# ISYE6414 2019 Summer Midterm

Get access to all your stats, your personal progress dashboard and smart study shortcuts with Quizlet Plus. Unlock Progress

## Terms in this set (166)

| | |
|---|---|
| The prediction intervals need to be corrected for simultaneous inference when multiple predictions are made jointly. | ... |
| The estimated versus predicted regression line for a given x* | ... |
| 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 larger the coefficient of determination or R-squared, the higher the variability explained by the simple linear regression model. | True |
| The estimators of the error term variance and of the | True |

ISYE6414 2019 Summer Midterm

| | |
|---|---|
| The one-way ANOVA is a linear regression model with one qualitative predicting variable. | True |
| We can assess the assumption of constant-variance in multiple linear regression by plotting the standardized residuals against fitted values. | True |
| If one confidence interval in the pairwise comparison includes zero under ANOVA, we conclude that the two corresponding means are plausibly equal. | True |
| We do not need to assume normality of the response variable for making inference on the regression coefficients. | False |
| Assuming the model is a good fit, the residuals in simple linear regression have constant variance. | True |
| We cannot estimate a multiple linear regression model if the predicting variables are linearly independent. | False |
| 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 |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| 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 |
| The prediction of the response variable has higher uncertainty than the estimation of the mean response. | True |
| In linear regression, outliers do not impact the estimation of the regression coefficients. | False |
| Multicolinearity in multiple linear regression means that the columns in the design matrix are (nearly) linearly dependent. | True |
| The statistical inference for linear regression under normality relies on large size of sample data. | False |
| If the non-constant variance assumption does not hold in multiple linear regression, we apply a transformation to the predicting variables. | False |
| The only assumptions for a linear regression model are linearity, constant variance, and normality. | False |
| In the regression model, the variable of interest for study is the predicting variable. | False |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| The constant variance is diagnosted using the quantile-quantile normal plot. | False |
| $\beta_1$ is an unbiased estimator for $\beta_0$ . | False |
| The estimator $\sigma$ ^ 2 is a fixed variable. | False |
| The linear regression 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 $\beta_1$ is a linear combination of normally distributed random variables. | True |
| A negative value of $\beta_1$ 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 $\beta_0$ , $\beta_1$ , $\sigma^2$ . | False |
| The regression coefficient is used to measure the linear dependence between two variables. | False |
| If the constant variance assumption in ANOVA does not hold, the inference on the equality of the means | 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 can 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 statistically significantly positive. | True |
| The number of degrees of freedom of the $\chi$ 2 (chi-square) distribution for the 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 |
| The ANOVA is a linear regression model with one or more qualitative predicting variables. | True |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| The sampling distribution for the variance estimator in ANOVA is χ 2 (chi-square) regardless of the assumptions of the data. | False |
| We assess the assumption of constant-variance by plotting the residuals against fitted values. | True |
| Prediction is the only objective of multiple linear regression. | False |
| The number of parameters to estimate in the case of a multiple linear regression model containing 5 predicting variables and no intercept is 6. | True |
| The equation to find the estimated variance of the error terms can be obtained by summing up the squared residuals and dividing that by n - p - 1, where n is the sample size and p is the number of predictors. | True |
| The regression coefficient corresponding to one predictor is interpreted in a multiple regression in terms of the estimated expected change in the response variable when there is a change of one unit in the corresponding predicting variable. | False |
| In case of multiple linear regression, controlling variables are used to control for sample bias. | True |

| | |
|---|---|
| Observational studies allow us to make causal inference. | False |
| For a given predicting variable, the estimated coefficient of regression associated with it will likely be different in a model with other predicting variables or in the model with only the predicting variable alone. | True |
| For estimating confidence intervals for the regression coefficients, the sampling distribution used is a normal distribution. | False |
| The regression coefficients that are estimated serve as unbiased estimators. | True |
| For testing if a regression coefficient is zero, the normal test can be used. | False |
| Studying the relationship between a single response variable and more than one predicting quantitative and/or qualitative variable is termed as Multiple linear regression. | True |
| Before making statistical inference on regression coefficients, estimation of the variance of the error terms is necessary. | True |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| Partial F-Test can also be defined as the hypothesis test for the scenario where a subset of regression coefficients are all equal to zero. | True |
| The causation of a predicting variable to the response variable can be captured using Multiple linear regression, conditional of other predicting variables in the model. | False |
| For a linearly dependent set of predictor variables, we should not estimate a multiple linear regression model. | True |
| An example of a multiple regression model is Analysis of Variance (ANOVA). | True |
| Given a categorial predictor with 4 categories in a linear regression model with intercept, 4 dummy variables need to be included in the model. | False |
| Assuming that the data are normally distributed, under the simple linear model, the estimated variance has the following sampling distribution: | Chi-square with n-2 degrees of freedom |
| The fitted values are defined as: | The regression line with parameters replaced with the estimated regression coefficients. |
| The estimators of the linear regression model are | Minimizing the sum of squared differences between observed and expected values |

| | |
|---|---|
| The estimators for the regression coefficients are: | Unbiased regardless of the distribution of the data. |
| The assumption of normality: | It is needed for the sampling distribution of the estimators of the regression coefficients and hence for inference. |
| The estimated versus predicted regression line for a given x*: | Have the same expectation |
| The variability in the prediction comes from: | The variability due to a new measurement and due to estimation. |
| We detect departure from the assumption of constant variance | When the residuals vs fitted values are larger in the ends but smaller in the middle. |
| The pooled variance estimator is: | The sample variance estimator assuming equal variances. |
| The total sum of squares divided by N-1 is | The sample variance estimator assuming equal means and equal variances |
| The mean squared errors (MSE) measures: | The within-treatment variability. |
| The objective of the residual analysis is | To evaluate departures from the model assumptions |
| The objective of the pairwise comparison is | To identify the statistically significantly different means. |
| The objective of multiple linear regression is | 1. To predict future new responses<br>2. To model the association of explanatory variables to a response variable accounting for controlling factors.<br>3. To test hypothesis using statistical inference on the model. |

| | |
|---|---|
| A multiple linear regression model with p predicting variables but no intercept has p model parameters. | False |
| The interpretation of the regression coefficients is the same whether or not interaction terms are included in the model. | False |
| Multiple linear regression is a general model encompassing both ANOVA and simple linear regression. | True |
| The regression coefficients can be estimated only if the predicting variables are not linearly dependent. | True |
| The estimated regression coefficient \beta^hat_i is interpreted as the change in the response variable associated with one unit of change in the i-th predicting variable . | False |
| The estimated regression coefficients will be the same under marginal and conditional model, only their interpretation is not. | False |
| Causality is the same as association in interpreting the relationship between the response and the predicting variables. | False |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| The estimated variance of the error term has a \chi^2 distribution regardless of the distribution assumption of the error terms. | False |
| The number of degrees of freedom for the \chi^2 distribution of the estimated variance is n-p-1 for a model without intercept. | False |
| The sampling distribution of the mean squared error is different of that of the estimated variance. | False |
| The sampling distribution of the estimated regression coefficients is centered at the true regression parameters. | True |
| The sampling distribution of the estimated regression coefficients is the t-distribution assuming that the variance of the error term is unknown an replaced by its estimate. | True |
| The sampling distribution of the estimated regression coefficients is dependent on the design matrix. | True |
| We can test for a subset of regression coefficients using the F statistic test of the overall regression. | False |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| We can test for a subset of regression coefficients only if we are interested whether additional explanatory variables should be considered in addition to the controlling variables. | False |
| We can test for a subset of regression coefficients to evaluate whether all regression coefficients corresponding to the predicting variables excluded from the reduced model are statistically significant. | False |
| The prediction intervals need to be corrected for simultaneous inference when multiple predictions are made jointly. | True |
| The prediction intervals are centered at the predicted value. | True |
| The sampling distribution of the prediction of a new response is a t-distribution. | True |
| In evaluating a multiple linear model the F test is used to evaluate the overall regression. | True |
| In evaluating a multiple linear model the coefficient of variation is interpreted as the percentage of variability in the response variable explained by the | True |

| | |
|---|---|
| In evaluating a multiple linear model residual analysis is used for goodness of fit assessment. | True |
| In the presence of near multicollinearity, the coefficient of variation decreases. | False |
| In the presence of near multicollinearity, the regression coefficients will tend to be identified as statistically significant even if they are not. | False |
| If the linearity assumption with respect to one or more predictors does not hold, then we use transformations of the corresponding predictors to improve on this assumption. | True |
| If the normality assumption does not hold, we transform the response variable, commonly using the Box-Cox transformation. | True |
| If the constant variance assumption does not hold, we transform the response variable. | True |
| The residuals have constant variance for the multiple linear regression model. | False |
| The residuals vs fitted can be used to assess the assumption of independence. | False |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| The residuals have a t-distribution distribution if the error term is assumed to have a normal distribution. | False |
| Independence assumption can be assessed using the residuals vs fitted values. | False |
| Independence assumption can be assessed using the normal probability plot. | False |
| Residual analysis can only be used to assess uncorrelated errors. | True |
| If a departure from normality is detected, we transform the predicting variable to improve upon the normality assumption. | False |
| If a departure from the independence assumption is detected, we transform the response variable to improve upon this assumption. | False |
| The Box-Cox transformation is commonly used to improve upon the linearity assumption. | False |
| In evaluating a simple linear model there is a direct relationship between coefficient of variation and the correlation between the predicting and response variables. | True |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| In evaluating a simple linear model the coefficient of variation is interpreted as the percentage of variability in the response variable explained by the model. | True |
| In evaluating a simple linear model residual analysis is used for goodness of fit assessment. | True |
| The means of the k populations is a model parameter in ANOVA. | False |
| The sample means of the k populations is a model parameter in ANOVA. | False |
| The sample means of the k samples is a model parameter in ANOVA. | False |
| If we reject the test of equal means, we conclude that all treatment means are not equal. | False |
| If we do not reject the test of equal means, we conclude that means are definitely all equal | False |
| If we reject the test of equal means, we conclude that some treatment means are not equal. | True |
| Residual analysis can only be used to assess | True |

| | |
|---|---|
| The variability in the prediction comes from | The variability due to a new measurement and due to estimation. |
| The estimated versus predicted regression line for a given x* | Have the same expectation |
| The fitted values are defined as | The regression line with parameters replaced with the estimated regression coefficients. |
| In evaluating a multiple linear model, the F test is used to evaluate the overall regression. | True |
| In evaluating a multiple linear model, the coefficient of variation is interpreted as the percentage of variability in the response variable explained by the model. | True |
| In evaluating a multiple linear model, Residual analysis is used for goodness of fit assessment. | True |
| In the presence of near multicollinearity, the coefficient of variation decreases. | False |
| In the presence of near multicollinearity, the regression coefficients will tend to be identified as statistically significant even if they are not. | False |
| In the presence of near multicollinearity, the | False |

| | |
|---|---|
| When do we use transformations? | If the linearity assumption with respect to one or more predictors does not hold, then we use transformations of the corresponding predictors to improve on this assumption.<br>If the normality assumption does not hold, we transform the response variable, commonly using the Box-Cox transformation.<br>If the constant variance assumption does not hold, we transform the response variable. |
| The sampling distribution of the estimated regression coefficients is Centered at the true regression parameters. | True |
| The sampling distribution of the estimated regression coefficients is the t-distribution assuming that the variance of the error term is unknown an replaced by its estimate. | True |
| The estimators for the regression coefficients are Uunbiased regardless of the distribution of the data. correct | True |
| Multiple linear regression is a general model encompassing both ANOVA and simple linear regression. correct | True |
| The regression coefficients can be estimated only if the predicting variables are not linearly dependent. | True |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| 1. The means of the k populations<br>2. The sample means of the k populations<br>3. The sample means of the k samples<br>are NOT all the model parameters in ANOVA | True |
| The only assumptions for a simple linear regression model are linearity, constant variance, and normality. | False |
| 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 |
| LaTeX: \beta_1 β 1 is an unbiased estimator for LaTeX: \beta_0 β 0 . | False |
| The estimator LaTeX: \hat \sigma^2 σ ^ 2 is a fixed variable. | False |
| The ANOVA model with a qualitative predicting variable with LaTeX: k k levels/classes will have LaTeX: k+1 k + 1 parameters to estimate. | True |
| Under the normality assumption, the estimator for LaTeX: \beta_1 β 1 is a linear combination of normally | True |

| | |
|---|---|
| A negative value of LaTeX: \beta_1 β 1 is consistent with an inverse relationship between LaTeX: x x and LaTeX: y y . | True |
| In the simple linear regression model, we lose three degrees of freedom because of the estimation of the three model parameters LaTeX: \beta_0,\:\beta_1,\sigma^2 β 0 , β 1 , σ 2 . | False |
| If the constant variance assumption in ANOVA does not hold, the inference on the equality of the means will not be reliable. | 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 LaTeX: \chi^2 χ 2 (chi-square) distribution for the pooled variance estimator is LaTeX: N-k+1 N - k + 1 where LaTeX: k k is the number of samples. | False |

| | |
|---|---|
| The sampling distribution for the variance estimator in ANOVA is LaTeX: \chi^2 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 LaTeX: R^2 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 LaTeX: R^2 R 2 are always preferred over models with lower LaTeX: R^2 R 2 . | True |
| For the model LaTeX: y=\beta_0+\beta_1x_1+...+\beta_px_p+\epsilon y = β 0 + β 1 x 1 + ... + β p x p + ϵ , where LaTeX: \epsilon\sim N(0,\sigma^2) ϵ ~ N ( 0 , σ 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 |
| The Partial F-Test can test whether a subset of | 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 |
| Let LaTeX: Y^ **Y ∗ be the predicted response at LaTeX: x^** x ∗ . The variance of LaTeX: Y^ **Y ∗ given LaTeX: x^** x ∗ depends on both the value of LaTeX: x^* x ∗ and the design matrix. | True |
| 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 |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| 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 |
| Analysis of Variance (ANOVA) is an example of a multiple regression model. | True |
| For a multiple regression model, both the true errors LaTeX: \epsilon ϵ and the estimated residuals LaTeX: \hat \epsilon ϵ ^ 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 |
| 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 |

## ISYE6414 2019 Summer Midterm

| | |
|---|---|
| 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 |