DISCOVERING STATISTICS USING IBM SPSS STATISTICS

4TH EDITION

The Beast of Bias

Lecture 04

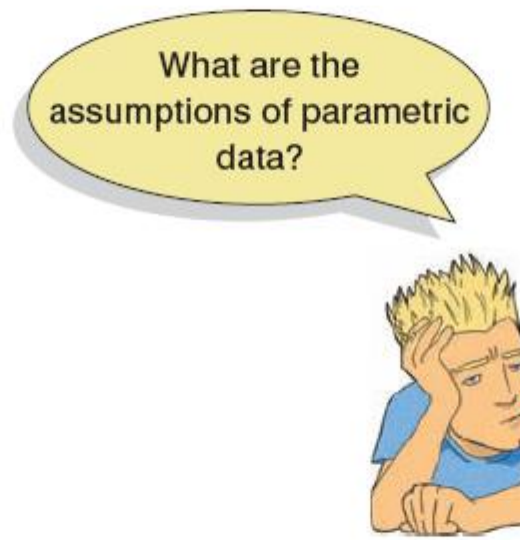
ANDY FIELD

Aims

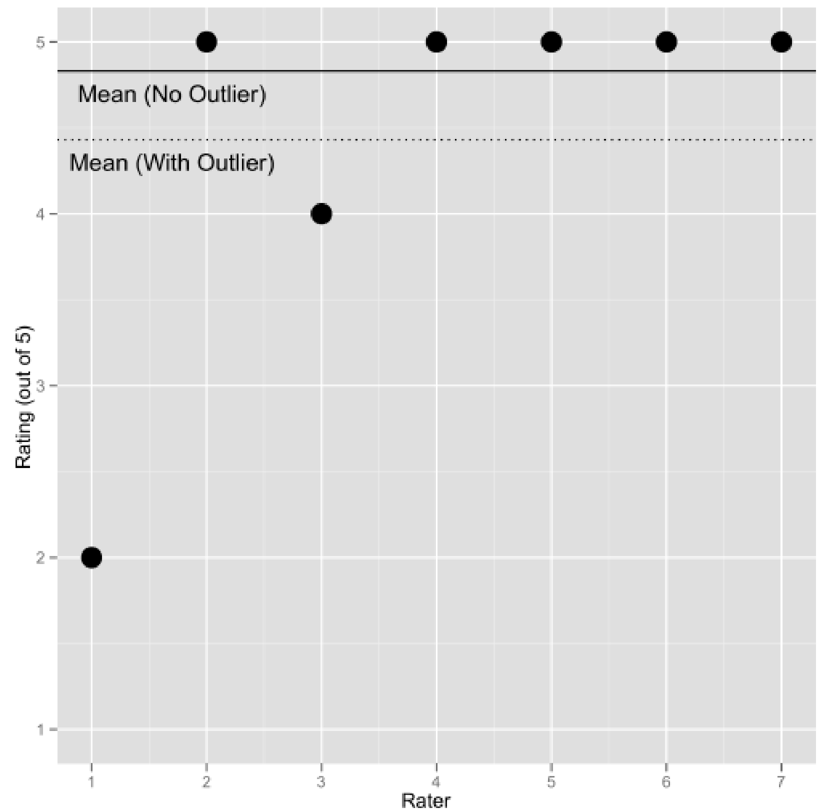
- Assumptions of parametric tests based on the normal distribution
- Understand the assumption of normality
  - Graphical displays
  - Skew
  - Kurtosis
  - Normality tests
- Understand Homogeneity of Variance
  - Levene's Test
- Know how to correct problems in the data
  - Log, Square Root and Reciprocal Transformations
  - Pitfalls and alternatives
  - Robust tests

Assumptions

- Parametric tests based on the normal distribution assume:
  - Additivity and linearity
  - Normality something or other
  - Homogeneity of Variance
  - Independence



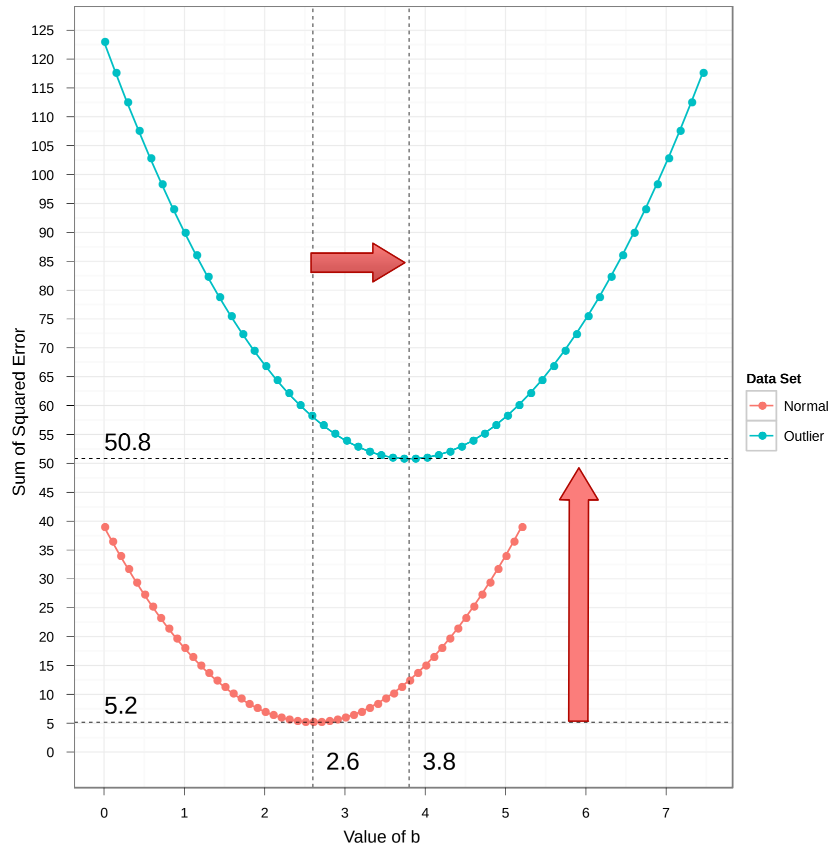
Outliers can Bias a Parameter Estimate



!

Figure' 5.2:' The!first!7!customer!ratings!of!this!book!on!www.amazon.co.uk!(in!about!2002).!The!first!score! biases!the!mean!

...and the Error associated with that Estimate



!

Figure+5.3:+The!effect!of!an!outlier!on!a!parameter!estimate!(the!mean)!and!it's!associated!estimate!of!error!(the!sum!of!squared!errors)!

ANDY FIELD

Additivity and Linearity

- The outcome variable is, in reality, linearly related to any predictors.
- If you have several predictors then their combined effect is best described by adding their effects together.
- If this assumption is not met then your model is invalid.

Normally Distributed Something or Other

- The normal distribution is relevant to:
  - Parameters
  - Confidence intervals around a parameter
  - Null hypothesis significance testing
- This assumption tends to get incorrectly translated as ‘your data need to be normally distributed’.

When does the Assumption of Normality Matter?

- In small samples.
  - The central limit theorem allows us to forget about this assumption in larger samples.
- In practical terms, as long as your sample is fairly large, outliers are a much more pressing concern than normality.

Homoscedasticity/ Homogeneity of Variance

- When testing several groups of participants, samples should come from populations with the same variance.
- In correlational designs, the variance of the outcome variable should be stable at all levels of the predictor variable.

Homoscedasticity/ Homogeneity of Variance

- Can affect the two main things that we might do when we fit models to data:
  - Parameters
  - Null Hypothesis significance testing

Assessing Homoscedasticity/ Homogeneity of Variance

Graphs (see lectures on regression)

- Levene's Tests
  - Tests if variances in different groups are the same.
  - Significant = Variances not equal
  - Non-Significant = Variances are equal
- Variance Ratio
  - With 2 or more groups
  - VR = Largest variance/Smallest variance
  - If VR < 2, homogeneity can be assumed.

Homogeneity of Variance

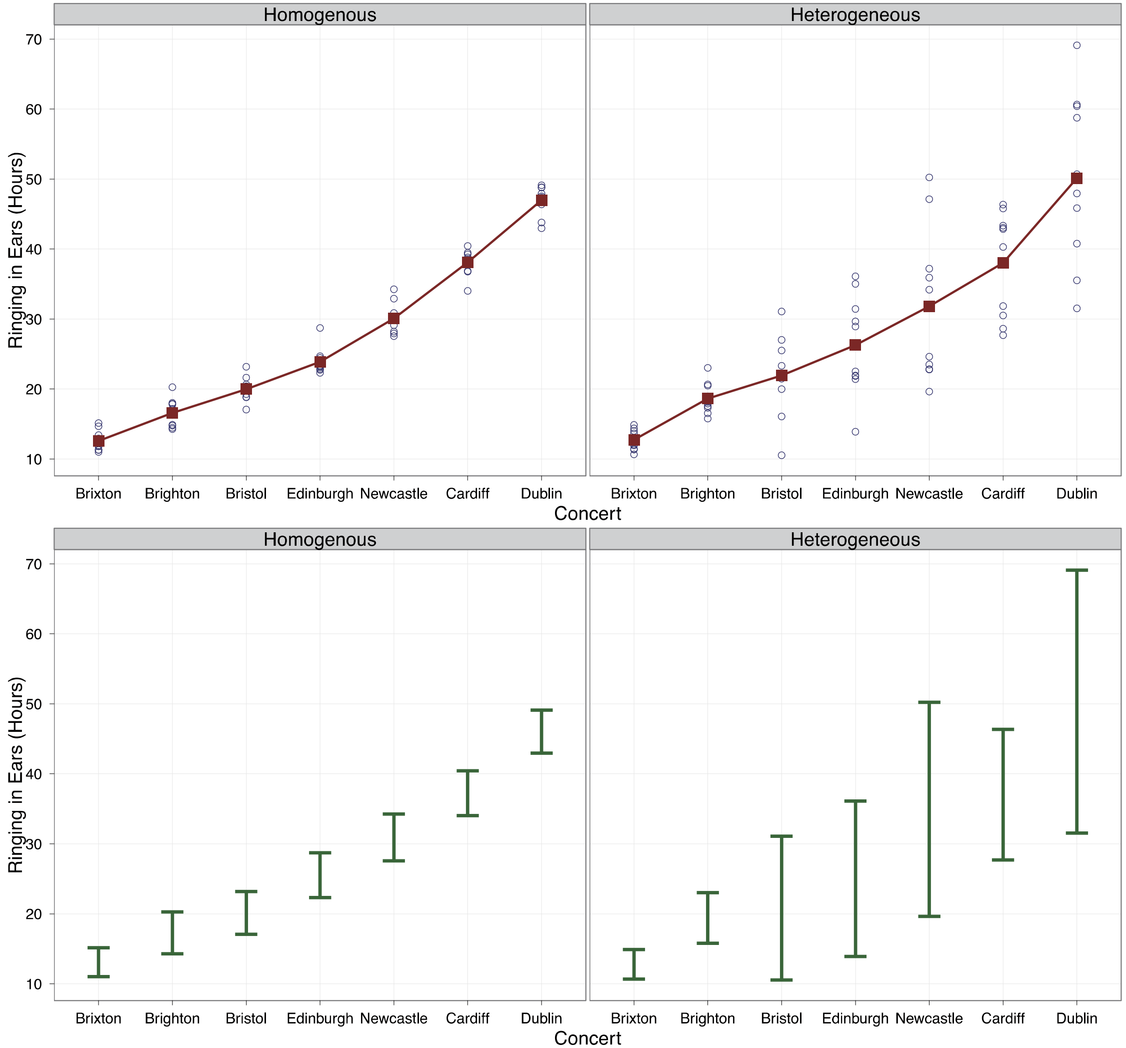


FIGURE 5.7 Graphs illustrating data with homogeneous (left) and heterogeneous (right) variances

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

4TH EDITION

ANDY FIELD

Independence

- The errors in your model should not be related to each other.
- If this assumption is violated:
  - Confidence intervals and significance tests will be invalid.
  - You should apply the techniques covered in Chapter 20.

Spotting Outliers Example

- A biologist was worried about the potential health effects of music festivals.
- Download Music Festival
- Measured the hygiene of 810 concert-goers over the three days of the festival.
- Hygiene was measured using a standardised technique :
  - Score ranged from 0 to 4
    - 0 = you smell like a corpse rotting up a skunk's arse
    - 4 = you smell of sweet roses on a fresh spring day

Spotting outliers With Graphs

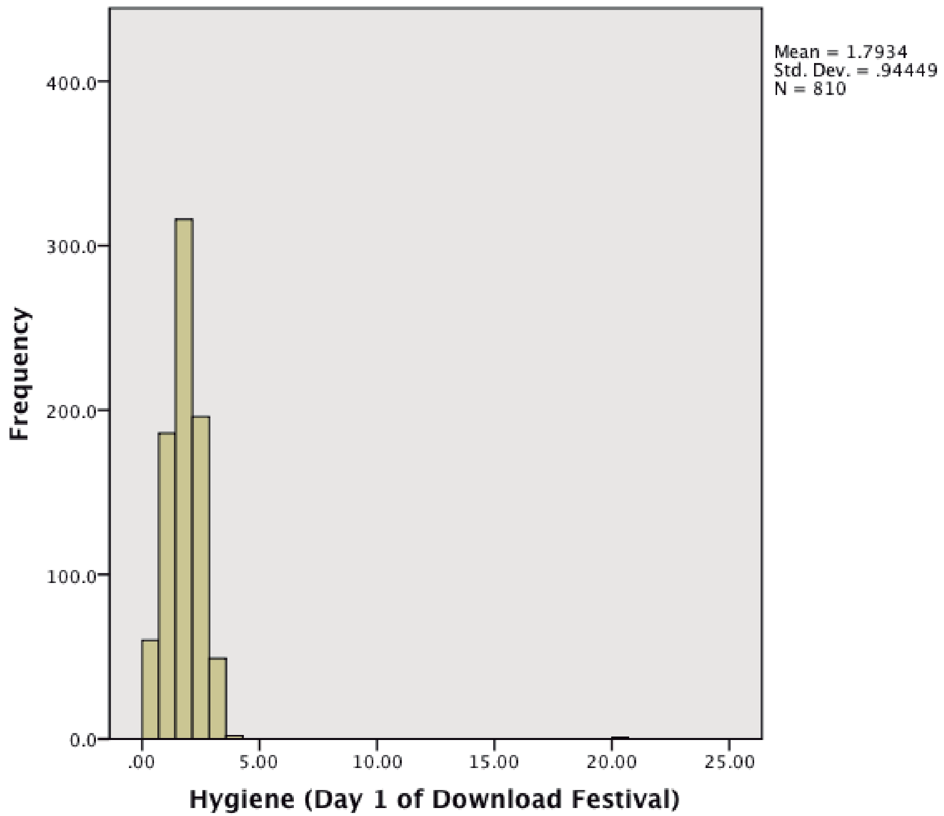


FIGURE 5.8 Histogram of the day 1 Download Festival hygiene scores

FIGURE 5.9
Boxplot of hygiene scores on day 1 of the Download Festival

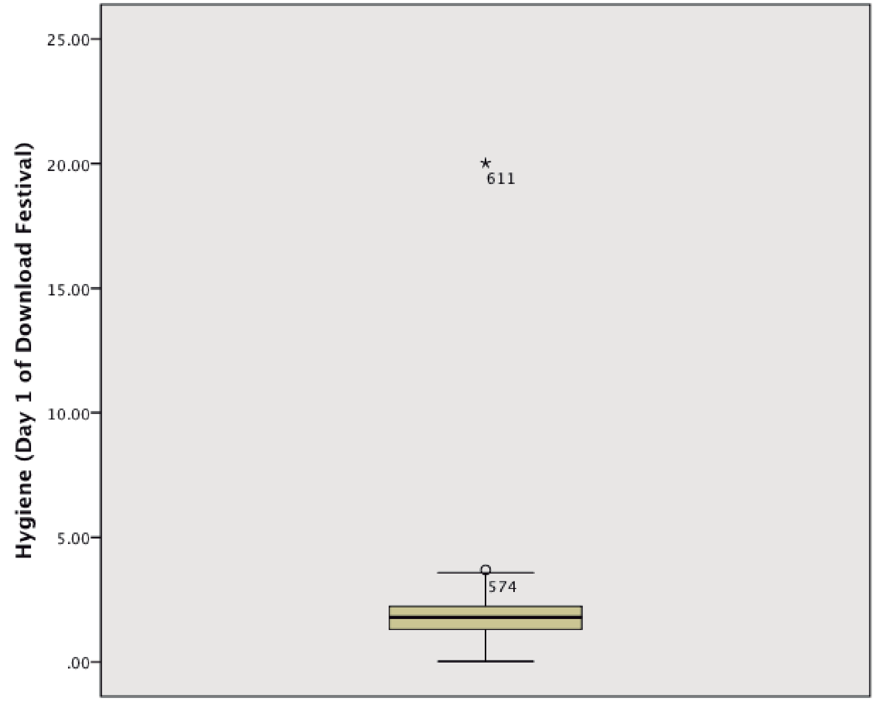
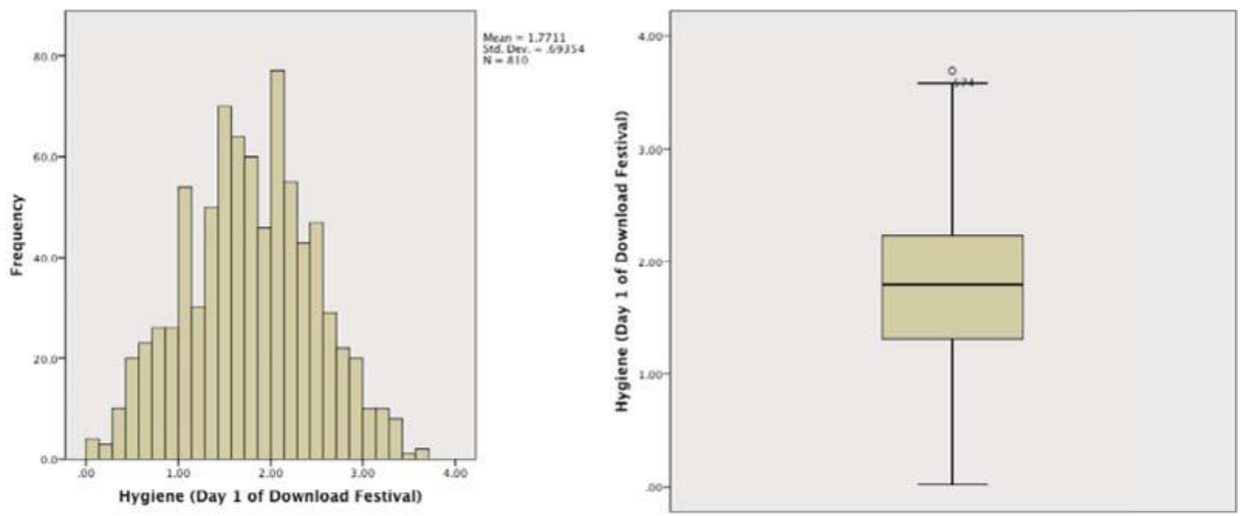


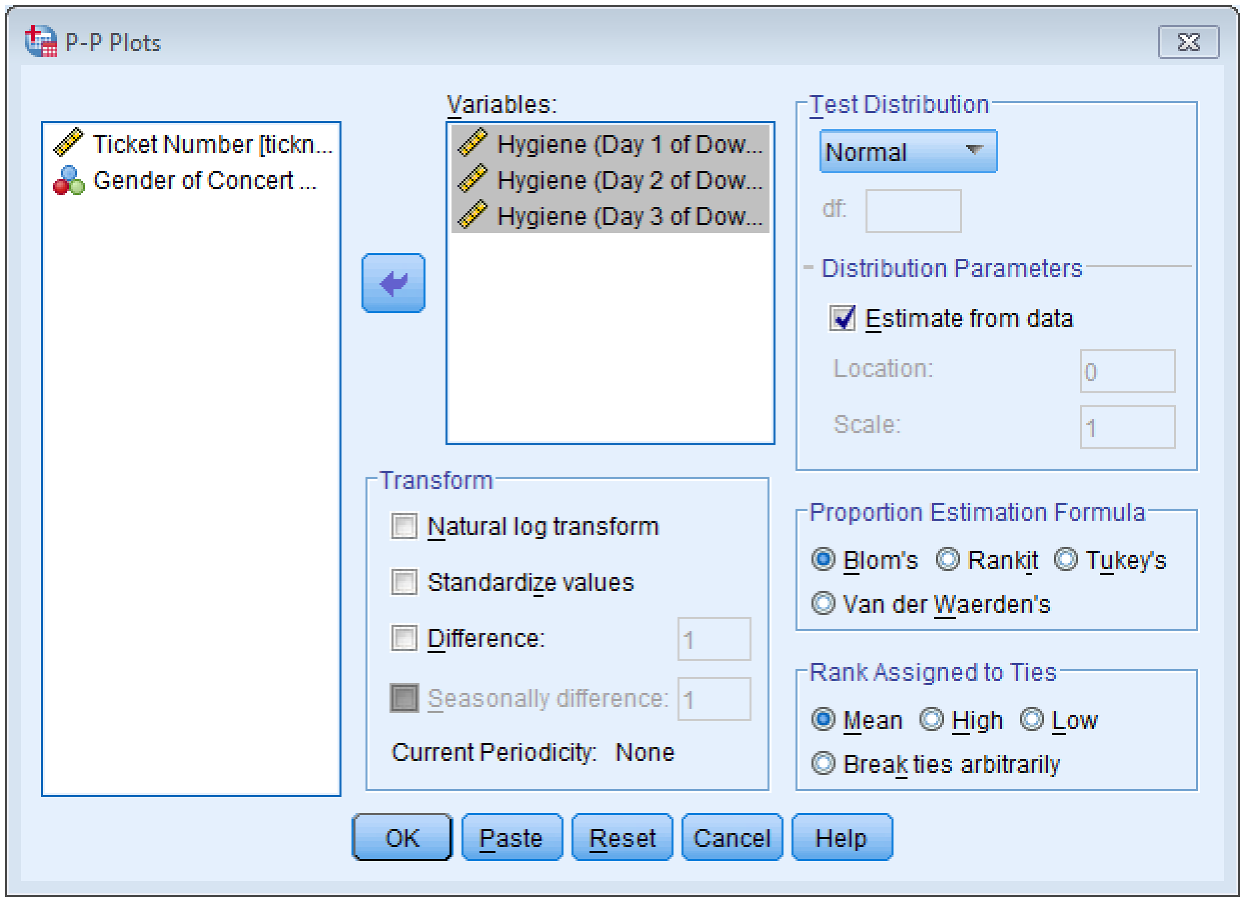
FIGURE 5.10
Histogram (left) and boxplot (right) of hygiene scores on day 1 of the Download Festival after removing the extreme score



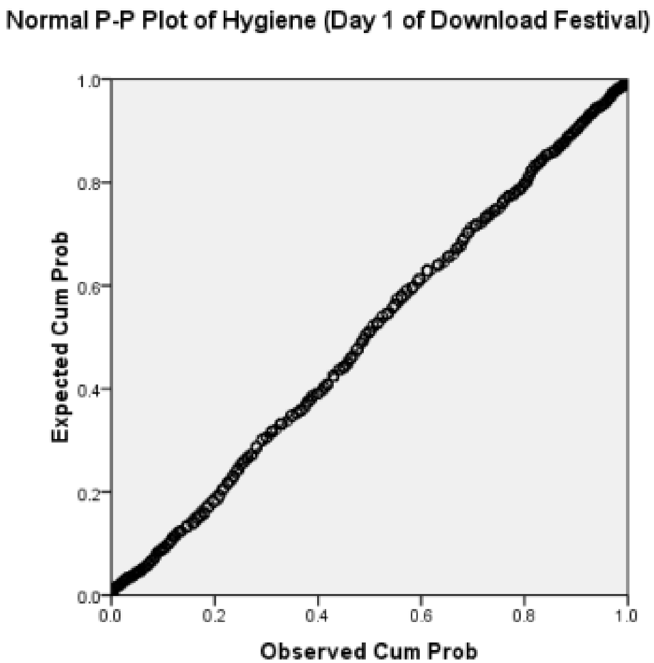
Spotting Normality

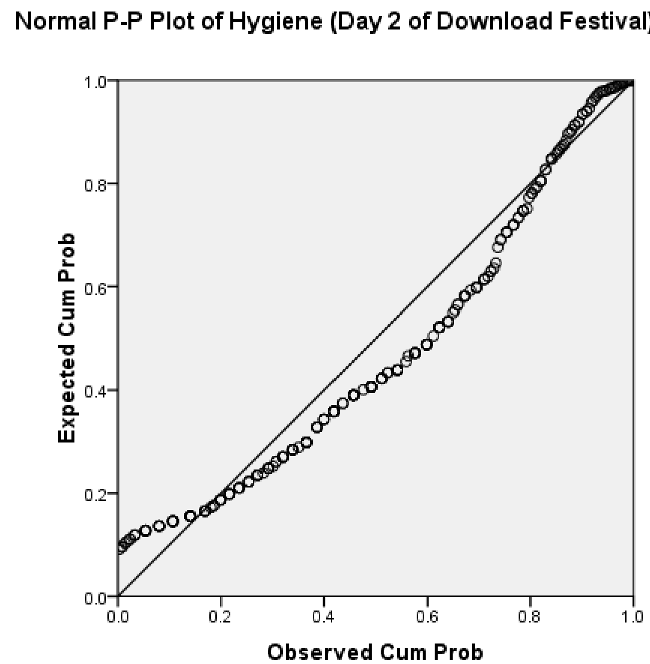
- We don't have access to the sampling distribution so we usually test the observed data
- Central Limit Theorem
  - If $N > 30$, the sampling distribution is normal anyway
- Graphical displays
  - P-P Plot (or Q-Q plot)
  - Histogram
- Values of Skew/Kurtosis
  - 0 in a normal distribution
  - Convert to z (by dividing value by *SE*)
- Kolmogorov-Smirnov Test
  - Tests if data differ from a normal distribution
  - Significant = non-Normal data
  - Non-Significant = Normal data

Spotting Normality: The P-P Plot



The P-P Plot





Normal **Not Normal**

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

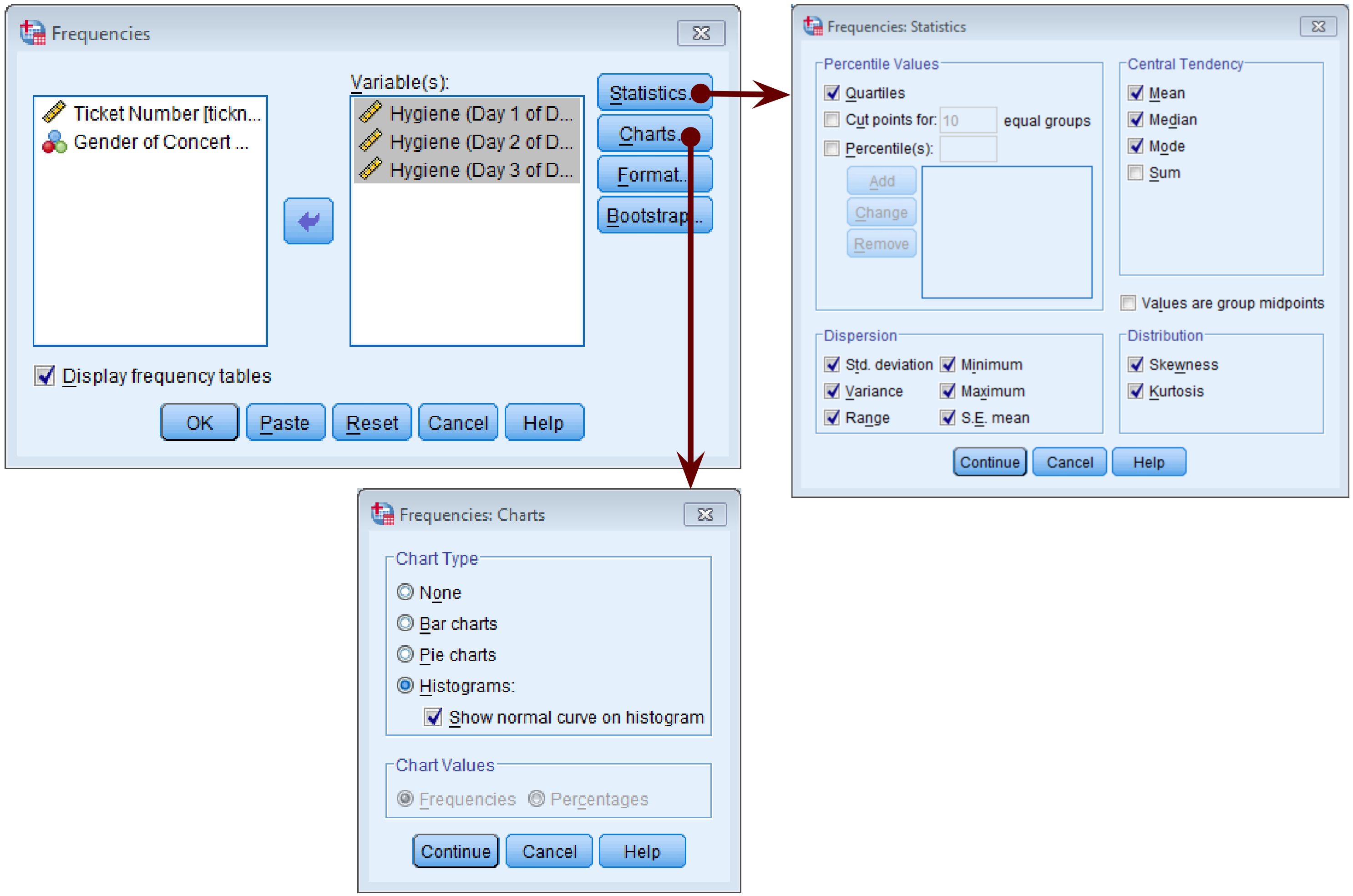
4TH EDITION

ANDY FIELD

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

4TH EDITION

Spotting Normality with Numbers: Skew and Kurtosis



ANDY FIELD

Assessing Skew and Kurtosis

Statistics

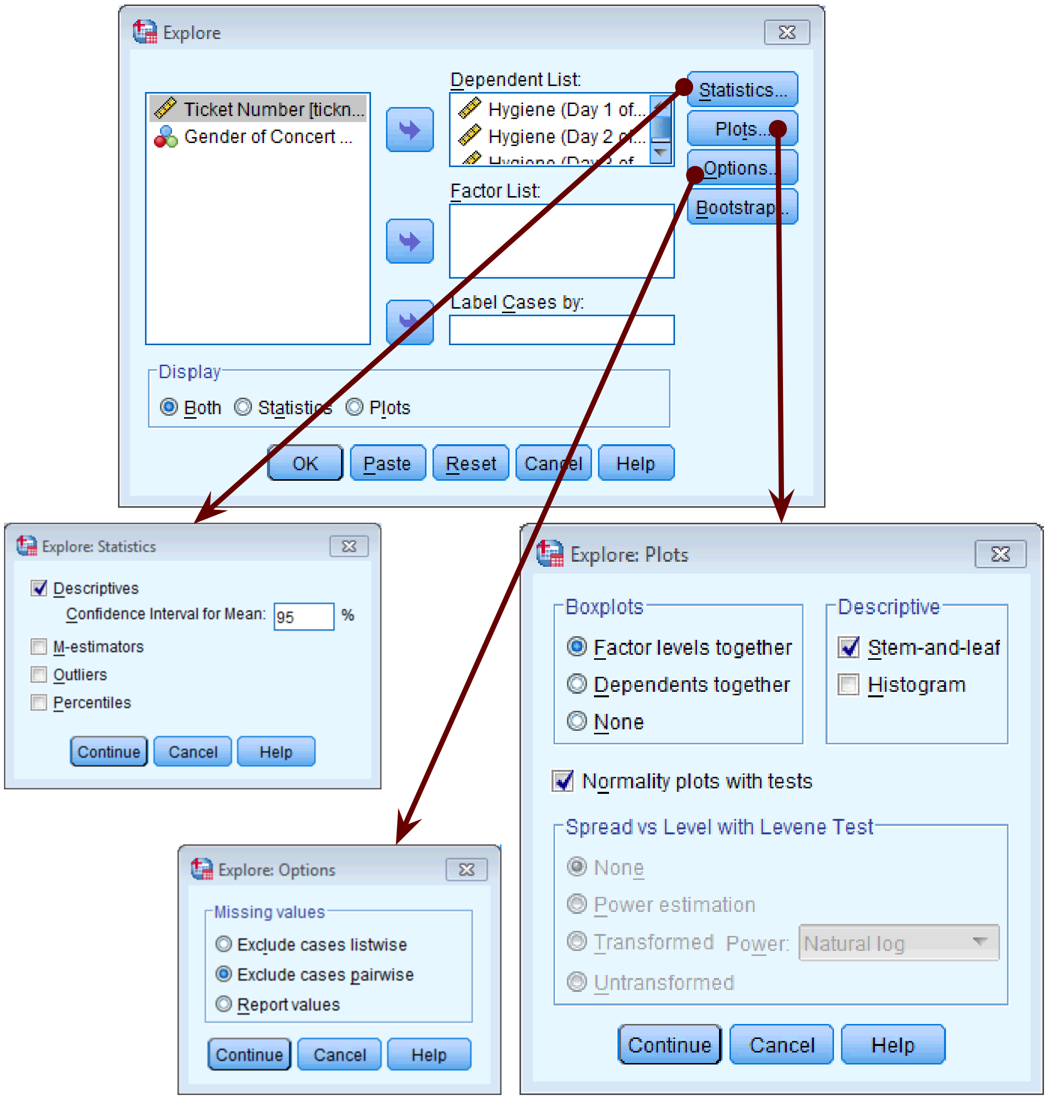
		Hygiene (Day 1 of Download Festival)	Hygiene (Day 2 of Download Festival)	Hygiene (Day 3 of Download Festival)
N	Valid	810	264	123
	Missing	0	546	687
Mean		1.7711	.9609	.9765
Std. Error of Mean		.02437	.04436	.06404
Median		1.7900	.7900	.7600
Mode		2.00	.23	.44[a]
Std. Deviation		.69354	.72078	.71028
Variance		.481	.520	.504
Skewness		-.004	1.095	1.033
Std. Error of Skewness		.086	.150	.218
Kurtosis		-.410	.822	.732
Std. Error of Kurtosis		.172	.299	.433
Range		3.67	3.44	3.39
Minimum		.02	.00	.02
Maximum		3.69	3.44	3.41
Percentiles	25	1.3050	.4100	.4400
	50	1.7900	.7900	.7600
	75	2.2300	1.3500	1.5500

a. Multiple modes exist. The smallest value is shown

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

4TH EDITION

Assessing Normality



ANDY FIELD

Tests of Normality

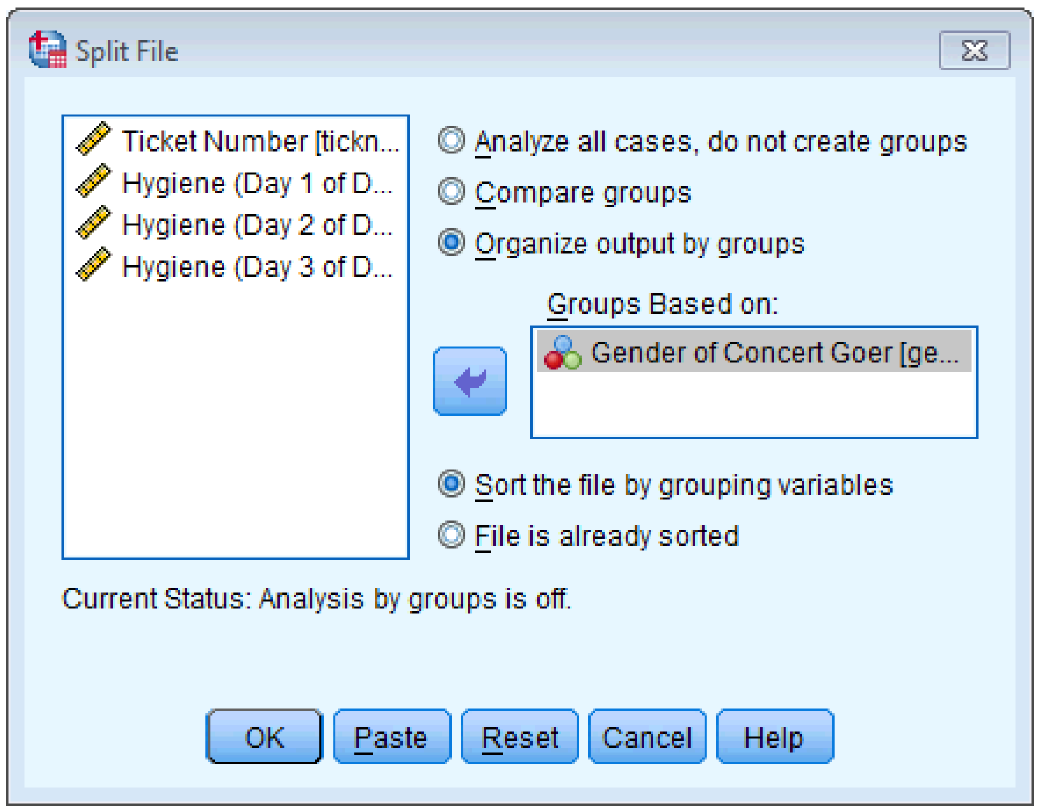
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Hygiene (Day 1 of Download Festival)	.029	810	.097	.996	810	.032
Hygiene (Day 2 of Download Festival)	.121	264	.000	.908	264	.000
Hygiene (Day 3 of Download Festival)	.140	123	.000	.908	123	.000

a. Lilliefors Significance Correction

Normality within Groups

- The Split File command



Normality Within Groups

Male

Statistics[a]

		Hygiene (Day 1 of Download Festival)	Hygiene (Day 2 of Download Festival)	Hygiene (Day 3 of Download Festival)
N	Valid	315	104	56
	Missing	0	211	259
Mean		1.6021	.7733	.8291
Std. Error of Mean		.03620	.05847	.07210
Median		1.5800	.6700	.7300
Mode		2.00	.23	.44
Std. Deviation		.64241	.59630	.53954
Variance		.413	.356	.291
Skewness		.200	1.476	.719
Std. Error of Skewness		.137	.237	.319
Kurtosis		-.101	3.134	-.268
Std. Error of Kurtosis		.274	.469	.628
Range		3.47	3.35	2.09
Minimum		.11	.00	.02
Maximum		3.58	3.35	2.11

a. Gender of Concert Goer = Male

Female

Statistics[a]

		Hygiene (Day 1 of Download Festival)	Hygiene (Day 2 of Download Festival)	Hygiene (Day 3 of Download Festival)
N	Valid	495	160	67
	Missing	0	335	428
Mean		1.8787	1.0829	1.0997
Std. Error of Mean		.03164	.06078	.09896
Median		1.9400	.8900	.8500
Mode		2.02	.85	.38
Std. Deviation		.70396	.76876	.81001
Variance		.496	.591	.656
Skewness		-.176	.870	.869
Std. Error of Skewness		.110	.192	.293
Kurtosis		-.397	.089	.069
Std. Error of Kurtosis		.219	.381	.578
Range		3.67	3.38	3.39
Minimum		.02	.06	.02
Maximum		3.69	3.44	3.41

a. Gender of Concert Goer = Female

Normality within Groups

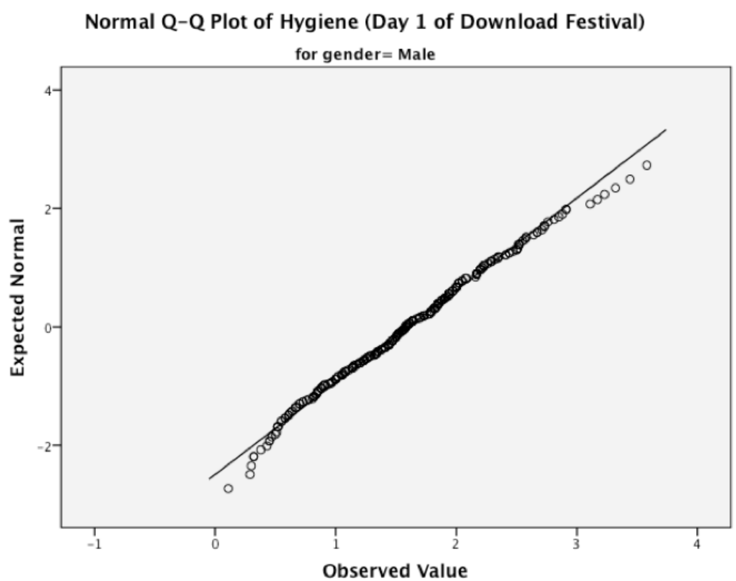
Tests of Normality

	Gender of Concert Goer	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Hygiene (Day 1 of Download Festival)	Male	.035	315	.200*	.993	315	.119
	Female	.053	495	.002	.993	495	.029

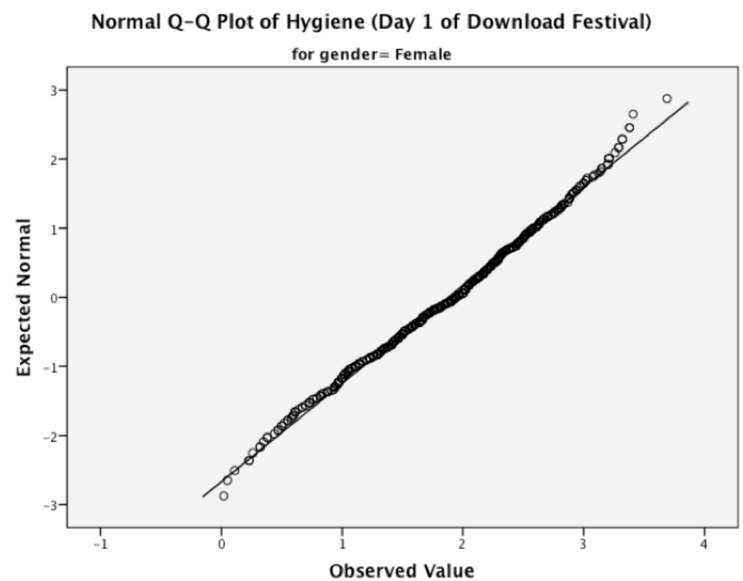
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

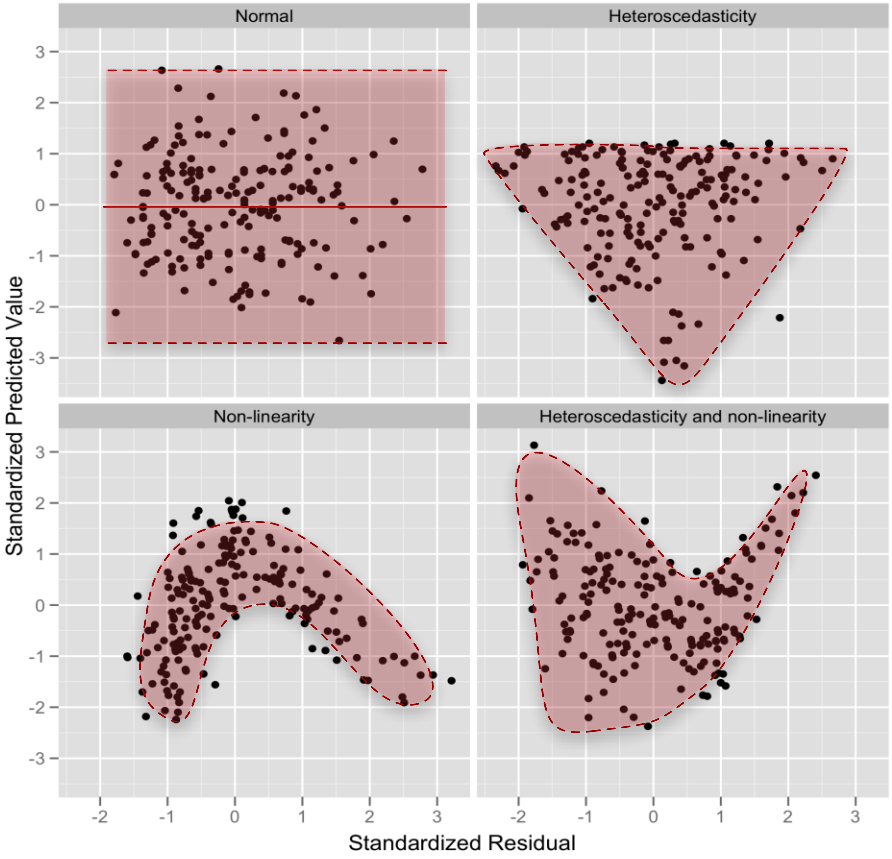
Male



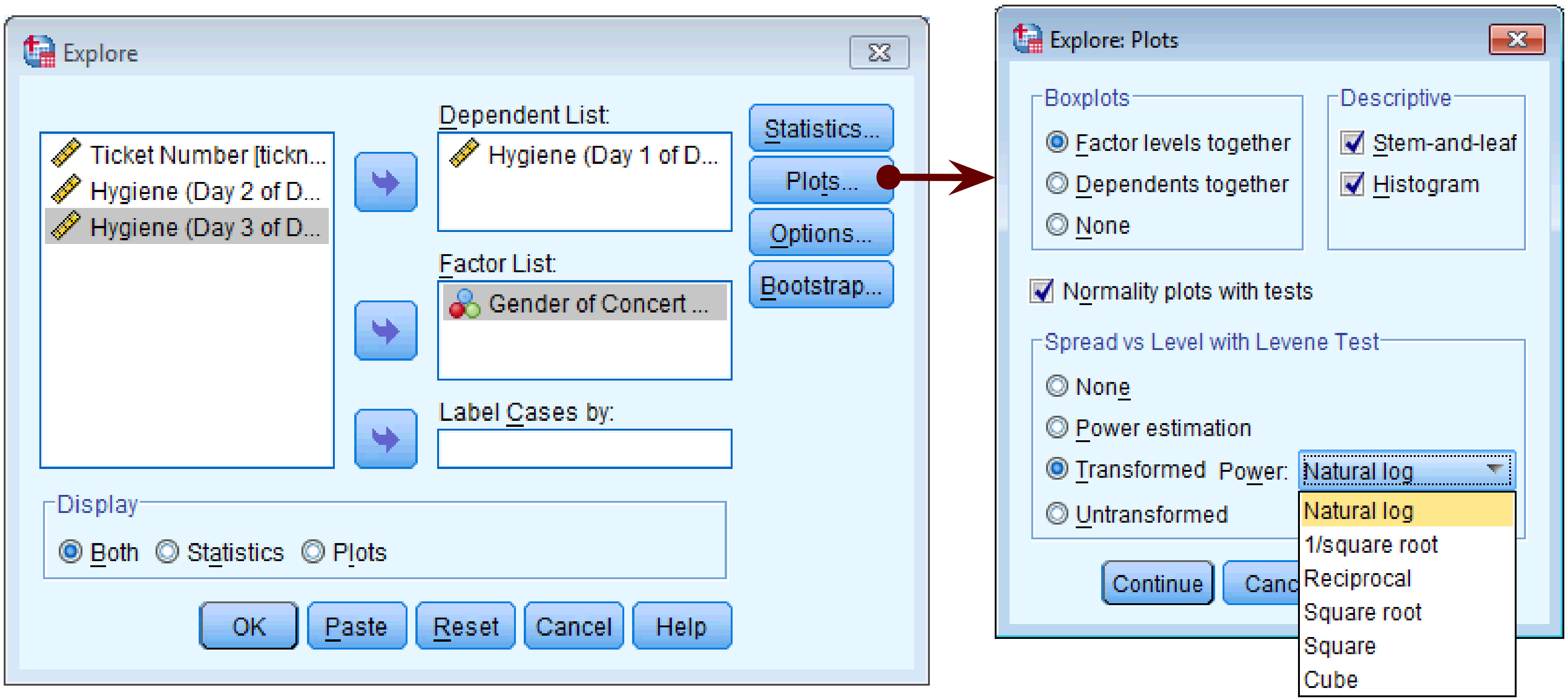
Female



Spotting problems with Linearity or Homoscedasticity



Assessing Homogeneity of Variance



Output for Levene's Test

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Hygiene (Day 1 of Download Festival)	Based on Mean	4.736	1	808	.030
	Based on Median	4.354	1	808	.037
	Based on Median and with adjusted df	4.354	1	805.066	.037
	Based on trimmed mean	4.700	1	808	.030

Reducing Bias

- Trim the data:
  - Delete a certain amount of scores from the extremes.
- Windsorizing:
  - Substitute outliers with the highest value that isn't an outlier
- Analyse with Robust Methods:
  - Bootstrapping
- Transform the data:
  - By applying a mathematical function to scores.

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

4TH EDITION

Trimming the Data

	Data		
	0 0 4 5 6 6 6 6 6 6 7 7 7 7 7 8 8 8 8 8		Ordered Data
5%	0 4 5 6 6 6 6 6 6 7 7 7 7 7 8 8 8 8	5%	5% Trim
10%	4 5 6 6 6 6 6 6 7 7 7 7 7 8 8 8	10%	10% Trim
20%	6 6 6 6 6 6 7 7 7 7 7 8	20%	20% Trim

ANDY FIELD

Robust Methods: Examples

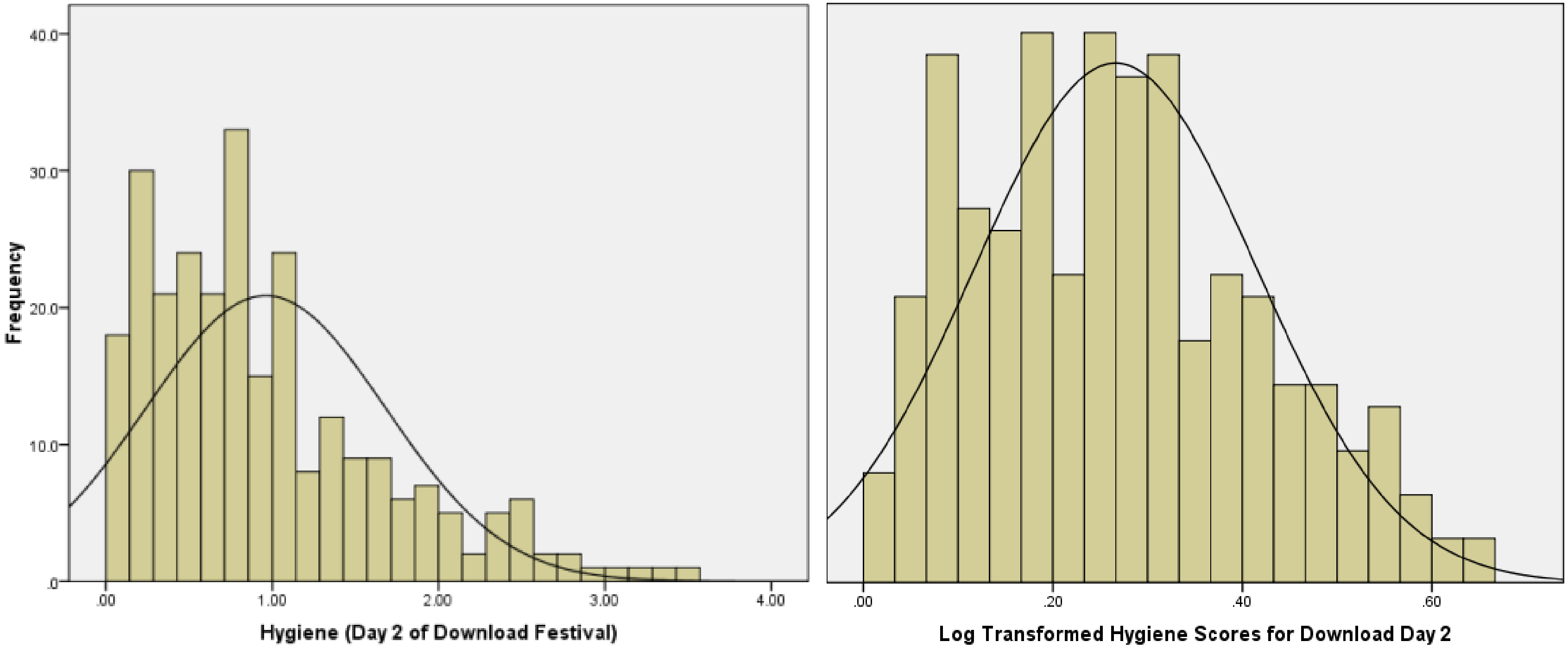
	Comparing Treatments	Relationships
Principle	Bootstrap	Bootstrap
	Trimmed Means	Least Trimmed Squares
	M-estimators	M-estimators
	Median	Least Median of Squares
Equivalent Tests	T-test	Correlation
	ANOVA (Including factorial)	Regression
	ANCOVA	ANCOVA
	MANOVA	

Transforming Data

- Log Transformation ($\log(X_i)$)
  - Reduce positive skew.
- Square Root Transformation ($\sqrt{X_i}$):
  - Also reduces positive skew. Can also be useful for stabilizing variance.
- Reciprocal Transformation ($1/X_i$):
  - Dividing 1 by each score also reduces the impact of large scores. This transformation reverses the scores, you can avoid this by reversing the scores before the transformation, $1/(X_{Highest} - X_i)$.

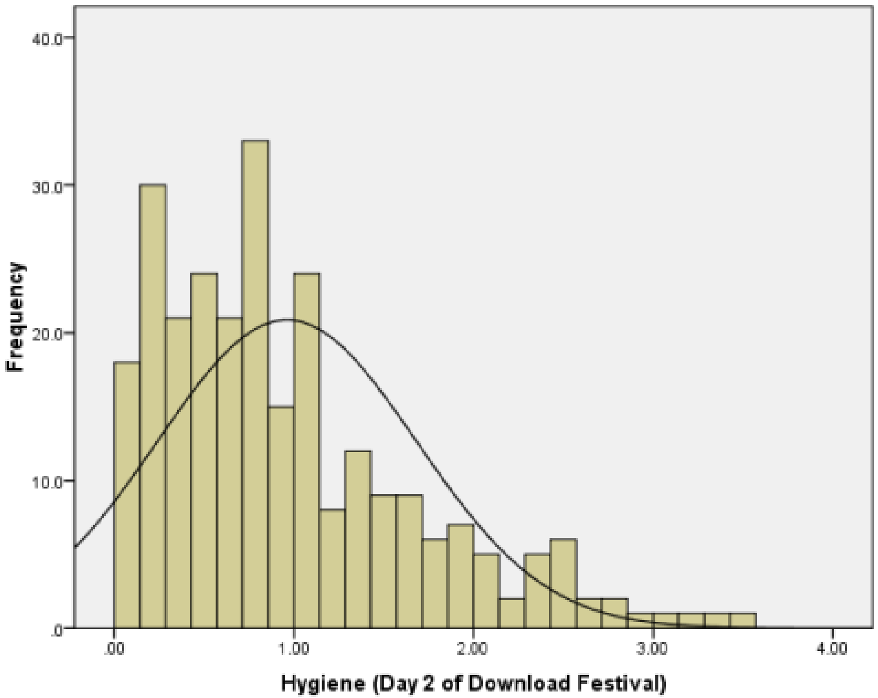
Log Transformation

Before After

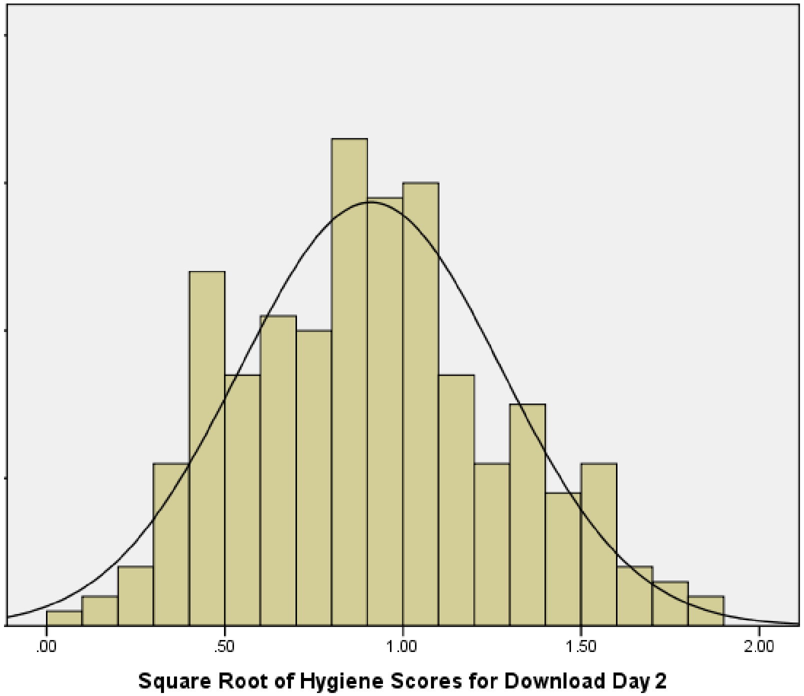


Square Root Transformation

Before

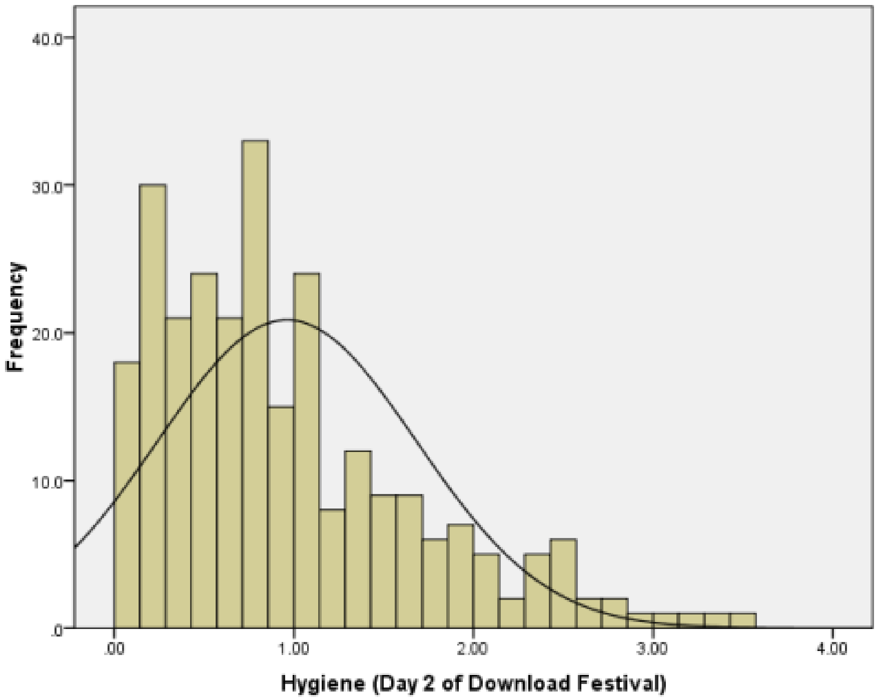


After

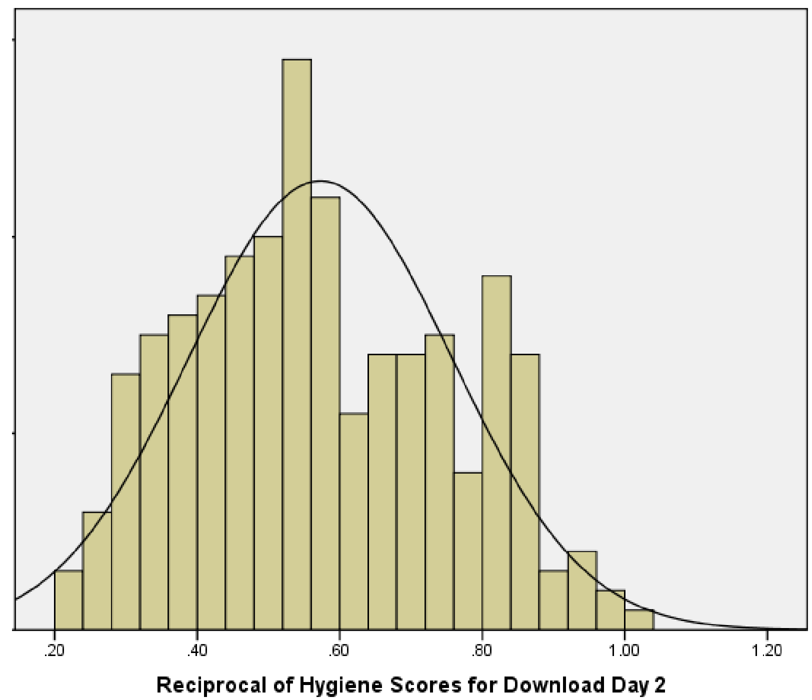


Reciprocal Transformation

Before

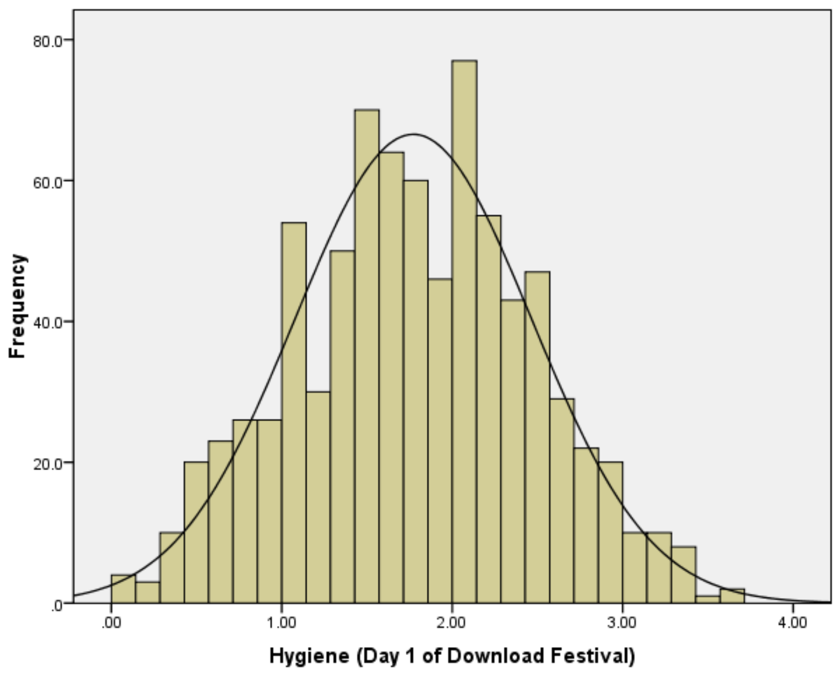


After

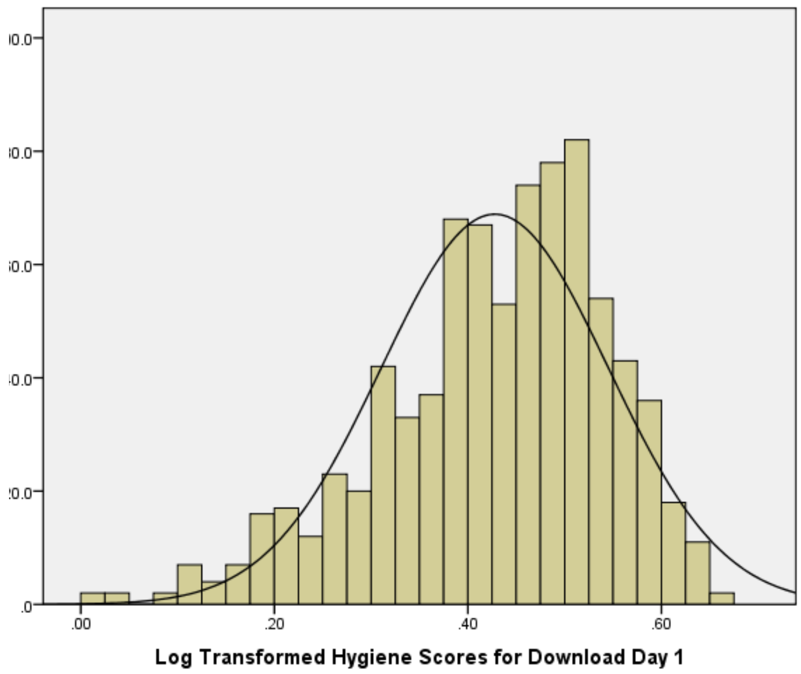


But …

Before



After



To Transform … Or Not

- Transforming the data helps as often as it hinders the accuracy of *F* (Games & Lucas, 1966).
- Games (1984):
  - The central limit theorem: sampling distribution will be normal in samples > 40 anyway.
  - Transforming the data changes the hypothesis being tested
    - E.g. when using a log transformation and comparing means you change from comparing arithmetic means to comparing geometric means
  - In small samples it is tricky to determine normality one way or another.
  - The consequences for the statistical model of applying the ‘wrong’ transformation could be worse than the consequences of analysing the untransformed scores.