ISYE 6414 Midterm Prep

Terms in this set (74)

We can assess the constant variance assumption in linear regression by plotting the residuals vs. fitted values.	True
If one confidence interval in the pairwise comparison in ANOVA includes zero, we conclude that the two corresponding means are plausibly equal.	True
The assumption of normality is not required in linear regression to make inference on the regression coefficients.	False (Explanation: is required)
We cannot estimate a multiple linear regression model if the predicting variables are linearly independent.	False (Explanation: linearly dependent)

ISYE 6414 Midterm Prep

If a predicting variable is a categorical variable with 5 categories in a linear regression model without intercept, we will include 5 dummy variables.	True
If the normality assumption does not hold for a regression, we may use a transformation on the response variable.	True
The prediction of the response variable has higher uncertainty than the estimation of the mean response.	True
Statistical inference for linear regression under normality relies on large sample size.	False (Explanation: small sample size is fine)
A nonlinear relationship between the response variable and a predicting variable cannot be modeled using regression.	False (Explanation: Nonlinear relationships can often be modeled using linear regression by including polynomial terms of the predicting variable, for example.)
Assumption of normality in linear regression is required for confidence intervals, prediction intervals, and hypothesis testing.	True
If the confidence interval for a regression coefficient contains the value zero, we interpret that the regression coefficient is plausibly equal to zero.	True

ISYE 6414 Midterm Prep

The smaller the coefficient of determination or R-squared, the higher the variability explained by the simple linear regression.	False (Explanation: The larger the R-squared)
The estimators of the variance parameter and of the regression coefficients in a regression model are random variables.	True
The standard error in linear regression indicates how far the data points are from the regression line, on average.	True
A linear regression model is a good fit to the data set if the R-squared is above 0.90.	False (Explanation: There are other things to check: assumptions, MSE, etc.)
In ANOVA, we assume the variance of the response variable is different for each population.	False (Explanation: is the same across all populations)
The F-test in ANOVA compares the between variability versus the within variability.	True
In testing for subsets of coefficients in a multiple linear regression, the null hypothesis we test for is that all coefficients are equal; $H_0: B_1 = B_2 = \ldots = B_{kf}$	False (Explanation: The null hypothesis is that all coefficients are equal to zero; none are significant in predicting the response.)
The only assumptions for a simple linear regression model are linearity, constant variance, and normality	False

ISYE 6414 Midterm Prep

In a simple linear regression model, the variable of interest is the response variable.	True
The constant variance assumption is diagnosed by plotting the predicting variable vs. the response variable.	False
β_1 is an unbiased estimator for β_0.	False
The estimator σ^2 is a fixed variable.	False
The ANOVA model with a qualitative predicting variable with k levels/classes will have k + 1 parameters to estimate.	True
Under the normality assumption, the estimator for β_1 is a linear combination of normally distributed random variables.	True
A negative value of β_1 is consistent with an inverse relationship between x and y.	True
In the simple linear regression model, we lose three degrees of freedom because of the estimation of the three model parameters β_0, β_1, σ_2.	False
The regression coefficient is used to measure the	False

If the constant variance assumption in ANOVA does not hold, the inference on the equality of the means will not be reliable.	True
If one confidence interval in the pairwise comparison does not include zero, we conclude that the two means are plausibly equal.	False
The mean sum of square errors in ANOVA measures variability within groups.	True
Only the log-transformation of the response variable should be used when the normality assumption does not hold.	False
If one confidence interval in the pairwise comparison includes only positive values, we conclude that the difference in means is positive, and statistically significant.	True
The number of degrees of freedom of the χ^2 (chi-square) distribution for the pooled variance estimator is $N - k + 1$ where k is the number of samples.	False
For assessing the normality assumption of the ANOVA model, we can use the quantile-quantile normal plot and the historgram of the residuals.	True

ISYE 6414 Midterm Prep

One-way ANOVA is a linear regression model with more than one qualitative predicting variables.	False
The sampling distribution for the variance estimator in ANOVA is χ^2 (chi-square) with N - k degrees of freedom.	False
In simple linear regression, we can diagnose the assumption of constant-variance by plotting the residuals against fitted values.	True
If response variable Y has a quadratic relationship with a predictor variable X, it is possible to model the relationship using multiple linear regression.	True
The R^2 value represents the percentage of variability in the response that can be explained by the linear regression on the predictors. Models with higher R^2 are always preferred over models with lower R^2 .	False
For the model $y = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$, there are p+1 parameters to be estimated	False
The F-test can be used to evaluate the relationship between two qualitative variables.	False

ISYE 6414 Midterm Prep

The Partial F-Test can test whether a subset of regression coefficients are all equal to zero.	True
In multiple linear regression, controlling variables are used to control for sample bias.	True
In a multiple regression model with 7 predicting variables, the sampling distribution of the estimated variance of the error terms is a chi-squared distribution with n-8 degrees of freedom.	True
There are four assumptions needed for estimation with multiple linear regression: mean zero, constant variance, independence, and normality.	False (why?)
Let $\hat{Y}^*$ be the predicted response at x^*. The variance of $\hat{Y}^*$ given x^* depends on both the value of x^* and the design matrix.	True (but the wording was confusing, so everyone got credit no matter what on this question)
Suppose x1 was not found to be significant in the model specified with lm(y ~ x1 + x2 + x3). Then x1 will also not be significant in the model lm(y ~ x1 + x2).	False
When estimating confidence values for the mean response for all instances of the predicting variables, we should use a critical point based on the F-distribution to correct for the simultaneous inference.	True

ISYE 6414 Midterm Prep

For estimating confidence intervals for the regression coefficients, the sampling distribution used is a normal distribution.	False
In a multiple linear regression model with quantitative predictors, the coefficient corresponding to one predictor is interpreted as the estimated expected change in the response variable when there is a one unit change in that predictor.	False
It is possible to produce a model where the overall F-statistic is significant but all the regression coefficients have insignificant t-statistics.	True. (explanation: This can happen when you have multicollinearity in two or more of the predictors. In that case, you have an overall model which is significant but on the level of individual predictor, they might not be since either of the collinear features could be included.)
Analysis of Variance (ANOVA) is an example of a multiple regression model.	True
For a multiple regression model, both the true errors ϵ and the estimated residuals ϵ-hat have a constant mean and a constant variance.	False
If the p-value of the overall F-test is close to 0, we can conclude all the predicting variable coefficients are significantly nonzero.	False

ISYE 6414 Midterm Prep

The causation effect of a predicting variable to the response variable can be captured using multiple linear regression, conditional of other predicting variables in the model.	False
A high Cook's distance for a particular observation suggests that the observation could be an influential point.	True
A no-intercept model with one qualitative predicting variable with 3 levels will use 3 dummy variables.	True
If the confidence interval for a regression coefficient contains the value zero, we interpret that the regression coefficient is definitely equal to zero.	False. The coefficient is plausibly zero, but we cannot be certain that it is
The larger the coefficient of determination or R-squared, the higher the variability explained by the simple linear regression model.	True. R-squared is the proportion of variability explained by the model.
The estimators of the error term variance and of the regression coefficients are random variables.	True. The estimators are $\hat{\beta}$, $\hat{\sigma}^2$, and $\hat{\varepsilon}$. These estimators are functions of the response, which is a random variable. Therefore they are also random.
The one-way ANOVA is a linear regression model with one qualitative predicting variable.	True. One-way ANOVA uses one qualitative variable, and it can be treated as a linear regression model (see Unit 2.2.3).

ISYE 6414 Midterm Prep

We can assess the assumption of constant-variance in multiple linear regression by plotting the standardized residuals against fitted values.	True. See Unit 3.3.1.
If one confidence interval in the pairwise comparison includes zero under ANOVA, we conclude that the two corresponding means are plausibly equal.	True. See Unit 2.2.1
We do not need to assume normality of the response variable for making inference on the regression coefficients.	False. The sampling distributions we use for inference rely on normality of the response.
Assuming the model is a good fit, the residuals in simple linear regression have constant variance	True. Goodness of fit refers to whether the model assumptions hold, one of which is constant variance.
We cannot estimate a multiple linear regression model if the predicting variables are linearly independent.	False. We cannot if they are linearly dependent.
If a predicting variable is categorical with 5 categories in a linear regression model without intercept, we will include 5 dummy variables in the model.	True. See Unit 2.2.3
In the ANOVA, the number of degrees of freedom of the chi-squared distribution for the variance estimator is N-k-1 where k is the number of groups.	False. This variance estimator has N-1 degrees of freedom.

ISYE 6414 Midterm Prep

If the non-constant variance assumption does not hold in multiple linear regression, we apply a transformation to the predicting variables.	False. We apply a transformation on the response.
The prediction of the response variable has higher uncertainty than the estimation of the mean response.	True. We have additional uncertainty from the newness of the observation (see Unit 3.2.4).
In linear regression, outliers do not impact the estimation of the regression coefficients.	False. Outliers can impact estimation, especially if they are also influential points.
Multicolinearity in multiple linear regression means that the columns in the design matrix are (nearly) linearly dependent.	True. See Unit 3.3.3
The statistical inference for linear regression under normality relies on large size of sample data.	False. As we are already assuming normality, we do not need to rely on a large sample size.