Example 6: Full Regression Example
Jamear is working for the car show in Monterey and he is very curious about what factors affects the time it takes to drive 1/4 of a mile. The following are the variables of interests:

- Gross horsepower
- Weight (1000 lbs)
- 1/4 mile time

Two regressions models were fitted.

Model 1:

- Response Variable: 1/4 mile time
- Predictor Variable: Weight (1000 lbs)

```
Call:
lm(formula = qsec ~ wt, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3638 -1.0766  0.2051  0.8655  5.0298

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.8753     1.1025  17.120   <2e-16 ***
wt           -0.3191     0.3283  -0.972    0.339
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.789 on 30 degrees of freedom
Multiple R-squared:  0.03053,Adjusted R-squared:  -0.00179
F-statistic: 0.9446 on 1 and 30 DF,  p-value: 0.3389
```

$$y_i = 18.8753 - 0.3191x_i \tag{35}$$

What b_1 in this problem?

What is b_0 in this problem?

Interpretation of b_1: as the independent variable increases by 1 unit, the dependent variable increases/decreases (depends on sign of b_1) by b_1

Interpretation of -0.3191: as the weight of the car increases by 1 unit, the speed decreases by 0.3191

Step 1 and 3 of Hypothesis Test for β (two-tailed):

- $H_0 : \beta_1 = 0$ and $H_1 : \beta_1 \neq 0$
- $TS = t = b_1/SE(b1) = -0.3191/0.3283 = -0.9720$, $df = n - (k-1)$ (to be exact)
- The proportion of variance explained by the model is $R^2 = 0.03053$.
- The proportion of variance explained by the model accounting for sample size and number of predictors is $Adj.\ \ R^2 = -0.00179$.
- **Without comparing the TS with the CV what would you do to H_0?**

Model 2:

- Response Variable: 1/4 mile time
- Predictor Variable: Weight (1000 lbs)
- Predictor Variable: Gross horsepower

```
Call:
lm(formula = qsec ~ wt + hp, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8283 -0.4055 -0.1464  0.3519  3.7030

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 18.825585   0.671867  28.020  < 2e-16 ***
wt           0.941532   0.265897   3.541  0.00137 **
hp          -0.027310   0.003795  -7.197 6.36e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.09 on 29 degrees of freedom
Multiple R-squared:  0.652,Adjusted R-squared:  0.628
F-statistic: 27.17 on 2 and 29 DF,  p-value: 2.251e-07
```

$$y_i = 18.8256 + 0.9415x_{1i} + -0.0273x_{2i} \tag{36}$$

where x_{1i} represents weight and x_{2i} represents horse power.

The R^2 value can used to compare models.

Which model performed better?

12 Multinomial Distribution and Contingency Tables

"I don't need it to be easy. I just want it to be worth it." - Dwayne Michael Carter Jr.

12.1 Multinomial Distribution

Imagine we have 1 die and we roll it 120 times. These are our outcomes:

Random Event	Observed	Expected
Die Roll is 1	26	
Die Roll is 2	30	
Die Roll is 3	10	
Die Roll is 4	17	
Die Roll is 5	18	
Die Roll is 6	19	

In this experiment the following must be satisfied:

1. Fixed number of trials (rolls)
2. Trials are independent of each other
3. Each trial is one of the known outcomes
4. The probability of an event is consistent throughout each trail

Multinomial Experiments: Goodness-of-Fit - Determine whether a distribution of a sample data agrees with or fits some claimed distribution.

The claim in the die example is that each event has a 1/6 probability of occurring for each roll.

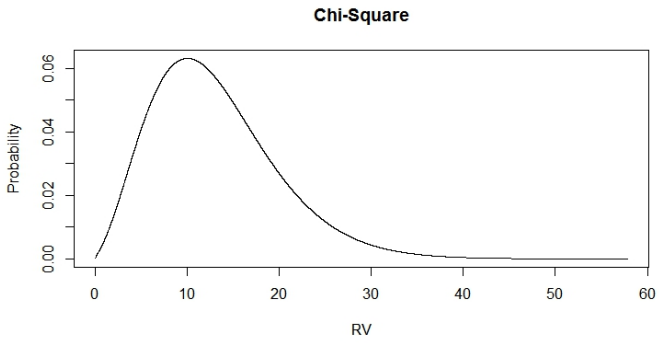
This means that of the 120 rolls, each side is expected so show up 20 times on average. Expectation can be seen as:

$$E = np$$

12.2 Hypothesis Testing for Goodness of Fit

Process for Hypothesis Testing for this class:

1. Identify and State the Statistical Question:
   - Determine the type variable(s) (i.e., quantitative or qualitative): qualitative
   - Identify and state the hypotheses (Null and Alternative Hypotheses) based on the question at hand $\boldsymbol{H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = p_6}$ and $\boldsymbol{H_1}$ **: At least one of the probabilities is different from the others**
2. Identify and state level of significance α (the probability of rejecting the H_0 when H_0 is true): **will be given to you, if not assume** $\alpha = 0.05$



   Really IMPORTANT:
   - α:
   - $df = k - 1$, k is number of categories
   - Critical Value:
3. Perform Statistical Test and Interpret Results

$$TS = \chi^2 = \sum \frac{(O - E)^2}{E} \tag{37}$$

   - O: observed frequency of outcome
   - E: expected frequency of outcome
   - k: number of different categories
   - n: number of trials
   - Test Statistic:
   - p-value:
4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms

Example 1: Dice Example

Random Event	O	E	$(O-E)$	$(O-E)^2$	$(O-E)^2/E$
Die Roll is 1	26	20	6	36	9/5
Die Roll is 2	30	20	10	100	5
Die Roll is 3	10	20	-10	100	5
Die Roll is 4	17	20	-3	9	9/20
Die Roll is 5	18	20	-2	4	1/5
Die Roll is 6	19	20	-1	1	1/20
Sum					12.5

$$TS = \chi^2 = 12.5$$

CV = Chi-Square table $k - 1 = 6 - 1 = 5$, $CV = 9.488$
Reject the H_0

12.3 Contingency Tables

Contingency tables are used to represent the distribution of two or more qualitative variables. The following is a contingency table for Helmet Usage and Type of Injury.

Table 4: My caption

	Helmet	No Helmet	Total
Facial Injuries	30	182	212
Other Injuries	83	236	319
Total	113	418	531

The question at hand can be seen as, is Type of Injury **independent** of Helmet usage? Before we do the Hypothesis test, lets talk about the observed and expected of each *bin*.

Table 5: My caption

	Level A	Level B	Total
Level 1	n_{1A}	n_{1B}	n_1
Level 2	n_{2A}	n_{2B}	n_2
Total	n_A	n_B	n_T

$$E_{1A} = \frac{n_A x n_1}{n_T} \qquad O_{1A} = n_{1A} \tag{38}$$

$$E_{2A} = \frac{n_A x n_2}{n_T} \qquad O_{2A} = n_{2A} \tag{39}$$

$$E_{1B} = \frac{n_B x n_1}{n_T} \qquad O_{1B} = n_{1B} \tag{40}$$

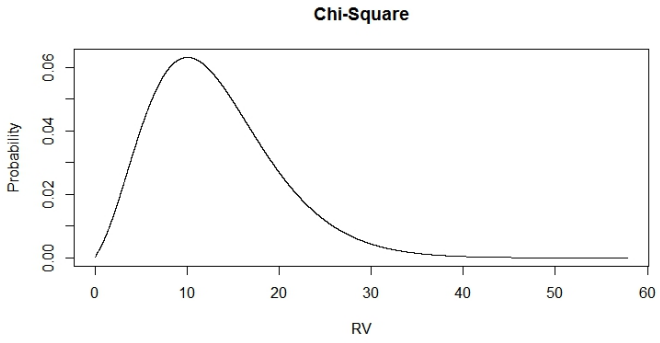
$$E_{2B} = \frac{n_B x n_2}{n_T} \qquad O_{2B} = n_{2B} \tag{41}$$

12.3.1 Hypothesis Testing: Test of Independence

12.4 Hypothesis Testing of Independence

Process for Hypothesis Testing for this class:

1. Identify and State the Statistical Question:
   - Determine the type variable(s) (i.e., quantitative or qualitative): qualitative
   - Identify and state the hypotheses (Null and Alternative Hypotheses) based on the question at hand $\boldsymbol{H_0}$ **: Row variable and column variable are independent** and $\boldsymbol{H_1}$ **: Row variable and column variable are dependent**
2. Identify and state level of significance α (the probability of rejecting the H_0 when H_0 is true): **will be given to you, if not assume** $\alpha = 0.05$



   Really IMPORTANT:
   - α:
   - $df = (r-1)(c-1)$, r is number of rows, c is the number of columns
   - Critical Value:
3. Perform Statistical Test and Interpret Results

$$TS = \chi^2 = \sum \frac{(O-E)^2}{E} \tag{42}$$

   - O: observed frequency of outcomes
   - E: expected frequency of outcomes
   - k: number of different categories
   - n: number of trials
   - Test Statistic:
   - p-value:
4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms

Example 2: Helmet Example

Random Event	O	E	$(O-E)$	$(O-E)^2$	$(O-E)^2/E$
Face Injury & Helmet					
Face Injury & No Helmet					
Face Injury & No Helmet					
Injury & No Helmet					
Sum					

$$TS = \chi^2 =$$

CV = Chi-Square table $(r-1)(c-1) =$, $CV =$
the H_0