Lab Assignment 11: Significance Testing

Use the *Movies.JMP* dataset to answer the following questions.

1. For the *World – Wide* variable, hypothesize an average value for the mean amount brought in by the top 216 movies of all time. Test whether the average is equal to your hypothesized value by selecting **Test Mean**. Go through each of your steps of hypothesis testing (ie, defining your hypotheses, stating your EV and SE, your test statistic, your p-value, and state conclusion with interpretation).

   This is going to vary greatly based on your own hypothesized value! The key to know is that your $H_0$ is that the mean of *World – Wide* = your hypothesized value. You can find the test statistic and p-value and make your conclusion based on whatever you chose for your value:)

2. Test for the difference in means of the continuous variable *Domestic* across types of movies. After selecting **Analyze Distribution**, choose type for the x-axis and *Domestic* for the y-axis. In difference of means tests, you should always place the nominal variable on the horizontal axis while the continuous variable goes on the vertical axis. Once your plot has shown up, select **Means/Anova/t-test** as one of your options. Go through all hypothesis testing steps as above.

   \[ H_0: \text{Average}_{\text{action}} = \text{Average}_{\text{comedy}} = \ldots = \text{Average}_{\text{drama}} \]
   \[ H_1: \text{Average}_{\text{action}} \neq \text{Average}_{\text{comedy}} \neq \ldots \neq \text{Average}_{\text{drama}} \]

   The test statistic is 2.84 with a p-value of 0.0254. Since 0.0254 < 0.05, then we *reject* $H_0$. So, we conclude that the mean amount movies made domestically is different across movie types. In other words, the difference is NOT due to chance, rather the differences in means are real. This conclusion does NOT tell us how important these differences are, however!

3. Create a new variable called *era*. I’ll explain this in class, but take notes because this will be valuable if your project dataset needs variables created! Test for independence between the type of movie and era released. Again, go through all steps of your hypothesis testing!

   Create a new *nominal* variable. Make sure you select **Formula** under **New Property**. Then, select conditional so that you can create an if statement. We want our 1st era to consist of movies released before 1980 (use the comparison selection to do this). Then, we need to select conditional again for our else statement so that we can set up 2 more eras (2 and 3–2 corresponding to those movies released between 1980 (inclusive) and 1990 (exclusive) ). Finally, we say our 2nd else statement is simply the 3rd era. Now that we’ve created this new *era* variable, we want to conduct a test of independence between era and type. After going to Fit Y by X (and selecting the above variables), we look at the Pearson ChiSquare test statistic of 19.43 and the corresponding p-value of 0.0127. Since this p-value is less that the 5% threshold, we *reject* $H_0$. Well, what was
our $H_0$: era is independent of movie type, while $H_1$: movie type is dependent of era. So, we conclude that era and movie type are dependent.

4. Now do a correlation significance test among all your continuous variables. Simply find your correlations as you normally would. Then, select **Pairwise Correlations** as one of your options. For this test, just state your hypotheses, your p-values, and which correlations are significant.

$H_0$: \( r = 0 \) (a correlation of 0 represents no association between variables); $H_1$: \( r \neq 0 \). Based on the p-values from the Pairwise Correlations, we see that there are 3 correlations that have p-values *greater* than .05 (meaning the corresponding variables are associated). Year Released and *Domestic* have a p-value of .0511, Year Released and *Foreign* have a p-value of .5537, and Year Released and *WorldWide* have a p-value of .7657. Thus, these are the only correlations that have significant associations.

5. Last test...let’s do a significance test of normality for the *Domestic* variable. Once you’ve got your histogram, overlay the normal distribution by selecting **Fit Distribution** → **Normal**, then select **Fitted Normal** → **Goodness of Fit**. State your hypotheses, test statistic, p-value, and conclusion.

$H_0$: *Domestic* is normally distributed, and $H_1$: *Domestic* is NOT normally distributed. The test statistic is .799 (you can round to .80), and the p-value is .0000. Based on this tiny p-value, we can reject $H_0$ and conclude that *Domestic* is NOT normally distributed.