Part I

1. a
2. b
3. b
4. d
5. a
6. c
7. b
8. d
9. d
10. b

Part II

1. c If X and Z are independent and their means are 0 then $E(Y|X)$ is still β_1
2. c The bias in K-fold CV comes from the training data being different from the full data set.

Part III

1. (a) Take derivative of $R(h)$, set to 0, and solve for h to get $\alpha = 1/5$. Plug in to get $\beta = 4/5$. Alternatively, set the two terms with h equal to each other and solve for $\alpha = 1/5$, then plug in as above

   (b) The variance goes down like one over the number of data values being averaged. Because the number of data points in a ball of radius h in p-dimensional space is on the order of $n * h^p$, $\gamma = p$

   c,d See Figure 4.1 in Shalizi (page 98) and Figure 2 of the lecture in week 8 when we revisited kernel regression

2. (a) Resample cases. Nonconstant residual variance indicates that the residuals are not iid, so neither residual nor parametric bootstraps are appropriate.

   (b) Let $i*$ be a sample of size n with replacement from $\{1, ..., n\}$. Use these cases as your bootstrap sample: $\{(X_i, Y_i) : i \in i*\}$.

   (c) H_0: this linear model fits well. H_1: a larger model (e.g. something nonparametric) fits better. Test statistic should be something like the difference in residual sum of squares between the linear model ($RSS0$) and a nonparametric model ($RSS1$) that generalizes the linear model. Call the difference $DRSS = RSS0 - RSS1$. Bootstrap parametrically from the fitted linear model and fit both the linear and nonparametric models each time. Compute the $DRSS*$ each time. Use the empirical distribution of $DRSS*$'s. Reject if real-data $DRSS$ is in upper α tail of $DRSS*$'s

3. (a) Treatment raises odds of improvement by a factor of $\exp(1.8)$, relative to Placebo's odds of improvement. 95% CI is $[\exp(0.75), \exp(2.9)]$.

   (b) Being Male reduces odds of improvement by a factor of $\exp(-1.5)$, relative to Female's odds of improvement. 95% CI is $[\exp(-2.7), \exp(-0.37)]$.

   (c) Yes, because the patients were assigned to treatments randomly

   (d) A 95% CI for the increase in odds of improvement associated with a one-year increase in age among Females is $[\exp(0.010), \exp(0.92)]$.

   (e) H_0: this GLM fits well. H_A: a larger model (e.g. saturated model) fits better. Test statistic should be residual deviance (92). (This is equivalent to the difference between residual and saturated deviances, since the saturated model's deviance is 0.) Under the null it is approximately chi-squared with residual df (80). Reject the null at level $\alpha = 0.05$ if 92 is in the upper alpha tail of the χ^2_{80} distribution.

4. In the HW, the response was $Y = \log(\text{death counts})$. We fit a GAM with Normal$(0, \sigma^2)$ errors and mean function

$$E(Y|X) = \beta_0 + \sum_{j=1}^{p} g_j(X_j).$$

So Y Normal$(\beta_0 + \sum_{j=1}^{p} g_j(X_j), \sigma^2)$ In a Poisson-response GAM, we would instead use the response Y = death counts and a log link function, with the same structure of mean function for log of death counts. Now the response is Y Poisson$(exp(\beta_0 + \sum_{j=1}^{p} g_j(X_j)))$. In other words, both have the same structure for conditional expectation of log of death counts, but they use different probability models for the variability of the response around that conditional mean function.