

Plots saved to plots/all_distributions.png

=====

ANALYSIS OF NUMERICAL FEATURES DISTRIBUTIONS

=====

=====

Feature: mean_radius

=====

Basic statistics:

Mean: 14.1273
Median: 13.3700
Standard deviation: 3.5210
Q1: 11.7000
Q3: 15.7800
IQR: 4.0800

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 0.9399)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 36.93%)
3. Similarity to normal distribution: moderately similar to normal
(Kurtosis coefficient: 0.8276)

Shapiro-Wilk test:

Statistic: 0.941069
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: mean_texture

=====

Basic statistics:

Mean: 19.2896
Median: 18.8400
Standard deviation: 4.2973
Q1: 16.1700
Q3: 21.8000
IQR: 5.6300

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 0.6487)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 40.04%)
3. Similarity to normal distribution: moderately similar to normal
(Kurtosis coefficient: 0.7411)

Shapiro-Wilk test:

Statistic: 0.976721
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====
Feature: mean_perimeter
=====

Basic statistics:

Mean: 91.9690
Median: 86.2400
Standard deviation: 24.2776
Q1: 75.1700
Q3: 104.1000
IQR: 28.9300

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 0.9880)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 38.41%)
3. Similarity to normal distribution: moderately similar to normal
(Kurtosis coefficient: 0.9532)

Shapiro-Wilk test:

Statistic: 0.936183
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====
Feature: mean_area
=====

Basic statistics:

Mean: 654.8891
Median: 551.1000
Standard deviation: 351.6048
Q1: 420.3000
Q3: 782.7000
IQR: 362.4000

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.6414)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 35.05%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 3.6098)

Shapiro-Wilk test:

Statistic: 0.858401
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====
Feature: mean_smoothness
=====

Basic statistics:

Mean: 0.0964
Median: 0.0959
Standard deviation: 0.0141

Q1: 0.0864
Q3: 0.1053
IQR: 0.0189

Distribution shape analysis:

1. Tail size and shape: symmetric
(Skewness coefficient: 0.4551)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 43.28%)
3. Similarity to normal distribution: moderately similar to normal
(Kurtosis coefficient: 0.8379)

Shapiro-Wilk test:

Statistic: 0.987487
p-value: 0.000086
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====
Feature: mean_compactness
=====

Basic statistics:

Mean: 0.1043
Median: 0.0926
Standard deviation: 0.0528
Q1: 0.0649
Q3: 0.1304
IQR: 0.0655

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.1870)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 38.97%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 1.6251)

Shapiro-Wilk test:

Statistic: 0.916978
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====
Feature: mean_concavity
=====

Basic statistics:

Mean: 0.0888
Median: 0.0615
Standard deviation: 0.0796
Q1: 0.0296
Q3: 0.1307
IQR: 0.1011

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.3975)

2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 42.49%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 1.9706)

Shapiro-Wilk test:

Statistic: 0.866831

p-value: 0.000000

Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====
Feature: mean_concave_points
=====

Basic statistics:

Mean: 0.0489

Median: 0.0335

Standard deviation: 0.0388

Q1: 0.0203

Q3: 0.0740

IQR: 0.0537

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.1681)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 44.70%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 1.0467)

Shapiro-Wilk test:

Statistic: 0.891650

p-value: 0.000000

Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====
Feature: mean_symmetry
=====

Basic statistics:

Mean: 0.1812

Median: 0.1792

Standard deviation: 0.0274

Q1: 0.1619

Q3: 0.1957

IQR: 0.0338

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 0.7237)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 37.88%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 1.2661)

Shapiro-Wilk test:

Statistic: 0.972589

p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: mean_fractal_dimension

=====

Basic statistics:

Mean: 0.0628
Median: 0.0615
Standard deviation: 0.0071
Q1: 0.0577
Q3: 0.0661
IQR: 0.0084

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.3010)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 37.99%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 2.9690)

Shapiro-Wilk test:

Statistic: 0.923284
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_radius

=====

Basic statistics:

Mean: 0.4052
Median: 0.3242
Standard deviation: 0.2771
Q1: 0.2324
Q3: 0.4789
IQR: 0.2465

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 3.0805)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 30.83%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 17.5212)

Shapiro-Wilk test:

Statistic: 0.745554
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_texture

=====

Basic statistics:

Mean: 1.2169
Median: 1.1080
Standard deviation: 0.5512
Q1: 0.8339
Q3: 1.4740
IQR: 0.6401

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.6421)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 38.29%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 5.2918)

Shapiro-Wilk test:

Statistic: 0.896168
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_perimeter

=====

Basic statistics:

Mean: 2.8661
Median: 2.2870
Standard deviation: 2.0201
Q1: 1.6060
Q3: 3.3570
IQR: 1.7510

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 3.4345)
2. IQR size relative to the rest of the distribution: small (less than 30% of interquartile range)
(IQR relative to 90% range: 29.63%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 21.2038)

Shapiro-Wilk test:

Statistic: 0.718164
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_area

=====

Basic statistics:

Mean: 40.3371
Median: 24.5300
Standard deviation: 45.4510
Q1: 17.8500
Q3: 45.1900
IQR: 27.3400

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 5.4328)
2. IQR size relative to the rest of the distribution: small (less than 30% of interquartile range)
(IQR relative to 90% range: 26.18%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 48.7672)

Shapiro-Wilk test:

Statistic: 0.563820

p-value: 0.000000

Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_smoothness

=====

Basic statistics:

Mean: 0.0070

Median: 0.0064

Standard deviation: 0.0030

Q1: 0.0052

Q3: 0.0081

IQR: 0.0030

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 2.3083)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 33.25%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 10.3675)

Shapiro-Wilk test:

Statistic: 0.838410

p-value: 0.000000

Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_compactness

=====

Basic statistics:

Mean: 0.0255

Median: 0.0204

Standard deviation: 0.0179

Q1: 0.0131

Q3: 0.0324

IQR: 0.0194

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.8972)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 36.77%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 5.0510)

Shapiro-Wilk test:
Statistic: 0.836878
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_concavity

=====

Basic statistics:
Mean: 0.0319
Median: 0.0259
Standard deviation: 0.0302
Q1: 0.0151
Q3: 0.0420
IQR: 0.0270

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 5.0970)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 35.62%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 48.4226)

Shapiro-Wilk test:
Statistic: 0.672070
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_concave_points

=====

Basic statistics:
Mean: 0.0118
Median: 0.0109
Standard deviation: 0.0062
Q1: 0.0076
Q3: 0.0147
IQR: 0.0071

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.4409)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 37.12%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 5.0708)

Shapiro-Wilk test:
Statistic: 0.919706
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_symmetry

=====

Basic statistics:

Mean: 0.0205
Median: 0.0187
Standard deviation: 0.0083
Q1: 0.0152
Q3: 0.0235
IQR: 0.0083

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 2.1893)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 35.82%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 7.8164)

Shapiro-Wilk test:

Statistic: 0.828386
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: se_fractal_dimension

=====

Basic statistics:

Mean: 0.0038
Median: 0.0032
Standard deviation: 0.0026
Q1: 0.0022
Q3: 0.0046
IQR: 0.0023

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 3.9136)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 35.88%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 26.0399)

Shapiro-Wilk test:

Statistic: 0.695079
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_radius

=====

Basic statistics:

Mean: 16.2692
Median: 14.9700
Standard deviation: 4.8290
Q1: 13.0100

Q3: 18.7900
IQR: 5.7800

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.1002)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 38.26%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 0.9253)

Shapiro-Wilk test:

Statistic: 0.913492
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_texture

=====

Basic statistics:

Mean: 25.6772
Median: 25.4100
Standard deviation: 6.1409
Q1: 21.0800
Q3: 29.7200
IQR: 8.6400

Distribution shape analysis:

1. Tail size and shape: symmetric
(Skewness coefficient: 0.4970)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 43.80%)
3. Similarity to normal distribution: close to normal
(Kurtosis coefficient: 0.2118)

Shapiro-Wilk test:

Statistic: 0.982562
p-value: 0.000003
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_perimeter

=====

Basic statistics:

Mean: 107.2612
Median: 97.6600
Standard deviation: 33.5730
Q1: 84.1100
Q3: 125.4000
IQR: 41.2900

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.1252)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)

- (IQR relative to 90% range: 39.78%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 1.0502)

Shapiro-Wilk test:
Statistic: 0.912585
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_area

=====

Basic statistics:
Mean: 880.5831
Median: 686.5000
Standard deviation: 568.8565
Q1: 515.3000
Q3: 1084.0000
IQR: 568.7000

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.8545)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 33.88%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 4.3473)

Shapiro-Wilk test:
Statistic: 0.816070
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_smoothness

=====

Basic statistics:
Mean: 0.1324
Median: 0.1313
Standard deviation: 0.0228
Q1: 0.1166
Q3: 0.1460
IQR: 0.0294

Distribution shape analysis:

1. Tail size and shape: symmetric
(Skewness coefficient: 0.4143)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 38.63%)
3. Similarity to normal distribution: moderately similar to normal
(Kurtosis coefficient: 0.5028)

Shapiro-Wilk test:
Statistic: 0.988620
p-value: 0.000210

Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_compactness

=====

Basic statistics:

Mean: 0.2543
Median: 0.2119
Standard deviation: 0.1572
Q1: 0.1472
Q3: 0.3391
IQR: 0.1919

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.4697)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 38.93%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 3.0021)

Shapiro-Wilk test:

Statistic: 0.891064
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_concavity

=====

Basic statistics:

Mean: 0.2722
Median: 0.2267
Standard deviation: 0.2084
Q1: 0.1145
Q3: 0.3829
IQR: 0.2684

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.1472)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 40.42%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 1.5906)

Shapiro-Wilk test:

Statistic: 0.917528
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_concave_points

=====

Basic statistics:

Mean: 0.1146

Median: 0.0999
Standard deviation: 0.0657
Q1: 0.0649
Q3: 0.1614
IQR: 0.0965

Distribution shape analysis:

1. Tail size and shape: symmetric
(Skewness coefficient: 0.4913)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 45.37%)
3. Similarity to normal distribution: moderately similar to normal
(Kurtosis coefficient: -0.5414)

Shapiro-Wilk test:

Statistic: 0.964835
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_symmetry

=====

Basic statistics:

Mean: 0.2901
Median: 0.2822
Standard deviation: 0.0618
Q1: 0.2504
Q3: 0.3179
IQR: 0.0675

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.4301)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 34.89%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 4.3951)

Shapiro-Wilk test:

Statistic: 0.916144
p-value: 0.000000
Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

=====

Feature: worst_fractal_dimension

=====

Basic statistics:

Mean: 0.0839
Median: 0.0800
Standard deviation: 0.0180
Q1: 0.0715
Q3: 0.0921
IQR: 0.0206

Distribution shape analysis:

1. Tail size and shape: right tail longer (positive skewness)
(Skewness coefficient: 1.6582)
2. IQR size relative to the rest of the distribution: medium (30-50% of interquartile range)
(IQR relative to 90% range: 36.20%)
3. Similarity to normal distribution: differs from normal
(Kurtosis coefficient: 5.1881)

Shapiro-Wilk test:

Statistic: 0.889546

p-value: 0.000000

Conclusion: Distribution DIFFERS from normal ($p \leq 0.05$)

SUMMARY TABLE OF RESULTS

	Feature	Skewness	Kurtosis	IQR/Range	Shapiro-Wilk	p-value	Normal?
	mean_radius	0.940	0.828	36.93%		0.000000	No
	mean_texture	0.649	0.741	40.04%		0.000000	No
	mean_perimeter	0.988	0.953	38.41%		0.000000	No
	mean_area	1.641	3.610	35.05%		0.000000	No
	mean_smoothness	0.455	0.838	43.28%		0.000086	No
	mean_compactness	1.187	1.625	38.97%		0.000000	No
	mean_concavity	1.397	1.971	42.49%		0.000000	No
	mean_concave_points	1.168	1.047	44.70%		0.000000	No
	mean_symmetry	0.724	1.266	37.88%		0.000000	No
mean_fractal_dimension		1.301	2.969	37.99%		0.000000	No
	se_radius	3.080	17.521	30.83%		0.000000	No
	se_texture	1.642	5.292	38.29%		0.000000	No
	se_perimeter	3.435	21.204	29.63%		0.000000	No
	se_area	5.433	48.767	26.18%		0.000000	No
	se_smoothness	2.308	10.368	33.25%		0.000000	No
	se_compactness	1.897	5.051	36.77%		0.000000	No
	se_concavity	5.097	48.423	35.62%		0.000000	No
	se_concave_points	1.441	5.071	37.12%		0.000000	No
	se_symmetry	2.189	7.816	35.82%		0.000000	No
	se_fractal_dimension	3.914	26.040	35.88%		0.000000	No
	worst_radius	1.100	0.925	38.26%		0.000000	No
	worst_texture	0.497	0.212	43.80%		0.000003	No
	worst_perimeter	1.125	1.050	39.78%		0.000000	No
	worst_area	1.854	4.347	33.88%		0.000000	No
	worst_smoothness	0.414	0.503	38.63%		0.000210	No
worst_compactness	1.470	3.002	38.93%		0.000000	No	
worst_concavity	1.147	1.591	40.42%		0.000000	No	
worst_concave_points	0.491	-0.541	45.37%		0.000000	No	
worst_symmetry	1.430	4.395	34.89%		0.000000	No	
worst_fractal_dimension	1.658	5.188	36.20%		0.000000	No	

Summary table saved to plots/distribution_summary.csv

ANALYSIS COMPLETED