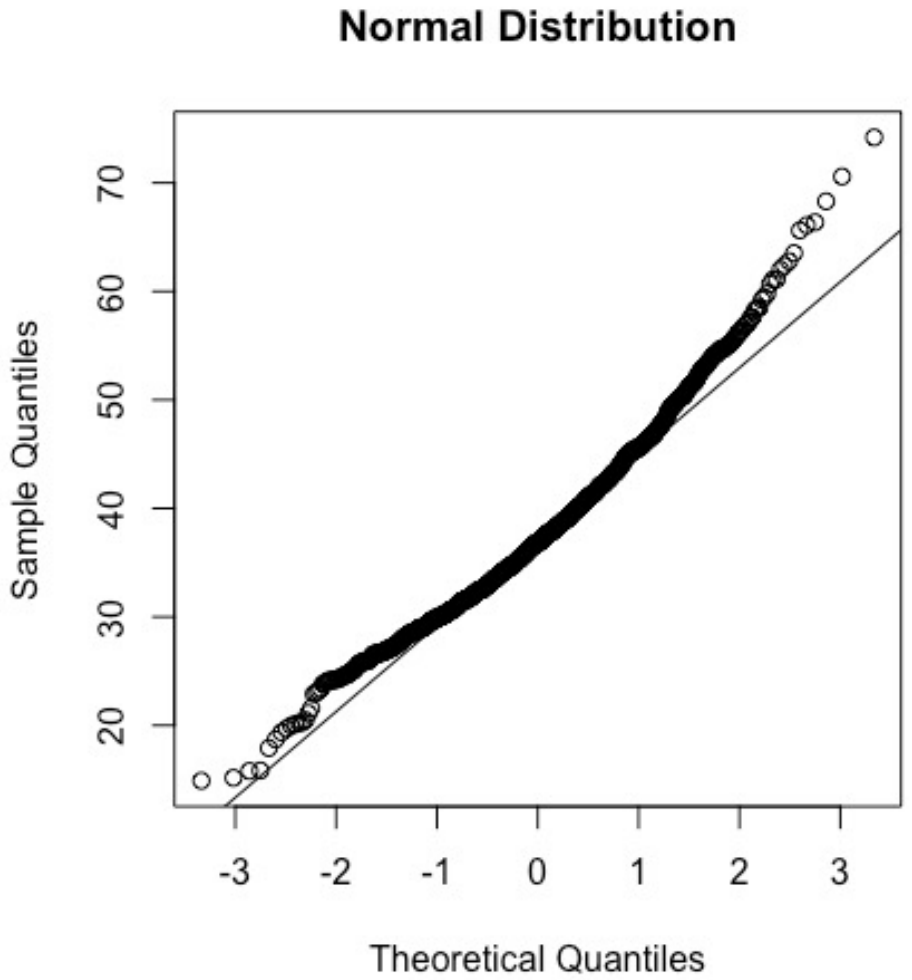


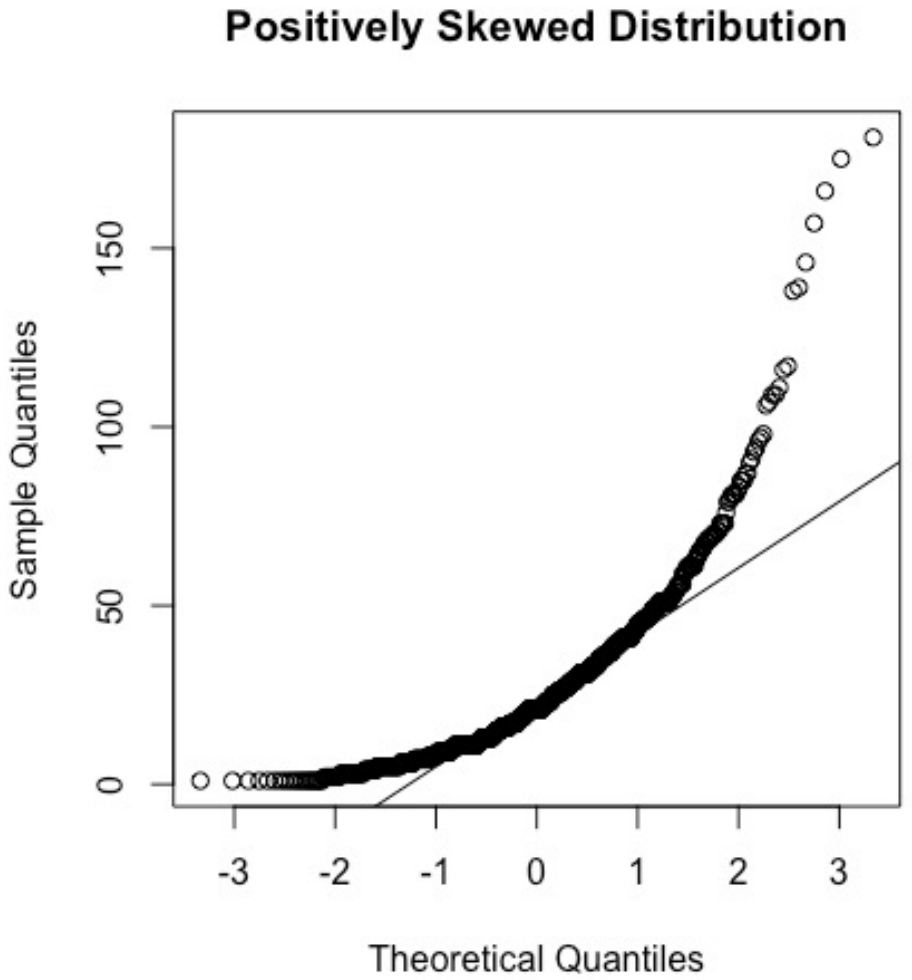
Quantile-Quantile Plots

- A tool for comparing the empirical distribution of data to the theoretical distribution

$$\Phi\left(\frac{i-0.5}{n}\right)$$ vs. sorted data (where *i* is the rank)

Quantile-Quantile Plots





Shapiro–Wilk Test of Normality

- A confirmatory tool for checking the normal distribution assumption

- **H_0: the population is normally distributed**
- H_1: the population is not normally distributed

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

$x_{(i)}$: the ith order statistic, i.e., the ith-smallest number in the sample
a_i: see reference

https://en.wikipedia.org/wiki/Shapiro%E2%80%93Wilk_test

Non-parametric Tests

- Often used when assumptions of parametric tests are not met
- **Robust with respect to the distribution of data**
- **Less assumptions**
  - e.g., they do not depend on the assumption of normality

- **Less statistical power** compared to parametric tests
  - Higher risk of type II errors (e.g., high probability of accepting there is no difference between the groups where there is a difference)

Non-parametric Tests

- **χ^2 test**
- **Wilcoxon rank-sum test (Mann–Whitney U test)** ~ Independent samples t-test
- **Kruskal-Wallis test** ~ one-way ANOVA
- **Mood’s Median Test** ~ one-way ANOVA
- **Friedman test** ~ two-way ANOVA
- **Spearman's rank correlation test** ~ Pearson correlation test
- …

Multiple Testing

H_0	Decision	
	Fail to reject	Reject
True	Correct decision	**Type I Error** α
False	**Type II Error** **ß**	Correct decision

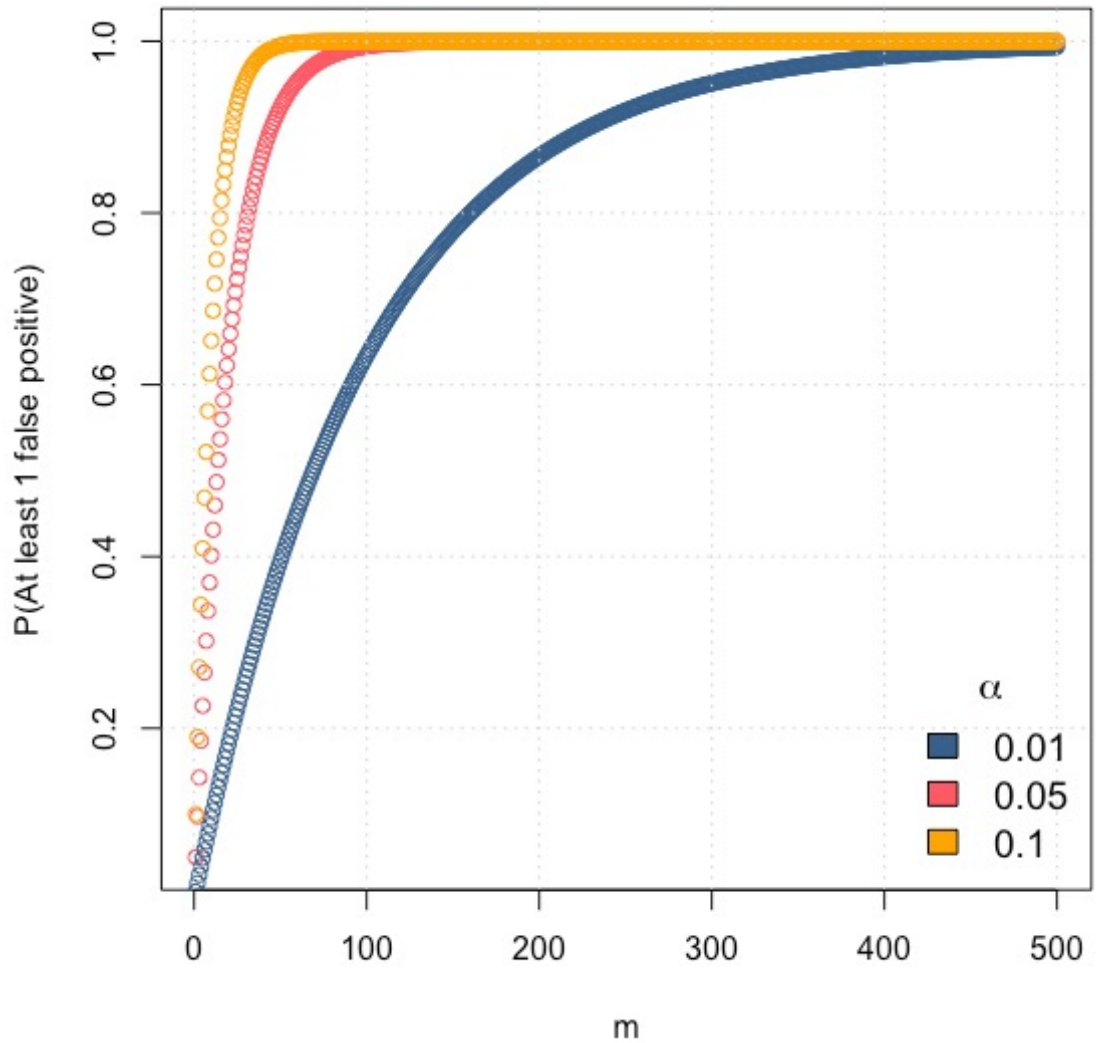
Multiple Testing - Example

- A typical microarray experiment might result in performing 10000 separate hypothesis tests
- If we use a standard p-value cut-off of 0.05, we'd expect ***500*** genes to be deemed "significant" by chance

Multiple Testing

- P(making a type I error) = α
- P(not making a type I error) = $1 - \alpha$
- P(not making a type I error in m tests) = $(1 - \alpha)^m$
- P(making at least 1 type I error in m tests) = $1 - (1 - \alpha)^m$

Multiple Testing



Correcting for Multiple Testing

- Controlling the Type I error rate
  - V = number of false positives out of all tests

Approaches to Control Type I Error Rate (V)

- Per comparison error rate (PCER)
- Per-family error rate (PFER)
- **Family-wise error rate (FWER)**
- **False discovery rate (FDR)**
- Positive false discovery rate (pFDR)

Family-wise Error Rate (FWER) Methods

- Bonferroni correction (single-step adjustment)
  - Rejects any hypothesis with p-value ≤ α/m

$$\widetilde{p_j} = \min\left(p_j \times m, 1\right)$$

  - If we want to have an experiment wide Type I error rate of 0.05 when we perform 10,000 hypothesis tests, we'd need a p-value of 0.05/10000 = 5 x 10^{-6} to declare significance

Family-wise Error Rate (FWER) Methods

- Holm's method (Sequential adjustments)

$$\widetilde{p_j} = \min[1, p_j \times (m - j + 1)]$$

e.g., m = 1000

$$\tilde{p}_1 = 1000 p_1,$$
$$\tilde{p}_2 = 999 p_2,$$
$$\dots,$$
$$\tilde{p}_m = 1 p_m$$

Family-wise Error Rate (FWER) Methods

- FWER is appropriate when you want to guard against **ANY** false positives

False Discovery Rate (FDR)

- Benjamini & Hochberg
  - To control FDR at level δ:
  1. Order the unadjusted p values in ascending order: $p_1 < \ldots < p_m$
  2. Find the test with the highest rank j for which:

$$p_j \leq \frac{j}{m}\delta$$

  3. Declare the tests of rank 1, …, j as significant

B&H FDR – Example

Controlling the FDR at $\delta = 0.05$

Rank (j)	P-value	(j/m)× δ	Reject H_0 ?
1	0.0008	0.005	1
2	0.009	0.010	1
3	0.165	0.015	0
4	0.205	0.020	0
5	0.396	0.025	0
6	0.450	0.030	0
7	0.641	0.035	0
8	0.781	0.040	0
9	0.900	0.045	0
10	0.993	0.050	0

Additional Reading

Noble WS. How does multiple testing correction work? Nat Biotechnol. 2009 Dec;27(12):1135–7: https://www.nature.com/articles/nbt1209-1135

Brief Summary

- Normality of a variable can be assessed using
  - Histogram
  - Q-Q plot
  - Shapiro-Wilk test
- Non-parametric tests have **fewer assumptions** but also have **less statistical power** compared to parametric tests
- Commonly used methods for multiple testing correction include:
  - Bonferroni correction
  - Holm’s method
  - Benjamini and Hochberg’s FDR