Chapter 6 Conclusion

The major issues addressed in this thesis are:

- an investigation of the relative power of common categorical goodness-of-fit test statistics for a number of practical null and alternative distributions;
- the creation of the Combined Kolmogorov-Smirnov test statistic for partially ranked categorical data - a new Empirical Distribution Function type goodness-of-fit test statistic;
- an investigation of the power of the new Combined Kolmogorov-Smirnov test statistic for a variety of practical null and alternative distributions which may be of use to applied researchers;
- the promotion of Monte Carlo approximations of the exact distribution of categorical goodness-of-fit test statistics.

Although the assessment of whether a sample of observations comes from a specified multinomial distribution is widespread throughout the research community, the use of categorical goodness-of-fit test statistics is often made without due reference to power. The literature associated with the powers of categorical empirical distribution function type goodness-of-fit test statistics is quite thorough when dealing with a uniform null distribution against a trend alternative distribution. However, it appears to be quite limited for null hypotheses defined by non-uniform distributions. Confusion exists in the literature relating to the relative power of categorical Empirical Distribution Function type goodness-of-fit test statistics and this confusion is addressed in this thesis with power studies of selected test statistics for a variety of null and alternative distributions.

This thesis investigates the powers of the Chi-Square and Empirical Distribution Function type test statistics for null and alternative hypotheses defined by uniform, continuous trend $A_1(\delta)$, step type $A_2(\delta)$, triangular A_3, platykurtic A_4, leptokurtic A_5 and bimodal A_6 distributions. Based on the results from the power studies in Chapter 4, it is impossible to make a general recommendation to always use a particular categorical goodness-of-fit test statistic, irrespective of whether the categorical data posses a natural ordering or not. However, should categorical data be completely ordinal, a summary of the recommendations, based on power, is made for:

- consistent trend (Table 6.1);
- step type trend (Table 6.2);
- triangular (Table 6.3);
- platykurtic (Table 6.4);
- leptokurtic (Table 6.5);
- bimodal (Table 6.6).

Table 6.1 Power recommendations for 6 categorical GOF test statistics – Consistent trend distributions.

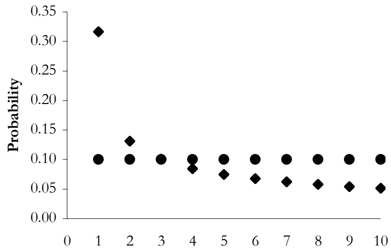
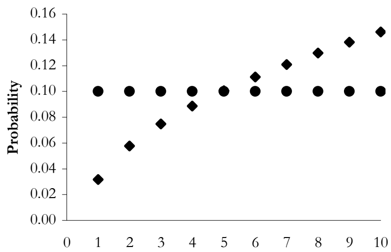
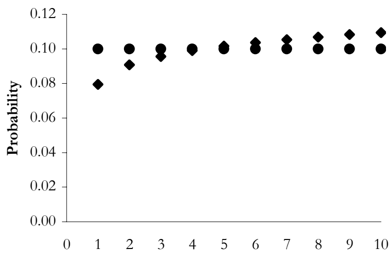
Null or Alternative Distribution	Graph	Recommended Test Statistic(s) with Higher Power for a Uniform Null	Recommended Test Statistic(s) with Higher Power for a Uniform Alternative
Steep decreasing consistent trend	 **Figure 6.1 Steep decreasing trend and uniform distribution**	Ordinal Kolmogorov-Smirnov (OrdKS) Anderson-Darling $\left(T_{A^2(Cat)}\right)$ Cramér-von Mises $\left(T_{W^2(Cat)}\right)$	Anderson-Darling $\left(T_{A^2(Cat)}\right)$ Cramér-von Mises $\left(T_{W^2(Cat)}\right)$
Steep increasing consistent trend	 **Figure 6.2 Steep increasing trend and uniform distribution**	Ordinal Kolmogorov-Smirnov (OrdKS) Anderson-Darling $\left(T_{A^2(Cat)}\right)$ Cramér-von Mises $\left(T_{W^2(Cat)}\right)$	Ordinal Kolmogorov-Smirnov (OrdKS) Anderson-Darling $\left(T_{A^2(Cat)}\right)$ Cramér-von Mises $\left(T_{W^2(Cat)}\right)$ Chi-Square
Near uniform consistent trend	 **Figure 6.3 Near uniform trend and uniform distribution**	None recommended as all are shown to generally have poor power for a uniform null against a near uniform alternative distribution.	Not investigated. However it is expected that the conclusion is similar to that for a uniform null against a near uniform trend alternative distribution. That is, none recommended with higher power.

Table 6.2 Power recommendations for 6 categorical GOF test statistics – Step type trend distributions.

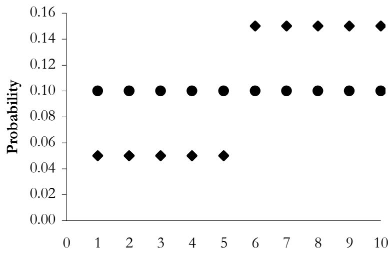
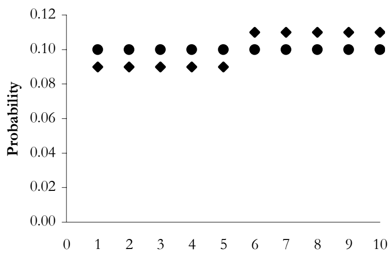
Null or Alternative Distribution	Graph	Recommended Test Statistic(s) with Higher Power for a Uniform Null	Recommended Test Statistic(s) with Higher Power for a Uniform Alternative
Step type trend	 **Figure 6.4 Step type trend and uniform distribution**	Ordinal Kolmogorov-Smirnov (OrdKS) Anderson-Darling $\left(T_{A^2(Cat)}\right)$ Cramér-von Mises $\left(T_{W^2(Cat)}\right)$ and, to a lesser extent Watson $\left(T_{U^2(Cat)}\right)$	Ordinal Kolmogorov-Smirnov (OrdKS) Anderson-Darling $\left(T_{A^2(Cat)}\right)$ Cramér-von Mises $\left(T_{W^2(Cat)}\right)$
Near uniform step type	 **Figure 6.5 Near uniform step type and uniform distribution**	None recommended as all are shown to generally have poor power for a uniform null against this near uniform step type alternative distribution.	Not investigated. However it is expected that the conclusion is similar to that for a uniform null against a near uniform step type alternative distribution. That is, none recommended with higher power.

Table 6.3 Power recommendations for 6 categorical GOF test statistics – Triangular distributions.

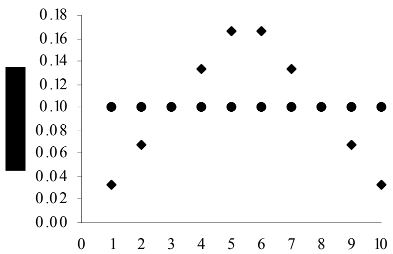
Null or Alternative Distribution	Graph	Recommended Test Statistic(s) with Higher Power for a Uniform Null	Recommended Test Statistic(s) with Higher Power for a Uniform Alternative
Triangular	 **Figure 6.6 Triangular and uniform distribution**	Watson $\left(T_{U^2(Cat)}\right)$ and, to a lesser extent Chi-Square Nominal Kolmogorov-Smirnov (NomKS)	Chi-Square Watson $\left(T_{U^2(Cat)}\right)$

Table 6.4 Power recommendations for 6 categorical GOF test statistics – Platykurtic distributions.

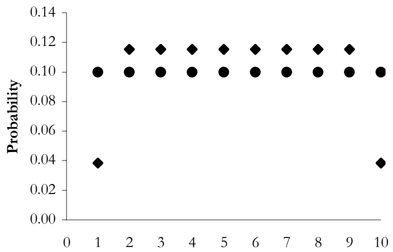
Null or Alternative Distribution	Graph	Recommended Test Statistic(s) with Higher Power for a Uniform Null	Recommended Test Statistic(s) with Higher Power for a Uniform Alternative
Platykurtic	 **Figure 6.7 Platykurtic and uniform distribution**	Chi-Square Nominal Kolmogorov-Smirnov (NomKS) Watson $\left(T_{U^2(Cat)}\right)$	Chi-Square Anderson-Darling $\left(T_{A^2(Cat)}\right)$

Table 6.5 Power recommendations for 6 categorical GOF test statistics – Leptokurtic distributions.

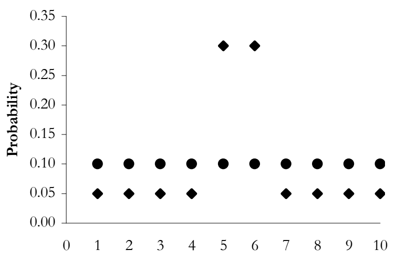
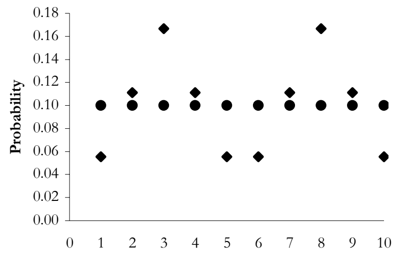
Null or Alternative Distribution	Graph	Recommended Test Statistic(s) with Higher Power for a Uniform Null	Recommended Test Statistic(s) with Higher Power for a Uniform Alternative
Leptokurtic	 **Figure 6.8 Leptokurtic and uniform distribution**	Watson $\left(T_{U^2(Cat)}\right)$ and, to a lesser extent Nominal Kolmogorov-Smirnov (NomKS)	Ordinal Kolmogorov-Smirnov (OrdKS) Anderson-Darling $\left(T_{A^2(Cat)}\right)$ Cramér-von Mises $\left(T_{W^2(Cat)}\right)$ Watson $\left(T_{U^2(Cat)}\right)$ Chi-Square Nominal Kolmogorov-Smirnov (NomKS)

Table 6.6 Power recommendations for 6 categorical GOF test statistics – Bimodal distributions.

Null or Alternative Distribution	Graph	Recommended Test Statistic(s) with Higher Power for a Uniform Null	Recommended Test Statistic(s) with Higher Power for a Uniform Alternative
Bimodal	 **Figure 6.9 Bimodal and uniform distribution**	Chi-Square Nominal Kolmogorov-Smirnov (NomKS) and, to a lesser extent Watson $\left(T_{U^2(Cat)}\right)$	Chi-Square Nominal Kolmogorov-Smirnov (NomKS) and, to a lesser extent Watson $\left(T_{U^2(Cat)}\right)$

The power studies use Monte Carlo approximations of the exact distributions of the selected test statistics for the specifically defined null and alternative distributions. Whilst it is shown, in this thesis and elsewhere, that Monte Carlo approximations with a large number of simulations can generally provide a more accurate approximation of the exact distribution of a test statistic than an asymptotic distribution, it is particularly beneficial as many test statistics do not have an asymptotic approximating distribution and because categories lacking a realistic meaning do not need to be created to simply satisfy an assumption. An historical review of categorical goodness-of-fit test statistics demonstrates that a significant amount of research has been undertaken to determine the conditions under which the asymptotic χ^2 distribution provides an accurate approximation of the exact null distribution of Chi-Square type test statistics - in particular Pearson's Chi-Square test statistic. It is demonstrated that only a relatively small number of simulations, generally about one thousand, are required to gain a reasonably accurate approximation of the exact distribution of the categorical goodness-of-fit test statistics used throughout this thesis.

With limited exceptions, most categorical goodness-of-fit test statistics used by applied researchers assume the data are either nominal or ordinal. A problem is identified where the categories are ranked but some different categories have an equivalent rank value. Ordinal goodness-of-fit test statistics such as the Kolmogorov-Smirnov and Cramér-von Mises are shown to produce vastly different *p*-values based on the ordering of the categories selected – this is the basic argument against using an Empirical Distribution type test statistic for nominal data. A new test statistic, called the Combined Kolmogorov-Smirnov, is proposed to make use of the information in the ranking of the categories for these problems. The new test statistic is shown to be more powerful than common categorical goodness-of-fit test statistics for trend null and alternative distributions. The null and alternative distributions which produce relatively lower power for the Empirical Distribution Function goodness-of-fit test statistics also tend to produce relatively lower power for the new Combined Kolmogorov-Smirnov test statistic. The new Combined Kolmogorov-Smirnov test statistic is recommended as a suitable goodness-of-fit test statistic in problems where an Empirical Distribution Function type test statistic, such as the Ordinal Kolmogorov-Smirnov, is preferred but the categories lack a complete natural ordering.

The major areas recommended for further study from the work done in this thesis are:

- modifications of the three Cramér-von Mises goodness-of-fit test statistics for partially ranked categorical data;
- power studies of the three categorical Cramér-von Mises test statistics formed by these modifications;

- an investigation of the powers of all the categorical goodness-of-fit test statistics discussed in this thesis for other commonly used null and alternative distributions;
- an expansion of the power studies for other significance levels.