Gambling in Deadwood, South Dakota: Analysis of Categorical Survey Data

by

Jill M. Trimble

STAT 575 Written Project

21 February 1995

Chapter 1

The Backround of Deadwood Gaming

In recent years, many states have introduced legislation to establish some form of gambling to correct perceived problems with the state economy. By the middle of 1992, thirty-three states and the District of Columbia operated some form of government-sanctioned gambling. Lotteries, scratch-and-play, horse racing, dog racing, river-boat gambling and full-scale casino gambling have been identified as new sources of revenue for failing state reserves. This rush to gambling was assisted in 1991 through a Supreme Court ruling that opened the opportunity for gambling on Indian lands throughout the United States. Many of these new gambling locations opened by this ruling and state legislation were in rural settings. One of these communities, Deadwood, South Dakota, made the decision to rely on gambling for its very existence.

1.1 Background of Deadwood

About forty miles west of Rapid City, in the northern Black Hills of South Dakota, gold mining provided the foundation for the development of a town that reflected the true American West. Deadwood grew as a twin community for Lead, so-named because it provided the lead seam in the geology that guided the search for Black Hills gold in the world famous Homestake mine. While Lead was a gold mining town, Deadwood was a center of commerce serving as a terminus on the Deadwood to Cheyenne stage route and providing the social life in that part of the American West.

Gambling, prostitution, other vices and excitement made Deadwood one of the necessary stops in the Old West. Wild Bill Hickok and Calamity Jane visited Deadwood on more than one occasion. During his last visit to the city as he gambled in the Number 10 Saloon, Wild Bill was shot holding the poker hand of aces and eights, the famous "Dead Man's Hand." Wild Bill and Calamity Jane are now buried in the Mt. Moriah cemetery on a hill overlooking Deadwood.

Deadwood also provided the inspiration for Edward L. Wheeler's dime novel hero, Deadwood Dick. In addition, Deadwood entertained many other characters of the Old West, some military, some Native American, some on this side of the law, some on that side.

The surrounding Black Hills, the Paha Sapa of the Sioux, provide a dramatic break in topography from the surrounding plains to the east. These hills also provide a temperate summer climate providing an excellent attraction for toursists.

With Mt. Rushmore, the Black Hills, the Black Hills National Forest, the history, the gold, the outdoor recreation, and more, Deadwood was a natural for tourism. Prostitution and gambling continued in various establishments and to various levels well into the 1980s. Yet by the middle of the 1980s Deadwood had declined to fewer than two thousand people. The community was perishing as many small rural towns had died, destined to become another western ghost town. Fires frequently destroyed portions of the town including historic buildings. Property values declined to a point at which in 1988 a downtown business location could be purchased for as little as $7,000.

Then, in 1989 the state legislature and the people of South Dakota approved the return of limited stakes gaming to the historic town of Deadwood. Poker, blackjack, slots, video lottery and keno were approved with a $5.00 limit per bet.

1.2 The New Deadwood

Limited stakes gaming in Deadwood was designed to produce a very different atmosphere than that found in Las Vegas or Atlantic City. Gambling was integrated into the historic fabric of the city, keeping the old hotels, the historic saloons, the brick streets and the western flavor.

Change occurred quite rapidly in historic Deadwood following the return of gambling in 1989. Property values escalated so that lots which were once available for as little as $7,000 now sold for hundreds of thousands of dollars within the span of a year. Property values increased by 7300% forcing residents to sell at the opportunity to make dream-world profits or because they could not afford the property taxes. Gas stations, pharmacies, grocery stores, churches and other traditional marks of a small town were replaced, first by slot machines on the premises, then by casinos. If a license could be acquired for gaming equipment, the potential profits from gambling far exceeded that from the prior small business operation.

Whether the return of gaming to Deadwood has had a positive or negative effect on the city, it is here to stay. As recently as November 1994, South Dakota voters were asked to raise the betting limit to $100. This bill was pushed heavily by actor Kevin Costner and other developers building multi-million dollar family resorts in and around Deadwood. Voters defeated the legislation indicating they thought Deadwood was changing fast enough and a slower and more cautious approach to the growth of gaming should be taken.

Chapter 2

The Survey

2.1 Goals of the Study

The Deadwood Economic Development Authority commissioned Drs. Jeffrey Zeiger and Lowell Caneday from the Center for the Advancement and Study of Tourism at Black Hills State University to study the type of people who visit Deadwood. This information is very important to the Authority since sustainable rural tourism relies on an understanding of the visitors' motivation, attitude and activity. This two year study of visitors to Deadwood was intended to provide a basis for understanding the visitor.

2.2 Questionnaire and Implementation

The questionnaire was distributed periodically between May 1991 and September 1992. The form and type of questions were evolving throughout the sampling period making incomplete questionnaires a very common feature of the data set. The questionnaire itself was nine pages long and read much like a flow chart. There were a total of 107 variables of interest by the end of the study.

The questionnaire was distributed to potential respondents in two ways:

1. The potential respondents were approached at Mount Moriah Cemetery and other locations in Deadwood and asked to fill out and return the questionnaire on the spot.

2. The potential respondents were given a pre-addressed, postage-paid questionnaire and asked to fill it out and return it as time permitted.

The importance of taking a random sample of all Deadwood visitors was taken into consideration before the questionnaire was distributed. Since the questionnaires were distributed two different ways, two different methods of randomization were used:

1. For the first method of distribution, the gaming district and historical sites of Deadwood were mapped. These locations were divided into one of two strata: gaming and historical. From these two strata, one historical site and two gambling sites were randomly chosen to use throughout

the study. Once these sites were chosen, the questionnaire distributors were asked to randomly hand out the questionnaires throughout the day. There was no information from the researchers on how this last step of randomization occurred.

2. For the second method of distribution, the same approach was taken, but there was a third strata: parking lots. Two parking lots were randomly selected and within the lots, a questionnaire was left on every second vehicle. These same two parking lots were used througout the study. If the vehicle was a touring coach, the questionnaire distributor left enough questionnaires for the tourists with the coach driver.

Although steps were taken in order to ensure randomization, the writers of the study expressed concern about the last stage of random sampling. In order to do any analysis on the data resulting from the questionnaire, it needs to be assumed that a random, or at least a representative, sample of Deadwood visitors was taken. The assumption of a random sample will-help to alleviate any bias problems in the study.

2.3 Questionnaire Problems and Suggested Changes

There were some definite problems with the questionnaire. Incompleteness was a very common feature of all the questionnaires. I think this was attributable to many factors.

First, the amount of time that the questionnaire took potential respondents to fill out was a definite problem. The questionnaire was nine pages long, taking an estimated ten minutes to accurately and completely fill it out. The length was a possible distraction to potential respondents in that it took too much time out of their sight-seeing or gambling to fill out in its entirety. This was evidenced by the very large number of incomplete questionnaires returned to the researchers. There were no questionnaires that were entirely completed. I think this was also a problem even when the return-mail questionnaires were used.

Second, the form of the questionnaire was another possible problem. The directions for the questionnaire were simple, but the actual reading of the questionnaire was similar to a flow-chart and somewhat confusing. More than likely a respondent was in a hurry and would skip over parts of the questionnaire by mistake or because it took too much time to determine what was being asked of them. This also could lead to problems of inaccuracy and non-completion.

Third, location of crucial questions was probably the biggest problem with questionnaire. Since there were problems with the length and form of the questionnaire, it seemed wise to ask the questions that would give the most important information at the beginning of the questionnaire. Unfortunately, many important questions (amount spent gambling, occupation, how you heard about Deadwood, etc.) were asked toward the end of the questionnaire. These questions were very often not answered as the result of the length and form problems.

Another idea that could lead to more accurate and complete questionnaires is rewarding the participants. The questionnaire distributors could make sure the potential respondent is interested in filling out the questionnaire and then rewarding them for their time with a parking, food, drink or gambling coupon.

If another study of this type was being planned, I would encourage the researchers to narrow their scope of interest. This would allow a shorter and more concise questionnaire which could lead to more accurate and complete data. The accuracy of the data could also be improved by making the form of the questionnaire easier and faster to fill out by the potential respondents. Even with a shorter questionnaire, I think it is important to place the most important questions at the beginning of the survey, rather than at the end.

The categories for the questions were often poorly chosen by the questionnaire writers. This did not create problems for the questionnaire respondents, but it did lead to a less informative data set than it could have been. These problems will be discussed in Chapter Three of the paper.

Chapter 3

Data Analysis

Since the bulk of the data produced from this study is categorical in nature, contingency tables, Pearson's Chi-Squared test for independence, Analysis of Deviance, and measures of concordance/discordance were employed for purposes of data analysis.

3.1 Tests for Independence Between Variables

3.1.1 Contingency Tables and Pearson's Statistic

One test for association between two variables is the Pearson statistic for the chi-square test of the hypothesis of no association (or independence) for two-dimensional contingency tables. This test involves comparing observed and expected frequencies in a contingency table. The formula for calculating

Pearson's X^2 is:

$$X^2 = \sum_i \sum_j \frac{(\text{observed}_{ij} - \text{expected}_{ij})^2}{\text{expected}_{ij}}$$

where observed_{ij} is the cell count formed from the tally of the data and expected_{ij} is the expected cell count. If the table of expected counts is of interest, it can be calculated using the following function written for S-Plus:

```
> expcount
function(obs)
{
 y <- matrix(rbind(apply(obs, 1, sum)), ncol = 1, byrow = T) %*% matrix(
        rbind(apply(obs, 2, sum)), nrow = 1, byrow = T)/sum(obs)
 return(y)
}
```

(where the variable "obs" is the matrix of observed counts from which a table of expected counts is to be computed.)

The table of individual X^2 cell values calculated from the table of expected counts can be very informative to look at. This table shows how the categories between the two variables contribute to the overall X^2 test statistic. The table can be calculated using the following function written for S-Plus:

```
> chitable
function(obs)
{
        y <- (expcount(obs) - obs)^2/expcount(obs)
        return(y)
}
```

The actual Pearson's X^2 test statistic is calculated by adding up the individual X^2 cell values in the above table. It can also be calculated using the following function written for S-Plus:

```
> chisquare
function(obs)
{
        y <- sum((expcount(obs) - obs)^2)/(expcount(obs))
        return(y)
}
```

This test statistic has an approximate χ^2 distribution under H_0 with $(s-1) \times (t-1)$ degress of freedom where s is the number of possible categorical responses for the variable S and t is the number of possible categorical responses for the variable T.

The returned p-value from this test should be interpreted carefully. It is a good idea to look at the table of expected counts to make sure the expected counts are adequate since the validity of this test depends heavily on the assumption that the expected cell counts are at least moderately large (an average cell count of five or more is often quoted as a rule of thumb). Even when the cell counts are adequate, the χ^2 is only a large-sample approximation to the true distribution of X^2 under the null hypothesis.

Establishing independence or the lack thereof between some variables is of interest in this study. This information is of interest to tourism professionals, who will decide which marketing campaigns will most benefit Deadwood. For example, an established statistical relationship between the variables AGE and AMOUNT GAMBLED would lead to a different marketing strategy than if no relationship was established. There could be a need to target a specific age group who seems to spend more gambling in Deadwood or boost gaming interest in an age group that is not spending a substantial amount of gambling dollars. An example of Pearson's Test can be shown in an analysis of the variables AGE and AMOUNT GAMBLED. In the following table, "NR" denotes the "No Response" category. This table is made up of the tallied counts provided by the survey. The categories of each variable are the same ones that survey respondent was able to choose from.

	AMOUNT GAMBLED					
AGE	NR	\$25	\$50	\$75	\$100	> \$100
NR	60	19	7	4	2	4
< 20	41	21	7	3	3	9
20-25	34	74	31	3	8	16
26-34	77	90	55	16	21	46
35-44	143	203	71	32	35	55
45-54	96	182	69	20	39	74
55-64	86	129	57	18	34	57
> 65	93	125	50	14	19	38

Table 3.1: Table of Observed Counts

There are a few interesting things to note in this table. First, almost 30% of the survey respondents chose not to answer one or both of the questions (NR). Secondly, the legal gambling age in Deadwood is 21, making some of the survey respondents illegal gamblers. This group would not seem a likely focus for a marketing strategy. If the surveyors are truly interested in the gambling habits of this group, it might be more informative to classify them into a group of their own rather than across two categories as is shown here.

The next table of interest is made up of expected cell counts. From the table, it can be seen that the assumption of a moderately large cell count is met:

AGE	NR	$25	$50	$75	$100	> $100
	AMOUNT GAMBLED					
NR	25.31	33.86	13.94	4.42	6.47	12.01
< 20	22.14	29.63	12.20	3.87	5.66	10.51
20-25	43.76	58.55	24.10	7.64	11.18	20.77
26-34	80.40	107.58	44.28	14.04	20.55	38.16
35-44	142.08	190.12	78.26	24.81	36.31	67.43
45-54	126.53	169.31	69.69	22.09	32.33	60.05
55-64	100.43	134.39	55.32	17.54	25.67	47.66
> 65	89.36	119.57	49.22	15.60	22.84	42.41

Table 3.2: Table of Expected Counts

The next step is to look at the table of individual X^2 cell counts:

AGE	NR	$25	$50	$75	$100	> $100
	AMOUNT GAMBLED					
NR	47.57	6.52	3.45	0.04	3.09	5.34
< 20	16.06	2.51	2.21	0.19	1.25	0.22
20-25	2.18	4.08	1.97	2.82	0.91	1.09
26-34	0.14	2.87	2.59	0.27	0.01	1.61
35-44	0.01	0.87	0.67	2.09	0.05	2.29
45-54	7.37	0.95	0.01	0.20	1.37	3.24
55-64	2.07	0.22	0.05	0.01	2.71	1.83
> 65	0.15	0.25	0.01	0.16	0.64	0.46

Table 3.3: Table of Individual Cell X^2 Values

The calculated X^2 statistic is 136.7 with 35 degrees of freedom, returning a p-value ≈ 0. This is testing the hypotheses:

H_0: AGE and AMOUNT GAMBLED are independent
vs.
H_A: AGE and AMOUNT GAMBLED are not independent

The very small p-value leads to the conclusion that H_0 should be rejected, implying that AGE and AMOUNT GAMBLED are not independent. This means that there is some sort of relationship between the two variables. It will be informative to examine the form of this relationship further.

The table of individual X^2 cell values shows that a large amount of the dependence comes from the "No Response" categories. This is determined by looking at the Table 3.3. The "No Response" categories make up almost 70% of the X^2 statistic. It appears that the "No Response" categories contribute the most to the conclusion that AGE and AMOUNT GAMBLED are dependent. If the dependence between the two variables is based entirely on this relationship, it is not a informative relationship in terms of gaining any valuable marketing information. If, however, the dependency relationship is not entirely due to the "No Response" category, then a relationship between the two variables AGE and AMOUNT GAMBLED has been established. Analysis of Deviance can be used to determine how influential the "No Response" categories are on the relationship between AGE and AMOUNT GAMBLED.

3.1.2 Analysis of Deviance

The Deviance Statistic (denoted G^2) tests for independence between two categorical variables. This statistic also makes use of a table of expected counts. In the case of the categorical variables AGE and AMOUNT GAMBLED, the G^2 statistic will be helpful in determining how much of an influence the "No Responses" had in the decision to reject H_0 : AGE and AMOUNT GAMBLED are independent, based on the Pearson's X^2 test statistic.

In order to determine the influence of the "No Response" categories, two tables will be needed: the table of observed counts and the condensed 2×2 table of "Response" by "No Response" for each categorical variable. A G^2 statistic will be calculated for each table. G_l^2 denotes the value of the statistic for the large table and G_s^2 the value for the smaller table. Both of these values are calculated using the same formula:

$$G^2 = 2 \sum_i \sum_j \text{observed}_{ij} \log \left(\frac{\text{observed}_{ij}}{\text{expected}_{ij}} \right)$$

The G^2 statistic can also be calculated using the following function written for S-Plus:

```
> deviant
function(obs)
{
        y <- 2 * sum(obs * log(obs/expcount(obs)))
        return(y)
}
```

The distribution of the G^2 statistic under H_0 is approximately χ^2 with $(s-1) \times (t-1)$ degrees of freedom where s is the number of categories in the S variable and t is the number of categories in the T variable. The G^2 statistic is testing the hypotheses:

H_0: variables T and S are independent
vs.
H_A: variables T and S are not independent

In the case of the variables AGE and AMOUNT GAMBLED, there are two tables of interest. The 8×6 table of observed counts:

	AMOUNT GAMBLED					
AGE	NR	\$25	\$50	\$75	\$100	> \$100
NR	60	19	7	4	2	4
< 20	41	21	7	3	3	9
20-25	34	74	31	3	8	16
26-34	77	90	55	16	21	46
35-44	143	203	71	32	35	55
45-54	96	182	69	20	39	74
55-64	86	129	57	18	34	57
> 65	93	125	50	14	19	38

Table 3.4: Table of Observed Counts

and the 2×2 condensed table of "Response" by "No Response" for AGE and AMOUNT GAMBLED:

		AMOUNT GAMBLED	
		No Response	Response
AGE	No Response	60	36
	Response	570	1724

Table 3.5: Condensed Table of Observed Counts

The individual G^2 statistics calculated for each table are: $G_l^2 = 126.9$ with 35 degrees of freedom, p-value ≈ 0 and $G_s^2 = 57.7$ with 1 degree of freedom, p-value ≈ 0. For both cases, the appropriate conclusion based on these p-values is the variables are not independent.

The difference in the G^2 statistics between the larger and the smaller tables is then calculated to test the association of the two variables after the effect of "No Response" versus "Response" has been removed. This difference, $G_l^2 - G_s^2$, is distributed approximately χ^2 with $(s-1)\times(t-1)-1$ degrees of freedom where s is the number of categories in the S variable and t is the number of categories in the T variable under the following null hypothesis:

H_0: variables T and S are independent *after* removal of NR effect
vs.
H_A: variables T and S are not independent *after* removal of NR effect

Now, the difference in the G^2 statistics $(G_l^2 - G_s^2)$ is calculated in order to determine if the dependence in the large table can be accounted for by the relationship of the "No Responses" and "Responses". $G_l^2 - G_s^2$=69.2 with 34 degrees of freedom, $p \approx 0.0003$. These values lead to the conclusion AGE and AMOUNT GAMBLED are not independent *after* removal of NR effect.

The establishment of dependence between the variables AGE and AMOUNT GAMBLED is interesting. There are many ways that the variables could be related. Two of the possible ways are: First, the amount spent gambling could be related to the age of the gambler in a positive relationship. The older a gambler is, the more they gamble (possibly because of a larger amount of disposable income). Or, second, the amount spent gambling could be related to the age of the gambler in a negative realtionship. The younger a gambler is, the more they spend on gambling (here, large amounts gambled by a younger person could be attributable to the recklessness of youth). In order to establish whether either of these relationships *does* exist, a look at the measure of concordance or discordance between the two variables is useful.

3.2 The Measure of Concordance/Discordance Between Two Variables

When the response of two subjects on each of two categorical variables, say T and S, are observed, the pair of subjects can be classified as concordant or discordant. The pair is classified as concordant if the survey respondent ranking higher on variable T also ranks higher on variable S. The pair is classified as discordant if the survey respondent ranking higher on variable T also ranks lower on variable S or if the survey respondent ranking lower on variable T ranks higher on variable S. The pair is tied if the survey respondents have the same classification on T and/or S. The measure of concordance/discordance should only be used with ordered categories.

The first step in determining this measure of association is calculating the total number of concordant pairs (C) and the total number of discordant pairs (D). A result showing $C > D$ would indicate the two variables are concordant while a result showing $C < D$ would indicate the two variables are discordant. The values of concordance and discordance can be calculated using the following functions written for S-Plus:

```
> concord
function(obs)
{
        d1 <- dim(obs)[1]
        d2 <- dim(obs)[2]
        conc <- 0
        for(i in 1:(d1 - 1)) {
          for(j in 1:(d2 - 1)) {
            conc <- conc + obs[i, j] * sum(obs[(i + 1):d1, (j + 1):d2])
        } }
        return(conc)
}

> discord
function(obs)
{
        d1 <- dim(obs)[1]
        d2 <- dim(obs)[2]
        disc <- 0
        for(i in 1:(d1 - 1)) {
          for(j in 2:d2) {
            disc <- disc + obs[i, j] * sum(obs[(i + 1):d1, 1:(j - 1)])
        } }
        return(disc)
}
```

While knowing the association between two variables is useful, it would be even more informative to know the strength of this relationship. An analagous situation would be to find the Pearson correlation coefficient (ρ) for continuous data. This value of ρ describes the strength of the linear relationship between two variables. By the nature of categorical variables, this type of relationship does not always make sense. However, since we are dealing only with ordered categories, it does make sense to look at the whether the relationship between the two variables is monotone. For example, T tends to increase as S does. The measure for categorical variables that describes the degree to which the relationship is monotone is γ and can be estimated by the formula:

$$\hat{\gamma} = \frac{C - D}{C + D}$$

Like the Pearson correlation coefficient for continuous data, $-1 \leq \hat{\gamma} \leq 1$. Here $\hat{\gamma} = 1$ denotes a monotone relationship between the two variables that is not necessarily strictly monotone. For instance, for $\hat{\gamma} = 1$, then for observations (X_a, Y_a) and (X_b, Y_b) on a pair of subjects a and b where $X_a < X_b$, it follows from above that $Y_a \leq Y_b$ (monotone increasing) but not necessarily that $Y_a < Y_b$ (strictly monotone increasing). The same holds true for $\hat{\gamma} = -1$, except the relationship is monotone decreasing. Finally,

$\hat{\gamma} = 0$ indicates that there is not a monotone relationship (either increasing or decreasing) between the two categorical variables.

3.3 Summary Analysis on Variables of Interest

The relationship between the variables AGE and AMOUNT GAMBLED are discussed in the previous sections. For these and the following variables of interest, I will look at:

- Pearson's X^2 Test for Independence
- The Deviance Statistic
- Measure of Concordance/Discordance

3.3.1 Income by Amount Gambled

Pearson's X^2 Test for Independence

The following table summarizes the observed counts, expected counts and individual cell X^2 values, respectively for the variables INCOME and AMOUNT GAMBLED. The results from the table are: $X^2 = 132.98$ degrees of freedom: 45 p-value ≈ 0. This small p-value indicates that INCOME and AMOUNT GAMBLED are not independent.

INCOME	AMOUNT GAMBLED NR	$25	$50	$75	$100	> $100
NR	262	224	75	25	33	60
	179.0	239.5	98.6	31.3	45.7	84.9
	38.5	1.0	5.6	1.3	3.5	7.3
< $15000	24	47	16	6	5	12
	29.0	38.8	16.0	5.1	7.4	13.8
	0.9	1.7	0.0	0.2	0.8	0.2
$15 – 20,000	22	43	22	8	10	16
	31.9	42.7	17.6	5.6	8.2	15.1
	3.1	0.0	1.1	1.1	0.4	0.0
$20 – 25,000	34	58	25	8	12	22
	41.9	56.1	23.1	7.3	10.7	19.9
	1.5	0.1	0.2	0.1	0.2	0.2
$25 – 35,000	60	95	47	20	20	34
	72.8	97.4	40.1	12.7	18.6	34.5
	2.2	0.1	1.2	4.2	0.1	0.0
$35 – 45,000	61	138	49	10	25	41
	85.4	114.3	47.0	14.9	21.8	40.5
	7.0	4.9	0.1	1.6	0.5	0.0
$45 – 55,000	58	93	55	12	24	31
	72.0	96.3	39.6	12.6	18.4	34.2
	2.7	0.1	6.0	0.0	1.7	0.3
$55 – 70,000	44	60	31	14	18	31
	52.2	69.8	28.7	9.1	13.3	24.8
	1.3	1.4	0.2	2.6	1.6	1.6
$70 – 85,000	28	41	12	2	7	18
	28.5	38.1	15.7	5.0	7.3	13.5
	0.0	0.2	0.9	1.8	0.0	1.5
> $85,000	37	44	15	5	7	34
	37.4	50.1	20.6	6.5	9.6	17.8
	0.0	0.7	1.5	0.4	0.7	14.8

Table 3.6: Table of Observed, Expected and Individual X^2

The Deviance Statistic

In order to calculate the $G_l^2 - G_s^2$, the 2×2 table of "Response" by "No Response" for INCOME and AMOUNT GAMBLED is needed:

		AMOUNT GAMBLED No Response	AMOUNT GAMBLED Response
INCOME	No Response	262	417
	Response	386	1325

Table 3.7: Condensed Table of Observed Counts

$G_l^2 = 126.6$, 45 degrees of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is the variables INCOME and AMOUNT GAMBLED are not independent in the large table.

$G_s^2 = 60.7$, 1 degree of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is the variables INCOME and AMOUNT GAMBLED are not independent in the condensed table.

$G_l^2 - G_s^2 = 65.9$, 44 degrees of freedom, p-value ≈ 0.02. Based on this p-value, the appropriate conclusion is the variables INCOME and AMOUNT GAMBLED are not independent *after* removal of NR effect at the 5% significance level.

Measure of Concordance/Discordance

$C = 1061391 \qquad D = 806523 \qquad \hat{\gamma} = 0.14$

Since $C > D$, the pair of variables is concordant. Also, the value of $\hat{\gamma} = 0.14$ indicates that there is a somewhat weak tendency in this sample for income levels to be higher at higher amount spent gambling levels.

3.3.2 Game by Amount Gambled

Pearson's X^2 Test for Independence

	AMOUNT GAMBLED					
GAME	NR	$25	$50	$75	$100	> $100
NR	109	9	0	0	2	5
	32.9	44.1	18.1	5.8	8.4	15.6
	175.5	27.9	18.1	5.8	4.9	7.2
NONE	365	8	0	0	0	15
	102.3	136.9	56.3	17.9	26.1	48.5
	674.9	121.3	56.3	17.9	26.1	23.2
POKER	23	39	33	20	14	31
	42.2	56.4	23.2	7.4	10.8	20.0
	8.7	5.4	4.1	21.7	1.0	6.0
BLACKJACK	34	114	69	25	42	55
	89.4	119.6	49.2	15.6	22.8	42.4
	34.3	0.3	8.0	5.7	16.1	3.7
SLOTS	88	658	241	63	100	179
	350.3	468.8	193.0	61.2	89.5	166.3
	196.4	76.4	12.0	0.1	1.2	1.0
VIDEO LOTTERY	6	12	4	2	1	14
	10.3	13.8	5.7	1.8	2.6	4.9
	1.8	0.2	0.5	0.0	1.0	17.1
KENO	5	3	0	0	2	0
	2.6	3.5	1.5	0.5	0.7	1.3
	2.1	0.1	1.5	0.5	2.6	1.3

Table 3.8: Table of Observed, Expected and Individual X^2

$X^2 = 1589.7$ degrees of freedom: 30 p-value ≈ 0

This small p-value indicates that GAME and AMOUNT GAMBLED are not independent.

The Deviance Statistic

In order to calculate the $G_l^2 - G_s^2$, the 2×2 table of "Response" by "No Response" for GAME and AMOUNT GAMBLED is needed:

		AMOUNT GAMBLED	
		No Response	Response
GAME	No Response	109	16
	Response	521	1744

Table 3.9: Condensed Table of Observed Counts

G_l^2 cannot be calculated because many of the observed and expected cell counts are zero. Adding 0.1 to each observed cell count will allow some information to be gained from the table. In this case, $G_l^2 = 1556.2$, 30 degrees of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is GAME and AMOUNT GAMBLED are not independent in the modified large table.

$G_s^2 = 218.4$, 1 degree of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is GAME and AMOUNT GAMBLED are not independent in the small table.

$G_l^2 - G_s^2$ cannot be calculated since G_l^2 cannot be calculated. Using the modified larger and smaller tables, $G_l^2 - G_s^2 = 1339.2$, 29 degrees of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is the variables GAME and AMOUNT GAMBLED are *not* independent *after* removal of NR effect.

Measure of Concordance/Discordance

A measure of concordance or discordance would not make sense for this pair of variables since there is no natural ordering to the variable GAME.

3.3.3 Occupation by Amount Gambled

Pearson's X^2 Test for Independence

The following table summarizes the observed counts, expected counts and individual cell X^2 values, respectively for the variables OCCUPATION and AMOUNT GAMBLED. The results from the table are: $X^2 = 168.3$ degrees of freedom: 30 p-value ≈ 0. This small p-value indicates that OCCUPATION and AMOUNT GAMBLED are not independent.

OCCUPATION	AMOUNT GAMBLED NR	$25	$50	$75	$100	> $100
NR	120	33	14	7	10	9
	50.9	68.1	28.0	8.9	13.0	24.1
	93.9	18.1	7.0	0.4	0.7	9.5
FULL-TIME	321	505	219	66	108	195
	372.7	498.7	205.3	65.1	95.3	176.9
	7.2	0.1	0.9	0.0	1.7	1.9
PART-TIME	19	45	16	7	4	15
	27.9	37.4	115.4	4.9	7.1	13.3
	2.9	1.5	0.0	0.9	1.4	0.2
RETIRED	129	184	73	26	32	57
	132.1	176.7	72.7	23.1	33.7	62.7
	0.1	0.3	0.0	0.4	0.1	0.5
STUDENT	19	25	6	1	1	7
	15.6	20.8	8.6	2.7	4.0	7.4
	0.8	0.8	0.8	1.1	2.2	0.0
HOMEMAKER	12	39	13	3	6	13
	22.7	30.3	12.5	4.0	5.8	10.8
	5.0	2.5	0.0	0.2	0.0	0.5
UNEMPLOYED	10	12	6	0	0	3
	8.2	10.9	4.5	1.4	2.1	3.9
	0.4	0.1	0.5	1.4	2.1	0.2

Table 3.10: Table of Observed, Expected and Individual X^2

The Deviance Statistic

In order to calculate the $G_l^2 - G_s^2$, the 2×2 table of "Response" by "No Response" for OCCUPATION and AMOUNT GAMBLED is needed:

		AMOUNT GAMBLED	
		No Response	Response
OCCUPATION	No Response	120	73
	Response	510	1687

Table 3.11: Condensed Table of Observed Counts

G_l^2 cannot be calculated because many of the observed and expected cell counts are zero. Adding 0.1 to each observed cell count will allow some information to be gained from the table. In this case, $G_l^2 = 154.9$, 30 degrees of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is OCCUPATION and AMOUNT GAMBLED are not independent in the modified large table.

$G_s^2 = 120.2$, 1 degree of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is the variables OCCUPATION and AMOUNT GAMBLED are not independent.

$G_l^2 - G_s^2$ cannot be calculated since G_l^2 cannot be calculated. Using the modified larger table, $G_l^2 - G_s^2 = 34.7$, 29 degrees of freedom, p-value ≈ 0.21. Based on this large p-value, the appropriate conclusion is to fail to reject H_0 the variables OCCUPATION and AMOUNT GAMBLED are independent *after* removal of NR effect.

Measure of Concordance/Discordance

A measure of concordance or discordance would not make sense for this pair of variables since there is no natural ordering to the variable GAME.

3.3.4 Gender by Amount Gambled

Pearson's X^2 Test for Independence

	AMOUNT GAMBLED					
GENDER	NR	\$25	\$50	\$75	\$100	> \$100
	97	49	15	3	8	15
NR	49.3	66.0	27.2	8.6	12.6	23.4
	46.2	4.4	5.4	3.7	1.7	3.0
	321	444	185	69	101	174
MALE	341.1	456.4	187.9	59.6	87.2	161.9
	1.2	0.3	0.0	1.5	2.2	0.9
	212	350	147	38	52	110
FEMALE	239.6	320.6	132.0	41.8	61.2	113.7
	3.2	2.7	1.7	0.4	1.4	0.1

Table 3.12: Table of Observed, Expected and Individual X^2

$X^2 = 79.9$ degrees of freedom: 10 p-value ≈ 0

This small p-value indicates that GENDER and AMOUNT GAMBLED are not independent.

The Deviance Statistic

In order to calculate the $G_l^2 - G_s^2$, the 2×2 table of "Response" by "No Response" for GENDER and AMOUNT GAMBLED is needed:

		AMOUNT GAMBLED	
		No Response	Response
GENDER	No Response	97	90
	Response	533	1670

Table 3.13: Condensed Table of Observed Counts

$G_l^2 = 73$, 10 degrees of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is the variables GENDER and AMOUNT GAMBLED are not independent in the large table.

$G_s^2 = 60.2$, 1 degree of freedom, p-value ≈ 0. Based on this small p-value, the appropriate conclusion is the variables GENDER and AMOUNT GAMBLED are not independent in the condensed table.

$G_l^2 - G_s^2 = 12.8$, 9 degrees of freedom, p-value ≈ 0.17. Based on this p-value, the appropriate conclusion is fail to reject H_0 : the variables GENDER and AMOUNT GAMBLED are independent *after* removal of NR effect at the 5% significance level.

Measure of Concordance/Discordance

A measure of concordance or discordance would not make sense for this pair of variables since there is no natural ordering to the variable GENDER.

3.4 Summary

The Pearson's X^2 Test for Independence indicated that the variable AMOUNT GAMBLED is not independent of the variables AGE, INCOME, GAME, OCCUPATION or GENDER. This dependence could be due to the high number of "No Responses" in the data set and the association between the "No Response" categories for different questions. The Difference Deviance Statistic will determine if this is the case.

The Difference Deviance Statistic, which describes the relationship between the variables *after* the removal of the "No Response" effect, indicated the dependence between the variable AMOUNT GAMBLED and the variables INCOME and GAME is not just due to the association between the "No Responses" on INCOME and GAME. The dependence between the variables AMOUNT GAMBLED and OCCUPATION is due to the association between the "No Responses" on OCCUPATION and AMOUNT GAMBLED. Also, the dependence between the variables AMOUNT GAMBLED and GENDER is due to the association between the "No Responses" on GENDER and AMOUNT GAMBLED.

The only two variables of interest that a measure of concordance/discordance makes sense for are AMOUNT GAMBLED and INCOME. It was determined that the pair of variables is concordant. This indicates that there is a tendency in this sample for income levels to be higher at higher amount spent gambling levels. The strength of this tendency was determined to be fairly weak.

Bibliography

Agresti, Alan. (1990) *Categorical Data Analysis.* John Wiley & Sons: New York, pp. 19-24.

Charlier, Marj. "Casino Gambling Saves Three Colorado Towns, But the Price is High." *The Wall Street Journal.* September 23, 1992.

Eadington, William R. (1986). "Impact of Casino Gambling on the Community: Comment on Pizam and Pokela." *Annals of Tourism Research.* 13: 279-285.

Koch, Gary G. and Vasant P. Bhapkar. (1982) "Chi-Square Tests". In *Encyclopedia of Statistical Sciences* Vol. I, Norman L. Johnson and Samuel Kotz, eds. John Wiley & Sons: New York, p. 442-457.

Fienberg, S.E. (1982) "Contingency Tables". In *Encyclopedia of Statistical Sciences* Vol. II, Norman L. Johnson and Samuel Kotz, eds. John Wiley & Sons: New York, p. 161-171.

South Dakota Business Review. Business Research Bureau. University of South Dakota. September 1992, Vol. L, No.5. December 1992, Vol. LI, No.2.