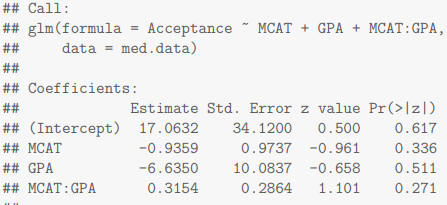
In one such study, 52 female-male pairs were measured and then observed mating. Does the size difference btwn the female and male spiders help explain whether or not cannibalism occurred? *In what proportion of the 52 matings did cannibalism occur?* This value is 0.2115. *Equation for predicting the log odds of cannibalism from the size difference btwn the female and male spiders.* Using the estimated intercept and slope from the output above, the equation for predicting the log odds of cannibalism from the size difference (xi) is: log(phat\_i / (1-phat\_i)) = -3.0890 + 3.0693xi *Equation for predicting the probability of cannibalism from the size difference btwn the female and male spiders.* For predicting the probability of cannibalism from the size difference (xi) is: phat\_i = (e-3.0890 + 3.0693x\_i)/(1 + e-3.0890 + 3.0693x\_i). *Interpret the slope and intercept of the logistic regression equation*. Slope: A 1mm increase in the size difference btwn the female and male spiders is associated with a e^3.0693 = 21.5268 times increase in the predicted odds of cannibalism. Intercept: For a 0mm size difference btwn the female and male spiders (they are the same size), the predicted odds of cannibalism is e^−3.0890 = 0.0455. *Find the predicted probability of cannibalism for a size difference btwn the female and male spiders of -0.2mm and 0.4mm.* For xi = −0.2: phat\_i = (e-3.0890 + 3.0693x\_i)/(1 + e-3.0890 + 3.0693x\_i) = 0.02406. For xi = 0.4: phat\_i = (e-3.0890 + 3.0693x\_i)/(1 + e-3.0890 + 3.0693x\_i) = 0.1346. *Find CIs for the probability of cannibalism for a size difference btwn the female and male spiders of 0mm and 0.8mm*. The CI is from 0.0089 to 0.1878. This means we have 95% confidence the probability of cannibalism in the population in matings that occur btwn the female and male spiders of the same size is btwn 0.0089 and 0.1878. The CI is from 0.1831 to 0.5568. This means we have 95% confidence the probability of cannibalism in the population in matings that occur btwn a female spider 0.8mm larger than the male spider is btwn 0.1831 and 0.5568. *Test for the statistical significance of the size difference btwn the female and male spiders in predicting the probability of cannibalism.* H0: B1 = 0, Ha: B1 != 0. The Wald test statistic comes from Coefficients table in the summary output of the model. The test statistic z = 3.057 with p-value = 0.0022. This means we have very strong evidence the size difference in the female and male spiders is associated with the probability of cannibalism in this population. The Likelihood ratio test statistic is 18.942 with a p-value < 0.0001. This means we have extremely strong evidence the size difference in the female and male spiders is associated with the probability of cannibalism in this population. *Calculate the pseudo R2 statistic for this logistic regression. Comment on its value.* A value of 0.3530 indicates a good model. *Conduct a goodness of fit test using the Hosmer-Lemeshow test statistic with the number of groups set to 5. Does this model appear to fit the data?* Null hypothesis: model is a good fit for the data. Alternative hypothesis: model is not a good fit for the data. Test statistic X2 = 3.0035 with a p-value = 0.3911. We conclude we have no evidence of lack of model fit. *Give the confusion table for the logistic regression model.* The figures indicate we correctly predicted 80.77% (Agreement) of the 52 observations. However, while we correctly predicted 95.12% (Specificity) cases when Cannibalism didn’t occur, we only correctly predicted 27.27% (Sensitivity) cases when Cannibalism did occur. *The area under this curve is:* This means that if I randomly select a case where Cannibalism occurred (success) and randomly select a case where Cannibalism didn’t occur (failure), the probability the success will have a higher predicted probability than the failure is 0.8869.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA computer code with numbers and symbols

Description automatically generated with medium confidenceData on 55 seniors who applied to medical school, all from the same liberal arts college in the Midwest. For each applicant, the variables GPA, MCAT, Gender, Apps, and Acceptance were collected. **Fit a logistic regression model for predicting the log odds of being accepted into medical school from the variables MCAT and GPA.** *Give the equation for predicting the log odds of acceptance from the variables MCAT and GPA.*  Log(phat\_i / 1-phat\_i) = −22.3727 + 0.1645 ∗ MCATi + 4.6765 ∗ GPAi. *Give the equation for predicting the probability of acceptance from the variables MCAT and GPA. Use this equation to predict the probability of acceptance for a student with a GPA of 3.54 and a MCAT score of 38.* The predicted probability of acceptance is: phat\_i = (e^(−22.3727 + 0.1645 ∗ MCATi + 4.6765 ∗ GPAi))/(1 + e^(−22.3727 + 0.1645 ∗ MCATi + 4.6765 ∗ GPAi)). when GPA = 3.54 and MCAT = 38, phat\_i = 0.6066. *Find and interpret a 95% CI for the probability of acceptance for a student with a GPA of 3.54 and a MCAT score of 38.* The CI is from 0.4159 to 0.7695. This means we are 95% confident a student from this population with a GPA of 3.54 and a MCAT of 38 has a probability of being accepted to medical school between 0.4159 and 0.7695. *Test for the significance of the overall model using the likelihood ratio test.* H0: B1 = B2 = 0. Ha: at least one Bj != 0. The test statistic is 21.777 with a p-value < 0.0001. We conclude we have extremely strong evidence that at least one of the explanatory variables is important in explaining the probability of acceptance to medical school in this population of students. *Test for the significance of GPA and MCAT scores separately using the Wald test.* The Wald test information is contained in the summary output of the model. For MCAT H0: B1 = 0. Ha: B1 != 0. Test Statistic: z = 1.595. p-value: 0.110786. Conclusion: We have little evidence that adding the MCAT score to the model that already contains GPA helps to explain the probability of acceptance to medical school in this population of students. For GPA H0: B2 = 0. Ha: B2 != 0. Test Statistic: z = 2.849. p-value: 0.004389. Conclusion: We have strong evidence that adding GPA to the model that already contains MCAT score helps to explain the probability of acceptance to medical school in this population of students. **Fit a logistic regression model for predicting the log odds of being accepted into medical school from the variables MCAT, GPA and Sex.** *Give the equation for predicting the log odds of acceptance from the variables MCAT and GPA for Females and the equation for predicting the log odds of acceptance from the variables MCAT and GPA for Males.* Using the output above and the coding Sex = 1 for Females and Sex = 0 for Males we have the two equations below. For females: log(phat\_i / 1 – phat\_i) = −23.9851 + 0.1809 ∗ MCATi + 5.1392 ∗ GPAi + 1.2580 ∗ (1) = −22.7271 + 0.1809 ∗ MCATi + 5.1392 ∗ GPAi. For males: log(phat\_i / 1-phat\_i) = −23.9851 + 0.1809 ∗ MCATi + 5.1392 ∗ GPAi + 1.2580 ∗ (0) = −23.9851 + 0.1809 ∗ MCATi + 5.1392 ∗ GPAi. Give the value of the coefficient for Sex in the model. *Calculate and interpret a 95% CI for the associated odds ratio.* The value of βˆ3 = 1.2580. The CI is from 0.8950 to 16.5025. This means we are 95% confident the odds of acceptance to medical school for females in this population of students is between 0.8950 to 16.5025 times the odds of acceptance to medical school for males in this population of students. *Test for the significance of the variable Sex in the model with GPA and MCAT using the Wald test.* H0: B3 = 0, Ha: B3 != 0. Test Statistic: z = 1.723. p-value: 0.084965. Conclusion: We have weak evidence that the variable Sex helps to explain the probability of acceptance to medical school for student in this population in the model that already contains MCAT and GPA. **Add an interaction term between GPA and MCAT to the logistic regression model.** *Give the equation for predicting the log odds of acceptance from the variables MCAT and GPA and their interaction.* log(phat\_i / 1 – phat\_i) = 17.0632−0.9359∗MCATi−6.6350∗GPAi + 0.3154∗MCATi∗GPAi. *Give the equation for predicting the probability of acceptance from the variables MCAT and GPA and their interaction.* phat\_i = (e^(17.0632 − 0.9359∗MCATi − 6.6350∗GPAi + 0.3154∗MCATi∗GPAi)) / (1+ e^(17.0632 − 0.9359∗MCATi − 6.6350∗GPAi + 0.3154∗MCATi∗GPAi)) *Find and interpret a 95% CI for the probability of acceptance for a student with a GPA of 3.54 and a MCAT score of 38.* The CI is from 0.4054 to 0.7780. This means we are 95% confident a student from this population with a GPA of 3.54 and a MCAT of 38 has a probability of being accepted to medical school between 0.4054 and 0.7780. The two confidence intervals are very similar. *Test for the significance of the interaction term in this model using the Wald test.* H0: B3 = 0. Ha: B3 != 0. Test Statistic: z = 1.101. p-value: 0.271. Conclusion: We have little to no evidence the interaction between MCAT and GPA helps to explain the probability of acceptance to medical school in this population of students.

A screenshot of a computer

Description automatically generatedA table with numbers and text

Description automatically generatedIn the 1998 General Social Survey, respondents were classified according to their sex, race, and belief in an afterlife. Below are the four conditional logit values for undecided versus no, one for each of the four combinations of sex and race. male.white.U<- -0.652994 - 0.1050684 + 0.2710218, female.white.U<- -0.652994 + 0.2710218, male.black.U<- -0.652994 - 0.1050684, female.black.U<- -0.652994. Below are the four conditional logit values for yes versus no, one for each of the four combinations of sex and race. male.white.Y<- 1.301614 - 0.4185635 + 0.3417684, female.white.Y<- 1.301614 + 0.3417684, male.black.Y<- 1.301614 - 0.4185635, female.black.Y<- 1.301614. *Use the model from part (a) to estimate the probabilities for each of the three categories for belief in afterlife for all four combinations of sex and race.* For white males: #Yes: exp(male.white.Y)/(1 + exp(male.white.Y) + exp(male.white.U)) ## [1] 0.6782693, #Undecided: exp(male.white.U)/(1 + exp(male.white.Y) + exp(male.white.U)) ## [1] 0.1224478, #No: 1/(1 + exp(male.white.Y) + exp(male.white.U)) ## [1] 0.1992829. For white females: #Yes: exp(female.white.Y)/(1 + exp(female.white.Y) + exp(female.white.U)) ## [1] 0.754562, #Undecided: exp(female.white.U)/(1 + exp(female.white.Y) + exp(female.white.U)) ## [1] 0.09956223, #No: 1/(1 + exp(female.white.Y) + exp(female.white.U)) ## [1] 0.1458757. For black males: #Yes: exp(male.black.Y)/(1 + exp(male.black.Y) + exp(male.black.U)) ## [1] 0.6221676, #Undecided: exp(male.black.U)/(1 + exp(male.black.Y) + exp(male.black.U)) ## [1] 0.1205539, #No: 1/(1 + exp(male.black.Y) + exp(male.black.U)) ## [1] 0.2572785. For black females: #Yes: exp(female.black.Y)/(1 + exp(female.black.Y) + exp(female.black.U)) ## [1] 0.7073575, #Undecided: exp(female.black.U)/(1 + exp(female.black.Y) + exp(female.black.U)) ## [1] 0.100176, #No: 1/(1 + exp(female.black.Y) + exp(female.black.U))

## [1] 0.1924665. *Test for the significance of sex in the model for predicting belief in the afterlife.* The test statistic is 7.1926 and the p-value is 0.0274. We will conclude we have moderate evidence a person’s sex helps to explain their belief in afterlife in the model that includes Race. *Test for the significance of race in the model for predicting belief in the afterlife.* The test statistic is 1.9942 and p-value is 0.3689. We will conclude we have no evidence a person’s Race helps to explain their belief in afterlife in the model that includes Sex.

A table with text on it

Description automatically generatedThe following table of counts was obtained from a random sample of 1,397 respondents from the population of adults (more then 18 years old) in the United States in 1982. *Fit the log linear model for the counts in the contingency table under the assumption of independence. Set the baseline categories for both variables to Oppose. Use the estimated coefficients to calculate the four expected counts in the table. Does this model fit the data?* pchisq(gpI.model$deviance, 1, lower.tail = F) ## [1] 0.02107415. We have moderately strong evidence the independence model does not fit the data. *Fit the saturated log linear model for the counts in the contingency table, again using the category Oppose as the baseline category. Use the model to obtain the odds ratio for the table and interpret its value.* exp(gpS.model$coefficients[4])

A table with numbers and words

Description automatically generated## Gun.RegisterFavor:Death.PenaltyFavor ## 0.7049975. Since the value of the predicted odds ratio is less than 1, interpret the reciprocal value. The predicted odds a person was in favor of the death penalty if they opposed gun registration as a part of gun control is 1/0.7050 = 1.4184 times the predicted odds a person was in favor of the death penalty if they were in favor of gun registration as a part of gun control.

The three-way cross-classification table for the three variables, Wet, Lather, and Rinse is given below. When Rinse is Toward: exp(showerS.model$coefficient[5]) ## WetAway:LatherAway ## 2.064243

For a student who faces toward the shower when they rinse their hair, the predicted odds they face away from the shower when they wet their hair if they also face away from the shower when they lather their hair is 2.0642 times the predicted odds they face Away from the shower when they wet

their hair if they face toward the shower when they lather their hair.